



---

# Quality Assurance Handbook: For JISC Digital Library Programmes

---

*This handbook provides advice and support for projects funded by JISC's digital library programmes. The handbook provides advice for projects in their choice of standards and best practices for their technical infrastructure. The handbook provides a quality assurance methodology which will help to ensure that projects funded by JISC's digital library programmes are interoperable and widely accessible.*

## Document details

Editor :	Brian Kelly
Contributors	Brian Kelly, Marieke Guy (UKOLN), Ed Bremner (ILRT), Hamish James and Gareth Knight (AHDS)
Date:	22 May 2004
Version:	1.1
Notes:	Minor updates made on 6 December 2012.

---

# Table Of Contents

---

1	Introduction	1
2	About The Handbook	3
3	Background to the QA Focus Work	3
4	QA Focus Technical Advice	5
4.1	Standards	6
4.2	Digitisation	19
4.3	Access / Web	81
4.4	Metadata	148
4.5	Software	159
4.6	General	176
4.7	Quality Assurance	196
4.8	Service Deployment	201
5	The QA Focus Toolkit	236
 <b>Appendices</b>		
1	Publications	245

# 1 Introduction

---

## Background

Welcome to QA Focus's "*Quality Assurance Handbook*". This handbook has been published by the JISC-funded QA Focus project. The handbook provides advice on compliance with the open standards and best practices which projects funded by JISC's digital library programmes should seek to implement.

## QA Focus Work

QA Focus was funded by the JISC to help develop quality assurance methodology which projects funded by JISC's digital library programmes should seek to implement in order to ensure that project deliverables comply with appropriate standards and best practices which. This will help to ensure that project deliverables are widely accessible and interoperable and to facilitate the deployment of deliverables into a service environment.

The approach taken by QA Focus has been *developmental*: rather than seeking to impose requirements on projects, which are being undertaken by many institutions across the country, with differing backgrounds and levels of funding and resources, we have sought to raise an awareness of JISC's commitment to use of open standards, to describe various technical frameworks which can help in deploying open standards and to outline ways of ensuring that selected standards are used in a compliance fashion.

We do, however, recognise the difficulties which projects may experience in implementing open standards (such as, for example, the immaturity of standards or the poor support for standards by tool vendors; the resource implications in implementing some of the standards; etc.). We have sought to address such concerns by developing a matrix framework to assist in the selection of standards which are appropriate for use by standards, in the light of available funding, available expertise, maturity of standard, etc.

We hope that the wide range of advice provided in this handbook will be valuable to projects. However the most important aspect of this handbook is the quality assurance (QA) methodology which is outlined in the handbook. The QA methodology has been developed with an awareness of the constraints faced by projects. We have sought to develop a light-weight QA methodology which can be easily implemented and which should provide immediate benefits to projects during the development of their deliverables as well as ensuring interoperability and ease of deployment into service which will help to ensure the maximum effectiveness of JISC's overall digital library development work.

## Scope Of QA Focus

QA Focus seeks to ensure technical interoperability and maximum accessibility of project deliverables. QA Focus therefore has a focus on the technical aspects of project's work.

Our remit covers the following technical aspects:

**Digitisation:** The digitisation of resources, including text, image, moving image and sound resources.

**Access:** Access to resources, with particular references to access using the Web.

**Metadata:** The use of metadata, such as resource discovery metadata.

**Software development:** The development and deployment of software applications.

**Service deployment:** Deployment of project deliverables into a service environment.

In addition to these core technical areas we also address:

**Standards:** The selection and deployment of standards for use by projects.

**Quality assurance:** The development of quality assurance procedures by projects.

QA Focus's was originally funded to support JISC's 5/99 programme. However during 2003 our remit was extended to support JISC's FAIR and X4L in addition to 5/99.

### About QA Focus

QA Focus began its work on 1 January 2002. Initially the service was provided by UKOLN and ILRT, University of Bristol. However, following ILRT's decision to re-focus on their core activities they left QA Focus and were replaced by the AHDS on 1 January 2003.

This handbook has been developed by the current QA Focus team members: Brian Kelly, UKOLN (QA Focus project leader), Marieke Guy, UKOLN (QA Focus officer), Hamish James, AHDS (QA Focus project leader at AHDS) and Gareth Knight (QA Focus officer).

The handbook was subsequently published on the University of Bath institutional repository, Opus.

### Acknowledgments

We wish to acknowledge the many contributors to this handbook and others who made it possible.

We acknowledge the funding provided by JISC which made the work possible.

We acknowledge the contributions to the project made by members of JISC, in particular Caroline Ingram, Balviar Notay and Rachel Bruce.

We acknowledge the contributions made in the early stages of the QA Focus team by Karla Youngs and Ed Bremner of ILRT.

We acknowledge the contributions to case studies.

---

# 2 About This Handbook

---

This handbook provides a comprehensive summary of QA Focus's work. The handbook has the following structure.

**Background to the QA Focus Work:**

The section summarises the approaches to quality assurance which were taken by the QA Focus project.

**QA Focus Technical Advice:**

This section provides the advice and recommendations made to projects to help projects in the selection of appropriate standards and best practices, to advice on implementation frameworks and common problems experienced, and on approaches to ensuring that chosen standards and best practices are implemented.

The advice is provided in the form of brief and focussed advisory documents. These are complemented with a selection of case studies, which describe the real-world experiences of projects in implementing best practices.

**QA Focus Toolkit:**

QA Focus has developed a toolkit which can help projects in developing the technical framework for their activities. The toolkit consists of a series of checklists covering the range of technical areas addressed by QA Focus.

**QA Focus Papers:**

QA Focus has sought to validate its approaches by submitting papers to peer-reviewed conferences. The papers which have been published are provided in this handbook.

# 3 Background to the QA Focus Work

---

After the QA Focus project had been launched and began its work it was recognised that it was important to engage with the user community (the projects we support) in order to establish the requirements of the user community and to identify how QA Focus can seek to satisfy those requirements.

We held a number of focus groups in order to identify the main barriers to the development and deployment of project deliverables. The following concerns were identified:

### **Standards:**

Comments were made that the standards were felt to be too dry and it was sometimes difficult to identify which standards were relevant to projects, which standards were mature and suited for mainstream use and which required substantial technical expertise to make use of.

### **Implementation issues:**

Comments were made on a range of implementation issues. Concerns over the poor support for Web standards by certain browsers were identified by many. Other comments indicated that there was a need to share experiences on successes and failures with different approaches to implementation.

### **Service deployment:**

There appeared in some cases to be a lack of awareness of how projects deliverables were to be deployed into a service environment and the challenges services may find in deploying projects deliverables.

In light of the feedback we received QA Focus adopted the following approach to its work:

### **Producing advisory documents:**

These documents would provide an explanation of relevant standards and best practices; the different possible approaches to making use of such standards and ways of ensuing compliance with standards and best practices.

### **Commissioning case studies:**

The case studies would provide an opportunity for the projects themselves (and others working in similar areas) to share their approaches to implementing standards and best practices.

### **Developing a QA methodology:**

The development of a quality assurance methodology for use by projects which would help them to ensure that their project deliverables were fit for their purpose.

The deliverables from this work are provided in this handbook.

# 4 QA Focus Technical Advice

## Format Of This Section

The technical advice provided in this handbook includes the following areas.

**Standards:** The selection and deployment of standards for use by projects.

**Digitisation:** The digitisation of resources, including text, image, moving image and sound resources.

**Access:** Access to resources, with particular references to access using the Web.

**Metadata:** The use of metadata, such as resource discovery metadata.

**Software development:** Development and deployment of software applications.

**Service deployment:** Deployment of project deliverables into a service environment.

**General:** Other areas not covered above.

**Quality assurance:** The development of quality assurance procedures by projects.

In each of these areas we include appropriate QA Focus briefing documents. This is followed by example of relevant case studies.

All of these documents are available separately on the QA Focus Web site. We provide the citation details, including the URL for all of these documents.

## About the Briefing Documents

The briefing documents have been designed to be used as A5 flyers. This is to enable the documents to be easily used in a range of contexts. You may, for example, print appropriate briefing documents for use in meetings, seminars, workshops, etc.

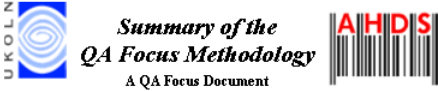
 <p><b>Background</b></p> <p>In order to provide value for money and a return on investment from the funders there is a need for project deliverables not only to be functional in their own right but also to be widely accessible, easily repurposed and deployed in a service environment.</p> <p>To achieve these aims projects should ensure that their deliverables comply with appropriate standards and best practices. Although it may be easy to require compliance, it may not always be easy to implement appropriate standards and best practices. In order to ensure that best endeavours are made it is recommended that projects should implement quality assurance (QA) procedures.</p> <p><b>QA Focus's Methodology</b></p> <p>Projects may be concerned that implementation of QA procedures can be time-consuming. The approach recommended by QA Focus is designed to be lightweight and to avoid unnecessary bureaucracy, while still providing a mechanism for implementation of best practices.</p> <p>The QA Focus methodology is based on the following:</p> <ul style="list-style-type: none"> <li><b>Documented policies on standards and best practices:</b> If standards and best practices are not documented it will be difficult to ensure best practices are implemented, especially in light of staff turnover, changing environments, etc.</li> <li><b>Documentation of the architecture used:</b> To ensure that the architecture used to implement the system is capable of complying with the standards.</li> <li><b>Documented exceptions:</b> There may be occasions when deviations from standards may be allowed. Such deviations should be documented and responsibility for this agreed.</li> <li><b>Systematic checking:</b> It is necessary to document systematic procedures for ensuring compliance with standards.</li> <li><b>Audit trails:</b> It can be helpful to provide audit trails which can help spotting trends.</li> </ul> <p>It is felt that use of this methodology should not only be beneficial to the projects themselves, but also help to minimise problems when project deliverables are re-used.</p> <p><small>Produced by QA Focus, a JISC-funded advice service supporting JISC 3500 projects Aug 2003</small></p>	<p><b>Example: QA For Web Sites</b></p> <p>As an example of implementation of this approach the QA policy for standards for the QA Focus Web site is given below.</p> <p><b>Area:</b> Web site: standards</p> <p><b>Standards:</b> The Web site will be based on the XHTML 1.0 and CSS 2.0 standards.</p> <p><b>Architecture:</b> The Web site will make use of PHP. XHTML 1.0 templates will be provided for use by authors, who will use simple HTML tools such as HTML-kit. Web site will provide access to an MS Access database. This will also comply with XHTML 1.0 and CSS 2.0 standards. The Web site will also host MS Word and MS PowerPoint files. These documents will also be available in HTML.</p> <p><b>Exceptions:</b> Resources converted from proprietary formats (such as MS Word and PowerPoint) need not necessarily comply with XHTML and CSS standards if doing so would be too time-consuming.</p> <p><b>Change Control:</b> The QA Focus project manager is responsible for changing this policy and addressing serious deviations from the policy.</p> <p><b>Checking:</b> When resources are created or updated the resource should be validated, usually using the <a href="#">validate</a> tool. When several resources are updated the <a href="#">validate</a> tool should be used.</p> <p><b>Audit trail:</b> A full audit should be carried out at least quarterly. The findings should be published on the QA Focus Web site, and deviations from the policy documented.</p> <p>A second example describes the QA policy for link checking of the QA Focus Web site.</p> <p><b>Area:</b> Web site: link checking</p> <p><b>Best Practice:</b> There should be no internal broken links. We should seek to fix broken links to external resources.</p> <p><b>Exceptions:</b> There may be broken links in historical documents or surveys. If remote Web sites are updated it may be too time-consuming to update links.</p> <p><b>Change Control:</b> The QA Focus project manager is responsible for changing this policy and addressing serious deviations from the policy.</p> <p><b>Checking:</b> When resources are created or updated the resource should be link-checked, usually using the <a href="#">checklink</a> tool. When several resources are updated the <a href="#">checklink</a> tool should be used.</p> <p><b>Audit trail:</b> A full audit should be carried out at least quarterly. Initially two tools should be used to spot deficiencies in the link-checking software. The findings should be published on the QA Focus Web site, and deviations from the policy documented.</p> <p>These two examples illustrate that developing QA policies need not be time-consuming. In addition implementation of these policies need not be time-consuming and can improve the quality of the Web site.</p> <p><small>For further information on QA Focus visit <a href="http://www.ukoln.ac.uk/qa-focus/">http://www.ukoln.ac.uk/qa-focus/</a></small></p>
---	---

Figure 1: Example Of A Typical Briefing Document

---

# 4.1 Standards

---

## Background

This section addresses the use of standards by projects.

JISC digital library development programmes seek to make use of open standards where possible in order to:

- Avoid application and platform dependencies
- Future-proof services
- Provide long-term access to resources

QA Focus seeks to ensure that projects are aware of this standards culture. However it is recognised that deployment of standards can prove difficult in certain circumstances for several reasons:

- Standards may be immature
- Authoring tools and viewers may not be widely available
- Deployment of standards may require technical expertise which is not readily available
- Projects may be building on existing systems which do not use appropriate standards
- Projects may find it difficult to ensure that they are complying fully with appropriate standards.

We are seeking to bridge this gap between the desire to make use of open standards and the implementation difficulties which projects may experience. We aim to do this by providing advice on the standards framework and on applicable standards in the technical areas covered by QA Focus; by sharing experiences on how projects addressed the section of standards and deployment of standards-based service and by developing a more flexible framework for use of open standards.

## Briefing Documents

The following briefing documents which address the area of standards have been produced:

- *What Are Open Standards?* (briefing-11)
- *Matrix for Selection of Standards* (briefing-31)

Advisory documents which cover specific technical areas are available within the section on the appropriate technical area.

## Case Studies

The following case studies which address the area of standards have been produced:

- *Creating Accessible Learning And Teaching Resources: The e-MapScholar Experience* (case-study-04)
- *ESDS Web Standards Policy* (case-study-16)



# What Are Open Standards?

Briefing

### About This Document

This document seeks to define what is meant by the term ‘open standards’.

### Citation Details

*What Are Open Standards?*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-11/>>

**Keywords:** standards, open standards, briefing

## Background

The *Standards and Guidelines to Build a National Resource* document (see <[http://www.jisc.ac.uk/index.cfm?name=projman\\_standards](http://www.jisc.ac.uk/index.cfm?name=projman_standards)>) lists many of the standards which projects funded by the JISC 5/99 and related programmes are expected to use.

The development programme for the Internet Environment (as the DNER is now known) is based on use of open standards. This raises two questions: “*Why open standards?*” and “*What are open standards?*”.

## Why Open Standards?

Open standards are required for several reasons:

- *Application Independence:*  
To ensure that access to resources is not dependent on a single application.
- *Platform Independence:*  
To ensure that access to resources is not restricted to particular hardware platforms.
- *Long-term Access:*  
To ensure that quality scholarly resources can be preserved and accessed over a long time frame.
- *Accessibility:*  
To ensure that resources can be accessed by people regardless of disabilities.
- *Architectural Integrity:*  
To ensure that the architectural framework for the Information Environment is robust and can be developed in the future.

## What Are Open Standards?

The term “open standards” is somewhat ambiguous. Open standards can mean:

- An open standards-making process
- Documentation freely available on the Web
- Use of the standard is uninhibited by licensing or patenting issues
- Standard ratified by recognised standards body

Some examples of recognised open standards bodies are given in Table 1.

Organisation	Comments
W3C	World Wide Web Consortium (W3C). Responsible for development of Web standards (known as <i>Recommendations</i> ). See W3C Recommendations at < <a href="http://www.w3.org/TR/">http://www.w3.org/TR/</a> >. Standards include HTML, XML, CSS, SVG, etc.
IETF	Internet Engineering Task Force (IETF). Responsible for the development of Internet standards (known as <i>IETF RFCs</i> ). See list of IETF RFCs at < <a href="http://www.ietf.org/rfc.html">http://www.ietf.org/rfc.html</a> >. Standards include HTTP, MIME, etc.
ISO	International Organisation For Standardization (ISO). See < <a href="http://www.iso.org/iso/en/stdsdevelopment/whowhenhow/how.html">http://www.iso.org/iso/en/stdsdevelopment/whowhenhow/how.html</a> >.
NISO	National Information Standards Organization (NISO). See < <a href="http://www.niso.org/">http://www.niso.org/</a> >. Relevant standards include Z39.50.
IEEE	Institute of Electrical and Electronics Engineers (IEEE). See < <a href="http://www.ieee.org/">http://www.ieee.org/</a> >.
ECMA	ECMA International. Association responsible for the standardisation of Information and Communication Technology Systems (such as JavaScript). See < <a href="http://www.ecma-international.org/">http://www.ecma-international.org/</a> >.

**Table 1: Examples Of Independent Standards Organisations**

### Other Types Of Standards

The term *proprietary* refers to formats which are owned by an organisation, group, etc. Unfortunately since this term has negative connotations, the term *industry standard* is often used to refer to a widely used proprietary standard. For example, the proprietary Microsoft Excel format is sometimes referred to as an industry standard for spreadsheets. To make matters even more confusing, the prefix is sometime omitted and MS Excel can be referred to as a standard.

To further confuse matters, companies which own proprietary formats may choose to make the specification freely available. Alternatively third parties may reverse engineer the specification and publish the specification. In addition tools which can view or create proprietary formats may be available on multiple platforms or as open source.

In all these cases, although there may appear to be no obvious barriers to use of the proprietary format, such formats should not be classed as open standards as they have not been approved by a neutral standards body. The organisation owning the format may chose to change the format or the usage conditions at any time. File formats in this category include Microsoft Office formats, Adobe's PDF, Macromedia Flash and Java.

# Matrix for Selection of Standards

Briefing

### About This Document

This document describes a selection matrix to help in choosing standards for use by projects.

### Citation Details

*Matrix for Selection of Standards*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-31/>>

**Keywords:** standards, selection, briefing

## Background

JISC provide advice on a wide range of standards and best practices which seek to ensure that project deliverables are platform and application-independent, accessibility, interoperable and are suitable for re-purposing.

The standards and best practices which JISC advisory service recommend have been developed with these aims in mind.

## Challenges

Although use of recommended standards and best practices is encouraged, there may be occasions when this is not possible:

### **Building on existing systems:**

Projects may be based on development of existing systems, which do not use appropriate standards.

### **Standards immature:**

Some standards may be new, and there is a lack of experience in their use. Although some organisations may relish the opportunity to be early adopters of new standards, others may prefer to wait until the benefits of the new standards have been established and many teething problems resolved.

### **Functionality of the standard:**

Does the new standard provide functionality which is required for the service to be provided?

### **Limited support for standards:**

There may be limited support for the new standards. For example, there may be a limited range of tools for creating resources based on the new standards or for viewing the resources.

### **Limited expertise:**

There may be limited expertise for developing services based on new standards or there may be limited assistance to call on in case of problems.

### **Limited timescales:**

There may be insufficient time to gain an understanding of new standards and gain experience in use of tools.

In many cases standards will be mature and expertise readily available. The selection of the standards to be deployed can be easily made. What should be done when this isn't the case?

### A Matrix Approach

In light of the challenges which may be faced when wishing to make use of recommended standards and best practices it is suggested that projects use a matrix approach to resolving these issues.

Area	Your Comments
<b>Standard</b>	
How mature is the standard?	
Does the standard provide required functionality?	
<b>Implementation</b>	
Are authoring tools which support the standard readily available?	
Are viewing tools which support the standard readily available?	
<b>Organisation</b>	
Is the organisation culture suitable for deployment of new standards?	
Are there strategies in place to continue development in case of staffing changes?	

Individual projects will need to formulate their own matrix which covers issues relevant to their particular project, funding, organisation, etc.

### Implementation

This matrix approach is not intended to provide a definitive solution to the selection of standards. Rather it is intended as a tool which can assist projects when they go through the process of choosing the standards they intend to use. It is envisaged that projects will document their comments on issues such as those listed above. These comments should inform a discussion within the project team, and possibly with the project's advisory or steering group. Once a decision has been made the rationale for the decision should be documented. This will help to ensure that the reasonings are still available if project teams members leave.

For examples of how projects have addressed the selection of standards see:

- *ESDS Web Standards Policy* case study:  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-16/>>
- *Standards for e-learning: The e-MapScholar Experience* case study:  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-05/>>.

# Creating Accessible Learning And Teaching Resources: The e-MapScholar Experience

Case study

### About This Document

This case study describes the rationale for complying with accessibility guidelines by the e-MapScholar project and their approaches to implementing this policy.

### Citation Details

*Creating Accessible Learning And Teaching Resources: The e-MapScholar Experience*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-04/>>

### Related Document

See also the *What Are Open Standards?* briefing document (briefing-11).

**Keywords:** standards, selection, Web, accessibility, e-MapScholar, eMapScholar, case study

## Background

**e-MapScholar** [1], a JISC 5/99 funded project, aims to develop tools and learning and teaching materials to enhance and support the use of geo-spatial data currently available within tertiary education in learning and teaching, including digital map data available from the EDINA **Digimap** service [2]. The project is developing:

- A range of Teaching Case Studies in various subject disciplines.
- An online Learning Resource Centre providing access to resources in Working with Digital Map Data, Integrating Spatial Data and Data Visualisation. These will include interactive tools for use by students that illustrate key concepts and perform some basic analytical tasks.
- A Content Management System that allows lecturers to customise online learning units and interactive tools through provision of discipline and place specific examples.
- A Virtual Work Placement - based on a wind farm development in Wales.

The *Disability Discrimination Act (1995)* (DDA) aimed to end discrimination faced by many disabled people. While the DDA focused mainly on the employment and access to goods and services, the *Special Education Needs and Disability Act (2001)* (SENDA) [3] amended the DDA to include education. The learning and teaching components of SENDA came into force in September 2002. SENDA has repercussions for all projects producing online learning and teaching materials for use in UK education because creating accessible materials is now a requirement of the JISC rather than a desirable project deliverable.

This case study describes how the e-MapScholar team has addressed accessibility in creating the user interfaces for the learning resource centre, case studies, content management system and virtual placement.

## 4.1 Technical Advice: Standards: Case Studies

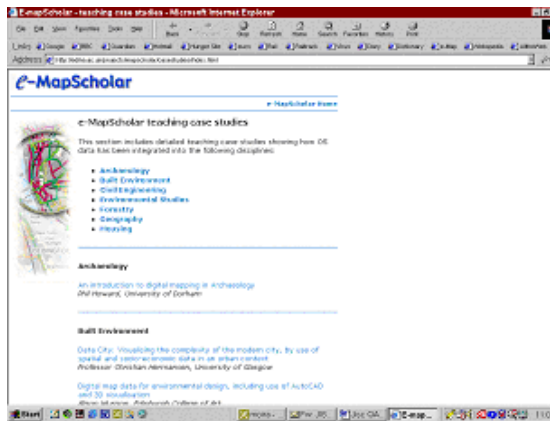


Figure 1: Screenshot of case study index



Figure 2: Screenshot of learning resource centre resource selection page

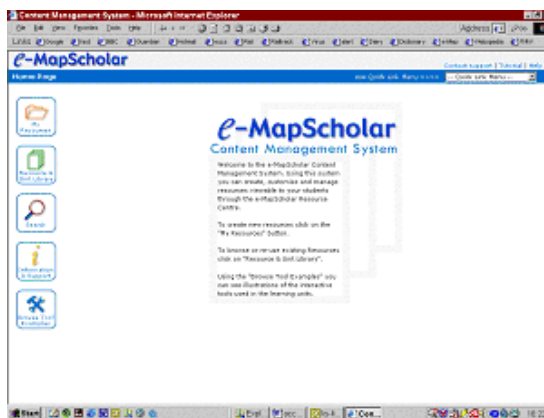


Figure 3: Screenshot of content management system



Figure 4: Screenshot of the Virtual Placement

## Why Online Learning And Teaching Materials Should Be Accessible

An accessible Web site is a Web site that has been designed so that virtually everyone can navigate and understand the site. A Web site should be informative, easy to navigate, easy to access, quick to download and written in a valid hypertext mark up language. Designing accessible Web sites benefits all users, not just disabled users.

Under SENDA the e-MapScholar team must ensure that the project deliverables are accessible to users with disabilities including mobility, visual or audio impairments or cognitive/learning issues. These users may need to use specialist browsers (such as speech browsers) or configure their browser to enhance the usability of Web sites (e.g. change font sizes). It is also a requirement of the JISC funding that the 5/99 projects should reach at least priority 1 and 2 of the Web Accessibility Initiative (WAI) guidelines [4] and where appropriate, priority 3.

## The Approach Taken

The project has been split into four major phases:

1. **Planning:** Early in the project the team looked at technical standards and constraints and went on to develop an Instructional Design document which looked partly at accessibility as well as good practice in terms of Web design, e-learning and online assessment. Accessibility therefore became an integral part of the project rather than an add on to be considered at a later date.

2. **Development** of tools and materials alongside user interfaces. Note that the design of the interfaces are being reviewed and updated as feedback from learning technologists and users suggest where improvements can or need to be made.
3. **Testing** with different browsers and also using tools such as Bobby [5].
4. **Evaluation** an evaluation team based at the Open University is evaluating software and tools directly with users and feeding back their findings to the developers and authors.

### Overall Content Design

While the CMS and learning units are inter-connected, the other components can exist separately from one another.

The project has employed a simple and consistent design in order to promote coherence and also to ease navigation of the site. Each part employs similar headings, navigation and design.

The basic Web design was developed within the context of the learning units and was then adapted for the case studies and virtual placement.

Summaries, learning objectives and pre-requisites are provided where necessary.

### User Support

Links to help page are provided. The help page will eventually provide information on navigation, how to use the interactive tools and FAQ in the case of learning units and details of any plug-ins/software used in the case studies.

### Text formatting

Font size will be set to be resizable in all browsers.

Verdana font has been used as this is considered the most legible font face.

CSS (Cascading style sheets) have been used to control the formatting of text; this cuts out the use of non-functional elements, which could interfere with text reader software.

### Navigation and Links

Navigation has been used in a consistent manner.

All navigational links are in standard blue text.

All navigation links are text links apart from the learning unit progress bar, which is made up of clickable images. These images include an ALT tag describing the text alternative.

The progress bar provides a non-linear pathway through the learning units, as well as providing the user with an indication of their progress through the unit.

The link text used can be easily understood when out of context, e.g. "Back to Resource" rather than "click here".

'Prev' and 'Next' text links provide a simple linear pathway through both the learning units and the case studies.

All links are keyboard accessible for non-mouse users and can be reached by using the tab key.

Where possible the user is offered a choice of pathway through the materials e.g. the learning units can be viewed as a long scrolling page or page-by-page chunks.

### **Colours**

Web safe colours have been used in the student interface.

The interface uses a white background ensuring maximum contrast between the black text, and blue navigational links.

### **Images**

Very few graphics have been used in the interface design to minimise download time.

Content graphics and the project logo have ALT tags providing a textual description.

Long descriptions will be incorporated where necessary.

Graphics for layout will contain "" (i.e. null) ALT tags so they will be ignored by text reader software.

### **Tables**

Tables have been used for layout purposes complying with W3C standards; not using structural mark-up for visual formatting.

### **Technical**

HTML 4.0 standards have been complied with.

JavaScript has been used for the pop-up help menu, complying with JavaScript 1.2 standards.

The user is explicitly informed that a new window is to be opened.

The new window is set to be smaller so that it is easily recognised as a new page.

Layout is compatible with early version 4 browsers, both in Netscape, Internet Explorer and Opera.

Specific software or plug-ins are required to view some of the case study materials e.g. GIS or AutoCAD software has been used in some of the case studies. Users will be advised of these and where possible will be guided to free viewers. Where possible the material will be provided in an alternative format such as screen shots, which can be saved as an image file.

Users are warned when non-HTML documents (e.g. PDF or MS Word) are included in the case studies and where possible documents are also provided in HTML format.

## **Problems Experienced**

Problems experienced have generally been minor.

- Input errors such as forgetting an ALT tag can be monitored by using testing software such as Bobby.
- Conflicts can arise between desired appearance/style and limitations imposed by creating accessible web pages. This is affected by the use of different browsers. One example of this has occurred when trying to set the font size. We originally set up font size using a fixed size:



```
.bodytext {font-family: verdana, arial; font-size: 10pts;
font-style: normal}
```

This could be changed by the user in Netscape but not in IE. This has been changed to:

```
.bodytext {font-family: verdana, arial; font-size: small;
font-style: normal}
```

However this solution also has problems. There are differences in the way that browsers view the page e.g. in Netscape the text is much smaller than in IE. When text is a legible size in Netscape it looks less appealing in IE.

- The case studies were originally developed with a navigation bar at the top and the bottom of the page containing about seven links. However it was pointed out that users with speech browsers would find this repetition annoying. This has been resolved with the inclusion of 'next','prev' and a 'back to top' link at the bottom of the page instead of the full button bar. Also a 'skip header' button has been included for users to jump the top navigation buttons if preferred.

### Things We Would Do Differently

Throughout the project accessibility has been thought of as an integral part of the project and this approach has generally worked well. Use of templates and CSS have helped to minimise workload when a problem has been noted and the materials updated.

It is important that time and thought goes into the planning stage of the project as it is easier and less time consuming to adopt accessible Web design in an early stage of the project than it is to retrospectively adapt features to make them accessible.

User feedback from evaluations, user workshops and demonstrations has been extremely useful in identifying potential problems.

### Resources

1. **e-MapScholar Project home page**, EDINA, <<http://edina.ac.uk/projects/mapscholar/>>
2. **EDINA Digimap**, EDINA, <<http://edina.ac.uk/digimap/>>
3. **Special Educational Needs and Disability Act 2001**, The Stationery Office, <<http://www.hmsso.gov.uk/acts/acts2001/10010--d.htm>>
4. **Web Content Accessibility Guidelines 1.0**, W3C, <<http://www.w3.org/TR/WCAG10/>>
5. **Bobby**, Watchfire Corporation, <<http://bobby.watchfire.com/bobby/html/en/>>

### Contact Details

Deborah Kent  
EDINA National Data Centre (St Helens Office)  
ICT Centre  
St Helens College  
Water St  
ST HELENS, WA10 1PZ  
Email: [dkent@ed.ac.uk](mailto:dkent@ed.ac.uk)

Lynne Robertson  
School of Earth, Environmental and  
Geographical Sciences  
The University of Edinburgh  
Drummond Street  
Edinburgh EH8 9XP  
Email: [lr@geo.ed.ac.uk](mailto:lr@geo.ed.ac.uk)

# ESDS Web Standards Policy

Case study

### About This Document

This case study describes the approaches to selection of Web standards by the ESDS.

### Citation Details

*ESDS Web Standards Policy*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-16/>>

### Related Documents

See also the *Matrix for Selection of Standards* (briefing-31) and *How To Evaluate A Web Site's Accessibility Level* (briefing-12) briefing documents.

**Keywords:** Web, standards, ESDS, case study

## Background

The **Economic and Social Data Service (ESDS)** [1] is a national data archiving and dissemination service which came into operation in January 2003.

The ESDS service is a jointly-funded initiative sponsored by the Economic and Social Research Council (ESRC) [2] and the Joint Information Systems Committee (JISC) [3].

## Problem Being Addressed

Many Web sites fail to comply with accessibility and usability guidelines or consist of valid code. Prior to setting up the ESDS Web site it was decided that a Web Standards Policy would be agreed upon and adhered to.

## The Approach Taken

The ESDS Web Standards Policy was released in June 2003 and applies to all newly constructed ESDS Web pages.

### ESDS Web Standards Policy

ESDS is committed to following agreed best standards and good practice in Web design and usability. The underlying code of the ESDS Web site achieves compliance with W3C guidelines for XHTML and Cascading Style Sheets (CSS). It strives to meet Web Content Accessibility Guidelines and be Special Educational Needs Disability Act (SENDA) compliant. Where this is not feasible or practical (e.g. proprietary software programs such as Nesstar Light is used) we will provide an alternative method for users to obtain the assistance they need from our user support staff. JISC and UKOLN recommendations have been reviewed for this policy.

### 1 XHTML and Accessibility Standards

Standards	Validation and Auditing Tools
XHTML 1.0 Transitional < <a href="http://www.w3.org/TR/xhtml1/">http://www.w3.org/TR/xhtml1/</a> >	W3C XHTML validation service < <a href="http://validator.w3.org/">http://validator.w3.org/</a> >
CSS Level 2 < <a href="http://www.w3.org/TR/REC-CSS2/">http://www.w3.org/TR/REC-CSS2/</a> >	W3C's CSS validation service < <a href="http://jigsaw.w3.org/css-validator/">http://jigsaw.w3.org/css-validator/</a> >
WCAG 1.0	A-prompt

Conformance Level: all Priority 1 checkpoints, most Priority 2 checkpoints, and some Priority 3 checkpoints.	<a href="http://aprompt.snow.utoronto.ca/Bobby">&lt;http://aprompt.snow.utoronto.ca/Bobby&gt;</a> <a href="http://bobby.watchfire.com/bobby/">&lt;http://bobby.watchfire.com/bobby/&gt;</a>
--	--

For more detailed information about accessibility standards and how best to implement them see:

**Bookmarklets,**

[<http://www.yourhtmlsource.com/accessibility/bookmarklets.html>](http://www.yourhtmlsource.com/accessibility/bookmarklets.html)

**Checklist of Checkpoints for Web Content Accessibility Guidelines 1.0, W3C,**

[<http://www.w3.org/TR/WAI-WEBCONTENT/full-checklist.html>](http://www.w3.org/TR/WAI-WEBCONTENT/full-checklist.html)

**Web Content Accessibility Guidelines 1.0 – Conformance, W3C,**

[<http://www.w3.org/TR/WCAG10/>](http://www.w3.org/TR/WCAG10/)

**WDG HTML Validator,** [<http://www.htmlhelp.com/tools/validator/batch.html>](http://www.htmlhelp.com/tools/validator/batch.html)

## 2 Non-HTML Formats

HTML is the recommended format for small documents and Web pages.

ESDS also provides access to significant amounts of lengthy documentation to users as part of its service. For these lengthier, more complex documents, we generally follow these JISC recommendations.

If a link leads to a non-HTML file, e.g., a zip or Adobe PDF file, this will be clearly indicated.

### Portable Document Format (PDF)

For documents provided in PDF, a link to the Adobe free viewer will be made available.

### Rich Text Format (RTF)

All leading word processing software packages include a standard facility for reading RTF and some documents may therefore be made available in this format.

## 3 Link Checking

The ESDS is committed to keeping the links on its pages as accurate as possible.

ESDS Web pages are checked using Xenu Link Sleuth [4] or an equivalent checker, on a monthly basis.

ESDS catalogue records are checked using Xenu Link Sleuth or an equivalent checker on a monthly basis.

ESDS Web page links are manually checked every six months to verify that the content of the pages to which they link is still appropriate.

## 4 Browser Compatibility

New templates and all pages are checked for use with these standard browsers:

- Internet Explorer 5.0 and 6.0
- Netscape 6.0 and 7.0
- Opera 7.0

### 5 Software/Hardware Platforms

We test our pages on PCs using Microsoft Windows operating systems. We do not have the equipment to test on an Apple Macintosh platform and rely on the standards we use to assure accessibility.

### References

1. **MIMAS**, <<http://www.esds.ac.uk/>>
2. **ESRC**, <<http://www.esrc.ac.uk/>>
3. **JISC**, <<http://www.jisc.ac.uk/>>
4. **Xenu's Link Sleuth**, <<http://home.snafu.de/tilman/xenulink.html>>

### Contact details

Diane Geraci and Sharon Jack  
Economic and Social Data Service  
UK Data Archive  
University of Essex  
Wivenhoe Park  
Colchester  
Essex  
UK  
CO4 3SQ

# 4.2 Digitisation

---

## Background

This section addresses the area of digitisation. The briefing documents seek to describe best practices in this area.

## Briefing Documents

The following briefing documents which address the area of digitisation have been produced:

- *Image QA in the Digitisation Workflow* (briefing-09)
- *QA Procedures For The Design Of CAD Data Models* (briefing-18)
- *Documenting Digitisation Workflow* (briefing-20)
- *QA For GIS Interoperability* (briefing-21)
- *Choosing A Suitable Digital Rights Solution* (briefing-22)
- *Recording Digital Sound* (briefing-23)
- *Handling International Text* (briefing-24)
- *Choosing a Suitable Digital Video Format* (briefing-25)
- *Implementing Quality Assurance For Digitisation* (briefing-27)
- *Choosing An Appropriate Raster Image Format* (briefing-28)
- *Choosing A Vector Graphics Format For The Internet* (briefing-29)
- *Transcribing Documents* (briefing-47)
- *Digitising Data For Preservation* (briefing-62)
- *Audio For Low-Bandwidth Environments* (briefing-65)
- *Producing And Improving The Quality Of Digitised Images* (briefing-66)
- *Implementing and Improving Structural Markup* (briefing-67)
- *Techniques To Assist The Location And Retrieval Of Local Images* (briefing-68)
- *QA Techniques For The Storage Of Image Metadata* (briefing-71)
- *Improving The Quality Of Digitised Images* (briefing-74)
- *Digitisation Of Still Images Using A Flatbed Scanner* (briefing-75)
- *Choosing A Suitable Digital Watermark* (briefing-76)

## Case Studies

The following case studies which address the area of digitisation have been produced:

- *Using SVG in the ARTWORLD Project* (case-study-07)
- *Crafts Study Centre Digitisation Project - and Why 'Born Digital'* (case-study-08)
- *Image Digitisation Strategy and Technique: Crafts Study Centre Digitisation Project* (case-study-09)
- *Digitisation of Wills and Testaments by the Scottish Archive Network (SCAN)*, (case-study-19)

## Image QA in the Digitisation Workflow

### About This Document

This briefing document describes the important of implementing quality controls when creating and storing images.

### Citation Details

Image QA in the Digitisation Workflow, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-09/>>

**Keywords:** digitisation, image QA, workflow, benchmark, briefing

## Introduction

Producing an *archive* of high-quality images with a server full of associated *delivery* images is not an easy task. The workflow consists of many interwoven stages, each building on the foundations laid before. If, at any stage, image quality is compromised within the workflow, it has been totally lost and can never be redeemed.

It is therefore important that image quality is given paramount consideration at all stages of a project from initial project planning through to exit strategy.

Once the workflow is underway, quality can only be lost and the workflow must be designed to capture the required quality right from the start and then safeguard it.

## Image QA

Image QA within a digitisation project's workflow can be considered a 4-stage process:

### 1 Strategic QA

*Strategic QA* is undertaken in the initial planning stages of the project when the best methodology to create and support your images, now and into the future will be established. This will include:

- Choosing the correct file types and establishing required sizes
- Sourcing and benchmarking all equipment
- Establishing capture guidelines
- Selecting technical metadata

### 2 Process QA

*Process QA* is establishing quality control methods within the image production workflow that support the highest quality of capture and image processing, including:

- Establishing best 'image capture' and 'image processing' methodology and then standardising and documenting this best practice
- Regularly calibrating and servicing all image capture and processing equipment

- Training operators and encouraging a pride in quality of work
- Accurate capture of metadata

#### 3 Sign-off QA

*Sign-off QA* is implementing an audited system to assure that all images and their associated metadata are created to the established quality standard. A QA audit history is made to record all actions undertaken on the image files.

- Every image must be visually checked and signed off with name and time recorded within audit history
- All metadata must be reviewed by operator and signed off with name and time
- Equipment must be calibrated and checked regularly
- All workflow procedures reviewed and updated as necessary

#### 4 On-going QA

*On-going QA* is implementing a system to safeguard the value and reliability of the images into the future. However good the initial QA, it will be necessary to have a system that can report, check and fix any faults found within the images and associated metadata after the project has finished. This system should include:

- Fault report system that allows faults to be checked and then if possible fixed
- Provision for ongoing digital preservation (including migration of image data)
- Ownership and responsibility for images, metadata and IMS
- A reliable system for the on-going creation of surrogate images as required

### QA in the Digitisation Workflow

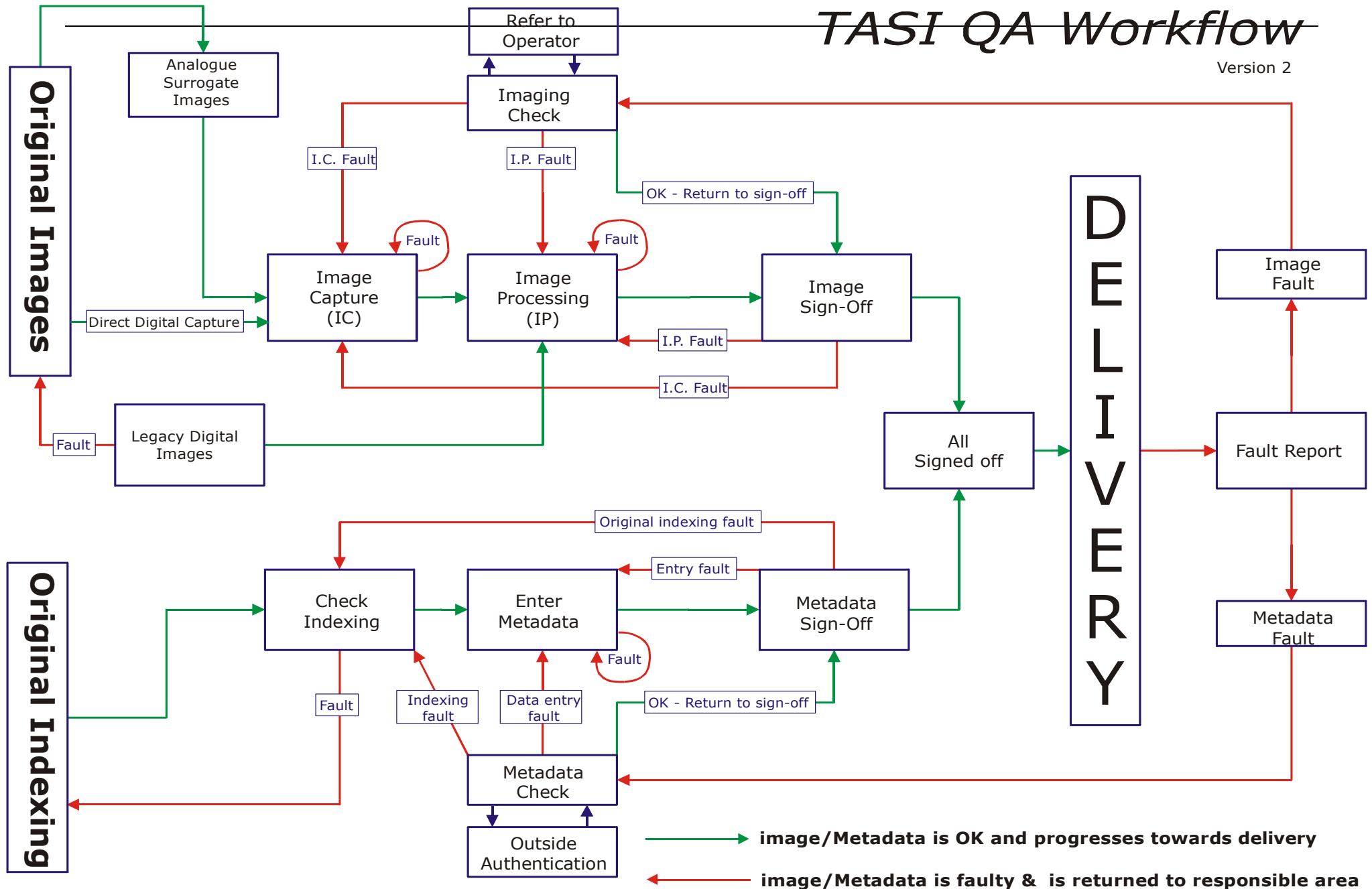
Much of the final quality of a delivered image will be decided, long before, in the initial '*Strategic*' and '*Process*' QA stages where the digitisation methodology is planned and equipment sourced. However, once the process and infrastructure are in place it will be the operator who needs to manually evaluate each image within the '*Sign-off*' QA stage. This evaluation will have a largely subjective nature and can only be as good as the operator doing it. The project team is the first and last line of defence against any drop in quality. All operators must be encouraged to take pride in their work and be aware of their responsibility for its quality.

It is however impossible for any operator to work at 100% accuracy for 100% of the time and faults are always present within a productive workflow. What is more important is that the system is able to accurately find the faults before it moves away from the operator. This will enable the operator to work at full speed without having to worry that they have made a mistake that might not be noticed.

The image digitisation workflow diagram in this document shows one possible answer to this problem.

# TASI QA Workflow

Version 2





# QA Procedures For The Design Of CAD Data Models

### About This Document

This briefing document describes procedures to reduce long-term manageability and interoperability problems in the design of CAD data models.

### Citation Details

*QA Procedures For The Design Of CAD Data Models*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-18/>>

**Keywords:** CAD, data model, conventions, geometry, briefing

## Background

The creation of CAD (Computer Aided Design) models is an often complex and confusing procedure. To reduce long-term manageability and interoperability problems, the designer should establish procedures that will monitor and guide system checks.

## Establish CAD Layout Standards

Interoperability problems are often caused by poorly understood or non-existent operating procedures for CAD. It is wise to establish and document your own CAD procedures, or adopt one of the national standards developed by the BSI (British Standards Institution) or NIBS (National Institute of Building Sciences). These may be used to train new members in the house-style of a project, provide essential information when sharing CAD data among different users, or provide background material when depositing the designs with a preservation repository. Particular areas to standardize include:

- Drawing sheet templates
- Paper layouts
- Text fonts, dimensions, line types and line weights
- Layer naming conventions
- File naming conventions

Procedures on constructing your own CAD standard can be found in the Construct IT guidelines (see references).

## Be Consistent With Layers And Naming Conventions

When creating CAD data models, a consistent approach to layer creation and naming conventions is useful. This will avoid confusion and increases the likelihood that the designer will be able to manipulate and search the data model at a later date.

- Create layers that divide the object according to pre-defined criteria. E.g. a model building may be divided into building part, building phase, site stratum, material, chronological standing, etc. The placement of too many objects on a single layer will increase the computational requirements to process the model and cause unexpected problems when moving objects between layers.
- Establish a layer name convention that is consistent and appropriate to avoid confusion in complex CAD model. Many users use 'wall', 'wall1', 'door', etc. to describe specific objects. This is likely to become confusing and difficult to

identify when the design becomes more complex. Layer names should be short and descriptive. A possible option is the CSA layer-naming convention that uses each character in the layer name to describe its position within the model.

### Ensure Tolerances Are Consistent

When exporting designs between different CAD applications it is common for model relationships to disintegrate, causing entities to appear disconnected or disappear from the design altogether. A common cause is the use of different tolerance levels – a method of placing limits on gaps between geometric entities. The method of calculating tolerance often varies in different applications: some use absolute tolerance levels (e.g. 0.005mm), others work to a tolerance level relative to the model size (e.g.  $10^{-4}$  the size), while others have different tolerances according to the units used. When considering moving a design between different applications it is useful to ensure the tolerance level can be set to the same value and identify potential problem areas that may be corrupted when the data model is reopened.

### Check For Illegal Geometry Definitions

Interoperability problems are also caused by differences in how the system identifies invalid geometry definitions, such as the three-sided degenerate NURBS surfaces. Some systems allow the creation of such entities, others will reject them, whereas some systems know that they are not permissible and in an effort to prevent them from being created, generate twisted four sided surfaces.

### Further Information

- *AHDS Guides to Good Practice: CAD*, AHDS, <<http://ads.ahds.ac.uk/project/goodguides/cad/>>
- *National CAD Standard*, <<http://www.nationalcadstandard.org/ncs2.htm>>
- *CAD Interoperability*, Kelly, D (2003). <[http://www.tech.purdue.edu/cg/courses/cgt226/cad\\_interop.pdf](http://www.tech.purdue.edu/cg/courses/cgt226/cad_interop.pdf)>
- *CAD Standards: Develop and Document*, <<http://www.arch.vuw.ac.nz/papers/bbsc303/assign2/198mc.htm>>
- *Construct I.T: Construct Your Own CAD Standard*, <<http://www.construct-it.org.uk>> (Note URL to resource is not available)
- *Common Interoperability Problems in CAD*, <[http://www.caduser.com/reviews/reviews.asp?a\\_id=148](http://www.caduser.com/reviews/reviews.asp?a_id=148)>

# Documenting Digitisation Workflow

### About This Document

This briefing document describes how to track workflow within a digitisation project.

### Citation Details

*Documenting Digitisation Workflow*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-20/>>

**Keywords:** digitisation, documentation, briefing

## Background

Digitisation is a production process. Large numbers of analogue items, such as documents, images, audio and video recordings, are captured and transformed into the digital masters that a project will subsequently work with. Understanding the many variables and tasks in this process – for example the method of capturing digital images in a collection (scanning or digital photography) and the conversion processes performed (resizing, decreasing bit depth, convert file formats, etc.) – is vital if the results are to remain consistent and reliable.

By documenting the workflow of digitisation, a life history can be built-up for each digitised item. This information is an important way of recording decisions, tracking problems and helping to maintain consistency and give users confidence in the quality of your work.

## What to Record

Workflow documentation should enable us to tell what the current status of an item is, and how it has reached that point. To do this the documentation needs to include important details about each stage in the digitisation process, and its outcome.

1. **What action was performed at a specific stage?** Identify the action performed. For example, resizing an image.
2. **Why was the action performed?** Establish the reason that a change was made. For example, a photograph was resized to meet pre-agreed image standards.
3. **When was the action performed?** Indicate the specific date the action was performed. This will enable project development to be tracked through the system.
4. **How was the action performed?** Ascertain the method used to perform the action. A description may include the application in use, the machine ID, or the operating system.
5. **Who performed the action?** Identify the individual responsible for the action. This enables actions to be tracked and identify similar problems in related data.

By recording the answers to these five questions at each stage of the digitisation process, the progress of each item can be tracked, providing a detailed breakdown of its history. This is particularly useful for tracking errors and locating similar problems in other items.

The actual digitisation of an item is clearly the key point in the workflow, and therefore formal capture metadata (metadata about the actual digitisation of the item) is particularly important.

### Where to Record the Information

Where possible, select an existing schema with a binding to XML:

- TEI (Text Encoding Initiative) and EAD (Encoded Archival Description) for textual documents
- NISO Z39.87 for digital still images.
- SMIL (Synchronized Multimedia Integration Language), MPEG-7 or the Library of Congress' METS A/V extension for Audio/Video.

### Quality Assurance

To check your XML document for errors, QA techniques should be applied:

- Validate XML against your schema or an XML parser
- Check that free text entries follow local rules and style guidelines

### Further Information

- *Encoded Archival Description*, <<http://www.loc.gov/ead/>>
- *A Metadata Primer*, <[http://www.cmswatch.com/Features/TopicWatch/FeaturedTopic/?feature\\_id=85](http://www.cmswatch.com/Features/TopicWatch/FeaturedTopic/?feature_id=85)>
- *Dublin Core Metadata Initiative*, <<http://dublincore.org/>>
- *MARC Standards*, <<http://www.loc.gov/marc/>>
- *MPEG- 7 Standard*, <<http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>>
- *Synchronized Multimedia*, <<http://www.w3.org/AudioVideo/>>
- *TEI Consortium*, <<http://www.tei-c.org/>>
- *Three SGML Metadata Formats: TEI, EAD, and CIMI*, <<http://hosted.ukoln.ac.uk/biblink/wp1/sgml/tei.rtf>>
- *Z39.87: Technical Metadata For Still Digital Images*, <[http://www.niso.org/standards/resources/Z39\\_87\\_trial\\_use.pdf](http://www.niso.org/standards/resources/Z39_87_trial_use.pdf)>

# QA for GIS Interoperability

### About This Document

This briefing document describes methods to improve interoperability between different GIS data.

### Citation Details

*QA for GIS Interoperability*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-21/>>

**Keywords:** Geographic Information System, GIS, data structure, measurement, briefing

## Background

Quality assurance is essential to ensure GIS (Geographic Information System) data is accurate and can be manipulated easily. To ensure data is interoperable, the designer should audit the GIS records and check them for incompatibilities and errors.

## Ensure Content Is Available In An Appropriate GIS Standard

Interoperability between GIS standards is encouraged, enabling complex data types to be compared in unexpected methods. However, the varying standards can limit the potential uses of the data. Designers are often limited by the formats available in different tools. When possible, it is advisable to use OpenGIS - an open, multi-subject standard constructed by an international standard consortium.

## Resolve Differences In The Data Structures

To integrate data from multiple databases, the data must be stored in a compatible field structure. Complementary fields in the source and target databases must be of a compatible type (Integer, Floating Point, Date, a Character field of an appropriate length etc.) to avoid the loss of data during the integration process. Checks should also be made that specific fields that are incompatible with similar products (e.g. dBase memo fields) are exported correctly. Specialist advice should be taken to ensure the memo information is not lost.

## Ensure Data Meet The Required Standards

Databases are often created in an *ad hoc* manner without consideration of later requirements. To improve interoperability the designer should ensure data complies with relevant standards. Examples include the BS7666 standard for British postal addresses and the RCHME Thesauri of Architectural Types, Monument Types, and Building Materials.

## Compensate For Different Measurement Systems

The merging of two different data sources is likely to present specific problems. When combining two GIS tables, the designer should consider the possibility that they have been constructed using different projection measurement systems (a method of representing the Earth's three-dimensional form on a two-dimensional plane and locate landmarks by a set of co-ordinates). The projection co-ordinate systems vary across nations and through time: the US has five primary co-ordinate systems in use that significantly differ with each other. The British National Grid removes this confusion by using a single co-ordinate, but can cause problems when merging contemporary with

pre-1940 maps that were based upon Cassini projection. This may produce incompatibilities and unexpected results when plotted, such as moving boundaries and landmarks to different locations that will need to be rectified before any real benefits can be gained. The designer should understand the project system used for each layer to compensate for inaccuracies.

### Ensure Precise Measurements Are Accurate

When recreating real-world objects created by two different people, the designer should note the degree of accuracy. One person may measure to the nearest millimetre, while the other measures to the centimetre. To check this, the designer should answer the following questions:

1. How many numbers are shown after the point (e.g. 2:12 cm)?
2. Is this figure consistent with the second designers' measurement methods?
3. Has the value been rounded up or down, or has a third figure been removed?

These subtle differences may influence the resulting model, particularly when designing smaller objects.

### Further Information

- *AHDS Guides to Good Practice*, AHDS,  
<<http://ads.ahds.ac.uk/project/goodguides/gis/>>
- *Geoplace – The Authoritative Source For Spatial Information*,  
<<http://www.geoplace.com/>>
- *GIS - Computational Problems*, <<http://www.cise.ufl.edu/~mssz/GIS/GIS-Basics.html>>
- *Using GIS To Help Solve Real-World Problems*,  
<<http://www.cadsystems.com/archive/0006f01.html>>
- *Open GIS Consortium, Inc.*  
<<http://www.opengis.org/>>

# Choosing A Suitable Digital Rights Solution

### About This Document

This briefing document defines criteria for choosing a digital rights solution and identifying how it may be implemented within your project.

### Citation Details

*Choosing A Suitable Digital Rights Solution*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-22/>>

**Keywords:** digitisation, digital rights, DRM, protect, watermark, briefing

## Background

Digital Rights Management (DRM) refers to any method for a designer to monitor, control, and protect digital content. It was developed primarily as an advanced anti-piracy method to prevent illegal or unauthorised distribution of digital data. Common examples of DRM include watermarks, licences, and user registration.

This document provides criteria for assessing a project's requirements for Digital Rights and guidance for choosing an appropriate solution.

## Do I Need Digital Rights Management?

Digital Rights Management is not appropriate for all projects. Some projects may find it useful to protect digital software or content, others may find it introduces unnecessary complexity into the development process, limit use and cause unforeseen problems at a later date.

Possible reasons for a project to implement DRM may include:

- You wish to identify digital content as your own work (i.e. via copyright notices).
- You are required to notify users of specific licence conditions.
- You wish to identify the users of your site and to track usage.

Before implementing a solution, you should that a) there is a convincing argument to implement digital rights within your project, and b) you possess sufficient time and finances to implement digital rights.

## DRM Workflow

To ensure Digital Rights are implemented in a consistent and planned manner, the project should establish a six-stage workflow that identifies the rights held and the method of protecting them.

1. **Recognition of rights:** Identify who holds rights and the type of rights held.
2. **Assertion of rights:** Identify legal framework or specific licensing conditions that must be considered.
3. **Expression of rights:** Provide human and machine-readable representation of these rights.
4. **Dissemination of rights:** Identify methods of storing rights information about the object?
5. **Exposure of rights** – How are rights to be made visible to the user?

6. **Enforcement of rights** – Identify the methods that will be used to legally enforce rights ownership.

### Expression And Dissemination Of Rights

The options available to express, disseminate and expose Rights information require an understanding of several factors:

- The type of content you wish to protect
- The technical measures available to protect the content.
- The purpose and type of protection that you wish to impose.

Projects in the education-sector are likely to require some method of establishing their rights, rather than restrict use. Self-describing techniques may be used to establish copyright ownership for digital derivatives (still images, audio, video) through a watermark, internal record (e.g. EXIF JPEG, TIFF) or unique code hidden within the file, or stored separately within a digital repository as a metadata record. Authors are encouraged to use the University Copyright Convention as a template:

© [name of copyright proprietor] [year of creation]

### Interoperability

To ensure Digital Rights can be identified and retrieved at a later date, data should be stored in a standard manner. It is therefore wise to be consistent when storing copyright information for a large number of files. Possible options are to store copyright notices in background noise of digital images or within readily identifiable elements within the metadata schema. The Dublin Core *Rights Management* element is a simple method to disseminate copyright notices when harvesting metadata for e-prints. Complex metadata schemas for media interchange, such as the eXtensible Media Commerce Language (XMCL), offer copyright information at an increased granularity by identifying rental, subscription, ownership, and video on demand/pay-per-view services. The XrML (eXtensible rights Markup Language) may also prove useful as a general-purpose grammar for defining digital rights and conditions to be associated with digital content, services, or other resources. The language is utilized as the basis for the MPEG-21 and Open eBook rights specifications.

### Summary

The implementation of Digital Rights is often costly and time-consuming. However, it does provide real benefits by establishing copyright ownership and providing restrictions on the possible uses. The project should choose the protection method that can be implemented within budget, without interfering with legitimate use.

### Further Information

- *Athens - Access Management Systems*, <<http://www.athens.ac.uk/>>
- *Directory for Social Issues in computing – Copy Protection*, <[http://www.rajivshah.com/directory/Content/Protection\\_Schemes/Windows/](http://www.rajivshah.com/directory/Content/Protection_Schemes/Windows/)>
- *How Much Is Stronger DRM Worth?*, Lewis, <[http://www.cpppe.umd.edu/rhsmith3/papers/Final\\_session1\\_lewis.pdf](http://www.cpppe.umd.edu/rhsmith3/papers/Final_session1_lewis.pdf)>
- *XMCL*, <<http://www.w3.org/TR/2002/NOTE-xmcl-20020919/>>
- *XrML*, <<http://www.xrml.org/>>



# Recording Digital Sound

### About This Document

This briefing document describes the influence of sample rate, bit-rate and file format upon digital audio and provides criteria for assessing their suitability for a specific purpose.

### Citation Details

*Recording Digital Sound*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-23/>>

**Keywords:** digitisation, recording sound, sample rate, bit-rate, encoding, briefing

## Background

The digitisation of digital audio can be a complex process. This document contains quality assurance techniques for producing effective audio content, taking into consideration the impact of sample rate, bit-rate and file format.

## Sample Rate

Sample rate defines the number of samples that are recorded per second. It is measured in Hertz (cycles per second) or Kiloherz (thousand cycles per second). The following table describes four common benchmarks for audio quality. These offer gradually improving quality, at the expense of file size.

Samples per second	Description
8kHz	Telephone quality
11kHz	At 8 bits, mono produces passable voice at a reasonable size.
22kHz	22k, half of the CD sampling rate. At 8 bits, mono, good for a mix of speech and music.
44.1kHz	Standard audio CD sampling rate. A standard for 16-bit linear signed mono and stereo file formats.

**Table 1: Description Of The Various Sample Frequencies Available**

The audio quality will improve as the number of samples per second increases. A higher sample rate enables a more accurate reconstruction of a complex sound wave to be created from the digital audio file. To record high quality audio a sample rate of 44.1kHz should be used.

## Bit-rate

Bit-rate indicates the amount of audio data being transferred at a given time. The bit-rate can be recorded in two ways – variable or constant. A variable bit-rate creates smaller files by removing inaudible sound. It is therefore suited to Internet distribution in which bandwidth is a consideration. A constant bit-rate, in comparison, records audio data at a set rate irrespective of the content. This produces a replica of an analogue recording, even reproducing potentially unnecessary sounds. As a result, file size is significantly larger than those encoded with variable bit-rates. Table 2 indicates how a constant bit-rate affects the quality and file size of an audio file.

Bit rate	Quality	MB/min
1411	CD audio	10.584
192	Near CD quality	1.440
128	Typical music level	0.960
112	Digital radio quality	0.840
64	FM quality	0.480
32	AM quality	0.240
16	Short-wave quality	0.120

**Table 2. Indication Of Audio Quality Expected With Different Bit-Rates**

### Digital Audio Formats

The majority of audio formats use lossy compression to reduce file size by removing superfluous audio data. Master audio files should ideally be stored in a lossless format to preserve all audio data.

Format	Compression	Streaming support	Bit-rate	Popularity
MPEG Audio Layer III (MP3)	Lossy	Yes	Variable	Common on all platforms
Mp3PRO (MP3)	Lossy	Yes	Variable	Limited support.
Ogg Vorbis (OGG)	Lossy	Yes	Variable	Limited support.
RealAudio (RA)	Lossy	Yes	Variable	Popular for streaming.
Microsoft wave (WAV)	Lossless	No	Constant	Primarily for MS Windows
Windows Media (WMA)	Lossy	Yes	Variable	Primarily for MS Windows

**Table 3. Common Digital Audio Formats**

Conversion between digital audio formats can be complex. If you are producing audio content for Internet distribution, a lossless-to-lossy (e.g. WAV to MP3) conversion will significantly reduce bandwidth usage. Only lossless-to-lossy conversion is advised. The conversion process of lossless-to-lossless will further degrade audio quality by removing additional data, producing unexpected results.

### What Is The Best Solution?

Whether digitising analogue recordings or converting digital sound into another format, sample rate, bit rate and format compression will affect the resulting output. Quality assurance processes should compare the technical and subjective quality of the digital audio against the requirements of its intended purpose.

A simple suite of subjective criteria should be developed to check the quality of the digital audio. Specific checks may include the following questions:

- Can listeners understand voices in recording?
- Can listeners hear quiet sounds?
- Can listener hear loud sounds without distortion?
- Can the listener distinguish between digitised audio and original recording?

Objective technical criteria should also be measured to ensure each digital audio file is of consistent or appropriate quality:

- Is there a documented workflow for creating the digital audio files?
- Is the file format and software used to compress the audio documented?
- Is the bit rate equal to or less than the available bandwidth?
- Does the sample and bit-rate of the digital audio match or exceed that of the original analogue recording (or is the loss of quality acceptable, see subjective tests above)?
- For accurate reproduction of an original analogue recording, is the digital audio master file stored in a lossless format?
- For accurate reproduction of the original sound is the sample rate at least twice that of the highest frequency sound?

### Further Information

- *MP3Pro Zone*,  
<<http://www.mp3prozone.com/>>
- *Measuring Audio Quality*,  
<<http://www.itworld.com/AppDev/1469/NWW1204revside4/>>
- *Ogg Vorbis*,  
<<http://www.vorbis.com/>>
- *PC Recording*,  
<<http://www.pcrecording.com/>>
- *Real Networks*,  
<<http://www.real.com>>
- *Xorys' MP3 FAQ*,  
<<http://webhome.idirect.com/~nuzhathl/mp3-faq.html>>

# Handling International Text

### About This Document

This briefing document describes common problems that occur when handling international text and methods of resolving them.

### Citation Details

*Handling International Text*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-24/>>

**Keywords:** digitisation, international text, Latin, ISO 10646, UTF-8, Unicode, encoding, briefing

## Background

Before the development of Unicode there were hundreds of different encoding systems that specific languages, but were incompatible with one another. Even for a language like English no single encoding was adequate for all the letters, punctuation, and technical symbols in common use.

Unicode avoids the language conversion issues of earlier encoding systems by providing a unique number for every character that is consistent across platforms, applications and language. However, there remain many issues surrounding its uses. This paper describes methods that can be used to assess the quality of encoded text produced by an application.

## Conversion to Unicode

When handling text it is useful to perform quality checks to ensure the text is encoded to ensure more people can read it, particularly if it incorporates foreign or specialist characters. When preparing an ASCII file for distribution it is recommended that you check for corrupt or random characters. Examples of these are shown below:

- Text being assigned random characters.
- Text displaying black boxes.

!^\*\$%'"&■  
■■■■■ ■■■

To preserve long-term access to content, you should ensure that ASCII documents are converted to Unicode UTF-8. To achieve this, various solutions are available:

1. **Upgrade to a later package** Documents saved in older versions of the MS Word or Word Perfect formats can be easily converted by loading them into later (Word 2000+) versions of the application and resaving the file.
2. **Create a bespoke solution** A second solution is to create your own application to perform the conversion process. For example, a simple conversion process can be created using the following pseudo code to convert Greek into Unicode:
  1. Find the ASCII value
  2. If the value is > 127 then
  3. Find the character in \$Greek737 ' DOS Greek
  4. Replace the character with the character in Unicode at that position
  5. End if
  6. Repeat until all characters have been done

7. Alternatively, it may be simpler to substitute the DOS Greek for \$GreekWIN.
3. **Use an automatic conversion tool** Several conversion tools exist to simplify the conversion process. Unifier (Windows) and Sean Redmond's Greek - Unicode converter (multi-platform) have an automatic conversion process, allowing you to insert the relevant text, choose the source and destination language, and convert.

### Ensure That You Have The Correct Unicode Font

Unicode may provide a unique identifier for the majority of languages, but the operating system will require the correct Unicode font to interpret these values and display them as glyphs that can be understood by the user. To ensure a user has a suitable font, the URL <<http://www.columbia.edu/kermit/utf8.html>> demonstrates a selection of the available languages:

If the client is missing a UTF-8 glyph to view the required language, they can be downloaded from <<http://www.alanwood.net/unicode/fonts.html>>.

### Converting Between Different Character Encoding

Character encoding issues are typically caused by incompatible applications that use 7-bit encoding rather than Unicode. These problems are often disguised by applications that “enhance” existing standards by mixing different character sets (e.g. Windows and ISO 10646 characters being merged into a ISO Latin document). Although these have numerous benefits, such as allowing Unicode characters to be displayed in HTML, they are not widely supported and can cause problems in other applications. A simple example can be seen below – the top line is shown as it would appear in Internet Explorer, the bottom line shows the same text displayed in another browser.

Microsoft Windows <sup>TM</sup>!  
Microsoft Windows   !

Although this improves the attractiveness of the text, the non-standard approach causes some information to be lost.

When converting between character encoding you should be aware of limitations of the character encoding.

Although 7-bit ASCII can map directly to the same code number in UTF-8 Unicode, many existing character encodings, such as ISO Latin, have well documented issues that limit their use for specific purposes. This includes the designation of certain characters as ‘illegal’. For example, the capital Y umlaut and a florin symbol. When performing the conversion process, many non-standard browsers save these characters through the range 0x82 through 0x95- that is reserved by Latin-1 and Unicode for additional control characters. Manually searching a document in a Hex editor for these values and examining the character associated with them, or the use of a third-party utility to convert them into a numerical character can resolve this.

### Further Information

- *Alan Wood's Unicode resources*, <<http://www.alanwood.net/unicode/>>
- *Unicode Code Charts*, <<http://www.unicode.org/charts/>>
- *Unifier Converter (Windows)*, <<http://www.melody-soft.com/>>

- *Sean Redmond's Greek - Unicode converter multi-platform CGI*,  
<<http://www.jiffycomp.com/smr/unicode/>>
- *On the Goodness of Unicode*,  
<<http://www.tbray.org/ongoing/When/200x/2003/04/06/Unicode>>
- *On the use of some MS Windows Characters in HTML*,  
<<http://www.cs.tut.fi/~jkorpela/www/windows-chars.html>>

# Choosing A Suitable Digital Video Format

### About This Document

This briefing document provides criteria for choosing the most appropriate method of storing digital video, by taking into account the compression rate, bandwidth requirements and special features offered by differing file formats.

### Citation Details

*Choosing A Suitable Digital Video Format*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-25/>>

**Keywords:** digitisation, digital video, bandwidth, distribution method, streaming, briefing

## Background

Digital video can have a dramatic impact upon the user. It can reflect information that is difficult to describe in words alone, and can be used within an interactive learning process. This document contains guidelines to best practice when manipulating video. When considering the recording of digital video, the digitiser should be aware of the influence of file format, bit-depth, bit-rate and frame size upon the quality of the resulting video.

## Choosing The Appropriate File Format

When choosing a file format for digital video, the following questions should be asked:

1. What type of distribution method will be used to deliver video?
2. What type of users are you aiming the video towards?
3. Do you wish to edit the video at a later stage?

The distribution method will have a significant influence upon the file format chosen. Digital video intended for static media (CD-ROM, DVD) are suited to progressive encoding methods that do not require extensive error checks. Video intended for Internet distribution should be encoded using one of the streaming formats. Streaming enables the viewer to watch the video after just a few seconds, rather than waiting for a download to complete. Quality is significantly lower than progressive formats due to compression being used.

Secondly, you should consider your target audience. Many computer users are, for various reasons, unable to view many digital video formats. If content is intended for Windows users primarily, a Microsoft streaming format (ASF and WMV) may be used. However, access may be difficult for Mac and Linux systems, which may prevent limit use. If the intent is to attract as many viewers as possible, an alternative cross-platform solution should be chosen. Possible formats include QuickTime, QuickTime Pro and RealMedia.

Finally, you should consider the project's needs for the digital video. Few compressed formats offer the ability to edit it extensively at a later date, so it will be important to store a master copy of the video in a format that supports spatial encoding. MJPEG spatial compression is one of the few mainstream examples that support this feature.

To summarise, Table 1 shows the appropriateness of different file formats for streaming or progressive recording.

NAME	PURPOSE OF MEDIA		
	Streaming	Progressive	Media
Advanced Streaming Format (ASF)	Y	N	
Audio Video Interleave (AVI)	N	Y	
MPEG-1	N	Y	VideoCD
MPEG-2	N	Y	DVD
QuickTime (QT)	Y	Y	
RealMedia (RM)	Y	Y	
Windows Media Video (WMV)	Y	Y	
DivX	N	Y	Amateur CD distribution
MJPEG	N	Y	

**Table 1: Intended purpose and compression used by different file formats**

### Video Quality

When creating digital video for a specific purpose (Internet, CD-ROM, DVD-Video), you should balance the desires for video quality (in terms of frame size, frame rate & bit-depth) with the facilities available to the end user. For reasons relating to file size and bandwidth, it is not always possible to provide the viewer with high-quality digital output. Static media (CD-ROM, DVD) are limited in their amount of data they can store. The creation of streaming video for Internet usage must also consider bandwidth usage. The majority of Internet content uses an 8-bit screen of 160 x 120 pixels, at 10-15 frames per second. Table 2 demonstrates how the increase in screen size, bit-depth and frames-per-second will affect the file size.

Screen Size	Pixels per frame	Bit depth (bits)	Frames per second	Bandwidth required per second (megabits)	Possible Purpose
640 x 480	307,200	24	30	221.184	DVD-Video
320 x 240	76,800	16	25	30.72	CD-ROM
320 x 240	76,800	8	15	9.216	CD-ROM
160 x 120	19,200	8	10	1.536	Broadband
160 x 120	19,200	8	5	0.768	Dial-up

**Table 2: Influence screen size, bit-depth and frames per second has on bandwidth**

### Potential Problems

When creating digital video, the designer should be aware of three problems:

1. **Hardware requirements** Captured digital video is often large and will require a large hard disk and sufficient amount of memory to edit and compress.
2. **Inability to decode video/audio stream** The user often requires third-party decoders to view digital video. Common problems include error messages, audio playback without the video, and corrupted treacle-like video. It is useful to inform the user of the format in which the video is saved and direct them to the relevant web site if necessary.



3. **Synchronicity** – Audio and video is stored as two separate data streams and may become out of sync- an actor will move their mouth, but the words are delayed by two seconds. To resolve the problem, editing software must be used to resynchronise the data.

### Further Information

- *Advanced Streaming Format (ASF)*,  
<<http://www.microsoft.com/windows/windowsmedia/format/asfspec.aspx>>
- *Apple Quicktime*, <<http://www.apple.com/quicktime/>>
- *DIVX*, <<http://www.divx.com/>>
- *Macromedia Flash*, <<http://www.macromedia.com/>>
- *MPEG working group*, <<http://www.chiariglione.org/mpeg/index.htm>>
- *Real Networks*, <<http://www.real.com>>
- *Microsoft Windows Media*,  
<<http://www.microsoft.com/windows/windowsmedia/default.aspx>>

# Implementing Quality Assurance For Digitisation

### About This Document

This briefing document describes techniques for implementing quality assurance within your digitisation project.

### Citation Details

*Implementing Quality Assurance For Digitisation*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-27/>>

**Keywords:** digitisation, audit, checks, briefing

## Background

Digitisation often involves working with hundreds or thousands of images, documents, audio clips or other types of source material. Ensuring these objects are consistently digitised and to a standard that ensures they are suitable for their intended purpose can be complex. Rather than being considered as an afterthought, quality assurance should be considered as an integral part of the digitisation process, and used to monitor progress against quality benchmarks.

## Quality Assurance Within Your Project

The majority of formal quality assurance standards, such as ISO9001, are intended for large organisations with complex structures. A smaller project will benefit from establishing its own quality assurance procedures, using these standards as a guide. The key is to understand how work is performed and identify key points at which quality checks should be made. A simple quality assurance system can then be implemented that will enable you to monitor the quality of your work, spot problems and ensure the final digitised object is suitable for its intended use.

The ISO 9001 identifies three steps to the introduction of a quality assurance system:

1. **Brainstorm:** Identify specific processes that should be monitored for quality and develop ways of measuring the quality of these processes. You may want to think about:
  - *Project goals:* who will use the digitised objects and what function will they serve.
  - *Delivery strategy:* how will the digitised objects be delivered to the user? (Web site, Intranet, multimedia presentation, CD-ROM).
  - *Digitisation:* how will data be analysed or created. To ensure consistency throughout the project, all techniques should be standardized.
2. **Education:** Ensure that everyone is familiar with the use of the system.
3. **Improve:** Monitor your quality assurance system and look for problems that require correction or other ways it may be improved.

## Key Requirements For A Quality Assurance System

First and foremost, any system for assuring quality in the digitisation process should be straightforward and not impede the actual digitisation work. Effective quality assurance can be achieved by performing four processes during the digitisation lifecycle:

1. The key to a successful QA process is to establish a clear and concise work timeline and, using a step-by-step process, document on how this can be achieved. This will provide a baseline against which actual work can be checked, promoting consistency, and making it easier to spot when digitisation is not going according to plan.
2. Compare the digital copy with the physical original to identify changes and ensure accuracy. This may include, but is not limited to, colour comparisons, accuracy of text that has been scanned through OCR software, and reproduction of significant characteristics that give meaning to the digitised data (e.g. italicised text, colours).
3. Perform regular audit checks to ensure consistency throughout the resource. Qualitative checks can be performed upon the original and modified digital work to ensure that any changes were intentional and processing errors have not been introduced. Subtle differences may appear in a project that takes place over a significant time period or is divided between different people. Technical checks may include spell checkers and the use of a controlled vocabulary to allow only certain specifically designed descriptions to be used. These checks will highlight potential problems at an early stage, ensuring that staff are aware of inconsistencies and can take steps to remove them. In extreme cases this may require the re-digitisation of the source data.
4. Finally, measures should be taken to establish some form of audit trail that tracks progress on each piece of work. Each stage of work should be ‘signed off’ by the person responsible, and any unusual circumstances or decisions made should be recorded.

The ISO 9001 system is particularly useful in identifying clear guidelines for quality management.

### Summary

Digitisation projects should implement a simple quality assurance system. Implementing internal quality assurance checks within the workflow allows mistakes to be spotted and corrected early-on, and also provides points at which work can be reviewed, and improvements to the digitisation process implemented.

### Further Information

- *TASI Quality Assurance*, <<http://www.tasi.ac.uk/advice/creating/pdf/qassurance.pdf>>

# Choosing An Appropriate Raster Image Format

### About This Document

This briefing document describes factors to consider when choosing a raster image format for archival and distribution.

### Citation Details

*Choosing An Appropriate Raster Image Format*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-28/>>

**Keywords:** digitisation, raster, image, bit-depth, lossless, lossy, compression, briefing

## Background

Any image that is to be archived for future use requires specific storage considerations. However, the choice of file format is diverse, offering advantages and disadvantages that make them better suited to a specific environment. When digitising images a standards-based and best practice approach should be taken, using images that are appropriate to the medium they are used within. For disseminating the work to others, a multi-tier approach is necessary, to enable the storing of a preservation and dissemination copy. This document will discuss the formats available, highlighting the different compression types, advantages and limitations of raster images.

## Factors To Consider When Choosing Image Formats

When creating raster-based images for distribution file size is the primary consideration. As a general rule, the storage requirements increase in proportion to the improvement in image quality. A side effect of this process is that network delivery speed is halved, limiting the amount that can be delivered to the user. For Internet delivery it is advised that designers provide a small image (30-100k) that can be accessed quickly for mainstream users, and provide a higher quality copy as a link or available on a CD for professional usage.

When digitising the designer must consider three factors:

1. File format
2. Bit-depth
3. Compression type.

## Distribution Methods

The distribution method will have a significant influence upon the file format, encoding type and compression used in the project.

- **Photograph archival** For photographs, the 24-bit lossless TIFF format is recommended to allow the image to be reproduced accurately. The side-effect is that file sizes will begin at 10Mb for simpler images and increase dramatically. This is intended for storage only, not distribution.
- **Photograph distribution** For photographs intended for Internet distribution, the lossy JPEG format is recommended. This uses compression to reduce file size dramatically. However, image quality will decrease.

## 4.2 Technical Advice: Digitisation – Case Studies

- **Simpler images** Simpler images, such as cartoons, buttons, maps or thumbnails, which do not require 16.8 million colours should be stored in an 8-bit format, such as GIF or PNG-8. Though 256 colours images can be stored correctly in a 24-bit format, a side effect of this process is the 8-bit file size is often equal or higher than 24-bit images.

To summarise, Table 1 shows the appropriateness of different file formats for streaming or progressive recording.

	Max. no. of colours	Compression Type	Suited for:	Issues
<b>BMP</b>	16,777,216	None	General usage. Common on MS Windows platforms	MS Windows rather than Internet format. Unsupported by most browsers.
<b>GIF87a</b>	256	Lossless	High quality images that do not require photographic details	File sizes can be quite large, even with compression
<b>GIF89a</b>	256	Lossless	Same as GIF87a, animation facilities are also popular	See above
<b>JPEG</b>	16,777,216	Lossy	High quality photographs delivered in limited bandwidth environment.	Degrades image quality and produces wave-like artefacts on image.
<b>PNG-8</b>	256	Lossless	Developed to replace GIF. Produces 10-30% smaller than GIF files.	File sizes can be large, even with compression.
<b>PNG-24</b>	16,777,216	Lossless	Preserves photograph information	File sizes larger than JPG.
<b>TIFF</b>	16,777,216	Lossless	Used by professionals. Redundant file information provides space for specialist uses (e.g. colorimetry calibration). Suitable for archival material.	Unsuitable for Internet-delivery

**Table 1: Comparison Table Of Image File Formats**

Once chosen, the file format will, to a limited extent, dictate the possible file size, bit depth and compression method available to the user.

### Compression Type

Compression type is a third important consideration for image delivery. As the name suggests, compression reduces file size by using specific algorithms. Two compression types exist:

1. **Lossless compression** Lossless compression stores colour information and the location of the pixel with which the colour is associated. The major advantage of this compression method is the image can be restored to its original state without loss of information (hence lossless). However, the compression ratios are not as high as lossy formats, typically reducing file sizes by half. File formats that use this compression type include PNG and GIF.
2. **Lossy compression** Offers significantly improved compression ratio, at the expense of image quality. Lossy compression removes superfluous image information that cannot be regained. The degree of quality loss will depend upon the amount of compression applied to the image (e.g., JPEG uses a percentage system to determine the amount of compression). Therefore it is possible to create an image that is 1:100 the size of the original file

As an archival format, lossy compression is unsuitable for long-term preservation. However, its small file size is used in many archives as a method of displaying lower resolution images for Internet users.

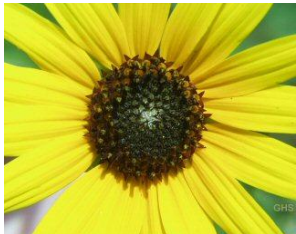



### Bit-depth

Bit-depth refers to the maximum number of colours that can be displayed in an image. The number of colours available will rise when the bit depth is increased. Table 2 describes the relationship between the bit depth and number of colours.

Bit depth	1	4	8	8	16	24	32
Maximum No. of colours	2	16	256	256	65,536	16,777,216	16,777,216

**Table 2: Relationship Between Bit-Depth And Maximum Number Of Colours**

The reduction of bit-depth will have a significant effect upon image quality. Figure 3 demonstrates the quality loss that will be encountered when saving at a low bit-depth.

	Bit-depth	Description
	24-bit	Original image
	8-bit	Some loss of colour around edges. Suitable for thumbnail images
	4-bit	Major reduction in colours. Petals consist almost solely of a single yellow colour.
	1-bit	Only basic layout data remains.

**Figure 3: Visual Comparison Of Different Bit Modes**

### Image Conversion Between Different Formats

Image conversion is possible using a range of applications (Photoshop, Paint Shop Pro, etc.). Lossless-to-lossless conversion (e.g. PNG-8 to GIF89a) can be performed without

quality loss. However, lossless-to-lossy (PNG-8 to JPEG) or lossy-to-lossy conversion will result in a quality loss, dependent upon the degree of compression used. For dissemination of high-quality images, a lossy format is recommended to reduce file size. Smaller images can be stored in a lossless format.

### Further Information

- *Gimp-Savvy*,  
<<http://gimp-savvy.com/cgi-bin/img.cgi?ufwsEFfOvPeIueo740>>
- *Raster Image Files*, <<http://scv.bu.edu/Tutorials/ImageFiles/image101.html>>

## Choosing A Vector Graphics Format For The Internet



### About This Document

This briefing document offers issues to consider when choosing an appropriate vector graphics format.

### Citation Details

*Choosing A Vector Graphics Format For The Internet*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-29/>>

**Keywords:** digitisation, vector, graphics, briefing

## Background

Vector graphics offer many benefits, allowing the screen size to be resized without the image becoming jagged or unrecognisable. However, there is often confusion on the correct format for the task. This document describes the suitable file formats available and the conventions that should be followed when editing vector files.

## Project QA

At the start of development it may help to ask your team the following questions:

1. What type of information will the graphics convey? (Still images, animation and sound, etc.)
2. Will the target audience be able to access and decode the content? (Older browsers and non PC browsers may have limited for XML languages.)
3. Will the format require migration after a few years?

The format that you choose should meet 2 or more of the criteria associated with these questions.

## File Formats

The choice of a vector-based file format should be derived from three different criteria: intended use of the format, availability of viewers and availability of specification. Several vector formats exist for use on the Internet. These construct information in a similar way yet provide different functionality:

Format	Availability		Viewers			Uses
	Open Standard	Proprietary	Internet browser	Browser plug-in	Application	
Scalable Vector Graphics (SVG)	✓		✓			Internet-based graphics
Shockwave / Flash		✓		✓		Multimedia requiring sound, video & text
Vector Markup Language (VML)	✓		✓		✓	GenericXML markup.
Windows Meta File (WMF)		✓			✓	Clipart

**Table 1: Summary Of vector Graphics Formats**

For Internet delivery of static images, the World Wide Web Consortium recommends SVG as a open standard for vector diagrams. Shockwave and Flash are also common if



the intent is to provide multimedia presentation, animation and audio. VML is also common, being the XML language exported by Microsoft products.

### XML Conventions

Although XML enables the creation of a diversity of data types it is extremely meticulous regarding syntax usage. To remain consistent throughout multiple documents and avoid future problems, several conventions are recommended:

1. Lower case should be used through. Capitalisation can be used for tags if it is consistent throughout the document.
2. Indent buried tags to reduce the time required for a user to recognise groups of information.
3. Avoid the use of acronyms or other tags that will be unintelligible for anyone outside the project. XML is intended as a human readable format, so obvious descriptions should be used whenever possible.
4. Avoid the use of white space when defining tags. If two word descriptions are necessary, join them via a hyphen (-). Otherwise concatenate the words by typing the first word in lower case, and capitalising subsequent words. For example, a creation date property would be called 'fileDateCreated'.

### Further Information

- *Official W3 SVG site*, W3C, <<http://www.w3.org/Graphics/SVG/Overview.htm>>
- *An Introduction to VML*,  
<<http://www.infoloom.com/gcaconfs/WEB/chicago98/wu.HTM>>
- *Flash and Shockwave*, Macromedia,  
<<http://www.macromedia.com/>>

---

# Transcribing Documents

### About This Document

This briefing document describes techniques to ensure transcribed documents are consistent and avoid common errors.

### Citation Details

*Transcribing Documents*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-47/>>

**Keywords:** digitisation, transcribing, briefing

## Digitising Text by Transcription

Transcription is a very simple but effective way of digitising small to medium volumes of text. It is particularly appropriate when the documents to be digitised have a complex layout (columns, variable margins, overlaid images etc.) or other features that will make automatic digitisation using OCR (Optical Character Recognition) software difficult. Transcription remains the best way to digitise hand written documents.

## Representing the Original Document

All projects planning to transcribe documents should establish a set of transcription guidelines to help ensure that the transcriptions are complete, consistent and correct.

Key issues that transcription guidelines need to cover are:

- What to do about illegible text
- How to record important information indicated by position, size, italics, bold or other visual features of the text
- What to do about accents, non-Latin characters and other language issues

It is generally good practice to **not** correcting factual errors or mistakes of grammar or spelling in the original.

## Avoiding Errors

Double-entry is the best solution – where two people separately transcribe the same document and the results are then compared. Two people are unlikely to make the same errors, so this technique should reveal most errors. It is, however often impractical because of the time and expense involved. Running a grammar and spell checker over the transcribed document is a simpler way of finding many errors (but assumes the original document was spelt and written according to modern usage).

## Transcribing Structured Documents

Structured documents, such as census returns or similar tabular material may be better transcribed into a spreadsheet package rather than a text editor. When transcribing tables of numbers, a simple but effective check on accuracy is to use a spreadsheet to calculate row and column totals that can be compared with the original table.

Transcriber guidelines for this type of document will need to consider issues such as:

## 4.2 Technical Advice: Digitisation – Case Studies

---

- What to do about ‘ditto’ and other ways of referring to an earlier entry in a list or table – should the value or the placeholder be transcribed?
- Should incorrect values be transcribed ‘as is’

It is good practice to record values, such as weights, distances, money and ages as they are found, but also to include a standardised representation to permit calculations (e.g. ‘baby, 6m’ should be transcribed verbatim, but an addition entry of 0.5, the age in years, could also be entered).

### Further Information

Many genealogical groups transcribe documents, and provide detailed instructions. Examples include:

- *The USGenWeb Census Project*,  
<<http://www.us-census.org/help/1910.html>>
- *The Immigrant Ships Transcribers Guild*,  
<<http://www.immigrantships.net/guild/typing2.html>>

# Digitising Data For Preservation

### About This Document

This briefing document describes QA techniques for improving the longevity of digital data and ensuring that content does not become inaccessible over time.

### Citation Details

*Digitising Data For Preservation*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-62/>>

**Keywords:**, digitisation, digitising, preservation, metadata, briefing

## Background

Digital data can become difficult to access in a matter of a few years. Technical obsolescence due to changes in hardware, software and standards, as well as media degradation can all affect the long-term survival of digital data.

The key to preserving digital data is to consider the long-term future of your data right from the moment it is created.

## Digitising for Preservation

Before beginning to digitise material you should consider the following issues:

1. What tools can I use to produce the content?
2. What file formats will be outputted by the chosen tools?
3. Will I have to use the specific tools to access the content?
4. What is the likelihood that the tools will be available in five years time?

The answer to these questions will vary according to the type of digitisation you are conducting, and the purpose of the digitised content. However, it is possible to make suggestions for common areas:

**Documents** – Documents that contain pure text or a combination of text and pictures can be saved in several formats. Avoid the use of native formats (MS Word, WordPerfect, etc.) and save them in Rich Text Format – a platform independent format that can be imported and exported in numerous applications.

**Images** – The majority of image formats are not proprietary, however they can be ‘lossy’ (i.e. remove image details to save file size). When digitising work for preservation purposes you should use a lossless format, preferably TIFF or GIF. JPEG is a lossy format and should be avoided.

**Audio** – Like images, audio is divided between lossless and lossy formats. Microsoft Wave is the most common format for this purpose, while MP3 is entirely unsuitable.

**Video** – Video is controversial and changes on a regular basis. For preservation purposes it is advisable to use a recognised standard, such as MPEG-1 or MPEG-2. These provide poor compression in comparison to QuickTime or DIVX, but are guaranteed to work without the need to track a particular software revision or codec.

### Preservation and Content Modification

It may be necessary to modify content at some point. This may be for a variety of reasons: the need to migrate to a new preservation format or production of distribution copies. At this stage there are two main considerations:

1. Do not modify the original content, create a copy and work on that.
2. Create a detailed log that outlines the differences between the original and the modified copy.

The extent of the detailed log is dependent upon your needs and the time period in which you have chosen to create it. A simple modification ‘log’ can consist of a text file that describes the modification, the person who performed it, when it was performed, and the reason for the changes. A more complex system could be encoded in XML and available online for anyone to access. Examples of both these solutions can be seen below.

#### A simple text file

**Data Conversion**

Description of the conversion process undertaken on the main data.

**Documentation Conversion**

Description of the conversion process undertaken on associated documentation

**Altered File Names**

Indication of changed file names that may differ between the original and modified version.

**Date**

The date on which the process was undertaken. Useful for tracking.

**Responsible Agent**

The person responsible for making the described changes.

#### TEI schema revision data

```
<revisionDesc>
<change>
<date>2002-02-07</date>
<respStmt>
<name>Colley, Greg</name>
<resp>Cataloguer</resp>
</respStmt>
<item>Header recomposed with TEI XML
header</item>
</change>
<change>
<date>1998-01-14</date>
<respStmt>
<name>Burnard</name>
<resp>Converter</resp>
</respStmt>
<item>Automatic conversion from OTA DTD to
TEI lite DTD</item>
</change>
</revisionDesc>
```

### Further Information

- *The Arts and Humanities Data Service*, <<http://ahds.ac.uk/>>
- *Technical Advisory Service for Images*, <<http://www.tasi.ac.uk/>>

# Audio For Low-Bandwidth Environments

### About This Document

This briefing document identifies criteria to consider when recording digital audio for a limited-bandwidth environment, such as those encountered by dial-up Internet users and mobile phones.

### Citation Details

*Audio For Low-Bandwidth Environments*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-65/>>

**Keywords:** digitisation, digitising, bandwidth, lossy, streaming, bit-rate, briefing

## Background

Audio quality is surprisingly difficult to predict in a digital environment. Quality and file size can depend upon a range of factors, including vocal type, encoding method and file format. This document provides guidelines on the most effective method of handling audio.

## Factors To Consider

When creating content for the Internet it is important to consider the hardware the target audience will be using. Although the number of users with a broadband connection is growing, the majority of Internet users utilise a dial-up connection to access the Internet, limiting them to a theoretical 56kbps (kilobytes per second). To cater for these users, it is useful to offer smaller files that can be downloaded faster.

The file size and quality of digital audio is dependent upon two factors:

1. File format
2. Type of audio

By understanding how these three factors contribute to the actual file size, it is possible to create digital audio that requires less bandwidth, but provides sufficient quality to be understood.

## File Format

File format denotes the structure and capabilities of digital audio. When choosing an audio format for Internet distribution, a lossy format that encodes using a variable bit-rate is recommended. Streaming support is also useful for delivering audio data over a sustained period without the need for an initial download. These formats use mathematical calculations to remove superfluous data and compress it into a smaller file size. Several popular formats exist, many of which are household names. MP3 (MPEG Audio Layer III) is popular for Internet radio and non-commercial use. Larger organisations, such as the BBC, use Real Audio (RA) or Windows Media Audio (WMA), based upon its digital rights support.

The table below shows a few of the options that are available.

Format	Compression	Streaming	Bit-rate
MP3	Lossy	Yes	Variable
Mp3PRO	Lossy	Yes	Variable
Ogg Vorbis	Lossy	Yes	Variable

RealAudio	Lossy	Yes	Variable
Windows Media Audio	Lossy	Yes	Variable

**Figure 1: File Formats Suitable For Low-Bandwidth Delivery**

Once recorded audio is saved in a lossy format, it is wise to listen to the audio data to ensure it is audible and that essential information has been retained.

Finally, it is recommended that a variable bit-rate is used. For speech, this will usually vary between 8 and 32kbps as needed, adjusting the variable rate accordingly if incidental music occurs during a presentation.

### Choosing An Appropriate Encoding Method

The audio quality required, in terms of bit-rate, to record audio data is influenced significantly by the type of audio that you wish to record: music or voice.

- **Music** Music data is commonly transmitted in stereo and will vary significantly from one second to the next. A sampling rate of 32-64kHz is appropriate for low-bandwidth environments, allowing users to listen to streamed audio without significant disruption to other tasks.
- **Voice** Voice is less demanding than music data. The human voice has a limited range, usually reaching 3-4kHz. Therefore, an 8-15kHz sampling rate and 8-32kbps bit-rate is enough to maintain good quality. Mono audio, transmitted through a single speaker, will also be suitable for most purposes. Common audio players ‘double’ the audio content, transmitting mono channel data as stereo audio through two speakers. This is equivalent to a short-range or AM radio, providing a good indication of the audio quality you can expect. By using these methods, the user can reduce file size for voice content by 60%+ in comparison to recording at a higher bit-rate without loss of quality.

### Assessing Quality Of Audio Data

The creation of audio data for low-bandwidth environments does not necessitate a significant loss in quality. The audio should remain audible in its compressed state. Specific checks may include the following questions:

- Can listeners understand voices in recording?
- Can listeners hear quiet sounds?
- Can listener hear loud sounds without distortion?

### Further Information

- *IaWiki: MP3*, <<http://www.infoanarchy.org/wiki/wiki.pl?MP3>>
- *MP3Pro Zone*, <<http://www.mp3prozone.com/>>
- *Measuring Audio Quality*, <<http://www.itworld.com/AppDev/1469/NWW1204revside4/>>
- *Ogg Vorbis*, <<http://www.vorbis.com/>>
- *PC Recording*, <<http://www.pcrecording.com/>>
- *Quality comparison for audio encoded at 64kbps*.  
<<http://audio.ciara.us/test/64test/presentation.html>>
- *Real Audio: Producing Music*,  
<<http://service.real.com/help/library/guides/production/htmfiles/audio.htm>>
- *Xorys' MP3 FAQ*, <<http://webhome.idirect.com/~nuzhathl/mp3-faq.html>>



## Producing And Improving The Quality Of Digitised Images

### About This Document

This briefing document describes QA techniques to ensure image quality is not reduced before and during the digitisation process.

### Citation Details

*Producing And Improving The Quality Of Digitised Images*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-66/>>

**Keywords:** digitisation, digitising, images, scan, straighten, image processing, briefing

## Introduction

To produce high-quality digital images you should follow certain rules to ensure that the image quality is sufficient for the purpose. This document presents guidance on digitising and improving image quality when producing a project Web site.

## Choose Suitable Source Material

Quality scans start with quality originals— high-contrast photos and crisp B&W line art will produce the best-printed results. Muddy photos and light-coloured line art can be compensated for, but the results will never be as good as with high-quality originals. The use of bad photos, damaged drawings, or tear sheets - pages that have been torn from books, brochures, and magazines - will have a detrimental effect upon the resultant digital copy. If multiple copies of a single image exist, it is advisable to choose the one that has the highest quality.

## Scan at a Suitable Resolution

It is often difficult to improve scan quality at a later stage. It is therefore wise to scan the source according to consistent, pre-defined specifications. Criteria should be based upon the type of material being scanned and the intended use. Table 1 indicates the minimum quality that projects should choose:

Use	Type	Dots Per Inch (dpi)
Professional	Text	200
	Graphics	600
Non-professional	Text	150
	Graphics	300

**Table 1: Guidelines To Scanning Source Documents**

Since most scans require subsequent processing, (e.g. rotate an image to align it correctly) that will degrade image quality, it is advisable to work at a higher resolution and resize the scans later.

Once the image has been scanned and saved to in an appropriate file format, measures should be taken to improve the image quality.



### Straighten Images

For best results, an image should lay with its sides parallel to the edge of the scanner glass. Although it is possible to straighten images that have been incorrectly digitised, it may introduce unnecessary distortion of the digital image.

### Sharpen the Image

To reduce the amount of subtle blur (or ‘fuzziness’) and improve visual quality, processing tools may be used to sharpen, smooth, improve the contrast level or perform gamma correction. Most professional image editing software contains filters that perform this function automatically.

### Correct Obvious Faults

Scanned images are often affected by many problems. Software tools can be used to remove the most common faults:

- Remove "red-eye" from a picture.
- Correct the colour balance
- Repair a tear or crease in a picture, or
- Remove a moiré pattern from a picture scanned from a book.

Be careful you do not apply the same effect twice. This can create unusual effects that distract the observer when viewer the picture.

### Further Information

- *Digital Album: Steps to Quality Scans*,  
<[http://www.city-gallery.com/digital/scan\\_steps.html](http://www.city-gallery.com/digital/scan_steps.html)>
- *JPEG Image Compression FAQ, part 1/2*,  
<<http://www.faqs.org/faqs/jpeg-faq/part1/preamble.html>>
- *How to Design Like a Pro*, <<http://www.prographicsllc.com/Digi/Scans.html>>
- *Scanning 101: Getting Great-Looking Line Art from Not-So-Great Sources*,  
<<http://www.creativepro.com/story/feature/6718.html>>

# Implementing and Improving Structural Markup

### About This Document

This briefing document describes methods for improving the quality of structural mark-up, with an emphasis upon use of recognised standards and conventions to ensure interoperability.

### Citation Details

*Implementing and Improving Structural Markup*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-67/>>

**Keywords:** structural mark-up, metadata mapping, standards, briefing

## Introduction

Digital text has existed in one form or another since the 1960s. Many computer users take for granted that they can quickly write a letter without restriction or technical considerations. This document provides advice for improving the quality of structural mark-up, emphasising the importance of good documentation, use of recognised standards and providing mappings to these standards.

## Why Should I Use Structural Mark-Up?

Although ASCII and Unicode are useful for storing information, they are only able to describe each character, not the method they should be displayed or organized. Structural mark-up languages enable the designer to dictate *how* information will appear and establish a structure to its layout. For example, the user can define a tag to store book author information and publication date.

The use of structural mark-up can provide many organizational benefits:

- **Easier to maintain** - allows modification to document structure without the need to directly edit the content. An entire site can be updated by changing a single CSS file.
- **Code reduction** – by abstracting the structural element to a separate file, the structural information can be used by multiple documents, reducing the amount of code required.
- **Portable** – The creation of well-formed documents will ensure the document will display correctly on browsers/viewers that support the markup language.
- **Interoperable** – Structural data can be utilized to access information stored in a third party database.

The most common markup languages are SGML and XML. Based upon these languages, several schemas have been developed to organize and define data relationships. This allows certain elements to have specific attributes that define its method of use (see Digital Rights document for more information). To ensure interoperability, XML is advised due to its support for contemporary Internet standards (such as Unicode).

## Improving The Quality Of Structural Mark-Up

For organisations that already utilise structural mark-up the benefits are already apparent. However, some consideration should be made on improving the quality of descriptive data. The key to improving data quality is twofold: utilise recognised

standards whenever possible; and establish detailed documentation on all aspects of the schema.

**Documentation:** Documentation is an important, if often ignored, aspect of software development. Good documentation should establish the purpose of structural data, examples, and the source of the data. Good documentation will allow others to understand the XML without ambiguity.

**Use recognised standards:** Although there are many circumstances where recognised schemas are insufficient for the required task, the designer should investigate relevant standards and attempt to merge their own bespoke solution with the various standard. In the long-term this will have several benefits:

1. The project can take advantage of existing knowledge in the field, allowing them to cover areas where they have limited or no experience.
2. Improve access to content by supporting proven standards, such as SVG.
3. The time required to map their data to alternative schemas used by other organisations will be reduced significantly.

TEI, Dublin Core, and others provide cross-subject metadata elements that can be combined with subject specific languages.

**Provide mappings to recognised standards:** Through the creation of different mappings the developer will standardise and enhance their approach to schema creation, removing potential ambiguities and other problems that may arise. In an organisational standpoint, the mappings will also allow improved relations between cooperating organisations and diversify the options available to use information in new ways.

**Follow implementation conventions:** In addition to implementing recognised standards, it is important that the developer follow existing rules to construct existing elements. In varying circumstances this will involve the use of an existing data dictionary, an examination of XML naming rules. Controlled languages (for example, RDF, SMIL, MathML, and SVG) use these conventions to implement specific localised knowledge.

### Further Information

- *Dublin Core Metadata Initiative*, <<http://dublincore.org/>>
- *TEI Home site*, <<http://www.tei-c.org/>>

# Techniques To Assist The Location And Retrieval Of Local Images

### About This Document

This briefing document describes techniques which can be used to assist the location and retrieval of local images.

### Citation Details

*Techniques To Assist The Location And Retrieval Of Local Images*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-68/>>

**Keywords:** digitisation, naming conventions, location, retrieval, images

## Summary

Use of a consistent naming scheme and directory structure, as well as controlled vocabulary or thesaurus improve the likelihood that digitised content captured by many people over an extended period will be organized in a consistent manner that avoid ambiguity and can be quickly located.

This QA paper describes techniques to aid the storage and successful location of digital images.

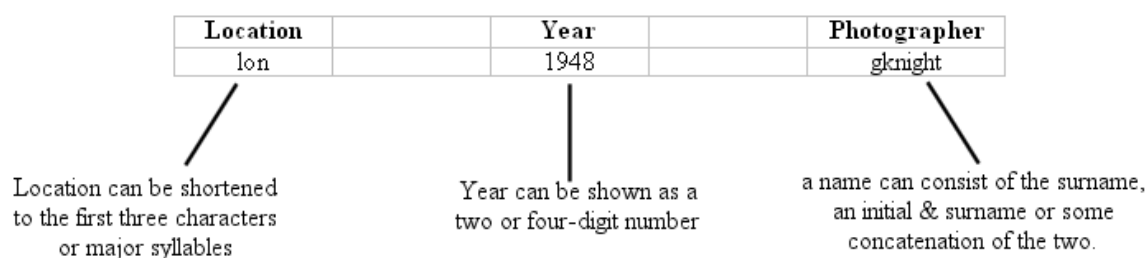
## Storing Local Images

Effective categorization of images stored on a local drive can be equally as important as storing them in an image management system. Digitisation projects that involve the scanning and manipulating of a large number of images will benefit from a consistent approach to file naming and directory structure.

An effective naming convention should identify the categories that will aid the user when finding a specific file. To achieve this, the digitisers should ask themselves:

- What type of information should be identified?
- What is the most effective method of describing this information in shorthand?

This can be better described with an example. A digitisation project is capturing photographs taken during wartime Britain. They have identified location, year and photographer as search criteria for locating images. To organize this information in a consistent manner the project team should establish a directory structure, common vocabulary and shorthand terms for describing specific locations. Figure 1 outlines a common description framework:



### Potential Problems

To avoid problems that may occur when the image collection expands or is transferred to a different system, the naming convention should also take account the possibility that:

- Some or all of this information may not be available (e.g. the year may be unknown)
- Several photographs are likely to exist that possess the same criteria – same location, year and photographer.
- Operating systems (OS) and Content Management Systems (CMS) treat lower case, upper case, and filename spaces in a different manner. To maintain consistency, filenames should be written in lower case and spaces should be avoided or replaced with underscores.
- Older operating systems or filing systems (e.g. ISO 9660) use the 8.3 DOS filename restrictions, which may cause problems when accessing these files.
- Some characters are illegal on different operating systems. Mac OS cannot use a colon in a filename, while DOS/Windows identifies ? [ ] / \ = + < > : ; " , as illegal.

Naming conventions will allow the project to avoid the majority of these problems. For example, a placeholder may be chosen if one of the identifiers is unknown (e.g. 'ukn' for unknown location, 9999 for year). Special care should be taken to ensure this placeholder is not easily mistaken for a known location or date. Additional criteria, such as other photo attributes or a numbering system, may also be used to distinguish images taken by the same person, in the same year, at the same location.

### Identification Of Digital Derivatives

Digital derivatives (i.e. images that have been altered in some way and saved under a different name) introduce further complications in how you distinguish the original from the altered version. This will vary according to the type of changes made. On a simple level, you may simply choose a different file extension or store files in two different directories (Original and modified). Alternatively you may append additional criteria onto the filename (e.g. \_sm for smaller images or thumbnails, \_orig and \_modif for original and modified).

### Further Information

- *Focusing Images for Learning and Teaching*, FILTER,  
<<http://www.filter.ac.uk/>>
- *MacWindows Tutorial*,  
< <http://www.macwindows.com/tutfiles.html>>
- *Controlling Your Language*, TASI,  
< <http://www.tasi.ac.uk/resources/vocabs.html>>
- *File Naming*, TASI,  
<<http://www.tasi.ac.uk/advice/creating/filenaming.html>>

# QA Techniques For The Storage Of Image Metadata

### About This Document

This briefing document describes techniques which can be used for the storage of images metadata.

### Citation Details

*QA Techniques For The Storage Of Image Metadata*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-71/>>

**Keywords:** digitisation, storage, image, metadata

## Background

The archival of digital images requires the consideration of the most effective method of storing technical and lifecycle information. Metadata is a common method used to describe digital resources, however the different approaches may confuse many users. This paper describes QA techniques for choosing a suitable method of metadata storage that takes into account the need for interoperability and retrieval.

## Choosing a Suitable Metadata Association Model

Metadata may be associated with an image in three ways:

1. **Internal Model:** Metadata is stored within the image file itself, either through an existing metadata mapping or attached to the end of an image file in an ad hoc manner. Therefore, it is simple to transfer metadata alongside image data without special requirements or considerations. However, support for a metadata structure differs between file formats and assignment of the same metadata record to multiple images causes inefficient duplication in comparison to a single metadata record associated with a group of images.
2. **External Model:** A unique identifier is used to associate external metadata with an image file. E.g. an image may be stored on a local machine while the metadata is stored on a server. This is better suited to a repository and is more efficient when storing duplicate information on a large number of objects. However, broken links may occur if the metadata record is not modified when an image is moved, or visa versa. Intellectual Property data and other information may be lost as a result.
3. **Hybrid Model:** Uses both internal and externally associated metadata. Some metadata (file headers/tags) are stored directly in the image file while additional workflow metadata is stored in an external database. The deliberate design of external record offers a common application profile between file formats and provides a method of incorporating format-specific metadata into the image file itself. However, it shares the disadvantages of internal & external models in terms of duplication and broken links.

When considering the storage of image metadata, the designer should consider three questions:

1. What type of metadata do you wish to store?
2. Is the file format capable of storing metadata?
3. What environment is the metadata intended to be stored and used within?

The answer to these questions should guide the choice of the metadata storage model. Some file formats are not designed to store metadata and will require supplementation

through the external model; other formats may not store data in sufficient detail for your requirements (e.g. lifecycle data). Alternatively, you may require IP (Intellectual Property) data to be stored internally, which will require a file format that supports these elements.

### Ensuring Interoperability

Metadata is intended for the storage and retrieval of essential information regarding the image. In many circumstances, it is not possible to store internal metadata in a format that may be read by different applications. This may be for a number of reasons:

- The file format does not define metadata placeholders (e.g. BMP), or does not use a metadata profile that the application uses.
- A standard image metadata definition and interchange format model does not exist (e.g. JPEG). As a result, the storage mechanism and metadata structure must be defined by each application.
- The metadata is stored in a proprietary file format that is not publicly defined.

Before choosing a specific image format, you should ensure the repository software is able to extract metadata and that editing software does not corrupt the data if changes are made at a later date. To increase the likelihood of this, you should take one of the following approaches:

- Convert image data to a file format that supports a known metadata structure (e.g. Exif, TIFF, SPIFF and Flashpix).
- Use a vendor-neutral and technology-independent, well-documented metadata standard, preferably one written in XML (e.g. DIG35, Z39.87 & MIX).
- Investigate the solutions offered by the DIG35 [1] and FILTER [2] projects which are developing a set of templates for consistent description of images.

Although this will not guarantee interoperability, these measures will increase the likelihood that it may be achieved.

### Structuring Your Image Collection

To organise your image collection into a defined structure, it is advisable to develop a controlled vocabulary. If providing an online resource, it is useful to identify your potential users, the academic discipline from which they originate, and the language they will use to locate images. Many repositories have a well-defined user community (archaeology, physics, sociology) that share a common language and similar goals. In a multi-discipline collection it is much more difficult to predict the terms a user will use to locate images. The US Library of Congress [2], the New Zealand “Time Frames” [3] and International Press Telecommunications Council (IPTC) [4] provide online examples of how a controlled vocabulary hierarchy may be used to catalogue images.

### References

1. *DIG35 Specification: Metadata for Digital Images*, Version 1.0, August 30, 2000, <<http://xml.coverpages.org/FU-Berlin-DIG35-v10-Sept00.pdf>>
2. *FILTER*, <<http://www.filter.ac.uk/>>
3. *Library of Congress Thesauri*, <<http://www.loc.gov/lexico/servlet/lexico/>>
4. *New Zealand Time Frames*, <<http://timeframes1.natlib.govt.nz/nlnz-browse>>
5. *International Press Telecommunications Council*, <<http://www.iptc.org/>>

# Improving the Quality of Digitised Images



## About This Document

This briefing document describes techniques for improving the quality of digitised images

## Citation Details

*Improving the Quality of Digitised Images*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-74/>>

**Keywords:** digitisation, master image, colour balance, artefacts, image

## Background

A digitised image requires careful preparation before it is suitable for distribution. This document describes a workflow for improving the quality of scanned images by correcting faults and avoiding common errors.

## Preparing Your Master Image

The sequence in which modifications are made will have a significant contribution to the quality of the final image. Although conformance to a strict sequence is not always necessary, inconsistencies may be introduced if the order varies dramatically between images. The Technical Advisory Service for Images (TASI) recommends the following order:

### 1. Does the image require rotation or cropping?

In many circumstances, the digitiser will not require the entire image. Cropping an image to a specific size, shape or orientation will reduce the time required for the computer to manipulate the image and prioritise errors to those considered important.

### 2. Are shades and colours difficult to distinguish?

Scanners and digital cameras often group colours into a specific density range. This makes it difficult to differentiate shades of the same colour. Use the Histogram function with Photoshop (or other software) and adjust the different levels to best use the range of available tones.

### 3. Is the colour balance accurate in comparison to the original?

Some colours may change when digitised, e.g. bright orange may change to pink. Adjust the colour balance by modifying the Red, Green & Blue settings. Decreasing one colour increases its opposite.

### 4. Are there faults or artefacts on the image?

Visual checks should be performed on each image, or a selection of images, to identify faults, such as dust specks or scratches on the image.

Once you are satisfied with the results, the master image should be saved in a lossless image format - RGB Baseline TIFF Rev 6 or PNG are acceptable for this purpose.



### Improving Image Quality

Subsequent improvements by resizing or sharpening the image should be performed on a derivative.

#### 1. Store work-in-progress images in a lossless format

Digitisers often get into the habit of making modifications to a derivative image saved in a ‘lossy’ format, i.e. a format that simplifies detail to reduce file size. This is considered bad practice, will reduce quality and cause compression ‘artefacts’ to appear over subsequent edits. When repeatedly altering an image it is advisable to save the image in a lossless format (e.g. TIFF, PNG) until the image is ready for dissemination. Once all changes have been made it can be output in a lossy format.

#### 2. Filter the image

Digitised images often appear ‘noisy’ or contain dust and scratches. Professional graphic manipulation (Photoshop, PaintShop Pro, etc.) possesses graphic processors that can be useful in removing these effects. Common filters include ‘Despeckle’ that subtly blurs an image to reduce the amount of ‘noise’ in an image and ‘median’ that blends the brightness of pixels and discards pixels that are radically different from adjacent pixels.

#### 3. Remove distracting effect

If you are funded to digitise printed works, moiré (pronounced more-ray) effects may be a problem. Magazine or newspaper illustrations that print an image as thousands of small coloured dots produce a noticeable repeating pattern when scanned. Blur effects, such as the Gaussian blur, are an effective method of reducing noticeable moiré effects, however these also reduce image quality. Resizing the image is also an effective strategy that forces the image-processing tool to re-interpolate colours, which will soften the image slightly. Although these effects will degrade image to an extent, the results are often better than a moiré.

### Further Information

- *Image Manipulation and Preparation*, TASI,  
<<http://www.tasi.ac.uk/advice/using/dimpmanipulation.html>>
- *Using Raster Images*, QA Focus briefing document,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-28/>>
- *Digital Imaging Basics*, TASI,  
<<http://www.tasi.ac.uk/advice/using/basics.html>>

# Digitisation Of Still Images Using A Flatbed Scanner

### About This Document

This briefing document describes the options to take into account when preparing to digitise still images using a flatbed scanner

### Citation Details

*Digitisation of still images using a flatbed scanner*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-75/>>

**Keywords:** digitisation, software, policy, quality, image

## Preparing For A Large-Scale Digitisation Project

The key to the development of a successful digitisation project is to separate it into a series of stages. All projects planning to digitise documents should establish a set of guidelines to help ensure that the scanned images are complete, consistent and correct. This process should consider the proposed input and output of the project, and then find a method of moving from the first to the second.

This document provides preparatory guidance to consider when approaching the digitisation of many still images using a flatbed scanner.

## Choose Appropriate Scanning Software

Before the digitisation process may begin, the digitiser requires suitable tools to scan & manipulate the image. It is possible to scan a graphic using any image processing software that supports TWAIN (an interface to connect to a scanner, digital camera, or other imaging device from within a software application), however the software package should be chosen carefully to ensure it is appropriate for the task. Possible criteria for measuring the suitability of image processing software include:

- The ability to perform batch operations upon many different images
- The ability to perform image processing upon an image.
- The digitisers familiarity with the software package

A timesaving may be found by utilizing a common application, such as Adobe Photoshop, Paintshop Pro, or GIMP. For most purposes, these offer functionality that is rarely provided by editing software included with the scanner.

## Check The Condition Of The Object To Be Scanned

Image distortion and dark shading at page edges are common problems encountered during the digitisation process, particularly when handling spine-bound books. To avoid these and similar issues, the digitiser should ensure that:

1. The document is uniformly flat against the document table.
2. The document is not accidentally moved during scanning.
3. The scanner is on a flat, stable surface.
4. The edges of the scanner are covered by paper to block external light, caused when the object does not lay completely flat against the scanner.

Scanning large objects that prevent the scanner lid being closed (e.g. a thick book) often causes discolouration or blurred graphics. Removing the spine will allow each page to be scanned individually, however this is not always an option (i.e. when handling valuable books). In these circumstances you should consider a planetary camera as an alternative scanning method.

### Identification Of A Suitable Policy For Digitisation

It is often costly and time-consuming to rescan the image or improve the level of detail in an image at a later stage. Therefore, the digitiser should ensure that a consistent approach to digitisation is taken in the initial stages. This will include the choice of a suitable resolution, file format and filename scheme.

#### **Establish a consistent quality threshold for scanned images**

It is difficult to improve low quality scans at a later date. It is therefore important to digitise images at a slightly higher resolution (measured in pixels per inch) and scan type (24-bit or higher for colour, or 8-bit or higher for grey scale) than required and rescale the image at a later date.

#### **Choose an appropriate image format**

Before scanning the image, the digitiser should consider the file format in which it will be saved. RGB Baseline TIFF Rev 6 is the accepted format of master copies for archival and preservation (although PNG is a possible alternative file format). To preserve the quality, it is advisable to avoid compression where possible. If compression must be used (e.g. for storing data on CD-ROM), the compression format should be noted (Packbits, LZW, Huffman encoding, FAX-CCITT 3 or 4). This will avoid incompatibilities in certain image processing applications.

Data intended for dissemination should be stored in one of the more common image formats to ensure compatibility with older or limited browsers. JPEG (Joint Photographic Experts Group) is suitable for photographs, realistic scenes, or other images with subtle changes in tone, however its use of 'lossy' compression causes sharp lines or letterings are likely to become blurred. When modifying an image, the digitiser should return to the master TIFF image, make the appropriate changes and resave it as a JPEG.

#### **Choose an appropriate filename scheme**

Digitisation projects will benefit from a consistent approach to file naming and directory structure that allows images to be organized in a manner that avoids confusion and can be quickly located. An effective naming convention should identify the categories that will aid the user when finding a specific file. For example, the author, year it was created, thematic similarities, or other notable factors. The digitiser should also consider the possibility that multiple documents will have the same filename or may lack specific information and consider methods of resolving these problems. Guidance on this issue can be found in related QA Focus documents.

### Further Information

1. *Acquiring Images, MicroFrontier Online*,  
<[http://www.microfrontier.com/support\\_area/faq031/qa002.html](http://www.microfrontier.com/support_area/faq031/qa002.html)>
2. *Scanner Quality problems, Epson*, <[http://support2.epson.net/manuals/english/scanner/perfection2450photo/REF\\_G/TRBLE\\_6.HTM](http://support2.epson.net/manuals/english/scanner/perfection2450photo/REF_G/TRBLE_6.HTM)>
3. *TWAIN Standard, Twain*, <<http://www.twain.org/>>

## 4.2 Technical Advice: Digitisation – Case Studies

---

4. *Resolving the Units of Resolution*, <<http://www.tasi.ac.uk/advice/creating/dpi.html>>
5. *A Rough Guide to Image Sizes*, TASI, <<http://www.tasi.ac.uk/advice/creating/roughguide.html>>
6. *Generic Image Workflow: TASI Recommended Best Practice for Digitisation Projects*, TASI, <[http://www.tasi.ac.uk/advice/managing/workflow\\_generic.html](http://www.tasi.ac.uk/advice/managing/workflow_generic.html)>
7. *The Book Scanning and Digitisation Process*, <<http://www.rodneep.co.uk/books/production/scan/>>

# Choosing A Suitable Digital Watermark

### About This Document

This briefing document describes considerations to be taken into account when choosing a digital watermark system.

### Citation Details

*Choosing a suitable Digital Watermark*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-76/>>

**Keywords:** digitisation, watermark, copyright, resilience, invisible, visible, image

## Summary

Watermarking is an effective technology that solves many problems within a digitisation project. By embedding Intellectual Property data (e.g. the creator, licence model, creation date or other copyright information) within the digital object, the digitiser can demonstrate they are the creator and disseminate this information with every copy, even when the digital object has been uploaded to a third party site. It can also be used to determine if a work has been tampered with or copied.

This paper describes methods for establishing if a project requires watermarking techniques and criteria for choosing the most suitable type.

## Purpose Of A Watermark

Before implementing watermarking within your workflow, you should consider its proposed purpose. Are you creating watermarks to indicate your copyright, using it as a method of authentication to establish if the content has been modified, or doing so because everyone else has a watermarking policy? The creation of a watermark requires significant thought and modification to the project workflow that may be unnecessary if you do not have a specific reason for implementing it.

For most projects, digital watermarks are an effective method of identifying the copyright holder. Identification of copyright is encouraged, particularly when the work makes a significant contribution to the field. However, the capabilities of watermarks should not be overstated – it is useful in identifying copyright, but is incapable of preventing use of copyrighted works. The watermark may be ignored or, given sufficient time and effort, removed entirely from the image. If the intent is to restrict content reuse, a watermark may not be the most effective strategy.

## Required Attributes Of A Watermark

To assist the choice of a watermark, the project team should identify the required attributes of a watermark by answering two questions:

1. To whom do I wish to identify my copyright?
2. What characteristics do I wish the watermark to possess?

The answer to the first question is influenced by the skills and requirements of your target audience. If the copyright information is intended for non-technical and technical users, a visible watermark is the most appropriate. However, if the copyright information is intended for technical users only or the target audience is critical of

visible watermarks (e.g. artist may criticise the watermark for impairing the original image), an invisible watermark may be the best option.

To answer the second question, the project team should consider the purpose of the watermark. If the intent is to use it as an authentication method (i.e. establish if any attempt to modify the content has been made), a fragile watermark will be a valued attribute. A fragile watermark is less robust towards modifications where even small change of the content will destroy embedded information. In contrast, if the aim is to reflect the owner's copyright, a more robust watermark may be preferential. This will ensure that copyright information is not lost if an image is altered (through cropping, skewing, warp rotation, or smoothing of an image)

### Choosing a resilient watermark

If resilience is a required attribute of a digital watermark, the project team has two options: invisible or visible watermark. Each option has different considerations that make it suitable for specific purposes.

#### Invisible Watermarks

Invisible watermarks operate by embedding copyright information within the image itself. As a rule, watermarks that are less visible are weaker and easier to remove. When choosing a variant it is important to consider the interaction between watermark invisibility and resilience. Some examples are shown in Table 1:

Name	Description	Resilience
Bit-wise	Makes minor alterations to the spatial relation of an image	Weak
Noise Insertion	Embed watermark within image “noise”	Weak
Masking and filtering	Similar to paper watermarks on a bank note, it provides a subtle, though recognisable evidence of a watermark.	Strong
Transform domain	Uses dithering, luminance, or lossy techniques (similar to JPEG compression) on the entire or section of an image.	Strong

**Table 1: Indication of resilience for invisible watermarks**

‘Bit-wise’ & ‘noise insertion’ may be desirable if the purpose is to determine whether the medium has been altered. In contrast, ‘transform domain’ and ‘masking’ techniques are highly integrated into the image and therefore more robust to deliberate or accidental removal (caused by compression, cropping, and image processing techniques) in which significant bits are changed. However, these are often noticeable to the naked eye.

#### Visible Watermarks

visible watermark is more resilient and may be used to immediately identify copyright without significant by the user. However, these are, by design, more intrusive to the media. When creating a visible watermark, the project team should consider its placement. Projects funded with public money should be particularly conscious that the copyright notice does not interfere with the purpose of the project – to entertain and



A  
effort

education the public. A balance should be reached between the need to make the watermark difficult to remove and its use to the user.

Both watermarks make them suitable for specific situations. If handling a small image collection, it may be feasible (in terms of time and effort) to use both watermarks as a redundant protection measure - in the event that one is removed, the second is likely to remain.

### Information stored within the watermark

If the project is using a watermark to establish their copyright, some thought should be made on the static information you wish to provide. For example:

- **Creator:** The forename and surname of the person who created the image, either as full text or their initials.
- **Organisation:** The project or organisation that holds copyright for the work.
- **Creation date:** The date of creation, either the exact date (e.g. 24/03/2004) or year (2004)
- **Identifiers:** A unique identifier to distinguish the image, distributor, creator, transaction, and other attributes.

Some content management systems are also able to generate dynamic watermarks and embed them within the image. This may record the file information (file format, image dimensions, etc.) and details about the download transaction (transaction identifier, download date, etc.). This may be useful for tracking usage, but may annoy the user if the data is visible.

### Implementing watermarks in the project workflow

To avoid unnecessary corruption of a watermark by the digitiser/creator themselves, the watermark creation process should be delayed until the final steps of the digitization workflow. Watermarks can be easily removed when the digitiser is modifying the image in any way (e.g. through cropping, skewing, adjustment of the RGB settings, or through use of lossy compression). If an image is processed to the degree that the watermark cannot be recognized, then reconstruction of the image properties may be possible through the use of an original image.

### Further Information

1. *DigiMarc*, <<http://www.digimarc.com/>>
2. *Digital Image Archive User Issues: Digital Watermarking*, TASI, <<http://www.tasi.ac.uk/advice/using/uissues.html#ui6>>
3. *Paint Shop Pro 5 - Watermarking Photographs*, <<http://www.grafx-design.com/24psp.html>>

# Using SVG in the ARTWORLD Project

Case study

### About This Document

This case study describes use of the SVG format which has been developed by the W3C.

### Citation Details

*Using SVG in the ARTWORLD Project*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-07/>>

**Keywords:** metadata, Artworld, SVG, case study

## About The ARTWORLD Project

**Artworld** [1] is a consortium project funded by JISC under the 5/99 funding round. The consortium consists of The Sainsbury Centre for Visual Arts (SCVA) at The University of East Anglia (UEA) and the Oriental Museum at the University of Durham.



The main deliverable for the project is its Web site which will include a combined catalogue of parts of the two collections and a set of teaching resources.

Object images are being captured using digital photography at both sites and some scanning at SCVA. Object data is being researched at both sites independently and is input to concurrent Microsoft Access databases. Image data is captured programmatically from within the Access database. Object and image data are exported from the two independent databases and checked and imported into a Postgres database for use within the catalogue on the Web site.

There are four teaching resources either in development or under discussion. These are African Art and Aesthetics, Egyptian Art and Museology, An Introduction to Chinese art and Japanese Art. These resources are being developed by the department of World Art Studies and Museology at UEA, The department of Archaeology, Durham and East Asian Studies, University of Durham respectively. The Japanese module is currently under negotiation. These resources are stored as simple XML files ready for publication to the Web.

The target audience in the first instance are undergraduate art history, anthropology and archaeology students. However, we have tried to ensure that the underlying material is structured in such a way that re-use at a variety of levels, 16 plus to post graduate is a real possibility. We hope to implement this during the final year of the project by ensuring conformance with IMS specifications.

## How Use Of SVG Came About

In the early days of the project we were trying very hard to find an IT solution that would not only fulfill the various JISC requirements but would be relatively inexpensive. After a considerable amount time researching various possibilities we selected Apache's **Cocoon** system as our Web publishing engine. To help us implement this we contracted a local internet applications provider Luminas [2].

The Cocoon publishing framework gives us an inexpensive solution in that the software is free so we can focus our resources on development.

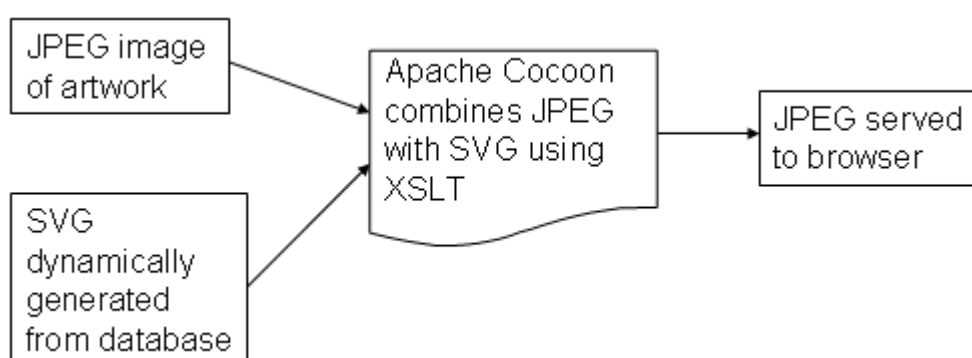


One area that we had inadvertently missed during early planning was how we represent copyright for the images whilst providing some level of protection. We considered using watermarking however this would have entailed re-processing a considerable number of images at a time when we had little resource to spare.

This issue came up in conversation with Andrew Savory of Luminas as early notification that all of the images already transferred to the server and in use through Cocoon would need to be replaced. As we talked about the issues Andrew presented a possible solution, why not insert copyright notices into the images "on the fly". This would be done using a technology called SVG (Scalable Vector Graphics). What SVG could do for us is to respond to a user request for an image by combining the image with the copyright statement referenced from the database and present the user with this new combined image and copyright statement.

We of course asked Luminas to proceed with this solution. The only potential stumbling block was how we represent copyright from the two institutions in a unified system. The database was based on the VADS/VRA data schema so we were already indicating the originating institution in the database. It was then a relatively simple task to include a new field containing the relevant copyright statements.

It should be noted that a composite JPEG (or PNG, GIF or PDF) image is sent to the end user - there is no requirement for the end user's browser to support the PNG format. The model for this is illustrated in Figure 1.



**Figure 1: Process For Creating Dynamic Images**

### Lessons Learnt

Although in this case we ended up with an excellent solution there are a number of lessons that can be derived from the sequence of events. Firstly the benefits of detailed workflow planning in the digitisation process cannot be understated. If a reasonable solution (such as water marking) had been planed into the processes from the start then a number of additional costs would not have been incurred. These costs include project staff time in discussing solutions to the problem, consultancy costs to implement a new solution. However, there are positive aspects of these events that should be noted. Ensuring that the project has a contingency fund ensures that unexpected additional costs can be met. Close relations with contractors with free flow of information can ensure that potential solutions can be found. Following a standard data schema for database construction can help to ensure important data isn't missed. In this case it expedited the solution.

### About Cocoon and SVG

Cocoon [3] is an XML Publishing Framework that allows the possibility of including logic in XML files. It is provided through the Apache software foundation.

SVG (Scalable Vector Graphics) [4] [5] is non proprietary language for describing two dimensional graphics in XML. It allows for three types of objects: vector graphic shapes; images and text. Features and functions include: grouping, styling, combining, transformations, nested transformations, clipping paths, templates, filter effects and alpha masks.

### References

1. **Artworld**, University of East Anglia, <<http://artworld.scva.uea.ac.uk/>>
2. **Luminas**, <<http://www.luminas.co.uk/>>
3. **Apache Cocoon**, Apache, <<http://xml.apache.org/cocoon/>>
4. **Introduction to SVG**, xml.com, <<http://www.xml.com/pub/a/2001/03/21/svg.html>>
5. **SVG Specification**, W3C, <<http://www.w3.org/TR/2000/CR-SVG-20001102/>>

### Contact Details

Paul Child  
ARTWORLD Project Manager  
Sainsbury Centre for Visual Arts  
University of East Anglia  
Norwich NR4 7TJ  
Tel: 01603 456 161  
Email: [p.child@uea.ac.uk](mailto:p.child@uea.ac.uk)

# Crafts Study Centre Digitisation Project - and Why 'Born Digital'

Case study

### About This Document

This case study describes the approaches to digitisation taken by the Crafts Study Centre Digitisation Project.

### Citation Details

*Crafts Study Centre Digitisation Project - and Why 'Born Digital'*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-08/>>

**Keywords:** metadata, craft study, digitisation, case study

## Background

The **Crafts Study Centre** (CSC) [1], established in 1970, has an international standing as a unique collection and archive of twentieth century British Crafts. Included in its collection are textiles, ceramics, calligraphy and wood. Makers represented in the collection include the leading figures of the twentieth century crafts such as Bernard Leach, Lucie Rie and Hans Coper in ceramics; Ethel Mairet, Phyllis Barron and Dorothy Larcher, Edward Johnston, Irene Wellington, Ernest Gimson and Sidney Barnsley. The objects in the collection are supported by a large archive that includes makers' diaries, documents, photographs and craftspeople's working notes.

## The Crafts Study Centre Digitisation Project

The Crafts Study Centre Digitisation Project [2] has been funded by the JISC to digitise 4,000 images of the collection and archive and to produce six learning and teaching modules. Although the resource has been funded to deliver to the higher education community, the project will reach a wide audience and will be of value to researchers, enthusiasts, schools and the wider museum-visiting public. The Digitisation Project has coincided with an important moment in the CSC's future. In 2000 it moved from the Holborne Museum Bath, to the Surrey Institute of Art & Design, University College, Farnham, where a purpose-built museum with exhibition areas and full study facilities, is scheduled to open in spring 2004.

The decision to create 'born digital' data was therefore crucial to the success not only of the project, but also in terms of the reusability of the resource. The high-quality resolutions that have resulted from 'born digital' image, will have a multiplicity of use. Not only will users of the resource on the Internet be able obtain a sense of the scope of the CSC collection and get in-depth knowledge from the six learning and teaching modules that are being authored, but the relatively large file sizes have produced TIFF files that can be used and consulted off-line for other purposes.

These TIFF files contain amazing details of some of the objects photographed from the collection and it will be possible for researchers and students to use this resource to obtain new insights into for example, the techniques used by makers. These TIFF files will be available on site, for consultation when the new CSC opens in 2004. In addition to this, the high-quality print out-put of these images means that they can be used in printed and published material to disseminate the project and to contribute to building the CSC's profile via exhibition catalogues, books and related material.

### Challenges

The project team were faced with a range of challenges from the outset. Many of these were based on the issues common to other digital projects, such as the development of a database to hold the associated records that would be interoperable with the server, in our case the Visual Arts Data Service (VADS), and the need to adopt appropriate metadata standards. Visual Resources Association (VRA) version 3.0 descriptions were used for the image fields. Less straightforward was the deployment of metadata for record descriptions. We aimed for best practice by merging Dublin Core metadata standards with those of the Museum Documentation Association (mda). The end produce is a series of data fields that serve firstly, to make the database compatible with the VADS mapping schema, and secondly to realise the full potential of the resource as a source of information. A materials and technique field for example, has been included to allow for the input of data about how a maker produced a piece. Users of the resource, especially students and researchers in the history of art and design will be able to appreciate how an object in the collection was made. In some records for example, whole 'recipes' have been included to demonstrate how a pot or textile was produced.

Other issues covered the building of terminology controls, so essential for searching databases and for achieving consistency. We consulted the Getty Art and Architecture Thesaurus (AAT) and other thesauri such as the mda's wordhord, which acts as a portal to thesauri developed by other museums or museum working groups. This was sometimes to no avail because often a word simply did not exist and we had to rely on terminology developed in-house by curators cataloguing the CSC collection, and have the confidence to go with decisions made on this basis. Moreover, the attempt to standardise this kind of specialist collection can sometimes compromise the richness of vocabulary used to describe it.

### Lessons Learnt

Other lessons learnt have included the need to establish written image file naming conventions. Ideally, all image file names should tie in with the object and the associated record. This system works well until sub-numbering systems are encountered. Problems arise because different curators when cataloguing different areas of the collection, have used different systems, such as letters of the alphabet, decimal and Roman numerals. This means that if the file name is to match the number marked on the object, then it becomes impossible to achieve a standardised approach. Lessons learnt here, were that we did not establish a written convention early enough in the project, with the result that agreement on how certain types of image file names should be written before being copied onto CD, were forgotten and more than one system was used.

The value of documenting all the processes of the project cannot be overemphasised. This is especially true of records kept relating to items selected for digitisation. A running list has been kept detailing the storage location, accession number, description of the item, when it was photographed and when returned to storage. This has provided an audit trail for every item digitised. A similar method has been adopted with the creation of the learning and teaching modules, and this has enhanced the process of working with authors commissioned to write the modules.

Lastly, but just as importantly, has been the creation of QA forms on the database based on suggestions presented by the Technical Advisory Services for Imaging (TASI) at the JISC Evaluation workshop in April 2002. This has established a framework for

checking the quality and accuracy of an image and its associated metadata, from the moment that an object is selected for digitisation, through to the finished product. Divided into two sections, dealing respectively with image and record metadata, this has been developed into an editing tool by the project's documentation officer. The QA forms allows for most of the data field to be checked off by two people before the image and record is signed off. There are comment boxes for any other details, such as faults relating to the image. A post-project fault report/action taken box has been included to allow for the reporting of faults once the project has gone live, and to allow for any item to re-enter the system.

The bank of images created by the Digitisation Project will be of enormous importance to the CSC, not only in terms of widening access to the CSC collection, but in helping to forge its identity when it opens its doors as a new museum in 2004 at the Surrey Institute of Art & Design, University College.

### References

1. **The Crafts Study Centre**, <<http://www.craftscentre.surrart.ac.uk/>>
2. **The Crafts Study Centre Digitisation Project**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=project\\_crafts](http://www.jisc.ac.uk/index.cfm?name=project_crafts)>

### Contact Details

Jean Vacher  
Digitisation Project Officer  
Crafts Study Centre  
Surrey Institute of Art & Design, University College

# Image Digitisation Strategy and Technique: Crafts Study Centre Digitisation Project

Case study

### About This Document

This case study describes the approaches taken for image capture by the Crafts Study Centre Digitisation Project.

### Citation Details

*Image Digitisation Strategy and Technique: Crafts Study Centre Digitisation Project*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-09/>>

**Keywords:** image, digitisation, image capture, case study

## Background

Information about the **The Crafts Study Centre (CSC)** [1] and the **The Crafts Study Centre Digitisation Project** [2] is given in another case study [3].

## Problem Being Addressed

At the outset of **The Crafts Study Centre (CSC)** Digitisation Project extensive research was undertaken by the project photographer to determine the most appropriate method of image capture. Taking into account the requirements of the project as regards to production costs, image quality and image usage the merits of employing either traditional image capture or digital image capture were carefully considered.

## The Approach Taken

The clear conclusion to this research was that digital image capture creating born digital image data via digital camera provided the best solution to meet the project objectives. The main reasons for reaching this conclusion are shown below:

All image capture and post-production is carried out in-house allowing precise control over quality and output.

Fast, safe and efficient delivery of captured digital image data for transfer to CSC database and subsequent Web delivery.

Photographer maintains full control over the final appearance of captured images particularly in respect of colour balance, tone, exposure and cropping thus helping to maintain high degrees of quality control.

Born digital image capture provides cost efficiency as compared to traditional image capture and associated processing and scanning costs.

Items from the CSC collection are identified by members of the project team and passed to the photographer for digitisation. Once the item has been placed in position and the appropriate lighting arranged, it is photographed by using a large format monorail camera (cambo) hosting a Betterlight digital scanning back capable of producing image file sizes of up to 137 megabytes without interpolation.

Initially a prescan is made for appropriate evaluation by the photographer. Any necessary adjustments to exposure, tone, colour, etc are then made via the camera software and then a full scan is carried out with the resulting digital image data being

automatically transferred to the photographers image editing program, in this case Photoshop 6.

Final adjustments can then be made, if required and the digital image then saved and written onto CDR for onward delivery to the project database.

The main challenges in setting up this system were mostly related to issues regarding colour management, appropriate image file sizes, and standardisation wherever possible.

To this end a period of trialling was conducted by the photographer at the start of the image digitisation process using a cross section of subject matter from the CSC collection.

Identifying appropriate file sizes for use within the project and areas of the digital imaging process to which a level of standardisation could be applied was fairly straightforward, however colour management issues proved slightly more problematic but were duly resolved by careful cross-platform (Macintosh/MS Windows) adjustments and standardisation within the CSC and the use of external colour management devices.

### References

1. **The Crafts Study Centre**, <<http://www.craftscentre.surrey.ac.uk/>>
2. **The Crafts Study Centre Digitisation Project**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=project\\_crafts](http://www.jisc.ac.uk/index.cfm?name=project_crafts)>
3. **Crafts Study Centre Digitisation Project - and Why 'Born Digital'**, QA Focus, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-08/>>

### Contact Details

David Westwood  
Project Photographer  
Digitisation Project Officer  
Crafts Study Centre Surrey Institute of Art & Design, University College

# Digitisation of Wills and Testaments by the Scottish Archive Network (SCAN)

Case study

### About This Document

This case study describes the approaches to digitisation taken by the Scottish Archive Network.

### Citation Details

*Digitisation of Wills and Testaments by the Scottish Archive Network (SCAN)*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-19/>>

**Keywords:** digitisation, SCAN, case study

## Background

**Scottish Archive Network (SCAN)** is a Heritage Lottery Funded project. The main aim of the project is to open up access to Scottish Archives using new technology. There are three strands to the project:

1. Creating collection level descriptions of records held by 52 archives throughout Scotland [1].
2. Provide all the resources that a researcher will need when accessing archive resources over the Internet [2].
3. Digitisation of the Wills and Testaments registered in Scotland from 1500s to 1901 [3].

The digitisation of the Wills and Testaments are the focus of this case study.

## Problem Being Addressed

The digitisation of the testaments is an ambitious undertaking. The main issues to be considered are:

- **Format:** all of the material is manuscript and mostly in bound volumes, but there is a significant amount of loose leaf material too
- **Conservation condition:** part of the reason for choosing the testaments was that some of them were in poor condition due to their popularity with users, which meant that they needed conservation prior to digitisation
- **Scale:** There are more than 3 million pages and three thousand volumes in the dates selected. This necessitated a high throughput together with quite a formal approach to the workflow involved.

## The Approach Taken

### Document Preparation

As digital objects, images of manuscript pages lack the obvious information given by a physical page bound in a volume. It is important for completeness and for sequence that the pages themselves are accurately paginated. This gives a visual indication of the page number on the image as well as being incorporated into the naming convention used to identify the file. As a result quality is improved by reducing the number of pages



missed in the digitisation process and by ensuring that entire volumes are captured and in the correct sequence.

### **Image Capture**

The image capture program (dCam) automated the file naming process thereby reducing operator error and automatically capturing metadata for each image. This included date, time, operator id, file name, camera id and so on which helped in identifying whether later problems related to operator training or to a specific workstation. The program also included simple options for retakes.

### **Post Image Capture**

We have instituted a secondary quality assurance routine. This involves an operator (different to the one who captured the images) examining a selection of the images for any errors missed by the image capture operator. Initially, 100% of the images were checked, but a 30% check was soon found to be satisfactory. The quality control is carried out within 24 hours of a volume being digitised, which means that the volume is still available in the camera room should any retakes be necessary. The QA operators have a list of key criteria to assess the image - completeness, colour, consistency, clarity and correctness. When operators find a defective image they reject it and select the reason from a standardised list. Although the images are chosen at random, whenever an error is found the QA program will present the next sequential image, as it is more likely for errors to be clustered together. A report is produced by the QA program which is then used to select any retakes. The reports are also analysed for any recurring problems that may be corrected at the time of capture. Further QA criteria: the quality of the cameras had been specified in terms of capacity (i.e. number of pixels), and we found that it is also possible to specify the quality of the CCD in terms of an acceptable level of defective pixels. This, however, does have a bearing on cost.

## **Problems Experienced**

### **Preparation**

This was a time consuming process, which was slower than capture itself. It was important to build up sufficient material in advance of the digitisation getting underway.

### **Capture**

We chose to capture colour images. The technique used was to take 3 separate colour images through red, green and blue filters and then to combine them into a single colour image. This worked well and produced very high quality colour images. \* However, it was very difficult to spot where there had been slight movement between the three colour shots. At a high level of magnification this produced a misregistration between the 3 colour planes. The QA process sometimes caught this but it was far more costly for this to be trapped later on. We discovered that where there had been slight movement, the number of distinct colours in an image was almost double the average. We used this information to provide a report to the QA operators highlighting where potential colour shift had taken place. In addition the use of book cradles helped reduce this problem as well as enabling a focused image to be produced consistently.

### Things We Would Do Differently

The project has achieved successful completion within budget. For the digital capture program it proved possible to capture an additional 1 million pages as the capture and quality control workflow worked well. It is clear that the process is well suited to high throughput capture of bound manuscript material. Loose-leaf material took far more conservation effort and a much longer time to capture.

### References

1. **Online Catalogues Microsite**, Scottish Archive Network,  
<<http://www.scan.org.uk/aboutus/indexonline.htm>>
2. **Scottish Archive Network**, <<http://www.scan.org.uk/>>
3. **Scottish Documents**, Scottish Archive Network,  
<<http://www.scottishdocuments.com/content/>>

### Contact details

Rob Mildren  
Room 2/1  
Thomas Thomson House  
99 Bankhead Crossway  
Edinburgh  
EH11 4DX

*for SCAN Business:*

Tel: 0131-242-5802  
Fax: 0131-242-5801  
Email: [rob.mildren@scan.org.uk](mailto:rob.mildren@scan.org.uk)  
URL: <http://www.scan.org.uk/>

*for NAS Business:*

Tel: 0131-270-3310  
Fax: 0131-270-3317  
Email: [rob.mildren@nas.gov.uk](mailto:rob.mildren@nas.gov.uk)  
URL: <http://www.nas.gov.uk/>

# 4.3 Access / Web

---

## Background

This section addresses access to resources, primarily through use of the Web..

The World Wide Web is the key delivery platform for many projects. There is an expectation that projects will comply with W3C's standards in this area, including HTML and CSS.

There is a need to ensure that systematic processes for checking compliance with such standards are used. It is not appropriate to rely on the appearance on a Web page in a Web browser as browsers are designed to be tolerant of errors. In addition we should seek to ensure that resources can be accessed by accessibility tools (such as speaking browsers) and by non-traditional devices (such as PDAs).

In addition to access to resources using traditional Web browsers there is an increasing expectation that resources will be processed by automated software or delivered in formats other than HTML.

The briefing documents seek to describe best practices in this area.

## Briefing Documents

The following briefing documents which address the area of access/Web have been produced:

- *Compliance With HTML Standards* (briefing-01)
- *Use Of Proprietary Formats On Web Sites* (briefing-03)
- *Approaches To Link Checking* (briefing-07)
- *Accessing Your Web Site On A PDA* (briefing-05)
- *Search Facilities For Your Web Site* (briefing-08)
- *404 Error Pages On Web Sites* (briefing-05)
- *Enhancing Web Site Navigation Using The LINK Element* (briefing-10)
- *The Purpose Of Your Project Web Site* (briefing-15)
- *How To Evaluate A Web Site's Accessibility Level* (briefing-12)
- *URI Naming Conventions For Your Project Web Site* (briefing-16)
- *Performance Indicators For Your Project Web Site* (briefing-17)
- *Use Of Cascading Style Sheets (CSS)* (briefing-34)
- *Mothballing Your Web Site* (briefing-04)

Advisory documents which cover specific technical areas are available within the section on the appropriate technical area.

## Case Studies

The following case studies which address the area of access/Web have been produced:

- *Edinburgh University Library Online: Work In Progress* (case-study-26)
- *Gathering Usage Statistics And Performance Indicators: The (NMAP) Experience* (case-study-06)

### 4.3 Technical Advice: Access/Web – Briefing Papers

---

- *Usability Testing for the Non-Visual Access to the Digital Library (NoVA) Project* (case-study-11)
- *Approaches to Accessibility at MIMAS* (case-study-15)
- *Standards and Accessibility Compliance for the FAILTE Project Web Site* (case-study-31)
- *Standards and Accessibility Compliance for the DEMOS Project Web Site* (case study-10)
- *Strategy For Fixing Non-Compliant HTML Pages On The QA Focus Web Site* (case study-23)
- *Providing Access to an EU-funded Project Web Site after Completion of Funding* (case-study-17)
- *Exploiting ACRONYM And ABBR HTML Elements* (case-study-29)
- *Using The ACRONYM And ABBR HTML Elements On The QA Focus Web Site* (case-study-30)

Case studies which cover specific technical areas are available within the section on the appropriate technical area.

# The Purpose Of Your Project Web Site

Briefing

### About This Document

This briefing document describes the importance of having a clear importance of the purpose of your Web site

### Citation Details

*The Purpose Of Your Project Web Site*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-15/>>

**Keywords:** Web, purpose, briefing

## Background

Before creating a Web site for your project you should give some thought to the purpose of the Web site, including the aims of the Web site, the target audiences, the lifetime, resources available to develop and maintain the Web site and the technical architecture to be used. You should also think about what will happen to the Web site once project funding has finished.

## Purposes

Your project Web site could have a number of purposes. For example:

- The Web site could provide information about the project.
- The Web site could provide access to the project deliverables.
- The Web site could be used to support communications with members of the project team.
- The Web site could act as a repository of information about the management of the projects, including minutes of meetings, reports to funders, etc.

Your Web site could, of course, fulfil more than a single role. Alternatively you may choose to provide more than one Web site.

## Why You Need To Think About The Different Purposes

You should have an idea of the purposes of your project Web site before creating it for a number of reasons:

- You may wish to have more stringent QA procedures for Web sites which are intended for a broad audience and which is intended to have a longer lifetime.
- You may wish to be proactive in promoting a Web site intended for public use.
- You may wish to be proactive in ensuring that a Web site which is **not** intended for public use does not become indexing by search engines or is not linked to be mistake.
- You may wish to allow a public Web site to be indexed and archived by third parties.
- You may wish to ensure that a Web site intended for use by project partners or other closed communities is not indexed or archived, especially if there may be confidentiality or data protection issues.

### **Web Site For Information About The Project**

Once funding has been approved for your project Web site you may wish to provide information about the project, often prior to the official launch of the project and before project staff are in post. There is a potential danger that this information will be indexed by search engines or treated as the official project page. You should therefore ensure that the page is updated once an official project Web site is launched so that a link is provided to the official project page. You may also wish to consider stopping search engines from indexing such pages by use of the Standard For Robot Exclusion [1].

### **Web Site For Access To Project Deliverables**

Many projects will have an official project Web site. This is likely to provide information about the project such as details of funding, project timescales and deliverables, contact addresses, etc. The Web site may also provide access to project deliverables, or provide links to project deliverables if they are deployed elsewhere or are available from a repository.

Usually you will be proactive in ensuring that the official project Web site is easily found. You may wish to submit the project Web site to search engines.

### **Web Site To Support Communications With Project Partners**

Projects with several partners may have a Web site which is used to support communications with project partners. The Web site may provide access to mailing lists, realtime communications, decision-making support, etc. The JISCMail service may be used or commercial equivalents such as YahooGroups. Alternatively this function may be provided by a Web site which also provides a repository for project resources.

### **Web Site As Repository For Project Resources**

Projects with several partners may have a Web site which is used to provide a repository for project resources. The Web site may contain project plans, specifications, minutes of meetings, reports to funders, financial information, etc. The Web site may be part of the main project Web site, may be a separate Web site (possibly hosted by one of the project partners) or may be provided by a third party. You will need to think about the mechanisms for allowing access to authorised users, especially if the Web site contains confidential or sensitive information.

### **References**

1. **robots.txt Robots Exclusion Standard**,  
<<http://www.robotstxt.org/>>

# URI Naming Conventions For Your Project Web Site

Briefing

### About This Document

This briefing document describes the importance of establishing URI naming policies for your Web site.

### Citation Details

*URI Naming Conventions For Your Project Web Site*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-16/>>

**Keywords:** Web, URI, URL, guidelines, briefing

## Background

Once you have agreed on the purpose(s) of your project Web site(s) [1] you will need to choose a domain name for your Web site and conventions for URIs. It is necessary to do this since this can affect (a) The memorability of the Web site and the ease with which it can be cited; (b) The ease with which resources can be indexed by search engines and (c) The ease with which resources can be managed and repurposed.

## Domain Name

You may wish to make use of a separate domain name for your project Web site. If you wish to use a .ac.uk domain name you will need to ask UKERNA. You should first check the UKERNA rules [2]. A separate domain name has advantages (memorability, ease of indexing and repurposing, etc) but this may not be appropriate, especially for short-term projects. Your organisation may prefer to use an existing Web site domain.

## URI Naming Conventions

You should develop a policy for URIs for your Web site which may include:

- Conventions on use of case (e.g. specifying that all resources should be in lower case), separators (e.g. a hyphen should be used to separate components of a URI) and permitted characters (e.g. spaces should not be used in URIs).
- Conventions on the directory structure. The directory structure may be based on the main functions provided by your Web site.
- Conventions on dates and version control. You may wish to agreed on a convention for including dates in URIs. You may also wish to agree on a convention for version control (which could make use of date information).
- Conventions for file names and formats.

## Issues

### Grouping Of Resources

It is strongly recommended that you make use of directories to group related resources. This is particularly important for the project Web site itself and for key areas of the Web site. The entry point for the Web site and key areas should be contained in the directory itself: e.g. use <http://www.foo.ac.uk/bar/> to refer to project *BAR* and not <http://www.foo.ac.uk/bar.html>) as this allows the *bar/* directory to be processed in its entirety, independently of other directories. Without this approach

automated tools such as indexing software, and tools for auditing, mirroring, preservation, etc. would process other directories.

### URI Persistency

You should seek to ensure that URIs are persistent. If you reorganise your Web site you are likely to find that internal links may be broken, that external links and bookmarks to your resources are broken, that citations to resources cease to work. You may wish to provide a policy on the persistency of URIs on your Web site.

### File Names and Formats

Ideally the address of a resource (the URI) will be independent of the format of the resource. Using appropriate Web server configuration options it is possible to cite resources in a way which is independent of the format of the resource. This should allow easy migration to new formats (e.g. HTML to XHTML) and, using a technology known as Transparent Content Negotiation [3] provide access to alternative formats (e.g. HTML or PDF) or even alternative language versions.

### File Names and Server-Side Technologies

Ideally URIs will be independent of the technology used to provide access to the resource. If server-side scripting technologies are given in the file extension for URIs (e.g. use of .asp, .jsp, .php, .cfm, etc. extensions) changing the server-side scripting technology would probably require changing URIs. This may also make mirroring and repurposing of resources more difficult.

### Static URIs Or Query Strings?

Ideally URIs will be memorable and allow resources to be easily indexed and repurposed. However use of Content Management Systems or databases to store resources often necessitates use of URIs which contain query strings containing input parameters to server-side applications. As described above this can cause problems.

## Possible Solutions

You should consider the following approaches which address some of the concerns:

- Using file extensions: e.g. foo refers to foo.html or foo.asp
- Using directory defaults: e.g. foo/ refers to foo/intro.html or foo/intro.asp
- Rewriting dynamic URIs to static URIs

## References

1. **The Purpose Of Your Project Web Site**, QA Focus,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-15/html/>>
2. **UKERNA**, <<http://www.ukerna.ac.uk>>
3. **Transparent Content Negotiation**, W3C,  
<<http://www.w3.org/Protocols/rfc2616/rfc2616-sec12.html>>



# Compliance With HTML Standards

Briefing

### About This Document

This briefing document summarises the importance of complying fully with HTML standards and approaches for checking compliance.

### Citation Details

*Compliance With HTML Standards*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-01/>>

**Keywords:** Web, HTML, standards, compliance, briefing

## Why Bother?

Compliance with HTML standards is needed for a number of reasons:

- HTML compliant resources are more likely to be accessible to a wide range of Web browsers including desktop browsers such as Internet Explorer, Netscape, Mozilla, Opera, Lynx and specialist browsers on PDAS, digital TVs, kiosks, etc.
- HTML compliant resources are more easily processed and repurposed by other applications.
- HTML compliant resources will be rendered more quickly by modern browsers.
- HTML compliance is required by the AAA W3C WAI accessibility guidelines.

## Which Standards?

The World Wide Web Consortium, W3C, recommend use of the XHTML 1.0 (or higher) standard. This has the advantage of being an XML application (allowing use of XML tools) and can be rendered by most browsers. However authoring tools which are widely deployed may not yet produce XHTML and there may be financial implications (licence costs, training, etc.) in upgrading. In such circumstances HTML 4.0 may be used.

Cascading style sheets (CSS) should be used in conjunction with XHTML/HTML to describe the appearance of Web resources.

## Approaches To Creating HTML Resources

Web resources may be created in a number of ways. Often HTML authoring tools such as DreamWeaver, FrontPage, etc. are used, although experienced HTML authors may prefer to use a simple editing tool. Another approach is to make use of a Content Management System. An alternative approach is to convert proprietary file formats (e.g. MS Word or PowerPoint). In addition sometimes proprietary formats are not converted but are stored in their native format.

## Monitoring Compliance

A number of approaches may be taken to monitoring compliance with HTML standards. For example you can make use of validation features provided by modern HTML authoring tools, use desktop compliance tools or Web-based compliance tools.

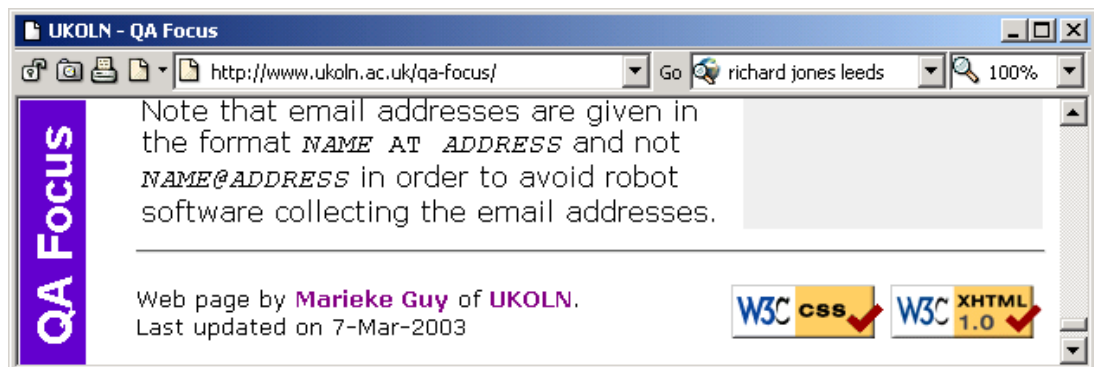
The different tools can be used in various ways. Tools integrated with an HTML authoring tool are used by the page author. It is important that the author is trained to use such tools on a regular basis. It should be noted that it may be difficult to address systematic errors (e.g. all files missing the DOCTYPE declaration) with this approach.

A popular approach is to make use of SSIs (server-side includes) to retrieve common features (such as headers, footers, navigation bars, etc.). This can be useful for storing HTML elements (such as the DOCTYPE declaration) in a manageable form. However this may cause validation problems if the SSI is not processed.

Another approach is to make use of a Content Management System (CMS) or similar server-side technique, such as retrieving resources from a database. In this case it is essential that the template used by the CMS complies with standards.

It may be felt necessary to separate the compliance process from the page authoring. In such cases use of a dedicated HTML checker may be needed. Such tools are often used in batch, to validate multiple files. In many cases voluminous warnings and error messages may be provided. This information may provide indications of systematic errors which should be addressed in workflow processes.

An alternative approach is to use Web-based checking services. An advantage with this approach is that the service may be used in a number of ways: the service may be used directly by entering the URL of a resource to be validated or live access to the checking service may be provided by including a link from a validation icon as used at <http://www.ukoln.ac.uk/qa-focus/> as shown in Figure 1 (this approach could be combined with use of cookies or other techniques so that the icon is only displayed to an administrator).



**Figure 1: Using Icons As ‘Live Links to Validation Services**

Another approach is to configure your Web server so that users can access the validation service by appending an option to the URL. For further information on this technique see <http://www.ukoln.ac.uk/,tools> and <http://www.ariadne.ac.uk/issue34/web-focus/>. This technique can be deployed with a simple option on your Web server’s configuration file.

# Use Of Proprietary Formats On Web Sites

Briefing

### About This Document

This briefing document describes approaches which can be taken when it is felt necessary to provide access to proprietary file formats on a Web site.

### Citation Details

*Use Of Proprietary Formats On Web Sites*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-03/>>

**Keywords:** Web, standards, proprietary formats, MS Word, MS PowerPoint, Microsoft, briefing

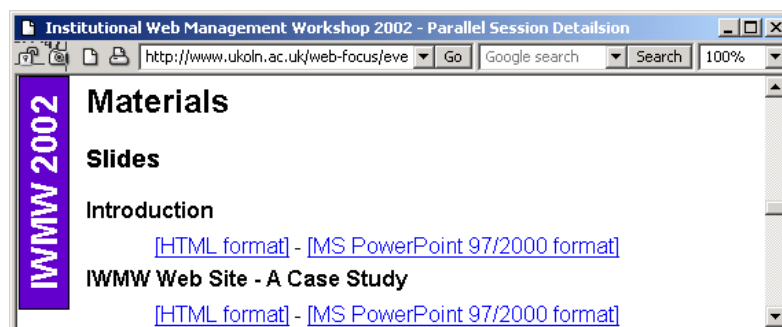
## Use Of Proprietary Formats

Although it is desirable to make use of open standards such as XHTML when providing access to resources on Web sites there may be occasions when it is felt necessary to use proprietary formats. For example:

- It is necessary to provide control over the page appearance which can only be achieved easily using formats such as MS Word, Adobe PDF, etc.
- Open standards are not yet available or tools are not yet widely deployed.
- Web access is used as a mechanism for delivering files.
- Conversion to an open format would be resource-intensive.
- The resource was created by a third-party and you do not have access to tools to convert the resource.

## URL Naming Conventions For Access To Proprietary Formats

If it is necessary to provide access to a proprietary file format you should not cite the URL of the proprietary file format directly. Instead you should give the URL of a native Web resource, typically a HTML page. The HTML page can provide additional information about the proprietary format, such as the format type, version details, file size, etc. If the resource is made available in an open format at a later date the HTML page can be updated to provide access to the open format – this would not be possible if the URL of the proprietary file was used.



An example of this approach is illustrated. In this case access to MS PowerPoint slides are available from a HTML page. The link to the file contains information on the PowerPoint version details.

Note that HTML resources which are generated in this way should be stored in their own directory so that they can be managed separately from other resources.

### Converting Proprietary Formats

Various tools may be available to convert resources from a proprietary format to HTML. Many authoring tools nowadays will enable resources to be exported to HTML format. However the HTML may not comply with HTML standards or use CSS and it may not be possible to control the look-and-feel of the generated resource.

Another approach is to use a specialist conversion tool which may provide greater control over the appearance of the output, ensure compliance with HTML standards, make use of CSS, etc.

If you use a tool to convert a resource to HTML it is advisable to store the generated resource in its own directory in order to be able to manage the master resource and its surrogate separately.

You should also note that some conversion tools can be used dynamically, allowing a proprietary format to be converted to HTML on-the-fly.

#### MS Word

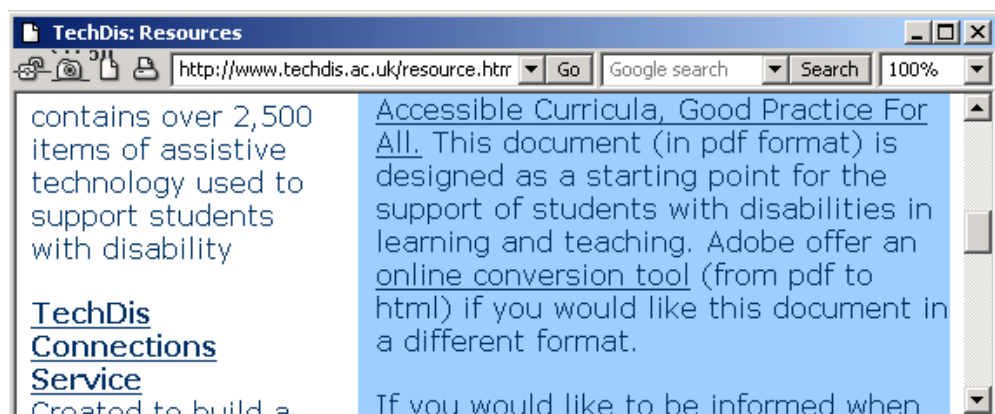
MS Word files can be saved as HTML from within MS Word itself. However the HTML that is created is of poor quality, often including proprietary or deprecated HTML elements and using CSS in a form which is difficult to reuse.

#### MS PowerPoint

MS PowerPoint files can be saved as HTML from within MS PowerPoint itself. However the `Save As` option provides little control over the output. The recommended approach is to use the `Save As Web Page` option and then to choose the `Publish` button. You should then ensure that the HTML can be read by all browsers (and not just IE 4.0 or later). You should also ensure that the file has a meaningful title and the output is stored in its own directory.

### Dynamic Conversion

In some circumstances it may be possible to provide a link to an online conversion service. Use of Adobe's online conversion service for converting files from PDF is illustrated.



It should be noted that this approach may result in a loss of quality from the original resource and is dependent on the availability of the remote service. However in certain circumstances it may be useful.

# Approaches To Link Checking

Briefing

### About This Document

This briefing document describes approaches to use of link checking tools and highlights possible limitations of such tools.

### Citation Details

*Approaches To Link Checking*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-07/>>

**Keywords:** Web, links, link checking, briefing

## Why Bother?

There are several reasons why it is important to ensure that links on Web sites work correctly:

- Web sites are based on hyperlinking, and if hyperlinks fail to work, the Web site can be regarded as not working correctly.
- Broken links reflect badly on the body hosting the Web sites.
- Hyperlinks are increasingly being used to deliver the functionality of Web sites, through links to JavaScript resources, style sheets files, metadata, etc. Broken links to these resources will result in the Web site not functioning as desired.

However there are resource implications in maintaining link integrity.

## Approaches To Link Checking

A number of approaches can be taken to checking broken links:

- Web site maintainer may run a link checking tool.
- A server-based link checking tool may send email notification of broken links.
- A remote link checking service may send email notification of broken links.
- Web server error log files may be analysed for requests for non-existent resources.
- Web server 404 error pages may provide a mechanism for users notifying the Web site maintainer of broken links.

Note that these approaches are not exclusive: Web site maintainers may choose to make use of several approaches.

## Policy Issues

There is a need to implement a policy on link checking. The policy could be that links will not be checked or fixed – this policy might be implemented for a project Web site once the funding has finished. For a small-scale project Web site the policy may be to check links when resources are added or updated or if broken links are brought to the project's attention, but not to check existing resources – this is likely to be an implicit policy for some projects.

For a Web site one which has a high visibility or gives a high priority to the effectiveness of the Web site, a pro-active link checking policy will be needed. Such a policy is likely to document the frequency of link checking, and the procedures for fixing broken links. As an example of approaches taken to link checking by a JISC service, see the article about the SOSIG subject gateway [1].

### Tools

Experienced Web developers will be familiar with desktop link-checking tools, and many lists of such tools are available [2] [3]. However desktop tools normally need to be used manually. An alternative approach is to use server-based link-checking software which send email notification of broken links.

Externally-hosted link-checking tools may also be used. Tools such as **LinkValet** [4] can be used interactively or in batch. Such tools may provide limited checking for free, with a licence fee for more comprehensive checking.

Another approach is to use a browser interface to tools, possibly using a **Bookmarklet** [5] although UKOLN's server-based **,tools** approach [6] [7] is more manageable.

### Other Issues

It is important to ensure that link checkers check for links other than `<a href="">`. There is a need to check that external JavaScript and CSS files (referred to by the `<link>` tag) and that checks are carried out on personalised interfaces to resources.

### References

- 1 **A Spring-Clean For SOSIG: A Systematic Approach To Collection Management**, Huxley, Place, Boyd and Cross, Ariadne, issue 33, <http://www.ariadne.ac.uk/issue33/planet-sosig/>
- 2 **Open Directory**, [http://dmoz.org/Computers/Software/Internet/Site\\_Management/Link\\_Management/](http://dmoz.org/Computers/Software/Internet/Site_Management/Link_Management/)
- 3 **Google Directory**, [http://directory.google.com/Top/Computers/Software/Internet/Site\\_Management/Link\\_Management/](http://directory.google.com/Top/Computers/Software/Internet/Site_Management/Link_Management/)
- 4 **LinkValet**, <http://www.htmlhelp.com/tools/valet/>
- 5 **Bookmarklets** <http://www.bookmarklets.com/>
- 6 **,tools**, <http://www.ukoln.ac.uk/,tools>
- 7 **Interfaces To Web Testing Tools**, Ariadne issue 34, <http://www.ariadne.ac.uk/issue34/web-focus/>

# Search Facilities For Your Web Site

Briefing

### About This Document

This briefing document describes approaches for providing search facilities on a Web site in order to improve its usability.

### Citation Details

*Search Facilities For Your Web Site*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-08/>>

**Keywords:** Web, searching, briefing

## Background

Web sites which contain more than a handful of pages should provide a search facility. This is important for several reasons:

- End users will not be aware of the size of your Web site, and so will wish to use a search tool to explore it.
- End users will not necessarily understand the browse structure of your Web site.
- Search engines are one of the most widely used tools by end users.

However there are resource implications in maintaining link integrity.

## Approaches To Providing Search Facilities

The two main approaches to the provision of search engines on a Web site are to host a search engine locally or to make use of an externally-hosted search engine.

### Local Search Engine

The traditional approach is to install search engine software locally. The software may be open source (such as **ht://Dig** [1]) or licensed software (such as **Inktomi** [2]). It should be noted that the search engine software does not have to be installed on the same system as the Web server. This means that you are not constrained to using the same operating system environment for your search engine as your Web server.

Because the search engine software can be hosted separately from the main Web server it may be possible to make use of an existing search engine service within the organisation which can be extended to index a new Web site.

### Externally-Hosted Search Engine

An alternative approach is to allow a third party to index your Web site. There are a number of companies which provide such services. Some of these services are free: they may be funded by advertising revenue. Such services include **Google** [3], **Atomz** [4] and **FreeFind** [5].

## Pros And Cons

Using a **locally-installed search engine** gives you control over the software. You can control the resources to be indexed and those to be excluded, the indexing frequency, the user interface, etc. However such control may have a price: you may need to have technical expertise in order to install, configure and maintain the software.

Using an **externally-hosted search engine** can remove the need for technical expertise: installing an externally-hosted search engine typically requires simply completing a Web form and then adding some HTML code to your Web site. However this ease-of-use has its disadvantages: typically you will lose the control over the resources to be indexed, the indexing frequency, the user interfaces, etc. In addition there is the dependency on a third party, and the dangers of a loss of service if the organisation changes its usage conditions, goes out of business, etc.

## Trends

Surveys of search facilities used on UK University Web sites have been carried out since 1998 [6]. This provides information not only on the search engines tools used, but also to spot trends.

Since the surveys began the most widely used tool has been **ht://Dig** – an open source product. In recent years the licensed product **Inktomi** has shown a growth in usage. Interestingly, use of home-grown software and specialist products has decreased – search engine software appears now to be a commodity product.

Another interesting trend appears to be in the provision of two search facilities; a locally-hosted search engine and a remote one – e.g. see the University of Lancaster [7].

## References

- 1 **ht://Dig**, <<http://www.htdig.org/>>
- 2 **Inktomi**, <<http://www.inktomi.com/>>
- 3 **Google**, <<http://services.google.com/googleuniv/login>>
- 4 **Atomz**, <<http://www.atomz.com/>>
- 5 **FreeFind** <<http://www.freefind.com/>>
- 6 **Surveys of Search Engines on UK University Web Sites**,  
<<http://www.ukoln.ac.uk/web-focus/surveys/uk-he-search-engines/>>
- 7 **University of Lancaster Search Page**,  
<<http://www.lancs.ac.uk/search.htm>>



# 404 Error Pages On Web Sites

Briefing

### About This Document

This briefing document describes how to improve the usability of a Web site by providing an appropriately configured 404 error page.

### Citation Details

*404 Error Pages On Web Sites*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-05/>>

**Keywords:** Web, 404, error page, briefing

## Importance Of 404 Error Pages

A Web sites 404 error page can be one of the most widely accessed pages on a Web site. The 404 error page can also act as an important navigational tool, helping users to quickly find the resource they were looking for. It is therefore important that 404 error pages provide adequate navigational facilities. In addition, since the page is likely to be accessed by many users, it is desirable that the page has an attractive design which reflects the Web sites look-and-feel.

## Types Of 404 Error Pages

Web servers will be configured with a default 404 error page. This default is typically very basic.

In the example shown the 404 page provides no branding, help information, navigational bars, etc.



An example of a richer 404 error page is illustrated. In this example the 404 page is branded with the Web site's colour scheme, contains the Web site's standard navigational facility and provide help information.



### Functionality Of 404 Error Pages

It is possible to define a number of types of 404 error pages:

#### **Server Default**

The server default 404 message is very basic. It will not carry any branding or navigational features which are relevant to the Web site.

#### **Simple Branding, Navigational Features Or Help Information**

The simplest approach to configuring a 404 page is to add some simple branding (such as the name of the Web site) or basic navigation features (link to the home page) or help information (an email address).

#### **Richer Branding, Navigational Features, Help Information Or Additional Features**

Some 404 pages will make use of the Web sites visual identity (such as a logo) and will contain a navigational bar which provides access to several areas of the Web site. In addition more complete help information may be provided as well as additional features such as a search facility.

#### **Full Branding, Navigational Features, Help Information And Additional Features**

A comprehensive 404 page will ensure that all aspects of branding, navigational features, help information and additional features such as a search facility are provided.

#### **As Above Plus Enhanced Functionality**

It is possible to provide enhanced functionality for 404 pages such as context sensitive help information or navigational facilities, feedback mechanisms to the page author, etc.

### Further Information

An article on 404 error pages, based on a survey of 404 pages in UK Universities is available at <<http://www.ariadne.ac.uk/issue20/404/>>. An update is available at <<http://www.ariadne.ac.uk/issue32/web-watch/>>.

## Enhancing Web Site Navigation Using The LINK Element

Briefing

### About This Document

This briefing document describes how the <LINK> element can be used to enhance the usability of a Web site.

### Citation Details

*Enhancing Web Site Navigation Using The LINK Element*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-10/>>

**Keywords:** Web, HTML, link, . navigation, usability, briefing

### About This Document

This document provides advice on how the HTML <link> element can be used to improve the navigation of Web sites.

### The LINK Element

#### About

The purpose of the HTML <link> element is to specify relationships with other documents. Although not widely used the <link> element provides a mechanism for improving the navigation of Web sites.

The <link> element should be included in the <head> of HTML documents. The syntax of the element is: <link rel="relation" href="url">. The key relationships which can improve navigation are listed below.

Relation	Function
next	Refers to the next document in a linear sequence of documents.
prev	Refers to the previous document in a linear sequence of documents.
home	Refers to the home page or the top of some hierarchy.
start	Refers to the first document in a collection of documents.
contents	Refers to a document serving as a table of contents.
help	Refers to a document offering help.
glossary	Refers to a document providing a glossary of terms that pertain to the current document.

**Table 1: Key Link Relations**

#### Benefits

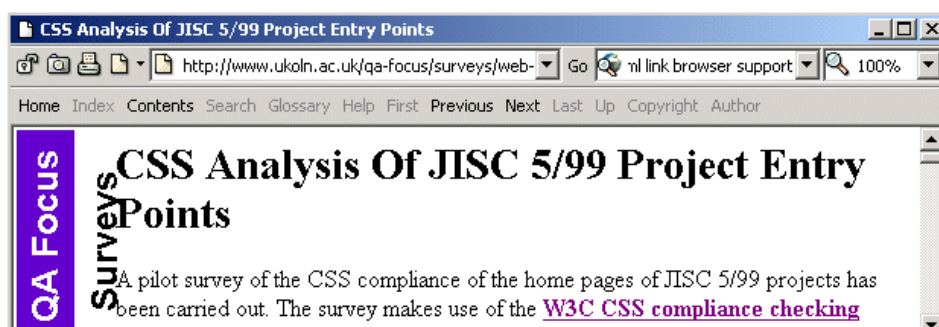
Use of the <link> element enables navigation to be provided in a consistent manner as part of the browser navigation area rather than being located in an arbitrary location in the Web page. This has accessibility benefits. In addition browsers can potentially enhance the performance by pre-fetching the next page in a sequence.

### Browser Support

A reason why `<link>` is not widely used has been the lack of browser support. This has changed recently and support is now provided in the latest versions of the Opera and Netscape/Mozilla browsers and by specialist browsers (e.g. iCab and Lynx).

Since the `<link>` element degrades gracefully (it does not cause problems for old browser) use of the `<link>` element will cause no problems for users of old browsers.

An illustration of how the `<link>` element is implemented in Opera is shown below.



**Figure 1: Browser Support For The `<link>` Element**

In Figure 1 a menu of navigational aids is available. The highlighted options (Home, Contents, Previous and Next) are based on the relationships which have been defined in the document. Users can use these navigational options to access the appropriate pages, even though there may be no corresponding links provided in the HTML document.

### Information Management Challenges

It is important that the link relationships are provided in a manageable way. It would not be advisable to create link relationships by manually embedding them in HTML pages if the information is liable to change.

It is advisable to spend time in defining the on key navigational locations, such as the Home page (is it the Web site entry point, or the top of a sub-area of the Web site). Such relationships may be added to templates included in SSIs. Server-side scripts are a useful mechanism for exploiting other relationships, such as Next and Previous – for example in search results pages.

# Accessing Your Web Site On A PDA

Briefing

### About This Document

This briefing document describes how making your Web site available on a PDA using a freely available service can help to identify possible problem areas when your Web resources are repurposed by tools.

### Citation Details

*Accessing Your Web Site On A PDA*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-05/>>

**Keywords:** Web, PDA, repurposing, Avantgo, briefing

### About This Document

With the growing popularity in use of mobile devices and pervasive networking on the horizon we can expect to see greater use of PDAs (Personal Digital Assistants) for accessing Web resources.

This document describes a method for accessing a Web site on a PDA. In addition this document highlights issues which may make access on a PDA more difficult.

## AvantGo

### About

AvantGo is a well-known Web based service which provides access to Web resources on a PDA such as a Palm or Pocket PC.

The AvantGo service is freely available from  
<<http://www.avantgo.com/>>.

Once you have registered on the service you can provide access to a number of dedicated AvantGo channels. In addition you can use an AvantGo wizard to provide access to any publicly available Web resources on your PDA.

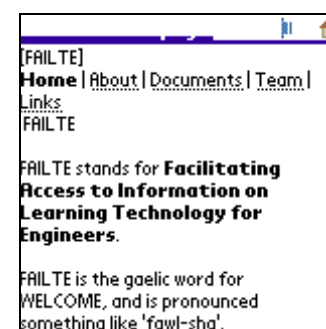
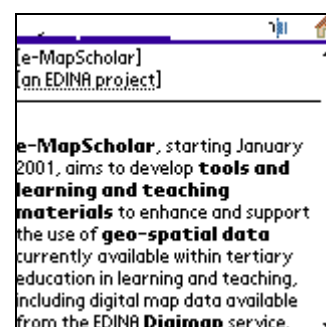
An example of two Web sites showing the interface on a Palm is illustrated.

### Benefits

If you have a PDA you may find it useful to use it to provide access to your Web site, as this will enable you to access resources when you are away from your desktop PC. This may also be useful for your project partners. In addition you may wish to encourage users of your Web site to access it in this way.

### Other Benefits

AvantGo uses robot software to access your Web site and process it in a format suitable for viewing on a PDA, which typically has more limited functionality, memory, and viewing area than a desktop PC. The robot software may not process a number of



features which may be regarded as standard on desktop browsers, such as frames, JavaScript, cookies, plugins, etc.

The ability to access a simplified version of your Web site can provide a useful mechanism for evaluating the ease with which your Web site can be repurposed and for testing the user interface under non-standard environments.

You should be aware of the following potential problem areas:

#### **Entry Point Not Contained In Project Directory**

If the project entry point is not contained in the project's directory, it is likely that the AvantGo robot will attempt to download an entire Web site and not just the project area.

#### **Frames**

If your Web site contains frames and you do not use the appropriate option to ensure that the full content can be accessed by user agents which do not support frames (such as the AvantGo robot software) resources on your Web site will not be accessible.

#### **Plugin Technologies**

If your Web site contains technologies which require plugins (such as Flash, Java, etc.) you will not be able to access the resources.

### **Summary**

As well as providing enhanced access to your Web site use of tools such as AvantGo can assist in testing access to your Web site. If your Web site makes use of open standards and follows best practices it is more likely that it will be usable on a PDA and by other specialist devices.

**You should note, however, that use of open standards and best practices will not guarantee that a Web site will be accessible on a PDA.**

# How To Evaluate A Web Site's Accessibility Level

Briefing

### About This Document

This briefing document describes approaches for evaluating the accessibility of a Web site.

### Citation Details

*How To Evaluate A Web Site's Accessibility Level*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-12/>>

**Keywords:** Web, accessibility, WAI, briefing

## Background

Many Web developers and administrators are conscious of the need to ensure that their Web sites reach as high a level of accessibility as possible. But how do you actually find out if a site has accessibility problems? Certainly, you cannot assume that if no complaints have been received through the site feedback facility (assuming you have one), there are no problems. Many people affected by accessibility problems will just give up and go somewhere else.

So you must be proactive in rooting out any problems as soon as possible. Fortunately there are a number of handy ways to help you get an idea of the level of accessibility of the site which do not require an in-depth understanding of Web design or accessibility issues. It may be impractical to test every page, but try to make sure you check the Home page plus as many high traffic pages as possible.

## Get A Disabled Person To Look At The Site

If you have a disability, you have no doubt already discovered whether your site has accessibility problems which affect you. If you know someone with a disability which might prevent them accessing information in the site, then ask them to browse the site, and tell you of any problems. Particularly affected groups include visually impaired people (blind, colour blind, short or long sighted), dyslexic people and people with motor disabilities (who may not be able to use a mouse).

If you are in Higher Education your local Access Centre [1] may be able to help.

## View The Site Through A Text Browser

Get hold of a text browser such as *Lynx* [2] and use it to browse your site. Problems you might uncover include those caused by images with no, or misleading, alternative text, confusing navigation systems, reliance on scripting or poor use of frames.

## Browse The Site Using A Speech Browser

You can get a free evaluation version of *IBM's Homepage Reader* [3] or *pwWebSpeak* [4] speech browsers used by many visually impaired users of the Web. The browsers "speak" the page to you, so shut your eyes and try to comprehend what you are hearing.

Alternatively, try asking a colleague to read you the Web page out loud. Without seeing the page, can you understand what you're hearing?

### Look At The Site Under Different Conditions

As suggested by the World Wide Web Consortium (W3C) Web Accessibility Initiative [5], you should test your site under various conditions to see if there are any problems including (a) graphics not loaded, (b) frames, scripts and style sheets turned off and (c) browsing without using a mouse. Also, try using *bookmarklets* or *favelets* to test your site under different conditions: a useful list of accessibility favelets can be found at [6].

### Check With Automatic Validation Tools

There are a number of Web-based tools which can provide valuable information on some potential accessibility problems. These include *Bobby* [7] and *The Wave* tools [8].

You should also check whether the underlying HTML of your site validates to accepted standards using the World Wide Web Consortium's *MarkUp Validation Service* [9] as non-standard HTML frequently creates accessibility barriers.

### Acting on Your Observations

Details of any problems found should be noted: the effect of the problem, which page was affected, plus why you think the problem was caused. You are unlikely to catch all accessibility problems in the site, but the tests described here will give you an indication of whether the site requires immediate attention to raise accessibility. Remember that improving accessibility for specific groups, such as visually impaired people, will often have usability benefits for all users.

### Commission an Accessibility Audit

Since it is unlikely you will catch all accessibility problems and the learning curve is steep, it may be advisable to commission an expert accessibility audit. In this way, you can receive a comprehensive audit of the subject site, complete with detailed prioritised recommendations for upgrading the level of accessibility of the site. Groups which provide such audits include the Digital Media Access Group, University of Dundee and the RNIB (who also audit for access by the blind).

### Acknowledgments

This document was written by David Sloan, DMAG, University of Dundee and originally published at <<http://www.techdis.ac.uk/resources/dsloan01.html>>. We are grateful for permission to republish this document.

### References

1. **Access Centres**, <<http://www.nfac.org.uk/>>
2. **Lynx**, <<http://lynx.isc.org/release/>>
3. **IBM's Homepage Reader**, IBM, <<http://www-3.ibm.com/able/hprtrial3.html>>
4. **pwWebSpeak**, <<http://www.soundlinks.com/pwgen.htm>>
5. **Accessibility Bookmarklets**, <<http://www.visionaustralia.org.au/webaccessibility/workshops/references.html#bm>>
6. **Web Content Accessibility Guidelines**, W3C. <<http://www.w3.org/TR/WAI-WEBCONTENT/>>
7. **Bobby**, <<http://bobby.watchfire.com/bobby/html/en/>>
8. **WAVE**, <<http://www.temple.edu/instituteondisabilities/piat/wave/>>
9. **W3C HTML Validator**, W3C, <<http://validator.w3.org/>>



# Performance Indicators For Your Project Web Site

Briefing

### About This Document

This briefing document describes a number of performance indicators for your Web site.

### Citation Details

*Performance Indicators For Your Project Web Site*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-17/>>

**Keywords:** Web, usage figures, statistics, performance indicators, briefing

## Background

It is desirable to measure usage of your project Web site as this can give an indication of its effectiveness. Measuring how the Web site is being used can also help in identifying the usability of the Web site. Monitoring errors when users access your Web site can also help in identifying problem areas which need to be fixed.

However, as described in this document, usage statistics can be misleading. Care must be taken in interpreting statistics. As well as usage statistics there are a number of other types of performance indicators which can be measured.

It is also important that consistent approaches are taken in measuring performance indicators in order to ensure that valid comparisons can be made with other Web sites.

## Web Statistics

Web statistics are produced by the Web server software. The raw data will normally be produced by default - no additional configuration will be needed to produce the server's default set of usage data.

The server log file records information on requests (normally referred to as a "hit") for a resource on the web server. Information included in the server log file includes the name of the resource, the IP address (or domain name) of the user making the request, the name of the browser (more correctly, referred to as the "user agent") issuing the request, the size of the resource, date and time information and whether the request was successful or not (and an error code if it was not). In addition many servers will be configured to store additional information, such as the "referrer" (sic) field, the URL of the page the user was viewing before clicking on a link to get to the resource.

## Tools

A wide range of Web statistical analysis packages are available to analyse Web server log files [1]. A widely used package in the UK HE sector is WebTrends [2].

An alternative approach to using Web statistical analysis packages is to make use of externally-hosted statistical analysis services [3]. This approach may be worth considering for projects which have limited access to server log files and to Web statistical analysis software.

## Configuration Issues

In order to ensure that Web usage figures are consistent it is necessary to ensure that Web servers are configured in a consistent manner, that Web statistical analysis packages process the data consistently and that the project Web site is clearly defined.

You should ensure that (a) the Web server is configured so that appropriate information is recorded and (b) that changes to relevant server options or data processing are documented.

## Limitations

You should be aware that the Web usage data does not necessarily give a true indication of usage due to several factors:

- Effects of caching.
- Effects of access from robots, off-line browsers, auditing tools, etc.
- Difficulties of measuring unique visitors, browser types, etc. accurately.
- Difficulties of defining terms such as sessions.

Despite these reservations collecting and analysing usage data can provide valuable information.

## Other Types Of Indicators

Web usage statistics are not the only type of performance indicator which can be used. You may also wish to consider:

- **Monitoring the number of links to your Web site:** tools such as LinkPopularity.com can report on the number of links to your Web site.
- **Numbers of resources indexed:** You can analyse the numbers of resources indexed by search engines such as Google.
- **Error log analysis:** Analysis of your server log error file can indicate problem areas.

With all of the indicators periodic reporting will allow trends to be detected.

## Conclusions

It may be useful to determine a policy on collection and analysis of performance indicators for your Web site prior to its launch.

## References

1. **Web server log files**, UKOLN,  
<<http://www.ukoln.ac.uk/nof/support/help/papers/performance/>>
2. **WebTrends**,  
<<http://www.netiq.com/webtrends/>>
3. **Externally-hosted statistical analysis services**, Exploit Interactive, issue 5,  
<<http://www.exploit-lib.org/issue5/indicators/>>

# Use Of Cascading Style Sheets (CSS)

Briefing

### About This Document

This briefing document describes the importance of CSS for Web sites.

### Citation Details

*Use Of Cascading Style Sheets (CSS)*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-34/>>

**Keywords:** Web, CSS, briefing

## Background

This document reviews the importance of Cascading Style Sheets (CSS) and highlights the importance of ensuring that use of CSS complies with CSS standards.

## Why Use CSS?

Use of CSS is the recommended way of defining how HTML pages are displayed. You should use HTML to define the basic structure (using elements such as <h1>, <p>, <li>, etc.) and CSS to define how these elements should appear (e.g. heading should be in bold Arial font, paragraphs should be indented, etc.).

This approach has several advantages:

**Maintenance:** It is much easier to maintain the appearance of a Web site. If you use a single CSS file updating this file allows the Web site look-and-feel to be altered easily; in contrast use of HTML formatting elements would require every file to be updated to change the appearance.

**Functionality:** CSS provides rich functionality, including defining the appearance of HTML pages when they are printed.

**Accessibility:** Use of CSS provides much greater accessibility, allowing users with special needs to alter the appearance of a Web page to suit their requirements. CSS also allows Web pages to be more easily rendered by special devices, such as speaking browsers, PDAs, etc.

There are disadvantages to use of CSS. In particular legacy browsers such as Netscape 4 have difficulty in processing CSS. However, since such legacy browsers are now in a minority the biggest barrier to deployment of CSS is probably inertia.

## Approaches To Use Of CSS

There are a number of ways in which CSS can be deployed:

**External CSS Files:** The best way to use CSS is to store the CSS data in an external file and link to this file using the <link> HTML element. This approach allows the CSS definitions to be used by every page on your Web site.

**Internal CSS:** You can store CSS within a HTML by including it using the <style> element within the <head> section at the top of your HTML file. However this approach means the style definitions cannot be applied to other files. This approach is not normally recommended.

**Inline CSS:** You can embed your CSS inline with HTML elements: for example `<p style="font-color: red">` uses CSS to specify that text in the current paragraph is red. However this approach means that the style definitions cannot be applied to other paragraphs. This approach is discouraged.

### Ensure That You Validate Your CSS

As with HTML, it is important that you validate your CSS to ensure that it complies with appropriate CSS standards. There are a number of approaches you can take:

**Within your HTML editor:** Your HTML editing tool may allow you to create CSS. If it does, it may also have a CSS validator.

**Within a dedicated CSS editor:** If you use a dedicated CSS editor, the tool may have a validator.

**Using an external CSS validator:** You may wish to use an external CSS validators, This could be a tool installed locally or a Web-based tool such as those available at W3C [1] and the Web Design Group [2].

Note that if you use external CSS files, you should also ensure that you check that the link to the file works.

### Systematic CSS Validation

You should ensure that you have systematic procedures for validating your CSS. If, for example, you make use of internal or inline CSS you will need to validate the CSS whenever you create or edit an HTML file. If, however, you use a small number of external CSS files and never embed CSS in individual HTML files you need only validate your CSS when you create or update one of the external CSS files.

### References

- 1 **Validator CSS**, W3C, <http://jigsaw.w3.org/css-validator/>
- 2 **CSSCheck**, WDG, <http://www.htmlhelp.com/tools/csscheck/>

# Mothballing Your Web Site

Briefing

### About This Document

This briefing document describes approaches to ‘mothballing’ a Web site once project funding finishes.

### Citation Details

*Mothballing Your Web Site*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-04/>>

**Keywords:** Web, mothballing, preservation, archiving, briefing

### About This Document

When the funding for a project finishes it is normally expected that the project’s Web site will continue to be available in order to ensure that information about the project, the project’s findings, reports, deliverables, etc. are still available.

This document provides advice on “mothballing” a project Web site.

### Web Site Content

The entry point for the project Web site should make it clear that the project has finished and that there is no guarantee that the Web site will be maintained.

You should seek to ensure that dates on the Web site include the year – avoid content which says, for example, “*The next project meeting will be held on 22 May*”.

You may also find it useful to make use of cascading style sheets (CSS) which could be used to, say, provide a watermark on all resources which indicate that the Web site is no longer being maintained.

### Technologies

Although software is not subject to deterioration due to aging, overuse, etc. software products can cease to work over time. Operating systems upgrades, upgrades to software libraries, conflicts with newly installed software, etc. can all result in software products used on a project Web site to cease working.

There are a number of areas to be aware of:

- If you are using unusual configuration features for the **Web server software**, the Web site may stop working if the server software is upgraded or replaced (you move from Microsoft’s IIS software to Apache).
- If you are using special features of the Web site’s **search engine software** aspects of the Web site may cease to work if the search engine software is upgraded or replaced.
- If you are using **online forms** on your Web site these may cease to work if the backend scripts are updated.
- If you are using a **Content Management System** or **server-side scripting technologies** (e.g. PHP, ASP, etc.) on your Web site these may cease to work if the backend technologies are updated.

- If you provide **automated feedback or annotation tools** which allow users to provide comments on resources on your Web site there is a danger that the tools may be used to submit spam or obscene messages. With popular feedback tools there may be automated devices which will submit inappropriate messages automatically.

## Process For Mothballing

We have outlined a number of areas in which a project Web site may degrade in quality once the project Web site has been “mothballed”.

In order to minimise the likelihood of this happening and to ensure that problems can be addressed with the minimum of effort it can be useful to adopt a systematic set of procedures when mothballing a Web site.

It can be helpful to **run a link checker** across your Web site. You should seek to ensure that all internal links (links to resources on your own Web site) work correctly. Ideally links to external resources will also work, but it is recognised that this may be difficult to achieve. It may be useful to provide a link to a report of the link check on your Web site.

It would be helpful to **provide documentation** on the technical architecture of your Web site, which describes the server software used (including use of any unusual features), use of server-side scripting technologies, content management systems, etc.

It may also be useful to **provide a mirror of your Web site** by using a mirroring package or off-line browser. This will ensure that there is a static version of your Web site available which is not dependent on server-side technologies.

## Contacts

You should give some thought to contact details provided on the Web site. You will probably wish to include details of the project staff, partners, etc. However you may wish to give an indication if staff have left the organisation.

Ideally you will provide contact details which are not tied down to a particular person. This may be needed if, for example, your project Web site has been hacked and the CERT security team need to make contact.

## Planning For Mothballing

Ideally you will ensure that your plans for mothballing your Web site are developed when you are preparing to launch your Web site!

## Edinburgh University Library Online: Work In Progress

Case study

### About This Document

This case study describes the approaches taken at the Edinburgh University Library towards ensuring that their Web site complied with accessibility guidelines.

### Citation Details

*Edinburgh University Library Online: Work In Progress*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-26/>>

### Related Documents

See also the *How To Evaluate A Web Site's Accessibility Level* (briefing-12) and *Use Of Cascading Style Sheets (CSS)* (briefing-34) briefing documents.

**Keywords:** Web, accessibility, Edinburgh University, Library Online, case study

## Background

**Library Online** [1] shown in Figure 1 is the main library Web site/portal for the University of Edinburgh [2]. Although clearly not a project site in itself, one of its functions is to provide a gateway to project sites with which the Library is associated [3].

In the last seven years or so it has grown to around 2,000 static pages plus an increasing amount of dynamic content, the main database-driven service being the related web-based Library Catalogue [4]. At the time of writing (October 2003), a proprietary Digital Object Management System has been purchased and is being actively developed.

This will no doubt impinge on some areas of the main site and, in time, probably the Catalogue: notably access to e-journals and other digital resources/collections. However, for the time being, Library Online and the Catalogue between them provide the basic information infrastructure.



**Figure 1: The Library Online Entry Point**

## Problem Being Addressed

The challenges include enhancing accessibility and usability; also maintaining standards as these develop. Problems exist with legacy (HTML) code, with increasingly deprecated layout designs and separating content from presentation. Addressing these issues globally presents real problems whilst maintaining currency and a continuous, uninterrupted service. It is, of course, a live site - and an increasingly busy one. There are currently over twenty members of staff editing and publishing with varying levels of expertise and no overall Content Management System, as such.

Policy has also been to maintain support for a whole range of older browsers, further complicating matters.

### The Approach Taken

Fortunately, the site design was based on Server-Side Includes (SSIs) and a great deal of effort was put into conforming to best practice guidelines as they were articulated over five years ago. The architecture appears to remain reasonably sound. So an incremental approach has been adopted generally, though some enhancements have been achieved quite rapidly across the board by editing sitewide SSIs. A recent example of the latter has been the introduction of the "Skip Navigation" accessibility feature across the whole site.

A fairly radical redesign of the front page was carried out within the last two years. This will need to be revisited before too long but the main focus is presently on the body of the site, initially higher level directories, concentrating on the most heavily-used key areas.

- Enhancements to accessibility and usability are documented in our fairly regularly updated accessibility statement [5]. These include:
- Font sizes are relative so users may resize text according to needs
- "Skip Navigation" feature introduced to allow text browsers and users of assistive technology to bypass repetitive elements (new for Autumn 2003)
- Stylesheets are being used extensively to separate content from presentation
- Content should still be visible to browsers with no support for stylesheets
- Pages are increasingly "resolution-independent" and able to expand or contract to fit users' screens
- Images have alternate text (alt) tags
- As far as possible, links are written to make sense out of context
- Main links have titles added which display in many visual browsers as "tooltips" but are also used by assistive technologies
- Where appropriate, text versions of pages are provided
- Page layouts still based on tables will have summaries amended to "null" to avoid confusion with data tables. Also table headers are not being used for layout tables
- Access keys have been introduced experimentally in an alternative contact form

None of these features should be contentious, though precise interpretations may vary. Many have been built in to the design since day one (e.g. "alt" tags); others have been applied retrospectively and incrementally. All are, we hope, worthwhile!

Additional functionality with which we are currently experimenting includes media styles, initially for print. The original site navigation design was quite graphically rich and not very "printer-friendly". Progress is being made in this area - but who knows what devices we may need to support in the future? Perhaps we shall eventually have to move to XML/XSLT as used within our Collections Gateway due for launch soon. Meanwhile, for Library Online, even XHTML remains no more than a possibility at present.

Our approach to site development is essentially based on template and stylesheet design, supported by Server-Side Include technology for ease of management and implementation. This largely takes care of quality assurance and our proposed approach to content management should underpin this. We are moving towards fuller adoption of



Dreamweaver (MX) for development and Macromedia Contribute for general publishing. Accessibility and usability quality assurance tools are already in regular use including LIFT Online and other resources identified on the site. It seems very likely that this will continue.

All this remains very much work in progress ... Upgrading legacy code, layout design, integration and interoperability with other information systems etc. Categorically, no claims are made for best practice; more a case of constantly striving towards this.

### Problems Experienced

The main problems experienced - apart from time and resources naturally - have been (a) the need to support older browsers; (b) the sheer number of static pages and (c) the amount of legacy code and use of tabular layouts

### Things We Would Do Differently

With the benefit of hindsight, perhaps stylesheets could have been implemented more structurally. Validity has always been regarded as paramount, while separation of true content from pure presentation might have been given equal weight(?) This is now being reassessed.

We might have derived some benefit from more extensive database deployment - and may well in the future - but we must constantly review, reappraise possibilities offered by new technologies etc. and, above all, listen to our users.

I have referred to some significant developments in prospect which present more opportunities to do things differently - but whether we get these right or wrong, there will always be scope for improvement on the Web, just as in the "real" world. Like politics, it seems to be the art of the possible - or should that be a science?

### References

1. **Library Online**, Edinburgh University Library, <<http://www.lib.ed.ac.uk/>>
2. **University of Edinburgh**, <<http://www.ed.ac.uk/>>
3. **Projects and Initiatives**, Edinburgh University Library, <<http://www.lib.ed.ac.uk/projects/>>
4. **Edinburgh University Library Catalogue**, <<http://catalogue.lib.ed.ac.uk/>>
5. **Design for Accessibility and Usability**, Edinburgh University Library, <<http://www.lib.ed.ac.uk/news/design.shtml>>

### Contact Details

Steve Scott  
Library Online Editor  
University of Edinburgh  
Edinburgh  
Email: [steve.scott@ed.ac.uk](mailto:steve.scott@ed.ac.uk)

# Gathering Usage Statistics And Performance Indicators: The (NMAP) Experience

Case study

### About This Document

This case study describes the use of Web site performance indicators by the NMAP project.

### Citation Details

*Gathering Usage Statistics And Performance Indicators: The (NMAP) Experience*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-06/>>

### Related Documents

See also the *Performance Indicators For Your Project Web Site* (briefing-17) briefing document.

**Keywords:** Web, usage, usage statistics, NMAP, case study

## Background

The NMAP project [1] was funded under the JISC 05/99 call for proposals to create the UK's gateway to high quality Internet resources for nurses, midwives and the allied health professions.

NMAP is part of the BIOME Service, the health and life science component of the national Resource Discovery Network (RDN), and closely integrated with the existing OMNI gateway. Internet resources relevant to the NMAP target audiences are identified and evaluated using the BIOME Evaluation Guidelines. If resources meet the criteria they are described and indexed and included in the database.

NMAP is a partnership led by the University of Nottingham with the University of Sheffield and Royal College of Nursing (RCN). Participation has also been encouraged from several professional bodies representing practitioners in these areas. The NMAP team have also been closely involved with the professional portals of the National electronic Library for Health (NeLH).

The NMAP service went live in April 2001 with 500 records.

The NMAP service was actively promoted in various journal, newsletters, etc. and presentations or demonstrations were given at a variety of conference and meetings. Extensive use was made of electronic communication, including mailing lists and newsgroups for promotion.

Work in the second year of the project included the creation of two VTS tutorials; the Internet for Nursing, Midwifery and Health Visiting, and the Internet for Allied Health.

## Problem Being Addressed

As one of the indicators of the success, or otherwise, in reaching the target group we wanted to know how often the NMAP service was being used, and ideally who they are and how they are using it.

The idea was to attempt to ensure we were meeting their needs, and also gain data which would help us to obtain further funding for the continuation of the service after the end of project funding.

There seems to be little standardisation of the ways in which this sort of data is collected or reported, and although we could monitor our own Web server, the use of

caching and proxy servers makes it very difficult to analyse how many times the information contained within NMAP is being used or where the users are coming from.

These difficulties in the collection and reporting of usage data have been recognised elsewhere, particularly by publishers of electronic journals who may be charging for access. An international group has now been set up to consider these issues under the title of project COUNTER [2] which issued a “Code of Practice” in January 2003.

### The Approach Taken

We took a variety of approaches to try to collect some meaningful data.

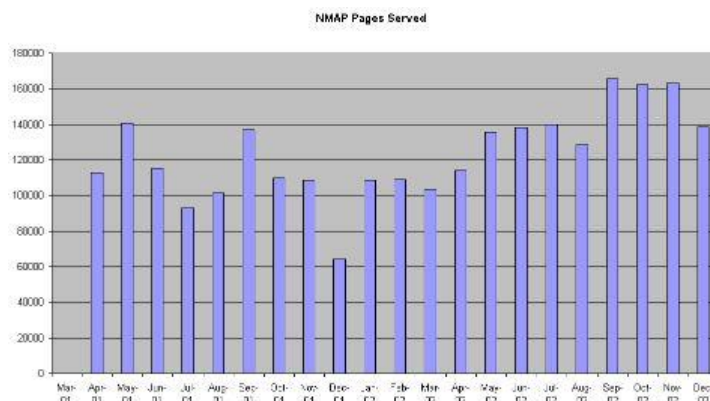
The first and most obvious of these is log files from the server which were produced monthly and gave a mass of data including:

- General Summary
- Monthly Report:
- Daily Summary
- Hourly Summary
- Domain Report
- Organisation Report
- Status Code Report
- File Size Report
- File Type Report
- Directory Report
- Request Report

A small section of one of the log files showing the general summary for November 2002 can be seen below. Note that figures in parentheses refer to the 7-day period ending 30-Nov-2002 23:59.

```
Successful requests: 162,910 (39,771)
Average successful requests per day: 5,430 (5,681)
Successful requests for pages: 162,222 (39,619)
Average successful requests for pages per day: 5,407 (5,659)
Failed requests: 2,042 (402)
Redirected requests: 16,514 (3,679)
Distinct files requested: 3,395 (3,217)
Unwanted logfile entries: 51,131
Data transferred: 6.786 Gbytes (1.727 Gbytes)
Average data transferred per day: 231.653 Mbytes (252.701 Mbytes)
```

A graph of the pages served can be seen in Figure 1.



**Figure 1: Pages served per month**

The log files also provided some interesting data on the geographical locations and services used by those accessing the NMAP service.

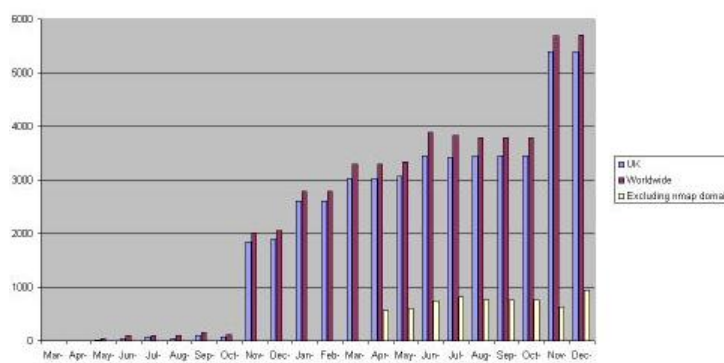
### 4.3 Technical Advice: Access/Web – Case Studies

Listing domains, sorted by the amount of traffic, example from December 2002, showing those over 1%.

Requests	% bytes	Domain
48237	31.59%	.com (Commercial)
40533	28.49%	[unresolved numerical addresses]
32325	24.75%	.uk (United Kingdom)
14360	8.52%	ac.uk
8670	7.29%	nhs.uk
8811	7.76%	.net (Network)
1511	1.15%	.edu (USA Educational)

**Table 1: Linking to NMAP**

A second approach was to see how many other sites were linking to the NMAP front page URL. AltaVista was used as it probably had the largest collection back in 2000 although this has now been overtaken by Google. A search was conducted each month using the syntax `link:http://nmap.ac.uk`. The results can be seen in Figure 2.



**Figure 2 - Number of sites linking to NMAP (according to AltaVista)**

The free version of the service provided by InternetSeer [3] was also used. This service checks a URL every hour and will send an email to one or more email addresses saying if the site is unavailable. This service also provides a weekly summary by email which, along with the advertising, includes a report in the format:

```
=====
Weekly Summary Report
=====
http://nmap.ac.uk
Total Outages: 0.00
Total time on error: 00:00
Percent Uptime: 100.0
Average Connect time*: 0.13
Outages- the number of times we were unable to access this URL
Time on Error- the total time this URL was not available (hr:min)
% Uptime- the percentage this URL was available for the day
Connect Time- the average time in seconds to connect to this URL
```

During the second year of the project we also conducted an online questionnaire with 671 users providing data about themselves, why they used NMAP and their thoughts on

its usefulness or otherwise, however this is beyond the scope of this case study and is being reported elsewhere.

## Problems Experienced

Although these techniques provided some useful trend data about the usage of the NMAP service there are a series of inaccuracies, partly due to the nature of the Internet, and some of the tools used.

The server log files are produced monthly (a couple of days in arrears) and initially included requests from the robots used by search engines, these were later removed from the figures. The resolution of the domains was also a problem with 28% listed as "unresolved numerical addresses" which gives no indication where the users are accessing from. In addition it is not possible to tell whether .net or .com users are in the UK or elsewhere. The number of accesses from .uk domains was encouraging and specifically those from .ac & .nhs domains. It is also likely (from data gathered in our user questionnaire) that many of the .net or .com users are students or staff in higher or further education or NHS staff who are accessing the NMAP service via a commercial ISP from home.

In addition during the first part of 2002 we wrote two tutorials for the RDN Virtual Training Suite (VTS) [4], which were hosted on the BIOME server and showed up in the number of accesses. These were moved in the later part of 2002 to the server at ILRT in Bristol and therefore no longer appear in the log files. It has not yet been possible to get access figures for the tutorials.

The "caching" of pages by ISPs and within .ac.uk and .nhs.uk servers does mean faster access for users but probably means that the number of users is undercounted in the log files.

The use of AltaVista "reverse lookup" to find out who was linking to the NMAP domain was also problematic. This database is infrequently updated which accounts for the jumps seen in Figure 2. Initially when we saw a large increase in November 2001 we thought this was due to our publicity activity and later realised that this was because it included internal links within the NMAP domain in this figure, therefore from April 2002 we collected another figure which excluded internal links linking to self.

None of these techniques can measure the number of times the records from within NMAP are being used at the BIOME or RDN levels of JISC services. In addition we have not been able to get regular data on the number of searches from within the NeLH professional portals which include an RDNi search box [5] to NMAP.

## Things We Would Do Differently

In the light of our experience with the NMAP project we recommend that there is a clear strategy to attempt to measure usage and gain some sort of profile of users of any similar service.

I would definitely use Google rather than AltaVista and would try to specify what is needed from log files at the outset. Other services have used user registration and therefore profiles and cookies to track usage patterns and all of these are worthy of consideration.

### References

1. **NMAP**, <<http://nmap.ac.uk>>
2. **Project COUNTER**, <<http://www.projectcounter.org/>>
3. **InternetSeer**, <<http://www.internetseer.com/>>
4. **RDN Virtual Training Suite**, RDN, <<http://www.vts.rdn.ac.uk/>>
5. **RDN-Include**, RDN, <<http://www.rdn.ac.uk/rdn-i/>>

### Contact Details

Rod Ward  
Greenfield Medical Library  
Queen's Medical Centre  
Nottingham NG7 2UH  
Email: [rw@biome.ac.uk](mailto:rw@biome.ac.uk)

# Usability Testing for the Non-Visual Access to the Digital Library (NoVA) Project

Case study

### About This Document

This case study describes approaches to usability and accessibility testing by the NoVA projects.

### Citation Details

*Usability Testing for the Non-Visual Access to the Digital Library (NoVA) Project*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-11/>>

### Related Documents

See also the *How To Evaluate A Web Site's Accessibility Level* (briefing-12) briefing document.

**Keywords:** Web, usability, accessibility, NoVA, case study

## Background

The **Non-Visual Access to the Digital Library (NoVA)** project was concerned with countering exclusion from access to information, which can all too easily occur when individuals do not have so-called 'normal' vision. Usability tests undertaken for the NoVA project provided an insight to the types of problems faced by users and interestingly, although the focus of the project was on the information seeking behaviour of blind and visually impaired people (generally using assistive technologies), the control group of sighted users also highlighted usability problems. This showed that although awareness of Web accessibility is increasing, all types of user could be faced with navigational problems, thus reinforcing the importance of involving a variety of different users in any design and development project.

Some problems experienced were due to accessibility and usability conflicts such as inappropriate or unhelpful use of alternative text, or poor use of language. Other problems were due to a lack of understanding of the different ways users interact and navigate around Web-based resources. Careful consideration must therefore be given not only to assure conflicts between accessibility and usability are addressed, but to the layout and navigation of a site and to the ways different assistive technologies interact with them.

This case study will look specifically at the usability testing phase of the NoVA project. The final report of the NoVA project [1] fully describes the methodology, findings and conclusions, and outlines a number of recommendations for digital library system design.

## Problem Being Addressed

Despite evidence of much good work to make interfaces accessible and on methods for accessibility checking (see for example: EDNER, 2002 [2] and the World Wide Web Consortium Accessibility Guidelines [3]), there is less work published on usability issues or how people using assistive technologies (such as screen readers) navigate around the Web interface.

Although sites may adhere to accessibility recommendations, users can still experience navigational problems. This is partly due to the fact that Web pages are increasingly designed for parallel or non-serial navigation, offering a variety of options within one page (frames, tables, drop down menus etc). Parallel design can cause problems for

users who are navigating the site using assistive technologies which force them down a serial (or linear) route, for example a screen reader reading out every hypertext link on a page one by one.

The overall objective of the NoVA project was to develop understanding of serial searching in non-serial digital library environments, with particular reference to retrieval of information by blind and visually impaired people. Serial searching was defined for the project a linear movement between pages, non-serial (or parallel) searching was defined as movements around a page, between frames or interacting with a number of options such as a table, dialog box or drop-down menu.

### **The Approach Taken**

Using a combination of desk research, task analysis and user interviews, the objectives of the study were to:

Review existing studies into Web-based information seeking behaviour.

Undertake a series of experiments with non-serial searching and retrieval, and subsequent use of digital content.

Map approaches so as to develop understanding of how searching and retrieval can be optimised in non-serial environments.

Report on findings and to make recommendations for digital library system design.

The NoVA usability tests used a combination of observation, transaction logging and verbal protocol, together with pre-and post-task questions.

### **The Sample**

A sample of 20 sighted and 20 blind and visually impaired people was used to undertake a series of usability experiments. Definitions of terms were set at the beginning of the project. The 'sighted' sample was made up of users who were all able to read a standard (14" - 15") screen. The term 'visually impaired' was defined for the NoVA project as people who needed to use assistive technology, or had to be very close to the screen to be able to 'read' it. The sample size for the NoVA project enabled comparative analysis to take place between two user groups, however it should be noted that Nielsen (2000) [4] suggests excellent results can be obtained from usability tests comprising as little as five users (although he recommends at least 15 users to discover all the usability design problems).

### **The Usability Experiments**

Four information-seeking tasks were set using four Web-based resources:

1. Search engine
2. Library OPAC
3. Online shopping site
4. Directory of Internet resources

Although not all of these might be viewed strictly as digital library resources, each resource displayed elements of parallelism in their design and were generally accessible, to greater and lesser degrees, according to the WAI recommendations.



Each of the tasks was consistently set so that comparative analysis could take place between the sighted and visually impaired users. For example, users were asked to search for a national and regional weather forecast using the same search engine.

It was recognised that success in performing searches could be influenced by previous knowledge or experience, either of the technology, the site visited, the subject matter of the task, or by individual interpretation and approach to a task. In an attempt to obtain a balanced picture, the tasks set covered a fairly broad subject base such as weather forecasts, shopping for clothes and travel information.

Every attempt was made to create a relaxed atmosphere and to dispel feelings among the users that they were being tested in any way (although inevitably this still occurred to some extent). This included an initial explanation of the purpose of the study, i.e. to highlight Web usability issues rather than to test information seeking skills. The observer also chatted informally prior to the tasks and offered the users tea/coffee and biscuits to put them at ease. Naturally, the users were ensured that all their responses would be kept strictly anonymous and only used for the stated purpose of the study.

To ensure everyone started from the same place, users were required to commence using the stated electronic resource, but were allowed to choose whether they used the search facility or browsed the site for relevant links. So for example, when asked to look for the national weather forecast for the UK, users were required to start with the search engine, either by typing in search terms or by browsing for a relevant weather link.

Users were not given a time limit to complete each task. At the beginning of the session they were told that they could stop the task at any time and were given examples such as "if you are satisfied that you have found the information", "if you are not satisfied, but think you have found all the information there is", or "if you are fed up with the task". The reason for this was to try and simulate real-life information searching behaviour, where information required by a user may or may not be found from within a specific resource and was not a judgment of the amount of information retrieved.

#### **Data Capture**

Data was gathered using a combination of on-screen data capture (Lotus ScreenCam which records on-screen activity and verbal dialog), sound recording and note taking. This method enabled each task to be recorded (either on-screen or by the sound of the assistive technology with verbal dialog) and backed up by note taking.

Users were asked to verbally describe what they were doing during each task. Users were also asked a set of pre- and post-task questions. These comprised general questions, such as how to tell a page is loading, initial comments about the interfaces and the type of information provided; and usability questions, such as their overall experience navigating around the resource. Both the verbal protocol and the pre- and post task questions provided a richer picture of the user's experience by enabling the observer to ascertain not only what they had done, but why they had done it, and how they felt about it.

Interviews were conducted before and after each task to help ensure the electronic resource and the task performed were still fresh in the user's mind before moving on to the next resource.

### Data Transcription

Data was transcribed in two ways:

1. Each step of the search process was logged and coded according to the action performed by the user. For example, the action of clicking on a link was coded as CO, tabbing down a page was coded as TD and the TI code was assigned to the action of typing in terms.
2. Pre- and Post-task questions were transcribed verbatim.

Data from the searches and questions were entered and coded into a Computer Assisted Qualitative Data Analysis tool (Atlas-ti) [5].

### Data Analysis

Data was analysed using Atlas-ti analysis tool, which provided an efficient method of data storage and retrieval. Although entering and coding data was initially time consuming, once completed it provided quick and easy access to the large amounts of data gathered for the project. It was then possible to generate queries and reports for data analysis and report writing.

Each step taken during a search was coded to show the number and type of keystroke used within each search task. This was used to compare the information seeking behaviour of the two samples (sighted and visually impaired) and to look at different trends within each.

Data from the pre- and post-task questions was grouped and coded into categories. This enabled comparisons to be made relating to specific questions. For example, coding quotes from users relating to the question 'How do you know if a page is loading?' revealed that only one of the sighted users mentioned the listening to the hard drive, whereas many of the visually impaired users said they relied on this clue to tell them that the page is loading.

## Problems and Solutions

Gathering volunteers for the study was a time-consuming process and could have been a problem if it had not been built in to the NoVA project time frame. It is therefore worth bearing in mind that a substantial amount of time and effort is needed to gather a suitable sample.

In order to obtain specific data on the way people search electronic sources, it was necessary to select a sample of people who were reasonably familiar with using the Internet and, where appropriate, were comfortable using assistive technology. This meant that it was not possible to gather a true random sample. Although this was not particularly problematic for the study, it did mean that the results could not be generalised to the population as a whole.

Data was gathered using a combination of on-screen data capture (Lotus ScreenCam [6]), sound recording and note taking. Initially it was hoped that ScreenCam could be used throughout, however the pilot tests revealed that ScreenCam can interfere with assistive technologies, so it was necessary to use a combination of sound recording and note taking for the visually impaired sample.

It was difficult to create a natural environment for the users to perform the tasks, and although every attempt was made to make the users feel comfortable and to dispel any

feelings that their ability was being tested, inevitably at times this did occur. However, this problem was probably unavoidable for the capture of qualitative data.

The observer attempted not to prompt subjects or give any instructions while the subject was performing the task. This proved difficult at times, particularly when it was evident that the user was becoming distressed. In some cases the observer had to provide a "hint" to enable the user to continue (it is suggested that this type of intervention is sometimes necessary in certain circumstances, as is prompting a user to ensure the transcription is accurately logged [7]).

## Conclusions

The usability tests undertaken for the NoVA project provided a rich picture of the types of problems faced by users when navigating around Web-based resources, particularly when using assistive technologies. It also provided evidence of the types of features users liked and disliked, how they overcame navigational problems and what types of features enhanced their searching experience, all of which can be fed back into recommendations for the design of digital library systems.

Although the sample chosen was appropriate for the NoVA study, for general usability studies it would be desirable to try to include users with a variety of disabilities such as motor impairments, hearing impairments and visual impairments. Testing with users with a mix of abilities will help ensure the site is usable as well as interesting and engaging. Usability testing should be used alongside accessibility checking to provide a rich picture of the accessibility and usability of a Web site, which will help designers and developers to ensure their sites embrace universal access and access for all.

The findings of the NoVA usability testing, together with conclusions and recommendations are described in the final report on the NoVA project Web site [1].

## References

1. **Non-visual access to the digital library: the use of digital library interfaces by blind and visually impaired people**, Craven, J. and Brophy, P., Library and Information Commission Research Report 145, Manchester: CERLIM, <<http://www.cerlim.ac.uk/projects/nova.html>>
2. **Web Accessibility Issues for Higher & Further Education. Issues Paper 6**, EDNER Project, <<http://www.cerlim.ac.uk/edner/ip/ip06.rtf>>
3. **Web Content Accessibility Guidelines 1.0**, W3C, <<http://www.w3.org/TR/WCAG10/>>
4. **Why you only need to test with 5 users**, Nielsen, J. (2000), Alertbox, <<http://www.useit.com/alertbox/20000319.html>>
5. **Atlas-ti**, <<http://www.atlasti.de/>>
6. **Lotus Screen Cam**, <<http://www.lotus.com/products/screencam.nsf>>
7. **Usability Engineering**, Nielsen, J. (1993). Academic Press, p.197.

## Contact Details

Jenny Craven  
CERLIM  
Manchester Metropolitan University  
Email: [j.craven@mmu.ac.uk](mailto:j.craven@mmu.ac.uk)

# Approaches to Accessibility at MIMAS

Case study

## About This Document

This case study describes the approaches taken by MIMAS to ensuring their projects and services complied with accessibility guidelines.

## Citation Details

*Approaches to Accessibility at MIMAS*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-15/>>

## Related Documents

See also the *How To Evaluate A Web Site's Accessibility Level* (briefing-12) briefing document.

**Keywords:** Web, accessibility, MIMAS, case study

## Background

MIMAS [1] is a JISC-funded service [2] which provides the UK higher education, further education and research community with networked access to key data and information resources to support teaching, learning and research across a wide range of disciplines. Services supported by MIMAS include the ISI Web of Science Service for UK Education, CrossFire, UK Census aggregate statistics, COPAC, the Archives Hub, JSTOR, a Spatial Data service which includes satellite data, and a new International Data Service (part of the Economic and Social Data Service) [3].

## Problem Being Addressed

This document describes the approaches which have been taken by MIMAS to ensure that its services provide levels of accessibility which are compatible with MIMAS's available resources and the services it provides.

The work was carried out in order to ensure that MIMAS services are compliant with the SENDA legislation wherever possible.

## The Approach Taken

A MIMAS Project Management Team called ACE (Accessibility Compliance Exercise) was set up with the remit of making recommendations on accessibility work. We were set the task of making the MIMAS Web site compliant at least with Priority 1 WAI guidelines [4] by 1 September 2002.

The ACE Team consisted of a coordinator and four members of MIMAS staff with a range of skills, and chosen so that each section manager (of which there are three) had at least one person on the team. We knew that it would take a great deal of effort from many members of staff and that it was crucial to have the support of all managers.

The coordinator reported to the MIMAS and Manchester Computing management team and left technical issues to the rest of the team.

The team went about identifying the services, projects, areas of the MIMAS Web sites and other items which are supported and/or hosted by MIMAS.

Usually the creator or maintainer, but in some cases the section manager was identified as the person responsible for each area and a member of the ACE team (the "ACE contact") was assigned to assist and monitor progress.

We drew a distinction between Web pages (information), data and applications. It took some time to get the message through to all staff that ACE (and SENDA) was concerned with all aspects of the services and that we needed to check applications as well as Web pages. We could all have some informed opinion about Web pages, but applications often required an expert in the area to think about the accessibility of a package. Managers were asked to request and gather statements from suppliers on the accessibility of their products.

A Web page was set up on the staff intranet to document the progress each service was making, who the main players were, and to summarise areas of concern that we needed to follow up. This helped the ACE Team to estimate the scope of any work which was needed, and managers to allocate time for staff to train up in HTML and accessibility awareness. We also provide notes on how to get started, templates, training courses etc., and we continue to add information that we think will help staff to make their pages more accessible.

The ACE team met fortnightly to discuss progress. Members of the team recorded progress on the staff intranet, and the coordinator reported to the management team and to others in the Department (Manchester Computing) engaged in similar work.

### Goals

The ACE team recommended that MIMAS Web resources should aim to comply with at least WAI Priority 1 guidelines. The ACE team had the following aims:

1. To ensure that all Web pages under the auspices of MIMAS comply to appropriate accessibility criteria by 1 September 2002. The minimum criteria for existing pages is WAI Priority 1, but higher levels will be sought where possible. WAI Priority 2 is the minimum level for new resources.
2. To establish local conventions (for example where alternatives are permitted within the WAI guidelines)
3. To provide guidance on accessibility issues, Web page design and HTML to MIMAS staff.
4. To liaise, where appropriate, with other members of the Manchester Computing department engaged in similar activities
5. To raise awareness of the issues and encourage staff to go on training courses such as those provided by Manchester Computing WWW Authoring courses [5] and Netskills courses [6].

### Implementing The Solution

Software (Macromedia's LIFT) was purchased to assist staff evaluate their pages, and extra effort was brought in to assist in reworking some areas accessible.

The ACE team set up an area on the staff intranet. As well as the ongoing progress report, and information about the project and the ACE team this contained hints and tips on how to go about evaluating the accessibility of Web pages, validating the (X)HTML, how to produce an implementation plan, examples of good practice etc.

Other information on the Staff intranet included:

- A case study on using Dreamweaver and LIFT to address accessibility issues for the Zetoc service

- Sample files showing:
- Recommended META tags for MIMAS resources (including Dublin Core)
- Examples of XHTML 1.0 transitional and HTML 4.01 Web pages
- Drop-down menus to access MIMAS services
- How to skip to main content
- Links to useful resources:
- Lists of Priority 1 checkpoints
- Approaches to accessibility for Java
- Tools for link checking
- Examples of cascading style sheets
- Recommendations for providing online access to documents

### Problems Experienced

The ACE team had their own pages to make accessible, whilst also being available to help staff who needed guidance with their own Web sites. We all had our usual day jobs to do, and time was short.

Some Web sites needed a lot of work. We brought in external help to rework two large sites and encouraged the systematic use of Dreamweaver in order to maintain the new standards. Using the Dreamweaver templates prepared by the ACE team helped those staff who were not that familiar with HTML coding.

Although Manchester computing and JISC put on Accessibility courses, not all staff were able to attend. Group meetings were used to get the message across, and personal invitations to attend the ACE workshops were successful in engaging HTML experts, managers, programmers and user support staff.

There was still a lot to do after September 2002. Not all sites could reach Priority 1 by September 2002. For these, and all services, we are recommending an accessibility statement which can be reached from the MIMAS home page. We continue to monitor the accessibility of Web pages and are putting effort into making Web sites conform to the local conventions that we have now set out in the MIMAS Accessibility Policy. This Policy is for staff use, rather than a public statement, and is updated from time to time. Accessibility statements [7] are specific to each service.

### Phase 2

In January and February 2003, the ACE team ran a couple of workshops to demonstrate key elements of the ACE policy - e.g. how to validate your pages, and to encourage the use of style sheets, Dreamweaver, and to discuss ways of making our Web sites more accessible generally.

We still have to ensure that all new staff are sent on the appropriate accessibility courses, and that they are aware of the standards we have set and aim to maintain.

### Things We Would Do Differently

Workshops help everyone to be more aware of the issues and benefited the ACE team as well as staff. Because of time constraints we were unable to prepare our own ACE workshops until January 2003, by which time most sites were up to Level 1. Other

people's workshops (e.g. the JISC workshop) helped those attending to understand the issues relating to their own sites, those maintained by others at MIMAS, and elsewhere.

Talking to staff individually, in small groups, and larger groups, was essential to keep the momentum going.

It would have been helpful to be more specific about the accessibility features we want built in to the Web sites. For example, we encourage "skip to main content" (Priority 3), and the inclusion of Dublin Core metadata.

## References

1. **MIMAS**, <<http://www.mimas.ac.uk/>>
2. **JISC**, <<http://www.jisc.ac.uk/>>
3. **ESDS**, <<http://www.esds.ac.uk/>>
4. **WAI Guidelines**, W3C, <<http://www.w3.org/WAI/>>
5. **Manchester Computing**, <<http://www.mc.man.ac.uk/>>
6. **Netskills**, <<http://www.netskills.ac.uk/>>
7. **MIMAS Accessibility Statements**, MIMAS, <<http://www.mimas.ac.uk/accessibility.html>>

## Contact Details

Anne McCombe  
MIMAS  
University of Manchester

# Standards and Accessibility Compliance for the FAILTE Project Web Site

Case study

### About This Document

This document describes a selection matrix to help in choosing standards for use by projects.

### Citation Details

*Standards and Accessibility Compliance for the FAILTE Project Web Site*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-31/>>

### Related Documents

See also the *Compliance With HTML Standards* (briefing-01) briefing document.

**Keywords:** Web, standards, accessibility, compliance, FAILTE, case study

## Background

The **FAILTE** project [1] was funded by JISC to provide a service which engineering lecturers in higher education could use to identify and locate electronic learning resources for use in their teaching. The project started in August 2000, and one of the first tasks was to set up a project Web site describing the aims, progress and findings of the project and the people involved.

## Problem Being Addressed

As an experienced Web author I decided to use this opportunity to experiment with two specifications which at that time were relatively new, namely cascading style sheets (CSS) and HTML. At the same time I also wanted to create pages which looked reasonably attractive on the Web browsers in common use (including Netscape 4.7 which has poor support for CSS) and which would at least display intelligible text no matter what browser was used.



Figure 1: The FAILTE home page

## Why Use XHTML and CSS?

Here is not the place for a detailed discussion of the merits of separating logical content markup from formatting, but I will say that I think that, since this is how HTML was envisaged by its creators, it works best when used in this way. Some of the immediate reasons at the time of starting the Web site were:

I had recently been through the tedious experience of adjusting the appearance of a Web site which had the formatting information held in each page, and having one file which held the formatting information for the entire site seemed an attractive proposition.



I wanted to increase the accessibility of the Web site by not using tables to control the layout of the pages

At that time several of my colleagues were telling me that CSS "just didn't work". I wanted to produce an example which showed to what extent it was possible to rely CSS alone for styling.

It has been my experience that avoiding browser-specific "enhancements" to HTML reduces the long-term maintenance requirements for Web sites.

### The Approach Taken

A quick investigation of the Web server log files from a related server which dealt with the same user community as our project targeted lead us to the decision that we should worry about how the Web site looked on Netscape 4.7, but not browsers with poorer support of XHTML and CSS (e.g. Netscape 4.5 and Internet Explorer 3).

The Web site was a small one, and there would be one contributor: me. This meant that I did not have to worry about the lack of authoring tools for XHTML at the time of setting up the Web site. I used HomeSite version 4.5, a text editor for raw HTML code, mainly because I was familiar with it. Divisions (`<div>` tags) were used in place of tables to create areas on the page (a banner at the top, a side bar for providing a summary of the page content), graphics were used sparingly, and colour was used to create a consistent and recognisable look across the site. It is also worth noting that I approached the design of the Web site with the attitude that I could not assume that it would be possible to control layout down to the nearest point.

While writing the pages I tested primarily against Netscape 4.7, since this had the poorest support of XHTML and CSS of the browsers which we wanted to make a good job of rendering the Web design. I also made heavy use of the W3C XHTML and CSS validation service [2], and against Bobby [3] to check for accessibility issues. Once the code validated and achieved the desired effect in Netscape 4.7 I checked the pages against a variety of browser platforms.

While it was never my aim to comply with a particular level of accessibility, the feedback from Bobby allowed me to enhance accessibility while building the pages.

### Problems Experienced

Most of the problems stemmed from the need to support Netscape 4.7, which only partially implements the CSS specification. This cost time while I tried approaches which didn't work and then looked for work-around solutions to achieve the desired effect. For example, Netscape 4.7 would render pages with text from adjacent columns overlapping unless the divisions which defined the columns had borders. Thus the `<div>` tags have styles which specify borders with `border-style: none;` which creates a border but doesn't display it.

The key problem here is the partial support which this version of Netscape has for CSS: older versions have no support, and so the style sheet has no effect on the layout, and it is relatively easy to ensure that the HTML without the style sheet makes sense.

Another problem was limiting the amount of white space around headings. On one page in particular there were lots of headings and only short paragraphs of text. Using the HTML `<h1>`, `<h2>`, `<h3>`, etc. tags left a lot of gaps and led to a page which was difficult to interpret. What I wanted to do was to have a vertical space above the headings but not below. I found no satisfactory way of achieving this using the standard

heading tags which worked in Netscape 4.7 and didn't cause problems in other browsers. In the end, I created class styles which could be applied to a `<span>` to give the effect I wanted e.g.:

```
<p><span class="h2">Subheading</span><br />
Short paragraph</p>
```

This was not entirely satisfactory since any indication that the text was a heading is lost if the browser does not support CSS.

### Pleasant Surprises

The Web site is now two years old and in that time I have started using two new browsers. I now use Mozilla as my main browser and was pleasantly surprised that the site looks better on that than on the browsers which I used while designing it. The second browser is an off-line Web page viewer which can be used to view pages on a PDA, and which makes a reasonable job rendering the FAILTE Web site - a direct result of the accessibility of the pages, notably the decision not to use a table to control the layout of the page. This is the first time that the exhortation to write Web sites which are device-independent has been anything other than a theoretical possibility for me (remember WebTV?)

### Things I Would Do Differently

I think that it is now much easier to use XHTML and CSS since the support offered by authoring tools is now better. I would also reconsider whether Netscape 4.7 was still a major browser: my feeling is that while it still needs supporting in the sense that pages should be readable using it, I do not think that it is necessary to go to the effort of making pages look attractive. In particular I would not create styles which imitated `<Hn>` in order to modify the appearance of headings. I look forward to the time when it is possible to write a page using standard HTML repertoire of tags without any styling so that it makes sense as formatted text, with clear headings, bullet lists etc., and then to use a style sheet to achieve the graphical effect which was desired.

### References

- 1 **FAILTE**, <http://failte.ac.uk/>
- 2 **HTML Validation Service**, W3C, <http://validator.w3.org/>
- 3 **Bobby**, WatchFire, <http://bobby.watchfire.com/bobby/>

### Contact Details

Phil Barker  
ICBL  
MACS  
Heriot-Watt University  
Edinburgh  
Email: [philb@icbl.hw.ac.uk/](mailto:philb@icbl.hw.ac.uk/)

## Standards and Accessibility Compliance for the DEMOS Project Web Site

Case study

### About This Document

This document describes a selection matrix to help in choosing standards for use by projects.

### Citation Details

*Standards and Accessibility Compliance for the DEMOS Project Web Site* QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-10/>>

### Related Documents

See also the *Compliance With HTML Standards* (briefing-01) briefing document.

**Keywords:** Web, standards, accessibility, compliance, DEMOS, case study

## Background

Funded by the Higher Education Funding Council for England under strand three of the initiative 'Improving Provision for Students with Disabilities', the aim of the **DEMOS Project** [1] was to develop an online learning package aimed at academic staff and to examine the issues faced by disabled students in higher education. The project was a collaboration between the four universities in the Manchester area - the University of Salford [2], the University of Manchester [3], the Manchester Metropolitan University [4] and UMIST [5].

## The Web Site

At the start of the project the purpose of the Web site was still unclear, which made it difficult to plan the information structure of the site. Of course, it would serve as a medium to disseminate the project findings, research reports, case studies... but for months the design and the information architecture of this site seemed to be in a never-ending state of change.

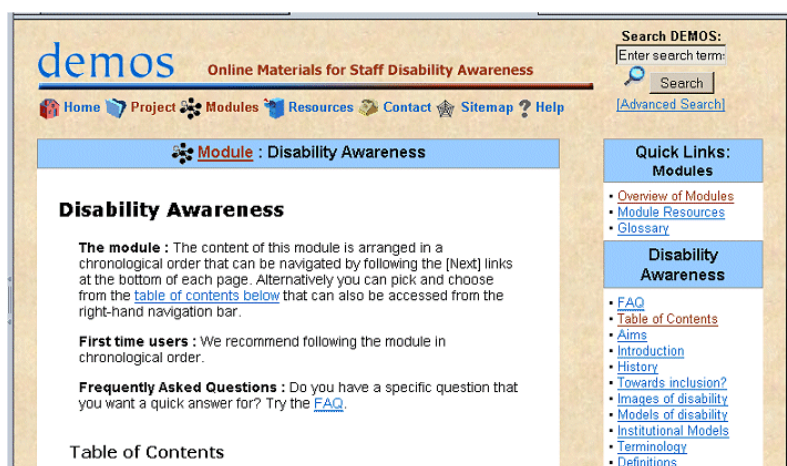


Figure 1: The Demos Web Site

In the early stage of the project virtual learning environments, such as WebCT, were tested and deemed unsuitable for delivering the course material, due to the fact that they did not satisfy the requirements for accessibility.

At this point it was decided that the Web site should carry the course modules. This changed the focus of the site from delivering information about the progress of the project to delivering the online course material.

In the end we envisioned a publicly accessible resource that people can use in their own time and at their own pace. They can work through the modules in the linear fashion

they were written in or they can skip through via the table of contents available on every page. There are also FAQ pages, which point to specific chapters.

Many academic institutions have already added links to the DEMOS materials on their own disability or staff development pages.

### **Problem Being Addressed**

To ignore accessibility would have been a strange choice for a site that wants to teach people about disability. Accessibility was therefore the main criteria in the selection of a Web developer.

I have been a Web designer since 1998 and specialised in accessibility from the very beginning. For me it is a matter of ethics. Now it is also the law.

The challenge here was to recreate, at least partially, the feeling of a learning environment with its linear structure and incorporating interactivity in form of quizzes and other learning activities without the use of inaccessible techniques for the creation of dynamic content.

### **The Approach Taken**

Accessibility techniques were applied from the beginning. But the site also represents an evolution in my own development as a Web designer, it always reflected my own state of knowledge. Where in the beginning accessibility meant eradicating the font tag, it now means standard-compliant code and tableless CSS layout.

#### **Valid and Accessible Design and Code**

This site was designed in compliance with the latest standards developed by the World Wide Web Consortium (W3C) [6] and using the latest accessibility techniques [7] as recommended by the Web Accessibility Initiative (WAI) [8] .

In December 2001 the code base of the site was switched over to XHTML 1.0 Transitional. In November 2002 the site was further improved by changing it to a CSS layout, which is used to position elements on the page without the use of tables. The only layout table left is the one holding the header: logo, search box and top navigation icons.

Stylesheets are also used for all presentational markup and a separate print stylesheet has been supplied.

The code was validated using the W3C Validation Service [9].

#### **Standards = Accessibility**

With the advent of standard-compliant (version 6 and 7) browsers, the Web developer community started pushing for the adoption of the W3C guidelines as standard practise by all Web designer. Now that version 3 and 4 browsers with all their proprietary markup were about to be consigned to the scrap heap of tech history, it was finally possible to apply all the techniques the W3C was recommending. Finally the makers of user agents had started listening to the W3C too and were making browsers that rendered pages designed according to standards correctly. (It turns out things weren't all that rosy but that's the topic for another essay.)

Standards are about accessibility, or, as the W3C phrases it, 'universal design'. They ensure that universal access is possible, i.e. that the information contained on a Web page can be accessed using:

- a modern standard-compliant browser on an up-to-date high spec machine
- an old browser on a slow connection and legacy hardware
- a text-only browser (often used as a basis for screen readers)
- assistive technology (screen readers, foot pedals, braille printers, etc.)
- a small device (PDA, mobile phone)

### User Control

The most important reason for designing according to standards is that it gives the user control over how a Web page is presented. The user should be able to increase font sizes, apply preferred colours, change the layout, view headers in a logical structure, etc.

This control can be provided by the Web designer by:

- using correct structural markup
- using stylesheets to specify presentational styles (such as fonts and colours)
- using CSS layout (instead of table layout)
- applying accessibility techniques

On the DEMOS site, all presentational styles are specified in stylesheets. The site 'transforms gracefully' when stylesheets are ignored by the user agent, which means that the contents of a page linearises logically. The user has control over background and link colours via the browser preferences and can increase or decrease font sizes.

The DEMOS Guide to Accessible Web Design contains a chapter on User Control [10], which describes how these changes can be applied in various browsers.

### Accessibility Techniques

(The links below lead to pages on the DEMOS site, more precisely: the DEMOS Guide to Accessible Web Design [11])

Some of the techniques used:

- A mechanism to skip to the main content was provided.
- Access keys were defined.
- Alternative descriptions were provided for all images. Purely decorative images contain an empty ALT attribute.
- Links are separated with printable characters.
- Stylesheets are used to allow user control.
- All acronyms and abbreviations are labelled.
- Icons and colours are used in a consistent manner to improve accessibility for users with cognitive or learning disabilities.
- The site can be navigated via the keyboard.
- Tables are only used for data, not for layout (except for the header table).
- Frames and dynamic content are avoided.

- Accessible interactivity was provided using PHP.
- Text is kept scannable, language clear and jargon is explained in glossaries.
- Language changes are declared for the benefit of screen readers.

More information and details: Accessibility techniques used on the DEMOS site [12] (listed by WAI checkpoints).

### Inclusive Design

Web developers sometimes believe that accessibility means providing a separate text-only or low-graphics version for the blind. First of all: I have always been on that side of the camp that believes that there should be only one version of a site and that it should be accessible.

*"Don't design an alternative text-only version of the site: disabled people are not second class citizens..."* (Antonio Volpon, [evolt.org](http://evolt.org) [13]).

Secondly, accessibility benefits not only blind people [14]. To be truly inclusive we have to consider people with a variety of disabilities, people with a range of visual, auditory, physical or cognitive disabilities, people with learning disabilities, not to forget colour blindness, senior citizens, speakers of foreign languages, etc.

Surely not all of them are part of the target audience, but you never know, and applying all available accessibility techniques consistently does not take that much more effort.

We tried to provide a satisfactory experience for everyone, providing user control, keyboard access, icons and colour to loosen things up, whitespace and headers to break up text in digestible chunks. And we encourage people to provide feedback, especially if they experience any problems.

### Accessibility Testing

To ensure accessibility the site was tested across a variety of user agents and on different platforms. A number of screenshots from these tests [15] can be found at the DEMOS site. The site has also been tested using the Bobby [16] and Wave [17] accessibility checkers. It is AAA compliant, which means that it meets all three levels of accessibility.

### Accessible Interactivity

One of the last things we finally solved to our satisfaction was the problem of creating interactive quizzes and learning activities for the course modules without the use of inaccessible techniques. Many of the techniques for the creation of dynamic and multimedia content (JavaScript, Java, Flash...) are not accessible.

Eventually we discovered that PHP, a scripting language, was perfect for the job. PHP is processed server-side and delivers simple HTML pages to the browser without the need for plug-ins or JavaScript being enabled.

## Problems Encountered

### Information Architecture and Navigation

As mentioned before, the Web site started without a clear focus and without a clear structure. Therefore there wasn't much planning and structured development. In the first months content was added as it was created (in the beginning mainly information about

the project) and the site structure grew organically. This caused some problems later when main sections had to be renamed and content restructured. From the Web development point of view this site has been a lesson in building expandability into early versions of Web site architecture.

Since there was so much uncertainty about the information architecture in the beginning, the navigation system is not the best it could be. The site grew organically and navigations items were added as needed. The right-hand navigation was added much later when the site had grown and required more detailed navigation - more detailed than the main section navigation at the top of the page underneath the logo and strapline.

But the right-hand navigation is mainly sub-navigation, section navigation, which might be confusing at times. At the same time, however, it always presents a handy table of contents to the section the visitor is in. This was especially useful in the course modules.

The breadcrumb navigation at the top of the main content was also added at a later date to make it easier for the visitor to know where they are in the subsections of the site.

#### **Netscape 4**

Already mentioned in Phil Barker's report on the FAILTE Web site [18], Netscape 4 was also my biggest problem.

Netscape 4 users still represent a consistent 12% of visitors in the UK academic environment (or at least of the two academic sites I am responsible for). Since this is the target audience for the DEMOS site, Netscape 4 quirks (i.e. its lack of support for standards) had to be taken into account.

Netscape understands just enough CSS to make a real mess of it. Older browsers (e.g. version 3 browsers) simply ignore stylesheets and display pages in a simpler fashion with presentational extras stripped, while standard-compliant browsers (version 6 and 7) display pages coded according to standards correctly. Netscape 4 is stuck right between those two scenarios, which is the reason why the DEMOS site used tables for layout for a long time.

Tables are not really a huge accessibility problem if used sparingly and wisely. Jeffrey Zeldman wrote in August 2002 in 'Table Layout, Revisited' [19]:

Table layouts are harder to maintain and somewhat less forward compatible than CSS layouts. But the combination of simple tables, sophisticated CSS for modern browsers, and basic CSS for old ones has enabled us to produce marketable work that validates - work that is accessible in every sense of the word.

Tables might be accessible these days because screenreader software has become more intelligent but standard-compliance was my aim and layout tables are not part of that.

Luckily techniques have emerged that allow us to deal with the Netscape 4 quirks.

One option is to prevent Netscape 4 from detecting the stylesheet, which means it would deliver the contents in the same way as a text-only browser, linearised: header, navigation, content, footer following each other. No columns, colours, font specifications. But an audience of 12% is too large to show a site to that has lost its 'looks'. The site still had to look good in Netscape 4.

The other option is to use a trick to get Netscape 4 to ignore some of the CSS instructions [20] . Deliver a basic stylesheet to Netscape 4 and an additional stylesheet with extra instructions to modern browsers. This required a lot of tweaking and actually consumed an enormous amount of time but only because I was new to CSS layout. I have converted a number of other sites to CSS layout in the meantime, getting better at it every time.

The DEMOS site now looks good in modern browsers, looks OK but not terrific in Netscape 4, and simply linearises logically in browsers older than that and in text-only browsers.

### To Do

There are still a few issues that need looking at, e.g. the accessibility of input forms needs improving (something I'm currently working on) and the structural mark-up needs improving so that headers are used in logical order starting with <h1>

There are also a few clashes of forms with the CSS layout. All forms used on the DEMOS site are still in the old table layout. I haven't had the time to figure out what the problem is.

Eventually I also plan to move the code to XHTML Strict and get rid of the remains of deprecated markup [21], which XHTML Transitional, the DOCTYPE [22] used at the moment, still forgives.

### The Site's Future

Of course it is important to keep the materials produced over the last two and a half years accessible to the public after the funding has run out. This will happen at the end of March 2003. This site will then become part of the Access Summit Web site (at time of writing still under construction). Access Summit is the Joint Universities Disability Resource Centre that was set up in 1997 to develop provision for and support students with disabilities in higher education in Manchester and Salford.

We currently don't know whether we will be able to keep the domain name, so keep in mind that the URL of the DEMOS site might change. I will do my best to make it possible to find the site easily.

### Resources

**DEMOS Web site**, <<http://www.demos.ac.uk/>>

**Jarmin.com Guide to Accessible Web Design**, <<http://jarmin.com/accessibility/>>

Tips, guidelines and resources from the author of this case study. Focuses on techniques but includes chapters on barriers to access, evaluation, legislation, usability, writing for the Web and more. Includes a huge resources section <<http://jarmin.com/accessibility/resources/>>. This section also contains resources that helped me understand the power of stylesheets <[http://jarmin.com/accessibility/resources/css\\_layout.html](http://jarmin.com/accessibility/resources/css_layout.html)>.

**DEMOS Accessibility Guide**, <<http://jarmin.com/demos/access/>>

Consists of the Techniques section from the Guide to Accessible Web Design <<http://jarmin.com/accessibility/>> plus information on accessibility techniques used on the DEMOS site <<http://jarmin.com/demos/access/demos.html>> and a number of demonstrations <<http://jarmin.com/demos/access/demos06.html>> on how the site looks under a variety of circumstances.



### References

1. **The DEMOS Project**, <<http://jarmin.com/demos/>>
2. **University of Salford**, <<http://www.salford.ac.uk/>>
3. **University of Manchester**, <<http://www.man.ac.uk/>>
4. **Manchester Metropolitan University**, <<http://www.mmu.ac.uk/>>
5. **University of Manchester Institute of Science and Technology (UMIST)**, <<http://www.umist.ac.uk/>>
6. **World Wide Web Consortium (W3C)**, <<http://www.w3.org/>>
7. **Guide to Accessible Web Design**, Iris Manhold, <<http://jarmin.com/accessibility/>>
8. **Web Accessibility Initiative (WAI)**, W3C, <<http://www.w3.org/wai/>>
9. **Validation Service**, W3C, <<http://validator.w3.org/>>
10. **DEMOS Accessibility Guide: User Control**, Iris Manhold, <<http://jarmin.com/demos/access/control.html>>
11. **DEMOS Accessibility Guide**, Iris Manhold, <<http://jarmin.com/demos/access/>>
12. **Accessibility techniques used on the DEMOS site** (listed by WAI checkpoints), <<http://jarmin.com/demos/access/demos.html>>
13. **The Lifecycle of Web Accessibility**, Antonio Volpon, evolt.org, <[http://evolt.org/article/The\\_Lifecycle\\_of\\_Web\\_Accessibility/20/50376/](http://evolt.org/article/The_Lifecycle_of_Web_Accessibility/20/50376/)>
14. **Benefits of accessibility**, Iris Manhold, <<http://jarmin.com/accessibility/access/index.html>>
15. **The DEMOS site tested**, <<http://jarmin.com/demos/access/demos06.html>>
16. **Bobby Accessibility Checker**, <<http://bobby.watchfire.com/bobby/html/en/index.jsp>>
17. **Wave Accessibility Checker**, <<http://www.temple.edu/instituteondisabilities/piat/wave/>>
18. **Standards and Accessibility Compliance for the FAILTE Project Web Site**, QA Focus case study, Phil Barker, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-02/>>
19. **Table Layout, Revisited**, Jeffrey Zeldman, A List Apart, 17 August 2001, <<http://www.zeldman.com/daily/0802d.html#livedandlovedlikefrank>>
20. **Tricking browsers and hiding styles**, Eric A. Meyer, <<http://www.ericmeyeroncss.com/bonus/trick-hide.html>>
21. **Deprecated markup**, <<http://jarmin.com/demos/access/deprecat.html>>
22. **Doctype**, <<http://jarmin.com/demos/access/doctype.html>>

### Contact Details

Please contact me if you have feedback, suggestions or questions about the DEMOS site, my design choices and accessible Web design.

# Strategy For Fixing Non-Compliant HTML Pages On The QA Focus Web Site

Case study

### About This Document

This case study describes how use of the W3C's Web log validator tools helps to identify key non-compliant HTML pages which need to be fixed.

### Citation Details

*Strategy For Fixing Non-Compliant HTML Pages On The QA Focus Web Site*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-23/>>

### Related Document

See also the *Compliance With HTML Standards* (briefing-01) briefing document.

**Keywords:** Web, standards, HTML, compliance, QA Focus, case study

## Problem Being Addressed

It is important that HTML resources comply with the HTML standard. Unfortunately in many instances this is not the case, due to limitations of HTML authoring and conversion tools, a lack of awareness of the importance of HTML compliance and the attempts made by Web browsers to render non-compliant resources. This often results in large numbers of HTML pages on Web sites not complying with HTML standards. An awareness of the situation may be obtained only when HTML validation tools are run across the Web site.

If large numbers of HTML pages are found to be non-compliant, it can be difficult to know what to do to address this problem, given the potentially significant resources implications this may involve.

One possible solution could be to run a tool such as Tidy [1] which will seek to automatically repair non-compliant pages. However, in certain circumstances an automated repair could result in significant changes to the look-and-feel of the resource. Also use of Tidy may not be appropriate if server-side technologies are used, as opposed to simple serving of HTML files.

This case study describes an alternative approach, based on use of W3C's Web Log Validator Tool.

## Solution Used

### W3C's Log Validator Tool

W3C's Log Validator Tool [2] processes a Web site's server log file. The entries are validated and the most popular pages which do not comply with the HTML standard are listed.

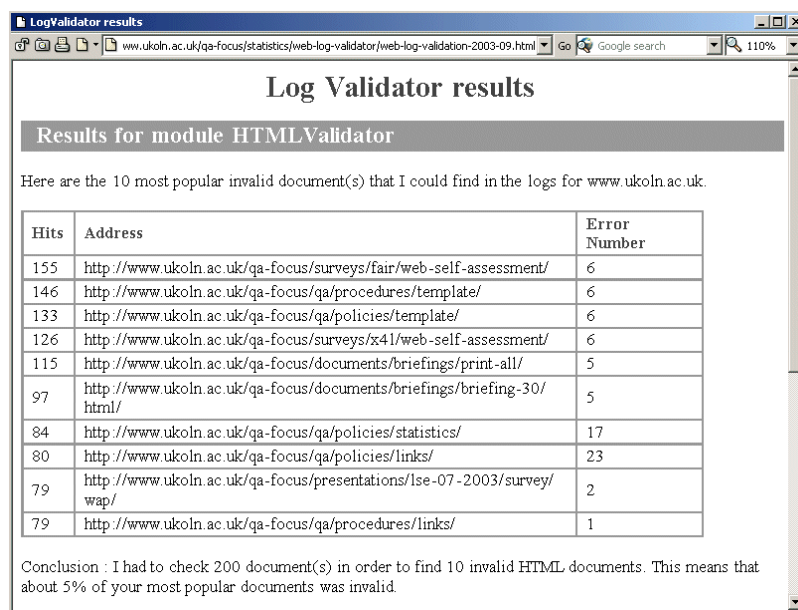
### Use On The QA Focus Web Site

The Web Log Validator Tool has been installed on the UKOLN Web site. The tool has been configured to process resources on the QA Focus area (i.e. resources within the <http://www.ukoln.ac.uk/qa-focus/> area).

The tool has been configured to run automatically once a month and the findings held on the QA Focus Web site [3]. An example of the output is shown in Figure 1.

When the tool is run an email is sent to the Web site editor and the findings are examined. We have a policy that we will seek to fix HTML errors which are reported by this tool.

This approach is a pragmatic one. It helps us to prioritise the resources to fix by listed the most popular pages which are non-compliant. Since only 10 non-compliant pages are listed it should be a relatively simple process to fix these resources. In addition if the errors reflect errors in the underlying template, we will be in a position to make changes to the template, in order to ensure that new pages are not created containing the same problems.



The screenshot shows a web browser window titled "LogValidator results". The address bar shows the URL "www.ukoln.ac.uk/qa-focus/statistics/web-log-validator/web-log-validation-2003-09.html". The page content is titled "Log Validator results" and "Results for module HTMLValidator". It states: "Here are the 10 most popular invalid document(s) that I could find in the logs for www.ukoln.ac.uk." Below this is a table with three columns: Hits, Address, and Error Number. The table lists 10 entries. At the bottom, a conclusion states: "Conclusion : I had to check 200 document(s) in order to find 10 invalid HTML documents. This means that about 5% of your most popular documents was invalid."

Hits	Address	Error Number
155	<a href="http://www.ukoln.ac.uk/qa-focus/surveys/fair/web-self-assessment/">http://www.ukoln.ac.uk/qa-focus/surveys/fair/web-self-assessment/</a>	6
146	<a href="http://www.ukoln.ac.uk/qa-focus/qa/procedures/template/">http://www.ukoln.ac.uk/qa-focus/qa/procedures/template/</a>	6
133	<a href="http://www.ukoln.ac.uk/qa-focus/qa/policies/template/">http://www.ukoln.ac.uk/qa-focus/qa/policies/template/</a>	6
126	<a href="http://www.ukoln.ac.uk/qa-focus/surveys/x41/web-self-assessment/">http://www.ukoln.ac.uk/qa-focus/surveys/x41/web-self-assessment/</a>	6
115	<a href="http://www.ukoln.ac.uk/qa-focus/documents/briefings/print-all/">http://www.ukoln.ac.uk/qa-focus/documents/briefings/print-all/</a>	5
97	<a href="http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-30/html/">http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-30/html/</a>	5
84	<a href="http://www.ukoln.ac.uk/qa-focus/qa/policies/statistics/">http://www.ukoln.ac.uk/qa-focus/qa/policies/statistics/</a>	17
80	<a href="http://www.ukoln.ac.uk/qa-focus/qa/policies/links/">http://www.ukoln.ac.uk/qa-focus/qa/policies/links/</a>	23
79	<a href="http://www.ukoln.ac.uk/qa-focus/presentations/lse-07-2003/survey/wap/">http://www.ukoln.ac.uk/qa-focus/presentations/lse-07-2003/survey/wap/</a>	2
79	<a href="http://www.ukoln.ac.uk/qa-focus/qa/procedures/links/">http://www.ukoln.ac.uk/qa-focus/qa/procedures/links/</a>	1

Conclusion : I had to check 200 document(s) in order to find 10 invalid HTML documents. This means that about 5% of your most popular documents was invalid.

**Figure 1: Output From The Web Log Validator Tool**

## Lessons Learnt

We have internal procedures for checking that HTML pages are compliant. However as these procedures are either dependent on manual use (checking pages after creation or updating) or run periodically (periodic checks across the Web site) it is useful to make use of this automated approach as an additional tool.

Ideally this tool would be deployed from the launch of the Web site, in order to ensure best practices were implemented from the start.

## References

- 1 **HTML Tidy Project Page**, <<http://tidy.sourceforge.net/>>
- 2 **Log Validator Tool**, W3C, <<http://www.w3.org/QA/Tools/LogValidator/>>
- 3 **QA Focus Statistics: Non-Compliant HTML Pages**, QA Focus, <<http://www.ukoln.ac.uk/qa-focus/statistics/web-log-validator/>>

## Contact Details

Brian Kelly  
UKOLN, University of Bath, BATH, UK, BA2 7AY  
Email: [B.Kelly@ukoln.ac.uk](mailto:B.Kelly@ukoln.ac.uk)

# Exploiting ACRONYM And ABBR HTML Elements

Case Study

### About This Document

This case study describes the development of a harvester which can be used to create an automated glossary.

### Citation Details

*Exploiting ACRONYM And ABBR HTML Elements*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-29/>>

**Keywords:** QA, Web. Web site, acronym. abbreviation, HTML, case study

## Background

The UK Centre for Materials Education [1] supports academic practice and innovative learning and teaching approaches in Materials Science and Engineering, with the aim of enhancing the learning experience of students. The Centre is based at the University of Liverpool, and is one of 24 Subject Centres of the national Learning and Teaching Support Network [2].

## Problem Being Addressed

Within any field, the use of discipline-specific language is widespread, and UK Higher Education is no exception. In particular, abbreviations are often used to name projects, programmes or funding streams. Whilst use of these initialisms can be an essential tool of discussion amongst peers, they can also reduce accessibility and act as a barrier to participation by others.

In this context, many individuals and organisations maintain glossaries of abbreviations. However, glossaries of this nature usually require manual editing which can be incredibly resource intensive.

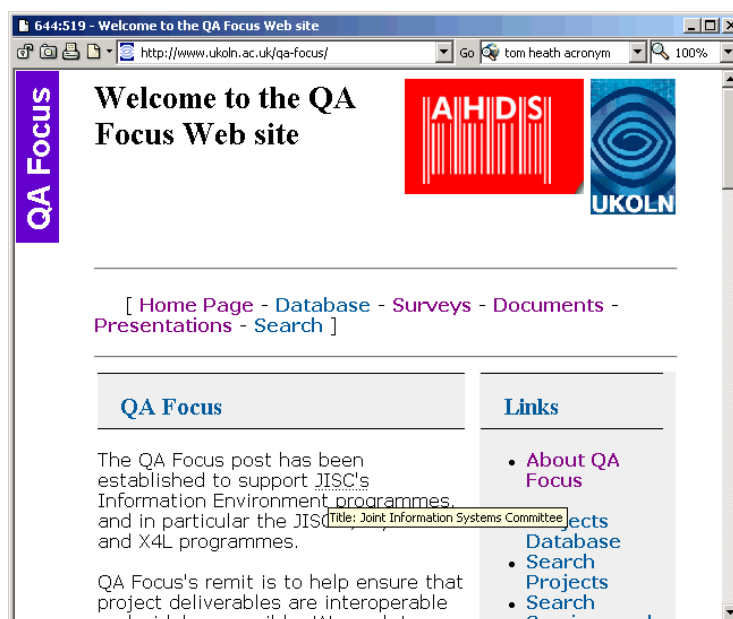
This case study describes a tool developed at the UK Centre for Materials Education to help demystify abbreviations used in the worlds of Higher Education, Materials Science, and Computing, through the use of an automated 'Web crawler'.

## The Approach Taken

The HTML 4 specification [3] provides two elements that Web authors can use to define abbreviations mentioned on their Web sites; `<abbr>` to markup abbreviations and `<acronym>` to markup pronounceable abbreviations, known as acronyms.

The acronyms and abbreviations are normally identified by underlining of the text. Moving the mouse over the underlined words in a modern browser which provides the necessary support (e.g. Opera and Mozilla) results in the expansion of the acronyms and abbreviations being displayed in a pop-up window, as illustrated in Figure 1.

Using this semantic markup as a rudimentary data source, the crawler retrieves Web pages and evaluates their HTML source code for instances of these tags. When either of the tags are found on a page, the initials and the definition provided are recorded in a database, along with the date/time and the URL of the page where they were seen.



**Figure 1: Rendering Of The <ACRONYM> Element**

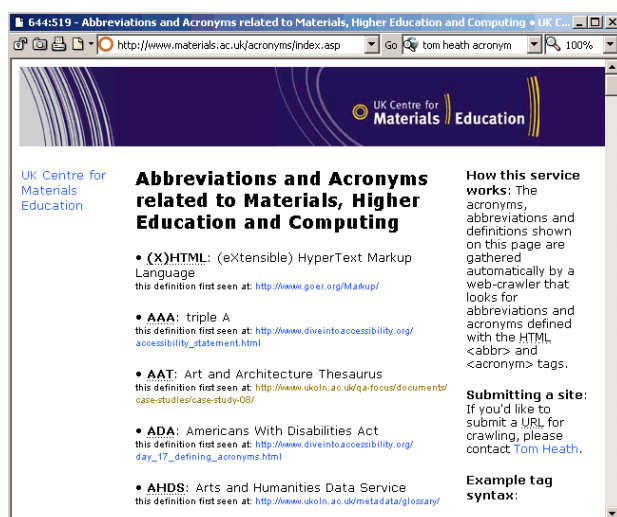
The pairs of abbreviations and definitions identified by the crawler are then made freely available online at [4] to allow others to benefit from the work of the crawler.

### Problems Experienced

The limiting factor first encountered in developing the crawler has been the lack of Web sites making use of the <abbr> and <acronym> tags. Consequently, the number of entries defined in the index is relatively small, and the subject coverage limited. Sites implementing the tags are predominantly those that address Web standards and accessibility, leading to a strong bias in the index towards abbreviations used in these areas.

A number of factors likely contribute to a lack of use of the tags. Firstly, many Web authors might not be aware of the existence of the tags. Even in the current generation of Web browsers, there is little or no support for rendering text differently where it has been marked up as an abbreviation or acronym within a Web page. Therefore there is little opportunity to discover the tags and their usage by chance.

The second major factor affecting the quality of the index produced by the crawler has been the inconsistent and occasionally incorrect definition of terms in pages that do use the tags. Some confusion also exists about the semantically correct way of using the tags, especially the distinction between abbreviations and acronyms, and whether incorrect semantics should be used in order to make use of the browser support that does exist.



**Figure 2: The Glossary Produced By Harvesting <ABBR> and <ACRONYM> Elements**

### Things We Would Do Differently/Future Developments

To provide a truly useful resource, the crawler needs to be developed to provide a larger index, with some degree of subject classification. How this classification might be automated raises interesting additional questions.

Crucially, the index size can only be increased by wider use of the tags. Across the HE sector as a whole, one approach might be to encourage all projects or agencies to 'take ownership' of their abbreviations or acronyms by defining them on their own sites. At present this is rarely the case.

In order to provide a useful service the crawler is reliant on more widespread deployment of <acronym> and <abbr> elements and that these elements are used correctly and consistently. It is pleasing that QA Focus is encouraging greater usage of these elements and is also addressing the quality issues [4].

Lastly, if sites were to produce their pages in XHTML [5] automated harvesting of information in this way should be substantially easier. XML parsing tools could be used to process the information, rather than relying on processing of text strings using regular expressions, as is currently the case.

### References

1. **UK Centre for Materials Education**, <<http://www.materials.ac.uk/>>
2. **Learning and Teaching Support Network**, <<http://www.ltsn.ac.uk/>>
3. **HTML 4.01 Specification**, W3C, <<http://www.w3.org/TR/html4/>>
4. **Using The ACRONYM And ABBR HTML Elements On The QA Focus Web Site**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-30/>>
5. **XHTML 1.0 The Extensible HyperText Markup Language (Second Edition)**, W3C, <<http://www.w3.org/TR/xhtml1/>>

### Contact Details

Tom Heath, Web Developer, UK Centre for Materials Education, Materials Science and Engineering, Ashton Building, University of Liverpool, Liverpool, L69 3GH  
Email [t.heath@liv.ac.uk](mailto:t.heath@liv.ac.uk)

## Using The ACRONYM And ABBR HTML Elements On The QA Focus Web Site

Case Study

### About This Document

This case study describes use of the ACRONYM and ABBR HTML elements.

### Citation Details

*Using The ACRONYM And ABBR HTML Elements On The QA Focus Web Site*, QA Focus, UKOLN, <http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-30/>

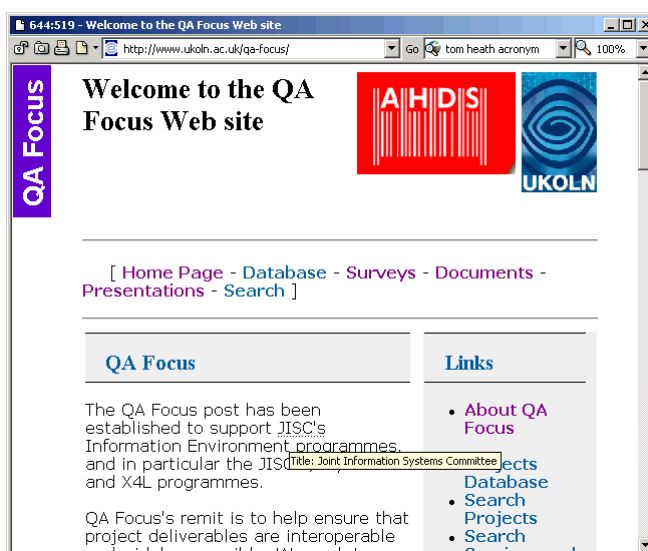
**Keywords:** QA, Web. Web site, preservation, long term access, Exploit Interactive, case study

## Background

After hearing about the automated tool which harvested <abbr> and <acronym> elements [1] it was decided to begin the deployment of these elements on the QA Focus Web site. This case study reviews the issues which needed to be addressed.

## Problem Being Addressed

The <abbr> and <acronym> elements were developed primarily to enhance the accessibility of Web pages, by allowing the definitions of abbreviations and acronyms to be displayed. The acronyms and abbreviations are normally identified by underlining of the text. Moving the mouse over the underlined words in a modern browser which provides the necessary support (e.g. Opera and Mozilla) results in the expansion of the acronym or abbreviation being displayed in a pop-up window. An example is illustrated in Figure 1.



**Figure 1: Rendering Of The <ACRONYM> Element**

As Tom Heath's case study describes, these elements can be repurposed in order to produce an automated glossary.

Since the QA Focus Web site contains many abbreviations and acronyms (e.g. Web terms such as HTML and SMIL, programme, project and service terms such as JISC, FAIR and X4L and expressions from the educational sector such as HE and HEFCE) it was recognised that there is a need for such terms to be explained. This is normally

done within the text itself e.g. "The JISC (Joint Information Systems Committee) ...". However the QA Focus team quickly recognised the potential of the <abbr> and <acronym> harvesting tool to produce an automated glossary of tools.

This case study describes the issues which QA Focus needs to address in order to exploit the harvesting tool effectively.

### The Approach Taken

The QA Focus Web site makes use of PHP which assemble XHTML fragments. The HTML-Kit authoring tool is used to manage the XHTML fragments. This approach was used to create <abbr> and <acronym> elements as needed e.g.:

```
<abbr title="Hypertext Markup Language">HTML</abbr>
```

In order to ensure that the elements had been used correctly we ensure that pages are validated after they have been updated.

### Problems Experienced

The harvesting tool processed pages on the UKOLN Web site, which included the QA Focus area. When we examined the automated glossary which had been produced [2] we noticed there were a number of errors in the definitions of abbreviations and acronyms, which were due to errors in the definition of terms on the QA Focus Web site.

Although these errors were quickly fixed, we recognised that such errors were likely to reoccur. We recognised the need to implement systematic quality assurance procedures, especially since such errors would not only give incorrect information to end users viewing the definitions, but also any automated glossary created for the Web site would be incorrect.

In addition, when we read the definitions of the <abbr> and <acronym> elements we realised that there were differences between W3C's definitions of these terms and Oxford English Dictionaries of these terms in English usage.

We also recognised that, even allowing for cultural variations, some terms could be classed either as acronyms or abbreviations. For example the term "FAQ" can either be classed as an acronym and pronounced "fack" or an abbreviation with the individual letters pronounced - "eff-ay-queue".

A summary of these ambiguities is available [3].

We recognised that the <abbr> and <acronym> elements could be used in a number of ways. A formally dictionary definition could be used or an informal explanation could be provided, possibly giving some cultural context. For example the name of the FAILTE project could be formally defined as standing for "Facilitating Access to Information on Learning Technology for Engineers". Alternatively we could say that "Facilitating Access to Information on Learning Technology for Engineers. FAILTE is the gaelic word for 'Welcome', and is pronounced something like 'fawl-sha'".

We realised that there may be common variations for certain abbreviations (e.g. US and USA). Indeed with such terms (and others such as UK) there is an argument that the meaning of such terms is widely known and so there is no need to explicitly define them. However this then raises the issue of agreeing on terms which do not need to be defined.



We also realised that there will be cases in which words which would appear to be acronyms or abbreviations may not in fact be. For example UKOLN, which at one stage stood for 'UK Office For Library And Information Networking' is now no longer an acronym. An increasing number of organisations appear to be no longer expanding their acronym or abbreviation, often as a result of it no longer giving a true reflection of their activities.

Finally we realised that we need to define how the <abbr> and <acronym> elements should be used if the terms are used in a plural form or contain punctuation e.g.: in the sentence:

*JISC's view of ...*

do we use:

```
<acronym title="Joint Information Systems  
Committee">JISC's<acronym> ...
```

or:

```
<acronym title="Joint Information Systems  
Committee">JISC<acronym>'s ...
```

## Our Solutions

### Policies

We recognised that we need to develop a policy on our definition of acronyms or abbreviations and QA procedures for ensuring the quality.

The policies we have developed are:

We will seek to make use of the <acronym> and <abbr> elements on the QA Focus Web site in order to provide an explanation of acronyms and abbreviations used on the Web site and to have the potential for this structured information to be re-purposed for the creation of an automated glossary for the Web site.

We will use the Oxford English Dictionary's definition of the terms acronyms and abbreviations. We treat acronyms as abbreviations which are normally pronounced in standard UK English usage as words (e.g. radar, JISC, etc.); with abbreviations the individual letters are normally pronounced (e.g. HTML, HE, etc.). In cases of uncertainty the project manager will adjudicate.

The elements will be used with the formal name of the acronyms and abbreviations and will not include any punctuation.

We will give a formal definition. Any additional information should be defined in the normal text.

We will not define acronyms or abbreviations if they are no longer to be treated as acronyms or abbreviations.

### Procedures

Implementing QA procedures is more difficult. Ideally acronyms and abbreviations would be defined once within a Content Management System and implemented from that single source. However as we do not have a CMS, this is not possible.

One aspect of QA is staff development. We will ensure that authors of resources on the QA Web site are aware of how these elements may be repurposed, and thus the importance of using correct definitions.

### Future Developments

We will liaise with Tom Heath, developer of the acronym and abbreviation harvester to explore the possibilities of this tool being used to display usage of <abbr> and <acronym> elements on the QA Focus Web site.

Although the issues explored in this case study are not necessarily significant ones the general issue being addressed is quality of metadata. This is an important issue, as in many cases, metadata will provide the 'glue' for interoperable services. We hope that the approaches described in this case study will inform the work in developing QA procedures for other types of metadata.

## References

1. **Exploiting ACRONYM And ABBR HTML Elements**, QA Focus. UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-29/>>
2. **Abbreviations and Acronyms related to Materials**, Higher Education and Computing, UK Centre for Materials Education, <<http://www.materials.ac.uk/acronyms/>>
3. **Abbreviations, Acronyms, Initialisms**, larsholst.info/blog, <<http://larsholst.info/blog/index.php?p=14&more=1#more14>>

## Contact Details

Brian Kelly  
UKOLN  
University of Bath  
BATH, BA1 3LY  
Email: B.Kelly@ukoln.ac.uk

# Providing Access to an EU-funded Project Web Site after Completion of Funding

Case Study

### About This Document

This case study describes how a Web site was ‘mothballed’ after the project funding had ceased.

### Citation Details

*Providing Access to an EU-funded Project Web Site after Completion of Funding*, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-17/>>

**Keywords:** QA, Web. Web site, preservation, long term access, Exploit Interactive, case study

## Background

*Exploit Interactive* [1] was a pan-European Web magazine, which was funded by the European Commission's Telematics for Libraries programme. The magazine was one of the project deliverable of the EXPLOIT project, which aimed to promote the results of EU library projects and to facilitate their take-up by the market of library and information systems. The magazine ran for seven issues between May 1999 and October 2000. During its lifetime the magazine developed and maintained a strong and involved community of *Exploit Interactive* readers, authors, project partners and information providers and provided a forum for discussion within the EU Telematics for Libraries community.

## Problem Being Addressed

Prior to the last issue being published it was recognised that maintaining the site could possibly be a problem. Funding would cease and there would no longer be a member of staff working on the site.

Note that this case study does not address the wider long-term preservation issues. In particular it does not address:

- The formats of the resources: if the file formats used (mainly HTML, CSS JPEG and GIF) become obsolete, we do not address how access can be obtained.
- The functionality of the service provided: if the service becomes unavailable (e.g. due to an operating system upgrade) we do not guarantee that the service will be restored.
- Access to resources if the host institution ceased to exist: we do not guarantee to provide access if our organisation ceases to exist.
- Access to resources due to catastrophe: we do not guarantee to provide access in the event of catastrophes, such as a fire which destroys the server and backups.

The case study provides a pragmatic approach to access to the Web site after the project funding has finished.

## The Approach Taken

It was decided to agree on a short-medium term access strategy for the *Exploit Interactive* Web site. This strategy would list policies and procedures for maintenance of the site for the next 10 years. It would also allow us to allocate money to certain activities.

10 years was decided upon primarily because the preservation industry rule of thumb is that data should be migrated every 10 years. It is unlikely that we will have the resources to migrate the *Exploit Interactive* Web site.

#### **Short-Medium Term Access Strategy**

The Web site's domain name will be kept for at least 3 years after the end of funding.

We will seek to ensure the Web site continues for at least 10 years after the end of funding.

We will seek to ensure that the Web site continues to function, although we cannot give an absolute commitment to this.

We will not commit to fixing broken links to external resources.

We will not commit to fixing non-compliant HTML resources.

We will use the following procedures:

We will have internal administrative procedures to ensure that the domain name bill is paid (even if current staff leave).

We will record the disk space usage of the Web site and provide an estimate of the cost of providing this disk space (in order to monitor the disk space usage costs).

We will run a link checker at least annually and record the number of internal broken links. We will keep an audit trail so that we can see if internal links start breaking.

We will use the UKOLN online calendar to remind staff when to rerun the check.

Any changes to the policy which would result in the Web site being removed need to be agreed by an appropriate management group. This would not be done for purely technical reasons.

The area on which *Exploit Interactive* is held was measured:

Disk Size: 3.92 Gb (3920 Mb)

*Exploit Interactive* live site: 62.9 Mb

*Exploit Interactive* development site: 70.3 Mb

*Exploit Interactive* log files: 292 Mb

*Exploit Interactive* currently takes up 425.4 Mg of disk space.

The cost of this space is negligible bearing in mind you can purchase 30 Gb disk drives for about £40.

We have established that the domain name has been paid for until 23rd October 2008. We feel this is a sufficiently long period of time.

#### **Problems Experienced**

Two years on from the end of funding there have been very few problems adhering to the access strategy. The domain name has been held and a regular link checking audit has been initiated [2] Time spent on the maintenance of the site, such as link checking, has been minimal (about 30 minutes per year to run a link check and provide links to the results).

There are a number of potential problems which we could face:

- The Web site could stop working due to technical difficulties (for example, if the operating system was upgraded it might be necessary to upgrade the backend scripting software).
- The Web site may fail WAI Web accessibility guidelines and there could be legal liabilities in continuing to host the Web site.
- The Web site may be too search-engine friendly and searches from Google, etc. may find out-of-date resources from the *Exploit Interactive* Web site rather than more up-to-date resources provided by UKOLN.

However in practice we think such possibilities are unlikely.

We are confident that we will be able to continue to host this resource for at least 3 years and for a period of up to 10 years. However this is, of course, dependent on our organisation continuing to exist during this period.

### References

1. **Exploit Interactive**, <<http://www.exploit-lib.org/>>
2. **Exploit Interactive Link Check**, <<http://www.exploit-lib.org/link-check/>>

### Contact Details

Marieke Guy  
QA Focus  
UKOLN  
University of Bath  
Tel: +44 1225 385105

# 4.4 Metadata

---

## Background

This section addresses the area of metadata.

The briefing documents seek to describe best practices in this area.

## Briefing Documents

The following briefing documents which address the area of metadata have been produced:

- *An Introduction To Metadata* (briefing-40)
- *Metadata Deployment* (briefing-41)
- *QA For Metadata* (briefing-42)
- *Choosing A Metadata Standard* (briefing-63)
- *Metadata And Subject Searching* (briefing-64)
- *IMS Question And Test Interoperability* (briefing-36)

Advisory documents which cover specific technical areas are available within the section on the appropriate technical area.

## Case Studies

The following case studies which address the area of metadata have been produced:

- *Managing And Using Metadata In An E-Journal*, (case-study-01)
- *Gathering the Jewels: Creating a Dublin Core Metadata Strategy*, (case-study-13)

Case studies which cover specific technical areas are available within the section on the appropriate technical area.

# An Introduction To Metadata

Briefing

### About This Document

This briefing document gives an executive overview of the role of metadata.

### Citation Details

*An Introduction To Metadata*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-40/>>

**Keywords:** metadata, briefing

## What is Metadata?

Metadata is often described as “data about data”. The concept of metadata is not new – a Library catalogue contains metadata about the books held in the Library. What is new is the potential that metadata provides in developing rich digital library services.

The term metadata has come to mean structured information that is used by automated processes. This is probably the most useful way to think about metadata [1].

## The Classic Metadata Example

The classic example of metadata is the library catalogue. A catalogue record normally contains information about a book (title, format, ISBN, author, etc.). Such information is stored in a structured, standardised form, often using an international standard known as MARC. Use of this international standard allows catalogue records to be shared across organisations.

## Why is Metadata So Important?

Although metadata is nothing new, the importance of metadata has grown with the development of the World Wide Web. As is well-known the Web seeks to provide universal access to distributed resources. In order to develop richly functional Web applications which can exploit the Web’s global information environment it is becoming increasingly necessary to make use of metadata which describes the resources in some formal standardised manner.

## Metadata Standards

In order to allow metadata to be processed in a consistent manner by computer software it is necessary for metadata to be described in a standard way. There are many metadata standards available. However in the Web environment the best known standard is the Dublin Core standard which provides an agreed set of core metadata elements for use in resource discovery.

The *Dublin Core* standard (formally known as the Dublin Core Metadata Element Set) has defined 15 core elements: Title, Creator, Subject, Description, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights [2].

The core element set is clearly very basic. A mechanism for extending Dublin Core elements has been developed. This allows what is known as *Qualified Dublin Core* elements to refine the core elements. For example `DC.Date.Created` refines the `DC.Date` element by allowing the date of creation of the resource to be described. `DC.Date.Modified` can be used to describe the date on which the resource was

changed. Without the qualifiers, it would not be possible to tell which date related to which event. Work is in progress in defining a common set of qualifiers.

### Using Metadata

The Dublin Core standard defines a set of core elements. The standard does not specify how these elements should be deployed on the Web. Initially consideration was given to using Dublin Core by embedding it within HTML pages using the <meta> element e.g. <meta name="DC.Creator" content="John Smith">. However this approach has limitations: initially HTML was not rich enough to all metadata schemes to be including (which could specify that a list of keywords are taken from the Library Of Congress list); it is not possible to define relationships for metadata elements (which may be needed if, for example, there are multiple creators of a resource) and processing the metadata requires the entire HTML document to be downloaded.

In order to address these concerns a number of alternative approaches for using metadata have been developed. RDF (Resource Description Framework) [3], for example, has been developed by W3C as a framework for describing a wide range of metadata applications. In addition OAI (Open Archives Initiative) [4] is an initiative to develop and promote interoperability standards that aim to facilitate the efficient dissemination of content.

In addition to selecting the appropriate standards use of metadata may also require use of a metadata management system and a metadata repository.

### References

- 1 **Metadata Demystified**, NISO,  
<[http://www.niso.org/standards/Metadata\\_Demystified.pdf](http://www.niso.org/standards/Metadata_Demystified.pdf)>
- 2 **Dublin Core Metadata Element Set**, DCMI,  
<<http://dublincore.org/documents/dces/>>
- 3 **Resource Description Framework (RDF)**, W3C, <<http://www.w3.org/RDF/>>
- 4 **Open Archives Initiative (OAI)**, <<http://www.openarchives.org/>>
- 5 **Information Environment Home**, JISC,  
<[http://www.jisc.ac.uk/index.cfm?name=ie\\_home](http://www.jisc.ac.uk/index.cfm?name=ie_home)>



# Metadata Deployment

**Briefing**

### About This Document

This briefing document gives an executive overview of the ways in which metadata can be deployed.

### Citation Details

*Metadata Deployment*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-41/>>

**Keywords:** metadata, deployment, briefing

## Introduction

This document describes the issues you will need to address in order to ensure that you make use of appropriate approaches for the deployment of metadata within your project.

## Why Do You Wish To Use Metadata?

The first question you should address is “Why do you wish to use metadata?”. You may have heard that metadata is important. You may have heard that metadata will help solve many problems you have with your project. You may have heard that others are using metadata and you don’t wish to be left behind. Although all of these points have some validity, they are not sufficient in isolation to justify the time and effort needed in order to deploy metadata effectively.

You should first specify the problem you wish to address using metadata. It may be that you wish to allow resources on your Web site to be found more easily from search engines such as Google. It may be that you wish to improve local searching on your Web site. It may be that you wish to interoperate with other projects and services. Or it may be that you wish to improve the maintenance of resources on your Web site. In all of these cases metadata may have a role to play; however different approaches may be needed to tackle these different problems and, indeed, approaches other than use of metadata may be more effective (for example, Google makes only limited use of metadata so an alternative approach may be needed).

## Identifying The Functionality To Be Provided

Once you have clarified the reasons you wish to make use of metadata you should identify the end user functionality you wish to provide. This is needed in order to define the metadata you will need, how it should be represented and how it should be created, managed and deployed.

## Choosing The Metadata Standard

You will need to choose the metadata standard which is relevant for your purpose. In many cases this may be self-evident – for example, your project may be funded to develop resources for use in an OAI environment, in which case you will be using the OAI application.

### Metadata Modelling

It may be necessary for you to decide how to model your metadata. For example if you wish to use qualified Dublin Core metadata you will have to choose the qualifiers you wish to use.

### Metadata Management

It is important that you give thought to the management of the metadata. If you don't you are likely to find that your metadata becomes out-of-date. Since metadata is not normally displayed to end users but processed by software you won't even be able to use visual checking of the metadata. Poor quality metadata is likely to be a major barrier to the deployment of interoperable services.

If, for example, you embed metadata directly into a file, you may find it difficult to maintain the metadata (e.g. the creator changes their name or contact details). A better approach may be use of a database (sometimes referred to as a metadata repository) which provides management capabilities.

#### Example Of Use Of This Approach

The *Exploit Interactive* [1] e-journal was funded by the EU and developed by UKOLN. Metadata was required in order to provide enhanced searching for the end user. The specific functionality required was the ability to search by issue, article type, author and title and by funding body. In addition metadata was needed in order to assist the project manager producing reports, such as the numbers of different types of articles. This functionality helped to identify the qualified Dublin Core elements required.

The MS SiteServer software used to provide the service provided an indexing and searching capability for processing arbitrary metadata. It was therefore decided to provide Dublin Core metadata stored in <META> tags in HTML pages. In order to allow the metadata to be more easily converted into other formats (e.g. XHTML) the metadata was held externally and converted to HTML by server-side scripts.

A case study which gives further information (and describes the limitations of the metadata management approach) is available [2].

### References

- 1 **Exploit Interactive**, <<http://www.exploit-lib.org/>>
- 2 **Managing And Using Metadata In An E-Journal**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-01/>>

# Quality Assurance For Metadata

Briefing

### About This Document

This briefing document gives an executive overview of quality assurance procedures for use with metadata applications.

### Citation Details

*Quality Assurance For Metadata*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefing/briefing-42/>>

**Keywords:** metadata, quality assurance, QA, briefing

## Introduction

Decisions on the use of metadata in your project should be based on the functionality to be provided by the metadata. The functionality required will influence the metadata standards to be used and the architecture for managing and deploying your metadata. However this is not the end of the matter. You will also need to ensure that you have appropriate quality assurance procedures to ensure that your metadata is fit for its purpose.

## What Can Go Wrong?

There are a number of ways in which services based on metadata can go wrong, such as:

**Incorrect content:** The content of the metadata may be incorrect or out-of-date. There is a danger that metadata content is even more likely to be out-of-date than normal content, as content is normally visible, unlike metadata which is not normally displayed on, say, a Web page. In addition humans can be tolerant of errors, ambiguities, etc. in ways that software tools normally aren't.

**Inconsistent content:** The metadata content may be inconsistent due to a lack of cataloguing rules and inconsistent approaches if multiple people are involved in creating metadata.

**Non-interoperable content:** Even if metadata is consistent within a project, other projects may apply different cataloguing rules. For example the date 01/12/2003 could be interpreted as 1 December or 12 January if projects based in the UK and USA make assumptions about the date format.

**Incorrect format:** The metadata may be stored in a non-valid format. Again, although Web browsers are normally tolerant of HTML errors, formats such as XML insist on validation against a DTD or schema.

**Errors with metadata management tools:** Metadata creation and management tools could output metadata in invalid formats.

**Errors with the workflow process:** Data processed by metadata or other tools could become corrupted through the workflow. As a simple example a MS Windows character such as © could be entered into a database and then output as an invalid character in a XML file.

### QA For Metadata Content

You should have procedures to ensure that the metadata content is correct when created and is maintained as appropriate. This could involve ensuring that you have cataloguing rules, ensuring that you have mechanisms for ensuring the cataloguing rules are implemented (possibly in software when the metadata is created). You may also need systematic procedures for periodic checking of the metadata.

### QA For Metadata Formats

As metadata which is to be reused by other applications is increasingly being stored in XML it is essential that the format is compliant (otherwise tools will not be able to process the metadata). XML compliance checking can be implemented fairly easily. More difficult will be to ensure that metadata makes use of appropriate XML schemas.

### QA For Metadata Tools

You should ensure that the output from metadata creation and management tools is compliant with appropriate standards. You should expect that such tools have a rich set of test suites to validate a wide range of environments. You will need to consider such issues if you develop your own metadata management system.

### QA For Metadata Workflow

You should ensure that metadata does not become corrupted as it flows through a workflow system.

### A Fictitious Nightmare Scenario

A multimedia e-journal project is set up. Dublin Core metadata is used for articles which are published. Unfortunately there are documented cataloguing rules and, due to a high staff turnover (staff are on short term contracts) there are many inconsistencies in the metadata (John Smith & Smith, J.; University of Bath and Bath University; etc.)

The metadata is managed by a homegrown tool. Unfortunately the author metadata is output in HTML as DC.Author rather than DC.Creator. In addition the tool output the metadata in XHTML 1.0 format which is embedded in HTNML 4.0 documents.

The metadata is created by hand and is not checked. This results in a large number of typos and use of characters which are not permitted in XML without further processing (e.g. £, —, “, ”, and &).

Rights metadata for images which describes which images can be published freely and which is restricted to local use becomes separated from the images during the workflow process.

# Choosing A Metadata Standard

Briefing

### About This Document

This briefing document describes the processes which can be used when selecting the standards for use with metadata.

### Citation Details

*Choosing A Metadata Standard*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-63/>>

**Keywords:** metadata, briefing

## Introduction

Resource discovery metadata is an essential part of any digital resource. If resources are to be retrieved and understood in the distributed environment of the World Wide Web, they must be described in a consistent, structured manner suitable for processing by computer software. There are now many formal standards. They range from simple to rich formats, from the loosely structured to the highly structured, and from proprietary, emerging standards, to international standards.

## How To Choose A Standard

There is no set decision-making procedure to follow but here are some factors that should normally be considered:

**Purpose of metadata:** A well-articulated definition of purposes at the outset can act as a benchmark against which to compare standards. Metadata may be for:

1. **Retrieval:** ‘Can I find the resource?’
2. **Identification:** Can I distinguish the resource from other similar resources (e.g. similar titles, or other editions or versions)?
3. **Access:** Can I use the resource (e.g. are there legal restrictions on access and usage and is it in a format I can handle)?

**Attributes of resource:** It is important that you also identify your resource type (e.g. text, image), its domain of origin (e.g. library, archive or museum), subject (e.g. visual arts, history) and the specific features that are essential to an understanding of it. Datasets, digital texts, images and multimedia objects, for instance, clearly have very different attributes. Does your resource have pagination or is it three-dimensional? Was it born digital or does it have a hard-copy source? Which attributes will the user need to know to understand the resource?

**Design of standard:** Metadata standards have generally been developed in response to the needs of specific resource types, domains or subjects. Therefore, once you know the type, domain and broad subject of your resource, you should be able to draw up a shortlist of likely standards. Here are some of the better-known ones:

1. **Anglo-American Cataloguing Rules (AACR2)** – Library resources.  
<<http://www.aacr2.org/index.html>>
2. **Data Documentation Initiative (DDI)** – Social sciences, datasets.  
<<http://www.icpsr.umich.edu/DDI/index.html>>

3. **Dublin Core (DC)** – All domains, resource types, and subjects.  
<<http://dublincore.org/>>
4. **Encoded Archival Description (EAD)** – Archives.  
<<http://www.tei-c.org/>>
5. **ISAD(G)** – Guidelines for the preparation of archival descriptions  
<[http://www.hmc.gov.uk/icacds/eng/ISAD\(G\).pdf](http://www.hmc.gov.uk/icacds/eng/ISAD(G).pdf) >
6. **MARC 21** -- Libraries, bibliographic records.  
<<http://www.loc.gov/marc/>>
7. **RSLP Collection-level description** – Collections of all subjects, domains and types.  
<<http://www.ukoln.ac.uk/metadata/cld/>>
8. **SPECTRUM** – Museum objects.  
<<http://www.mda.org.uk/spectrum.htm>>
9. **Text Encoding Initiative (TEI)** – Digital texts.  
<<http://www.tei-c.org/>>
10. **VRA Core 3.0** – Visual art images.  
<<http://www.vraweb.org/vracore3.htm>>

The key attributes of your resource can be matched against each standard in turn to find the best fit. Is there a dedicated element for each attribute? Are the categories of information relevant and at a suitable level of detail?

**Granularity:** At this point it is worth considering whether your metadata should (as is usual) be created at the level of the text, image or other such item or at collection level. Collection-level description may be provided where item-level metadata is not feasible or as an additional layer providing an overview of the resource. This could be valuable for large-scale digitisation projects or portals where item-level searching may retrieve an unmanageable number of ‘hits’. Digital reproductions may be grouped like their real world sources – e.g. by subject or provenance - or be assigned to multiple ‘virtual collections’. The RSLP Collection Level Description is emerging as the leading format in this area.

**Interoperability:** It is important, wherever possible, to choose one of the leading standards (such as those listed above) from within your subject community or domain. This should help to make your resource accessible beyond the confines of your own project. Metadata that is in a recognisable common format may be harvested by subject or domain-wide portals and cross-searched with resources from many other institutions. In-house standards may be tailored to your precise needs but are unlikely to be compatible with other standards and should be used only where nothing suitable already exists. If your over-riding need is for interoperability across *all* domains or subjects, Dublin Core may be the most suitable standard but it may lack the richness required for other purposes. Care should be taken to ensure that in-house standards at least map to Dublin Core or one of the DC Application profiles.

**Support:** Using a standard that is well supported by a leading institution can also bring cost benefits. Implementation guidance, user guidance, examples, XML/RDF schemas, crosswalks, multi-lingual capacity, and software tools may pre-exist, thus easing the process of development, customisation and update.

**Growth:** Consider too whether the standard is capable of further development? Are there regular working groups and workshops devoted to the task?

**Extensibility:** Also, does the standard permit the inclusion of data elements drawn from other schemas and the description of new object types?' It may be necessary to 'mix and match' elements from more than one standard.

**Reputation:** Funding bodies will be familiar with established, international standards – something, perhaps, to remember when applying for digitisation grants.

**Ease of use:** Be aware that the required level of expertise can vary greatly between standards. AACR2 and MARC 21, for instance, may produce rich bibliographic description but require the learning of rules. The simpler Dublin Core may allow creators to produce their own metadata records with no extensive training.

**Existing experience:** Have staff at your organisation used the metadata standard before? If so, the implementation time may be reduced.

### Summary

There is no single standard that is best for all circumstances. Each is designed to meet a need and has its own strengths and weaknesses. Start by considering the circumstances of the individual digital project and identify the need(s) or purpose(s) that the metadata will need to satisfy. Once that is done, one can evaluate rival metadata schemas and find the best match. A trade-off will normally have to be made between the priorities listed above.

### Further Information

- *Digital Libraries: Metadata Resources*. Sophie Felfoldi, International Federation of Library Associations and Institutions.  
<<http://www.ifla.org/II/metadata.htm>>
- *Application Profiles: Mixing And Matching Metadata Schemas*. Rachel Heery and Manjula Patel, *Ariadne*, No. 25, 24 September 2000.  
<<http://www.ariadne.ac.uk/issue25/app-profiles/>>

# Metadata And Subject Searching

Briefing

### About This Document

This briefing document describes how metadata can be used for subject searching.

### Citation Details

*Metadata And Subject Searching*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-64/>>

**Keywords:** metadata, subject searching, briefing

## Introduction

Digital collections are only likely to make an impact on the Web if they are presented in such a way that users can retrieve their component parts quickly and easily. This is true even if they have been well selected, digitised to a suitable standard and have appropriate metadata formats. Subject-based access to the collection through searching and/or browsing a tree-like structure can greatly enhance the value of your resource.

## Subject Access – Some Options

Subject-based access can be provided in several ways:

**Keywords:** A simple but crude method is to anticipate the terms that an unguided searcher might intuitively choose and insert them into a keyword field within relevant records. For instance, the text of *Ten days that shook the world* [1], a classic narrative of the events of 1917, is more likely to be retrieved if the keywords *Russian Revolution* are added by the cataloguer (based on his/her analysis of the resource and subject knowledge) and if the keyword field is included in the search. In the absence of an agreed vocabulary, however, variant spellings (*labor* versus *labour*), and synonyms or near synonyms (*Marxist* versus *Communist*) that distort retrieval are likely to proliferate.

**Thesauri and subject schemes:** Controlled vocabularies, known as thesauri, can prevent inconsistent description and their use is recommended. They define preferred terms and their spelling. If the thesaurus structure is shown on the search interface, users may be guided through broader-term, narrower-term and associated-term relationships to choose the most appropriate keyword with which to search. Take care to choose a vocabulary appropriate to the scope of your resource. A broad and general collection might require a correspondingly universal vocabulary, such as the Library of Congress Subject Headings (LCSH) [2]. A subject-specific vocabulary, such as the Getty Art and Architecture Thesaurus (AAT) [3], may provide a more limited but detailed range of terms appropriate for a tightly focused collection.

**Classification schemes:** Keywords and thesauri are primarily aids to searching but browsing can often be a more rewarding approach - particularly for users new to a given subject area. Thesauri are not always structured ideally for browsing – as when related or narrower terms are listed alphabetically rather by topical proximity. Truly effective browsing requires the use of a subject classification scheme. A classification scheme arranges resources into a hierarchy on the basis of their subject but differs from a thesaurus in using a sophisticated alphanumeric notation to ensure that related subjects will be displayed in close, browsable, proximity. A well-designed classification scheme



should present a navigable continuum of topics from one broad subject area to another and in this way guide the user related items that might otherwise be missed, as in this example from the Dewey Decimal Classification (DDC) [4].

700	Arts, fine and decorative
740	Drawing and decorative arts
745	Decorative arts
745.6	Calligraphy, heraldic design, illumination
745.66	Heraldic design

**Table 1: Example Of Dewey Decimal Classification**

The notation does not necessarily have to be displayed on screen, however. The subject terms, rather than their respective numbers, may mean more to the user. Another tip is to assign multiple classification numbers to any item that crosses subjects. Digital items can have several ‘virtual’ locations, unlike a book, which is tied to a single position on a shelf.

Keywords, thesauri and classification can be used in combination or individually.

### Choosing A Classification Scheme

The most important consideration when choosing a classification scheme is to select the one that best fits the subject, scope and intended audience of your resource.

**Universal classification schemes:** These are particularly appropriate where collections and their audiences span continents, subjects and languages. Dewey Decimal Classification (DDC) [5], for instance, is the most widely recognised scheme worldwide, whilst UDC (Universal Decimal Classification) [6] is predominant in Europe and Asia. Well-established schemes of this sort are most likely to have user-friendly online implementation tools.

**National or subject-specific schemes:** More specific collections are usually best served by schemes tailored to a single country (e.g. BC Nederlandse Basisclassificatie) [7], language, or subject (e.g. NLM National Library of Medicine) [8]. If nothing suitable exists, a scheme can be created in-house.

**Homegrown schemes:** Project-specific schemes can be flexible, easy to change and suited wholly to one’s own needs so that there are no empty categories or illogical subject groupings to hinder browsing. However, the development process is costly, time-consuming and requires expert subject-knowledge. Users are sure to be unfamiliar with your categories and, perhaps worst of all, such schemes are unlikely to be interoperable with the broader information world and will hinder wider cross searching. They should be regarded very much as a last resort.

**Adapting an existing scheme:** A better approach is normally to adapt an existing scheme – by rearranging empty classes, raising or lowering branches of the hierarchy, renaming captions, or extending the scheme. Be aware, though, that recurring notation may be found within a scheme at its various hierarchical levels or the scheme might officially be modified over time, both of which can lead to conflict between the official and customised versions. Take care to document your changes to ensure consistency through the lifetime of the project. Some well-known Internet search-services (e.g.

Yahoo!) [9] have developed their own classifications but there is no guarantee that they will remain stable or even survive into the medium term.

**Double classification:** It may be worthwhile classifying your resource using a universal scheme for cross-searching and interoperability in the wider information environment and at the same time using a more focused scheme for use within the context of your own Web site. Cost is likely to be an issue that underpins all of these decisions. For instance, the scheme you wish to use may be freely available for use on the Internet or alternatively you may need to pay for a licence.

### References

1. *Ten Days That Shook the World*. Reed, J. New York: Boni & Liveright, 1922; Bartleby.com, 2000. <<http://www.bartleby.com/79/>>
2. *Library of Congress Authorities*, <<http://authorities.loc.gov/>>
3. *Art & Architecture Thesaurus Online*, <[http://www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)>
4. *Dewey Decimal Classification and Relative Index*, Dewey, M. in Joan S. Mitchell et-al (ed.), 4 vols, (Dublin, Ohio: OCLC, 2003), vol.3, p.610.
5. Dewey Decimal Classification, <<http://www.oclc.org/dewey/>>
6. *Universal Decimal Classification*, <<http://www.udcc.org/>>
7. *Nederlandse Basisclassificatie (Dutch Basic Classification)*, <http://www.kb.nl/dutchess/>
8. *National Library of Medicine*, <<http://wwwcf.nlm.nih.gov/class/>>
9. *Yahoo!*, <<http://www.yahoo.com/>>

### Further Information

- *DESIRE Information Gateways Handbook*, Cross, P. et al <<http://www.desire.org/handbook/2-5.html>>
- *Controlled Vocabularies, Thesauri And Classification Systems Available In The WWW. DC Subject (1996 and ongoing)*, Koch, T. <<http://www.ub2.lu.se/metadata/subject-help.html>>
- *The Role Of Classification Schemes In Internet Resource Description And Discovery*, Koch, T. & Day, M. <<http://www.ukoln.ac.uk/metadata/desire/classification/>>
- *Beyond Bookmarks: Schemes For Organising The Web*, McKiernan, G. <<http://www.iastate.edu/~CYBERSTACKS/CTW.htm>>

# IMS Question And Test Interoperability

Briefing

### About This Document

This briefing document describes an international specification for computer based questions and tests, suitable for those wishing to use computer based assessments in courses.

### Citation Details

*IMS Question And Test Interoperability*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-36/>>

**Keywords:** metadata, IMS, briefing

This document describes an international specification for computer based questions and tests, suitable for those wishing to use computer based assessments in courses.

## What Is IMS Question And Test Interoperability?

Computers are increasingly being used to help assess learning, knowledge and understanding. IMS Question and Test Interoperability (QTI) is an international specification for a standard way of sharing such test and assessment data. It is one of a number of such specifications being produced by the IMS Global Learning Consortium to support the sharing of computer based educational material such as assessments, learning objects and learner information.

This new specification is now being implemented within a number of assessment systems and Virtual Learning Environments. Some systems store the data in their own formats but support the export and import of question data in IMS QTI format. Other systems operate directly on IMS QTI format data. Having alternative systems conforming to this standard format means that questions can be shared between institutions that do not use the same testing systems. It also means that banks of questions can be created that will be usable by many departments.

## Technical Details

The QTI specification uses XML to record the information about assessments. XML is a powerful and flexible markup language that uses 'tags' rather like HTML. The IMS QTI specification was designed to be pedagogy and subject neutral. It supports five different type of user response (item selection, text input, numeric input, xy-position selection and group selection) that can be combined with several different input techniques (radio button, check box, text entry box, mouse xy position dragging or clicking, slider bar and others). It is able to display formatted text, pictures, sound files, video clips and even interactive applications or applets. How any particular question appears on the screen and what the user has to do to answer it may vary between different systems, but the question itself, the knowledge or understanding required to answer it, the marks awarded and the feedback provided should all remain the same.

The specification is relatively new. Version 1.2 was made public in 2002, and a minor upgrade to version 1.2.1 was made in 2003 that corrected errors and ambiguities. The specification is complex comprising nine separate documents. Various commercial assessment systems (e.g. Questionmark, Granada, MedWeb, Canvas Learning) have implemented some aspect of IMS QTI compatibility for their assessments. A number of academic systems are also being developed to comply with the specification, including the TOIA project which will have editing and course management facilities, the STOMP

system which was used with students for the first time in the autumn of 2002, and a Scottish Enterprise system called Oghma which is currently being developed.

### Discipline Specific Features

A disadvantage of such a standard system is that particular features required by some disciplines are likely to be missing. For example, engineering and the sciences need to be able to deal with algebraic expressions, the handling of both accuracy and precision of numbers, the use of alternative number bases, the provision of randomised values, and graphical input. Language tests need better textual support such as the presetting of text entry boxes with specific text and more sophisticated text based conditions. Some of these features are being addressed by groups such as the CETIS assessment SIG.

### What This Means To Academics

If you are starting or planning to use computer based tests, then you need to be aware of the advantages of using a standard-compliant system. It is a good idea to choose a system that will allow you to move your assessments to other systems at a later time with the minimum of effort or to be able to import assessments authored elsewhere.

A consideration to bear in mind, however, is that at this early stage in the life of the specification there will be a range of legacy differences between various implementations. It will also remain possible with some 'compliant' systems to create non-standard question formats if implementation specific extensions are used. The degree of conformity of any one system is a parameter that is difficult to assess at any time. Tools to assist with this are now beginning to be discussed, but it will be some time before objective measures of conformance will be available. In view of this it is a good idea to keep in touch with those interested in the development of the specification, and the best way within UK HE is probably via the CETIS Assessment SIG Web site.

It is important that the specification should have subject specific input from academics. The needs of different disciplines are not always well known and the lack of specific features can make adoption difficult. Look at the examples on the CETIS Web site and give feedback on areas where your needs are not being met.

### Further Information

Further information is available at <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-36/>>.

### Acknowledgments

This document was originally written by Niall Sclater and Rowin Cross of CETIS and adapted by Dick Bacon, Department of Physics, University of Surrey, consultant to the LTSN Physical Sciences Centre.

The original briefing paper (PDF format) is available on the CETIS Web site. The version available on this Web site was originally published in the LTSN Physical Science News (Centre News issue 10).

# Managing And Using Metadata In An E-Journal

Case study

### About This Document

This case study describes the approaches to choosing and deploying Dublin Core metadata.

### Citation Details

*Managing And Using Metadata In An E-Journal*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-01/>>

### Related Documents

See also the *An Introduction To Metadata* (briefing-41) and *Deploying Metadata* (briefing-41) briefing documents.

**Keywords:** Web, metadata, Dublin Core, Exploit Interactive, case study

## Background

The *Exploit Interactive* e-journal [1] was funded by the EU's Telematics For Libraries programme to disseminate information about projects funded by the programme. The e-journal was produced by UKOLN, University of Bath.

*Exploit Interactive* made use of Dublin Core metadata in order to provide enhanced local search facilities. This case study describes the approaches taken to the management and use of the metadata, difficulties experienced and lessons which have been learnt.

## The Need For Metadata

Metadata needed to be provided in order to provide richer searching than would be possible using standard free-text indexing. In particular it was desirable to allow users to search on a number of fields including Author, Title and Description

In addition it was felt desirable to allow users to restrict searches by issues by article type (e.g. feature article, regular article, news, etc.) and by funding body (e.g. EU, national, etc.) These facilities would be useful not only for end users but also by the editorial team in order to collate statistics needed for reports to the funders.

## The Approach Taken

The metadata was stored in a `article_defaults.ssi` file which was held in the directory containing an article. The metadata was held as a VBscript assignment. For example, the metadata for the *The XHTML Interview* article [2] was stored as:

```
doc_title = "The XHTML Interview"
author="Kelly, B."
title="WebWatching National Node Sites"
description = "In this issue's Web Technologies column we ask
Brian Kelly to tell us more about XHTML."
article_type = "regular"
```

This file was included into the article and converted into HTML <META> tags using a server-side include file.

Storing the metadata in a neutral format and then converting it into HTML <META> tags using a server-side script meant that the metadata could be converted into other formats (such as XHTML) by making a single alteration to the script.

### The Service Provided

It was possible to index the contents of the <META> tags using Microsoft's SiteServer software in order to provide enhanced search facilities, as illustrated.

As illustrated in Figure 1 it is possible to search by issue, article type, project category, etc.



Figure 1: Standard Search Interface

Alternative approaches to providing the search interface can be provided. An interface which uses a Windows-explorer style of interface is shown in Figure 2.

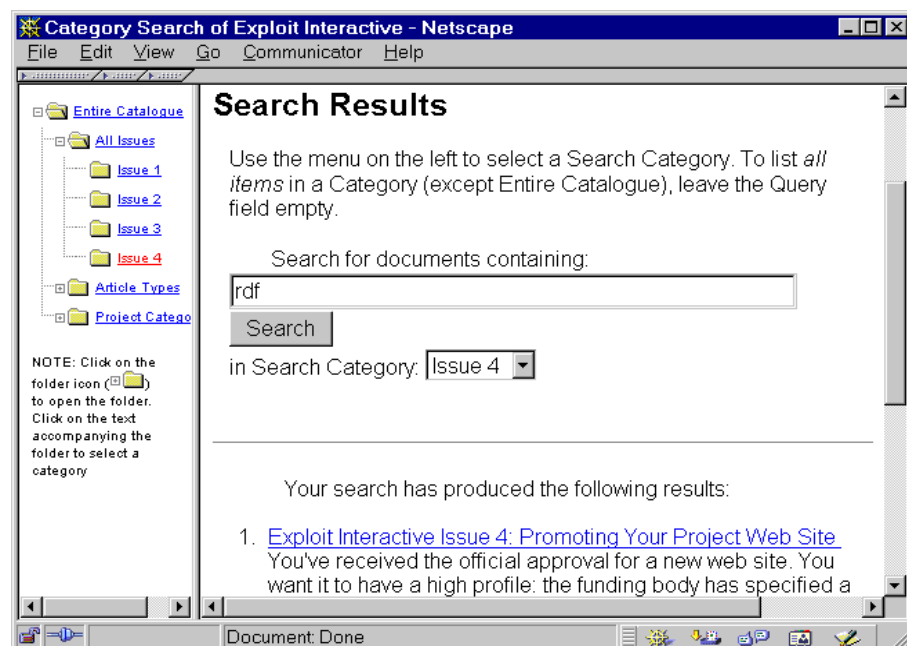


Figure 2: Alternative Search Interface

### Problems Experienced

Initially when we attempted to index the metadata we discovered that it was not possible to index <META> tags with values containing a full stop, such as <meta name="DC.Title" content="The XHTML Interview">.

However we found a procedure which allowed the <META> tags to be indexed correctly. We have documented this solution [3] and have also published an article describing this approach [4].

During the two year lifetime of the *Exploit Interactive* e-journal three editors were responsible for its publication. The different editors are likely to have taken slightly different approaches to the creation of the metadata. Although the format for the author's name was standardised (surname, initial) the approaches to creation of keywords, description, etc. metadata was not formally documented and so, inevitably, different approaches will have been adopted. In addition there was no systematic checking for the existence of all necessary metadata fields and so some may have been left blank.

### Things We Would Do Differently

The approaches which were taken provided a rich search service for our readers and enabled the editorial team to easily obtain management statistics. However if we were to start over again there are a number of changes we would consider making.

Although the metadata is stored in a neutral format which allows the format in which it is represented to be changed by updating a single server-side script, the metadata is closely linked with each individual article. The metadata cannot easily be processed independently of the article. It is desirable, for example, to be able to process the metadata for every article in a single operation - in order to, for example, make the metadata available in OAI format for processing by an OAI harvester.

In order to do this it is desirable to store the metadata in a database. This would also have the advantage of allowing the metadata to be managed and allow errors (e.g. variations of author's names, etc.) to be cleaned.

Use of a database as part of the workflow process would enable greater control to be applied for the metadata: for example, it would enable metadata such as keywords, article type, etc. to be chosen from a fixed vocabulary, thus removing the danger of the editor misspelling such entries.

### References

1. **Exploit Interactive**, <<http://www.exploit-lib.org/>>
2. **The XHTML Interview**, *Exploit Interactive*, issue 6, 26th June 2000, <<http://www.exploit-lib.org/issue6/xhtml/>>
3. **Configuring MS SiteServer To Support Dublin Core Metadata**, UKOLN, <<http://www.ukoln.ac.uk/web-focus/papers/www9/resource-discovery/tech-spec/>>
4. **A Lightweight Approach To Support Of Resource Discovery Standards**, Kelly, B., Poster Proceedings, Ninth International World Wide Web Conference, May 2000, <<http://www.ukoln.ac.uk/web-focus/papers/www9/resource-discovery/>>

### Contact Details

Brian Kelly  
UKOLN  
University of Bath  
BATH  
Email: [b.kelly@ukoln.ac.uk](mailto:b.kelly@ukoln.ac.uk)

# Gathering the Jewels: Creating a Dublin Core Metadata Strategy

Case study

### About This Document

This case study describes the approaches to choosing and deploying Dublin Core metadata.

### Citation Details

*Gathering the Jewels: Creating a Dublin Core Metadata Strategy*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-13/>>

### Related Documents

See also the *An Introduction To Metadata* (briefing-40) and *Deploying Metadata* (briefing-41) briefing documents.

**Keywords:** metadata, Dublin Core, Gathering the Jewels, case study

## Background

**Gathering the Jewels** [1] was established by a consortium of the following bodies: National Library of Wales, Society of Chief Librarians (Wales), National Museums and Galleries of Wales, Federation of Welsh Museums, Archives Council Wales, Royal Commission of Ancient and Historic Monuments Wales, Council of Museums in Wales, Wales Higher Education Libraries Forum and the Welsh County Archivists Group. The goal of the project was to digitise 23,000 items from approximately 190 libraries, museums and archives all over Wales and to present them on the Internet by means of a searchable database.

## The Approach Taken

The nature of the project has four important consequences for the way we approach the collection of metadata:

- We do not hold the materials being digitised, so we are dependent on what information (metadata) the holding institutions can supply.
- There is no standard way of cataloguing across libraries, museums and archives - and even within professions there are variations, so we have to accept metadata in a variety of different forms.
- Very few institutions catalogue to the Dublin Core standard.
- The Web site has to be bilingual.

## Problems Experienced

When we first looked at the question of metadata, and came face to face with the reality of the difficulties listed above, the problem seemed massive. To make things worse, the Dublin Core elements apparently needed their own glossary to make them intelligible. These were dark days. However, things very quickly improved.

In the first place, we talked to professionals from the National Library of Wales's metadata unit, who reassured us that the Dublin Core elements could be translated into English. But more importantly than that, they showed us that the elements could be made to work for us: that there is a degree of flexibility about what many of the



elements can be taken to mean; that the most important thing is to be consistent, however you interpret a particular element.

For example, there is a Dublin Core element called "Publisher". The National Library would interpret this as the organisation publishing the digital material on the Internet - i.e., us; we, on the other hand, would prefer to use it for the institution providing us with the material. Both interpretations are apparently valid, so long as they are used consistently. We also interpret the "Title" element in a way that will let us use it as a caption to the image when it is displayed on the Internet.

We also made a couple of key decisions. We were not here to catalogue 23,000 items to the Dublin Core standard. Also, the output of the whole project was to be a Web site linked to a searchable database – so the bare minimum metadata we had to collect was defined by the requirements of the Web site and search mechanisms for the database. In other words, an image appearing on a user's computer screen had to have a certain amount of information associated with it (a caption, a date, a credit to the institution that gave it to us, as well as subject and place-name keywords, etc.); any other metadata we could collect would be nice (the 'extent' or size, the 'medium', etc.) but not essential.

This was also our "Get Out Of Jail Free" card with regard to the bilingual aspects of the Web site. Anything which the user will see or search on has to be in English and Welsh. Other Dublin Core elements are recorded in English only. (I should emphasise that this decision was taken on the advice of the National Library of Wales and is based entirely on the limitations of existing computer systems and the amount of time that fully bilingual metadata would take to translate and enter; it has nothing to do with political preferences for one language or the other.)

As a result we have divided our metadata into four categories. Core elements are those that are mandatory, and which will be viewed or searched by the user, together with copyright information; Important elements are those which we may not get from an institution but which we will supply ourselves, such as a detailed interpretative description of the image. Technical elements are those which record how the material was digitally captured; we do not regard these as a high priority but as they are easy to enter in batches we always make sure we complete them. And finally Useful elements are the other Dublin Core elements that we will collect if the institution can supply them easily, but which we will otherwise leave blank until such time as cataloguing to the Dublin Core standard becomes the norm.

## Metadata Schema

### Core Elements

<b>Title English</b>	A caption for the item, no more than one line
<b>Title Welsh</b>	As above, in Welsh
<b>Identifier</b>	Unique ID of item, e.g., accession or catalogue number
<b>Location</b>	Place name most significantly associated with the image
<b>Period</b>	Period of subject depicted
<b>Copyright</b>	Brief details of copyright ownership and clearance

### Important Elements

<b>Creator</b>	Institution/individual that produced the original
<b>Date</b>	Date of production, e.g., when a painting was painted
<b>Description</b>	Max. 200 word description of the resource and its content
<b>Description Welsh</b>	As above, in Welsh

### Technical Elements

<b>Capture device</b>	e.g., the scanner or camera used to capture the image
<b>Capture history</b>	e.g., the software employed
<b>Manipulation history</b>	file format master created in, quality control checks, etc.
<b>Resolution of master</b>	number of pixels (e.g., 3,400 x 2,200)
<b>Compression</b>	compressed or uncompressed
<b>Bit depth of master</b>	e.g., 24 bit
<b>Colour profiles</b>	e.g., Apple RGB embedded
<b>Greyscale patch</b>	e.g., Kodak Q13 greyscale

### Useful Elements

<b>Type</b>	Type of resource, e.g., “image”, “text”
<b>Extent</b>	Size, quantity, duration e.g.: “1 vol., 200 pages”
<b>Medium</b>	Example, “photograph”
<b>Language</b>	Example, “EN” , “CY” ,”FR”
<b>Relationship</b>	Example, “is part of collection ....”
<b>Location alt.</b>	Bilingual place name variants
<b>Publisher</b>	usually repository name
<b>GIS Reference</b>	Eastings, Northings of place most significantly associated with the image
<b>OS NGR</b>	OS National Grid Reference of place most significantly associated with the image
<b>Credit Line</b>	Where additional credit line is required for a record. Defaults to repository name

### References

1. **Gathering the Jewels**, <<http://www.gsj.org.uk/>>

### Contact Details

Allison Coleman, Gathering the Jewels Ltd, National Library of Wales, Aberystwyth, Ceredigion, SY23 3BU

---

## 4.5 Software

---

### Background

This section addresses the area of Software.

The briefing documents seek to describe best practices in this area.

### Briefing Documents

The following briefing documents which address the area of Software have been produced:

- *Software Code Development*, (briefing-13)
- *Making Software Changes to a Web Site*, (briefing-19)

### Case Studies

The following case studies which address the area of Software Development have been produced:

- *Error Detection on the UKOLN Web Site*, (case-study-14)

# Software Code Development

Briefing

### About This Document

This briefing document provides high-level advice for software developers.

### Citation Details

*Software Code Development*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-13/>>

**Keywords:** software, software development, briefing

## About

This document gives high-level advice for people who develop software for use either internally within a project or for use externally as a project deliverable.

## Background

Each computer programming language has its own coding conventions. However there are a number of general points that you can follow to ensure that your code is well organised and can be easily understood by others. These guidelines are not in any way mandatory but attempt to formalise code so that reading, reusing and maintaining code is easier. Most coding standards are arbitrary but adopting some level of consistency will help create better software.

The key point to remember is that good QA practice involves deciding on and recording a number of factors with your programming team before the onset of your project. Having such a record will allow you to be consistent.

## Documentation

In order for programmers to use your software it is important that you include clear documentation. This documentation will take the form of internal and external documentation.

- **External documentation:** should explain how the software will be used. Internal documentation explains how to implement the code.
- **Comments:** Comments should be added to your code to explain implementation details of the source code. Avoid adding obvious or lengthy information. Prior to your project you should agree on how frequent comments should be and their location, format and length in the file. These conventions may need to be agreed on for block, single-line and end-of-line comments.
- **Readme file:** Every software package should contain a readme file describing and the purpose and functionality of the software and information on external dependencies.

## Naming Conventions

Naming conventions of files, procedures and variables etc. should be sensible and meaningful and agreed on before the projects starts. Use of capitalisation may vary in different programming languages but it is sensible to avoid names that differ only in case or look very similar. Also avoid names that conflict with standard library names.

### Code

There are a number of key points to remember when writing your code:

- **Linearity:** If using a procedural language ensure your code is linear and starts at the first executable statement and continues to a final return or end of block statement.
- **If constructs:** Avoid complicated "if" constructs. It is better to use several simpler nested "if" constructs than a complicated compound one. Keep it simple.
- **Layout:** Code layout is very important. It should be formatted to provide visual clues to the flow of the implementation. It is useful to agree on factors such as indentation, location of brackets, use of white space and line spacing used before the project starts. For example how long will your lines be? Will you use tabs or spaces?
- **External Constants:** Define constant values outside of the code as this makes maintenance easier. Changing hard-coded constants can be time-consuming and prone to human error.
- **Error Handling:** It is also important that you write in some form of error handling into the code.
- **Portability:** Portable code allows the source file to be compiled with any compiler and executed on any machines and operating system. Creating portable code is fairly complex. It is useful to keep machine dependent and machine independent code in separate files.

### Standards

Standards are “documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose” (ISO 1997). The aim of international standards is to encapsulate most appropriate current practice. The International Organization for Standardization (ISO) is the head of all national standardisation bodies. The most relevant ISO standard for software code development is ISO 9000-3: 1997 (QA for the development, supply, installation and maintenance of computer software). For other relevant standards also check the Institute of Electrical and Electronics Engineers and the American National Standards Institute.

### Project QA

At the start of development it may help to ask your team the following questions:

- Do you have local guidelines for writing code?
- Are your software staff aware of the conventions to be used?
- Do you have procedures in place for use when creating local software?

### References

1. **International Organization for Standardization (ISO)**, <<http://www.iso.ch/>>
2. **Institute of Electrical and Electronics Engineers (IEEE)**, <<http://www.ieee.org/>>
3. **American National Standards Institute (ANSI)**, <<http://www.ansi.org/>>

Further information on software QA is available from Sticky Minds,  
<<http://www.stickyminds.com/>>

# Creating and Testing Web Forms

Briefing

### About This Document

This briefing document provides advice on the design of Web forms.

### Citation Details

*Creating and Testing Web Forms*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-14/>>

**Keywords:** software, software testing, forms, briefing

## Background

A Web form is not dissimilar in appearance and use to a paper form. It appears on a Web page and contains special elements called controls along with normal text and mark up. These controls can take the form of checkboxes, text boxes, radio buttons, menus, etc. Users generally fill in the form by entering text and selecting menu items and then submit the form for processing. The agent could be an email or Web server.

Web forms have a variety of uses and are a way to get specific pieces of information from the user. Web sites with forms have their own specific set of problems and need vigorous testing to ensure that they work.

## Designing Forms

Some of the key things to consider when designing your form are:

### Mandatory Fields

Making fields compulsory can cause problems. Occasionally a user may feel that the question you have asked is inappropriate in context or they just can't provide the information. You need to decide if the information needed is absolutely necessary. Users will be more likely to give information if you explain why the data that you're asking for is needed. It is acceptable to ask the user if they meant to leave out a piece of information and then accept their answer.

Validation of forms can be carried out either client side or by processing on the server. Client side validation requires the use of a scripting language like JavaScript and can be problematic if the user's browser disallows scripting. However server side validation can be more complicated to set up.

### Drop Down Lists

Sometimes the categories you offer in a drop down list do not match the answer that the user wants to give you. Sites from the USA often ask for states, which UK users cannot provide. If you want to use a drop down list make sure that your error messages are helpful rather than negative and allow users to select an 'other' option. If you have given a good selection of categories then you should rarely get users picking this.

Also consider if the categories you have provided are subjective enough. There may be issues over the terms used to refer to particular countries (for example if a land area is disputed. If you have to provide a long drop down list then it might be worth offering the common categories first. You could also try sub dividing the categories into two-drop downs where the selection from the first dynamically creates the options in the second.

### Separate Display

You may wish to have the user see a new page or sidebar when filling in a form. A new page may be easier to look at but can be annoying if it is perceived as a diversion or, even worse, an advertisement. It may also be prevented from opening by window blocking software available on newer browsers.

### User Errors

Users will often make typing or transcription errors when filling a form in. These errors can occur in any free text fields on the form.

Occasionally users will press the Submit or Send button either deliberately or inadvertently when only part-way through the form. Make sure that you have an appropriate error message for this and allow users to go back to the unfinished form. Users also often fill in part of a form and then click on the back button. They may be doing this to lose the data they have filled in, to check previous data or because they think they have finished. These activities suggest poor user interface design.

It is important to provide a helpful message on the submission screen explaining the form has been submitted successfully. You could also give replicate the details inputted for users to print out as hard copy.

## Testing Forms

Once you have created your Web form you need to test it thoroughly before release. There are a number of different free software products available that will help you with your testing. Tools such as **Roboform** [1] are freely available and can be used to store test data in and automatically fill in your forms with data.

When testing your form it is worth bearing in mind some problem areas

- **Character sets:** If you require users to fill in their names you will have to be ready to deal with different character sets. Creating different characters to test with can be problematic but services such as **BabelMap** [2] can help with this.
- **Checking Scripts:** Be sure to check you common gateway interface (CGI), server-side scripts and client-side scripts by submitting and resetting form data.
- **Tab Order:** Often when creating a form information is moved about. That is why it is important that you check the tab order of your form. Tab order is especially important for people using screen readers.

## References

1. **Roboform**, <<http://www.roboform.com/>>
2. **BabelMap**,  
<<http://uk.geocities.com/BabelStone1357/Software/BabelMap.html>>

# Making Software Changes to a Web Site

Briefing

### About This Document

This briefing document describes the approaches you should take when you need to make software changes to a Web site.

### Citation Details

*Making Software Changes to a Web Site*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-19/>>

**Keywords:** software, software changes, Web, briefing

## Background

It is desirable to minimise the time a Web site is unavailable. However it may be necessary to bring a Web server down in order to carry out essential maintenance. This document lists some areas to consider if you wish to minimum down time.

## Planning

The key to any form of critical path situation is planning. Planning involves being sure about what needs to be done and being clear about how it can be done. Quality Assurance is vital at this stage and final ‘quality’ checking is often the last act before a site goes live or a new release is launched.

### Prior to Down Time:

- Collect statistics on your site and find out what time and on what day it has the least number of visitors and activity. If it is possible (staff allowing) choose this time to take the site offline
- Before bringing the site down hold a meeting. At this draw up a schedule of what needs to be done and who will complete each task. Create a checklist of testing to be done.
- If the whole site will be offline post a maintenance page stating that our site will be down temporarily. Make sure this page is created in advance.

### During Down Time:

- Make all modifications to the site such as installing new software, changing databases, etc.
- Review all configuration settings and check that the correct files are in place.
- Check all server services are running. Ensuring services such as secure sockets layer (SSL) are running is vital if you are working in a business environment.
- Check any INI, property files, or other configuration files that may have been changed. A list of configuration file that may change could be created prior to down time.
- Use the checklist to check all other relevant change areas. You should look at:
  - Visual changes – you could add icons to new pages
  - Functionality changes



- Run some general user tests, such as ordering a book, retrieving information from the database, submitting a form. It is worth anticipating in advance some Scenario-specific check areas that can be looked at.

### After Down Time:

- Again run some general user tests. These should be run from inside and outside your firewall and on a variety of PCs, browsers etc.
- Check that all links to third parties are running.
- Keep an eye on how the site runs for the next few days and watch for cracks.
- It is important that all the technical support team are notified of any changes that have been made and a problem reporting system is in place.
- Have some form of software or system installed that can inform you of unexpected down time or errors [1]. A list of the type of tools that can be used is available from the QA Focus database [2].

### Conclusions

Advance preparation is vital if you want to minimise time your site down time and avoid confusion when installing new releases.

### References

1. **Error Detection on the UKOLN Web site**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-14/>>
2. **QA Focus Database - Web Site Testing Tools**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/qa-projects/searchtoolkit.asp/>>

---

# 4.6 General

---

### Background

This section addresses areas not covered elsewhere.

The briefing documents seek to describe best practices in this area.

### Briefing Documents

The following briefing documents which address general areas have been produced:

- *Implementing A Technical Review* (briefing-33)

### Case Studies

The following case studies which address general areas have been produced:

- *Managing a Distributed Development Project: The Subject Portals Project* (case-study-03)
- *Implementing a Communications Infrastructure* (case-study-12)
- *Assessing Learning Materials from JISC-funded Resource Databases* (case-study-24)

Case studies which cover specific technical areas are available within the section on the appropriate technical area.

# Implementing A Technical Review

Briefing

### About This Document

This briefing document describes how planning a technical review for your project can provide an opportunity to re-evaluate your technical approaches.

### Citation Details

*Implementing A Technical Review*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-33/>>

**Keywords:** software, software testing, forms, briefing

## Background

When projects submit an initial proposal the project partners will probably have an idea as to the approaches which will be taken in order to provide the project deliverables. During the project's life it may be desirable to review the approaches which were initially envisaged and, if necessary to make changes. This document describes possible approaches to periodic reviews.

## Reasons For A Review

There are a number of reasons why a technical review may be necessary:

**Technological issues:** There may be changes with underlying technologies. For example the software which was initially envisaged being used may be found to be inappropriate or alternative software may be felt to provide advantages.

**Staffing issues:** There may be staffing changes. For example key technical staff may leave and are difficult to replace.

**Organisational issues:** There may be changes within the organisation which is providing the project.

**Changing requirements:** There may be changes in the requirements for the project, following, say, a user needs requirements survey.

**Ensure that deliverables comply with standards and best practices:** It may be necessary to ensure that the project has implemented quality assurance processes to ensure that project deliverables comply with appropriate standards and best practices.

A project review may, of course, also address non-technical issues.

## Approaches To A Review

Projects may find it useful to allocate some time during the project life span to a technical review of the project.

**Review by development team:** The project development team may wish to reflect on the approaches they have taken. They may be encouraged to provide a report to the project manager.

**Review by project partners:** The project partners may be involved in the review process.

**Review involving third parties:** The project team may wish to invite external bodies to participate in the review.

**Comparison with one's peers:** You may chose to compare your deliverables with your peers, such as similar projects. This approach is particular suited for reviewing publicly available deliverables such as Web sites.

When organising a project review you should take care to ensure that the review is handled in a constructive manner.

### Outputs From A Review

It is important to note that any improvements or changes which may have been identified during a view need **not** necessarily be implemented. There may be a temptation to implement *best* practices when good practices are sufficient, and that implementation of best practices may take longer to implement than envisaged. The outputs from a review may be:

**Better understanding:** The review may have an educational role and allow project partners to gain a better understanding of issues.

**Enhanced workflow practices:** Rather than implementing technical changes the review may identify the need for improvements to workflow practices.

**Documenting lessons:** The review may provide an opportunity to document limitations of the existing approach. The documentation could be produced for use by project partners, or could be made more widely available (e.g. as a QA Focus Case Study).

**Deployed in other areas:** The recommendations may be implemented in other areas which the project partners are involved in.

**Implemented within project:** The recommendations may be implemented within the project itself. If this is the case it is important that the change is driven by project needs and not purely on technical grounds. The project manager should normally approve significant changes and other stakeholders may need to be informed.

### Conclusions

It can be useful to allocate time for a mid-project review to ensure that project work is proceeding satisfactorily. This can also provide an opportunity to reassess the project's technical architecture.

## Managing a Distributed Development Project: The Subject Portals Project

Case Study

### About This Document

This case study ...

### Citation Details

*Managing a Distributed Development Project: The Subject Portals Project*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-03/>>

**Keywords:** QA, distributed development, software development SPP, case study

### The Project

The RDN's **Subject Portals Project** (SPP) is funded under the JISC 5/99 Development Programme. There were two proposals, SAD I (Subject Access to the DNER) and SAD II. The original SAD I proposal was part of a closed JISC DNER call, 'Enhancing JISC Services to take part in the DNER'. The SAD II proposal was successful under the JISC 5/99 call, 'Enhancing the DNER for Teaching and Learning'. The original project proposals are available [1].



The aim of the project is to improve the functionality of five of the RDN hub sites to develop them into subject portals. Subject portals are filters of Web content that present end users with a tailored view of the Web within a particular subject area. In order to design software tools that simultaneously satisfy the needs of a variety of different sites and make it easier for institutional portals to embed our services in the future, we are designing a series of Web "portlets". One portlet will be built for each of the key portal functions required, focussing initially on authorisation and authentication (account management); cross-searching; and user profiling; but including eventually a range of "additional services" such as news feeds, jobs information, and details of courses and conferences. The project is committed to using open source software wherever possible.

The hub sites involved in the SPP are EEVL (based at Heriot Watt University, Edinburgh), SOSIG (University of Bristol), HUMBUL (University of Oxford); BIOME (University of Nottingham) and PSIGate (University of Manchester). The project is managed from UKOLN based at the University of Bath, and the technical development is led from ILRT at the University of Bristol.

### Distributed Development: The Problems

The fact that the SPP partners are geographically dispersed has posed a number of challenges. Since the objective of the SPP is the enhancement of the existing hub sites, hub representatives have naturally wished to be closely involved, both on the technical and on the content management sides of the project. At the last count, 38 people are involved in the project, devoting to it varying percentages of their time. But this means that physical meetings are difficult to organise and costly: since work began in December 1999 on the SAD II project, only two full project meetings have been held, with another planned for the beginning of 2003. Smaller physical meetings have been held by the technical developers at ILRT and the five hubs, but these again are extremely time-consuming.

We also faced the problem that many of the project partners had never worked together before. Not only was this a challenge on a social level, it was also likely to prove difficult to find where the skills and experience (and software preferences) of the developers overlapped, and at the beginning of 2002, the then project manager Julie Stuckes commissioned a skills audit to discover the range and extent of these skills and where the disparities lay. It was also likely to be hard to keep track at the project centre of the different development activities taking place in order to produce a single product, and to reduce the risk of duplicating effort, or worse, producing incompatible work. We also thought moreover that it was desirable to develop a method of describing the technical work involved in the project in a way easily understood by the content managers and non-technical people outside of the project.

### The Solutions

We tackled the problem of communication across the project by the use of a project JISCmail mailing list [2]. The list is archived on the private version of the SPP Web site [3] where other internal documents are also posted.

The developers have their own list (`spp-dev@dev.portal.ac.uk`) and their own private Web site [4] which is stored in a versioning system (CVS - Concurrent Version System [5]) which gives any authenticated user the ability to update the site remotely.

In addition the developers hold weekly live chat meetings using IRC (Internet Relay Chat [6]) software (as shown in Figure 1), the transcripts of which are logged and archived on the developers' Web site.

Using IRC means the developers are able to keep each other informed of their activities in a relaxed and informal manner; this has aided closer working relationships.

As well as holding the developers' Web site, CVS also contains the project's source code and build environment.

This takes the form of a central repository into and out of which developers check code remotely, ensuring that their local development environments are kept in step. A Web interface also provides the option of browsing the code, as well as reviewing change histories. Automatic e-mail notification alerts the developers to updates checked into the CVS repository, and all changes are also logged. This has proved an essential tool when co-ordinating distributed code development.

The other part of the software development infrastructure is providing a build environment that takes care of standard tasks, allowing the team members to concentrate on their coding. Using a combination of open-source tools (e.g. **ant** [7] and **junit** [8]) a system has been created that allows the developer to build their code automatically, run tests against it, and then configure and deploy it into their test server. As well as this, the build system will also check for new versions of third-party

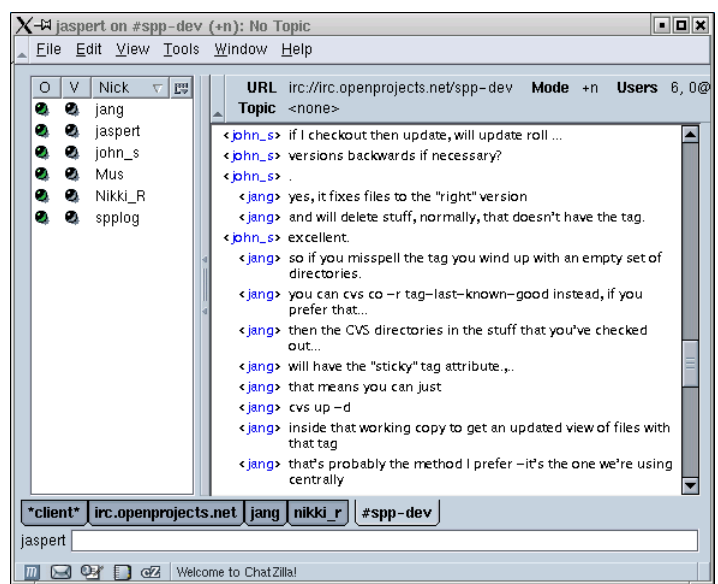
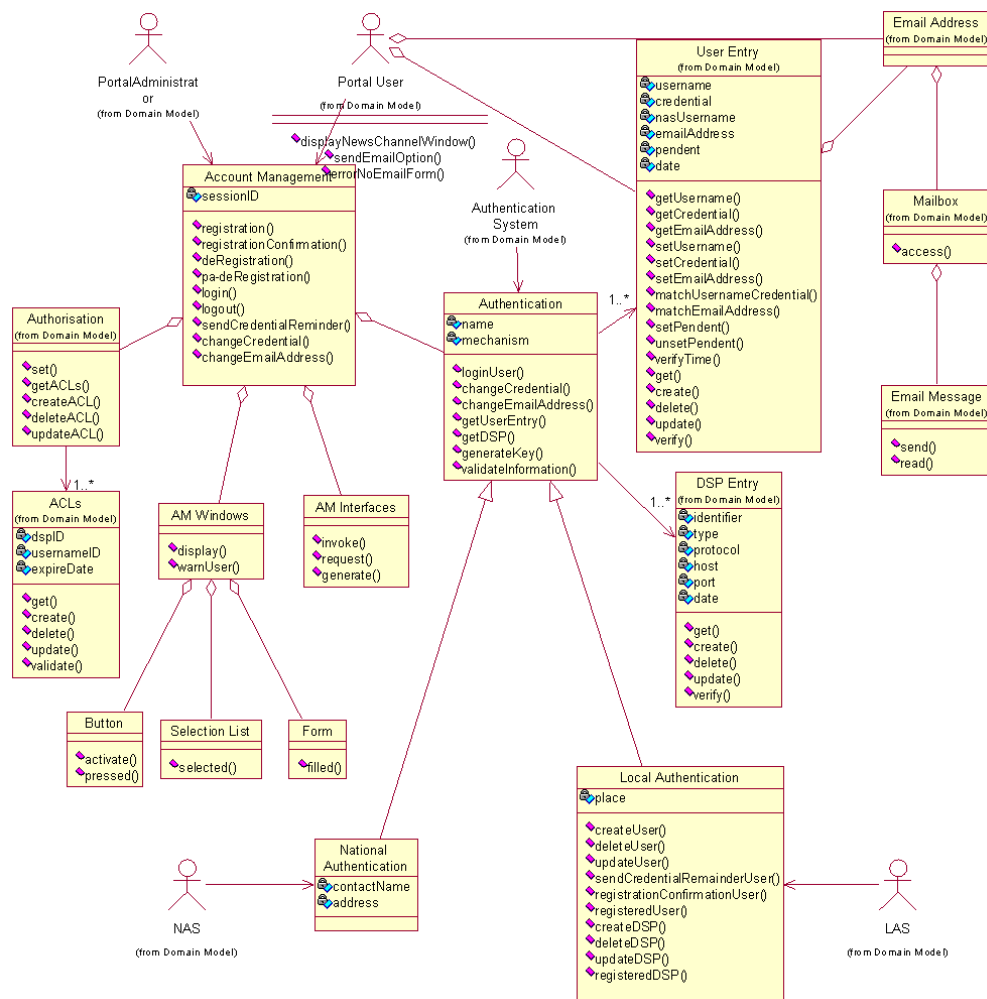


Figure 1: Example IRC Session

packages used by the project, updating them automatically if necessary. This system is also managed by current project down from the central repository, build, configure and deploy it, having it running in a matter of minutes.

Because of the widely dispersed team, the difference in software preferences and the mixed technical ability across the project, we looked around for a design process that would best record and standardise our requirements. UML (Unified Modelling Language [9]) is now a widely accepted standard for object oriented modelling, and we chose it because we felt it produced a design that is clear and precise, so making it easy to understand for technical and non-technical minds alike. UML gave us a means to visualise and integrate use cases, integration diagrams and class models. Moreover using UML modelling tools, it was possible to generate code from the model or update the model whenever the code was further developed.



### Figure 2: Example UML Diagram

Finding UML software that had all the features needed was a problem: there are plenty of products available but none quite met all our requirements, especially when it came to synchronising the work being done by different authors. Eventually we opted to use the **ICONIX** process [10]. This is a simplified approach to UML modelling, which uses a core subset of diagrams. This enabled us to move from use cases to code quickly and efficiently using a minimum number of steps, thus giving the technical side of the project a manageable coding cycle.

Additional funding was obtained from the JISC in order to bring one of the authors of the ICONIX process (Doug Rosenberg) over from California to run a three day UML training course. Although this course was specially designed for SPP, places were offered to other 5/99 projects in order to promote wider use of this methodology across the JISC community. Unfortunately, despite early interest, no other project was represented at the training, although Andy Powell, the technical co-ordinator for the RDN, attended the course. Additional funding was also received from the JISC to purchase licences for **Rational Rose** [11], which we had identified as the most effective software available to produce the design diagrams

Finally, to provide greater structure to the project, a timetable of activities produced using MS Project is posted on the private project Web site and is kept continually up to date. A message is posted to the project mailing list to alert partners of any major changes to the timetable.

### What We Would Have Done Differently

It would have been sensible for us to have adopted a process for software development at an earlier stage in the project: it was perhaps a need that we could have anticipated during the SAD I project phase. Also, it is worth noting from our experiences that getting the communications and technical support infrastructure in place is a job in itself, and should be built into the initial planning stage of any large and dispersed project.

### Continuing Problems

Electronic communication is still no substitute for face-to-face meetings so the SPP development team continue to try to meet as regularly as possible. Time is inevitably a major problem wherever project partners have other work commitments: all the project partners based at the hub sites have to juggle SPP work which is for the project as a whole, with that which relates particularly to their own hub's adoption of the project's outcomes. Increasingly, as the project develops, less work will be required from the project "centre" and more at the hubs, leading to an eventual handover of the subject portal developments to the hubs for future management.

### The Future

It is our plan to make use of UML diagrams in the final project documentation to describe the design and development process. They will offer a detailed explanation of our decision making throughout the project and will give future projects an insight into our methodology. Andy Powell was also so impressed with UML that he is planning to use it across development work for the RDN in the future.

The future development of the SPP beyond the end of the project is likely to be led by the technical development partners, for instance in the continued development of the portlets to enable them to be installed into alternative open source software platforms to make the technology as compatible with existing systems as possible. It is therefore greatly to the benefit of the project that they have become such an effective and close working team.

### Contacts:

Ruth Martin  
Former SPP Project Manager, based at UKOLN



Jasper Tredgold, SPP Technical Co-ordinator  
ILRT  
University of Bristol  
10 Berkeley Square  
Bristol  
BS8 1HH  
Email: [jasper.tredgold@bris.ac.uk](mailto:jasper.tredgold@bris.ac.uk)

### References:

1. **Subject Portals Project**, RDN,  
<<http://www.portal.ac.uk/spp/documents/>>
2. **Archives of SPP Mailing List**, JISCMAIL,  
<<http://www.jiscmail.ac.uk/archives/spp.html>>
3. **RDN portal development project**, RDN,  
<<http://www.portal.ac.uk/spp/private/>> (Note username and password required)
4. **dev.portal.ac.uk**,  
<<http://dev.portal.ac.uk/>>
5. **CVS - Concurrent Versions System**,  
<<http://www.gnu.org/software/cvs/>>
6. **Internet Relay Chat (IRC)**,  
<<http://www.irchelp.org/>>
7. **ant**, Apache<<http://jakarta.apache.org/ant/>>
8. **junit - Testing Resources for Extreme Programming**,  
<<http://www.junit.org/>>
9. **UML**,  
<<http://www.uml.org/>>
10. **ICONIX**,  
<<http://www.iconixsw.com/>>
11. **Rational Rose**,  
<<http://www.rational.com/products/rose/>>

# Implementing a Communications Infrastructure

Case Study

## About This Document

This case study ...

## Citation Details

*Implementing a Communications Infrastructure*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-12/>>

**Keywords:** QA, QA Focus, communications, case study

## Background

QA Focus is a distributed project, with team members based in UKOLN (University of Bath) and the AHDS (Kings College, London). There is a need to provide effective communications for the team members. This case study describes the communications infrastructure which is employed.

## Problem Being Addressed

The distributed nature of the QA Focus team means that a good communications infrastructure is an essential part of working practice. The communications tools chosen for use need to be both efficient and easy to maintain, as well as being freely available.

## The Approach Taken

The QA Focus communications infrastructure has been built around a number of separate but complimentary tools.

## MyYahoo! Web site

One of the first communication mechanisms established was a shared file space. MyYahoo [1] is a highly customisable, shared repository. The site allows you to choose exactly what sort of information is relevant to you from a variety of different information tools including news, bookmarks, maps, calendar and email. The briefcase area allows online storage of files that can then be accessed from anywhere, either by visiting the site or by clicking on links to items. Any type of file can be stored providing they fall within certain content and size guidelines provided by Yahoo. Using MyYahoo the QA Focus team can manage files from work, home or any other location. Yahoo currently provides 30MB worth of free space.

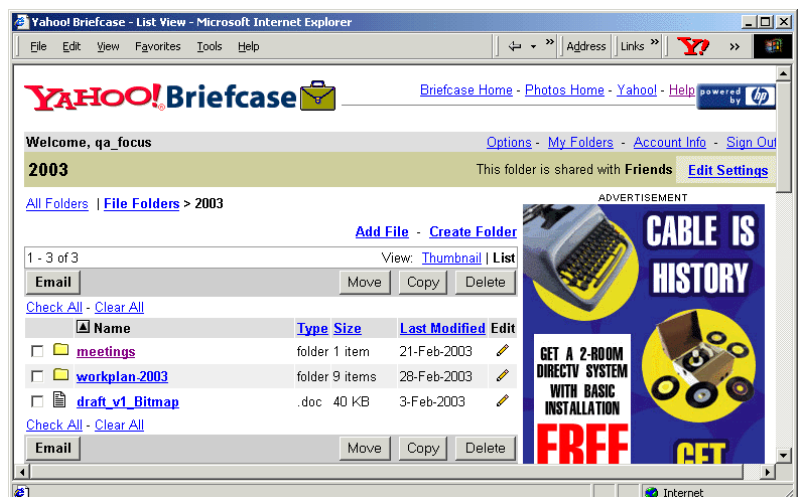


Figure 1: The MyYahoo briefcase

### Yahoogroups Mailing List

A Yahoogroups mailing list [2] called qa-focus-team was also set up for internal QA Focus use. Yahoo encourage the use of user profiles allowing all their communication methods to be linked together. On setting up your own email account or a MyYahoo Web site you create a profile, this profile can then be assigned an email. Setting up the mailing list involved selecting a Yahoo! Groups Category and deciding on a group email address. Members are then enrolled or invited to join the group. Each individual member can then be configured, allowing them to post, receive and/or receive a copy of all messages to their usual daytime email addresses. The list is maintained and customised by the list owner and lists can be set up for public use or private use only. The main advantage of using the list is the creation of a comprehensive archive. This means that all email information is in an open space and not only held in one person's email box, whom may be on holiday or have changed jobs.

### Blog

A Blog is a Web log held on the Internet which is updated frequently. Blogging has taken off in recent years and there is now a variety of free software allowing you to set up your own blog without any programming skills. One of the most famous blogs is blogger.com [3]. In order to record activities, ideas etc a blog was set up by the QA Focus team using Movabletype [4], a decentralised, Web based personal publishing system. The blog is currently only accessible internally and is used as a record of activities carried out during the week. These summaries will help with keeping note of work carried out and compiling reports. It is hoped that at a later date the blog will be open for external viewing.

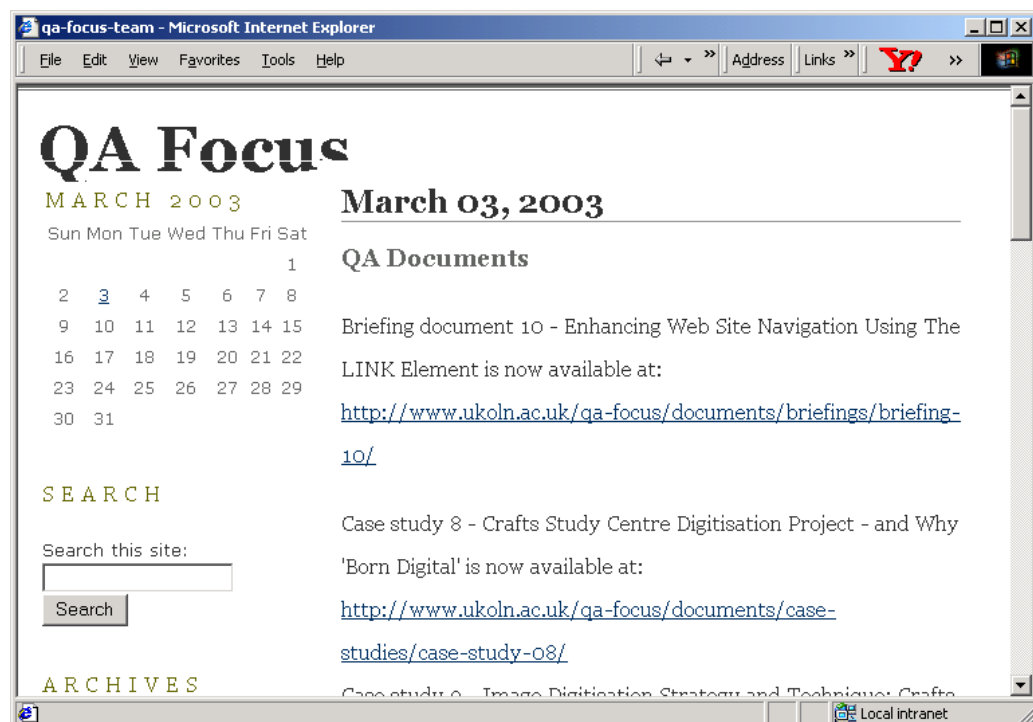


Figure 2: The QA Focus Blog

### Instant Messaging

The QA Focus team have also experimented with forms of instant messaging. These are services that provide users with instantaneous contact with other Internet users. The main advantages of instant messaging are that you can carry out a real-time

conversation while involved with other tasks. There is a much higher level of synchronicity than that achieved with an e-mail conversation, so it is useful for high priority work that needs group input.

Previously the messenger used by the team was MSN Instant Messenger. MSN has some potentially useful features, such as the ability to share desktop applications. However use of MSN waned after a change of a member of staff. It is now felt that using Yahoo IM might integrate better with the other Yahoo tools.

### Problems Experienced

Setting up the various communication tools is fairly straightforward but can be time-consuming. The real problem is getting users or members of a team to actually use the tools. The core QA Focus team only consists of three people so encouraging use has not been that much of an issue, but occasionally you do find yourself slipping back into old ways of working and using solely email.

Having a good communications infrastructure is key when working in a team, especially when members are distributed remotely. The important factor in establishing use is to document procedures and use the tools diligently at the start so use becomes second nature.

### Things We Would Do Differently

The nature of QA Focus means that all experience and experimentation in a Web related area is always useful and gives us knowledge of both the problems and success areas. However if we had to repeat the process then maybe we would spend more time investigating the different tools available and document their advantages and disadvantages. Unfortunately as most people working on a project will know there is never enough time for research as anyone would like.

### References

1. **MyYahoo**, <<http://my.yahoo.com/>>
2. **Yahoogroups mailing list**, <<http://groups.yahoo.com/>>
3. **Blogger.com**, <<http://www.blogger.com/>>
4. **Movabletype**, <<http://www.movabletype.org/>>

# Assessing Learning Materials from JISC-funded Resource Databases

Case Study

### About This Document

This case study ...

### Citation Details

*Assessing Learning Materials from JISC-funded Resource Databases*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-24/>>

**Keywords:** QA, evaluation, case study

## Background

The aim of the **Healthier Nation** [1] project is to identify suitable learning materials from JISC-funded resource databases and/or content repositories and re-purpose a sample of the material as learning chunks to support Health and Social Care curriculum delivery (at FE and HE Levels). The project has specifically focused on 'the Big Four' diseases affecting Scotland: Cancer, Coronary Heart Disease, Stroke and Mental Illness.

## Problem Being Addressed

The first Strand of the project has been concerned with research and evaluation. The objectives of strand 1 have been to identify and evaluate relevant materials on the "big 4" diseases by:

- Carrying out an electronic search of defined and agreed databases and content repositories
- Evaluating the resources for appropriateness to Health Care related programmes at Scottish Credit and Qualifications Framework (SCQF) levels 4 to 7 [2]
- Identifying keywords and signposting information (levels, courses, delivery options) to facilitate the development of metadata
- Highlighting areas where little or no learning materials exists to support learning in the identified subject areas at the appropriate levels.

## The Approach Taken

### Research Methodology

Academic experts from each partner institution prepared a 'mapping grid' to assist the research team to identify relevant resources. The grid included information on key subject areas, specific keywords and exclusions, courses which could use the material (including the level) and any relevant learning outcomes.

As the emphasis for the academic staff was on finding resources that could be used in teaching situations at FE and initial HE levels - possibly with some element of re-purposing - the research team concentrated on the relevant subject gateways, rather than bibliographic sources and indexes.

To provide a structured framework for the evaluation of the learning material, resource sheets were used to record relevant details. The sheets ensured that the evaluation criteria used by all partner institutions were consistent.

The researcher team evaluated all the learning material on their content (clarity, authority, bias, level) and their style and functionality. Copyright details were also recorded for future re-purposing. Restricted vocabularies were used whenever possible to assist metadata tagging of learning objects. Resource sheets were then passed to academic staff to evaluate their appropriateness for teaching and to indicate how they could be used (delivery type, mapped to level/course, any re-purposing that would be required).

The intention was to carry out accessibility evaluations on a selection of the resources during this part of the project. A key issue that has affected this work has been the lack of agreed criteria for accessibility evaluations. One of the project partners, RNC, has been working with TechDis to develop a model for evaluating resources and the accessibility evaluation of the resources will now be carried out at a later stage.

Following evaluation, the materials were:

- Included in a final list of resources to be uploaded to a repository
- Included in a list of supplementary resources that could be used to support teaching
- Rejected as unsuitable

## Problems Experienced

### Findings

The research team had difficulty in retrieving relevant material for the project using the search options in RDN subject gateways. Whenever two subject terms were combined the number of hits was drastically reduced. The search term "Heart" for example, retrieved 312 sites in BIOME; by adding a second term "Heart physiology" this was reduced to 8 sites. Search terminology was often restricted to key areas only, e.g. neoplasm, then the researchers trawled through the numerous hits to find materials – a lengthy process, but the only way to ensure that useful material had not been missed.

Searching under the sub-headings that had been provided by the academic staff produced few or in some cases no hits. A BIOME search for "ACE inhibitors" only retrieved 1 site. To provide enough material for the future strands of the project, Google was also used to locate materials for both mental illness and coronary heart disease/stroke.

On average only one in 10 of the resources located were passed to the academic staff for evaluation. The majority were too advanced, predominantly text based and therefore had no advantages over a textbook, or did not cover the project's subject areas (particularly the sub-headings/keywords).

Over 500 resources were evaluated by academic staff, but only 46% made it to the final repository of resources, or as a supplementary teaching resource. The main reasons for the rejection of the remainder were that the material was:

- Too advanced
- Not applicable/not helpful
- Repeating information found elsewhere
- US biased
- Not able to be accessed/opened.

Academic staff felt that, while some of the resources were excellent (particularly for cancer), in general the resources have not been as good as expected and there were not enough graphic or interactive materials for re-purposing. Mental health resources were geared towards the layperson and had a heavy emphasis on organisations.

Most of the resources went through a secondary evaluation stage to ensure that comments made by FE academic staff were applicable for HE and vice versa. In the secondary evaluation, there was general agreement between the academics in FE and HE about the usefulness of the resources. Although some materials were either too high or too low a level, others were rejected because of their similarity or due to problems of access.

All of the academics involved in the project, felt that they would use alternative sources to locate material. Google was their preferred option as it gave access to relevant material more easily than the subject gateways and has the advantage of advanced search strategies, including searches for images, applying multiple search terms, restricting searches by country of origin.

### Conclusions and Recommendations

1. From a tutor's point of view, finding suitable resources via JISC repositories/subject gateways to support our target level of course may not be an attractive or viable option.
2. More resources are required to meet the needs of FE and the initial HE levels for the subjects covered by this project. In each of the key areas there were only 20 – 47 resources that could be used, with some element of re-purposing, to support teaching. The most common problem for exclusion from the repository was that the materials were either too advanced or not suitable for FE students.
3. In the subject areas of mental illness coronary heart disease and stroke, the JISC resources alone were too few in number to support the other strands of work in the project and had to be supplemented by resources located using other search engines like Google.
4. One of the objectives of our project is to look at ways of encouraging tutors to contribute their own resources to the digital repository which may assist to build a better bank of resources in the longer term. On a practical level, one of our specific deliverables is to prepare guidance on how to load resources into a digital repository.
5. The subject indexing and metadata tagging of resources needs to be more detailed to meet the needs of academic staff and learners. In areas where the staff knew they required images – e.g. the respiratory system – it was impossible to search for images alone. Searching involved going to each resource individually to check their contents.
6. Many resource were rejected due to US bias and because they were not of a suitable quality to support a teaching situation. Improved metadata tagging could also assist with ensuring that tutors only access high quality material that will meet their selection criteria.
7. We expect the search facilities provided by the JORUM repository will give us a lot more flexibility. Evidence from the repositories presented so far indicates that this will be the case.

8. The metadata tagging of resources in subject gateways would have to be amended to enable advanced search techniques to be applied. Some descriptions were more detailed than others, so some consistency is required.
9. We believe that the metadata collected during this strand of our project is already of a much higher quality and will certainly address some of the problems described in the report.
10. The advanced search options in the subject gateways need to be improved as they very limited compared to standard search engines. As any searches for specific sub-headings or combinations of subjects produced very few results, the emphasis was on larger key areas only. With a hit rate of approximately 20 resources per day, this method would be a deterrent to academic staff with some experience of standard search engines.
11. Addressing points 3 and 4 above should significantly reduce the time tutors would have to spend on searching for resources.

### References

1. **Healthier Nation**, <<http://extranet.lauder.ac.uk/x4l>>
2. **Scottish Credit and Qualifications Framework**, <<http://www.scqf.org.uk/>>

### Contact Details

Heather Sanderson  
Project Manager  
Email: [heather@kerson1.freemove.co.uk](mailto:heather@kerson1.freemove.co.uk)



# INHALE And INFORMS Case Study

Case Study

### About This Document

This case study ...

### Citation Details

*INHALE And INFORMS Case Study*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-31/>>

**Keywords:** QA, INFORM, INHALES, case study

## Background

A brief case study on the INHALE project, September 2000- March 2003 and the subsequent INFORMS project, October 2002-August 2003.

## Introduction

The INHALE Project (one of the JISC 5/99 DNER Programme projects) had a number of aims:

- To utilise the ubiquity of the Internet to create a set of learning materials which embed a variety of electronic services within a series of web-based units. These services consist of those available nationally via the DNER and those subscribed to locally.
- To integrate the open world of the DNER with the relatively closed world of the virtual learning environment (VLE).
- To demonstrate how the INHALE materials can equip students with transferable information skills to allow them to select, locate, use and evaluate information for their studies.

The outcomes from the project were:

- A set of standalone web based information skills units, which are accessible to HE and FE institutions in the UK, each of these units being designed to assist users to acquire the necessary skills to find and use quality information sources.
- A preaxial demonstration of how the materials can be successfully embedded within a VLE such as Blackboard or WebCT.
- An iterative product methodology showing how live Web sites can be exploited to maximize interactivity.
- A database of (learning) objects created from the initial standalone materials which would be accessible, searchable and extensible to other HE/FE institutions allowing individuals to recreate a tailored set of materials.
- Reports and journal articles describing the processes and experiences of the work of the project.
- Quantitative and qualitative evaluations of the utilization of the resources by students at the University of Huddersfield and at the project's partner institutions.

At the outset of the project the vision of what was required was set out clearly in the project plan and this was closely adhered to throughout.

During the first year the project team successfully created a standalone set of information skills units for students on Nursing and Health courses. The JISC's technical guidelines on interoperability and accessibility guided the Web developer in the creation of the online resources. The new information skills materials using the JISC's DNER resources as well as freely available web resources were tested within pilot modules in the School of Health Sciences at Huddersfield. Evaluation reports from these were written. These evaluations fed into the continuous "product" development.

During the second year (September 2001- September 2002) additional information skills units were created and some of the initial units were customised. Some of the new resources were based around subscription information databases and were cascaded out for use within the partner institutions, Leeds Metropolitan University and the University of Central Lancashire.

Meanwhile at Huddersfield the resources were being embedded at different levels within Blackboard in new pilot modules.

Running parallel to the delivery of the resources within modules was the continuing development of the INHALE "database". The "database" was seen as the key to enable customisation of the initial set of INHALE materials and the generation of new units by all the partner institutions. This required the disaggregation of all the original materials into objects. Fortunately, from the outset, the vision was that the end result would be a database of learning "objects" and all the materials were created with this concept to the fore. Thus the disassembly was not as onerous a task as it may seem.

Dissemination of the project's learning and outcomes began early in the project and had two key strands. The first was to involve stakeholders in the delivery of information skills within the institutions. Workshops and meetings were held internally and attended by academic teaching staff, librarians, learning technologists, computing service staff and learning and teaching advisors. The second strand was to disseminate to the wider UK HE and FE community and various events were held beginning with an event that was to be repeated "E-Resources for E-Courses".

By July 2002 interest in the use of the INHALE resources had grown. In September 2002, the submission of a proposal for a project within the DiVLE Programme to continue the work of the INHALE project was successful. The new project was named INFORMS and from October 2002 to March 2003, the INHALE project and INFORMS projects ran concurrently. During this time the University of Loughborough and the University of Oxford, (the new INFORMS project partners), were able to test the transferability and viability of all the INHALE project materials and models as well as inputting new ideas for developing the resources.

By the end of the INHALE project in March 2002 there were over 200 units within the new database and a number of new institutions were also testing and using the database.

At this point an exit strategy was written for the INHALE project. The project team felt that there was a possible "market" for the INHALE/INFORMS information skills database within the HE/FE community. However the JISC Programme Managers considered that the database of units required more "market testing" within the HE/FE community. To some extent the INFORMS project has allowed the team to begin the process of market testing.

The INFORMS project officially completed in August 2003 and there are now over 400 units in the INFORMS database and 17 institutions have portfolios of units across the range of subjects studies across the HE/FE community. Usage of the resources can be tracked via a web log analysis tool developed in house that is linked to the database.

Librarians (and some academic teaching staff) institutions are creating their own online, interactive innovative information skills teaching and learning resources without any knowledge of web authoring. The database allows instant editing and updating, it automatically produces accessible and printable versions of the units. The 400 plus units in the database are shared across all the participating institutions. Units copied across institutions are tracked via an audit trail. A User Guide, Editorial Policy and Conditions of Use Agreement are all essential documents that have been produced to support users of the database.

### Problems Encountered

#### Project Creep

There was some initial hold-up in getting the project started and by the time the Project Co-ordinator joined the team in January 2001, the project was approximately 2-3 months behind in writing the initial units, rolling out the baseline evaluation and writing the evaluation instruments. Delivery to the students in the first pilot module was set for mid-February 2001 and this deadline was met.

**Action Taken:** The Project Co-ordinator was employed full time instead of part-time. This was funded from both the unspent salary from the project for 3 months prior to January 2001 and topped up from the salary of the Project Co-ordinator's substantive post.

#### Loss of a Project Partner

Manchester Metropolitan University pulled out of the project in June 2001 when they were successful with obtaining funding from the JISC for the BIG BLUE Project.

**Action Taken:** None, but subsequent events mitigated against the loss of the partner.

#### Loss of Project Staff

The Project Director left to take up another post at the University of Central Lancashire in September 2001. The loss of the Director's role as the stakeholder for the project within the Library Management Team and amongst the Academic Librarians had a detrimental effect on the uptake of the resources across the institution that is still being addressed.

The loss of someone else with information skills expertise to bounce off ideas and to provide another point of view on the project's development, as well as mutual support, has been a problem for the subsequent Project Director.

However the move by the Project Director to the University of Central Lancashire was beneficial as UCLan was invited to take the place of Manchester Metropolitan Library on the project and the input from that institution was invaluable.

**Action Taken:** The Project Co-ordinator took over the role of Director.

In November 2001, two months after the demise of the Project Director the project's Web developer was recruited to an internal position in the library. This could have

proved disastrous but in fact a new web developer was recruited from the interviewees for the internal post and began work on the project only 10 days after the original developer had moved.

**Action Taken:** Recruited new Web developer from suitable candidates already interviewed for another post.

### **Territorialism**

The new Project Director encountered internal political problems that have constantly hampered the uptake of the resources.

**Action Taken:** Sought the support of line manager.

### **Local Area Network Problems**

In September 2001 the University of Huddersfield experienced a severe problem with the load on its network. The project was unable to continue development on integrating video and audio into the resources.

**Action Taken:** Abandoned using video and audio.

### **Changes in University Infrastructure**

The problem with the LAN traffic had a knock-on effect. The central service managing the Blackboard resources plus the learning and teaching support for this was re-organised. Key stakeholders in this support area within the University left so the necessary key personnel to champion the uptake of the INHALE resources in Blackboard were lost. Eventually some new posts have been created.

**Action Taken:** Contact made with new staff and process of rolling-out started again.

### **Running Two Projects Concurrently**

The Project Director misjudged the demands that running the two projects (INHALE & INFORMS) during the period October 2002 to March 2003 would make.

**Action Taken:** None. With hindsight it might have been more feasible to have a different financial balance between the three partners to have allowed additional staff to be recruited at Huddersfield.

## **Future Developments**

The JISC only require projects to make their web sites available to the rest of HE/FE for 3 years after the end of the project. Thus if a resource has a potential for further uptake and development then the project will need to produce a strategy to enable this.

The University of Huddersfield is not in a position to fund user support for the database. The institution is still in the early days of recovery after its re-organisation of the technical and teaching and learning support infrastructure for Blackboard.

The INFORMS (INHALE) project team have been pursuing a number of possible strategies:

### **Commercialisation**

The INFORMS Project team think that there is a commercial potential for the INFORMS software beyond the HE/FE sector and have been successful in a bid for

funding to investigate and pursue this further over the next 12 months via a University of Huddersfield Commercial Fellowship.

It is planned that any profit will eventually be used to provide support for the INFORMS database. (Staffed support for HE/FE users of the INFORMS database, support of the web server hosting the database, support to implement new developments).

### **Collaboration with other JISC Projects**

The location of the INFORMS resources within an Information Skills Portal alongside the VTS, Big Blue & the Resource Guides etc. would be an ideal scenario and one that has been suggested already by the Big Blue project.

### **Collaboration with other Key Players**

Both the Open University and Sheffield Hallam have products that may benefit from the technical developments of the INHALE/INFORMS projects.

### **Hosting by JISC Services**

If demand for portfolios in the database grows then the capacity of the web server at Huddersfield will be over-reached. So one possible strategy could be to move the database to either EDINA or MIMAS.

Mirroring the database at Edina has been explored and this may be possible in 12 months if the return from the commercialisation of the software is sufficient.

### **Providing Individual Institutions with the Software**

It may be possible to give away the software to HE/FE institutions to run on their own servers and develop should they wish to do so. The main disadvantage of this is the loss of the shared resources.

### **Additional Support from the JISC**

A case has been put to the JISC, the reply has been that additional evidence of a need must be gathered through “market testing”.

## **Conclusion**

For the time being the new INFORMS (Commercial) Project is the route being taken by the ex Project Director of INHALE/INFORMS to create supportive funding in the long term for the INFORMS database of information skills teaching and learning resources. The new INFORMS (Commercial) project began officially on 1<sup>st</sup> October 2003 and will run for 12 months. One of its first successes has been to secure a place at a reception in the House of Commons being held by the Set for Britain group who are promoting start-up, spin-out, spin-off commercialisation of UK University research. At the reception we will be delivering a poster presentation for the MP's, Peers and various other attendees of the proposed commercialisation of the INHALE/INFORMS software.

---

## 4.7 Quality Assurance

---

### Background

This section addresses the area of quality assurance.

The briefing documents seek to describe best practices in this area.

### Briefing Documents

The following briefing documents which address the area of quality assurance have been produced:

- *Summary of the QA Focus Methodology*, (briefing-30)
- *Top 10 QA Tips*, (briefing-40)

Advisory documents which cover specific technical areas are available within the section on the appropriate technical area.

# Summary of the QA Focus Methodology

Briefing

### About This Document

This briefing document gives a summary of the quality assurance methodology which QA Focus has developed for JISC's digital library programmes.

### Citation Details

*Summary of the QA Focus Methodology*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-30/>>

**Keywords:** QA, quality assurance, briefing

## Background

In order to provide value for money and a return on investment from the funders there is a need for project deliverables not only to be functional in their own right but also to be widely accessible, easily repurposed and deployed in a service environment.

To achieve these aims projects should ensure that their deliverables comply with appropriate standards and best practices. Although it may be easy to require compliance, it may not always be easy to implement appropriate standards and best practices. In order to ensure that best endeavours are made it is recommended that projects should implement quality assurance (QA) procedures.

## QA Focus's Methodology

Projects may be concerned that implementation of QA procedures can be time-consuming. The approach recommended by QA Focus is designed to be lightweight and to avoid unnecessary bureaucracy, while still providing a mechanism for implementation of best practices.

The QA Focus methodology is based on the following:

- **Documented policies on standards and best practices:** If standards and best practices are not documented it will be difficult to ensure best practices are implemented, especially in light of staff turnover, changing environments, etc.
- **Documentation of the architecture used:** To ensure that the architecture used to implement the system is capable of complying with the standards.
- **Documented exceptions:** There may be occasions when deviations from standards may be allowed. Such deviations should be documented and responsibility for this agreed.
- **Systematic checking:** It is necessary to document systematic procedures for ensuring compliance with standards.
- **Audit trails:** It can be helpful to provide audit trails which can help spotting trends.

It is felt that use of this methodology should not only be beneficial to the projects themselves, but also help to minimise problems when project deliverables are re-used.

### Example: QA For Web Sites

As an example of implementation of this approach the QA policy for standards for the QA Focus Web site is given below.

**Area:** Web site: standards

**Standards:** The Web site will be based on the XHTML 1.0 and CSS 2.0 standards.

**Architecture:** The Web site will make use of PHP. XHTML 1.0 templates will be provided for use by authors, who will use simple HTML tools such as HTML-kit. Web site will provide access to an MS Access database. This will also comply with XHTML 1.0 and CSS 2.0 standards. The Web site will also host MS Word and MS PowerPoint files. These documents will also be available in HTML.

**Exceptions:** Resources converted from proprietary formats (such as MS Word and PowerPoint) need not necessarily comply with XHTML and CSS standards if doing so would be too time-consuming.

**Change Control:** The QA Focus project manager is responsible for changing this policy and addressing serious deviations from the policy.

**Checking:** When resources are created or updated the resource should be validated, usually using the `, validate` tool. When several resources are updated the `, rvalidate` tool should be used.

**Audit trail:** A full audit should be carried out at least quarterly. The findings should be published on the QA Focus Web site, and deviations from the policy documented.

A second example describes the QA policy for link checking of the QA Focus Web site.

**Area:** Web site: link checking

**Best Practice:** There should be no internal broken links. We should seek to fix broken links to external resources.

**Exceptions:** There may be broken links in historical documents or surveys. If remote Web sites are updated it may be too time-consuming to update links.

**Change Control:** The QA Focus project manager is responsible for changing this policy and addressing serious deviations from the policy.

**Checking:** When resources are created or updated the resource should be link-checked, usually using the `, checklink` tool. When several resources are updated the `, rchecklink` tool should be used.

**Audit trail:** A full audit should be carried out at least quarterly. Initially two tools should be used to spot deficiencies in the link-checking software. The findings should be published on the QA Focus Web site, and deviations from the policy documented.

These two examples illustrate that *developing* QA policies need not be time-consuming. In addition *implementation* of these policies need not be time-consuming and can improve the quality of the Web site.



# Top 10 Quality Assurance Tips

Briefing

### About This Document

This briefing document gives QA Focus's top 10 tips for deploying quality assurance procedures.

### Citation Details

*Top 10 Quality Assurance Tips*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-37/>>

**Keywords:** QA, quality assurance, briefing

## The Top 10 Tips

### 1 Document Your Policies

You should ensure that you document policies for your project – remember that it can be difficult to implement quality if there isn't a shared understanding across your project of what you are seeking to achieve. For example, see the QA Focus policies on Web standards and link checking [1] [2].

### 2 Ensure Your Technical Infrastructure Is Capable Of Implementing Your Policies

You should ensure that your technical infrastructure which is capable of implementing your policies. For example, if you wish to make use of XHTML on your Web site you are unlikely to be able to achieve this if you are using Microsoft Word as your authoring tool.

### 3 Ensure That You Have The Resources Necessary To Implement Your Policies

You should ensure that you have the resources needed to implement your policies. This can include technical expertise, investment in software and hardware, investment in training and staff development, etc.

### 4 Implement Systematic Checking Procedures To Ensure Your Policies Are Being Implemented

Without systematic checking procedures there is a danger that your policies are not implemented in practice. For example, see the QA Focus checking procedures for Web standards and link checking [3] [4].

### 5 Keep Audit Trails

You should seek to provide audit trails which provide a record of results of your checking procedures. This can help to spot trends which may indicate failures in your procedures (for example, a sudden growth in the numbers of non-compliant HTML resources may be due to deployment of a new authoring tool, or a lack of adequate training for new members of the project team).

### 6 Learn From Others

Rather than seeking to develop quality assurance policies and procedures from scratch you should seek to learn from others. You may find that the QA Focus case studies [5] provide useful advice which you can learn from.

### 7 Share Your Experiences

If you are in the position of having deployed effective quality assurance procedures it can be helpful for the wider community if you share your approaches. For example, consider writing a QA Focus case study [6].

**8 Seek ‘Fitness For Purpose’ – Not Perfection**

You should seek to implement ‘fitness for purpose’ which is based on the levels of funding available and the expertise and resources you have available. Note that perfection is not necessarily a useful goal to aim for – indeed, there is a danger that ‘seeking the best may drive out the good’.

**9 Remember That QA Is For *You* To Implement**

Although the QA Focus Web site provides a wide range of resources which can help you to ensure that your project deliverables are interoperable and widely accessible you should remember that you will need to implement quality assurance within your project.

**10 Seek To Deploy QA Procedures More Extensively**

Rather than seeking to implement quality assurance across your project, it can be beneficial if quality assurance is implemented at a higher level, such as within your department or organisation. If you have an interest in more widespread deployment of quality assurance, you should read about the ISO 9000 QA standards [7].

## References

- 1 **Policy on Web Standards**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/qa/policies/web/>>
- 2 **Policy on Linking**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/qa/policies/links/>>
- 3 **Procedure for Web Standards**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/qa/procedures/web/>>
- 4 **Procedures for Linking**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/qa/procedures/links/>>
- 5 **Case Studies**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/>>
- 6 **Contributing To Case Studies**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/#contributing>>
- 7 **Selection and Use of the ISO 9000:2000 family of standards**, ISO,  
<[http://www.iso.ch/iso/en/iso9000-14000/iso9000/selection\\_use/selection\\_use.html](http://www.iso.ch/iso/en/iso9000-14000/iso9000/selection_use/selection_use.html)>

---

## 4.8 Service Deployment

---

### Background

This section addresses the area of Service Deployment.

The briefing documents seek to describe best practices in this area.

### Briefing Documents

The following briefing documents which address the area of Service Deployment have been produced:

- *Top Tips For Service Deployment* (briefing-40)
- *From Project To Production Service* (briefing-38)
- *Planning An End-User Service* (briefing-39)

Advisory documents which cover specific technical areas are available within the section on the appropriate technical area.

### Case Studies

The following case studies which address the area of Service Deployment have been produced:

- *Launching New Database Services: The BIDS Experience* (case-study-27)
- *Providing Access To Full Text Journal Articles* (case-study-28)
- *Merging Data Sources* (case-study-33)
- *Porting the UnCover Service* (case-study-34)
- *Approaches To 'Spring Cleaning' At SOSIG* (case-study-25)

# Top 10 Tips For Service Deployment

Briefing

### About This Document

This briefing paper gives QA Focus's top 10 tips for ensuring project deliverables can be deployed into a service environment with the minimum of difficulties.

### Citation Details

*Top 10 Tips For Service Deployment*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-40/>>

**Keywords:** QA, service, service deployment, tips, briefing

### About This Document

This document provides top tips which can help to ensure that project deliverables can be deployed into a service environment with the minimum of difficulties.

### The Top 10 Tips

- 1 Document The Technical Architecture For Your Project**  
Provide a description of the technical architecture of aspects of your project which are intended for deployment into service. The description will be helpful for the service provider. In addition it can help the funders in gaining an appreciation of the technical approaches being taken by projects across a digital library programme as well as being of value to your project team (especially if staff leave).
- 2 Document Any Deviations From Use Of Recommended Standards Or Best Practices**  
If you fail to make use of recommended standards or best practices you should document the decision and the justification.
- 3 Document Use Of Unusual Or Innovative Aspects Of Your Project**  
If you are making use of any new standards or unusual technologies you should document this, and explain the reasons for your choice. This could include use of emerging standards (e.g. SVG, SMIL), use of Content Management Systems, etc.
- 4 Have An Idea Of Where You Envisage Your Project Deliverables Being Deployed**  
Give some thought to where your deliverables will be deployed. This could be by a JISC Service, within your institution, within other institutions or elsewhere.
- 5 Seek To Make The Service Provider Aware Of Your Project**  
You should seek to make contact with the service provider for your deliverables. You should seek to gain an understanding of their requirements (e.g. see [1] [2]). In addition it can help if the service provider is aware of your work and any special requirements associated with your project.
- 6 Be Aware Of Legal, IPR, etc. Barriers To Service Deployment**  
The service provider will need to ensure that there are no legal barriers to the deployment of your deliverables. This can include clarifying copyright, IPR and accessibility issues.
- 7 Ensure Your Have Any Documentation Which Is Necessary To Assist Service Deployment**  
You should ensure that you provide installation documentation which should list dependencies on other software and cover any security or performance issues. As well as the installation documentation you should also provide user documentation which can help the service provide support for end users.

### 8 Remember To 'Let Go'

Although it can be helpful if your project team is in a position to provide advice to the service provider after the end of the project, the project team should also be willing to relinquish control over the project if, for example, the service provider needs to make changes to your deliverables.

### 9 Learn From Others

Learn from the experiences of others. For example, read the case studies which provide various examples of porting systems into a service environment [3] [4].

### 10 Share Your Experiences

Be willing to share your experiences. For example, consider writing a case study for QA Focus [5].

## References

- 1 **From Project To Production Service**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-38/>>
- 2 **Planning An End User Service**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-39/>>
- 3 **Launching New Database Services: The BIDS Experience**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-27/>>
- 4 **Providing Access To Full Text Journal Articles**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-28/>>
- 5 **Contributing To Case Studies**, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/#contributing>>

# From Project To Production Service

Briefing

### About This Document

This briefing paper ...

### Citation Details

*From Project To Production Service*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-38/>>

**Keywords:** QA, service, service, BIDS, briefing

## Background

Projects deliverables are normally expected to be deployed in a service environment. The deliverables could be passed on to an existing JISC service provider. In some cases, however, a project may evolve into a service. This document outlines some of the issues that need to be considered to facilitate such a transition.

If evolving to a service is not relevant to your project, the issues services need to address when deploying your project deliverables may still be of interest.

## Hosting

Hosting of your project deliverables is one of the first issues to be considered. A prototype service may be developed on in-house equipment and in an environment which may not be appropriate for long term production. Issues to consider include:

- Host in-house or on a national data centre or other host? A factor to consider is the speed of your connection to JANET. If your service is likely to be delivering large volumes of data (e.g. large graphics files, moving images etc), it may be important to locate the service at a site directly connected to JANET.
- If you are not going to host the service yourself, is the software environment of the data centre/other host compatible with the development environment? This includes versions of operating systems and other infrastructure products.

## Data Feeds

Your service may require regular updates of the raw data which the service is delivering to users. Issues to consider when moving into a production environment include:

- How frequently will the database be updated. Data which is updated on a regular basis (e.g. weekly, monthly) should be updated according to a published schedule if at all possible. Users then know when to revisit the site for new information.
- Will the data be ‘pushed’ to the service site, or ‘pulled’ from the source? Each has its advantages and disadvantages. If the data is pushed to the service, a service operator has to check that it has arrived and loaded correctly. Setting up processes to automatically pull the data from the source and load it (for example overnight) is more satisfactory, but an alerting process needs to be built in for instances when the data load fails.

### Gateway Links

The JISC supports a range of subject-specific gateway services. Decide which gateway, if any, your service fits into. The subject matter of your service may span more than one area and therefore need to be incorporated in more than one gateway.

Review the RDN [1] and see where a description and link to your service may fit.

Arrange for your service to be made visible. The more links that are established to your service, the more likely it is to become visible to search engines such as Google and the more successful it is likely to be in terms of awareness and take-up.

### Legal Issues

When an experimental or development system is turned into a production service, there will be copyright, licensing and other legal issues that need to be carefully considered.

Does your service contain any material that is subject to copyright or IPR legislation? This could include such things as images, artwork, extracts from publications, sound or movie clips and so on. If it does, you will need to get permission before you can ‘publish’ your site.

Have you considered how accessible your service is to those with special needs or disabilities? There are now legal obligations that need to be taken into account before releasing a new system.

The JISC TechDis service [2] provides information on how to make your Web site conform. Also consult the appropriate QA Focus document on Accessibility Testing [3] and the JISC Legal Information Service [4] for a range of advice on issues such as Data Protection, Freedom of Information, Disability and the Law, Intellectual Property and much else.

### Managing Expectations

As soon as you have a reliable release date, publicise the fact on relevant JISCmail and other lists. Keep people informed of the progress of the new service as launch day approaches.

As soon as delays appear inevitable, let people know, even if a revised date hasn’t been fixed. This will help front-line staff, who will have to support your service, decide on their own local information strategy.

### Launching the Service

The move of an experimental or development service into a full production service provides a ‘hook’ for raising its profile. Things to consider include:

- Think about marking the start of a new service with a launch event. Try to get a relevant high profile personality, chair of a relevant JISC committee or some other significant person to give a presentation. Make the event worth a day out of the office by including hard information and, if possible, include live demonstrations. The event should be free, but take bookings so you can estimate numbers for catering purposes. Try to get a report in the press by inviting journalists - remember that academics read newspapers too.
- Alternatively launch the service within the context of another event such as a national conference or exhibition. Try to time the launch so it doesn’t clash with other events, national holidays or the start of the academic year.

### Support and Publicity

Consider the kind of support and publicity materials that are appropriate for your service. Examples include:

- Promotional flyers
- Posters
- User guides
- PowerPoint files
- Examples of using the service to answer specific questions (self-help guides)
- More detailed reference material
- Other promotional items (pens, mugs, etc.)

Think about the target audience for the material. You may want to produce different versions for users from different backgrounds and experience. Consider which items may be worth printing (as opposed to being made available on the web). For example posters and flyers are useful for distribution at events such as conferences and exhibitions. Review what other JISC services have done and discuss their experiences with them.

You should also seek advice from the JISC's Communications and Marketing Team [5] who maintain a register of key events and are able to help with such things as preparing and issuing press releases.

### Service Development

Once your service is in production there will be a requirement to improve or update the service and to fix problems. User feedback on suggested service improvements or errors should be gathered through a contact publicised on the service's Web site.

Presentations and demonstrations provide forums for discussion and constructive criticism. Find out if there is an existing user group who will extend their remit to cover your service.

When changes are identified and implemented, ensure that the change is publicised well in advance. Unless the change is an important bug fix, try to make the changes infrequently, preferably to coincide with term-breaks.

### Service Monitoring

Check if your service will come under the remit of the JISC's Monitoring Unit [6]. If it does, you will need to agree a service level definition with them. Typically you will also need to:

- Provide quarterly reports in an agreed format
- Measure usage levels
- Record availability and downtime (scheduled and unscheduled)
- Measure turnaways
- Measure helpdesk performance
- Record how quickly updates are added to the service



### References

1. **Resource Discovery Network (RDN),**  
<<http://www.rdn.ac.uk/>>
2. **JISC TechDis Service,**  
<<http://www.techdis.ac.uk/>>
3. **Accessibility Testing, QA Focus, UKOLN.**  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-02/>>
4. **JISC Legal Information Service,**  
<<http://www.jisc.ac.uk/legal/>>
5. **JISC's Communications and Marketing Team, JISC**  
<[http://www.jisc.ac.uk/index.cfm?name=people\\_masterlist#outreach](http://www.jisc.ac.uk/index.cfm?name=people_masterlist#outreach)>
6. **JISC Monitoring Unit,**  
<<http://www.mu.jisc.ac.uk/>>

# Planning An End-User Service

Briefing

### About This Document

This briefing paper provides advice on how to plan for the development of a service to end users.

### Citation Details

*Planning An End-User Service*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-39/>>

**Keywords:** QA, service, service, BIDS, case study

## Background

For some projects, it will be clear from the start that the intention is to transition the project into an end-user service, either hosted by the project itself, or by another host such as a national data centre.

Other projects may have the potential for development into a production service, but without this being a declared aim of the project.

In both cases, it is sensible to think carefully about how the system might fit into a service environment at the planning and design stage, to avoid costly re-engineering and retro-fitting of features later on.

## Software Environment

The software regime that may seem most appropriate for an experimental development environment may not be the best choice when running a large-scale end-user service. Issues to think about include:

- Software versions; does the software development environment you are intending to use match the versions that are in general use on service delivery platforms?
- Do you have a strong reason for using commercial products (e.g. database management systems)? If so, check to see if there are likely to be high costs when employed in an environment with large numbers of concurrent users. Try to select industry standard public domain products for preference.
- Are the systems you are intending to use supported by the major national service providers? This will be an important issue if they are expected to adopt your project and host it for you.
- Is your project likely to have to integrate with a family of similar products or services? If so, try to ensure that they have compatible operating environments.
- When in doubt, consult with others. Make sure that you do this before development work starts to avoid costly reversals of policy at a later stage.

## Consultation

A key factor in the success of any project is careful preparation and planning. If you intend your project to develop into an end-user production service, it is worth spending time and effort in the early stages of the project testing your ideas and designs. It is easier to rewrite a specification document than to re-engineer a software product.

Depending on the nature of the project, the following issues may be worth considering:

- **Surveys:** carrying out a survey of needs and expectations from typical potential users/customers.
- **Brainstorming sessions:** getting together a group of people interested in the outcome (representing your customers) and carrying out exercises to identify the features and facilities that are most important to them.
- **Consult other JISC service creators:** their experience may help you avoid pitfalls
- **Wireframes:** mocking up a series of screens with active links so that the general functionality of the service can be demonstrated is a powerful way of testing the structure of the service before committing to a full implementation.
- **Prioritising:** not all functions and features will be equally important. Rank them so that you ensure that the most important ones are implemented first. You may decide to relegate some of them to a ‘further development’ phase.
- **Document your design:** ensure that all parties concerned agree to a written specification of what you are aiming to create.

### Authentication and Authorisation

Controlling access to your service may not be an issue when it is in an experimental or development phase, but will become an important consideration if it is released into service.

Some issues to review include:

- Is your service likely to be free or charged for? If it is likely to be free, is this open-ended, or does it depend on central funding? If the latter, what will happen when the funding stops? Will you then need to introduce a subscription fee and, therefore, access controls?
- Even if you expect your service to be free, there may be restrictions on who can use it. For example, the funding of the project may require you to limit access to UK only or higher and further education only.
- Bear in mind that, even if the service is free to UK users, there may be an option for charging for access by non-UK education sector users.
- If you decide you do need to build in access control mechanisms, are you going to use Athens? Athens [1] now supports single sign on (AthensSSO), meaning that users can access several different compliant services with only one password challenge.
- This is a developing area and, depending on the timescale of your project, you may need to keep a watching brief on issues such as the potential for using digital certificates, and Internet 2 related activities such as Shibboleth [2].
- Are any of the remote resources your service depends on IP authenticated? This can cause confusion for users, especially if they are accessing the service from off-campus.
- Even if you don’t expect your project to become a service with controlled access, it would be wise to bear in mind that this could happen in the future, and to structure your service so that an access control mechanism can be easily fitted in at a later stage.

### Legal Issues

When your project reaches the stage of being turned into an production service with large numbers of users, consideration will need to be given to issues which are less important during the development phase.

It is helpful to be aware of these at an early stage in the planning and design of the project to avoid difficult problems later. Some things you should think about include:

- Copyright material: are you thinking of incorporating items such as images, artwork, extracts from publications, sound or movie clips. If so, are there going to be copyright or IPR implications of making this material more generally available? This can apply not only to website material but also printed promotional material. If you have a choice, try to select clipart etc that is in the public domain.
- Accessibility: there is government legislation that needs to be taken into account when designing a new system. You need to think carefully about how your system might be used by those with special needs or disabilities. The TechDis service [3] advice in this area.
- Consult the QA Focus document on *Accessibility Testing* [4] and the JISC Legal Information Service [5] for a range of advice on issues such as Data Protection, Freedom of Information, Disability and the Law, Intellectual Property and much else.

### Planning for Maintenance

It is to be expected that a Web-based user service will require maintenance, revision and updating during its lifetime. There may be requests for new features, or for modifications to the way existing facilities work.

Bear in mind that the people doing this work may not be the original project team that created the service. It is important that the end-products are designed and structured in such a way as to allow parts of the system to be modified and updated by others who are less familiar with the system without unexpected consequences.

Therefore, when starting to develop a new system:

- Ensure that you structure the system in a modular fashion
- Document as the work proceeds, not after the project is complete
- Note any software environment dependencies or support products including versions/releases

### References

1. **Athens access management services**, <<http://www.athens.ac.uk/>>
2. **Internet 2**, <<http://middleware.internet2.edu/>>
3. **JISC TechDis Service**, <<http://www.techdis.ac.uk/>>
4. **Accessibility Testing**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-02/>>
5. **JISC Legal Information Service**, <<http://www.jisc.ac.uk/legal/>>

# Launching New Database Services: The BIDS Experience

Case study

### About This Document

This case study describes the issues the BIDS service faced when taking a deliverable from a third party and deploying it into service.

### Citation Details

*Launching New Database Services: The BIDS Experience*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-27/>>

**Keywords:** QA, service, service, BIDS, case study

## Background

This case study focuses on two of the database services that were created and operated as part of the **BIDS** [1] service: the ISI service and the IBSS service. It describes the experience of launching and supporting these services and discusses lessons learned that might be of value to the creators of other services. The experience demonstrates how a professional approach to service design, support and delivery reflects well on service creators, service operators and the sponsoring bodies. It also demonstrates how community involvement at all stages has a range of important benefits.

The ISI service was the first, and is probably still the best known of the services sponsored by the JISC. Launched in 1991, the BIDS service provided access to the bibliographic databases (or citation indexes) supplied by the Institute for Scientific Information. It was the first large scale, national, end-user service of its kind anywhere in the world. Originally launched as a telnet service, a Web interface was introduced in 1997. This service was replaced in 2000 by the Web of Science operated by MIMAS in which both the data and the service interface are supplied by ISI.

The BIDS IBSS service was launched in 1995. Introduced with a similar interface to the BIDS ISI service, it provided access to bibliographic data indexed and supplied by the International Bibliography of the Social Sciences [2] team based in the library at the London School of Economics (the BLPES). This service was also originally launched as a telnet service, with a Web interface being introduced in 1997. It continues to be operated by BIDS under Ingenta management.

Although these events took place several years ago, much of the experience gained continues to be of relevance and importance to new services being launched today.

## The Problem Being Addressed

When the first ISI service was launched, there were no similar services available, and this presented a major challenge when it came to developing a strategy to bring this new end-user service to the attention of those who could benefit from it. By the time the IBSS service was launched in 1995, it was a more mature market place and the process had become somewhat easier, though a number of the original challenges still remained.

The basic problem was that of creating effective communication channels where none existed before.

### Approaches Taken

The solutions adopted to these issues of communications were manifold and arguably each had a role. New services should consider which of the methods described could be effective for their particular needs.

#### Consultation

For both the services mentioned, consultation took place at a number of different levels. As soon as BIDS had been established, a **Steering Group** was formed. Chaired by a senior librarian, it had representatives from major research libraries, politicians and the BIDS service itself. It provided a valuable sounding board for strategic development of the service, but more importantly from a promotional point of view, it provided an information channel back to the home institutions of the members on current and future service developments. As IBSS is itself a JISC-funded service, it has its own steering group which provides a valuable sounding board for service performance and developments and a forum for announcements.

When the original BIDS ISI service was launched, it wasn't clear how to set about designing an end user service for such a disparate audience. An important contribution to the success of the service was the establishment of a **working group** to help with the service design, including functionality and screen design. Drawn from a large number of different institutions, most of the members were librarians with experience in the use of online and CD-ROM based search services. As well as supplying important knowledge and experience of good design, like the Steering Group members, they became natural ambassadors for the new service.

Shortly after the first BIDS service was launched, a **BIDS User Group** was formed. Subsequently, when MIMAS and EDINA were established the group widened its remit and became known as the JIBS user group [3]. Again this group became a two-way communication channel, lobbying for change and improvement, but also becoming a natural route for disseminating information about the services and their development to intermediaries and thence to end-users.

In summary, one of the important factors in the success of these services was **community involvement** in all stages of service development, as well as the actual service launch.

#### Launching The Services

The BIDS ISI service first became publicly available on 18th February 1991. In the run-up to the launch, regular **news bulletins** were sent to a new mailing list set up in Newcastle on the service which became known as Mailbase (now replaced by JISCmail [4]). This helped keep all the supporters of the new service in touch with progress on how the service was being developed prior to launch. This enabled us to advertise a series of **launch events** and demonstrations which were held at various locations around the UK near the time of the start of service.

When the IBSS service agreement was signed in 1994, a more sophisticated approach was taken with a formal **press release** being issued by the library at LSE (BLPES). When the service itself was launched in January 1995, a series of joint IBSS/BIDS launch events were held around the UK including live demonstrations and 'hands-on' sessions.

### Managing Expectations

It is very common for project timescales to become invalidated by a range of different factors, many outside the service-provider's control. It is important to keep key stakeholders informed as soon as practical of any likely changes to published timescales, even if new dates are difficult to confirm. **Regular postings to mailing lists** provided a very useful vehicle for keeping people up to date on progress with service launches or the introduction of new features. This was especially important when published dates were close to the start of the academic year.

### Training And Publicity Materials

A key feature of the early BIDS services was **community involvement** in the design and writing of a wide range of support material. Although these experiences may not be directly applicable to the present day environment, the general principle of involvement is still valid as an important factor in successful launch and support of new services.

With funding from a central training materials initiative, a suite of ISI service support material (flyers, posters, user guides, training materials) was developed with the aid of around 24 volunteers from 20 different institutions. The design of these materials formed the basis of similar materials developed later for additional services such as IBSS.

### Service Development

Both of these services (ISI and IBSS) went through a series of developments during their lifetime. Managing service development is an important issue. In general at any given time there will be pressure from a number of parties for changes to be made to a service. Each potential development will have **costs** as well as **benefits**. The benefits have to be assessed and prioritised in the context of their value to the service and its users as well as the cost in terms of time and effort. Developments can be typically either low or high value and low or high cost. Relatively low value improvements may still be worth making if their cost is also low. On the other hand, high value developments may still not be justified if their cost is judged to be excessive (or there may be a case for additional funding to carry out the development).

For the ISI service, an example of an early and important, though costly development, was citation searching (searching for all the papers in the database that have identified or cited a particular work in their list of references). Because this was a unique feature of the database (the indexing of all the citations for each paper), all concerned felt it was vital that the facility should be created. It is interesting to note that monitoring of user activity after citation searching was made available showed disappointingly low levels of use of the facility, despite extensive publicity and the creation of documentation and training support materials, etc. In practice the vast majority of searches are simple words or phrases from titles or abstracts, or author names.

### Service Monitoring

Services such as BIDS-ISI and BIDS-IBSS need to be monitored. Funding bodies are keen to establish whether their investment has been wisely spent, and service providers need to judge performance against an agreed set of criteria. The JISC's Monitoring and Advisory Unit (now the Monitoring Unit [5]) drafted **Service Level Agreements** for each database service. Quarterly reports demonstrate how the services have delivered against the benchmarks agreed. These include usage levels, help desk activity,

registrations, documentation and support material, promotion and marketing activity and hardware availability.

This monitoring has been very useful in establishing the high levels of popularity of these services and demonstrating the quality of service delivery. The figures can also be used to extrapolate likely future usage growth and permit planned increases in resources before the service starts to deteriorate.

### **Lessons Learnt**

In this short document it is only possible to draw a limited number of conclusions from more than a decade of experience of running these services.

#### **Consultation**

The chances of success of any new service will be greatly enhanced by making constructive use of widespread consultation. This can cover a wide range of activities including such things as service functionality, interface design, and desirable facilities.

#### **Launching A New Service**

Consideration should be given to marking the launch of a new service in some suitable manner. Depending on the type of service, its intended clientele, and the predicted take-up rate, it may be appropriate to hold a formal launch event at a strategic location; this often means London. Alternatively, the announcement could be made at a suitable conference or exhibition. It is unwise to rely on only one or two methods of announcing a new service. Try to think of as many different appropriate routes to both decision makers and potential end-users. The timing of a launch is also important. Be aware of the academic year cycle, and try to avoid the period immediately before the start of the academic year. Probably the optimum time is mid-late summer term, before staff go on holiday but after most of the student pressure is off.

#### **Managing Expectations**

It is a good strategy to keep people informed as to the progress of a new service as launch day approaches. As soon as delays appear inevitable, let people know, even if a revised date hasn't been fixed.

#### **Training And Publicity Materials**

This is one area where the world has changed significantly from the days of the first BIDS services. Most students and many (though not all) staff are much more computer literate and are frequent network users. The general expectation is that a network service should be intuitive to use and not require extensive training or help. Nevertheless it is important that potential subscribers (if it is a paid for service) and users are aware of the scope and limitations of a new service. So some form of descriptive publicity or promotional material is still relevant, and serious consideration should be given to at least some paper-based material as well as online information. Using professional designers for paper-based material is well worth considering.

#### **Service Development**

Launching a new service is only the beginning, not the end. Mechanisms for feedback from users and purchasers should be established. The service should contain contact details for help and advice. Presentations and demonstrations provide forums for



discussion and constructive criticism. Find out if there is an existing user group who will extend their remit to cover the service.

When changes are identified and implemented, ensure that the change is publicised well in advance. Unless the change is an important bug fix, try to make the changes infrequently, preferably to coincide with term-breaks.

### **Service Monitoring**

Discuss with the JISC Monitoring Unit a suitable set of parameters for measuring performance. Benchmarks will normally be established in a Service Level Agreement. Set up procedures for recording the information and then delivering it to the MU at quarterly intervals.

### **References**

1. **BIDS**, <<http://www.bids.ac.uk/>>
2. **International Bibliography of the Social Sciences**, <<http://www.ibss.ac.uk/>>
3. **JIBS User Group**, <<http://www.jibs.ac.uk/>>
4. **JISCMail Service**, <<http://jiscmail.ac.uk/>>
5. **JISC Monitoring Unit**, <<http://www.mu.jisc.ac.uk/>>

### **Contact Details**

Terry Morrow  
Tel: +44 1373 830686  
Email: [tm\\_morrow@yahoo.co.uk](mailto:tm_morrow@yahoo.co.uk)

Terry was Director of the well-known BIDS service from 2000 until September 2003. Previously he had been the Marketing and Training Manager for BIDS from the beginning of the service in 1990. He is now an independent consultant and a member of the UK Serials Group Executive Committee.

# Providing Access To Full Text Journal Articles

Case study

### About This Document

This case study describes the issues faced by BIDS when launching a new service

### Citation Details

*Providing Access To Full Text Journal Articles*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-28/>>

**Keywords:** QA, service, BIDS, JournalsOnline, case study

## Background

In 1996, the BIDS [1] team launched a new service called **JournalsOnline**. The service was one of the first to provide a single point of entry to electronic versions of articles published in academic journals by more than one publisher. The service was a development of two initiatives, both of which emerged out of the 1993 Higher Education Libraries Review [2] more commonly referred to as the "Follett Report".

The first of these was **eLib** [3] whose aim was to fund a variety of experiments which would collectively make progress in the sphere of electronic access to library materials.

The second was an experiment known as the **Pilot Site Licence Initiative** [4] which was designed to try and find a way of stabilising the problem of rising journal subscription costs and to sponsor experiments in providing network access to the full text of these journals.

The BIDS JournalsOnline service brought together these two initiatives into one service. In 1998 management responsibility for BIDS passed to the newly created company known as Ingenta, and JournalsOnline was renamed IngentaJournals.

## The Problem Being Addressed

By the mid 1990s it was common for students and researchers carrying out literature searches to have searched one or more of the growing number of networked bibliographic databases to identify an article of interest. Having discovered an article they wished to read, they had to note down the reference and either try to find a physical copy of the journal containing the article in their local library, or ask the library to order a copy under arrangements commonly known as 'inter-library loan'. In practice this often meant ordering (and paying for) a copy from the British Library's Document Supply Centre. The costs were either covered by the library's budget, or recovered from the enquirer's department.

Around that time, some of the larger publishers were beginning to establish their own Web sites which provided network access to electronic versions of their journals.

The challenge was to find a way of enabling end-users to find articles in these journals, given that they were unlikely to know which publisher would be likely to own journals covering their particular area of interest.

## Approaches Taken

JournalsOnline was a synthesis of two separate activities involving the BIDS team: the eLib-funded Infobike project and the Pilot Site Licence Initiative.

### **Infobike**

BIDS was a successful applicant for an eLib grant to develop a system for online access to a range of electronic journals from a variety of different publishers. The project, with the rather unlikely title of **Infobike** [5], had a remit to develop and test in general service a system architecture that would allow end-users (as opposed to intermediaries such as librarians) to identify articles of interest by searching or browsing bibliographic databases or publishers catalogues, to check on the status of the enquirer in relation to institutional subscription rights, and to deliver the full text article, either free of charge or for an on-screen payment.

The original partners in the project included Academic Press, Blackwell Science, CALIM (the Consortium of Academic Libraries in Manchester), ICL and the Universities of Keele, Staffordshire and Kent.

### **Pilot Site Licence Initiative (PSLI)**

This initiative was the outcome of discussions which took place in 1994 between a small number of publishers and the Higher Education Funding Council for England (HEFCE). The result was a three year experiment (later extended) involving four publishers - Academic Press, Blackwell Science, Blackwell Publishing and the Institute of Physics Publishing. These publishers offered access to their entire journal collection for between 60% and 70% of the normal price. BIDS submitted a proposal to provide a single point of access to the material from the PSLI publishers and three of the four (IOPP declined) agreed to participate.

### **JournalsOnline**

By taking the technology that was developed under the Infobike project, and combining it with the material that was covered by the PSLI proposal, BIDS was able to create an entirely new service which was christened JournalsOnline. This was launched in November 1996.

The service consisted of a merged publishers' catalogue of bibliographic details of published articles, including titles, authors, affiliations and the full text of any available abstract. Access to the catalogue was set up so that it could be searched either as a registered user or as a guest user. Alternatively the 'contents pages' of journal issues could be browsed to identify articles of interest.

When the user requested the full text, the administration software checked their status. If s/he was registered as belonging to a site that had a subscription to the electronic form of the selected journal, the article was delivered immediately to the screen — typically as a PDF. If the user was from a site that did not subscribe to the selected title, they were given the option of paying for the article, either by account (if one has been set up) or by credit/debit card. Similarly, guest searchers were given the option of article delivery with payment by credit card.

### **Bibliographic Databases**

A further development was to take existing bibliographic database services, such as the original BIDS ISI service or IBSS service, and use these for the resource discovery phase. The search systems carried out a check to see which, if any, of the bibliographic search results matched articles in the full text catalogue. Where there was a match, the user was shown a hypertext link to follow to the full text. Special 'fuzzy matching'

software was used to cater for minor discrepancies between the titles provided by indexing services such as ISI, and the titles supplied by the publishers.

Resource limitations meant that it was impractical to carry out a subscription status check for every search hit, so the 'full text' link only meant that the article existed in full text form in the collection. Accessibility could only be tested by the user following the link — the system would then carry out the subscription check and either offer to deliver the full text (usually PDF) or sell a copy of the article.

### Lessons Learnt

There were numerous issues uncovered by these experiments. There isn't space in this short document to describe all of these in detail, though a number of articles have been published covering many of them. The following summarises some of the major ones.

#### The Myth Of The One-Stop-Shop

One of the original goals of the JournalsOnline service was to provide a one-stop-shop, a single web entry point which was the network equivalent of a well-stocked library. From here it should be possible to find 90% of the material needed to support teaching and research.

The reality has turned out to be a bit different. There are tensions between the requirements of the players, including libraries, funding agencies, commercial publishers, and academic researchers. Each has a different ideal solution and optimum economic model. JournalsOnline explored one model, namely a service largely paid for by commercial publishers to provide a shop window for their material. They are also charged for hosting the full text where this is part of the contract.

In 1998 the JISC awarded the National Electronic Site Licence Initiative (NESLI) managing agent contract to a consortium of MIMAS and Swets Blackwell. Part of their remit was to provide another resource discovery service for searching and retrieving electronic articles. Many of the journals covered by NESLI continued to be available in parallel via JournalsOnline.

In the meantime libraries continue to complain about the high costs of journals (paper and electronic), while commercial publishers say that their costs have risen because of the growing amount of material submitted for publication and the additional costs of parallel publishing. At the same time researchers want peer-group recognition for their work in the recognised leading journals in their sphere (usually commercially published), but also want free and instant access to everyone else's publications. They would also prefer to be able to find them with only one search operation (the Google effect).

A number of major publishers have developed their own end-user services, and there has been a growing tendency for smaller publishers to be taken over by, or merge with, their larger peers. While this is going on, parts of the research community are testing out new models of publishing, including self-publishing, institutional publishing and pre-print publishing.

So the original goal of JournalsOnline of providing the user community with a genuine one-stop-shop was unsuccessful. It did however (and still does in the form of the IngentaJournals service) provide an extremely useful service for identifying and delivering a large body of full text material.

### Commercial Publishing Vs Community Initiatives

As noted earlier, commercial publishing is only one possible model for exposing the results of research. There are numerous experiments for alternative models being carried out, including self-publishing on the Web, publishing of their research output by individual institutions, and pre-print archives. Some references for more information about this work are listed at the end of this article [6], [7], [8].

### Plagiarism

An apparently growing problem for teachers in higher and further education is the ease with which network publishing has made it possible for students, especially undergraduates, to copy sections of material from already published articles and to portray the work as original. The JISC has set up a unit to provide advice and guidance on this difficult issue [9].

### Migrating A Project Into A Commercial Service

Not all JISC initiatives have the potential for developing into a full commercial service. JournalsOnline and, more recently, HERON [10], have provided examples of a successful transition from funded experiment to profitable commercial product.

One lesson to learn from this is that the possibility of eventual commercialisation should be thought about whenever a new JISC project is commenced. If it is thought likely that the resulting service could have a commercial application, then even greater care needs to be taken with issues such as choice of development platform and the integration of community-developed material which may have been made available in the spirit of mutual sharing.

Another issue to consider is who owns the intellectual property rights of any software, data or other products that may emerge from a project. Even if the JISC has provided funds for the project, the IPR typically belong to the major grant receiving organisation. But you should check carefully to see what the situation is with your project.

## References

1. **BIDS**, <<http://www.bids.ac.uk/>>
2. **Joint Funding Council's Libraries Review Group: Report**, <<http://www.ukoln.ac.uk/services/papers/follett/report/>>
3. **eLib: The Electronic Libraries Programme**, <<http://www.ukoln.ac.uk/services/elib/>>
4. **Report on Phase I of the Evaluation of the UK Pilot Site Licence Initiative**, <[http://www.hefce.ac.uk/pubs/hefce/1997/m3\\_97.htm](http://www.hefce.ac.uk/pubs/hefce/1997/m3_97.htm)>
5. **Infobike**, <<http://www.ukoln.ac.uk/services/elib/projects/infobike/>>>
6. **ePrints UK**, RDN, <<http://www.rdn.ac.uk/projects/eprints-uk/>>
7. **BioMed Central Agreement**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=biomed\\_pr\\_190603](http://www.jisc.ac.uk/index.cfm?name=biomed_pr_190603)>
8. **Focus on Access to Institutional Resources (FAIR) Programme**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_fair](http://www.jisc.ac.uk/index.cfm?name=programme_fair)>
9. **JISC Plagiarism Advisory Service**, <[http://online.northumbria.ac.uk/faculties/art/information\\_studies/Imri/JISCPAS/site/jiscpas.asp](http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/JISCPAS/site/jiscpas.asp)>
10. **Welcome to HERON**, <<http://www.heron.ingenta.com/>>

### Contact Details

Terry Morrow

Phone: +44 1373 830686

Email: [tm\\_morrow@yahoo.co.uk](mailto:tm_morrow@yahoo.co.uk)

Terry was Director of the well-known BIDS service from 2000 until September 2003. Previously he had been the Marketing and Training Manager for BIDS from the beginning of the service in 1990. He is now an independent consultant and a member of the UK Serials Group Executive Committee.

# Approaches To 'Spring Cleaning' At SOSIG

Case study

### About This Document

This case study describes the systematic approaches taken by a service in order to ensure that its metadata is of sufficient quality to provide a quality end-user service.

### Citation Details

*Approaches To 'Spring Cleaning' At SOSIG*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-25/>>

**Keywords:** QA, service, metadata, SOSIG, case study

## Background

The JISC and ESRC-funded **SOSIG** service [1] is one of the longest running RDN subject gateways. SOSIG provides access to high quality, peer-reviewed records on Internet resources in the area of Social Science, Business and Law.

Many projects will be providing metadata which describes projects' deliverables, which may include resource discovery or educational metadata.

In order for projects to gain an understanding of the importance which JISC services place on the quality of metadata, this case study has been written which describes the approach to 'spring-cleaning' which SOSIG has carried out as one of its quality assurance procedures in order to ensure that its records continued to provide high quality information.

## The Collection

The core of the SOSIG service, the Internet Catalogue, holds over 26,000 structured metadata records describing Internet resources relevant to social science teaching, learning and research. Established in 1994, SOSIG is one of the longest-running subject gateways in Europe. The subject section editors have been seeking out, evaluating and describing social science Internet resources, developing the collection so that it now covers 17 top-level subject headings with over 1,000 sub-sections. Given the dynamic nature of the Internet, and the Web in particular, collection development is a major task. Collection management (i.e. weeding out broken links, checking and updating records) at this scale can also be something of a challenge.

The SOSIG core team, based at ILRT in Bristol, devotes considerable resource to removing or revising records with broken links (human checks based on reports from an automated weekly link-checking programme). Subject section editors, based in universities and research organisations around the UK, also consider durability and reliability of resources as part of the extensive quality criteria for inclusion in the Catalogue. They regularly check records and update them: however, the human input required to do this on a systematic and comprehensive scale would be beyond current resources. SOSIG has therefore recently embarked on a major 'spring cleaning' exercise that it is hoped will address this issue and keep the records current. We describe below the method, and outcomes to date.

### Why Bother?

There are several reasons why such collection management activity is important. User feedback indicates that currency of the resource descriptions is one of the most appreciated features of the SOSIG service. SOSIG and other RDN hubs are promoted on the basis of the quality of their records: offering out-of-date descriptions and other details is likely to frustrate users and, in the long term, be detrimental to their perceptions and therefore use of the service. Recent changes in data protection legislation also emphasise the obligation to check that authors/owners are aware of and happy with the inclusion of their resources in SOSIG. Checking with resource owners also appears to have incidental public relations benefits and is helping to develop the collection by identifying new resources from information publishers and providers.

### The Approach

How did we go about our spring-clean? Each of the metadata records for the 26,000 resources catalogued in SOSIG contains a field for 'administrative email' - the contact email address of the person or organisation responsible for the site. We adapted an existing Perl script (developed in ILRT for another project), which allowed a tailored email to be sent to each of these addresses. The message includes the URL of the SOSIG record(s) associated with the admin email. Recipients are informed that their resources are included in SOSIG and are asked to check the SOSIG record for their resource (via an embedded link in the message) and supply corrections if necessary. They are also invited to propose new resources for addition to the Catalogue.

### Phasing The Process

We first considered a mass, simultaneous mailout covering all 26,000 records. The script sends one message per minute to avoid swamping the servers. However we had no idea of the level of response likely to be generated and wanted to avoid swamping ourselves! We therefore decided to phase the process, running the script against batches of 2,000 records on a roughly monthly basis, in numerical order of unique record identifiers, these were grouped notifications so that an administrator would get one email referring to a number of different sites/pages they were responsible for. The process was run for the first time at the end of July 2002 and, on the basis of low-numbered identifiers, included records of resources first catalogued in SOSIG's early days. The SOSIG technical officer oversaw the technical monitoring of the process, whilst other staff handled the personal responses, either dealing with change requests or passing on suggestions for additional resources to Section Editors responsible for specific subject areas on SOSIG.

### Some Results

#### A range of responses

In total we received 950 personal responses (approximately 4%) from email recipients. A further 3,000 or so automated 'bounced' responses were received. Those of us who are regular and long-term users of the Web are well aware of the fairly constant evolution of Web resource content and features. The SOSIG spring clean exercise also highlights the extent of change in personnel associated with Web resources. As mentioned above, of the emails sent relating to the first 4,000 records, over a quarter 'bounced' back. Although a very small proportion of these were automated 'out of office' replies, most were returned because the address was no longer in use.



The majority of the personal responses requested a change in the URL or to the administrative email address recorded for their resource. Many had stopped using personal email addresses and had turned to generic site or service addresses. Others reported that they were no longer responsible for the resource. As the first batches included older records, it will be interesting to see whether the proportion of bounced and changed emails reduces over time, or whether people are really more volatile than the resources.

We have to assume that the remaining email recipients have no cause for complaint or change requests. In fact, we were very pleased at the overwhelmingly positive response the exercise has generated so far. Many simply confirmed that their records were correct and they were pleased to be included. Others noted minor corrections to descriptions, URLs and, as mentioned, admin email addresses. Many also took the time to recommend new resources for addition to the Catalogue. Only one or two concerns were raised about the inclusion of certain data in the recorded, although there were several queries which highlighted changes needed to the email message for the second and subsequent batches.

One of these arose as a result of the de-duplication process, which only operates within each batch of 2,000 records. Where the same admin email address is included in records excluded from that batch, the de-duplication process ignores it. Some recipients therefore asked why we had apparently included only some of their resources, when they are actually on SOSIG, just not in that particular set of records.

Only one major issue was raised, that of deep-linking. It seems that this is a problem for one organisation, and raises questions about the changing nature of the Web - or perhaps some companies' difficulty in engaging with its original principles. Time will tell whether this is an issue for other organisations: to date it has been raised only once.

### **Handling the responses**

Spring-cleaning in domestic settings always involves considerable effort, and the SOSIG spring clean is no exception. SOSIG staff spent about a week, full-time, dealing with the personal responses received after each batch of 2,000 records were processed. The first batch of messages all had the same subject line, so it was impossible to distinguish between responses appearing in the shared mailbox used for replies. In the second 2,000, the subject line includes the domain of the admin email address, which makes handling the responses much easier.

Bounced messages create the most work, because detective skills are then necessary to check resources 'by hand' and search for a replacement admin email address to which the message can then be forwarded. Minor corrections take little time, but the recommendation of new resources leads to initiation of our usual evaluation and cataloguing processes which can be lengthy, depending on the nature and scale of the resource.

We realised that timing of the process could have been better: initiating it in the middle of Summer holiday season is likely to have resulted in more out-of-office replies than might be expected at other times. Emails are now sent as routine to owners of all new additions to the catalogue: this complies with the legal requirements but is also an additional quality check and public relations exercise. Once informed of their inclusion in the gateway, resource owners may also remember to notify us of changes in future as has already been the case!

Although time-consuming, the spring clean is still a more efficient way of cleaning the data than each Section Editor having to trawl through every single record and its associated resource. Here we are relying on resource owners to notify us of incorrect data as well as new resources: they are the ones who know their resources best, and are best-placed to identify problems and changes.

### Implications For Projects

If you are providing metadata which will be passed on to a JISC service for use in a service environment the JISC service may require that the metadata provided is still up-to-date and relevant. Alternatively the service may need to implement validation procedures similar to those described in this document.

In order to minimise the difficulties in deploying metadata created by project into a service environment, projects should ensure that they have appropriate mechanisms for checking their metadata. Ideally projects will provide documentation of their checking processes and audit trails which they can make available to the service which may host the project deliverables.

### References

1. **SOSIG**, <<http://www.sosig.ac.uk/>>

### Contact Details

This document is based on an Ariadne article entitled "*Planet SOSIG - A spring-clean for SOSIG: a systematic approach to collection management*" originally written by Lesley Huxley, Emma Place, David Boyd and Phil Cross (ILRT). The article was edited for inclusion as a QA Focus case study by Brian Kelly (UKOLN) and Debra Hiom (ILRT).

Contact details for the corresponding authors is given below.

Debra Hiom  
ILRT  
University of Bristol  
Bristol  
email: [d.hiom@bristol.ac.uk](mailto:d.hiom@bristol.ac.uk)

Brian Kelly  
UKOLN  
University of Bath  
BATH  
Email: [b.kelly@ukoln.ac.uk](mailto:b.kelly@ukoln.ac.uk)

# INHALE And INFORM Case Study

Case study

### About This Document

This case study describes the experiences of the INHALE and INFORM projects.

### Citation Details

*INHALE And INFORM Case Study*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-31/>>

**Keywords:** QA, service, INHALE, INFORMS, case study

## Background

This case study describes the experiences of the INHALE project (which ran from September 2000- March 2003) and the subsequent INFORMS project (which ran from October 2002-August 2003).

## Introduction

The INHALE Project (one of the JISC 5/99 Programme projects) had a number of aims:

- To utilise the ubiquity of the Internet to create a set of learning materials which embed a variety of electronic services within a series of Web-based units. These services consist of those available nationally via the DNER and those subscribed to locally.
- To integrate the open world of JISC's Information Environment (IE) with the relatively closed world of the virtual learning environment (VLE).
- To demonstrate how the INHALE materials can equip students with transferable information skills to allow them to select, locate, use and evaluate information for their studies.

The outcomes from the project were:

- A set of standalone Web based information skills units, which are accessible to HE and FE institutions in the UK, each of these units being designed to assist users to acquire the necessary skills to find and use quality information sources.
- A preaxial demonstration of how the materials can be successfully embedded within a VLE such as Blackboard or WebCT.
- An iterative product methodology showing how live Web sites can be exploited to maximize interactivity.
- A database of (learning) objects created from the initial standalone materials which would be accessible, searchable and extensible to other HE/FE institutions allowing individuals to recreate a tailored set of materials.
- Reports and journal articles describing the processes and experiences of the work of the project.
- Quantitative and qualitative evaluations of the utilization of the resources by students at the University of Huddersfield and at the project's partner institutions.

At the outset of the project the vision of what was required was set out clearly in the project plan and this was closely adhered to throughout.

During the first year the project team successfully created a standalone set of information skills units for students on Nursing and Health courses. The JISC's technical guidelines on interoperability and accessibility guided the Web developer in the creation of the online resources. The new information skills materials using the JISC's IE resources as well as freely available Web resources were tested within pilot modules in the School of Health Sciences at Huddersfield. Evaluation reports from these were written. These evaluations fed into the continuous "product" development.

During the second year (September 2001- September 2002) additional information skills units were created and some of the initial units were customised. Some of the new resources were based around subscription information databases and were cascaded out for use within the partner institutions, Leeds Metropolitan University and the University of Central Lancashire.

Meanwhile at Huddersfield the resources were being embedded at different levels within Blackboard in new pilot modules.

Running parallel to the delivery of the resources within modules was the continuing development of the INHALE "database". The "database" was seen as the key to enable customisation of the initial set of INHALE materials and the generation of new units by all the partner institutions. This required the disaggregation of all the original materials into objects. Fortunately, from the outset, the vision was that the end result would be a database of learning "objects" and all the materials were created with this concept to the fore. Thus the disassembly was not as onerous a task as it may seem.

Dissemination of the project's learning and outcomes began early in the project and had two key strands. The first was to involve stakeholders in the delivery of information skills within the institutions. Workshops and meetings were held internally and attended by academic teaching staff, librarians, learning technologists, computing service staff and learning and teaching advisors. The second strand was to disseminate to the wider UK HE and FE community and various events were held beginning with an event that was to be repeated "E-Resources for E-Courses".

By July 2002 interest in the use of the INHALE resources had grown. In September 2002, the submission of a proposal for a project within the DiVLE Programme to continue the work of the INHALE project was successful. The new project was named INFORMS and from October 2002 to March 2003, the INHALE project and INFORMS projects ran concurrently. During this time the University of Loughborough and the University of Oxford, (the new INFORMS project partners), were able to test the transferability and viability of all the INHALE project materials and models as well as inputting new ideas for developing the resources.

By the end of the INHALE project in March 2002 there were over 200 units within the new database and a number of new institutions were also testing and using the database.

At this point an exit strategy was written for the INHALE project. The project team felt that there was a possible "market" for the INHALE/INFORMS information skills database within the HE/FE community. However the JISC Programme Managers considered that the database of units required more "market testing" within the HE/FE community. To some extent the INFORMS project has allowed the team to begin the process of market testing.

The INFORMS project officially completed in August 2003 and there are now over 400 units in the INFORMS database and 17 institutions have portfolios of units across the range of subjects studies across the HE/FE community. Usage of the resources can be tracked via a Web log analysis tool developed in house that is linked to the database.

Librarians (and some academic teaching staff) institutions are creating their own online, interactive innovative information skills teaching and learning resources without any knowledge of Web authoring. The database allows instant editing and updating, it automatically produces accessible and printable versions of the units. The 400 plus units in the database are shared across all the participating institutions. Units copied across institutions are tracked via an audit trail. A User Guide, Editorial Policy and Conditions of Use Agreement are all essential documents that have been produced to support users of the database.

### Problems Encountered

#### Project Creep

There was some initial hold-up in getting the project started and by the time the Project Co-ordinator joined the team in January 2001, the project was approximately 2-3 months behind in writing the initial units, rolling out the baseline evaluation and writing the evaluation instruments. Delivery to the students in the first pilot module was set for mid-February 2001 and this deadline was met.

**Action Taken:** The Project Co-ordinator was employed full time instead of part-time. This was funded from both the unspent salary from the project for 3 months prior to January 2001 and topped up from the salary of the Project Co-ordinator's substantive post.

#### Loss of a Project Partner

Manchester Metropolitan University pulled out of the project in June 2001 when they were successful with obtaining funding from the JISC for the BIG BLUE Project.

**Action Taken:** None, but subsequent events mitigated against the loss of the partner.

#### Loss of Project Staff

The Project Director left to take up another post at the University of Central Lancashire in September 2001. The loss of the Director's role as the stakeholder for the project within the Library Management Team and amongst the Academic Librarians had a detrimental effect on the uptake of the resources across the institution that is still being addressed.

The loss of someone else with information skills expertise to bounce off ideas and to provide another point of view on the project's development, as well as mutual support, has been a problem for the subsequent Project Director.

However the move by the Project Director to the University of Central Lancashire was beneficial as UCLan was invited to take the place of Manchester Metropolitan Library on the project and the input from that institution was invaluable.

**Action Taken:** The Project Co-ordinator took over the role of Director.

In November 2001, two months after the demise of the Project Director the project's Web developer was recruited to an internal position in the library. This could have

proved disastrous but in fact a new Web developer was recruited from the interviewees for the internal post and began work on the project only 10 days after the original developer had moved.

**Action Taken:** Recruited new Web developer from suitable candidates already interviewed for another post.

### **Territorialism**

The new Project Director encountered internal political problems that have constantly hampered the uptake of the resources.

**Action Taken:** Sought the support of line manager.

### **Local Area Network Problems**

In September 2001 the University of Huddersfield experienced a severe problem with the load on its network. The project was unable to continue development on integrating video and audio into the resources.

**Action Taken:** Abandoned using video and audio.

### **Changes in University Infrastructure**

The problem with the LAN traffic had a knock-on effect. The central service managing the Blackboard resources plus the learning and teaching support for this was re-organised. Key stakeholders in this support area within the University left so the necessary key personnel to champion the uptake of the INHALE resources in Blackboard were lost. Eventually some new posts have been created.

**Action Taken:** Contact made with new staff and process of rolling-out started again.

### **Running Two Projects Concurrently**

The Project Director misjudged the demands that running the two projects (INHALE & INFORMS) during the period October 2002 to March 2003 would make.

**Action Taken:** None. With hindsight it might have been more feasible to have a different financial balance between the 3 partners to have allowed additional staff to be recruited at Huddersfield.

## **Future Developments**

The JISC only require projects to make their Web sites available to the rest of HE/FE for 3 years after the end of the project. Thus if a resource has a potential for further uptake and development then the project will need to produce a strategy to enable this.

The University of Huddersfield is not in a position to fund user support for the database. The institution is still in the early days of recovery after its re-organisation of the technical and teaching and learning support infrastructure for Blackboard.

The INFORMS (INHALE) project team have been pursuing a number of possible strategies:

### **Commercialisation**

The INFORMS Project team think that there is a commercial potential for the INFORMS software beyond the HE/FE sector and have been successful in a bid for

funding to investigate and pursue this further over the next 12 months via a University of Huddersfield Commercial Fellowship.

It is planned that any profit will eventually be used to provide support for the INFORMS database. (Staffed support for HE/FE users of the INFORMS database, support of the Web server hosting the database, support to implement new developments).

### **Collaboration with other JISC Projects**

The location of the INFORMS resources within an Information Skills Portal alongside the VTS, Big Blue & the Resource Guides etc. would be an ideal scenario and one that has been suggested already by the Big Blue project.

### **Collaboration with other Key Players**

Both the Open University and Sheffield Hallam have products that may benefit from the technical developments of the INHALE/INFORMS projects.

### **Hosting by JISC Services**

If demand for portfolios in the database grows then the capacity of the Web server at Huddersfield will be over-reached. So one possible strategy could be to move the database to either EDINA or MIMAS.

Mirroring the database at Edina has been explored and this may be possible in 12 months if the return from the commercialisation of the software is sufficient.

### **Providing Individual Institutions with the Software**

It may be possible to give away the software to iHE/FE institutions to run on their own servers and develop should they wish to do so. The main disadvantage of this is the loss of the shared resources.

### **Additional Support from the JISC**

A case has been put to the JISC, the reply has been that additional evidence of a need must be gathered through “market testing”.

## **Conclusions**

For the time being the new INFORMS (Commercial) Project is the route being taken by the ex Project Director of INHALE/INFORMS to create supportive funding in the long term for the INFORMS database of information skills teaching and learning resources. The new INFORMS (Commercial) project began officially on 1<sup>st</sup> October 2003 and will run for 12 months. One of its first successes has been to secure a place at a reception in the House of Commons being held by the Set for Britain group who are promoting start-up, spin-out, spin-off commercialisation of UK University research. At the reception we will be delivering a poster presentation for the MPs, Peers and various other attendees of the proposed commercialisation of the INHALE/INFORMS software.

# Sustainability: The TimeWeb Experience

Case study

### About This Document

This case study describes the experiences of the TimeWeb project.

### Citation Details

*Sustainability: The TimeWeb Experience*, QA Focus, UKOLN,  
<<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-32/>>

**Keywords:** QA, service, sustainability, TimeWeb, case study

## Background

The **TimeWeb** (Time Series Data on the Web) project was a joint project between Biz/ed [1] at the Institute for Learning and Research Technology [2] at the University of Bristol and the JISC/ESRC supported MIMAS service [3] at Manchester Computing at the University of Manchester. The central aim of the project was to develop the key national and international macro-economic time series data banks, such as the OECD Main Economic Indicators, held at MIMAS into a major learning and teaching resource.

The key deliverables of the TimeWeb project were:

1. The development of extensive learning and teaching materials [4] covering most essential data handling skills. Included in the materials are explanations, illustrations, worksheets and a reference section with full glossary. The materials are split into three main parts to reflect the different skills involved in handling time series data.
2. The TimeWeb Explorer – an advanced web-based application that allowed users to explore the OECD Main Economic Indicators, a database containing thousands of comparative statistics for the major economies of the world. The TimeWeb Explorer enabled users to browse series descriptions, subset data by country, subject or keyword, plot selected series and download data.

The TimeWeb learning and teaching materials and the TimeWeb Explorer were successfully launched into service on 14<sup>th</sup> February 2002 [5]. Through the use of shared style sheets and a common design, movement between the learning and teaching materials developed by Biz/ed and the TimeWeb Explorer web site developed at MIMAS appeared seamless to the user. Thus Timeweb provided an integrated package of both data and learning and teaching materials.

## Problem Being Addressed

This case study describes the approaches adopted by Biz/ed and MIMAS to deliver the TimeWeb project deliverables to users and also to embed those deliverables into a service environment to facilitate long term maintenance and support.

## The Approach Taken

In order for the JISC to be successful in its stated aim of enhancing JISC services for learning and teaching, it was imperative that the deliverables from the TimeWeb project were released to users and embedded in a service environment. Both MIMAS and Biz/ed fully understood the importance of releasing the deliverables and promoting their long term use in order to maximise JISC's investment.



In the original project plan, it was intended that the release of prototype interfaces and learning and teaching materials for user testing and evaluation would take place at various stages during the development phase. The objective was that final release of the TimeWeb Explorer and the associated learning and teaching materials would coincide with the end of the project. Once the project ended it was anticipated that the ongoing support and maintenance of the TimeWeb Explorer and learning and teaching materials would be absorbed by the existing MIMAS and Biz/ed service infrastructures.

At the time, these aims were felt to be realistic as both MIMAS and ILRT had considerable experience in transferring project deliverables into services. Whilst MIMAS and Biz/ed successfully achieved the objective of releasing the deliverables into service at the end of the project, the long term support and maintenance has proved more problematic than originally anticipated.

### Problems Experienced

The TimeWeb team encountered a range of problems which had to be overcome in order to achieve the twin objectives of releasing the project deliverables to users and also to embed these deliverables in a service environment to facilitate long term maintenance and support. The following is a summary of the problems encountered and how the Biz/ed and MIMAS teams overcame them:

#### 1. TimeWeb Explorer:

##### *Early Intentions*

MIMAS encountered a range of technical problems that needed to be overcome before the TimeWeb Explorer could be officially released to users. To avoid the normal problems associated with the long term support and maintenance of software developed 'in house' MIMAS decided to use a proprietary solution for the development of the web based interfaces to the time series databanks. The selected solution was SAS AppDev Studio [6] which had been developed by the SAS Institute [7]. The intention was to use the visual programming environment provided by SAS to build a lightweight Java based interface to the time series databanks.

##### *Development Problems*

Whilst Java facilitated the development a sophisticated and interactive interface it also resulted in a series of major development problems which had to be resolved. For example, the Java sandbox security model typically does not allow data files to be written to the server or client, an essential step for data downloads. Such development problems were compounded as the TimeWeb Explorer was one of the most advanced projects ever written with SAS AppDev studio, and SAS themselves were limited in the technical help they could provide. The additional staff effort required to resolve the unanticipated technical problems significantly held up development work and prevented MIMAS from releasing the interface for user testing until towards the end of the project. It also resulted in MIMAS shelving plans for the more advanced user interface.

##### *Deployment Delays*

When the TimeWeb Explorer was released for initial user testing a number of unanticipated deployment problems were encountered which caused significant delays. Firstly, the use of the applet required users to install a particular version of the Sun Java Plug-in (Sun's newer releases of the plug-in are unfortunately not backward compatible

with earlier versions). AppDev Studio tends to lag behind the latest version of the plug-in produced by Sun and, moreover, different versions of the plug-in could not co-exist on the same PC. This created problems for users unable to install software on their PC due to network restrictions, or for cluster users where the latest version of the plug-in had already been installed. Much work went into finding the best compromise, resulting in a parallel version of Timeweb that ran on later versions of the plug-in also being created. A second deployment problem resulted from the many variations amongst user systems (such as operating system, browser version, download permissions, cache settings or network connection), all of which had some influence on the operation of the TimeWeb Explorer. All these deployment problems had to be fully investigated and documented to allow a wide range of users as possible to use the Timeweb Explorer reliably. Resolution of these technical problems required significant additional development effort towards the end of the project which further delayed the release of the TimeWeb Explorer into service.

Before the TimeWeb Explorer was released to users as a new service, it was necessary to embed it within the existing MIMAS Macro-Economic Time Series Databank Service. As the OECD MEI was updated monthly it was necessary to establish data loading procedures which existing support staff could use. As part of the service integration, it was also necessary to implement and test the access management system required to restrict access to authorised users as required under the terms and conditions of the OECD data redistribution agreement.

### ***Training and Support***

It was also necessary to develop a range of support and promotional materials to coincide with the release of the TimeWeb Explorer. MIMAS launched the Explorer alongside an accompanying website containing help pages, detailed information on running requirements and links to the metadata for the OECD MEI databank. In addition to email announcements sent out to various lists, a TimeWeb Explorer factcard [8] and an A3 TimeWeb publicity poster were produced and widely distributed. The creation of these publicity materials required assistance from other support staff within MIMAS. In addition, it was also necessary to provide training to MIMAS Helpdesk staff to enable them to deal with initial queries relating to the use of the TimeWeb Explorer.

### ***On-Going Maintenance***

Having transitioned the TimeWeb Explorer into a supported MIMAS service it soon became apparent that additional effort was required for both on-going maintenance and development of the interface. For example, additional software engineering effort would be required to respond to user feedback/bug reporting and – more importantly – to extend the TimeWeb Explorer interface to provide access to other time series databanks. The loss of dedicated software engineering effort at the end of the project - due to the absence of continuation funding - made the on-going maintenance and development of the interface very problematic.

### ***A New Solution***

When the TimeWeb project started in 2000, there were no proprietary systems available that could have been used to provide the required flexible Web-based access to aggregate time series. By the time the project had ended, the Beyond 20/20 Web Data Server (WDS) [9] had emerged as a standard tool for the publication and dissemination

of international time series databanks over the web and was starting to be used by many of the world's largest international and national governmental organisations, such as OECD and the Office for National Statistics (ONS). Not only did the Beyond 20/20 WDS offer the required functionality, it could also be used to import data in a range of different formats. More significantly, the WDS runs in a standard Web browser (IE 4.01/Netscape 4.5 and above) with Javascript enabled thus avoiding the problems associated with Java plug-ins which had been encountered with the TimeWeb Explorer.

In 2002/2003, the MIMAS Macro-economic Time Series Data Service underwent a major transformation as part of the establishment of the new ESRC/JISC funded Economic and Social Data Service (ESDS) [10]. In January 2003, the new ESDS International Data Service [11] based at MIMAS was launched. In order to provide flexible Web-based access to a much larger portfolio of international time series databanks statistics produced by organisations, such as the International Monetary Fund, and to minimise in-house interface development overheads, a strategic decision was taken to standardise on the Beyond 20/02 WDS interface. As a result, an internal project team was set up to plan and oversee the transition from the TimeWeb Explorer to Beyond 20/20 WDS. The project team benefited considerably from the lessons learnt when introducing the TimeWeb Explorer interface into service and the transition to Beyond 20/20 was completed in April 2003.

## 2. TimeWeb Learning and Teaching Materials

### ***Where to Begin?***

One of the most significant problems faced in the creation of the learning materials was the sheer breadth of potential data handling skills that exist. There is a wide variety of contexts and qualifications that involve data skills. The Biz/ed team was aware that whilst the Higher Education market was the chief target, the materials would have maximum effectiveness if they addressed other audiences. It follows that supporting the needs of different users is difficult when the user base can be drawn from such a variety of backgrounds.

### ***Supporting the Materials***

The main problem faced by the Biz/ed team was in relation to the need for sample data to support the learning and teaching materials under development. This need having been identified, it was necessary to source the datasets and agree terms for their release by the data provider. In this case it was felt appropriate that UK data would be sampled. UK National Statistics were approached in order to gain their approval for a small number of datasets to be held within the TimeWeb suite of learning and teaching materials.

### ***Approval to Use Sample Data***

During the period of negotiations with National Statistics there was a change in policy at Governmental level which had the effect of removing all barriers to the use of official data, on the proviso that commercial benefit was not to be obtained. As Biz/ed is a free educational service, this did not pose a problem. However, getting hold of the data codes for the sample datasets added extra delays in being able to finally release the TimeWeb learning and teaching materials.

### ***Maintenance Issues***

In preparation for TimeWeb moving into service, it was recognised that the maintenance of up-to-date data was crucial. This involved technical work in creating scripts to run out the data from National Statistics. This occurs on an annual basis. However problems continue to emerge as the codes applied to the data by National Statistics appear to be changed on every update. Thus, on-going maintenance continues to be an issue.

As a non-JISC service at the time of the project, the materials were placed within Biz/ed as a stand-alone resource. Given that Biz/ed became a JISC service in late 2002, there are now issues around the integration of the TimeWeb resource into the service and how they are maintained

## **Things We Would Do Differently**

### **Learning Objects**

One of the key things to come out of the project was how difficult it was to respond to emerging standards and changing requirements both during the development phase and once deliverables have been transferred into a service environment. For example, since the completion of the TimeWeb project, learning objects have emerged as a major theme in e-learning. Migrating the TimeWeb materials to a learning object model and ensuring compliance with new metadata standards (e.g. IEEE LOM) so that that they are reusable and form part of a true resource discovery environment would be a major undertaking which would require additional funding. However, it is very difficult to respond to new funding opportunities, such as X4L [12], when teams and associated expertise have dispersed.

### **Exit Strategies**

We believe that TimeWeb would have benefited from closer examination of possible project exit strategies at various points during the project. When the project finished in February 2002 there was very little guidance from JISC about future directions. An optimal solution would have been for the project partners – in their roles as service providers - to seek continuation funding for the materials to be updated and the data interface to be maintained. For instance, the sample datasets used within the learning materials could have been adapted to reflect changing interests and events. Whilst we demonstrated successfully that project deliverables could be delivered into service through existing service providers it was clear that additional resources were going to be required for long term support and maintenance. As a project, we should have been more proactive at an earlier stage in terms of making a case to JISC for additional funding.

### **Into Service**

The detailed planning of the transfer of project deliverables into service was left until towards the end of the project. It would have been better to start the planning at a much earlier stage. It would have also have been advisable to have defined the transfer of deliverables to service as a separate work package in the original project plan. This work package would have needed to be kept under review during the course of the project to reflect changes and developments. However, it was clear from our experience that we had underestimated the amount of software engineering effort required to transfer ‘project quality’ software to ‘service quality’. We also underestimated the

amount of additional work that would have to be provided by other support staff to assist with the transfer to service.

### Technical Issues

Whilst Java held out the promise of developing a sophisticated and interactive interface to time series that would meet the needs of researchers and students alike, we had not fully anticipated the technical problems that would arise. Had we been aware of the pitfalls of the Java route, we would have probably adopted a simpler and more robust database driven approach to delivering time series data across the web. Rather than trying to fully exploit leading edge technology we should have focused on a less challenging software solution that would have been easier to transfer into service and subsequently maintain.

### Outcomes

Whilst the TimeWeb Explorer had a limited service life and was eventually replaced by a commercial system, this does not mean that it was a failure. During the year in service it resulted in a significant increase in the use of the OCED MEI – much of it for teaching and learning. Developing the TimeWeb Explorer gave MIMAS invaluable insights into what was required to deliver international macro-economic time series via an interface that was suitable for both researchers and students. Therefore, TimeWeb has played an important role in the establishment of ESDS International as a major new UK academic data service.

## References

1. **Biz/ed**, <<http://www.bized.ac.uk/>>
2. **ILRT**, <<http://www.ilrt.bris.ac.uk>>
3. **MIMAS**, <<http://www.mimas.ac.uk/>>
4. **TimeWeb Learning and Teaching Materials**,  
<<http://www.bized.ac.uk/timeweb/>>
5. **TimeWeb launch announcement**,  
<<http://www.mimas.ac.uk/news/timeweb-feb02.html>>
6. **AppDev Studio**, <<http://www.sas.com/technologies/bi/appdev/appdev/>>
7. **SAS**, <<http://www.sas.com/>>
8. **TimeWeb Factcard**, <<http://www.mimas.ac.uk/docs/paper/twebleaflet.pdf>>
9. **Beyond20/20 Web Data Server**,  
<<http://www.beyond2020.com/products/webdataserver.html>>
10. **ESDS**, <<http://www.esds.ac.uk/>>
11. **ESDS International Data Service**, <<http://www.esds.ac.uk/international/>>
12. **Exchange for Learning Programme**, JISC  
<[http://www.jisc.ac.uk/index.cfm?name=programme\\_x4l](http://www.jisc.ac.uk/index.cfm?name=programme_x4l)>

## Contact Details

Keith Cole  
Deputy Director/Services Manager  
MIMAS  
University of Manchester  
Email: [Keith.Cole@man.ac.uk](mailto:Keith.Cole@man.ac.uk)

Andy Hargrave  
Biz/ed Research Officer  
ILRT  
University of Bristol  
Email: [Andy.Hargrave@bristol.ac.uk](mailto:Andy.Hargrave@bristol.ac.uk)

---

## 5 The QA Focus Toolkit

---

### About The Toolkit

One of the project deliverables for QA Focus is a toolkit to help projects in implementing QA procedures. The toolkit will cover the technical areas which have been addressed by QA Focus:

**Digitisation:** The digitisation of resources, including text, image, moving image and sound resources.

**Access:** Access to resources, with particular references to access using the Web.

**Metadata:** The use of metadata, such as resource discovery metadata.

**Software development:** The development and deployment of software applications.

**Service deployment:** Deployment of project deliverables into a service environment.

**Standards:** The selection and deployment of standards for use by projects.

**Quality assurance:** The development of quality assurance procedures by projects

### Using The Toolkit

The toolkit aims to provide more than a set of resources to be read passively. Instead the toolkit is meant to be used actively. The toolkit could be used, for example, within a project consortium meeting, as part of a workshop, etc.

The toolkit is available in two formats: as a set of documents (available in MS Word, PDF and HTML formats) and as an online interactive set of Web pages. The former format is intended for printing and use in workshops and the latter for individual use online.

## Standards Toolkit

The standards toolkit seeks to address the issues of your section of standards for use in your project.

<b>1. Ownership of Standard</b>	
The standard is owned by an acknowledged open standards body?	
The standard is owned by a neutral body, but not (yet) formally adopted as an open standard (e.g. Dublin Core)?	
The standard is owned by a company (i.e. a proprietary standard)	
<b>2. Openness of Proprietary Format</b>	
If the standard is proprietary:	
There is an open development process (e.g. Sun's Java)?	
The specification is published openly (e.g. Microsoft's RTF)?	
The specification has been published by third parties reverse-engineering the specification (e.g. Microsoft's Word)?	
The specification has not been published?	
<b>3 Availability of Viewers</b>	
Are viewers for the format:	
Available free of charge?	
Available on multiple platforms?	
Available as open source?	
<b>4 Availability of Authoring Tools</b>	
Are authoring tools for the format:	
Available free of charge?	
Available on multiple platforms?	
Available as open source?	

<b>5 Functionality of the Standard</b>	
Does the standard provide:	
Rich functionality?	
Basic functionality?	
<b>6 User Requirements</b>	
Does the standard:	
Largely provide the functionality required by end users of the service?	
Adequately provide the functionality required by end users of the service?	
Insufficiently provide the functionality required by end users of the service?	
Largely fail to provide the functionality required by end users of the service?	
<b>7 Fitness for Purpose</b>	
Is the standard:	
Ideal for the purpose envisaged?	
Appropriate for the purpose envisaged?	
Not particularly appropriate for the purpose envisaged?	
<b>8 Resource Implications</b>	
Will use of the standard:	
Have significant staffing implications for development and maintenance?	
Have relatively few staffing implications for development and maintenance?	
Have significant financial implications for development and maintenance?	
Have relatively few financial implications for development and maintenance?	
<b>9 Preservation</b>	
Is the format:	
Easy to preservation?	
Difficult to preserve?	
You are unsure of the ease of preservation?	



<b>10 Migration</b> Is the format:	
Easy to migrate to alternative formats?	
Difficult to migrate to alternative formats?	
<b>11 Stability of Standard</b> Does the standard change:	
Often?	
Occasionally?	
Rarely?	
<b>12 Cultural Factors</b> As well as the various technical issues addressed above, there is also a need to consider the organisational culture of the developers. For example is your organisation:	
Keen to make use of innovative developments?	
Keen to make use of mature solutions?	
Is the use of open source software prevalent in the organisation?	

## Web Site Self Assessment Toolkit

The standards toolkit seeks to address the issues of approaches to self assessment for your Web site.

<b>1. Purpose Of Your Web Site</b>	
Have you identified the purpose of your Web site. Has the purpose been agreed by all partners?	
<b>2. Standards For Your Web Site</b>	
Have you chosen the standards to be used on your Web site? Have you identified permissible exceptions?	
<b>3 Technical Architecture</b>	
Have you defined the technical architecture for your Web site? Does the chosen architecture allow you to implement your chosen standards?	
<b>4 Usability And Accessibility</b>	
Have you defined usability and accessibility policies?	
<b>5 Checking Procedures</b>	
Have you defined procedures which will ensure that your policies are complied with?	
<b>6 Resourcing Issues</b>	
Have you the resources needed to implement the above?	

## Metadata Toolkit

The metadata toolkit seeks to address quality assurance issues for use of metadata within for your project.

<b>1. Purpose Of Your Metadata</b>	
Have you a clear idea of the intended purpose of your metadata? If so, please give brief details.	
<b>2. Functionality To Be Provided By Your Metadata</b>	
Have you defined the functionality to be provided by use of metadata within your project? If so, please give brief details.	
<b>3. Metadata Standards</b>	
Have you defined the standards, specifications, schemas, etc. to be used to provide the required functionality? If so, please give brief details.	
<b>4. Metadata Modelling</b>	
Have you carried out the metadata modelling needed to provide the required functionality? If so, please give brief details.	
<b>5. Implementation Architectures</b>	
Have you defined the architecture which will be used to create, manage and deploy your metadata? If so, please give brief details.	
<b>6. Metadata Content Rules</b>	
Have you defined the rules governing the content of your metadata? If so, please give brief details.	

<b>7. Checking Procedures</b>	
Have you defined checking procedures to ensure that the metadata content and formats comply with defined rules?	
<b>8. Interoperability Issues</b>	
Have you identified possible third parties with whom your metadata may need to be interoperable? Have you engaged in communications with them to address interoperability issues?	
<b>9. Training And Staff Developments Issues</b>	
Have you developed a training strategy for staff involved in creating and maintaining your metadata.	
<b>10. Resourcing Issues</b>	
Have you the resources needed to implement the above?	

## Software Toolkit

The software toolkit seeks to address quality assurance issues for development or use of software applications within for your project.

<b>1. Purpose Of Your Software</b>	
Have you defined the functionality to be provided by the software you will develop/use within your project? Do you have a systems specification? Have you implemented change control procedures?	
<b>2. Approach To Software Development</b>	
What approaches to the development or use of software do you intend to take?	
<b>3 Use Of Third Party Software</b>	
If you are dependent on significant use of third party software, have you identified any potential problem areas?	
<b>4 Database Modelling, Design and Implementation</b>	
If you intend to use database software have you implemented procedures for the database modelling, design and implementation?	
<b>5 Software Deployment</b>	
Have you identified how software you have developed will be deployed? Have you identified any dependencies or potential difficulties?	
<b>6 Legal Issues</b>	
Have you identified any potential legal issues which may hinder the deployment of software?	
<b>7 Resourcing Issues</b>	
Have you the resources needed to implement the above?	

## Service Deployment Toolkit

The metadata toolkit seeks to address quality assurance issues for use of metadata within for your project.

<b>1. Your Intended Area For Service Deployment</b>	
Have you identified the intended recipients of your project deliverables?	
<b>2. Awareness Of The Technical Architecture</b>	
Have you documented the technical architecture so that the intended recipients are in a position to establish the suitability at an early stage?	
<b>3 Potential Areas Of Concern</b>	
Have you identified any areas of potential concern to the recipients?	
<b>4 Technical Documentation, Support, etc.</b>	
Have you documented the technical architecture, installation requirements, etc, so that the intended recipients of your project deliverables are in a position to deploy the deliverables?	
<b>5 Security, Performance, etc. Issues</b>	
Have you documented any security or performance issues which may be of concern to service deployment staff?	
<b>6 Legal, IPR, etc. Issues</b>	
Have you documented any legal or IPR issues which may be of concern to service deployment staff?	
<b>7 Resourcing Issues</b>	
Have you identified the resources needed to (a) provide the documentation, liaison, etc. and (b) deploy the deliverables?	

---

## Appendix 1 QA Focus Publications

---

The following peer-reviewed papers were produced by the QA Focus team:

- *Ideology or Pragmatism? Open Standards and Cultural Heritage Web Sites*, B. Kelly, A. Dunning, M. Guy and L. Phipps, ichim03 Proceedings (CDROM), <<http://www.ukoln.ac.uk/qa-focus/documents/papers/ichim03/>>
- *Developing A Quality Culture For Digital Library Programmes*, B. Kelly, M. Guy and H. James, EUNIS 2003 conference proceedings, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/eunis-2003/>>
- *Deployment Of Quality Assurance Procedures For Digital Library Programmes*, B. Kelly, A. Williamson and A. Dawson, IADIS Internet/WWW 2003 Conference Proceedings, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/iadis-2003/paper/>>
- *A Proposal For Consistent URIs For Checking Compliance With Web Standards*, B. Kelly, IADIS Internet/WWW 2003 Conference Proceedings, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/iadis-2003/poster/>>

In addition to these peer-reviewed publications the following articles have also been published:

- *Implementing A Quality Assurance Methodology For Digital Library Programmes*, B. Kelly, VINE, Vol. 1, No. 2, 2005, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/vine-01-2004/>>
- *Quality Assurance Procedures in the JISC 5/99 Programme*, M. Napier, Ariadne 32, July 2003, <<http://www.ariadne.ac.uk/issue32/qa-focus/>>
- *JISC QA Focus*, M. Napier and E. Bremner, VINE 126, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/vine-126/>>

# Ideology or Pragmatism?

## Open Standards and Cultural Heritage Web Sites

Brian Kelly\*, Alastair Dunning<sup>+</sup>, Marieke Guy\* and Lawrie Phipps<sup>#</sup>

\* UKOLN, University of Bath, Claverton Down, Bath, UK

+ AHDS, King's College London, London, UK

# TechDis, The Network Centre, 4 Innovation Close, York Science Park, York, UK

<B.Kelly@ukoln.ac.uk>, <Alastair.Dunning@ahds.ac.uk>,

<M.Guy@ukoln.ac.uk>, <Lawrie@techdis.ac.uk>

### Abstract

The importance of open standards for providing access to digital resources is widely acknowledged. Bodies such as the W3C are developing the open standards needed to provide universal access to digital cultural heritage resources. However, despite the widespread acceptance of the importance of open standards, in practice many organisations fail to implement open standards in their provision of access to digital resources. It clearly becomes difficult to mandate use of open standards if it is well-known that compliance is seldom enforced. Rather than abandoning open standards or imposing a stricter regime for ensuring compliance, this paper argues that there is a need to adopt a culture which is supportive of use of open standards but provides flexibility to cater for the difficulties in achieving this.

**Keywords:** open standards, proprietary formats, policies

### 1 About This Paper

The World-Wide Web is widely accepted as the key platform for providing access to digital cultural heritage resources. The Web promises universal access to resources and provides flexibility (including platform- and application-independence) through use of open standards. In practice however, it can be difficult to achieve this goal. Proprietary formats can be appealing and, as we learnt during the “browser wars”, software vendors can state their support for open standards while deploying proprietary extensions which can result in services which fail to be interoperable.

Many digitisation programmes which seek to provide access to digital cultural heritage resources will expect funded projects to comply with a variety of open standards. However if, in practice, projects fail to implement open standards this can undermine the premise that open standards are essential and would appear to threaten the return of application- and platform-specific access to resources.

Although a commitment to Web development based on open standards is desirable in practice it is likely that there will be occasions when use of proprietary solutions may be needed. But the acceptance of a mixed economy in which open standards and proprietary formats can be used as appropriate can lead to dangers. So should we mandate strict compliance with open standards or should we tolerate a mixed economy? This paper seeks to explore these issues.

### 2 An Open Standards Culture: Two Case Studies

We will now review two digital library programmes in more detail and expand on the standards framework and project monitoring and technical support services which seek to ensure that the deliverables from funded projects are interoperable and comply with appropriate standards and best practices. Three of the authors of this paper have been involved in providing the technical support services to these programmes. The experiences gained in providing this support have helped to inform the writing of this paper.

#### JISC's Learning and Teaching (5/99) Programme

Within the UK the Higher Education community has a culture which is supportive of open standards in its digitisation programmes. An early digital library programme known as eLib ran from 1995 until 2001. A set of guidelines known as the *eLib Standard Guidelines* (JISC-1) defined the standards which funded projects were expected to implement were produced.

In 1999 the Joint Information Systems Committee (JISC) established a digital library programme with the intention of improving the applicability of its collections and resources for learning and teaching. Although digital information and data resources had been created in previous JISC-funded programmes so far they had mainly been used for research and their



learning and teaching value had not been widely utilised. The JISC Learning and Teaching Programme (5/99) was aimed at increasing the use of online electronic resources by integrating them into the JISC's Information Environment through deployment into a service environment. Alongside this increased need for usage of resources in new areas was the recognition that if the digital resources created on programmes were to be widely accessible, interoperable, durable and represent value for money, their technical development should be rigorous and based on best practices. To ensure that the project deliverables could be easily deployed into a service environment the JISC expected projects to make use of standards documented in the *Standards and Guidelines To Build A National Resource* document (JISC-2) which was based on an update of the *eLib Standard Guidelines* document.

The JISC were aware that although projects funded by the eLib programme were expected to comply with the eLib standards document, in practice compliance was never checked. This may have been appropriate for the eLib programme as, when the programme commenced in 1995, it was not necessarily clear that the Web would turn out to be the killer application for delivery of resources. However there is now an awareness that the Web is the killer application for access to digital resources. There is also a realisation that compliance with standards will be necessary in order for digital resources to be widely interoperable. In response to such needs the JISC funded a new post: QA Focus. The QA Focus post was initially established as a support mechanism solely for the 5/99 programme (although recently its remit has been expected to cover additional programmes). The aim of QA Focus is to ensure that projects comply with standards and recommendations and make use of appropriate best practices by deploying quality assurance procedures.

An initial QA Focus activity was organising focus group meetings which provided feedback on the standards framework. The feedback received included: (a) a lack of awareness of the Standards document; (b) difficulties in seeing how the standards could be applied to projects' particular needs; (c) concerns that the standards would change during the project lifetime; (e) lack of technical expertise and time to implement appropriate standards; (f) concerns that standards may not be sufficiently mature to be used; (g) concerns that the mainstream browsers may not support appropriate standards and (h) concerns that projects were not always starting from scratch but may be building on existing work and in such cases it would be difficult to deploy appropriate standards.

Following the focus group meetings surveys of project Web sites were carried out in order to gain an understanding of the approaches taken by projects in their provision of project Web sites and to identify examples of best practices and areas in which improvements could be made. The surveys analysed compliance with HTML and CSS standards and with W3C WAI guidelines. The findings showed that few project entry points appeared to comply fully with open standards (QA-Focus-1). A number of reasons for this have been expressed: (a) the surveys may have analysed Web pages *about* the project, rather than the actual project Web site; (b) the surveys may have analysed Web pages aimed at project partners rather than end users; (c) the surveys may have been carried out at an early stage of development and (d) the focus of the projects deliverables may have been on digitisation or software development and on providing information on a Web site.

It should be noted that such comments appear to indicate that strict compliance with standards is felt to be difficult or that there may be occasions when compliance is not felt to be necessary or would be unnecessarily expensive to implement. These comments appear to show reservations as to the applicability or scope of compliance with open standards.

### The NOF-digitise Programme

The NOF-digitise programme (NOF-1) is the second of these case studies. Supported by public funding of about £50 million, the programme forms part of a larger initiative (the New Opportunities Fund or NOF) that distributed funding to education, health and environment projects throughout the United Kingdom, with a focus on providing for those in society who are most disadvantaged. The NOF-digitise element was, as the title suggests, was dedicated to funding and supporting universities, local government, museums and other public sector organisations in digitising material from their collections and archives and making this cultural heritage available on the Web.

Emphasis on the need for standards and good practice began early in the lifespan of the programme. This was for two reasons. Firstly, few of the funded projects had much experience of digitisation and a fair degree of education was required to inculcate the importance of standards. Secondly, it was realised that the public funding of a large-scale digitisation programme entailed the creation of material that needed to be preserved and made accessible not just in the present, but for future generations. Therefore the NOF-digitise programme elected to formulate a set of standards based on open standards. In addition a Technical Advisory Service (NOF-2) was established which would be able to offer technical assistance to the projects as they applied these standards.

The standards developed for NOF-digitise projects (NOF-3) were split into five areas: creation, management, collection development, access and re-use. In many cases defining the open standards in these areas was a relatively straightforward matter. Thus those projects that were digitising textual material needed to do so in XML or HTML; those creating digital images had to use formats such as TIFF, GIF, JPEG (JFIF) or PNG.

But almost immediately the difficulty of applying purely open standards became apparent. At the time of the standards' initial creation, there was no suitable open standard for the creation of audio or video files – thus the programme had to adopt a more pragmatic outlook, accepting formats such as MPEG4 for video and MP3 for audio. The problem was even

more acute when it came to access to data using software such as Macromedia Flash or Adobe PDF. Macromedia Flash's SWF format provides many features attractive to projects. Projects, quite rightly, considered that the ability to create stylish graphics and animations would be an important feature in attracting users to their Web sites, especially younger users. Adobe's PDF format offered presentational advantages, especially of some historical documents, that made it easier to apply than HTML.

But the use of such proprietary formats presented two problems. Firstly, in terms of accessibility: as the NOF-digitise programme was committed to being inclusive in delivering digital material, it had to take account users who would not be able to access, for example, resources created in Flash or PDF. Secondly, there were preservation issues related to the creation of material in such proprietary formats. To what extent would such material be accessible in five, ten, twenty years? Would there also be the possibility that future users would have to begin to have to pay for the plug-ins needed to access such materials?

The programme was therefore faced with the problem addressed in this paper: should it enforce strict compliance or cater for a mixed economy? It was decided to adopt a pragmatic approach. The development of resources in such proprietary formats was not forbidden but projects had to ensure that the creation of any part of the resource in a format such as PDF or SWF was accompanied by various safety checks that ensured data would become locked into solely proprietary formats. The crucial stipulation was that any significant digital resource created in a proprietary format had also to be presented in an open standard as well. Thus documents created in PDF also had to be available in HTML or XML. Extra functionality could be provided by proprietary formats but the project had to ensure that the core content was accessible in an open standard as well. The same was true for Flash – if a project was using its digitised content to create a resource in SWF the project had to ensure that key content used was available to those users without Flash. So if a project was developing a game or animation using the content they had digitised, they had to make sure that the content was available without Flash as well (even if the functionality of the game itself was not replicated in an open standard).

Additionally, users who were unable to access the Flash-based resources would have to be informed, using short notices on the project Web site, of the resources which could not be accessed. This ensured that even if a user could not access a particular part of the Web site they would not be left in the dark as to what was available to other users. This technique was also required for resources existing in other proprietary formats: audio files, for example, needed to be accompanied by textual transcriptions or descriptions of what the file contained.

Finally NOF stipulated that projects should devise migration strategies to ensure the existence of their digital material in the medium to long term. The kernel of such migration strategies was to be the continued survey of the possibility of migrating the data currently held in proprietary format to an open standard. So in the case of Flash resources, projects need to review on the possibility of transferring to, say, SMIL (W3C-1) or explore the opportunities afforded by the publication of the SWF specification (SWF).

The NOF-digitise programme has been committed to helping develop digital resources in open standards, thus increasing the chances of developing. But it has done so within a framework that has tried to understand and accommodate the advantages provided by proprietary formats.

### 3 Implementation Challenges

As we have seen communities which have expressed commitments to use of open standards are currently failing to comply with such standards. We can speculate on a number of reasons for this:

**Bad experiences with standards:** Organisations may have sought to implement standards in the past and experienced difficulties, which may have been costly. Within the UK Higher Education community those with long memories will remember the edict that the community must strive towards OSI networking protocols through use of Coloured Book software (JNT).

**Lack of awareness of standards:** There is a danger that although awareness of standards may be widespread amongst certain sectors of the Web development community, other developers may have a focus on Web development applications and not the underlying standards they support.

**Difficulties in monitoring compliance:** Even in cases in which there is an awareness of the importance of open standards and a commitment to their use we can find that Web sites fail to comply with standards. This may be due to the difficulties in monitoring compliance with standards. Compliance testing services such as W3C's HTML validator (W3C-2) and CSS validator (W3C-3) are not particularly easy to use, requiring a cumbersome manual process which is not cleanly integrated with a publishing process or scalable for validating large numbers of resources.

**Limitations of the tools:** Many authoring tools fail to comply with open standards. In addition many authoring tools fail to implement best practices, and generate deprecated features such as HTML elements used for formatting rather than using cascading style sheets to define the appearance of resources. Open source advocates argue that there are open source

authoring tools which do provide better support for open standards. However replacement of existing tools will inevitably result in hidden costs such as training and support costs.

**If it's not broken ...:** Developers of the current generation of Web sites may argue that the Web resources are accessible in the current generation of browsers. Some will argue that their Web sites have been tested across a range of browsers and operating system environments; others will point out that their Web sites have been tested under the most popular browsers and this is an adequate testing regime, especially in light of the costs of testing and the diminishing returns gained by testing under the more esoteric environments.

**Maturity of standards:** Although some organisations may welcome the opportunity to be early adopters of new standards, others may not wish to make use of new standards until they have been adequately tested and a wide range of tools which support the standards are available.

**Standards wars:** There are occasions when there are competing standards. For example the news feed syndication standards RSS has two competing standards – one based on XML (RSS-1) and one on RDF/XML (RSS-2).

**We have a problem – let's invent a new standard:** When a standard is found to have limitations, there seems to be a temptation to use this as an opportunity to develop a new standard. This can happen before the flawed standard has yet been widely deployed and is still being promoted. An example of this is XHTML 2.0. Although XHTML 1.0 provides many advantages, effective deployment is hindered by the requirement of current browsers to attempt to display resources which do not comply with standards. A recent survey has shown that many XHTML 1.0 documents are not compliant (Goer). Such document may be displayed, but as they are not valid XML documents, they cannot be processed as XML. In an attempt to address this W3C are developing XHTML 2.0 which will not be expected to be backwards compatible. This leaves Web developers uncertain whether to move from HTML to XHTML 1.0 or wait until XHTML 2.0 becomes available. Moving from HTML 4.0 to XHTML 1.0 and the XHTML 2.0 would appear to be a resource-intensive operation. As Mark Pilgrim put it “*Someday, I'll upgrade myself from 'SHOULD NOT chase after bleeding edge technologies that don't solve real world problems' to 'MUST NOT chase after bleeding edge technologies that don't solve real world problems.'*” (Pilgrim, 2002).

## 4 How Should We Proceed?

### Possible Strategies

We know the benefits which use of open standards has to offer. But, as we have seen, many organisations are simply not complying with open standards such as HTML and there are a number of reasons why this is the case. So what should we be doing? Possible strategies include:

**Lobbying For Open Standards:** The traditional approach is to attempt to argue more persuasively more open standards. This is the approach taken by the Web Standards Project (WaSP) which acts as a lobby organisation for use of W3C standards. W3C itself has set up a Quality Assurance (QA) activity (W3C-4). As well as addressing the QA for W3C standards, this group is also promoting use of open standards and provides access to a library of resources on Web Site Quality (W3C-5).

**Name And Shame:** The advocacy activities of such bodies are complemented by bodies which survey various communities and publish their findings, often highlighting examples on non-compliance. For example Marko Karppinen's has surveyed of W3C member organisations home pages (Karppinen, 2003) and Business2WWW has carried out surveys of UK Local Authority Web site (Business2WWW-1) and e-Government Web sites (Business2WWW-2).

**Stricter Guidelines And Enforcement:** Another approach to the lack of failure to comply with open standards is to provide stricter guidelines and to mandate compliance with guidelines. For example the New Zealand Government Web Guidelines (New Zealand) state that “*The primary format for all content available on government websites must be HTML. The HTML must validate to the HTML 4.01 Transitional specification or earlier HTML specifications.*”

Although the New Zealand Government Web Guidelines have formal requirements for compliance with standards the document gives no indication of measures for assessing compliance. Some organisations are developing self-assessment toolkit approaches. In the UK the Government is developing a proforma to be used by local government bodies to document their compliance with appropriate standards (UK, 2003) which requires organisations to state their compliance with the Government Interoperability Framework, the Guidelines for UK Government Web sites and with W3C WAI guidelines.

### The Problems With Mandating Compliance

The approaches of providing greater encouragement to comply with open standards or of mandating compliance do not address many of the difficulties which have been outlined previously. Such approaches do not take into account conflicts within standards organisations, the dangers facing earlier adopters, the resource implications in deploying new tools, etc.

It should also be pointed out that we may see developments in the marketplace in response to the needs of the community which open standards seek to address. For example:

**Need to define DOCTYPE:** HTML standards mandate that compliant HTML documents use a DOCTYPE to define the version of HTML used. In practise, however, Web browsers can render documents which do not have a DOCTYPE. In addition, other tools, such as search engine robots, transformation tools, etc. are capable of processing documents which do not have a DOCTYPE. In light of the vast numbers of documents which do not contain a DOCTYPE one could argue that heuristics approaches can be taken to compensate.

**Need to define Character Encoding:** HTML standards mandate that compliant HTML documents define the character encoding of characters used. As mentioned above, vast numbers of documents do not define the character encoding used and one could argue that heuristics approaches can be taken to compensate.

**Need to use relative sizing:** W3C WAI guidelines require HTML elements to be defined using relative positioning and sizes. This is to enable visually impaired readers to resize resources to an appropriate size. In practice accessibility aids are available which will allow users to resize not only text on a Web page but everything on the computer display.

It may be argued that there is a proven difference between real world standards which, for example, require an electric plug to be of a particular size and characteristics in order to function correctly. In an IT environment it is possible for software to compensate for deviations from standards. One should avoid taking this example too far: it is not intended to argue that any proprietary formats or deviation from a standard can or should be processed correctly. The point being made is that in today's Web environment a great many resources do not comply with standards and yet the services are functional.

### Guidelines or Stealth Standards?

There is a need to address the applicability of mandating strict compliance in 'softer' areas such as accessibility. Widespread access to digitised resources has been important for the two case studies described in Section 2. However the implementation of Web accessibility has led to much discussion initially focussed on the accessibility of proprietary formats. However there is a wider issue which needs to be addressed: whether Web accessibility guidelines are regarded as a formal standard or guidelines which provide sensible suggestions in many cases, but which can be interpreted and applied on a case-by-case basis.

The recognition of discriminatory practices in society has led to a range of initiatives and legislation to prevent disabled people being treated unfairly. As information systems such as the Internet developed guidelines such as the W3C WAI Content Accessibility Guidelines (W3C-6) have been established to help developers ensure that their methods and materials were not excluding disabled people.

However the authors argue that these guidelines should be treated as exactly that: a set of guiding principles, rather than absolute and fixed standards. They are not and cannot be 'hard and fast' rules because of the very nature of the community they wish to serve, which is diverse and sometimes has conflicting needs, for example dyslexic users of the Web may use a very visual interface and prefer very rich multimedia Web sites, whilst this is of less use to a blind user. This is not to say that the rich multimedia site cannot be made accessible, just that the designers may not be able to please everybody, and though the guidelines have a caveat that if something is not accessible then there should be an alternative, it is often the case that even if the content is dynamically driven and placed in a user's own interface, it is a different experience to that envisaged by the originator.

In addition to the problems of having a set of guidelines that services such a diverse community, are the 'subjective' and 'user' checks that are needed to claim adherence to the guidelines. These relate to a range of issues such as plain and simple language, the use of appropriate colour and style. All of which can be perceived differently, not only by developers but also users. However, without these elements to the guidelines they become much less effective, a mere technical shadow of the purpose for which they were envisaged; an inclusive user experience.

There is now a trend to cite these guidelines as standards, or at least use them as the basis; in the UK an industry based group the Digital Content Forum as recently put together an 'Industry Action Group' to look at Web accessibility standards in the UK. Their co-chair commented: *"Despite the talk, there is currently little genuine understanding of accessibility related issues in the UK Web design community. And worryingly, there is even less practical experience of building sites that meet the highest recognised standards in accessibility. Our first objective is therefore to widen understanding within the industry of the relevant standards that already exist, and then to foster a shared approach to overcoming the technical issues relating to making existing Web-based technologies meet these standards."*

Whilst the rhetoric is about ensuring accessibility, there is a worrying sub-text, that of the 'recognised standards' and meeting standards. At best, this approach suggests a misunderstanding of the use of the guidelines, at worst there is a worry that the community for which the guidelines were written are now being taken over and rewritten in a form that is more suitable for a standards-driven technical community. Guidelines are in place for a reason, they are a guide only, and recognise that there is a diverse set of needs for users - not a standard that can be used as a 'one size fits all', and certainly not a standard that is developed and imposed. If there is to be a standard in this area, then it is essential that the community and not industry in isolation drive it. Furthermore, until there is (or if there is) a standard the use and abuse of the term should be treated with the utmost caution, lest a 'stealth standard' is imposed before we notice.

### 5 An Alternative Approach: An Open Standards Culture

If a simple commitment to use of open standards is difficult to implement and an abandonment of open standards will lead to difficulties in providing universal access to resources, what should we do? The solution advocated in this paper is based on a developmental approach which recognises the desirability of supporting open standards, but the difficulties in doing so. The approach recognises that developers are constrained by a wide range of factors, such as resources, expertise, timescales and organisational culture. Rather than mandating a single approach for all, it is proposed that digitisation programmes should recognise such complexities, but rather than abandoning a commitment to open standards, provide a developmental culture which is supportive of open standards but does not mandate open standards in all cases. This approach has grown from our experiences in supporting the NOF-digitise and JISC 5/99 programmes.

#### A Matrix Approach

On reflection it would appear that an approach based on simply advocating use of open standards is not necessarily desirable. It is felt that there are several factors which need to be addressed, which are listed in the following table.

Area	Comments
Ownership	Is the standard owned by a recognised neutral open standards body or by a company.
Development process	Is there is community process for development of a proprietary standard.
Availability	Has the proprietary standard has been published openly or reverse-engineered.
Viewers	Are viewers (a) available for free, (b) available as open source and (c) available on multiple platforms
Authoring tools	Are authoring tools (a) available for free, (b) available as open source and (c) available on multiple platforms
Fitness for purposes	Is the standard appropriate for the purpose envisaged
Resource implications	What are the resource implications in making use of the standard?
Complexity	How complex is the standard?
Interoperability	How interoperable is the standard?
Ease of service deployment	How easy will it be to deploy the deliverable in a service environment?
Ease of long term preservation	Is the standard suitable for long term preservation?
Organisational culture	Is the organisational cultural appropriate for use of the standard?
Approaches to migration	What approaches can be taken to migrating to more appropriate standards in the future?
Approaches to assessing compliance	What approaches can be taken to measuring compliance?

**Table 1: A Matrix For Use When Choosing Standards**

#### A QA Approach

This matrix approach can be supported by appropriate quality assurance (QA) procedures. The QA approach requires provision of documentation on the policies regarding the standards to be implemented, the architecture use to implement

the standards, compliance measures to ensure that policies are correctly implemented which may include audit trails providing details of compliance. An example of a QA policy is illustrated below.

<p><b>Policy On Standards For QA Focus Web Site</b></p> <p><b>Area:</b> Web</p> <p><b>Policy:</b> The Web site will be based on XHTML 1.0.</p> <p><b>Justification:</b> Compliance with appropriate standards should ensure that access to Web resources is maximised and that resources can be repurposed using tools such as XSLT.</p> <p><b>Exceptions:</b> Resources which are derived automatically from other formats (such as MS PowerPoint) need not comply with standards. In cases where compliance with this policy is felt to be difficult to implement the policy may be broken. However in such cases the project manager must give agreement and the reasons for the decision must be documented.</p> <p><b>Compliance measures:</b> When new resources are added to the Web site or existing resources update the , validate tool will be used to check compliance. A complete compliance survey will be carried out quarterly.</p> <p><b>Audit trail:</b> Reports from the monthly audit will be published on the Web site in order to monitor trends.</p>
---

**Figure 1: QA Policy For QA Focus Web Site**

This approach has been developed by QA Focus and is documented at (QA-Focus-2).

## 7 Implementing This Approach

Three of the authors of this paper are involved in providing support services for JISC and NOF-digitise programmes. Our work will include making recommendations for support work in future programmes. Our recommendations are linked to include the following.

### **Providing Information On Standards And Technical Architectures In Bidding Processes**

We will recommend that future programme calls require bids to include information on the standards to be used by the project and the technical architecture which will be used to support the standards.

### **Providing Information On Quality Assurance In Bidding Processes**

We will recommend that future programme calls require bids to include information on their quality assurance procedures which will ensure that projects comply with their stated policies.

### **Providing Information On Compliance With Standards In Project Reports**

We will recommend that future programme calls require projects to provide details of their compliance with their stated policies in periodic reports to funders. This may include information on their approaches to self-assessment on compliance with standards and best practices including deviance from agreed standards and best practices.

### **Providing Information On Service Deployment**

We will recommend that future programme calls require projects to provide information on the expected service delivery platform for their project deliverables.

## 8 Conclusions

The importance of use of open standards is widely recognised within the cultural heritage sector. However in practice many digital cultural heritage Web resources fail to comply with open standards. On consideration of the reasons for this it would appear to be counter productive merely to impose greater pressure on developers to comply with standards. Rather there is a need to ensure that players within the community have an understanding of the importance of open standards but also have some degree of flexibility to provide access to resources which acknowledges the challenges in implementing fully compliant services. This paper provides a model based on the deployment of documented quality assurance processes, self assessment and liaison with funders which encourages a standards-based approach while still allowing the flexibility needed to allow for the complexities of Web development.

## References

- Business2WWW-1** Local Government website improves with SiteMorse, Business2WWW, <[http://www.business2www.com/news\\_article.html?news\\_start=4&news\\_current=5909&news\\_all=>](http://www.business2www.com/news_article.html?news_start=4&news_current=5909&news_all=>)

<b>Business2WWW-2</b>	'e-venturing.gov.uk' worst of tested government websites., Business2WWW, < <a href="http://www.business2www.com/news_article.html?news_start=4&amp;news_current=5708&amp;news_all=">http://www.business2www.com/news_article.html?news_start=4&amp;news_current=5708&amp;news_all=</a> >
<b>Goer</b>	The XHTML 100, April 2003 archives, goer.org, < <a href="http://www.goer.org/Journal/2003/Apr/index.html#29">http://www.goer.org/Journal/2003/Apr/index.html#29</a> >
<b>JISC-1 (1998)</b>	eLib Standards Guidelines, < <a href="http://www.ukoln.ac.uk/services/elib/papers/other/standards/">http://www.ukoln.ac.uk/services/elib/papers/other/standards/</a> >
<b>JISC2 (2001)</b>	Standards and Guidelines to Build a National Resource, < <a href="http://www.jisc.ac.uk/index.cfm?name=projman_standards">http://www.jisc.ac.uk/index.cfm?name=projman_standards</a> >
<b>JNT</b>	A Brief History of the Joint Network Team 1979 - 1994, < <a href="http://www.cms.brookes.ac.uk/modules/web_tech/p08770/archives/jnt.htm">http://www.cms.brookes.ac.uk/modules/web_tech/p08770/archives/jnt.htm</a> >
<b>Karppinen (2003)</b>	Steady As She Goes, Marko Karppinen, < <a href="http://www.markokarppinen.com/20030224">http://www.markokarppinen.com/20030224</a> >
<b>New Zealand</b>	New Zealand Government Web Guidelines Version 2.1, < <a href="http://www.e-government.govt.nz/docs/web-guidelines-2-1/">http://www.e-government.govt.nz/docs/web-guidelines-2-1/</a> >
<b>NOF-1</b>	NOF-digitise, < <a href="http://www.nof-digitise.org/">http://www.nof-digitise.org/</a> >
<b>NOF-2</b>	NOF-digitise Technical Advisory Service, < <a href="http://www.ukoln.ac.uk/nof/support/">http://www.ukoln.ac.uk/nof/support/</a> >
<b>NOF-3</b>	NOF-digitise technical standards and guidelines, < <a href="http://www.peoplesnetwork.gov.uk/content/ts_index.asp">http://www.peoplesnetwork.gov.uk/content/ts_index.asp</a> >
<b>Pilgrim (2002)</b>	A Warning To Others, Dive Into Mark, Thursday, November 21, 2002, < <a href="http://diveintomark.org/archives/2002/11/21/a_warning_to_others.html">http://diveintomark.org/archives/2002/11/21/a_warning_to_others.html</a> >
<b>RSS-1</b>	RSS 2.0, < <a href="http://backend.userland.com/rss">http://backend.userland.com/rss</a> >
<b>RSS-2</b>	RDF Site Summary (RSS) 1.0, < <a href="http://web.resource.org/rss/1.0/">http://web.resource.org/rss/1.0/</a> >
<b>SWF</b>	OpenSWF.org, < <a href="http://www.openswf.org/">http://www.openswf.org/</a> >
<b>UK (2003)</b>	Implementing Electronic Government Return 20022003 (IEG3), < <a href="http://www.localegov.gov.uk/Nimoi/sites/ODMP/resources/IEG3%20Draft%20Guidance%20for%20consultation%20-%20May%202003.doc">http://www.localegov.gov.uk/Nimoi/sites/ODMP/resources/IEG3%20Draft%20Guidance%20for%20consultation%20-%20May%202003.doc</a> >
<b>WaSP</b>	The Web Standards Project, < <a href="http://www.webstandards.org/">http://www.webstandards.org/</a> >
<b>QA-Focus-1 (2002)</b>	QA Focus Surveys Of Project Web Sites - October 2002, < <a href="http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/">http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/</a> >
<b>QA-Focus-2 (2003)</b>	Developing A Quality Culture For Digital Library Programmes, B. Kelly, M. Guy and H. James, EUNIS 203 Conference Proceedings, < <a href="http://www.ukoln.ac.uk/qa-focus/documents/papers/eunis-2003/">http://www.ukoln.ac.uk/qa-focus/documents/papers/eunis-2003/</a> >
<b>W3C-1</b>	SMIL, W3C, < <a href="http://www.w3.org/AudioVideo/">http://www.w3.org/AudioVideo/</a> >
<b>W3C-2</b>	W3C Markup Validation Service, W3C, < <a href="http://validator.w3.org/">http://validator.w3.org/</a> >
<b>W3C-3</b>	CSS Validator, W3C, < <a href="http://jigsaw.w3.org/css-validator/">http://jigsaw.w3.org/css-validator/</a> >
<b>W3C-4</b>	W3C - Quality Assurance, W3C, < <a href="http://www.w3.org/QA/">http://www.w3.org/QA/</a> >
<b>W3C-5</b>	The W3C QA Library, W3C, < <a href="http://www.w3.org/QA/Library/#eo">http://www.w3.org/QA/Library/#eo</a> >
<b>W3C-6</b>	Web Content Accessibility Guidelines 1.0, W3C, < <a href="http://www.w3.org/TR/WCAG10/">http://www.w3.org/TR/WCAG10/</a> >

### About The Authors

**Brian Kelly** provides the JISC-funded UK Web Focus post, which provides support for the UK Higher and Further Education communities on Web issues. He is also the project manager for QA Focus and the NOF Technical Advisory Service. He is based in UKOLN – a national centre of excellence in digital information management, based at the University of Bath.

**Marieke Guy** is a member of the QA Focus team which supports JISC's Information Environment programme by ensuring that funded projects comply with standards and recommendations and make use of appropriate best practices. Marieke has previously worked as a NOF-digitise Advisor and co-ordinated technical support and advice services to the NOF national digitisation programme. Marieke is based in UKOLN – a national centre of excellence in digital information management, based at the University of Bath.

**Alastair Dunning** works at the Arts and Humanities Data Service (AHDS) as Communications Manager, responsible for the AHDS's publications, Web site, events and training workshops. The AHDS aids the discovery, creation and preservation of digital collections in the arts and humanities. Alastair has also been working as a member of the NOF-digitise Technical Advisory Service, advising many projects on standards and good practice in the creation and dissemination of their digital resources.

**Lawrie Phipps** is a Senior Advisor (Higher Education) to the JISC TechDis service which aims to enhance provision for disabled students and staff in further and higher education through technology. Lawrie's background is in learning technology as a developer and he retains a research interest in virtual fieldwork. Currently he is working on e-learning and accessibility, and how mobile computing can be of benefit to disabled students.



## Developing A Quality Culture For Digital Library Programmes

Brian Kelly\*, Marieke Guy<sup>+</sup> and Hamish James<sup>#</sup>

\* UKOLN, University of Bath, BATH, UK,  
<B.Kelly@ukoln.ac.uk>

+ UKOLN, University of Bath, BATH, UK,  
<M.Guy@ukoln.ac.uk>

# AHDS, Kings College London, UK,  
<Hamish.James@ahds.ac.uk>

### Abstract

*In this paper the authors describe approaches for the development of quality assurance (QA) procedures for a digital library programme. The authors argue that QA procedures are needed in order to ensure that deliverables from digital library programmes will be interoperable and give be easily deployed and repurposed. The adoption of open standards is acknowledged as essential in digital library programmes but in a distributed development environment it can be difficult to ensure that programme deliverables actually implement appropriate standards and best practices. The authors describe the approaches to the development of a quality culture based on encouragement of use of QA by project holders in one digital library programme which is funded by the JISC in the UK.*

**Keywords:** quality assurance, QA

### 1 Background

The World-Wide Web is accepted as the key delivery platform for digital library services. The Web promises universal access to resources and provides flexibility, including platform- and application-independence, though use of open standards. In practice however, it can be difficult to achieve this goal. Proprietary formats are appealing and, as we learnt during the “browser wars”, software vendors can promise open standards while deploying proprietary extensions which can result in services which fail to be interoperable. Developers can be unsure as to which standards are applicable to their area of work: there is a danger that simple standards, such as HTML, are used when richer standards, such as XML, could provide greater interoperability.

The JISC (Joint Information Systems Committee) has funded a QA Focus post which aims to ensure that projects make use of QA (quality assurance) procedures which will help ensure interoperability through use of appropriate standards and best practices.

A summary of the work of QA Focus is provided in this paper. The paper describes the background to IT development in the UK’s Higher Education community, the role of standards and the approaches taken by QA Focus. The paper concludes by outlining future work for QA Focus and the potential for use of similar approaches by other digital library programmes.

### 2 IT Development Culture

The UK’s Higher Education community has a culture which is supportive of open standards in its IT development programmes. Within the eLib Programme, for example, the eLib Standards Guidelines [1] defined the standards funded projects were expected to implement.

Although the Standards Guidelines document was available shortly after the start of the programme, compliance was not enforced. There was recognition of the dangers of enforcing standards too rigidly in those early days of the Web: if the programme had started a few years earlier use of Gopher could well have been chosen as the standard delivery mechanism! In addition the UK Higher Education community had previously attempted to standardise on Coloured Books networking protocols, which subsequently failed to be adopted widely and were eventually superseded by Internet protocols.

The eLib programme encouraged a certain amount of diversity: this approach of letting a “thousand flowers bloom” was probably appropriate for the mid-1990s, before it was clear that the Web would be the killer application which, with hindsight, we recognise that it is. This approach also reflected the culture of software development in a HE environment, in which strict management practices aren’t the norm and there has been a tendency to allow software developers a fair amount of freedom.

Nowadays, however, there is increased recognition of the need to have a more managed approach to development. The Web is now recognised as the killer application. Project deliverables, which are often Web-based, can no longer be treated as self-contained services – there is a need for them to interoperate. Also stricter compliance with standards will be needed: Web browsers have been tolerant of errors in HTML resources, but this will be different in a world in which “Web Services” technologies will be reliant on well-structured resources for machine processing. Finally, JISC has moved on from a research and experimental approach and is now funding programmes in which project deliverables are normally expected to be deployed in a service environment.

### 3 The JISC Information Environment

The JISC's Information Environment (IE, formerly DNER) [2] seeks to provide seamless access to scholarly resources which are distributed across a range of providers, including centrally-funded JISC services, commercial providers and the institutions themselves. The *Standards and Guidelines To Build A National Resource document* [3] was written to define the standards which form the basis for the IE. The standards document is supported by an IE Architecture [4] which describes the technical architecture of the IE.

The JISC has funded a number of programmes in order to develop the IE, including 5/99 [5] which was followed by the FAIR [6] and X4L [7] programmes.

### 4 The QA Focus Post

JISC has recognised that there is a need for the JISC-funded programmes to be supported by a post which ensures that projects comply with standards and best practices. The QA Focus post has been funded for two years (from 1 January 2002) to support the JISC 5/99 programme. Initially the post was provided by UKOLN (University of Bath) and ILRT (University of Bristol), but, following a decision to refocus on other areas, in January 2003 ILRT were replaced by AHDS (Arts and Humanities Data Service).

### 5 Approaches To QA

QA Focus aims to provide a support service to 5/99 projects: the emphasis is on advice and support, based on close links with the projects, rather than a policing role. An important deliverable will be the development of a self-assessment toolkit which can be used by the projects themselves for validation of the project deliverables.

QA Focus is addressing a range of technical areas which include digitisation, Web (including accessibility), digitisation, metadata, software development and service deployment.

The areas of work which are being carried out by QA Focus include:

- Providing advice on standards and best practices.
- Carrying out surveys across projects, looking at compliance with standards and best practices.
- Commissioning case studies which provide examples of best practices.
- Providing documentation on best practices, approaches to compliance checking, etc.
- Developing a Self-Assessment Toolkit

Although QA Focus places an emphasis on its role in supporting projects in developing their own QA procedures, in cases of severe interoperability problems QA Focus will be expected to make contact with the project concerned and seek to ensure that concerns are

addressed. If this does not result in a satisfactory solution, the issue will be passed on to the JISC.

### 6 QA Focus Work To Date

#### 6.1 Links With Projects

A number of workshop sessions have been held with a selection of the projects. The first two workshops aimed to obtain feedback from the projects on (a) the Standards document, (b) implementation experiences and (c) deployment of project deliverables into a service environment.

The workshops provided valuable feedback which has helped to identify key areas which need to be addressed. Useful information was obtained about the Standards document including a lack of awareness of the standards document in some cases, concerns over the change control of the document (since new standards may be developed and other standards may fail to gain acceptance) and some uncertainties as to the appropriateness of some of the standards and deployment difficulties in other cases, especially projects which were reliant on third party software development of existing systems which cannot easily be modified. The feedback on implementation experiences raised several predictable issues, including the poor support for Web standards in many widely-used browsers. The lack of a technical support infrastructure was highlighted by several projects, mainly those based in academic departments or in smaller institutions.

#### 6.2 Surveys

A meeting of 5/99 projects was held at the University of Nottingham, on 30 October - 1 November 1 2002. Prior to the meeting QA Focus carried out a survey of various aspects of 5/99 project Web sites. The survey findings [8] were made available and formed the basis for discussions at the QA Focus workshop sessions.

The surveys made use of a number of freely available tools, all of which had a Web-interface. This meant methodology was open and tools could be used by projects themselves without the need to install software locally. The survey findings were published openly. This allowed examples of best practices to be seen, trends to be monitored and areas which projects found difficult to implement to be identified.

The surveys were complemented by a number of brief advisory documents. In addition a number of case studies have been commissioned which allows the projects themselves to describe their approaches to compliance with standards and best practices and any difficulties they have experienced and lessons they have learnt.

Survey	Tool	Information
HTML Compliance	W3C's HTML validator	Does the home page comply with HTML standards? What DTDs are used?

CSS Compliance	W3C's CSS validator	Does the home page use CSS? Does the CSS comply with standards?
Accessibility	Bobby	Does the home page comply with W3C WAI guidelines?
404 page	Manual observation	Does the 404 page provide navigational facilities and support?
Internet Archive	Manual observation	Is the Web site available in the Internet Archive?
PDA Access	AvantGo	Can the Web site be accessed by a PDA?
XHTML Conversion	W3C's Tidy tool	Can the Web site be converted to XHTML without loss of functionality?
WML Conversion	Google WAP conversion service	Can the Web site be converted to WML without loss of functionality?
HTTP Headers	Dundee's HTTP analysis tool	Are correct HTTP headers sent? What Web server environment is used?
Metadata	W3C's Tidy and RDF validator & UKOLN's DC-dot tools	Is Dublin Core metadata used? Does it comply with standards?

**Table 1: Initial QA Focus Surveys**

The surveys aimed to establish how well project Web sites complied with standards and best practices. The surveys addressed several areas related to Web technologies including compliance with HTML and CSS standards and compliance with W3C's Web Accessibility Initiatives (WAI) guidelines for project entry points. The HTTP Headers were analysed and details of the Web server platform recorded (together with details of invalid HTTP headers).

As well as testing compliance with well-defined standards the survey also used a number of tools which helped to see if the Web sites allowed repurposing. This included checking availability of project Web sites in the Internet Archive, using the AvantGo service to test access to project Web sites on a PDA and converting the Web site to WML and viewing in the Opera browser which provides a WAP emulator.

The survey also used a simple usability test by reporting on the approach taken to the Web site 404 error page: whether the 404 error page was branded, provided helpful information and appropriate links, etc.

Metadata embedded in project Web site entry points was tested and any Dublin Core metadata found was validated using a Dublin Core validation tool developed in UKOLN. In addition the Dublin Core metadata was converted to RDF format and then visualised allowing an alternative display of the metadata to be viewed.

### 6.3 Limitations of Methodology

There is a danger that the publication of the findings can be perceived as threatening to projects. Where the findings indicate lack of compliance with standards or failure to implement best practices projects may point out particular features of their project which the surveys fail to acknowledge, limitations of the tools used the timing of the survey and the available resources.

There is an element of truth in such concerns. The projects are addressing a diverse set of areas, including digitising content, enhancing existing services and software development. The project Web sites will also have a diverse set of objectives, including providing communications with project partners, providing information about the project and providing access to project deliverables. The projects will have different levels of funding, start and completion dates and technical expertise.

Despite these reservations it is felt that significant benefits can be gained from the QA Focus approach. The openness seeks to facilitate dialogue with projects and sharing of best practices. The approach also takes what can be perceived as a dry standards document and places it more centrally in the activities of the projects. It also helps to provide feedback on the standards; if a particular standard has not been adopted this may indicate that the standard is too esoteric or a lack of tools or expertise. Such considerations can be fed back to the authors of the Standards document.

### 6.4 Documentation

An important role of QA Focus is to ensure that appropriate documentation is provided for the projects. The approach that has been taken is to produce short advisory documents which address specific problems. This approach has the advantage that documents can be written more quickly and can be easily updated.

A summary of the documents published to date is given in Table 2. The documents can be accessed at [9].

Document	Area
<i>Checking Compliance With HTML and CSS Standards</i>	Summarises a number of approaches for checking that HTML resources comply with HTML and CSS standards
<i>Use Of Automated Tools For Testing Web Site Accessibility</i>	Describes tools such as Bobby and summarises the implications of common problem areas
<i>Use Of Proprietary Formats On Web Sites</i>	Provides suggestions for techniques when using common proprietary formats
<i>404 Error Pages On Web Sites</i>	Describes ways of providing user-friendly 404 error pages
<i>Accessing Your Web Site On A PDA</i>	Describes an approach for making a Web site available on a PDA

<i>Approaches To Link Checking</i>	Describes approaches for link-checking, including links to CSS & JavaScript files
<i>Search Facilities For Your Web Site</i>	Describes different approaches for providing search facilities on project Web sites
<i>Enhancing Web Site Navigation Using The LINK Element</i>	Provides advice on use of the HTML <link> element to provide enhanced Web site navigation
<i>Image QA In The Digitisation Workflow</i>	Provides advice on QA for images
<i>What Are Open Standards?</i>	Gives an explanation of open standards
<i>Mothballing Your Web Site</i>	Provides advice on "mothballing" a Web site, when funding ceases
<i>How To Evaluate A Web Site's Accessibility Level</i>	Describes approaches for checking Web accessibility

**Table 2: QA Focus Advisory Documents**

The advisory documents are complemented by case studies which are normally written by the project developers themselves. The case studies provide a solution to the common request of "Can you tell me exactly what approaches I should be using?". It is not possible to provide answers to this question as there are many projects, addressing a range of areas and with their own background and culture. It is also not desirable to impose a particular solution from the centre. The case studies allow projects to describe the solution which they adopted, the approaches they took, any problems or difficulties they experienced and lessons learnt.

The case studies which have been published to date include:

- *Managing And Using Metadata In An E- Journal*
- *Standards and Accessibility Compliance in the FAILTE Project Web Site*
- *Managing a Distributed Development Project: The Subject Portals Project*
- *Creating Accessible Learning And Teaching Resources: The e-MapScholar Experience*
- *Standards for e-learning: The e-MapScholar Experience*
- *Gathering Usage Statistics And Performance Indicators: The NMAP Experience*
- *Using SVG In The ARTWORLD Project*
- *Crafts Study Centre Digitisation Project - and Why 'Born Digital'*
- *Image Digitisation Strategy and Technique: Crafts Study Centre Digitisation Project*
- *Standards and Accessibility Compliance for the DEMOS Project Web Site*
- *Implementing a Communications Infrastructure*

- *Usability Testing for the Non-Visual Access to the Digital Library (NoVA) Project*

Access to these documents is available at [10].

## 7 Next Steps

Once the QA Focus work in the Web area has been finalised work will move on to a number of other areas including digitisation, multimedia, metadata, software development and deployment into service.

The initial work carried out by QA Focus made use of automated tools to monitor compliance with standards and best practices. In the areas listed above there will be a need to address the use of manual QA processes as well as use of automated tools. For example, the use of correct syntax for storing metadata can be checking using software, but ensuring that textual information is correct cannot be done using only automated processes.

As well as providing advice and support for the projects, QA Focus will also provide advice to JISC on best practices for the termination of the programme and for setting up new programmes. This will include development of FAQs (along the lines of those which have been developed by UKOLN to support its role in providing the Technical Advisory Service for the nof-digitise programme [11]).

## 8 Digitisation

Digitisation is the first stage in the creation of a resource, and it represents the link between the analogue and digital worlds. The consequences of poor quality digitisation will flow through the entire project, reducing the value of all later work.

QA for digitisation is therefore very important, but, with the exception of the digitisation of bitmap images [12] there is relatively little advice and support that is accessible to the non-specialist. QA Focus will provide QA guidance for image digitisation, but will also deal with other types of material including text, audio and moving images. We will also link the process of capturing data to the next step, organising the data once it is in digital form, by providing QA for databases and XML applications in particular.

### 8.1 Workflow

Digitisation typically has some of the qualities of a production line with analogue originals being retrieved, digitised and returned while digital files are created, edited and stored. Rigorous procedures can make sure that this process goes smoothly ensuring that originals are not missed, or mislaid, and the status of digital files (particularly what post-processing has occurred to them and which original or originals they relate to) is tracked. This type of quality assurance for the digitisation workflow is well established for images, and we aim to provide analogous advice for projects digitising other types of material, including checklists and model procedures to follow.

Ensuring consistent quality, and keeping records that demonstrate this, is a vital part of digitisation that indirectly affects interoperability by ensuring, that however the final resource is accessed, users can make informed use of it.

Structured metadata provides a useful mechanism for recording aspects of the digitisation process. QA Focus will review relevant existing and emerging standards. We will also investigate tools for the semi-automatic or automatic creation of technical metadata about digitised material.

### 8.2 Fitness for Purpose

Before any material is digitised, projects need to define their requirements for the digitised material. QA Focus advocates that projects' take active responsibility for these decisions and avoid allowing the capabilities of available technology set these decisions.

A key part of QA for digitisation is the development of objective, measurable criteria for judging if the digitised material is 'fit for purpose'. Determining what is fit for purpose involves consideration of the acceptable level of accuracy in digitisation in relation to the intended purpose of the digitised material. For example, a low resolution image may be suitable for a Web page, but a product also available on CD-ROM could include higher resolution images. Very similar situations occur with the digitisation of audio and moving images, but we will also address less obviously similar situations, such as rules for the standardisation of place names or the transliteration of text during transcription.

### 8.3 Rights

Digital files are easily copied and distributed, so it is important for projects to ensure that they have obtained any necessary rights to use the originals. Projects may also want to protect their own rights in the digitised material.

Intellectual Property law is a complex area and QA Focus will not be able to provide definitive answers, but we hope to produce a series of case studies that demonstrate how a project can best minimise the risk of running afoul of copyright infringement. We will liaise with JISC's Legal Information Service [13] which has expertise in this area.

## 9 Metadata

Metadata has a key role to play in ensuring the projects deliverables can be interoperable. However unless QA procedures are deployed which ensure that the metadata content is correct, the metadata is represented in an appropriate format, complies with appropriate standards and can be processed unambiguously we are likely to encounter difficulties in service deployment.

While resource discovery metadata is central to interoperability, we will also investigate requirements for workflow, technical and rights metadata that support the digitisation process and deployment into service.

We are currently planning focus group sessions in which we will obtain feedback from groups with experience in metadata activities. This should provide us with examples of the type of approaches which can be recommended in order to ensure that metadata is interoperable.

Approaches we are currently considering include:

- Checking syntax, encoding, etc. for metadata embedded in HTML and XML resources. This may include documenting the methodology employed in the survey of Dublin Core metadata embedded in project home pages [14] and employing use of XSLT [15].
- Ensuring that the metadata deployed is appropriate for the purposes for which it will be used.
- Ensuring projects have appropriate cataloguing rules for their metadata and processes in place for implementing the rules and monitoring compliance.
- Ensuring that metadata can interoperate with third party services.
- Using techniques for checking metadata such as use of spell-checkers, checking against lists of controlled vocabularies, etc.

The QA procedures will be applied to metadata which is used in various ways including metadata embedded in HTML and XML resources, OAI metadata, educational metadata, RSS newsfeeds, etc.

A case study which describes the use of metadata in an e-journal, including details of the metadata elements used, the purpose of the metadata, the architecture for managing the metadata and the limitations of the approach has been published [16].

## 10 Software Development

QA is crucial in the development of quality software. It is fundamental to the entire software development process from the initial systems analysis and agreement on standards through to problem handling and testing and software deployment. Once established, QA processes form a thread through the software development lifecycle and help developers focus on possible problem areas and their prevention.

### 10.1 Development

Before the onset of a software development project the project team should produce a detailed set of specifications that document what exactly the software will do. Questions need to be asked about the purpose of the software and whether this purpose reflects the requirements of the user.

QA Focus will be providing case studies and briefing papers on these areas. Consideration of one possible design process for recording specific software development requirements, Unified Modelling Language (UML), is given in a case study provided by the Subject Portals Project [17].

### 10.2 Documentation

QA Focus will be providing advice on standards for software documentation, both public and internal. Having clear documentation is especially important in a digital library programme in which short term contracts and high staff turnover are the norm [18]. In the long term good documentation can improve usability, reduce support costs, improve reliability and increase ease of maintenance. Throughout a project's lifetime information should be recorded on the software environment a package has been developed in, language systems used and the libraries accessed.

Project teams will need to agree on standards used when writing software code. This should be done prior to development. QA Focus have produced a briefing paper which provides advice on how projects do this [19].

In the later stages of development work user documentation may be required. Writing documentation is a useful process that can show up bugs which have been missed in testing. Ideally the documentation writers are a different team of people from the developers and provide a different perspective on the software.

### 10.3 Testing

A software product should only be released after it has gone through a proper process of development, testing and bug fixing. Testing looks at areas such as performance, stability and error handling by setting up test scenarios under controlled conditions and assessing the results.

Before commencing testing it is useful to have a test plan which gives the scope of testing, details on the testing environment (hardware/software) and the test tools to be used. Testers will also have to decide on answers to specific questions for each test case such as what is being tested? How are results documented? How are fixes implemented? How are problems tracked? QA Focus will be looking mainly at automated testing which allows testers reuse code and scripts and standardise the testing process. We will also be considering the documentation that is useful for this type of testing such as logs, bug tracking reports, weekly status report and test scripts. We recognise that there are limits to testing, no programme can be tested completely. However the key is to test for what is important. We will be providing documentation on testing methodologies which projects should consider using.

As part of the testing procedure it is desirable to provide a range of inputs to the software, in order to ensure that the software can handle unusual input data correctly. It will also be necessary to check the outputs of the software. This is particularly important if the software outputs should comply with an open standard. It will be necessary not only to ensure that the output template complies with standards, but also that data included in the output template complies with standards (for example special characters such as '&' will need to be escaped if the output format is HTML).

### 11 Deployment Into Service

The final area QA Focus will be looking at is the deployment of project deliverables in a service environment. It is unlikely that project will migrate into a service directly – the intention is that many of the project deliverables will be transferred to a JISC service who will be responsible for deploying the deliverables into a service environment. In addition to the deployment into a service environment for use by end users project resources may also need to be preserved. This is another area in which we will provide appropriate advice. Other work will address the issues involved in deploying software deliverables, digitised resources, Web sites, etc. into a service environment.

There are a number of scenarios for the deployment of projects deliverables: the deliverables may be hosted by a national service, within an institution or on the user's desktop.

It may be necessary to consider any special requirements for the user's desktop PC. For example will the service require a minimum browser version, will it require use of browser plugin technologies, are there any security issues (e.g. use of JavaScript), could institutional firewalls prevent use of the service, etc.

Inevitably there are resource implications for the deployment of project deliverables into a service environment: consideration needs to be given of the time taken for deployment and possible impact on other services (such as security, performance and compatibility issues). As well as these technical and resource issues there will be human aspects, including the potential resistance to change or reluctance to make use of work carried out by others.

An interesting approach which sought to provide a simple syndication tool has been carried out by the RDN. The RDN-include tool provides access to subject gateways and allows the institution to control the look-and-feel of the gateway. However, as this tool is implemented as a CGI script it requires System Administration privileges in order to be deployed. It was felt that System Administrators may be reluctant to deploy the tool, due to concerns over potential security problems. In order to address such concerns RDNi-Lite was developed, which provides similar functionality but, as it is implemented using JavaScript, can be used by an HTML author: no special System Administration privileges are required. This example illustrates an approach which acknowledges potential deployment difficulties and provides an alternative solution. Further information on this approach is available [20].

An important aspect of this work will be to ensure that projects describe the development environment at an early stage, in order to ensure that services are aware of potential difficulties in deploying deliverables in a service environment. One could envisage, for example, a project which made use of innovative technologies, open source tools, etc. which the service had no expertise in.

This could potentially make service deployment a costly exercise, even if open standards and open source products are used.

In addition to considerations of the deployment technologies, there is also a need to address the licence conditions of digitised resources. Again it would be possible to envisage a scenario in which large numbers of resources were digitised, some with licenses which permitted use by all and some which limited use to the project's organisation. In this scenario it is essential that the right's metadata allows the resources which can be used freely is made available to the service and that the production service can be deployed without making use of resources with licence restrictions.

### 12 Preservation Of Project Results

Even if a project has a clear idea of its final service deployment environment, there may be additional requirements during the project's development. Within the context of the JISC 5/99 programme there is now an expectation that learning objects funded by the programme will be stored in a learning object repository. The Jorum+ project [21] has been set up to provide repositories of the learning objects.

There is also discussion of the need to provide a records management service to ensure that project documentation, such as project reports, are not lost after the end of the programme.

In both of these areas QA Focus is well-positioned to advise JISC and the projects on appropriate strategies, based on its work in advising on technical interoperability.

### 13 The QA Focus Toolkit

An important QA Focus deliverable will be a QA Self Assessment Toolkit which will allow projects to check their project QA procedures for themselves.

A pilot version of the toolkit is currently being tested. The pilot covers the QA requirements when mothballing a project Web site and other project deliverables once the project has finished and funding ceases [22].

The toolkit consists of a number of checklists with pointers to appropriate advice or examples of best practice. The toolkit is illustrated below.

Subjective QA		
Web Site Content		
1.	Is it clear on the entry page that your project has finished?	Yes <input type="radio"/> No <input type="radio"/>
2.	Do all the dates on the site show a year?	Yes <input type="radio"/> No <input type="radio"/>
3.	Do you have a watermark that shows the site is no longer maintained?	Yes <input type="radio"/> No <input type="radio"/>
4.	Do your Web pages have persistent URLs?	Yes <input type="radio"/> No <input type="radio"/>

Figure 1: Toolkit For Mothballing Web Sites

The toolkit aims to document the importance of standards in a readable manner, which can be understood by project managers as well as technical developers. The toolkit will make use of case studies which have been commissioned and appropriate advisory documents. Most importantly the toolkit will provide a checklist and, in a number of cases, a set of tools which will allow projects to assess project deliverables for themselves.

The structure of the toolkit is illustrated below.

#### QA Self-Assessment Toolkit

**Area:** Access (e.g. Web resources, accessibility, ...).

**Importance:** Describe the importance of standards and best practices, including examples of things that can go wrong.

**Standards:** Describe relevant standards (e.g. XHTML 1.0).

**Best Practices:** Describe examples of best practices.

**Tools:** Will describe tools which can be used to measure compliance with standards and best practices.

**Responsibility:** Person responsible for policy and compliance.

**Exceptions:** Description of allowable exceptions from the policy.

**Compliance:** Description of approaches for ensuring compliance.

Figure 2: QA Self-Assessment Toolkit Structure

In the area of standards compliance for Web resources software tools can be used to check for compliance with standards. An article on "*Interfaces To Web Testing Tools*" describes the use of "bookmarklets" and a server-based interface to testing tools [23].

In a number of areas the use of software tools will be documented. The documentation will include a summary of the limitations of the tools, and ways in which the tools can be used for large-scale deliverables. This may include testing of significant deliverables, sampling techniques, etc.

### 14 Applying QA To QA Focus Web Site

We are using the methodologies described in this paper for in-house QA for the QA Focus Web site. This is being done in order to ensure that the Web site fulfils its role, to test our own procedures and guidelines and to gain experience of potential difficulties.

The approach used is to provide a series of policy documents [24]. The policies follow a standard template, which describes the area covered, the reason for the policy, approaches to checking compliance, allowable exceptions and audit trails, as illustrated below.

#### Policy On Standards For QA Focus Web Site

**Area:** Web

**Policy:** The Web site will be based on XHTML 1.0.

**Justification:** Compliance with appropriate standards should ensure that access to Web resources is maximised and that resources can be repurposed using tools such as XSLT.

**Responsibilities:** The QA Focus project manager is responsible for this policy. The Web editor is responsible for ensuring that appropriate procedures are deployed.

**Exceptions:** Resources which are derived automatically from other formats (such as MS PowerPoint) need not comply with standards. In cases where compliance with this policy is felt to be difficult to implement the policy may be broken. However in such cases the project manager must give agreement and the reasons for the decision must be documented.

**Compliance measures:** When new resources are added to the Web site or existing resources update the , validate tool will be used to check compliance. A batch compliance audit will be carried out monthly.

**Audit trail:** Reports from the monthly audit will be published on the Web site. The QA Focus Blog will be used to link to the audit.

**Further information:** Links to appropriate QA Focus documents.

Figure 3: QA Policy For QA Focus Web Site

### 15 Applying QA Methodology In Other Contexts

Although the approach to QA described in this paper is meant to be developmental, it is likely that projects will, to some extent, feel obligated to deploy the methodologies described. Use of the methodology from projects which are not funded under the JISC 5/99 programme will help to establish the effectiveness of the approach and should provide valuable feedback.

A presentation on the QA Focus work was given to staff from the Centre For Digital Library Research (CDLR) based at the University of Strathclyde in April 2003 [25]. Shortly afterwards CDLR staff felt sufficiently motivated to investigate the potential of the methodology for two

digital library projects: a digitisation project funded by the NOF-digitise programme which is currently under development and a regional digital library project which has been completed with no funding available for additional work.

The following conclusions were drawn:

“CDLR staff attempted to follow QA Focus guidelines retrospectively and to implement appropriate recommendations. This exercise showed that the extent of compliance with guidelines could be categorised into four areas: (1) areas of full compliance, where the project had already made decisions in accordance with QA guidelines; (2) areas in which compliance could be achieved with little extra work or with minor changes to workflow procedures; (3) areas in which QA guidelines were considered desirable but impracticable or too expensive and (4) areas where QA guidelines were not considered appropriate for the project.

The conclusion from the project managers involved was that consideration of the QA guidelines improved the value, flexibility and accessibility of the digital library deliverables, provided they were interpreted as guidelines and not rules. Rather than the QA process imposing additional constraints, the exercise validated decisions that had been made to vary from recommended standards, provided the issues had been considered and the decisions documented. What had been seen as a potentially burdensome exercise was regarded in retrospect as beneficial for the user service, for accessibility, interoperability, future flexibility and even for content management. It was felt that there are a number of areas in which simple developments to scripts or use of tools can provide a significant development to interoperability.” [26].

### 16 The Open Standards Philosophy

The JISC promotes the use of open standards in its development programmes. However feedback from projects indicates that there is not necessarily a clear understanding of what is meant by open standards.

QA Focus has produced a briefing document which seeks to clarify the term ‘open standards’ [27]. However there is still an unresolved issue as to the role that proprietary standards have in development programmes and the processes needed to evaluate open and proprietary standards and perhaps, in certain circumstances, chose a proprietary standards rather than an open one due to issues such as resources implications, maturity of standard, etc.

On reflection it would appear that an approach based on a simply advocating use of open standards is not necessarily desirable. It is felt that there are several factors which need to be addressed, including:

- Ownership of the standard (owned by an open standards body or by a company).



- In cases of proprietary standards, whether there is a community process for development of the standard.
- In cases of proprietary standards, whether the standard has been published openly or reverse-engineered.
- Whether viewing tools are available, available for free, available as open source and available on multiple platforms.
- Whether authoring tools are available, available for free, available as open source and available on multiple platforms.
- The fitness for purpose of the standard.
- Resource implications in use of the standard.
- Complexity of the standard.
- Interoperability of the standard.
- Organisational culture of the project's organisation.

It is felt that use of a matrix approach when choosing the standards for use in a development programme is well suited to the developmental culture prevalent in many digital library programmes and is preferable to a strict requirement that only open standards may be used.

The approach will, of course, require documentation outlining the decisions made and justification of deviation from use of accepted open standards and best practices.

### 17 Team Working Within QA Focus

QA Focus is provided by UKOLN and the AHDS, which are located in Bath and London respectively. In order to support working by a distributed team and minimise unnecessary travel team members make use of a number of collaborative tools, including My.Yahoo as a shared repository of resources. YahooGroups for managing the team mailing list and the MSN instant messenger to provide real time communications. We are also making use of a 'Blog' to provide news on QA Focus activities.

This approach appears to be working well. In order to share the experiences with other projects and to highlight potential problems (e.g. reliance on an unfunded third party) a case study has been produced [28].

### 18 What Next For QA Focus?

Although QA Focus funded is due to finish on 31<sup>st</sup> December 2003 we will be seeking additional funding to continue our work. We feel that QA for JISC's development programmes will be an ongoing activity, and, indeed, will grow in importance as "Web Service" technologies are developed which will require more rigorous compliance with standards.

We would hope to maintain the resources on the QA Focus Web site and produce new ones in appropriate areas. Additional activities we could engage in could include the deployment, development or purchase of testing tools and services. One possibility would be

hosting a JISC compliance service, along the lines of the UK Government's eGIF Compliance Service [29].

As well as providing advice to projects, QA Focus will also advise JISC on approaches to future programmes. We will be well-placed to provide advice prior to the start of project work, which will help to ensure that best practices are deployed from the start. We will recommend that, in addition to providing training on project management when new programmes begin that training is provided on best practices for ensuring that project deliverables are interoperable in a broad sense. We will also advise on contractual issues, including advice on the persistency of Web sites once project funding has finished. Advice will also be provided for evaluators of project proposals to ensure that consideration is given to issues such as QA procedures as well as technical feasibility.

## 19 Conclusions

The paper has described the work of the QA Focus project which supports JISC development activities by providing advice and support for projects in ensuring that project deliverables will be widely accessible, are interoperable and can be deployed into a service environment with the minimum of effort.

JISC will not be alone in giving a higher profile to quality assurance and compliance with standards and best practices for its development programmes. Within the UK two examples of standards-based programmes should be mentioned: (1) the e-government interoperability framework (e-GIF) defines the "internet and World Wide Web standards for all government systems" [30]; (2) the New Opportunities Fund's NOF-digitise programme provides funding to digitise cultural heritage resources [11].

We will be exploring the possibilities of shared approaches to QA with these bodies. The author welcomes feedback from those involved in similar activities in the international digital library community.

## References

- 1 **eLib Standards Guidelines**, <<http://www.ukoln.ac.uk/services/elib/papers/other/standards/>>
- 2 **Welcome To The DNER**, <[http://www.jisc.ac.uk/index.cfm?name=dner\\_home](http://www.jisc.ac.uk/index.cfm?name=dner_home)>
- 3 **Standards and Guidelines to Build a National Resource**, <[http://www.jisc.ac.uk/index.cfm?name=projman\\_standards](http://www.jisc.ac.uk/index.cfm?name=projman_standards)>
- 4 **JISC Information Environment Architecture**, <<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/>>
- 5 **Learning and Teaching (5/99) Programme**, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_learning\\_teaching](http://www.jisc.ac.uk/index.cfm?name=programme_learning_teaching)>

- 6 **FAIR Programme**, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_fair](http://www.jisc.ac.uk/index.cfm?name=programme_fair)>
- 7 **X4L Programme**, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_x4l](http://www.jisc.ac.uk/index.cfm?name=programme_x4l)>
- 8 **QA Focus Surveys Of Project Web Sites**, <<http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/>>
- 9 **QA Focus Briefing Documents**, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/>>
- 10 **QA Focus Case Studies**, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/>>
- 11 **NOF-digitise Technical Advisory Service**, <<http://www.ukoln.ac.uk/nof/support/>>
- 12 **TASI**, <<http://www.tasi.ac.uk/>>
- 13 **Legal Information Service**, <<http://www.jisc.ac.uk/legal/>>
- 14 **Metadata Analysis Of JISC 5/99 Project Entry Points**, <<http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/dc-metadata/>>
- 15 **Approaches To Validation Of Dublin Core Metadata Embedded In (X)HTML Documents**, WWW 2003 Poster, P. Johnston, B. Kelly and A. Powell, <<http://www.ukoln.ac.uk/web-focus/papers/www2003/>>
- 16 **Managing And Using Metadata In An E-Journal**, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-01/>>
- 17 **Managing a Distributed Development Project**, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-03/>>
- 18 **Resolving the Human Issues in LIS Projects**, <<http://www.rslp.ac.uk/circs/2002/LIS.htm>>
- 19 **Guidelines for Writing Software Code**, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-13/>>
- 20 **RDN-Include: Re-branding Remote Resources**, WWW 10 Poster, B. Kelly, P. Cliff and A. Powell, <<http://www.rdn.ac.uk/publications/www10/>>
- 21 **Jorum+**, <<http://www.jorum.ac.uk/>>
- 22 **QA Self-Assessment Toolkit: Mothballing**, <<http://www.ukoln.ac.uk/qa-focus/toolkit/mothballing/>>
- 23 **Interfaces To Web Testing Tools**, Ariadne issue 34, Jan 2003, <<http://www.ariadne.ac.uk/issue34/web-focus/>>
- 24 **Inhouse QA**, <<http://www.ukoln.ac.uk/qa-focus/qa/>>
- 25 **QA For Digital Library Projects**, <<http://www.ukoln.ac.uk/qa-focus/presentations/strathclyde-04-2003/>>
- 26 **Deployment Of Quality Assurance Procedures For Digital Library Programmes**, B. Kelly, A. Dawson and A. Williamson, paper submitted to IADIS 2003 conference, <<http://www.ukoln.ac.uk/web-focus/papers/iadis-2003/>>
- 27 **What Are Open Standards?**, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-11/>>
- 28 **Implementing a Communications Infrastructure**, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-12/>>
- 29 **e-GIF Compliance Assessment Service**, <<http://www.egifcompliance.org/>>
- 30 **e-GIF**, UK Gov Talk <<http://www.govtalk.gov.uk/interoperability/egif.asp>>

### About The Authors

**Brian Kelly** provides the JISC-funded UK Web Focus post, which provides support for the UK Higher and Further Education communities on Web issues. He is also the QA Focus project manager.

Brian has been involved in Web activities since 1993, when he was helped establish one of the first Web sites in the UK Higher Education community. He is now based in UKOLN – a national centre of excellence in digital information management, based at the University of Bath.

**Marieke Guy** is a member of the QA Focus team which supports JISC's Information Environment programme by ensuring that funded projects comply with standards and recommendations and make use of appropriate best practices.

Marieke has previously worked as a NOF-digitise Advisor and co-ordinated technical support and advice services to the NOF national digitisation programme. Prior to this she was the editor of *Exploit Interactive* and *Cultivate Interactive* Web magazines, both funded by the European Commission and the deputy editor of *Ariadne* Web magazine.

Marieke is based in UKOLN – a national centre of excellence in digital information management, based at the University of Bath.

**Hamish James** is the Collections Manager for the Arts and Humanities Data Service and is the AHDS member of the QA Focus team.

From 1999 to 2002 Hamish was User Support and Collections Management Officer for the History Data Service, a Service Provider for the AHDS. Hamish leads the AHDS's work in developing policies and practices for digital preservation, and contributes to workshops and other advice services offered by the AHDS.

Hamish is based in the AHDS Executive, at Kings College London – the AHDS is a distributed national service that aids the discovery, creation and preservation of digital collections in the arts and humanities.

# DEPLOYMENT OF QUALITY ASSURANCE PROCEDURES FOR DIGITAL LIBRARY PROGRAMMES

Brian Kelly

*UKOLN, University of Bath, Bath, BA2 7AY  
b.kelly@ukoln.ac.uk*

Andrew Williamson

*Centre for Digital Library Research, University of Strathclyde  
101 St.James Road, Glasgow, G4 0NS  
a.williamson@strath.ac.uk*

Alan Dawson

*Centre for Digital Library Research, University of Strathclyde  
101 St.James Road, Glasgow, G4 0NS  
alan.dawson@strath.ac.uk*

## ABSTRACT

Many digital library programmes have a development philosophy based on use of open standards. In practice, however, projects may not have procedures in place to ensure that project deliverables make use of appropriate open standards. In addition there will be occasions when open standards are not sufficiently mature for deployment in a service environment or use of open standards will require expertise or resources which are not readily available.

The QA Focus project has been funded to support a digital library development programme by advising on QA procedures which help to ensure that project deliverables are interoperable. Although the methodology developed by QA Focus is aimed primarily at one particular programme, the ideas and approaches are being made freely available and deployment of the approaches by others is being encouraged. This short paper provides an outline of the work of the QA Focus project and an analysis of the relevance and feasibility of QA Focus recommendations from two contrasting digital library projects.

## KEYWORDS

Quality Assurance, Standards, Digital Libraries

## 1. BACKGROUND

The JISC (Joint Information Systems Committee) provides funding for a wide range of digital library development projects. In recent years it has funded development of an ambitious strategy originally known as the DNER (Distributed National Electronic Resource) but now known as the Information Environment (IE) [JISC-1]. Projects funded under the IE programme are expected to comply with a set of documented standards and best practices. The *Standards and Guidelines to Build a National Resource* document [JISC-2] requires use of a range of open standards such as XML, HTML, CSS, etc.

The experience of previous programmes has shown that projects will not necessarily follow recommendations. There are a number of reasons for this: there may be a lack of awareness of the standards document; projects may find it difficult to understand the standards which are relevant to their work; there may be a temptation to make use of proprietary solutions which appear to provide advantages over open standards, and there may be concerns that use of open standards will require resources or expertise which are not readily available.

### 2. QA FOCUS

The QA Focus project was funded under the JISC 5/99 programme [JISC-3] to ensure that projects funded under this programme complied with appropriate standards and best practices in order to maximise interoperability and access to resources. QA Focus is addressing areas such as access, digitisation, metadata, software development and service deployment. A description of the QA Focus work has been published elsewhere [KELLY].

The approach taken by QA Focus is *developmental* which seeks to (a) explain the importance of standards and best practices; (b) review the approaches taken by projects in order to profile the community and obtain examples of best practices and areas where improvements may be made; (c) provide documentation, especially in areas where problems have been observed and (d) encourage projects which have implemented best practices to document their approaches and share their experiences within the community.

QA Focus is also developing a self-assessment toolkit which will provide a checklist for projects to validate their own QA procedures. A self-assessment toolkit designed for use when a project Web site is to be 'mothballed', which will form part of the final toolkit, is currently being tested [QA-FOCUS-1].

Although QA Focus is funded to support JISC's 5/99 programme the QA Focus deliverables are freely available on the QA Focus Web site [QA-FOCUS-2]. Related organisations are encouraged to make use of these methodologies, as this will provide valuable feedback, help refine the work of the project team, validate the methodology and help to ensure that deliverables from other programmes will interoperate with the deliverables of JISC 5/99 projects.

We will now describe the experiences of an organisation which is seeking to deploy the QA Focus methodology across a selection of its own projects. The two case studies have been provided by staff in the Centre for Digital Library Research [CDLR] at the University of Strathclyde. This work was initiated following a presentation on QA Focus work given to CDLR staff [QA-FOCUS-3].

### 3. CASE STUDY 1: VICTORIAN TIMES

Victorian Times [VICTORIAN-TIMES] is a large digitisation project funded by the New Opportunities Fund [NOF]. The project is digitising a range of textual and pictorial resources relating to social, political and economic conditions in Victorian Britain (1837-1901). These are supplemented with educational resources written by subject specialists. The project is required to be accessible by a variety of browsers, platforms, automated programs and end users.

NOF provide extensive guidance on the use of open standards, supported by online discussion forums, and access to a technical advisory team. They also require quarterly reports from projects documenting their implementation of standards and, when decisions have been taken to set aside standards, to provide strategies for migrating to suitable standards in the future.

This case study provides a brief account of some of the QA issues faced by the project, highlighting instances where it was necessary to compromise on adoption of standards for financial, technical, or service quality reasons.

Based on NOF guidance for **creation standards** the project decided that high-quality digital master images should be created in uncompressed TIFF format at 400 dpi resolution. This would meet preservation requirements and maximise options for creating digital surrogates as new open standards emerged. Later consideration of QA Focus guidelines shows this decision to be in full accord with recommendations.

Three surrogate formats were identified for **delivery formats** of the digitised resources; JPEG image files, plain-text OCR output, and PDF text files. It was also decided that Web content would be delivered in HTML 4, utilising cascading style sheets, and meeting W3C WAI accessibility criteria where possible.

Although PDF is a proprietary format it was judged to be acceptable since free viewers were available and materials would also be offered in other open formats.

It soon became apparent that the quality of OCR output from the digitisation was variable and highly dependent on the quality of the source materials. The option of manual correction of the text

proved prohibitively expensive. As the OCR output was being used to support free-text searching, the variation in quality was accepted as inevitable in the short-term. The high-quality digital masters allowed the possibility of repeating the process if there were significant improvements in the technology.

Comparison of project delivery formats with QA Focus recommendations showed mixed results, with areas of full compliance, partial compliance and non-compliance. However, the realistic and flexible nature of the guidelines meant that it was possible to comply with the QA framework even where recommended standards were not being followed, provided suitable procedures were followed and documented.

NOF guidelines on **resource identification** specify that digitised resources should be 'unambiguously identified and uniquely addressable'. This posed difficulties for the Victorian Times project, since its content is delivered by a bespoke content management system (CMS), with pages being dynamically generated based on the profiles of individual users. URIs are therefore lengthy strings of characters which are unique - if largely meaningless - to the user. It is, however, possible for users to uniquely address individual images from the collection, though this removes the images from the context of the Web service.

In this area it was necessary to balance the benefits to users of strict adherence to the identification standards against the richer service quality which implementation of the CMS would deliver. As the project would also be implementing an Open Archives Initiative [OAI] gateway to its digitised resources, it was decided that it would be appropriate in this case to set aside the standard. Again, the decisions made did not follow QA Focus recommendations, but could still be regarded as following best practice as a cost-benefit analysis had been carried out and informed decisions were made after considering the alternatives available. In addition consideration of the QA Focus recommendations helped to raise awareness of the issues.

#### 4. CASE STUDY 2: GLASGOW DIGITAL LIBRARY

The Glasgow Digital Library [GDL] has a long-term aim to create a wholly digital resource to support teaching, learning, research and public information at all levels in the city of Glasgow, bringing together material separated by ownership and physical location. Funding was obtained for two years to research the feasibility of a co-operative and distributed approach to developing a regional digital library, but not to provide an ongoing service.

By early 2003 the library had a collection of around 5,000 publicly available digital objects, and is being supplemented by further collections as small amounts of funding are obtained for specific digitisation projects. However, unlike the Victorian Times project, no funding is available for technical support, content management or ongoing maintenance and development.

There is a need to consider the extent to which it is feasible to apply the recommendations and procedures of the QA Focus project to an existing digital library with little time or money available. Many aspects of QA Focus guidance for Web sites and digital libraries have been considered, but particular attention is given to the use of open standards, the migration from HTML to XHTML, compliance with accessibility guidelines, and implementation of the <LINK> element to assist navigation, as recommended by QA Focus [QA-FOCUS-4].

In view of the importance of XHTML and the potential of languages such as XSLT for repurposing XML resources, it was felt desirable to migrate the Glasgow Digital Library from HTML to XHTML format. The GDL Web site consists of a large number of static but *automatically generated* web pages and a small number of *manually created* pages. In order to migrate the automatically generated pages, one page was converted manually, so that all the changes were understood. Once this had been validated, the programs and templates used to generate multiple pages were modified to produce the desired results, after which a small random sample was tested to validate the migration. In contrast, for the manually created pages, a batch conversion tool was used to carry out the migration from HTML to XHTML. Both approaches were feasible, but the exercise emphasised the value of automatic generation over manual creation for quality assurance as well as content maintenance.

### 5. CONCLUSIONS

The experiences in addressing QA in the context of real-world issues have helped QA Focus to refine its methodologies. It is clear that the approaches taken by projects to the use of open standards and best practices will be strongly influenced by issues such as resource implications, time scales and technical expertise.

Two digital library projects (one under development, with a CMS being implemented, one largely complete with a large collection of static pages) have attempted to follow QA Focus guidelines retrospectively and to implement appropriate recommendations. This exercise showed that the extent of compliance with guidelines could be categorised into four areas: (1) Areas of full compliance, where the project had already made decisions in accordance with QA guidelines; (2) Areas where compliance could be achieved with relatively little extra work or with minor changes to workflow procedures; (3) Areas where QA guidelines were considered desirable but impracticable or too expensive and (4) Areas where QA guidelines were not considered appropriate for the project.

The conclusion from the project managers involved was that consideration of the QA guidelines improved the value, flexibility and accessibility of the digital library deliverables, provided they were interpreted as guidelines and not rules. Rather than the QA process imposing additional constraints, the exercise validated decisions that had been made to vary from recommended standards, provided the issues had been considered and the decisions documented. What had been seen as a potentially burdensome exercise was regarded in retrospect as beneficial for the user service, for accessibility, interoperability, future flexibility and even for content management. It was felt that there are a number of areas in which simple developments to scripts or use of tools can provide a significant development to interoperability.

The developmental approach taken by QA Focus appears to have been largely validated in recommending that any compromises taken are *documented* and agreed with funding bodies, steering committees, etc. rather than *mandating* strict compliance with open standards. The feedback on real-world deployment issues is being addressed by QA Focus through a number of internal QA Focus documents which will provide examples of documentation which describe compromises which may be necessary [QA –FOCUS-5].

### REFERENCES

- [CDLR] **Centre For Digital Library Research**, University of Strathclyde, <<http://cdlr.strath.ac.uk/>>
- [GDL] **Glasgow Digital Library**, University of Strathclyde, <<http://gdl.cdler.strath.ac.uk/>>
- [JISC-1] **Information Environment Home**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=ie\\_home](http://www.jisc.ac.uk/index.cfm?name=ie_home)>
- [JISC-2] **Standards and Guidelines to Build a National Resource**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=projman\\_standards](http://www.jisc.ac.uk/index.cfm?name=projman_standards)>
- [JISC-3] **Learning and Teaching (5/99) Programme**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_learning\\_teaching](http://www.jisc.ac.uk/index.cfm?name=programme_learning_teaching)>
- [KELLY] **Developing A Quality Culture For Digital Library Programmes**, Brian Kelly, EUNIS 2003, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/eunis-2003/>>
- [NOF] **New Opportunities Fund**, <<http://www.nof.org.uk/>>
- [OAI] **Open Archives Initiative**, <<http://www.openarchives.org/>>
- [QA-FOCUS-1] **Mothballing Self-Assessment QA**, QA Focus, <<http://www.ukoln.ac.uk/qa-focus/toolkit/mothballing/>>
- [QA-FOCUS-2] **QA Focus**, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/>>
- [QA-FOCUS-3] **QA For Digital Library Projects**, QA Focus, <<http://www.ukoln.ac.uk/qa-focus/presentations/strathclyde-04-2003/>>
- [QA-FOCUS-4] **Enhancing Web Site Navigation Using The LINK Element**, QA Focus, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-10/>>
- [QA-FOCUS-5] **QA Focus In-House QA**, QA Focus, <<http://www.ukoln.ac.uk/qa-focus/qa/>>
- [VICTORIAN-TIMES] **Victorian Times**, <<http://www.victoriantimes.org/>>

# A PROPOSAL FOR CONSISTENT URIS FOR CHECKING COMPLIANCE WITH WEB STANDARDS

Brian Kelly, UK Web Focus, UKOLN, University of Bath, Bath, UK  
<B.Kelly@ukoln.ac.uk>

## ABSTRACT

This paper describes a technique for providing access to HTML validation services using a URI interface to validation services. Details of the implementation in Apache is given. The author proposes a small set of standard URI interfaces to validation and testing services which would ensure a consistent interface for users as well as authors.

## KEYWORDS

Quality Assurance, Standards, Digital Libraries

## 1 Enhancing Web-Based Validation

The importance of compliance with Web standards is well-understood. However although tools such as W3C's HTML and CSS validation services [W3C] are valuable use of the tools is not integrated with normal Web publishing process: using the interactive mode requires the URI to be copied, going to the validation service and pasting the URI. It is desirable to deploy a simpler interface.

An alternative approach is to provide an interface to the validation tool which can be accessed using a URI. This approach has been deployed on the UKOLN Web server [UKOLN]. Access to the HTML validation service can be obtained by appending `, validate` to any URI on the UKOLN Web server. This has been extended to include additional validation and testing services (e.g. CSS validation and link checking).

The interface is implemented using a simple server-side redirect. On the Apache Web server the `, validate` tool is implemented using the following code:

```
RewriteRule    /(.*) , validate
               http://validator.w3.org/check?
               uri=http:// www.ukoln.ac.uk/$1    [R=301]
```

This approach has the advantages that it can be used from anywhere: no software needs to be installed. It is easily maintained as a change to the testing service requires a single change to the server configuration file. A disadvantage with this approach is the reliance on a third party service and the associated dangers that the third party may change its conditions of use. However if solutions based on open source software are used it will be possible to migrate the service from the original host to a local service if necessary.

## 2 Proposal For Standard URIs

This technique will work on appropriately configured Web servers. Since the approach is easy to implement and has clear benefits it is hoped that the approach will be widely deployed. There will be a temptation to provide one's own naming conventions for the tools. However providing a standard naming convention has a number of advantages: a consistent interface across different domains used within an organisation; support help desk activities by providing consistent approach for collecting data and promotion of use of standards.

The author proposes the following standard set of URI for core validation and testing services: `, validate`, `, rvalidate`, `, cssvalidate`, `, checklink`, `, rchecklink` and `, http-headers`.

## References

- [UKOLN] Web Site Validation and Auditing Tools, UKOLN,  
<<http://www.ukoln.ac.uk/,tools>>  
[W3C] Markup Validation Service, W3C, <<http://validator.w3.org/>>

---

# Implementing A Quality Assurance Methodology For Digital Library Programmes

---

by Brian Kelly, UKOLN, University of Bath

*The JISC vision for the Information Environment seeks to provide users with seamless access to quality resources which are distributed across a range of providers, including JISC services, the institutions themselves and commercial vendors. The vision is based on use of open standards, which will allow developers and end user institutions freedom of choice in the application they use to develop and provide access to resources. This approach is reliant on use of open standards to ensure interoperability. This paper outlines the work of JISC's QA Focus advisory service which has been developing a quality assurance methodology and support service which aims to ensure that project deliverables will be interoperable.*

## Background

Although there is an awareness of the importance of open standards across many institutions and particularly those involvement in development work for the JISC there has not been a culture of rigorous checking to ensure that project deliverables comply with open standards. This is due in part to the developmental culture within the higher education sector, which is supportive of self-motivation and willingness to experiment.

This approach was probably sensible in the early days of Web development: if the eLib programme [1] had begun in the early 1990s use of Gopher rather than the Web could well have been mandated. We would then have faced difficulties similar to those which arose when use of the OSI networking standard and Coloured Book software was mandated and institutions were discouraged from using Internet protocols.

Fortunately however we are now in a more stable environment: the Internet and the World Wide Web have been accepted as the killer applications for the development of a rich set of distributed network services. The underlying architectural framework for the Web has also matured, and it is widely acknowledged that

XML provides the meta format for the development of new data formats.

In light of the growing maturity of the network environment infrastructure we are now in a position to progress from the experimental phase and seek to adopt more rigorous approaches to ensuring that project deliverables are interoperable and future-proofed.

Such an approach will be necessary in order to implement the seamless access to resources which the JISC's *Information Environment* [2] seeks to provide. The development of self-contained Web sites (the approach is the late 1990s) is no longer desirable; instead resources will need to be capable of being processed in a consistent manner by automated tools. This is a significant contrast with the development of Web pages for processing by Web browsers: an environment in which browsers were tolerant of errors.

## QA Focus

In light of the growing need for more rigorous compliance with standards the JISC-funded QA Focus to support initially JISC's 5/99 [3] and later the Facilitating Access to Institutional Resources (FAIR) [4] and Exchange for Learning (X4L) [5] programmes. The aim was the development of a quality assurance methodology to help ensure that project deliverables were interoperable through the deployment of appropriate quality assurance procedures.

QA Focus was launched in January 2002. Initially it was provided by UKOLN [6] and ILRT [7], University of Bristol. However, following ILRT's decision to refocus on their core activities, in January 2003 the AHDS [8] replaced ILRT, strengthening its work by being able to exploit AHDS's broad range of service experiences and extensive knowledge in the area of digitisation and service provision.



## A Developmental Approach

From the start QA Focus felt the need to take a developmental approach to its work. A hardline policing approach, in which project deliverables would be closely checked for compliance with standards and, in cases of non-compliance, recommendations to JISC that project funding should cease, was not felt to be appropriate.

The approach taken is developmental. We seek to ensure that projects have an understanding of the importance of open standards. Although we are not in a position to advise on best ways of implementing solutions, we have developed an infrastructure which allows projects to share their approaches. We also encourage projects to share the problems they have experienced and the limitations of their solutions.

## User Feedback

Prior to beginning our work it was clearly important to talk to our users – project developers funded by the JISC 5/99 programme – in order to get an understanding of the challenges projects faced in implementing standards-based solutions.

A questionnaire and two focus group meetings sought to gain feedback on (a) the standards framework [9]; (b) implementation issues and (c) service deployment. The responses indicated a number of concerns in implementing the standards:

**Lack of awareness of standards:** In a small number of cases there appeared to be a lack of awareness of the Standards document.

**Difficulties in choosing appropriate standards:** There was more widespread concern that it could be difficult to establish which standards were applicable to projects.

**Concerns over maturity of standards:** There were concerns that in some cases the standards may not be sufficiently mature for deployment.

**Concerns over change control of the standards document:** There were concerns that the standards framework may change during the project lifetime.

**Concerns over lack of tools:** There were concerns that in some cases tools which implement the standards may not be widely available.

**Difficulties in checking compliance with standards:** There were concerns over the difficulties in ensuring that standards were being used correctly.

The feedback on implementation issues had many overlaps with the concerns listed above. The poor support for standards by some browsers, for example, was identified as a concern for many.

Although useful feedback on standards and implementation challenges was provided, it was noticeable that the issue of deployment of project deliverables into a service environment did not appear to have been given as much thought. There was an exception in the case of projects being undertaken by JISC Services themselves. In other cases, there appeared to be a feeling that deploying project deliverables was an area to be addressed by the service providers and this was not a top priority for projects themselves.

## Surveying The Community

The user feedback was complemented with a number of semi-automated surveys [10] which helped us to profile the approaches taken by the projects in the provision of their Web sites. The surveys also helped us to identify common problem areas. This helped us to prioritise the areas in which advice needed to be provided.

## Providing Advice

Our approach to providing advice has been to produce brief, focussed documents. The documents seek to provide either an explanation of a standard, approaches to using the standard, common problems encountered with a standard and approaches to checking compliance.

To date (June 2004) we have published 70 briefing papers, covering areas of standards, digitisation, Web provision, metadata, software development and service deployment.

## Sharing Best Practices

The focus groups identified the need for specific advice on the deployment of standards and on appropriate implementation frameworks. Due to the wide range of areas being addressed by projects and the different approaches they may take and the different organisational cultures to be found across the institutions and organisations involved in project work it is neither possible nor desirable to recommend a particular implementation frameworks.

Our approach has been to encourage the community to document how they have approached use of standards and best practices. The case studies we have commissioned have been brief describing the issue being addressed in the case study, the solution chosen, the effectiveness of the approach chosen and details of lessons learnt or things that would be done differently in the future.

To date (June 2004) we have published 34 of these case studies, covering the areas of standards, digitisation, Web provision, metadata, software development and service deployment.

## The QA Focus Methodology

Although the feedback on the resources we have made available has been positive, our ultimate aim has been wider than this: our goal was to develop a quality assurance (QA) infrastructure which projects could deploy in order to embed best practices within their development work.

The QA methodology we have developed is based on well-established approaches to QA which can be implemented within the technical development framework for the projects. We are advising projects that they should adopt the following framework:

**Documented policies:** Projects should document their choice of standards and architectural framework.

**Compliance checking:** Projects should document their approaches to ensuring that they comply with their policies.

**Audit trails:** Projects should provide an audit trail which documents their compliance monitoring.

We recognise that this may be felt to be time-consuming to implement. In order to address such concerns and to illustrate that this framework can be implemented in a lightweight fashion the following examples have been provided.

### Policy Area: Web Standards

**Policy:** The QA Focus Web site is primarily based on XHTML 1.0 and CSS 2.0. Web pages should comply with these standards.

**Framework:** The Web site uses PHP to include HTML fragments. Part of the Web site provides access to an SQL Server database. Simple HTML editing tools (e.g. HTML-kit) are used to create and maintain the Web site.

**Exceptions:** Files automatically derived from other applications (e.g. MS PowerPoint) need not comply with HTML standards until conversion tools which generate compliant HTML are readily available.

**Change Control:** The project manager is responsible for the policy, ensuring policies are implemented and for changes to policies.

In order to ensure that the policies in this policy document are implemented it is necessary to document the compliance testing procedures. For example:

### Compliance Area: Web Standards

**Compliance Testing:** When pages are created or updated they should be checked for HTML compliance using the `, validate` tool.

When new CSS files are created or CSS is embedded within a page, the `, cssvalidate` tool should be used.

At least quarterly a survey of the Web site should be carried out using the `, rvalidate` (or equivalent) tool.

W3C's Web log validator tool should be run monthly to report on the top 10 pages which are not compliant.

**Audit Trail:** The output from the periodic bulk audits should be published.

We hope these examples illustrate that QA procedures need not be time-consuming to develop. We also hope that the implementation of such QA procedures will be seen as a normal part of ensuring that a Web site is functioning correctly and are not excessively time-consuming to implement, especially if they are implemented from the start of a project.

## A Matrix Approach For Standards Selection

We have outlined our recommendations on QA policies and procedures for standards and best practices. In addition to this, we have also produced a matrix for the selection of the standards and best practices.

Although in an ideal world the richest open standards and best practices would be deployed in reality it is often necessary to make compromises: the best choices may be difficult to implement due to lack of times, skills or resources.

There is a need to acknowledge such issues, without losing sight of the underlying principles of use of open standards. In order to ensure that open standards are not ignored because projects can't be bothered, but have legitimate reasons for a compromise solution we recommend a matrix approach in which the following issues are addressed.

**Openness of format:** Is the file format to be used open or proprietary?

**Openness of proprietary format:** If the file format is proprietary has the specification been published openly?

**Availability of viewers:** Are viewers available for free and/or as open source? Are viewers available on all relevant platforms?

**Availability of authoring tools:** Are authoring tools available for free and/or as open source? Are authoring tools available on all relevant platforms?

**Maturity of standard:** Is the standard mature or new?

**Richness of standard:** Is the standard rich and capable of being used to support complex applications?

**Complexity of standard:** Is the standard complex or relatively simple to understand and use?

**Resource implications:** Does the organisation have the resources necessary to make effective use of it?

**Organisational culture:** Does use of the standard reflect the organisation's culture?

Clearly addressing such issues has a subjective element and there may be conflicts (e.g. richness versus complexity). However if projects address such issues at an early stage in the project's life, it can help ensure that there is an awareness of the decisions made, the reasons for the decisions and the implications.

## The QA Focus Toolkit

In order to help projects embed a QA approach within their work, we have developed a toolkit [11] which seeks to ensure that we provide more than a static repository of documents, but also provide an interactive aspect to our service.

An example of the toolkit is illustrated below.

Web Site Content	
1. Is it clear on the entry page that your project has finished?	Yes <input type="radio"/> No <input type="radio"/>
2. Do all the dates on the site show a year?	Yes <input type="radio"/> No <input type="radio"/>
3. Do you have a watermark that shows the site is no longer maintained?	Yes <input type="radio"/> No <input type="radio"/>
4. Do your Web pages have persistent URLs?	Yes <input type="radio"/> No <input type="radio"/>

Figure 1: The QA Focus Toolkit

## Testing Tools

Although the remit of QA Focus's work is primarily in the development of a QA methodology and does not cover software development, we have addressed the issues of tools and approaches for checking compliance with standards. The work has focussed on tools which can check that Web sites comply with standards and best practices since this is an

area for which remote testing can be carried out.

We have sought to overcome the lack of integration of many Web testing tools with the publication process by describing an approach which provides authors with an interface to a range of testing services which can be accessed using the URL area of a Web browser [12]. This has been implemented by a simple change to the Apache configuration file on the UKOLN Web server, enabling HTML validation to be carried out by appending **,validate** to any URL on the UKOLN Web site.

In addition to documenting this approach we have also highlighted the limitations of commercial tools. For example, some link checkers, tools fail to check links to external resources such as JavaScript or CSS files; some link checkers and HTML validation tools cannot process resources which make use of features, such as frames, redirects, personalised interfaces, etc. There is a danger that use of such tools could give the impression that a Web site is compliant when this is not the case.

### Service Deployment

The main purpose of quality assurance procedures is to ensure that project deliverables can be deployed in a service environment easily, that deliverables are future-proofed against new developments and can be accessed in a wide range of environment.

We have been working with JISC services to ensure that an awareness of the challenges which services face in taking project deliverables and deploying them is gained across the development community. It appears to be not widely appreciated that even if projects comply fully with standards and best practices that there may still be potential difficulties in deploying the deliverables: for example, if a project makes use of a specialist content management system it may be resource intensive to deploy this application within a service environment. Use of open source software does not necessarily overcome such barriers as there is still a potential learning curve to be overcome.

As well as deployment by JISC services, there are a number of other environments in which project deliverables may be deployed, such as within institutions, for example as services to be managed within institutions or desktop applications; reports may need to be archived by a records management system or learning objects deposited in a repository. There is a need for the recipients to consider issues such as security and performance implications; legal issues; resource implications and the relevance to the institution.

As well as the deployment of project deliverables there are also long term preservation and records management issues which need to be addressed. It has been observed that project Web sites funded under eLib and the EU's Telematics For Libraries programme have disappeared shortly after funding has finished [13] [14]. This is an area of relevance to QA Focus. We have provided a number of recommendations on the availability of project Web sites after funding finishes [15] and provided a case study illustrating various procedures which can be used prior to 'mothballing' a project Web site [16].

### Acceptance Within The Wider Community

It is clearly desirable that the QA methodology outlined in this article is embedded within the working practices within organisations involved in JISC project work. In order to help to gain wider acceptance we have sought to disseminate our work across institutions. The main focus for this has been a workshop session in the Institutional Web Management Workshop in 2003 [17], although a number of other seminars have also been given.

### Gaining International Acceptance

We have sought international recognition of the approaches to QA outlined in this article. We are pleased to report that papers have been accepted at four peer-reviewed international conferences: a description of the QA Focus work was given at the EUNIS 2003 conference in a paper on "*Developing A Quality Culture For Digital Library Programmes*" [18]; the approach to the selection of standards was

described at the ichim03 conference in a paper on “*Ideology Or Pragmatism? Open Standards And Cultural Heritage Web Sites*” [19]; the deployment of the QA Focus methodology was described at the IADIS 2003 conference in a paper on “*Deployment Of Quality Assurance Procedures For Digital Library Programmes*” [20] and a paper on “*Interoperability Across Digital Library Programmes? We Must Have QA!*” [21] will be presented at the ECDL 2004 conference to be held at the University of Bath in September 2004.

## What Next?

QA Focus has developed a repository of support materials which can help projects in ensuring their project deliverables are compliant with standards and best practices. More importantly we have developed a QA methodology which we feel can be deployed by projects without providing too onerous a burden on the projects.

The JISC is looking to integrate aspects of the QA Focus work into the JISC Technical Standards Framework. The QA Focus outputs continue to be of relevance to ensuring the quality and interoperability of digital resources.

## References

- 1 **eLib**, UKOLN, <<http://www.ukoln.ac.uk/services/elib/>>
- 2 **Information Environment**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=about\\_info\\_env](http://www.jisc.ac.uk/index.cfm?name=about_info_env)>
- 3 **5/99 Learning and Teaching Programme**, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_learning\\_teaching](http://www.jisc.ac.uk/index.cfm?name=programme_learning_teaching)>
- 4 **Facilitating Access to Institutional Resources**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_fair](http://www.jisc.ac.uk/index.cfm?name=programme_fair)>
- 5 **Exchange for Learning**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=programme\\_x4l](http://www.jisc.ac.uk/index.cfm?name=programme_x4l)>
- 6 **UKOLN**, <<http://www.ukoln.ac.uk>>
- 7 **Institute for Learning & Research Technology (ILRT)**, <<http://www.ilrt.bris.ac.uk/>>
- 8 **Arts and Humanities Data Service (AHDS)**, <<http://www.ahds.ac.uk/>>
- 9 **Standards and Guidelines to Build a National Resource**, JISC, <[http://www.jisc.ac.uk/index.cfm?name=projman\\_standards](http://www.jisc.ac.uk/index.cfm?name=projman_standards)>
- 10 **QA Focus Surveys**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/surveys/>>
- 11 **Self Assessment Toolkit**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/toolkit/>>
- 12 **A Proposal For Consistent URIs For Checking Compliance With Web Standards**, B. Kelly, IADIS Internet/WWW 2003 Conference, <<http://www.ukoln.ac.uk/qa-focus/documents/papers/iadis-2003/poster/>>
- 13 **WebWatching eLib Project Web Sites**, *Ariadne* issue 26, 2001, <<http://www.ariadne.ac.uk/issue26/web-watch/>>
- 14 **URLs for Telematics for Libraries Project Page**, *Exploit Interactive* issue 1, 1999, <<http://www.exploit-lib.org/issue1/urls/>>
- 15 **Mothballing Your Web Site**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-04/>>
- 16 **Providing Access to an EU-funded Project Web Site after Completion of Funding**, QA Focus, UKOLN, <<http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-17/>>
- 17 **Catching Mistakes: Using QA to Address Problems on your Web Site**, <<http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2003/sessions/#workshop-14>>
- 18 **Developing A Quality Culture For Digital Library Programmes**, B. Kelly,

M. Guy and H. James, EUNIS 2003  
Conference Proceedings,  
<[http://www.ukoln.ac.uk/qa-  
focus/documents/papers/eunis-2003/](http://www.ukoln.ac.uk/qa-focus/documents/papers/eunis-2003/)>

- 19 **Ideology Or Pragmatism? Open  
Standards And Cultural Heritage Web  
Sites**, B .Kelly, M. Guy, A. Dunning, and  
L. Phipps, ichim03 Conference  
Proceedings, <[http://www.ukoln.ac.uk/qa-  
focus/documents/papers/ichim03/](http://www.ukoln.ac.uk/qa-focus/documents/papers/ichim03/)>
- 20 **Deployment Of Quality Assurance  
Procedures For Digital Library  
Programmes**, B. Kelly, A. Dawson and A.  
Williamson, IADIS Internet/WWW 2003  
Conference, <[http://www.ukoln.ac.uk/qa-  
focus/documents/papers/iadis-2003/paper/](http://www.ukoln.ac.uk/qa-focus/documents/papers/iadis-2003/paper/)>
- 21 **Interoperability Across Digital Library  
Programmes? We Must Have QA!**, B.  
Kelly, ECDL 2004 Conference,  
<[http://www.ukoln.ac.uk/qa-  
focus/documents/papers/ecdl-2004/](http://www.ukoln.ac.uk/qa-focus/documents/papers/ecdl-2004/)>

## **Contact Details**

Brian Kelly  
UKOLN  
University of Bath  
BATH  
BA2 7AY  
Tel: 01225 383943  
Email: B.Kelly@ukoln.ac.uk

---

# QA Focus

---

Marieke Napier

## Introduction to the QA Focus Post

The JISC QA (Quality Assurance) Focus post [1], which came into being in January 2002, was detailed in full in the last issue of Vine [5]; but for those unfamiliar with the post a brief introduction follows.

The new QA Focus post is promoting a Quality Assurance framework to ensure a more rigorous approach to the establishment of consistent, high quality standards for all the JISC DNER 5/99 projects and their associated 'products'. The decision by JISC to create the post was the result of the culmination of a number of significant factors. Over the past five years projects and programmes involved in the creation of learning materials have expanded rapidly across the FE and HE sectors. In particular, through the DNER and associated initiatives, a range of different digital assets and products have been produced. These include distinct types of digital assets (such as image archives, video or audio clips, and discrete learning materials), and software/applications (such as computer-aided learning packages, support systems, databases and portals). Throughout the creation of these materials there have been no established Quality Assurance procedures against which the quality of these materials can be assessed. It is anticipated that the QA Focus post will reverse this position for the JISC.

The post itself is being provided by an equal partnership of ILRT [2] (University of Bristol) and UKOLN [3] (University of Bath) and is jointly held by Ed Bremner (ILRT) and Marieke Napier (UKOLN). A collaborative approach to working has been established but the two partners are also responsible for individual areas. Marieke covers quality assurance for Web sites; aspects include accessibility, provision of Web sites, access by non-standard browsers and other user agents, compliance with accessibility guidelines and standards documentation, metadata, re-use/repackaging of Web sites and preservation of Web sites. She is also responsible for deployment of project deliverables in a service environment and Quality Assurance for software development. Ed Bremner covers quality assurance for digitisation; aspects include digital images, technical metadata, digital sounds, moving images and multi media resources. He is also responsible for the creation and delivery of learning and

teaching packages, modules and objects. The QA Focus post is given strategic support from Brian Kelly and Karla Youngs, Project Managers in UKOLN and ILRT respectively.

## Background to the QA Focus Questionnaire

One of the first QA Focus outings was to the 3rd Joint Programmes meeting for all DNER projects held at the Manchester Conference Centre in late January. At this meeting QA Focus handed out a questionnaire that aimed to both gain some preliminary understanding of what Quality Assurance procedures were currently being carried out within the DNER 5/99 projects and to obtain a better insight into project's expectations and hopes for the QA Focus post.

The initial questions given on the questionnaire were used to acquire information about the individual project, such as the name of it, the project's goals, the materials intended to be produced and how they would be produced. General questions were also asked in order to get projects personnel to think about the procedures carried out in their projects. In total we received 22 replies, which accounts for just under half of the current projects.

## The Questionnaire

### Products

The products which are being created by the projects are fairly varied. They include images (both 2D and 3D graphics), sound and video material, e-texts, html pages, case studies, interfaces, teaching and learning materials, online tutorials, databases and datasets, metadata and some reusable tools. Almost all of these materials are being created in-house. However some projects are also using the help of partners in the FE and HE sectors. A few projects have used external companies for specific pieces of work such as creating CDs but only one project stated that they have used an external organisation for all of their resource creation.

### Goals

A fair number of projects have had to modify their goals in some way. Most have had to scale down the number of images or resources created. Many

have also have found that their project has progressed slower than they initially expected, this was mainly due to a late start date or staffing problems. As a result of this a number of projects will be running for a longer period than initially expected. Projects seemed keen for the QA Focus to recognise that there is a need for more understanding of the difficulties involved in the type of projects covered by the DNER 5/99 programme and that there may possibly have been the necessity for lowering expectations at the start of some projects.

### **Quality Assurance Procedures**

Almost all projects that replied to the questionnaire felt that they had some quality assurance procedures in place already. Many saw these procedures as being carried out through some form of internal monitoring and self assessment. A number of projects have set up internal focus groups dedicated to quality assuring resources created. Others have regular steering group meetings and management evaluation sessions that consider how their project is running. These progress reviews are found by projects to be very useful.

Projects have also been using peer review systems to assess the quality of their resources and have been asking for feedback from other DNER projects as well as from students studying the relevant subject area.

A number of projects mentioned user rating systems. These were explained to be a process of carrying out usability testing on resources to study how the users use them and to also locate any problem areas, such as bad Web site navigation. It is interesting that these particular projects see quality assurance as a system of user evaluation as opposed to internal assessment.

The area of work in which quality assurance procedures have been the most clearly identified is in image creation. For image work projects often have testing mechanisms or a form of quality control in force. Images are often internally vetted, although a number of projects explained that they had also used external consultants to come in and evaluate their projects. This type of evaluation was expensive and had usually been written into the original project plan.

It also seems that metadata is being vigorously validated by a high number of projects. Some projects explained that metadata standards were followed and records were then checked by at least 2 people.

### **ADVISORY SERVICES USED**

An assortment of advisory services were mentioned by projects including TASI, VADS, HEDS, LTSN, PADS, UKOLN, TechDis, and the CD Focus; although worryingly a number of projects claimed that they had not consulted any yet. There seems to be consensus between projects that further information is needed on the specific roles of advisory service. Many projects found that they were unsure of the exact nature of work carried out by some of the services and the overlap between them. Some projects were also unsure of how advisory services could specifically help them with their work.

### **Standards Documentation**

Of the 22 projects who replied to the survey only one was not aware of the DNER technical standards. Although this information in itself is worrying it is possible that the answer is due to a misunderstanding of the question. All other projects stated that they have found the advice and guidelines offered useful, but were quick to point out that the standards had not been available at the start of the programme. QA Focus will be doing further investigation into this area.

### **QA Focus Role**

As was anticipated there is some apprehension about the QA Focus role. Projects are unsure of what to expect from the position and how the QA Focus work will affect them. This conclusion is probably reasonable as the role is a dynamic one.

The main areas in which project personnel felt that the QA Focus could help them were:

1. As a sounding board, a 'critical friend'. Projects felt that because of the unique position that the QA Focus was in, of being neither project nor service, they could be a point of formal reference that projects could turn to. They would be able to see the whole picture and provide support and understanding.
2. In demonstrating systematic approaches to processes. This could be achieved through online check lists, information papers, templates and the Quality Assurance toolkit.
3. By using knowledge about the programme and other projects to come up with benchmarking criteria, rating systems and methods (and examples) of best practice.
4. As a communication service that put projects in touch with others working on similar areas or suffering related problems. Also in informing projects about new developments.



The projects felt that although there would be significant benefits to their final resources in having an active QA Focus there were also potential problems. A number of the pros and cons of the role were dealt with in the Vine article but the main one quoted by projects was the extra time and money needed to document quality assurance procedures. This was felt to be particularly pertinent if the project had already finished. However it was noted that the experiences learnt would be beneficial for future project funding and on future programmes. Projects also felt that they would have problems articulating the quality assurance procedures they already had in place and that learning 'quality assurance speak' would take time.

The replies in the completed questionnaire indicated that there was also confusion on how the QA Focus role actually differs from the other main JISC services.

### FROM WORKPLANS

In the evaluation process carried out the quality assurance information given in the questionnaires was used alongside quality assurance information given in project's original workplans. Many projects had given some thought to quality assurance and its role within the project from the start; though as was displayed in the questionnaires Quality Assurance processes were usually defined as summative evaluation undertaken by external organisations.

### Conclusion

So what did we learn from the questionnaires? We learnt that quality assurance is already seen as an area of potential importance for projects but that at the moment how this quality assurance should be implemented is unclear to people. People seem unsure of what quality is and what quality assurance is. They almost view the two as some sort of add on that comes at the end of a project rather than a core thread running through a project.

The role of the QA Focus will be to encourage the integration of quality assurance procedures into general working practices. This will mean attempts to move away from post project evaluations carried out by external companies and instead concentrate on ongoing assurance of quality. The QA Focus does recognise that external evaluation has its role within a programme and individual projects, but feels that continual assessment and 'QAing' is preferable. Quality assurance, as the QA Focus sees it, examines the processes that shape a project and when good quality assurance is implemented there should be improvement in the

overall way a project runs as well as the final deliverables.

QA Focus has established 4 key areas of quality assurance for implementation within a project:

- Strategic Quality Assurance - carried out before development takes place - involves establishing best methodology for your project.
- Workflow Quality Assurance - carried out as formative Quality Assurance before and during development - involves establishing and documenting a workflow and processes.
- Sign-off Quality Assurance - carried out as summative Quality Assurance once one stage of development has been carried out - involves establishing an auditing system where everything is reviewed.
- On-going Quality Assurance - carried out as summative Quality Assurance once one stage of development has been carried out - involves establishing a system to report check, fix any faults found etc.

Consideration of these key areas as the basis of quality assurance would be highly beneficial to ongoing and future projects. They were used in a recent workshop on QA for Web sites given at the institutional Web Managers Workshop at Strathclyde [4]. Further discussion of their implementation will be discussed in future QA Focus articles.

### Contact Information

Ed I Bremner  
ILRT  
[Ed.Bremner@bristol.ac.uk](mailto:Ed.Bremner@bristol.ac.uk)  
Phone: 0117 - 9287109

Marieke Napier  
UKOLN  
[M.Napier@ukoln.ac.uk](mailto:M.Napier@ukoln.ac.uk)  
Phone: 01225 - 386354

### References

1. JISC QA Focus  
URL: <[http://www.jisc.ac.uk/qa\\_focus/](http://www.jisc.ac.uk/qa_focus/)>
2. ILRT, Institute for Learning and Research Technology  
URL: <<http://www.ilrt.bris.ac.uk/>>
3. UKOLN, a national focus of expertise in digital information management  
URL: <<http://www.ukoln.ac.uk/>>

QA Focus Area of the UKOLN Web site  
URL: <<http://www.ukoln.ac.uk/qa-focus/>>

4. QA for Web sites

URL: <<http://www.ukoln.ac.uk/web-focus/events/workshops/webmaster-2002/materials/napier/>>

5. JISC QA Focus Published in *Vine*, Volume 126.

URL:  
<<http://susanna.catchword.com/vl=1050753/cl=35/nw=1/rpsv/catchword/litc/03055728/v126n1/contp1-1.htm>>

### Author Details

*Marieke Napier*

QA Focus  
UKOLN  
University of Bath,  
Bath  
England, BA2 7AY

## Appendix - QA Focus Questionnaire

What is the name of your project?

What is your project start and end date?

Give a concise outline of your goals and aims for the project?

What digital products/materials have or will be produced by the project? (i.e. what are the deliverables?)

Have your digital products been created in-house or through an external agency? If so who?

Since the project bid have you had to change or modify any of these goals in any way? If so how.

What Quality Assurance processes does your project currently use, if any?

Are you aware of the DNER quality standards (DNER Standards document) and have these been useful in establishing standards for your project?

Which advisory services, if any, have you consulted so far?

What Quality Assurance issues do you feel are especially important to your project?

What do you expect from the QA Focus role?

Is there any particular way in which you hope the QA Focus will be able to help you?

---

## JISC QA Focus

---

Marieke Napier and Ed Bremner

An introductory article written by Marieke Napier and Ed Bremner introducing the new JISC Quality Assurance (QA) Focus post, discussing what it will be doing, the benefits it will provide, the challenges it faces and how it will affect projects.

### Why a QA Focus post?

In January 2002 the JISC Quality Assurance (QA) Focus post came into being. The decision by JISC to create the post was the result of the culmination of a number of significant factors. Over the past five years digitisation projects and programmes have expanded rapidly across the HE and FE sectors. In particular, through the DNER and associated initiatives, a range of different digital assets and products have been produced. These include distinct types of digital assets (such as image archives, video or audio clips, and discrete learning materials), and software/applications (such as computer-aided learning packages, support systems, databases and portals).

Throughout the creation of these materials there have been no consistent established QA procedures against which the quality of these materials can be assessed. The new QA Focus post will develop and promote a QA framework to ensure a more rigorous approach to the establishment of consistent, high quality standards for all the JISC DNER projects and services and their associated 'products'. Although the quality of project deliverables are usually of a very high standard, the projects themselves have not necessarily adhered to the technical standards.

### Why are Standards Important?

Standards have a crucial role to play in the creation of any information resources.

Creating products to an established and appropriate standard is critical in enabling their accessibility, interoperability, preservation and reusability. These are all important factors in the accountability of public funding that should take place in a programme like the DNER. Within the DNER Technical Standards and Guidelines it states that "*Adherence to standards plays an essential role in improving access to the information resources accessible online (and that) without the implementation of agreed approaches throughout the DNER, the aspiration of timely and usable networked information for use in education will not be realised.*"

Along with the requirement for more consistency in use of standards the JISC programme managers also recognize that there is a need to consolidate and expand the remit of the current JISC Services in order to provide a coherent approach to future digitisation work. With the QA Focus post these activities will be co-ordinated. The QA Focus will be overseeing, liaising and establishing QA practice rather than undertaking actual project by project work. The intention is for QA Focus to encourage and initialise QA practices and procedures within projects, which will then be carried out by the projects themselves.

### What is Quality Assurance?

It may be useful at this point to define what exactly quality is. Quality has been seen by those in the commercial world, and increasingly by those in the Higher Education

and Further Education communities, as the ability of your product or service to satisfy your customers. Quality assurance (QA) is the process that demonstrates that your product or service will satisfy your customers (users or preferred term); however this satisfaction level ideally needs to be agreed upon in advance. QA therefore is a collective process by a programme or organisation that ensures that the quality of its processes are maintained to the standards it has set itself. Such a process can be carried out by review by an external quality-auditing panel, in which case accountability becomes the key, or by self-evaluation, where improvement is the priority. The QA Focus will help projects to carry out internally driven QA of their processes, digital resources and Web sites by self evaluation, with the ultimate aim of improving the overall service for projects customers/end users. This self evaluation will be based on structured processes and will allow the formulation of distinct plans on how improvement can be effected.

To understand what makes a good digital collection or set of resources it is useful to be familiar with the life-cycle model and associated procedures and guidelines. Some of the key principles that need to be considered include collection development policy, digital resource creation, metadata descriptions, sustainability, accessibility, Intellectual Property Rights (IPR), measurement of use, reusability and how the resource fits into the larger context.

The QA Focus will consider a QA methodology for all these areas and significant lessons learnt from the application of the processes across all DNER projects will be disseminated to all DNER stakeholders.

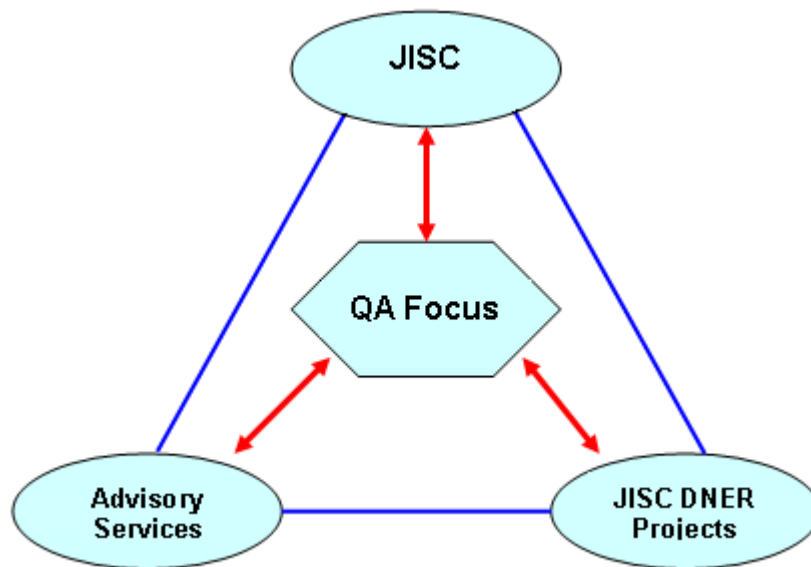
### **Who are the QA Focus Team?**

The QA Focus will be provided by an equal partnership of ILRT<sup>i</sup> (University of Bristol) and UKOLN<sup>ii</sup>, (University of Bath) and is jointly held by Ed Bremner (ILRT) and Marieke Napier (UKOLN). A collaborative approach to working has been established but the two partners will also be responsible for individual areas. Marieke will be covering QA for Web sites; aspects include access, provision of Web sites, access by non-standard browsers and other user agents, compliance with accessibility guidelines and standards documentation, metadata, re-use/repackaging of Web sites and preservation of Web sites. She will also be responsible for deployment of project deliverables in a service environment and QA for software development. Ed Bremner will be covering QA for digitisation; aspects include digital images, technical metadata, digital sounds, moving images and multi media resources. He will also be responsible for the creation and delivery of learning and teaching packages, modules and objects. Further details on these specific areas of work will be given later in this article. The QA Focus post will also be given strategic support from Brian Kelly and Karla Youngs, Project Managers in UKOLN and ILRT respectively.

### **Who will be affected by the QA Focus post?**

The QA Focus post is being funded to support and oversee QA within the 5/99 DNER projects; its existence will therefore have an effect on key stakeholders involved in any way with these projects, from creator to end-user. To start, the potential of the post is already being appreciated by the JISC and DNER management teams who now feel re-assured that the projects that they have funded are being helped and encouraged to create products to a universal high level of quality. The use of common standards across all projects will provide a greater likelihood of interoperability of the project deliverables and their subsequent

deployment into a service environment. The JISC advisory services will be affected by the existence of the new post because of their need to provide guidance and support for the projects. In time the projects will be able to create products that easily exceed the QA standards which the DNER have set. Eventually the effect will filter down to the end user, who will be provided with a much better product than could ever have been produced without any QA. The QA Focus team will sit between all stakeholders within the DNER 5/99 programme and provide the systems and communications to put QA process into place.



### **What are the benefits of a QA culture?**

Taking a systematic approach to QA with the appropriate standards will provide a host of obvious as well as subtle benefits to the DNER projects and all other stakeholders:

- The introduction of a QA culture into a project can only improve the quality of the product. Whatever the product is, establishing the appropriate standards and establishing a workflow that assures that your project complies to them will guarantee their quality.
- An established QA system assuring the product has been made to the appropriate standards can provide a useful ‘proof’ of the quality of the projects product supporting subsequent use within a service environment.
- QA culture is based around a concept of improving communication between all the stakeholders of the DNER projects. QA Focus will provide an active hub to encourage improved communication on all matters pertaining to QA.
- It is important to everyone involved with the DNER projects that every effort is taken to make sure that the projects are successful, however it is perhaps the funders who will most appreciate the approaches you have taken.
- Being able to ‘prove’ the quality of the project will provide a greater chance of future funding from other funding sources in the future.
- Once established within a team or organisation QA Culture will become second nature to those involved and can therefore be reused in any future project, providing benefit for all subsequent projects.
- Having a QA system within the workflow and creating products that have been made to an established standard will make it much easier to establish the

interoperability of project deliverables & then deploy them into a service environment.

- Last, but hardly least, you will provide your self with the self satisfaction that that comes from knowing that you have done everything that you can to make sure that your product is as good as it possibly can be.

### **What challenges does establishing a QA culture provide?**

The advantages of QA can be considered pretty self-evident, but on the other hand, what do the projects stand to lose by introducing the process? On balance it is certainly hoped that the challenges that are presented by establishing a QA culture (if not already present) will be outweighed by the benefits, but there are still some possible issues:

- Projects that had no provision for any QA system will now need to use up precious resources to review the quality of their work so far, which in turn can make it harder to ensure that the project remains on schedule.
- A QA culture demands a high level of communication both internally as well as with external bodies. This is not always easy and can certainly be time consuming. However, it is only by sharing our failures as well as our successes that as a group we will have any chance of learning at a sufficient rate to keep up with the advances in current technology and best practice.
- Developing a QA culture with a self-analytical basis within your project's team or supporting institute can be a challenge due to resistance from over-stretched staff and QA sceptics.
- QA is based on the establishment and adherence to a set of qualitative standards. Without these standards, QA can quickly become a very amorphous concept! However the speed at which the underlying technology within digitisation moves and the growth of our knowledge and experience means that best practices and indeed the standards themselves need to also be continually updated.

While it is understood that these factors do provide real challenges to the projects, QA Focus believes that the advantages greatly outweigh the disadvantages. QA Focus will do everything that it can to support DNER projects and JISC services, making the whole process as easy as possible for all concerned.

### **What is QA Focus Going to Do?**

As mentioned earlier, QA Focus intends to provide all the necessary methodology to enable the projects to undertake all aspects of QA through self-evaluation. It is not the intention of QA Focus to actually undertake any of the QA work itself. Given that there are over 50 projects such an undertaking would be both costly and time consuming. QA Focus will manage and oversee the whole process by taking a dynamic and reactive standpoint, from which they will be able to interact with all the relevant stakeholders to help all the DNER projects create products of the highest quality.

### **QA Focus will be able to do this by:**

**Creating a channel for information between the JISC-DNER (funders), the JISC Services and the DNER 5/99 projects themselves.** If there is one key element that is common through the whole of QA, it is communication. The QA Focus' primary purpose is to act as a conduit to enable cross-programme communication on all quality standards and best practices.

**Developing and disseminating a wealth of information on current good QA practices.** Keeping abreast of changes and developments within digitisation as they occur. QA Focus will then disseminate this information to the JISC services and the DNER projects as they need it, using the most appropriate method.

## **Establishing key areas of interest and expertise within QA**

**These areas will include:**

- QA for Digitisation, which will involve the evaluation of digitisation methodologies and workflow for digital images, moving images and sound. There will be consideration of choice of 'size' & 'quality' of resource, of chosen file type, compression, bit depth and choice of the appropriate size for the purpose of archiving and delivering images.
- QA for Learning materials, which will involve looking at the pedagogy and teaching methods used in the creation of resource and such areas as how e-learning packages are chunked.
- QA for Metadata, which will involve consideration of the metadata created for resources and the Web site etc.
- QA for Access, which will involve looking at the accessibility and usability of Web sites, compliance with HTML, CSS and other relevant standards. There will also be consideration of dissemination strategies and Web statistics.
- QA for Software, which will involve looking at code quality, process documentation and other allied issues.
- QA for Service Deployment, which will involve looking at deploying project deliverables within JISC services, preservation etc.

**Producing a QA toolkit.** This will provide a checklist for projects to undertake QA evaluation of their own work. The checklist will include links to tools that allow quick evaluation of the resources created. It is hoped that the toolkit will become part of our core knowledge for further projects.

At this early stage of QA Focus, many of the finer details of approach and methodology for this work have yet to be established and as such, are still open to discussion and feedback. QA Focus intends to give a more detailed breakdown of this work shortly in an issue of the Web magazine, Ariadneiii.

## **Conclusion**

So what does all this mean for you, if you are working on a project? What should you do now? To start with, if you aren't already, QA Focus suggests that you review your QA processes and make sure they are completely documented. Begin by considering your compliance to the current DNER Architectural Framework documentation and what procedures you have in place to help you assure your product complies with those standards. Such activities may seem tiresome right now but will provide useful long term benefits.

Remember...the QA Focus post has been established to help you work better and in the long run you have much to gain from it. If you have any further questions, please do not hesitate to get in contact with either of the post-holders<sup>iv</sup>.

## Contact Information

Ed I Bremner

ILRT

Ed.Bremner@bristol.ac.uk

Phone: 0117 9287109

Marieke Napier

UKOLN

M.Napier@ukoln.ac.uk

Phone: 01225 386354

---

<sup>i</sup> ILRT, Institute for Learning and Research Technology

URL: <<http://www.ilrt.bris.ac.uk/>>

<sup>ii</sup> UKOLN, a national focus of expertise in digital information management

URL: <<http://www.ukoln.ac.uk/>>

<sup>iii</sup> Ariadne

URL: <<http://www.ariadne.ac.uk/>>

<sup>iv</sup> QA Focus area of the UKOLN Web site

URL: <<http://www.ukoln.ac.uk/qa-focus/>>