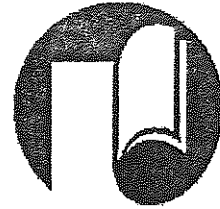


NATIONAL PRESERVATION  
OFFICE



RESEARCH  
LIBRARIES  
GROUP

## GUIDELINES FOR DIGITAL IMAGING

Papers given at the  
Joint National Preservation Office and Research Libraries Group  
Preservation Conference in Warwick  
28<sup>th</sup> – 30<sup>th</sup> September 1998

Organised by the National Preservation Office, the Research Libraries Group, the  
Joint Information Systems Committee and the United Kingdom Office for Library  
and Information Networking

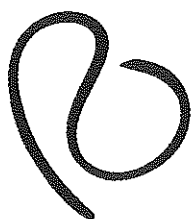
*NPO is supported by*  
The British Library, The Public Record Office,  
The National Library of Scotland, Trinity College Library Dublin,  
The Consortium of University Research Libraries,  
Cambridge University Library, The National Library of Wales,  
The Bodleian Library, Oxford University

© The Contributors 1998

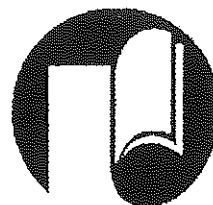
First Published 1998 by  
The National Preservation Office  
The British Library  
96 Euston Road  
London  
NW1 2DB

ISBN 0-7123-4655-4

All website addresses correct as of September 1999



NATIONAL PRESERVATION  
OFFICE



RESEARCH  
LIBRARIES  
GROUP

## GUIDELINES FOR DIGITAL IMAGING

Papers given at the  
Joint National Preservation Office and Research Libraries Group  
Preservation Conference in Warwick  
28<sup>th</sup> – 30<sup>th</sup> September 1998

Organised by the National Preservation Office, the Research Libraries Group, the  
Joint Information Systems Committee and the United Kingdom Office for Library  
and Information Networking

*NPO is supported by*  
The British Library, The Public Record Office,  
The National Library of Scotland, Trinity College Library Dublin,  
The Consortium of University Research Libraries,  
Cambridge University Library, The National Library of Wales,  
The Bodleian Library, Oxford University



© The Contributors 1998

First Published 1998 by  
The National Preservation Office  
The British Library  
96 Euston Road  
London  
NW1 2DB

ISBN 0-7123-4655-4

All website addresses correct as of September 1999



# GUIDELINES FOR DIGITAL IMAGING

Papers given at the  
Joint National Preservation Office and Research Libraries Group  
Preservation Conference in Warwick  
28<sup>th</sup> – 30<sup>th</sup> September 1998





# Joint NPO and RLG Preservation Conference

## Guidelines for Digital Imaging 28<sup>th</sup>-30<sup>th</sup> September 1998

<b><u>Introduction to Joint Conference Publication</u></b>	Robin L. Dale	
<b><u>Keynote Address</u></b>		
Guidelines Versus Guidance for Digital Imaging: The Opportunities Before Us	Anne R. Kenney	1
<b><u>Conference Papers – Guidelines for Selection</u></b>		
Selection Guidelines for Preservation	Janet Gertz	11
Guidance for Selecting Materials for Digitisation	Paul Ayris	21
<b><u>Conference Papers – Guidelines for Preparation</u></b>		
Preparation of Materials for Digitisation	Ann Swartzell	29
Protecting the Physical Form	John McIntyre	33
<b><u>Conference Papers – Guidelines for Image Capture</u></b>		
Guidelines for Image Capture	Stephen Chapman	39
Guidelines for Image Capture: a UK Perspective	Jane Williams	61
Issues and Approaches to Preservation Metadata	Michael Day	73
Issues and Approaches to Digital Archiving	Margaret Hedstrom	85
<b><u>Conference Papers – International Reaction</u></b>		
Canadian Overview	Karen Turko	89
An Australian Perspective	Colin Webb	95
<b><u>Closing Address</u></b>		
Weaving the Threads	Peter Fox	103
<i>Appendices</i>		
Appendix 1 – Supporting documents		
Appendix 2 - Responses to preparation questionnaire		
Appendix 3 – Participants list		
Appendix 4 – Conference report	Nancy Elkington & Neil Beagrie	



**Introduction to Joint Conference Publication**  
**Robin L. Dale**  
**Research Libraries Group**

Digital imaging has been a part of libraries, archives and museums for more than ten years. Groundbreaking research completed in the early- to mid-1990s has led to greater understanding, increasingly better imaging, and has contributed to the emergence of digital collections around the world. Yet because of the breathtaking speed of technological change, there has been little time to develop consensus on the "best practices" institutions should follow to create high quality images which can meet the demands of their respective user populations over the long-term. What materials should be selected for digitisation? Are some collections and/or source documents better suited to digital imaging than others? How should the materials be digitised? And to what level of quality? How much metadata is appropriate? Can these new electronic resources be maintained and remain accessible over time?

In September 1998, the Research Libraries Group (RLG) and the National Preservation Office (NPO) joined forces to address those very questions. The result was the NPO and RLG Joint Preservation Conference: **Guidelines for Digital Imaging**, a joint effort between two organisations with remits related to preservation and digital imaging. Designed to be a working conference, the event created a forum for the discussion of current practices related to digital imaging for cultural heritage institutions with an eye toward documenting the lessons learned from the collective experience.

Representatives from Australia, North America, and nearly a dozen European nations attended and participated in the conference. Most of the delegates had extensive experience with different types of imaging projects. Attendees represented a broad spectrum of interests and included librarians, archivists, imaging specialists, corporate/vendor representatives, preservation consultants, and academic researchers. There was clearly international interest in comparing notes, practices, and strategies.

Three plenary sessions addressed issues of selection for digital imaging, preparation of materials for digitisation, and digital image capture. Following presentations by a UK and US speaker on each of the three main topics, participants selected one of the three topics for intensive discussion. The group discussions were the heart of the conference, where participants' experience and expertise were brought together in focused dialogues. Can best practice be agreed upon or should guidance be offered instead of prescriptive guidelines? Can international agreement on basic technical requirements for high-quality digital images be reached?

In the end, participants endorsed the recommendation highlighted by Anne R. Kenney in her keynote address – when dealing with the technical tools which are likely to evolve, and when considering the range of cultural heritage materials which repositories will want to digitise, it is much better to offer *guidance* than construct generalised guidelines. To quote Kenney,

*"... guidelines are contextual - they are specific to institutional settings, and may not scale beyond those constraints or to other environments... Guidance, on the other hand, can be developed to enable an institution to step through a process with the end result leading to the creation of their own specific set of guidelines. Guidelines should be a by-product of this process, not the starting point. Deciding on a course of action, what to do and what not to do, should be based on the reasons for digitising, the nature of the source documents, the institutional mission, available resources, the technical infrastructure, and users requirements and capabilities. Guidance begins with understanding the context, provides a process for data gathering and decision-making,*

*points to available guidelines and a means for assessing their applicability, and finally leads to the development of guidelines that are specific to institutional needs."*

Conference speakers and participants were able to agree on many basic requirements related to selection, preparation of materials for digitisation and for digital imaging. At the same time, participants identified the numerous challenges that remain on the digital imaging front. These issues, and others such as digital preservation – the long term retention of and access to digital information – continue to be addressed by those working in concert with the Research Libraries Group and the National Preservation Office.

Research Libraries Group and National Preservation Office wish to thank the speakers and discussion leaders who inspired and challenged participants throughout the two-day conference. Particular thanks are extended to the Joint Information Systems Committee's Committee on Electronic Information and Kodak for their financial support which enabled both the international and the UK-based speakers costs to be met. Finally, our collective gratitude to Hazel Gott (UK Office for Library and Information Networking) for organising such an effective and rewarding event.

The electronic version of the papers, delivered at the conference are also available on the following websites:

RLG (UK):	<a href="http://www.rlg.ac.uk/preserv/joint/">http://www.rlg.ac.uk/preserv/joint/</a>
RLG (non-UK):	<a href="http://www.rlg.org/preserv/joint/">http://www.rlg.org/preserv/joint/</a>
UKOLN:	<a href="http://www.ukoln.ac.uk/events/sept-conf98/">http://www.ukoln.ac.uk/events/sept-conf98/</a>
NPO:	<a href="http://www.bl.uk/services/preservation">http://www.bl.uk/services/preservation</a>

## Guidelines vs. Guidance for Digital Imaging: The Opportunity before Us

Anne R. Kenney

Associate Director for Preservation

Cornell University

We have an unprecedented opportunity for action here in the next several days. This conference offers us the means to codify, on an international basis, issues associated with the use of digital imaging technology to retrospectively convert to digital form paper and film based sources. I believe that we are here to act, not just listen, and that it is an opportune time to act - the stars are indeed aligned. Some of those stars are here in the audience. Look to your left and to your right - the level of experience and expertise assembled is impressive. In fact it strikes me that we couldn't have had this conference much earlier because there were too few institutions working in this area - so the timing is right.

I'm also struck by the collaborative nature of the conference. This assemblage represents a powerful combination of players. We have an international representation, with delegates from thirteen countries. Institutional perspectives are well represented, as are critical consortia and organisations. It is fitting that this conference is co-sponsored by RLG and NPO. RLG brings a stellar reputation in laying down guidelines in image reformatting and collaborative action; and the NPO, as RLG, is committed to act on behalf of its constituencies to establish best practices and disseminate widely the results of consensus-based working groups. We also have representation from JISC, UKOLN, and *TASi*, and the recent spate of publications coming out of these groups on all manner of things digital have made those of us on the other side of the pond sit up and listen.

So we are called together at the right time, at the right place, with the right players and the right mission. It behoves us, therefore, not to let this opportunity get away from us. What we do here can have a pronounced impact on local, regional, national, and international programs over the next coming years. We have an opportunity and an obligation to do good and to do it well. We will be praised if we do so, and criticised if we do not. One might almost suggest that we are on the treadmill of opportunity.

**But How Shall We Act?** And what is the best course of action? We are not here to review and endorse proposed guidelines, as had been the intent when this conference was originally conceived. In many ways this is a good thing, because a lot has happened in the interim between the development of the Conceptual Table of Contents for a manual on preservation digitisation by the RLG Working Group on Digital Image Capture in the spring of 1997 and our meeting here today. We have the opportunity to take advantage of solid work already accomplished, which can serve as the foundation for our deliberations. The real question before us is what is the best role for us to take as a deliberative body - we are here to reach consensus - poised to move, but in what direction? In an effort to help us consider how best to optimise this opportunity, I have come up with ten proposed guiding principles.

### Proposed Guiding Principles.

- **Do not reinvent the wheel - or any of its spokes.** Should we spend our time developing a set of RLG/NPO guidelines for digital imaging? And if so, why? As we all know, interest in things digital occupies the attention of many people, some of whom have already given a lot of thought to suggesting guidelines for digital imaging. A number of the better ones have been produced by those assembled here. If we are to proceed in this direction, we must answer the following questions: Is what is available inadequate? Incomplete? Inaccurate?

Outdated? Obscured? Or even appropriate? However we decide to act, it will be our role to acknowledge this work, to take the very best from it, and to build on it, not just replicate it. Our first task then should be to assess what is out there, where the gaps are, and what is the commitment from other bodies to create, update, and maintain important information.

- **At least not an old wheel.** If we do decide to move forward in creating a set of RLG/NPO guidelines for digital imaging, we should take care not to reinvent old wheels. We must clearly be mindful of what is come before, but also assume a critical distance from it. We must be sensitive to differences not fully comprehensible today between things physical and things digital. It is natural to look to other image reformatting guidelines because they have served us well and can serve us again, but only with an understanding that we are dealing with something altogether different. Consider the differences between microfilm and digital imagery. Microfilm represents a physical object, consisting of content fixed on a *medium*. In the case of digital imagery, an emerging definition of the digital object embodies both content and services. Unlike microfilm, the structuring, context, and provenance of a digital file must be "explicitly captured and documented as it is created." Or consider that in a microfilm version, contrast is fixed, but in a digital image it isn't. Digital images are dynamic, where access and preservation requirements are more process oriented than with microfilm. In moving from analogue to digital, there is a significant translation going on, and current models can not anticipate all of the results. We need to consider these compelling differences as we borrow from the old to meet the new.
- **Acknowledge the full continuum.** As others have made clear, digital image conversion is intricately entwined with other functions, particularly those associated with management, preservation, and use. We must appreciate how decisions made upstream at the point of capture can have ramifications all along the digitisation chain, affecting functions and roles in unexpected ways. So if we seek to define guidelines for digital imaging, we must consider them in a very broad context. We must strive to understand the technical environment in which these materials will be accessed and used, as Columbia University has in providing technical recommendations for digital imaging projects conducted by faculty, students, and staff. And, in a networked world, we will be held accountable publicly for our decisions made; we need to look no further than the current trial of Clinton by the American public to know how true this can be. So we must think about how well significant informational content is conveyed, including what the Germans call the "testimonial" evidence that supports or clarifies the content through external formal features. But we must also consider how the file lends itself to processing (OCR, visual searching), to navigation, resource discovery, and versioning for access and use, to "dynamic updating with richer encoding," and to long term maintenance. We know that we want to provide files of "enduring value" but we must also ensure that they convey "endearing value." So we must concern ourselves with all manner of things technical: quality metrics, file formats, header information, structural metadata, the use, type, and level of compression, resolution, bit depth, and so on. All of this would be hard enough in a static environment, but in a constantly changing one, it is more than a little mind-boggling. We must strive to go beyond meeting current service objectives to envisioning capabilities on the horizon, such as visual searching, vectorisation, improvements in OCR, on the fly conversions. We need a lot of help from our friends.
- **Beware of the hype.** In an article on the continuing role of books, RLG's Walt Crawford warns against the "great technological hand wave." He refers to a tendency to believe that all things are technologically possible and within our reach, and to minimise shortcomings, placing faith in a belief that they will be overcome in "just a couple more years." How many of us have heard that the technology just keeps getting cheaper and cheaper, so that by tomorrow it will be free and presumably by the next day we will be paid to use it? It is important to keep in mind that technology is not our friend. If things appear to be difficult it is probably because they are difficult. While this seems common sense, we must also not be paralysed by it, and we should beware the naysayers as well. Even some of our most forward

thinkers can have their worldview impaired by a lack of appreciation for change - recall that in 1981 Bill Gates declared that "640K ought to be enough for everybody." Things do change, and often for the better. We must strike a delicate balance between wishing problems away, and being too wedded to the way things work now. This conflict leads us to another principle.

- **Own the technology, but do not let it own you.** One of the key differences between microfilm and digital technology is the pace of change. With microfilm, we do not have to know that much about the specifics of the technology, because we can have faith in its stability and its rules. But as consumers of digital technology, where there is no stability and the ground rules change all the time, we must assume a greater responsibility to understand what is here, what is coming, what is transient, and to be able to see technological change in terms of library and archival interests.

Near term improvements in the post-processing of image information, for instance, may facilitate a shift in thinking about how to create the highest possible image quality for a given collection. A suggested new capture architecture has the appropriate raw grayscale or colour data collected from any scanner whose document handling capabilities suit the peculiarities of a particular item, such as a bound volume, a 35mm slide, or a 40 inch wide architectural drawing. The scanner choice can be made on the basis of its physical suitability and the quality of its raw image data. All special processing from these various sources would then be performed in an off-line, largely scanner-independent manner. If this were possible, we would not be constrained by the variable and inconsistent processing offered within the many different scanners that are needed to overcome the physical peculiarities of each item in a collection. This work would be particularly important in developing the means for capturing bound volumes without having to resort to disbinding, to endangering the volume, to creating photo-intermediates, or to a lowering of our quality requirements.

We need to own the technology, yet not let it own us. This suggests to me that as we seek to define requirements, they should be as independent as possible from any specific technological approach. The constraints of today may not be there tomorrow, and we should avoid building an approach that becomes quickly outdated or superseded. A strong concern at Corbis, a commercial digital archive, is to be able to maintain the usability of digital photographic reproductions on future unknown delivery and output systems. As Paul Conway reminded us at RLG's first digital preservation conference five years ago, we should follow the advice of our colleague Elsie Freeman Finch. When we need a quarter inch hole, we should say so, rather than define our needs in terms of a particular drill bit.

- **Become a sceptic, especially about hard and fast guidelines.** As noted earlier there is a plethora of information being presented as digital imaging guidelines, but a lot of it is contradictory, misleading, incomplete, or self-serving. To paraphrase the *New York Times*, we need to follow all the advice that's fit to take. Consider, for instance, recommendations regarding resolution. Some say scan at the highest possible resolution, others recommend that you avoid scanning images at very high resolution because it can cause an image to be blurred. Several imaging "experts" argue that there is no sense in scanning colour photographic prints at resolutions above 200 dpi because there is no additional information to be captured. It is true that a colour print has far less information than the negative. But you still have to look at it, see how big it is, and determine the level of significant detail contained in it. One can make a fairly decent scan at 200 dpi for most consumer variety colour prints produced at the local drug store on smaller print sizes. But the same does not hold true for higher quality colour prints generated through fine cameras and printed onto larger formats via a professional lab for prints generated from transparencies above 35mm. As James Reilly put it, "you cannot be absolutist about any dictum that only talks about dpi for a whole (rather undefined) class of photos." It is important to consider all the aspects governing resolution choice with a colour print or anything else.

- **Come to terms with the notion of preservation in digital reformatting.** It has been suggested that digital imaging can be used for preservation, for improved access, or for both. I believe this is not the case today. The primary use of digital imaging into the near future will be to improve access. It may also be used for preservation *and* access, but not just for preservation purposes alone. If only preservation purposes are to be served, other reformatting options to ensure continued viability of the informational content are probably better choices to make. We will hear later in the conference an update on digital preservation efforts, but we all know that we do not *decrease* the preservation problem by relying on digital information, we only *increase* it. As Terry Kuny put it, "being digital means being ephemeral." A German report on the intrinsic value of archives and library material explicitly rejects the use of digital imaging for preservation purposes because of its lack of fixity, but does acknowledge its use for improving access. In fact, there is renewed interest in a hybrid approach that marries microfilm for preservation and digital imaging for access. In the United States, where the brittle books program opened the door to a definition of preservation as the permanence of information documents, there has even been a resurgent demand for retaining the physical objects as well. In some American universities, digital imaging is used to replace originals when an analogue paper or film copy is created, and the University of Michigan is experimenting with digitisation and on-line access only. A growing number of governmental agencies and businesses are turning to the technology to conduct office back file conversions, with or without an analogue backup, and with or without official archival sanction. With few exceptions, though, the primary motivation has been to improve access, even when preservation is also a concern.

So if we are to assume digital imaging is primarily motivated by access considerations, where does that leave the preservation side? We should acknowledge that the notion of preservation is relative, and turn our attention to creating digital objects that we are prepared to keep indefinitely (however long or short that might be). Even if we are not thinking long term retention, immediate applications and current uses require that we create a fairly rich file or set of files, hence the need to focus on the requisite image quality and metadata for creating "access masters." These criteria should be based on how well the files lend themselves for processing and for presentation in a variety of formats and on a range of devices and media. At this point, if we think preservation digital master at all, we should be judging it by whether the file can produce preservation quality analogue backups - preferably COM - and how to preserve the accompanying requisite metadata in an eye-readable form.

I like the Australian assumption that resources should be devoted to retaining digital materials "only for as long as they are judged to have continuing value and significance." We should strive to create access master files in a way that makes them worthy candidates for long term retention - so that disposition decisions are based on continuing value and functionality, not limited by technical decisions made at the point of conversion or anywhere else along the digitisation chain. The role of those concerned with preservation should encompass the creation of image files that are "preservation worthy." Peter Lyman and Brewster Kahle distinguish between "born" digital and "borne-again" digital - files created by converting from other media. Once in digital form, these distinctions may not be that significant, at least for digital imagery. I would argue we just do not want to create "still born" digital files - that we need to create files that are "fit for purpose" as our UK colleagues would say, and cost-effective over their full life cycle. Neil Beagrie and Daniel Greenstein of the Arts and Humanities Data Service argue in their strategic policy framework for creating and preserving digital collections that "how data is created will impinge directly upon how it can be managed, used and preserved at any future date." They suggest that we need case studies to "demonstrate the benefits of any additional investment toward long-term preservation during data creation in terms of efficiencies and use later in the life cycle of the resource."



- **Presume Web or Web-like delivery.** Deanna Marcum, President of the Council on Library and Information Resources in the US once said, "The notion of a stand-alone digital library seems anathema to its intent." Digitisation is a process of separating information from physicality of form and place - and by implication digital copying and digital distribution, which today and into the foreseeable future equates to network access. With digital conversion, cultural institutions become concerned about distributed resources and distributed users. This leads us to consider quality and utility at the point of use much more than we ever had to face in the analogue world. Just as we are concerned about quality at the start of the process, so must we be concerned at its end. We must evaluate the effects of technical choices made in creation *and* presentation of digital image surrogates for a range of document types.

Although there is much anecdotal evidence, and a number of studies conducted involving digital image use, there has been no systematic assessment of the cumulative effects of technological choices on the networked display of digital image material. In 1990, Michael Ester published the results of an influential Getty study on users' assessments of digital image quality of works of art, based on such factors as resolution and bit depth. Subsequent studies have noted that the quality of digital masters will directly affect the quality, utility, and expense associated with creating derivative images for on-screen display, printing, and image processing. A number of institutions, the Library of Congress and the National Archives amongst them, have promulgated guidelines for the creation of access images for Web display. The National Archives and Corbis, Inc. have both gone so far as to create electronic grayscale targets designed to assist users in calibrating their monitors to optimise the presentation of delivered digital files. In addition, there are some helpful sites, including the YaleC/AIM Web Style Guide, which provide useful recommendations on choosing file formats and optimising graphics. A number of companies and institutions have also produced Web-accessible reports or graphic examples of the effects on image quality of specific technical choices, such as the type and level of compression used.

Despite their valuable contributions to the field, these resources have largely ignored how the range of technical choices - file formats, compression processes, scripting routines, transfer protocols, web browsers, and the like - can affect perceived image quality at a networked workstation. So too must we consider the lag in technology adoption at the user's end. For instance, the Museum Education Site License Project, co-sponsored by the Getty Institute, involved the creation of fine access images of paintings, photographs, and three-dimensional objects. These images met many of the researchers' needs and expectations at high end user workstations, but were almost useless in their largely posterised form when received on an older machine with a VGA monitor incapable of supporting more than 256 levels of colour.

As cultural institutions increasingly turn to the web to make retrospective resources accessible to a broader public, issues associated with image quality, utility, and delivery at the user's end get raised. Creating high quality digital image files is not an absolute guarantee that this will happen. The preservation community must become involved in decisions made at the point of image manipulation, scaling, compression, and other technical matters in order to safeguard the user's rights to access satisfactory digital information.

- **Acknowledge the gap between the ideal and the possible.** Having said all that has come before, I do think that we have to exercise judgement and do the best we can, given various resource constraints and the technological limitations of the here and now. Something can be conceptually strong, but just not work. Consider for instance the case of the Library of Congress, which went out with a Request for Proposals (an Invitation To Tender) for the imaging of graphic materials with a stated preference for image files that were 12 bits per channel. The staff discovered that it was difficult to store and use these files, and have used them instead to create, as does Corbis, good quality 8-bit per channel images - but they do not retain the larger bit files.

There is also an emerging debate within the library imaging community in the US over the use of compression in the digital master. Some argue that any compression is bad, but others argue that tough choices have to be made - do we save fewer items by making our requirements so demanding? The authors of the previously cited Strategic Policy Framework for Creating and Preserving Digital Collections concluded:

*"The standards and best practices which promised to facilitate and reduce the cost of a data resource's long-term preservation were not always those which promised best to facilitate and reduce the cost of its intended use. The standards and best practices, which promised to ensure a data resource's maximum fitness for purpose, were also not always affordable or technically achievable. Accordingly, the selection of standards and best practice frequently involved a range of compromises between data creation aims and costs."*

We will often need to make decisions that conflict with our quality requirements on the one hand or our financial capabilities on the other, this is to be expected, but we should be making those choices in an informed manner.

We need more information on lessons learned about when things go wrong (people are more than happy to tell you when they go right!). A good example of this is found in the TASI report, *Creating Digital Image Archives*. TASI visited several digital imaging library projects to identify what worked and what did not, and to determine whether current standards could be practically implemented. One particular problem, experienced by the project team of the Higher Education Library for Image eXchange (HELIX), arose when trying to buy scanners: "the advice given by suppliers and bureau's was technically wrong. For example, one supplier stated firmly that a scanner could be PC driven, when in fact it could only be run from a Mac." This, together with other technical problems led to some delays within the project.

- **Undertake the doable.** In a presentation to the 1992 annual meeting of the Research Libraries Group, Don Waters suggested a number of enabling principles with which to view the challenges of imaging technology. Among them were the KISS (Keep it Simple, Stupid) principle and the adoption of an incremental approach. These two points are still valid today. We should strive to undertake what is doable, from the perspective of our collective energies and knowledge base. Consider for instance the case of the International Colour Consortium, which undertook several years ago the challenge of solving all the issues associated with colour management. The task was too daunting - at least in the very near term, and the ICC strategy has now shifted to an incremental approach, addressing what is doable now and developing building blocks for an iterative effort. If we become too ambitious, we may end up further behind than if we had attempted nothing at all. I cannot stress enough the need for us to succeed in our deliberations. Let's plan for it.

- **Consider guidance rather than guidelines.** All these principles lead me astray from my appointed rounds to suggest that we consider guidance instead of guidelines. I suggest this for a number of reasons:

1. *One size doesn't fit all.* In a world with multiple stakeholders and multiple perspectives it becomes difficult to create guidelines that suit all circumstances. Institutional purposes vary, so it seems problematic to create hard and fast answers, to standardise technical choices that can scale across all institutions, all collections, and all use objectives. The imaging guidelines proposed by both the National Archives and the Library of Congress are similar in many ways, but differ in significant ways, based on conclusions experts at each of these institutions reached regarding their institutional mission, their collections, their users, and current technological constraints. In other words there is no one right way to do things that is applicable in all circumstances.

2. *Reasonable, well informed people can disagree.* We must acknowledge that differences of opinion in an uncertain world are a healthy sign. We need to experiment with a range of approaches and evaluate their impact over the long haul. But if experts can disagree on such

fundamental principles as compression and bit depth, and the use of enhancements at the point of capture, how do we sort out the "truth?"

3. *Digital imaging is a process not a prescription.* I am more of a believer in an approach than essential truths. Digital imaging is not so much a recipe as a series of choices that have to be made, in which we must balance competing requirements along the way. Guidelines by their very nature are static or restricted to the specifics of the environment in which they were produced, whereas guidance looks at the way things flow and adjusts accordingly. As we utilise this technology, we want to make informed decisions, to understand the consequences of choices made upstream to what flows thereafter, and to be prepared to right our course occasionally as things change.

4. *But people want simple answers, not a process.* And they will interpret guidelines more broadly than they should. Take for instance the recommendations coming out of Cornell for the use of 600 dpi bi-tonal scanning. This recommendation is based on an examination of a specific body of material (brittle books containing text and simple line art) that exhibits remarkably uniform characteristics typical of the commercial printing processes of the 19th and early 20th century. It does not extend to office documents, to other publications, to manuscripts, or to graphic and photographic materials - yet time and time again, I hear of institutions requiring 600 dpi for their imaging projects that have nothing to do with brittle books, because that is the Cornell "standard." Six hundred dpi is not for all things, and may not be required in a program where access alone is the driving force.

5. *It is often difficult to state requirements in terms of precise objective characteristics.* Sometimes what we want is for the image to look right, and that cannot always be quantified or characterised in terms of precise requirements. If we rely too closely on technical definitions, we may end up being too wedded to the numbers and sacrifice quality to consistency. Our goal is not to reproduce the grayscale bar but to convey the essence of the original object. Or it may not be that at all. Photo-historians refer to the concept of "rendering intent" in reformatting photographs - what the photographer meant to convey, rather than what is recorded on the photographic medium, which calls for an interpretative assessment. This, I would argue leads back to stating requirements in terms of what we want, rather than prescribing how to get there. Let me not be misunderstood. For some things we can and should take a definitive stand - we should require objective measurements, but understand them as *indicators* rather than as guarantees of image quality. In photographs, tone reproduction is the most important attribute of image quality: "If you get your neutrals right, most of your colour management work is done." We should require the use of targets, but understand that in relying on an evenly distributed grayscale, the image will look darker and muddier on screen and in print. The final decision on image quality should be made by a person sitting in front of a monitor, who decides what looks "good." As Supreme Court Justice Potter Stewart reputedly once said about pornography, "I cannot define it, but I know it when I see it."

6. *Our knowledge is not evenly developed.* We know more about some kinds of scanning than we do others, and we might be close in reaching agreement on best practices for some categories of materials. For other materials, specialists in their fields suggest that we are far from consensus on how best to achieve digitisation. And for some materials, the most that can be said may be that we know what we do not want. Given this situation, we can not be uniformly definitive in stating requirements for digital imaging. But we can be knowledgeable about the trade-offs associated with different courses of action.

I cannot help but wonder if we would not be better served by assuming the role of Sherpa rather than chef. If you will forgive the mixed metaphors, one can navigate with any number of sea charts. But the chances of arriving safely are greater if there is someone along who has made this trip many times, and knows how to avoid the shoals, depending upon the prevailing winds, the vessel used, and the behaviour of the gods. Where experts disagree or institutional missions conflict, our role may be the most beneficial if we can help institutions

interpret and then synthesise the differences of opinion, and the circumstances under which one approach may be better than another. In some areas, we should be prepared to challenge assumptions, or to offer a "Good Housekeeping Seal of Approval" for a particularly well reasoned methodological approach. We should also point out where solid work remains to be done, and alert others via a "watch this space" to initiatives underway that may shed some light on particularly thorny issues. Why, for instance, should we spend the next three days agreeing on guidelines that include imaging pictorial materials if the Visual Arts Data Service draft guidelines for "Creating digital information for the visual arts" offer us all we need? And, if they do not, then the Council on Library and Information Resources is also preparing to undertake this task. In other words, let us provide guidance rather than guidelines. Let us develop a product that allows us to continue to retain relevancy as the technology and our understanding of its uses change, and as others develop solid guidelines for specific aspects of this work. I believe that RLG/NPO are in an ideal position to vet for us the work performed by others: to examine, investigate, and evaluate it in a thorough manner - via a process in which we can have a great deal of confidence and trust.

### **So how does guidance work?**

It is pretty easy to suggest that we focus on guidance rather than guidelines, but just how do we go about this? We know how to do guidelines - we have a pretty good track record in this area - and the process itself represents a comfortable, well-trodden path. Developing guidance is more challenging, representing a paradigm shift in what we do, and making this shift will be challenging. I do, however, believe that we have no choice. Our heart may be in creating guidelines, but our head should be with guidance. We will ultimately fail in the attempt to create generalisable guidelines because guidelines are contextual - they are specific to institutional settings, and may not scale beyond those constraints or to other environments. As Sabinne Susstrunk from Corbis points out, "choosing the correct scan resolution is dependent on the purpose of the archive," not on some absolute. Even though the temptation is strong to extrapolate from one situation to another, the reality is that what works for the Library of Congress probably will not work for a local historical society.

Guidance, on the other hand, can be developed to enable an institution to step through a process with the end result leading to the creation of their own specific set of guidelines. Guidelines should be a by-product of this process, not the starting point. Deciding on a course of action, what to do and what not to do, should be based on the reasons for digitising, the nature of the source documents, the institutional mission, available resources, the technical infrastructure, and users requirements and capabilities. Guidance begins with understanding the context, provides a process for data gathering and decision-making, points to available guidelines and a means for assessing their applicability, and finally leads to the development of guidelines that are specific to institutional needs. Guidance should lay out a uniform process in which prejudices are stated up front, decisions are well defined and documented, and evaluation is built in all along the way. Guidance is an inductive process that enables institutions to specify, create, and implement their own guidelines, based on the specifics of their environment instead of relying on a pre-defined, limited set of options. One such model may be represented in the work of the Science Museum in the UK. The Museum is in the process of creating a digital image policy that includes "cost-benefit analysis of storage, access, and image quality options to guide staff in selecting appropriate standards and formats for image content which is being considered for digitisation."

As we break for the next few days into separate groups, we should think about providing a model process for guidance to be used by institutions to develop guidelines for digital imaging. We should think about the impressive amount of work already accomplished, and consider where agreement on basic requirements can be reached, for instance in the area of metadata. We should acknowledge where there is divergence of opinion based on solid reasoning, and strive to put this in the context of decisions made rather than universal truths.

We should recognise that preservation may be used in some instances and not in others, but that long-term utility should be a guiding force. We should consider the best "bang for the buck" in what we come out with, and come out with something doable - the world is watching. We should focus on a product that will best meet the needs of those of us assembled here today, because if we can meet our needs, I guarantee we will go along way in meeting the needs of others.



**Guidelines for Digital Imaging**  
**Janet Gertz**  
**Director for Preservation**  
**Columbia University**

### **Introduction**

There are digitally converted images everywhere today. Some projects are well designed and rich in information; others lack meaningful - or usable - content. Making worthwhile selection decisions means careful evaluation of the physical and scholarly nature of the materials, the intellectual property rights vested in them, and our ability to carry out a technically correct and cost-effective conversion project. Every topic to be discussed at this seminar plays its role in the initial decision of whether and what to digitise.

### **Selection for Preservation**

Models for selection derive from traditional preservation reformatting via microfilm (and also photocopying). Preservation selection criteria amount to four questions:

1. Is the item or collection damaged or endangered? This is the essential question for preservation.
2. Does it have sufficient enduring value to justify preservation, as indicated by:
  - artifactual features, e.g. bindings, illustrations,
  - historic importance,
  - distinguished intellectual content for teaching and/or research,
  - consonance with the mission of the institution,
  - support for historically important areas at the institution?
3. Can a microfilm or other preservation copy successfully capture its content and support current and predicted future use, or are there factors like use of colour, poor contrast, missing pages that might make reformatting inappropriate?
4. Does the cost of preservation fit the value of the item?

These criteria have been applied to three general classes of materials.<sup>(1)</sup> First are the artifactually valuable objects which must be permanently retained in original form - rare and unique materials, archives, visual materials, and so forth. Reformatting provides surrogates to prevent excessive handling, and to broaden access by distributing copies to distant readers.

Second are actively used damaged items with relatively low artifactual value, whose importance consists in their intellectual content. These are "higher use items that are currently in demonstrable demand for curriculum and research purposes.... Class 2 preservation is, in fact, really only an extension of, or supplement to, the core building and maintenance done by most selectors in most libraries. It is, in a manner of speaking, simply current selection by other means."<sup>(2)</sup> Preservation replaces the damaged item with a copy that will serve immediate demand, often a reprint or photocopy.

Third are the little-used, mostly brittle items which are preserved for their content, not their physical attributes. Many are scholarly materials which have fallen temporarily (or perhaps permanently) out of fashion. Little or no immediate demand is anticipated. The goal of reformatting is a microfilm of the original that will endure until some future researcher discovers an interest in the materials.

Unfortunately, although microfilm is a fine long-term storage medium, readers find it awkward and unpleasant to work with. Digital files are easy to access, but is digital conversion a form of preservation? After all, digital files reside on unstable media and must

survive repeated processes of migration or perhaps emulation which have yet to be fully worked out, and which require long-term commitment to funding of perpetual maintenance. If economic crises hit in future, will money be spent to migrate non-essential files?

The University of Michigan and a number of other libraries feel that in fact digital files should be employed as preservation master copies.

*"We consider all digital images captured, processed and stored to our specifications - done either in-house or outsourced to a vendor - to be archival. ... The Library and the University are committed to providing enduring access to these images, with funding to support system maintenance and staff. The TIFF images will be migrated forward to new storage and access systems as necessary. ... This Library's participation in Making of America, Phase I, was designed from the start to retain the digital file and withdraw the original. A paper facsimile from the digital file will be printed only when a user requests that service...."*<sup>(3)</sup>

In other words, only the digital version is intended to exist into the future.

Most other libraries hold the view expressed at the 1995 RLG digital selection symposium, that "digitisation appears to have a preservation role in reducing use of originals ... but appears not to be suitable for preservation of information that is preserved in no other form."

<sup>(4)</sup>The consensus among most American preservation officers is that digitisation alone does not constitute preservation <sup>(5)</sup>:

- digitisation is access-driven,
- it is not a replacement for microfilming or other preservation activities,
- selection for digitisation does not arise from the same motivation as selection for preservation,
- the original object or analogue copy serves as the long-lived preservation medium, while the digital version serves extremely well as the access medium,
- special collections and visual materials take priority for digitisation because the technology suits them and they suit the technology.

To cite several examples, Harvard uses digital versions for access, and analogue media for long-term storage; Columbia and Chapel Hill apply scanning for access to primary sources which are retained afterwards and for access to high-use course-related materials; Indiana uses digitisation to print out acid-free copies instead of photocopying; and Yale's Project Open Book created digital versions from already-existing preservation microfilm.

In selection for digitisation, the need for preservation appears secondary to improving access. Tamara Swora of the Library of Congress suggests that *"in digital reformatting, we must learn to think first about defining the context as a product, considering how it will be disseminated and used, and then selecting materials for conversion.... We must also shift our attention to access as the dominant objective. In many of our preservation reformatting programs, we may be required to move away from our traditional priority of selecting materials at risk because of their physical condition to materials where use is the decisive factor. Digital technology forces us to consider access before preservation...."*<sup>(6)</sup>

At the Library of Congress digitisation is "primarily for improving access to materials that should be put on-line because of high informational and visual value. Digitisation is aimed mainly at Special Collections and non-book collections."<sup>(7)</sup> The National Library of Medicine "doesn't really select materials for digitisation"; they do not "consider this activity to be preservation other than in the sense it reduces use of the originals."<sup>(8)</sup> The New York Public Library likewise selects collections which "need to be made widely available," regardless of condition.<sup>(9)</sup>



### Selection for Digitisation

Preservation deals with materials which are endangered or already damaged. The basic purpose of preservation reformatting is to create as accurate as possible a reproduction of deteriorating originals on a long-lasting medium, and of course to record the existence of the film in international databases like RLIN. Digital conversion deals with materials which are in demand, regardless of their condition. The basic purpose of digitisation is to create reproductions that can be viewed by as many people as possible as easily as possible, and also to enhance those reproductions and the ability to search and manipulate them. Preservation reformatting and digital conversion overlap where materials are both endangered and in demand. Applied to the three classes of materials discussed earlier:

- **Class 1:** Digitisation provides high quality, easily accessed surrogates for valuable and fragile materials. Preservation is served through the protection from handling that endangered materials receive. The originals are retained.
- **Class 2:** Easily accessed digital images make good sense for high use materials. Where materials are deteriorating and need permanent replacement, Michigan and some others are willing to depend on the survival of the digital files. Most other libraries prefer to produce microfilm or other durable copy in addition to the digital access version.
- **Class 3:** There is little enthusiasm for digital conversion of low-use materials, since quick and broad access is not required. Where the materials are embrittled and must be replaced by preservation copies, most libraries still prefer microfilm, since it is stable, and cheap and easy to store.

To make intelligent decisions about digitisation and preservation, we must know why we want to digitise. Is the goal to provide a quick overview or a browsing tool for a collection? To create a research tool with high resolution images and full text searching? To provide better access to originals which will be or have already been preserved? Is it designed to protect originals by obviating the need to handle them? Is it designed to create master digital files to permanently replace the originals? There must be clear understanding of - and agreement on - the goals of the project, because those goals will drive decisions about selection of materials, image quality, indexing and searching, and preservation of the digital files.

Selection criteria for digitisation comes down to whether the materials should be digitised and whether they can be digitised:

- Does the item or collection have sufficient value?
- Is there a current audience to justify digitisation?
- Do we have the legal rights to create a digital version?
- Can the materials be digitised safely and successfully?
- Do we have the infrastructure to carry out all aspects of a digital project?
- Is the cost appropriate, and can we afford it?

### Value of Content

The first criterion is the same basic question that holds for any kind of selection - new acquisitions, collection review, digitisation, or preservation. Does the content of the materials justify the expenditure of effort and resources? Specific definitions of value will vary from institution to institution, but in general terms:

- Is the proposed item or collection unique?
- What is its intellectual, historic, physical value?
- Is it one of the best in the subject?
- Is its content broad, deep and accurate?
- How does it fit with other materials on the same subject?
- Does it further the work of the institution and its clients?

- Does it support historic strengths of the institution?
- Does it have active current users?

All selection guidelines focus on enduring value of content. Some libraries speak very generally, while others seek to specify in some detail what materials are of particular value to them. For instance, consider general phrases like "unique combination of information sources"<sup>(10)</sup> or "accurate, representative, insightful".<sup>(11)</sup> The National Agricultural Library's guidelines are among the more specific in defining what is important to its clients.<sup>(12)</sup>

Research value may reside in the individual object or in an aggregation of objects. Selection has to evaluate whether a project should include all, or only portions, of a collection. If the individual titles or documents are separately of significant value, then even a small sample is worthwhile. With ephemera, archival materials, or photograph collections, often the aggregate has greater research value than the individual items. Will a project to scan representative samples taken out of context result in a meaningful product?

#### **Demand for the Materials**

Next, is there a demand for the materials? The level of current demand or likeliness that materials will see significant use once they are digitised is of great concern for selection. RLG SHARES representatives in 1994 reported that they "preferred use over other types of selection methods for materials to be held in the form of digitised records."<sup>(13)</sup> Strong potential for an immediate audience can be predicted from various factors. Does the collection support current high priority activities, for instance materials used in teaching core courses which get heavy use and are seen as central to the work of the institution? Is there an active but widely, perhaps internationally, dispersed audience which needs better access to the materials? A specialised collection is not necessarily low-use, since a small but intense group of specialists can generate a great deal of activity. Or, do restrictions on the handling of fragile or valuable originals create a source of immediate demand?

*In general, "collection managers must always look at the local community and its needs. When these needs coincide with a broader good to the scholarly community at large, the decision for digitising a particular collection gains greater weight.... We need to be assured... that there is a user community for these digital collections that will make the expenditure of funds and effort worthwhile...."*<sup>(14)</sup>

#### **Intellectual Property Rights**

Once we have established that the materials merit digitisation, we must then determine if we have the legal right to make digital copies, and if not, whether we can get clearance from the holder of the copyright.

Preservation microfilming has concentrated on printed matter which is no longer under copyright. Most institutions have filmed to replace or protect damaged items, and have done so primarily for local use or provision of copies for clearly research purposes. Digitisation increasingly focuses on unique archival and visual materials, many of them unpublished. Copyright rules are more complex, and we run into unclear histories of ownership and multiple layers of authorship.

Intellectual property issues become murkier as online versions join the mix, because they are universally accessible, not open only to scholarly purposes. To protect intellectual property from unauthorised use, some libraries limit access to institutional affiliates. Other institutions mount low resolution images which are inappropriate for printing. Unfortunately,

the resolution may also be too low for serious research use. The tension between protecting rights and broadening access must be carefully weighed before materials are selected.

### **Safe, Successful Scanning and the Necessary Infrastructure**

Next is the question of whether the materials can be captured adequately, whether digital versions can be made that will actually fulfil readers' needs, and whether digitisation can be accomplished without causing damage to the originals, for instance through exposure to strong light or the need to disbind. These topics will be covered in detail tomorrow, but they are also part of selection. An informed selection decision can only be made with a solid understanding of whether digital versions of adequate quality can in fact be made, stored, and distributed. Selectors must have some knowledge of what it takes to produce archival-quality resolution and colour fidelity in digital images, how to cope with materials characterised by large size or fine detail and real-world limits on manipulation of very large file sizes.

Users must be able to access files without unusual delays or crashing machines. But projects must also be designed for future viewers with better equipment, or today's efforts will quickly become outdated. One answer is to capture images at high resolutions necessary to replicate the content of the material, but provide lower quality derivatives which match the capacity of current equipment. Selection includes thinking about not only whether we can capture materials at high quality, but also whether we can deliver legible images conveniently through commonly used equipment and software.

Furthermore, an informed selection decision cannot be made without understanding whether a project is feasible in terms of preparation, cataloguing and intellectual control, creation of metadata, navigational tools, and other work which makes digital images functional. Again, these issues will be discussed in detail tomorrow, but it is essential to factor into selection all of the tasks which are not part of image capture, but which are basic to an effective project, because achieving them consumes a significant amount of resources.

Beyond asking if the institution has the know-how and infrastructure to digitise the materials, there is the question of added functionality or added value. In contrast to microfilming, which simply produces reproductions of the originals, digitisation is expected to do more. "The decision regarding what to digitise is absolutely interconnected with the question of how we digitise it. Digital capture will not ultimately be sufficient."<sup>(15)</sup> Selection guidelines ask whether "the product have more value than the source material because of indexing, links, finding aid availability, added text or images or other features?"<sup>(16)</sup>

An important aspect of digital images is their ability to improve documents' legibility. Part of selection is considering whether such enhancements ought to be made to the images, and whether this might impinge on their historical accuracy. Should both accurate and enhanced, more legible versions be provided? How should the changes be documented? What are the implications of enhanced images of visual materials?

Another crucial improvement is better ability to search texts. OCR and other added search and manipulation capacities appear to be a necessity when converting printed matter. At the RLG selection symposium "it was felt that researchers would not want to sit at the terminal to read large quantities of text and that, unless the text were searchable, digitisation added little value to the research potential of the materials. Indeed, capturing large bodies of text as image would provide primarily a storage medium, for which microfilm is probably better adapted."<sup>(17)</sup> Yet another wrinkle for the selector to take into account.

### Cost

Once we have ascertained that the digital project's design is appropriate and feasible, we next deal with the cost of digitisation compared with other options, avoiding duplication of effort, and determining if funding is available. At present digitisation appears to cost more than microfilming when all costs for image capture, preparing the materials, and intellectual control activities are counted. Certainly that has been the experience at my institution, and at the National Agricultural Library, whose guidelines state<sup>(18)</sup>:

*The purpose of these Guidelines is the selection of materials for electronic preservation, but a consideration of microfilming is still in order... There was general agreement that (1) microfilming is cheaper than digitising, in the long run; (2) digitising provides better access and cheaper distribution; (3) it is cheaper and easier to scan older materials from microfilm than from the paper source; (4) digital information is the wave of the future.*

Selectors must weigh whether the materials could be adequately reproduced through other means at lower expense, but unfortunately it is truly difficult to find accurate cost figures. In 1997, for instance, Library of Congress estimated a price of \$500 per 300-page book<sup>(19)</sup>, while Cornell in the same year estimated approximately \$60 per 300-page book<sup>(20)</sup>. It is not easy to discover what causes such wide discrepancies because there are so many variables involved.

Availability of funding is of course a deciding factor. Occasionally an institution may recoup its expenses by creating a marketable product, or by contracting with a publisher who will pay royalties for digital images of high-profile and rare items. It seems likely, however, as Demas says, that publishers and other for-profit organisations will convert most items with any commercial value, and libraries will be left converting the "lesser-used, more obscure, alternative, or specialised materials - in other words, materials judged important to scholarship for reasons other than high demand."<sup>(21)</sup>

Much project funding comes from grant agencies or co-operative programs. The institution must determine if a collection of high local priority can meet the selection criteria established by an external funding organisation. Looked at from the other direction, is any collection that happens to fit the funding organisation's criteria also important enough locally to justify applying for and carrying out a project? Desire to receive grant funds can easily skew local priorities into digitising something of secondary importance, while top priority collections languish for lack of money because they do not fit any grant organisation's criteria.

### Preservation of Digital Resources

Finally, there is preservation of the digital files we are producing. "Selection for preservation in digital form is not a one-time choice made near the end of an item's life, but rather an ongoing process intimately connected to the active use of the digital files."<sup>(22)</sup> We are all depending on current investigations to tell us how to preserve digital resources - for instance, the work of the Consortium of University Research Libraries (CURL) CURL exemplars in digital archives (CEDARS) project which will "address the strategic, methodological and practical issues and will provide guidance for libraries in best practices for digital preservation."<sup>(23)</sup>

Digital selection guidelines call for genuine commitment to preserving the files, and for the willingness and ability to develop the infrastructure necessary to assure that digital preservation activities will be carried out routinely over the long-term. The recent Council on

Library and Information Resources (CLIR) publication, *Selecting Research Collections for Digitisation*, sums it up well <sup>(24)</sup>:

*Digital preservation ... requires a supporting organisation and infrastructure dedicated to storing the electronic files and to migrating them to new formats and/or media as technologies change. Unless these capacities are all in place, digital files cannot be regarded as permanent. Creating an enduring digital preservation master file is a multidimensional task with long-term implications. Hybrid projects, in which digital files are complemented by copies on microfilm, alkaline paper, or some other stable medium, provide the insurance that exclusively electronic projects do not.*

### Conclusion

In conclusion, huge masses of material deserve preservation and improved accessibility. Not all are well suited for digitisation, and funds are limited. Guidelines for selection will help us decide intelligently what to preserve and what to digitise. *"The overriding concern is a fit of the nature of the content and the uses to which it will be put.... It becomes clear that no common answer suits all uses and all users. The library's responsibility is to weave all of the aspects of use - need and purpose for the content, the content itself, the medium by which the content is delivered - into the provision of a service that meets the user's need.... Also implicit in the above notion is the importance of distinguishing the possible from the desirable.... A particular medium, in and of itself (whether it be print on paper or hypermedia), does not necessarily contain the most effective uses of content."* <sup>(25)</sup>

As we work toward determining what materials are worth preserving and digitising, as we investigate intellectual property rights and evaluate our ability to create well-designed digital and preservation projects, and as we figure out how to preserve the digital products, I hope that this seminar will give all of us tools to make the right decisions.

### REFERENCES

1. Originally discussed by Ross Atkinson, "Selection for Preservation: A Materialistic Approach," in *Library Resources & Technical Services* 30:341-62 (1986).
2. Ross Atkinson, "Selection for Preservation: A Materialistic Approach," in *Library Resources & Technical Services* 30:341-62 (1986), p. 346.
3. Carla Montori, Head of the Preservation Division, University of Michigan Library, e-mail posting to Conservation OnLine, December 15, 1997.
4. Barclay Ogden, "The Preservation Perspective", in *Selecting Library and Archive Collections for Digital Reformatting: Proceedings from an RLG Symposium held November 5-6, 1995, in Washington, DC*. Ed. Ricky Erway. Mountain View, CA : RLG, 1996: 115-117, p. 116.
5. Summary of comments made June 28, 1998 at the Preservation and Reformatting Discussion Group meeting at the annual meeting of the American Library Association.
6. Tamora Swora, "The Preservation Context", in *Selecting Library and Archive Collections for Digital Reformatting: Proceedings from an RLG Symposium held November 5-6, 1995, in Washington, DC*. Ed. Ricky Erway. Mountain View, CA : RLG, 1996: 47-52, p.48

7. Debra McKern, Chief, Binding and Collections Care Division, Library of Congress. Handout distributed at American Library Association meeting, June 1997.
8. Margaret Byrnes, Head, Preservation and Collection Management Section, National Library of Medicine. E-mail, April 20, 1998.
9. Joan Gatewood, Associate Chief for Preservation Reformatting, New York Public Library. Telephone conversation, April 1, 1998. Arizona State University Library, "Digitisation Program Pilot Project Application," March 1998.
10. Smithsonian Institution Libraries, "Guidelines for Developing an SIL Digital Project," October 1997.
11. Selection Task Force of the Electronic Preservation Committee, "Selection Criteria and Guidelines: Selecting Materials for Digital Preservation at the National Agricultural Library," National Agricultural Library, November 1995
12. Paula De Stefano, "Use-Based Selection for Preservation Microfilming," in *College & Research Libraries* 56:409- 418, 1995, fn. 2, p. 417.
13. Linda Matthews, "The Collection Management Perspective", in *Selecting Library and Archive Collections for Digital Reformatting: Proceedings from an RLG Symposium held November 5-6, 1995, in Washington, DC*. Ed. Ricky Erway. Mountain View, CA: RLG, 1996:99-103, p.100-101.
14. Margo Crist and John Price-Wilkin, "The Access Context", in *Selecting Library and Archive Collections for Digital Reformatting: Proceedings from an RLG Symposium held November 5-6, 1995, in Washington, DC*. Ed. Ricky Erway. Mountain View, CA: RLG, 1996:29-38, p.37.
15. Smithsonian Institution Libraries, "Guidelines for Developing an SIL Digital Project," October 1997.
16. "Selection Exercise: Discussion", in *Selecting Library and Archive Collections for Digital Reformatting: Proceedings from an RLG Symposium held November 5-6, 1995, in Washington, DC*. Ed. Ricky Erway. Mountain View, CA: RLG, 1996:119-122, p.120.
17. Selection Task Force of the Electronic Preservation Committee, "Selection Criteria and Guidelines: Selecting Materials for Digital Preservation at the National Agricultural Library," National Agricultural Library, November 1995.
18. Debra McKern, Chief, Binding and Collections Care Division, Library of Congress. Handout distributed at American Library Association meeting, June 1997.
19. Anne Kenney, *Digital to Microfilm Conversion: A Demonstration Project 1994-1996. Final Report to the National Endowment for the Humanities*. Ithaca, NY: Cornell University Library, 1997. Available at <http://www.library.cornell.edu/preservation/com/comfin.html>
20. Samuel Demas, "What Will Collection Development Do?", in *Selecting Library and Archive Collections for Digital Reformatting: Proceedings from an RLG Symposium held November 5-6, 1995, in Washington, DC*. Ed. Ricky Erway. Mountain View, CA: RLG, 1996, p.18.

21. Society of American Archivists, "Statement on the Preservation of Digitized Reproductions", available at [http://www.archivists.org/governance/handbook/app\\_k.htm#Preservation](http://www.archivists.org/governance/handbook/app_k.htm#Preservation), approved by the SAA Council June 9, 1997.
22. Consortium of University Research Libraries, CEDARS Project website, available at <http://www.curl.ac.uk/cedarsinfo.shtml>
23. Dan Hazen, Jeffrey Horrell, Jan Merrill-Oldham, *Selecting Research Collections for Digitisation*, Washington, DC: Council on Library and Information Resources, 1998, p.14.
24. John Budd and Bart Harloe, "Collection Development and Scholarly Communication in the 21st Century: From Collection Management to Content Management," in *Collection Management for the 21st Century: A Handbook for Librarians*, ed. G.E. Gorman and Ruth Miller. Westport, CT: Greenwood Press, 1997 (pp.3-25), p.18.
25. John Budd and Bart Harloe, "Collection Development and Scholarly Communication in the 21st Century: From Collection Management to Content Management," in *Collection Management for the 21st Century: A Handbook for Librarians*, ed. G.E. Gorman and Ruth Miller. Westport, CT: Greenwood Press, 1997 (pp.3-25), p.18.





## Guidance for selecting materials for digitisation

Paul Ayris

Director of Library Services

University College London

### The Context

It is timely for the Information profession to consider the managerial implications of the digitisation process. Within existing academic libraries, which is my own area of interest and expertise, well-developed mechanisms are to be found for selecting print on paper materials. For digitisation to take its place in the Information world as a standard tool in the armoury of professionals, similar arrangements must exist in the electronic arena.

A number of published studies set the need for the selection of materials in the context of the whole digitisation process. A recent JISC/NPO study into long-term archiving concluded that, for preservation, responsibility for drawing up selection guidelines could be assigned to a number of different agencies depending on the type of material under consideration. The basis for selection should be the permanent value of the data or product. The authors maintained that the same criteria used for selecting print materials could be used for electronic publications. The responsibility for developing detailed acquisition policies should lie with the legal deposit libraries or their agents<sup>(1)</sup>. In a more recent study, Beagrie and Greenstein also see the selection process as being firmly embedded in the policy framework regarding digitisation.<sup>(2)</sup> A recent investigation by the Data Archive at the University of Essex looked at selection in the context of unpublished research materials.<sup>(3)</sup> The prevalent view of selection to emerge from the Data Archive's survey can be summarised as a series of difficulties:

- How to predict what will be useful in the future
- How to know when data can be acquired for preservation and access if researchers are still working on it
- How to ensure the integrity of, and responsibility for, data which needs constant updating

A thorough review of selection in the digitisation process has recently appeared in the United States, a study which rightly sees selection of materials as deeply embedded in the whole digitisation process. The conclusion deserves to be quoted here:

*The process of deciding what to digitise anticipates all the major stages of project implementation. Digital resources depend on the nature and importance of the original source materials, but also on the nature and quality of the digitising process itself - on how well relevant information is captured from the original, and then on how the digital data are organised, indexed, delivered to users, and maintained over time. Disciplined efforts to address the themes and questions outlined in this essay will help ensure that new digitising projects fulfil the expectations of libraries, students, and scholars.<sup>(4)</sup>*

At an operational level, a number of institutions have developed local policies for the selection of material for digitisation. Such policies are hard to find, since few are published. In a search of the World Wide Web, the following policies have come to light. The National Digital Library Program at the Library of Congress has a project planning checklist, which divides the selection process into a series of six interlinked steps.<sup>(5)</sup> The University of California has well-developed selection criteria for digitisation, which are divided into two main processes containing a series of twenty steps.<sup>(6)</sup> The University of Columbia has also developed a set of selection criteria for digital imaging projects, which are divided into three categories:<sup>(7)</sup>

- Collection development
- Handling and use
- Added value

It is a significant, and probably an accurate, sign of a general lack of activity that no guidelines have been discovered on institutional web servers at universities in the UK. Harvard has adopted an interesting approach by developing a decision-making matrix for selecting materials for digitisation. This matrix arises naturally out of the conditions which obtain in that university. There is a cluster of nine questions with further elaboration of some of the issues involved. The questions can be summarised as:

- Does the material have sufficient intrinsic value to ensure interest in digitisation?
- Will digitisation significantly enhance access or increase use by an identifiable constituency?
- What goals will be met by digitisation?
- Does a product exist that meets identified needs?
- Are rights and permissions for electronic distribution securable?
- Does current technology yield images of sufficient quality to meet stated goals?
- Does technology allow digital capture from a photo intermediate?
- Are costs supportable? Does an institution have sufficient expertise in project management? Is the local organisational and technical infrastructure adequate?
- Can the project be re-defined to recast objectives? Can infrastructure needs be addressed?

The approach is interesting, since at any stage in the process the digitiser can answer 'No' and so halt the work.<sup>(8)</sup>

### **The Task**

Most libraries have collection development policies for traditional print on paper materials. Such documents form the core of a library's collection management strategy. In a digital world, however, such documents will not in themselves give sufficient help to those who wish to select materials for digitisation. What is needed is some guidance, or a set of guidelines, to further this process. In truth, it can only be guidance rather than guidelines because relatively little is stable in a digital environment. Where, for example, are the costing models which help us study the economic issues in the digitisation process? What is given here, therefore, is guidance in the form of a decision-making matrix which can underpin the selection process for materials which are candidates for digitisation. There also follows a case study in my own institution, where the effects of implementing this matrix are evaluated.

### **Categories**

The sorts of questions which need to be addressed in formulating such guidance can be grouped under the following heads:

- Assessment
- Gains
- Standards
- Administrative issues

### **Assessment**

1. What level of support is there amongst target user groups?
2. Is digitisation consistent with local collection development policies?
3. Would digitisation make a contribution to local or national electronic resources?
4. Is there another product which meets these needs?
5. Is this digitisation for preservation or digitisation to enhance access?

Assessment is really the prelude to all other activities in the selection process and should naturally come first. Let us study one or two of these questions in greater depth. There is no point in selecting materials for digitisation if there is no support for using the resource amongst target user groups (see no. 1 above). This tenet is identical to guidance in a conventional collection development policy. No paper material would be purchased by a library if potential use could not be identified amongst library users. Again, take the point about existing products (see no. 4 above). There is no point undertaking digitisation activity locally if there is a product elsewhere which fits the bill. This idea is also to be found in conventional collection development policies, where the needless purchase of duplicated material is to be avoided.

### Gains

1. Does digitisation significantly reduce the handling of fragile originals?
2. Material which has been digitised should enhance the academic use of the work by
  - the creation of finding aids
  - links to bibliographic resources
  - links to online records
  - creation of training materials
3. Navigation should be easy
4. Where collections are split amongst different sites, the result should be a virtual collection which unites disparate material
5. Where originals are damaged, text and images should enrich the academic use of the collections

For digitisation to be a success, there have to be palpable gains in undertaking such work in the first place and this is the theme of the points in this group. Take just one example, namely the creation of training material (see no. 2 above). No one would expect a user, who could not read Chinese characters, to cope with material in that language and script. The same is true of digital material, where the necessary training materials should accompany any digital image. Use of the text, or images, and links to embedded resources should be covered as well as use of standard software tools, which can be used to study the digitised resources.

### Standards

1. Do the standards being used meet national/international standards, yielding images of suitable quality?
2. Will the resources thus digitised be available from the variety of hardware platforms supported by your institution?
3. Is the software used to deliver the materials readily available and easy to use?
4. Does the metadata conform to agreed international standards, e.g. Dublin Core?
5. What are the requirements for archiving in terms of hardware, software and data migration?

Archiving (see no. 5 above) is a big issue, particularly in academic research libraries. Such libraries acquire print-on-paper, secure in the knowledge that such materials will continue to be available in fifty years time. No such security of thought can be present in the digitisation process. Technology changes too fast, with hardware and software becoming obsolescent very quickly. In the UK at least, libraries are only just beginning to grapple with the problem of electronic archiving and few university institutions currently have systems in place to cope with it.

### **Administrative issues**

1. Is there sufficient finance to meet the costs of digitisation, and does the outcome of the digitisation process meet the requirements of the funding body?
2. Have copyright permissions be obtained and rights issues addressed?
3. Does the institution have sufficient expertise to carry through the project?
4. Does the digitised resource enable you to create a partnership with a commercial provider?
5. Do the benefits of digitisation justify the costs of doing it?

Many of these issues are so obvious as to require no further comment here. The proposed cost-benefit analysis (see no. 5 above) is an interesting area. For it to work, all costs should be taken into account and there should be a comparison with the cost-benefits of acquiring and storing print-on-paper. It is unlikely that many academic libraries have the figures immediately at hand to undertake such work.

### **UK-based Case Study**

What would be the effect today of using this guidance in a UK academic library? For the purposes of this study, I will take my own institution at University College London. During the last ten years, College has doubled in size due to a complex series of mergers with other institutions in London. There are now approximately 15,000 students and 5,000 staff.

### **Digitised teaching material?**

During the same period, transactions at the Issue Desks have increased fivefold. Clearly, the library service is far busier than it was a decade ago. One of the possible solutions to delivering services in this environment is to digitise more teaching material and make it available over the campus-wide network. There is a three-tiered library committee structure in College:

- College Library Committee
- 8 Faculty Library Committees
- Departmental Library Committees

This committee structure is relatively new, but is already working well. As Director of Library Services, I would have to consult all these committees to find out:

1. If electronic resources are embedded in the curriculum
2. Whether the electronic delivery of material to support taught-course provision would be acceptable

### **Co-ordination of work within Library Services**

If the guidance on action outlined above were accepted, there would certainly be a problem of co-ordination since many of the tasks embedded in the digitisation process cut across traditional library boundaries. The following categories of staff would be affected:

- 23 Subject Librarians
- Staff in the Central Cataloguing Unit and Periodicals Department
- Staff in Planning & Resources
- Staff in IT Services
- Staff in the Subject Support Unit

What would this mean in practice? Subject Librarians are responsible for academic liaison and would clearly be involved in the process of selecting material, liaising with academics and for providing training in the use of the final resource. The Subject Support Unit in UCL

is directly responsible for taught-course support and would be responsible for dealing with the necessary rights issues. Cataloguing and Periodicals would be involved in metadata and licence issues. Planning & Resources would be involved in financial matters, ensuring that all costs could be met. IT Services would be involved in technical issues, ensuring that the resource could be networked centrally and that it could be made available on all necessary hardware platforms. How can all this work be co-ordinated and who is responsible for ensuring that all steps in the chain of actions have been taken?

### **Value-added features**

One of the gains to be made from digitising original materials is the creation of value-added features, which enhance access to damaged originals. An instance of this can be found in Cambridge amongst the digitised fragments of the Taylor-Schechter collection. The Taylor-Schechter Genizah comprises tens of thousands of fragments, dating from the early Middle Ages, which were found in Cairo. There are religious texts, but also an enormous quantity of other material which sheds shafts of penetrating light on the Mediterranean world of Judaism at this time. Many of the fragments are damaged and digitisation has given the Taylor-Schechter Unit the ability to move fragments around on the screen, avoiding damaging contact with the originals, in order to try to re-create the original form of the manuscripts. The work of the Taylor-Schechter Unit can be found at <http://www.lib.cam.ac.uk/Taylor-Schechter/>.

### **Long-term archiving**

One of the issues discussed above is the need to ensure long-term archiving of the digitised materials. When libraries buy books and periodicals in conventional format, they do so in the expectation that the originals will still be accessible in fifty years time. The same model needs to apply to electronic libraries, at least as far as research libraries are concerned. In many universities, however, libraries have no control over the technical process of digital archiving. This falls under the remit of the central Computing Service, who will take an institution-wide strategic decision on how best to meet this need. In addition, many institutions in the UK are only just beginning to grapple with this problem. CURL, the Consortium of University Research Libraries, is undertaking a study of digital archiving as it affects research-based libraries. They are looking at CD-ROMs, websites, dynamic e-mail discussion lists and other sorts of resources to try and identify models for archiving each type of resource.

### **Decision-making Matrix**

Using the twenty issues/questions outline above under categories, it is possible to arrive at a basic decision-making matrix to inform the selection process in a digital arena. This matrix is given in full at the end of this article. The matrix is, of course, not in itself a collection development policy. Rather it is a decision-making tool, which informs the process of selection, which itself forms part of a statement about collecting strengths. It would be perfectly possible for the decision-making matrix to be applied to the same resource in two different libraries and for different decisions to be made as to whether the resource should be digitised. This would be quite in order, as a similar outcome could be expected when selecting conventional paper materials for purchase. It might also be possible for individual libraries to assign different weights, or values, to each of the questions in the matrix in order to ensure that they reflect local conditions as accurately as possible. What *is* important, however, is that libraries validate their selection procedures for digitisation with reference to external, internationally agreed, criteria. Digital libraries are best formed through national and international collaborations, especially since collection development is increasingly a collaborative activity between libraries. In addition, external funding is most likely to be

available where proposed digitisation programmes meet agreed criteria in terms of preparation, selection and image capture. It is important that the selection process does not become isolated from national and international debate. The decision-making matrix is offered, therefore, as an aid to decision-making, and as a tool which can be applied in a variety of libraries and archives, but which is sensitive to the context of all.

### Conclusion

This paper has identified a number of recent studies which have considered the role of selection in the process of digitisation. It is clear that some recent essays have taken the question very seriously and are treating it as a core part of the management process in a digitisation project. A number of operational guidelines also exist on websites, although little such activity is to be found in the UK. This study suggests a decision-making matrix of twenty questions grouped around four issues to aid the selection process:

- Assessment
- Gains
- Standards
- Administrative issues

Methods for implementing the matrix are also suggested. For co-operative digitisation projects to succeed, and for funding bodies to consider substantial financial investment in such projects, it is desirable that all digitisation projects validate their activity by reference to agreed international criteria. A decision-making matrix is not in itself a collection development policy, but it *does* form a contribution to the creation of such a policy. A case study is then offered for the proposed implementation of the decision-making matrix in a typical research library in the UK. It can be seen that many of the decisions required by implementing the matrix cut across traditional administrative boundaries in libraries. In addition, the demands of long-term archiving place an almost insupportable burden of responsibility on local institutions. In the UK, few academic institutions are equipped to implement the decision-making matrix without changes to their internal structures and decision-making processes. The future is exciting, but it poses some interesting challenges for any body wishing to invest seriously in digitisation to deliver materials to their users.

### REFERENCES

1. Haynes, D. Streatfield, T. Jowett and M. Blake, *Responsibility for digital archiving and long term access to digital data*, Elib study P1 (London, 1997), p. 44.
2. N, Beagrie and D. Greenstein, *A strategic policy framework for creating and preserving digital collections*, British Library Research and Innovation report no. 107 (London, 1998), section 4.5.
3. The Data Archive, University of Essex, *An investigation into the digital preservation needs of universities and research funders: the future of unpublished research materials*, British Library Research and Innovation report no. 109 (London 1998), pp. 10-11.
4. D. Hazen, J. Horrell, J. Merrill-Oldham, *Selecting research collections for digitisation* (Washington, 1998), p. 18.
5. NDLP, *Project planning checklist*, available at <http://lcweb2.loc.gov/ammem/prjplan.html>

6. University of California, *Selection criteria for digitisation*.

7. University of Columbia, *Selection criteria for digital imaging projects*, available at <http://www.columbia.edu/cu/libraries/digital/criteria.htm>

8. Hazen, Horrell and others, *Selecting research collections*, p. [19].

### Decision-making matrix to support selection activities in the digitisation process

	Question 1	Question 2	Question 3	Question 4	Question 5
Assessment	Is there user support?	What are local collection development policies?	Does this form a national or international contribution?	Does a similar project already exist elsewhere?	Is this conservation or preservation?
Gains	Does digitisation reduce wear on the originals or open up access?	Is the intellectual content of the work enhanced?	Is navigation easy?	Are disparate collections unified?	Is use of the damaged original material enriched?
Standards	Have suitable standards been followed?	Are the originals available from a variety of hardware platform	Is the software available and easy to use?	Does the metadata conform to agreed standards?	What are the archiving requirements?
Administrative Issues	Do you have enough money?	Have copyright and rights issues been secured?	Does your institution have enough expertise?	Is there a partnership with a commercial provider?	Do the benefits justify the costs?





**Preparation of Materials for Digitisation**  
**Ann G. Swartzell**  
**Senior Preservation Program Officer**  
**Harvard University Library Preservation Centre**

I will address the preparation of materials for digitisation, particularly using a model based on my experience in preservation microfilming production workflow planning. I contend that the reformatting process in the preservation realm of activity, usually moving materials from a brittle paper format to a more stable medium, is valuable as a model for conversion of 'legacy documents' in the world of digital resources. My role in preservation microfilming projects has been as facilitator of projects, working between the curators and bibliographers who had responsibility for embrittled paper based collections and the filming service bureaus and in-house labs that produce preservation microfilm. I feel this same role is necessary and useful in the conversations between curators and bibliographers and the Information Technologists who maintain our servers and web spaces.

The five 'classic steps' in the preservation microfilming workflow, as outlined in the title *Preservation Microfilm...* edited by Nancy Gwinn as well as the subsequent RLG microfilming manuals are:

1. Selection/identification
2. Preparation of materials
3. Filming and Quality control
4. Access and bibliographic control
5. Long term maintenance and storage

Here I will focus on the fine points of editorial preparation of materials as other papers address the physical handling and preparation of materials.

My first experience with a digital production plan of work was a project with some units of the library of the University of California, Berkeley, where project goals were to replace preservation photocopying with a scanning process, providing both hard copy text to shelve and a digital file for patron access via the web. Details of this project, one outcome of which was the Document Type Definition (DTD) known as Ebind, are available at Berkeley Digital Library SunSITE, (<http://sunsite.berkeley.edu/Ebind/>).

**My personal epiphany of the project: digital files are very different from microfilm!**

Digital resources are not the same as microfilm. The bits and bytes generated by the computer, in whatever format, on whatever platform, mean nothing without an operating system and application software - microfilm requires only magnification (or very good eyes) and a light source to be used. The microfilm frames appear in a sequence which offers the user a first attempt at how to access the order of items; there is nothing 'intuitive' about ordering of electronic files. Though some question this point, I suggest that a good way to learn about how to make microfilm is to look at a variety of microfilm, produced by commercial vendors as well as non-profit institutions, and 'reverse engineer' the product you would like to make for your own microfilm. In the world of digital resources, it is not as direct a process to see the product and know how it was made/how to make the same for your product, particularly in regard to extra functionality of the digital resources (an attribute entirely lacking from microfilm). In the world of microfilm, you can produce a not very good product that is still very usable (especially as the image quality issues are based on technology that has been understood since the mid-19th century and standardised in production equipment since the 1930's), whereas with digital resources no doubt, we could all provide some citations of products that fall far below even a minimal standard of usefulness.

But I would still like to suggest that the microfilm model is useful to us, making us to think through an entire plan of work, particularly keeping the end product and eventual user of the product in mind at all points of the planning and production workflow.

This seminar has already taken a stand that we will suggest guidance not guidelines for digital imaging, let me therefore suggest a series of ideas to keep in mind. Editorial preparation is the planning of pre-digitising production processes, keeping in mind:

- Productivity is the goal, for mass treatment verses one off treatment of items/images
- Staff skills may not match new demands of the production workflow (particularly coming from the model of brittle book projects)
- The source artefact has much information in it, conveyed via content as well as physical attributes of type faces, page layout and sequence etc. that should be considered, to be confirmed/carried forward or adapted in the new digital product.

My point of planning a production workflow is that you have reached the point where one must implement procedures, not merely debate philosophy and plans. As example, "as sophisticated browsing and search and retrieval techniques will be further developed to assist in the location of required resources" does not tell us how images are tagged, nor the system implemented or accessed. From the short survey of activity done prior to this conference, answers of how navigation and access would be provided within and between digital items brought answers of Common Gateway Interface (CGI) and Perl scripts and proprietary software reminding me of the cartoon of the mathematical formula, with the complex equations of right and left and a big circle in the middle: "a miracle happens here". We know in general terms what we want; I will suggest that the pre-digitisation stage, with planning, offers great opportunities for carrying forward information in an efficient manner.

In planning the workflow, one must know what that final product is (including its functionality) and how best to prepare the materials to reach that end; in the preparation stage, one has the opportunity for implementing processing procedures that could avoid extensive per-image labour later. In particular, many characteristics of a printed book's structure or a folder of processed manuscripts leaves are clearly evident in the physical attributes of the item; capturing those characteristics for the digital resource in some manner during initial scanning is the most efficient manner of carrying that information forward to the user product. Not wanting to really beat a dead horse, this prep work may be similar to some targeting practices of the preservation microfilming workflow....

In very broad terms, one can list three types of digital resources that might result from the conversion of legacy documents:

- A set or series of image files
- OCR'd or re-keyed text
- A combination of image and text files

There are obviously pros and cons in evaluating the usability of each model in terms of the service to user: with bit-mapped images, the information is unsearchable, simply 'digital photographs' with no machine intelligible information, so added text will be required to provide indexing and access to the images. Images in this case are dependent on some external system to manage their relationship to one and other; this is not entirely unknown to the library world, as most of us are familiar with a card catalogue or OPAC and its relationship to the catalogued collection (and we are also aware that the cost of the book is not the only library cost in making that item usable: cataloguing is essential). Even a simple structuring to allow a 'next page' or 'previous page' progression within these images requires a sequence of image file names that are labelled in related or predictable fashion. And in approaching image files, it is not to be underestimated the impact of the sheer quantity of

images that can accumulate, all needing to be organised, labelled and processed for use and maintenance.

OCR'd or keyed text loses all look and feel of the source materials, which may be fine for accessing a corpus of works for literary study, allowing elaborate text comparisons and searching, but the First Folios will still hold significance for the Shakespearean scholar, and author's annotation on a manuscript (or a printer's proof copy) can not be replaced even by very clever layering of electronic text - the combination of images and electronic text seems most desirable goal for most products.

With a plan for the form of the final product of the conversion process, and some goals of functionality, one can choose from a variety of methods of containing the information that will be necessary to make the digital resources function:

- Each image file may be uniquely named and labelled
- Plans may include relationships of image files expressed in folder and/or directory hierarchy
- Information may be inserted in the file headers of individual images
- Information may be gathered to travel beside images

In some elaboration of how information may be gathered to travel 'beside' a digital file, one could think of the control file approach, as used in the world of ASCII, the SGML approach or even the fixed field of the MARC record. Another manner of handling the information necessary for use of image files is to create a separate database (in simple flat file format or full blown relational database). With the database approach, one must remain aware of keeping a separate system accurately linked to the image collection, as well as the system overhead.

Up to this point I have worked carefully to avoid a term that immediately creeps into almost any conversation of digital technology: METADATA. It is clear that all I have discussed up to this point is the classic metadata, data about data. In particular, I have been focused on structural metadata, which helps in navigation in the digital object as well as relating one object to another - this is not the descriptive metadata for that initial discovery and retrieval of an object nor the concerns of rights management, provenance or file maintenance, which might be gathered in an area of administrative metadata.

Production planning needs to consider structural metadata elements, and how to integrate their identification/markings in the initial image capture process.

In an article about 'just in time' file conversion versus 'just in case' storage of converted images, John Price Wilkins made these comments which I think relate to the goal of editorial preparation of materials as well:

*"We cannot reliably predict which materials will be used or relevant for research. Effective digital libraries will be those that make their resources available in ways that do not influence research by using predictive methods that penalise the user who steps outside the mainstream. [This..] positions the digital library in ways that allow us to take advantage of future capabilities without losing access to historical collections."*

It is necessary to add structuring to digital files to allow their use, and the pre-digitisation editorial preparation stage is a good point to do it; while the complete and finite use of a digital resource can not be anticipated, certain basic minimum points can be seen as useful and most efficiently made part of this production planning.

Pulling many of these threads together, I will suggest that minimum pre-digitisation activity should include consideration of:

- A file naming system to allow correlation of the source materials to the digital images (if only for purpose of production workflow and quality control).
- Use of unique image file names, which may relate to a hierarchical structure of the source materials.
- Capture of self-referencing structural elements of the source materials, to be carried forward through imaging work for later access/use of the digital resource created.
- Getting these elements on the initial project planning checklist will assure efficiency in production workflow as well as prompting discussion of the final product and its use early in the project.

We can all agree that there is great interest in the creation of digital products at this point in time, for the new access that is possible to a corpus of materials in digital form as well as the simple fact that if one institution does it, others feel that they should do the same (and probably better). As my colleague has presented a case to consider the physical impact of digitisation, I would like to stress efficiency in production planning so that the best resource is made in most efficient manner, such that the digital resource is both useful and economical, negating a requirement for more reformatting work in the next technological wave.

#### BIBLIOGRAPHY

Preservation Microfilming: a guide for Librarians and Archivists. Ed. Nancy E Gwinn for the Association of Research Libraries. Chicago, IL: American Library Association, 1987.

RLG Preservation Microfilming Handbook. Ed Nancy E Elkington. Mountain View CA: The Research Libraries Group, 1992. And RLG Archives Microfilming Manual. Ed. Nancy E Elkington, Mountain View, CA: The Research Libraries Group 1994.

Price-Wilkin, John. "Just-in-time Conversion, Just-in-case Collections: Effectively Leveraging Rich Document Formats for the WWW." D-Lib Magazine, May 1997, available at <http://www.dlib.org/dlib/may97/michigan/05pricewilkin.html>.

**Protecting the Physical Form**  
**John E McIntyre**  
**Head of Preservation**  
**National Library of Scotland**

### **Introduction**

It is vital that we consider carefully what risk we might be exposing original material to during any proposed scanning procedure. Collection materials are at risk in any handling situation but when an item is being considered for digital scanning, it is likely to be subjected to more handling than usual as it passes through the various stages from selection through preparation to the digitisation process. The physical process of capturing digital images from library and archive materials can be compared to capturing images from the same material on microfilm, or even electrostatic copying. Our concerns are the same, we must be aware of the need to safely accommodate various forms on the image capture mechanisms of systems if we are to protect the information, its carrier and, at times, the artefact. To do this we can take lessons from the experience gained in creating microforms in modern times, and more lately of advances made in safer electrostatic copying mechanisms. Many of the threats are the same with the addition of some new ones or variations of old ones.

In short, we must seek to identify risk, apply measures to remove risk where possible and reduce any remaining risk to an acceptable level. If remaining risk is unacceptable then we should be prepared to take a decision not to digitise. It would be quite wrong to negate any preservation benefit that might be gained from making a digital surrogate available by damaging the original in the process.

### **Assessment of condition**

- Identify vulnerable material

There must be an initial judgement as to whether the chosen material is able to withstand the physical process. This first stage following selection should take account of the following.

- Any deteriorated condition or **fragility of the substrate** and whether or not it will be able to withstand the physical strains to which it may be subjected during any necessary handling stages and the chosen scanning process.

- Bound material needs to be assessed to judge any special needs for support of the form to enable the image capture area to be presented to the scanning device without presenting risk to the structure and to enable a satisfactory quality of image to be captured. Bindings may themselves be examples of important historical information and at times be more important than the text they enclose. There must be concern for tight, deteriorated and delicate **binding structures** and the need to protect them. In most cases, there is likely to be a need to preserve the integrity of the item that will include considerations related to the binding structure.

- Flaking, powdering or otherwise **unstable/deteriorating pigments** that risk further damage through handling or contact with a potentially damaging surface or exposure to heat gain. These need to be identified so that appropriate recommendations/decisions can be made.

- **Photographs** need special consideration and specification if they are to be protected. Many photographs are simply placed on flat bed scanners with no thought for damage to the emulsion layer through abrasion. Special needs and choice of scanning method need to be specified.

- **Light sensitive text or images** need to be identified and any restrictions special requirements relating to the scanning process specified.

All of the above are important elements in the first stage of preparation that is essentially an exercise in identifying risk associated with handling or with an available/chosen process and in taking appropriate action to remove or reduce risk to an acceptable level. If risk requirements cannot be achieved then at this point a decision not to digitise may be taken.

### **Preservation/conservation records**

The established preservation/conservation practice of creating a record of condition, including detail of existing damage and vulnerability should be adopted. Such a record created at the stage of preparation may be attached to existing record systems or form a separate reference file according to how precious the item might be and whether relevant records already exist. Such a record serves as a reference for pre-digitising condition and may be used to judge whether damage has been sustained but also to alert personnel involved at the various stages of processing that an assessment of condition and need has been made. This serves both as a reminder of the need for care and as a deterrent against carelessness.

### **Procedures for safe handling**

There should be procedures in place at institutional level and if necessary contractor level. These procedures should be designed for the purpose of protecting the physical form from the rigours of movement from storage through digitisation, including associated handling procedures as it travels from point to point. To achieve this:

- **Basic handing procedures** should be laid down.
- **Staff and contractors need to be informed** of institutional policy and requirements on handling collection items and have received training where necessary.
- **Any special needs specified.**
- Special requirements need to be **communicated** in a way that ensures they are understood.
- **Intervention to identify a suitable scanning method**  
The potential for damage that might be associated with unsuitable scanning methods and/or equipment must be recognised. In particular there should be concern for the following;
  - The risk of **surface abrasion** damage from using a flat bed scanner,
  - Accommodation needs of **large format** material. Apart from different handling techniques use of line, drum or digital camera scanning equipment needs to be judged carefully. Limitations on available equipment may mean that scanning has to be performed in segments with later decisions being taken on whether or not to "stitch" files together.
  - Support for **bound material** and the provision of special support equipment. It is generally agreed that bound volumes should not be scanned on a flat bed scanner, reflecting perhaps experience in the use of flat bed photocopying machines.
  - If available scanning methods cannot safely accommodate needs, then an **intermediate stage** may need to be considered e.g. a photograph produced which is then scanned. If photography is chosen and no satisfactory photographic form already exists, then the same assessment of the process to be used for obtaining the photograph needs to be undertaken to ensure that one unacceptable risk is not replaced by another.

### **Safe handling procedures**

The already mentioned safe handling procedures need to be established at institutional level, perhaps with three levels of implementation, formulation, assessment and monitoring.

- Policy and procedures must first be **formulated**.
- There should be an assessment of procedures to evaluate them followed by **implementation**.

- Continuous **monitoring** to some degree is required to ensure established policy and procedures are being recognised, used and continue to serve institutional needs.

### **The disbinding issue**

Returns from the survey the Preparation Group's questionnaire suggest that disbinding in order to scan a volume is common, in most cases so that a flat bed scanner can be used. The level of protection needed will of course vary according to an institution's responsibilities, the sector within which it operates and whether the volume has any value as an artefact. Disbinding was sometimes justified because there was a second copy available or because the condition of the paper was so poor that it was not holding together as a volume anyway. These can be valid reasons for disbinding but such a decision should not be taken lightly and to have available some guidelines that reflect current professional and ethical opinions would be helpful.

If disbinding is to be at times accepted, the question of how this should be done must be raised. Is the book to be rebound afterwards? and can this be achieved without destroying historical evidence? The question of guillotining the spine should also be raised and needs to be taken seriously as there can be no question of preserving the integrity of an item if this happens. There are several key questions that are central to the issue.

- **How ?**
  - Is there any intrinsic value?
  - Is the book or binding of national and/or historical importance?
  - Does it need to be carefully pulled apart in order to allow recovery?
- **Guillotining**
  - Can guillotining be allowed?
- **Reconstruction**
  - Is it to be simply rebound, or is it to be rebuilt carefully to preserve the integrity of the item ?
  - Or is it simply to be boxed?
- **Disposal**
  - Can it be discarded?
  - Does a preservation surrogate exist?
  - Are copies held elsewhere?

In considering disposal after digitisation we must consider carefully how long the digital form can exist and for how long it is likely to be accessible. Both questions are presently unlikely to receive answers that satisfy the needs of the preservation community.

### **Methods of scanning**

This is important when considering what the physical form, and the image or text, can withstand in terms of risk. Where necessary we must exercise options and if risk to the physical form is a concern, the chosen method of scanning should take this into account.

The survey questionnaire revealed that the use of flat bed scanners was common but around half also have a digital camera of sorts and stated that they were selective in their use according to what was being scanned. In contrast only three institutions said they used digital cameras exclusively. With a common reference to the Minolta PS3000 being made it may be that many people regard this equipment as a digital camera. This is wrong. While it might technically be described as such it does not have the versatility that a high resolution digital camera mounted on a copy stand or similar structure has.

In considering a specification for a preferred method of scanning we need to take account of the following:

- **Different material formats**
  - Bound volumes
  - Single leaf items and folios
  - Large formats, plans, maps etc.

- Photographic materials, negatives, prints, glass slides, early processes etc.
- **Special needs**
  - Fragile pigments etc.
- **An intermediary stage e.g. microfilm, photograph, PhotoCD**
  - Use of an intermediary stage can serve to protect vulnerable items but only if the intermediary stage itself is free from unacceptable risk.

### **Book cradles**

The bound form and the use of book cradles has already been mentioned. Support of some sort for a bound book must be available in order to reduce risk. Few off-the-shelf systems that might be described as book cradles for use in digital imaging or photography (with a digital camera the needs are very similar) are available. Some are beginning to emerge designed to hold the volume at a restricted opening with a page presented to the scanning device. They are particularly valuable for this reason. However, development is needed to improve cradle systems for some structures and to provide methods of holding a page or leaf under control without contacting the surface of the text or image being scanned. The latter is a particularly important requirement for delicate pigments.

Around half the people responding to the survey questionnaire said they were using a cradle of some sort. Others said they built up support using foam pieces. Some had not encountered the book form in projects reminding us that we are at an early stage with limited experience. For books we should consider the need for adequate support during the scanning process.

- **Supporting the form**
  - Horizontal platforms, are they split level?
  - Is support needed at an angle? i.e. restricted opening
  - Will foam wedges provide a solution?
- **Supporting delicate/tight bound structures**
  - Is a more sophisticated type of book cradle needed?
  - Do we need further development?

### **Lighting considerations**

The two concerns we have with lighting during digitisation are heat gain, which is likely to affect the substrate and some pigments, and exposure to radiating energy which may damage the information content, text or image. The dangers of exposing archive and library materials to light are well known and documented. In modern times this has developed into the formation of opinions and strategies that address the problem of *total* exposure to incident light rather than just the *intensity* of it.

Digital imaging needs light with the correct colour temperature if colour images are being captured, usually that is around 3,500 Kelvin. To achieve this it has been common to use halogen lighting but this risks damage to the items due to rapid dimensional changes as the surface of the item is subjected to the very high temperatures resulting from radiated heat. A much better alternative for normal camera scanning requirements is to use high frequency low voltage fluorescent lamps. These radiate very little heat and have a colour temperature suited to colour scanning.

Tests carried out at the National Library of Scotland when we were considering lighting systems and their suitability indicated high frequency, low voltage fluorescent lights to be safe. Our system has 2 cowls each with 3 single-ended fluorescent Biax L lamps and a high frequency ballast. Each lamp generates an output of 40W whilst consuming only 10W. Temperature, UV and lux meters were placed on the capture surface and the lighting system turned on at its highest level. Readings taken every 10 minutes over 4 hours indicated a temperature rise of only 1° Celsius and a change in relative humidity of only 1-2%. Both changes are within the error margins allowed for in measurement and for tolerated fluctuations in the air conditioning system controlling the area. What was revealed, but not discernible to the eye, was that the lights took a full 15 minutes to reach maximum intensity.



This means that during a days scanning the lights have to be left on continuously risking unnecessary exposure of collection material to light. This has been controlled by adding baffle plates to the lamp cowls to enable us to limit exposure as required for scanning. Flash exposure used with a linear array camera is an alternative way of controlling exposure to light and has been used successfully in digital programmes.

Whatever lighting system is chosen we must take account of the following.

- **Type of lighting**
  - We should choose the type of lighting with regard to the potential for damage from the light source.
- **UV radiation**
  - This comes from the part of the wavelength that contains the most energy and therefore is the most damaging.
  - We do not need it and should attempt to eliminate it.
- **Temperature**
  - Lighting systems radiate heat.
  - Radiated heat must be reduced to a minimum with concern being for the temperature at the scanning surface.
  - Parchment, vellum and other materials, including pigments, are susceptible to damage from heat.
- **Lux levels**
  - It is important to take account of lux levels, particularly with sensitive inks, pigments and photographic materials.
- **Duration**
  - We must control the amount of exposure to light during scanning and associated processes taking into account exposure to light sources to which the item may have previously been subjected.

### **In-house operations**

The question of whether to run a digitising project with scanning methods carried out in-house or at an establishment such as a bureau will depend on many things, such as the availability of capital expenditure, institutional experience and expertise, the size of the project, even space. Associated with any decision to out-house work must be the concern for a contractor to deliver the institute's requirements in terms of meeting set specifications and of safeguarding the material being processed, both during storage and processing.

- **In-house versus bureau?**
  - It is easier to monitor the provision of protection during in-house operations compared to out-house.
  - Does the bureau have sufficient environmental control to safeguard the type of materials being processed?
  - Does the bureau have adequate security?
  - Is the specialist equipment necessary to protect certain materials and structures available?
- **Part in-house, part bureau?**

Perhaps a mixture of both serves some institutions. Low capital outlay in-house coupled with the use of high capital investment out-house, e.g. microfilm scanning, less important material or that which can be disposed of afterwards.
- **Criteria for choosing a contractor**
  - Criteria for choosing a contractor needs to be established, e.g.
    - Relevant experience
    - Awareness and understanding of special needs
    - Security
    - Specialist equipment
    - Environment

There can be no doubt that commercial establishments have an important role to play in providing digitisation services but there is a need to establishment requirement in terms of qualification and provision.

### **Post scanning access to originals**

The question of providing access to originals when a digital forms exists is perhaps a matter outside the preparation stage but it is closely linked to the protection of the physical form and therefore I believe worth a mention here. The digital form is not yet considered to be a preservation medium but it does have a preservation value in that while a digital form exists and is used for access, protection is being gained by the original not being subjected to the stresses associated with handling processes and consultation. With access often being enhanced by the digital form, we must consider the need to place restrictions on access to the original in the interest of preservation. Since we preserve in order to provide continued access, any controls imposed should restrict and not deny access.

The survey questionnaire returns indicated no real reduction in demand for the original when a digital form exists. Only eight institutions said demand had decreased, twenty-seven said it had stayed the same but eleven said demand had actually increased. It is perhaps too early to predict the effect digitisation may have on demand. At present it is considered by many to be something of a novelty and can have the effect of drawing attention to the item in question, even generate curiosity as to how close a copy the digital form is to the original. There is little doubt though that a preservation policy should provide control to limit or prevent unnecessary access to original material, particularly with precious and/or fragile items. In any access policy we will need to make decisions based on:

- There being good reason for access to the original
- Restriction due to condition
- Restriction due to rate of use
- Change in the rate of use
- Restriction for security reasons.

### **Conclusion**

It is clear that while digitisation can provide many benefits including helping to preserve original material, if there is not an awareness of the potential for damage during digitisation and associated processes, then there can be a net loss as far as preservation is concerned. We must therefore identify all the areas of risk associated with digital scanning and where possible eliminate them, or reduce risk to an acceptable level. To do this we will need to establish guidance, and at the appropriate time, Guidelines.

**Guidelines for Image Capture**  
**Steven Chapman**  
**Preservation Librarian for Digital Initiatives**  
**Harvard University**

The reasons to scan historic collections range from enhancing catalogue records with thumbnail images to creating facsimile reprints. Given this broad array of goals it is reasonable to ask, "How can we find consensus?" In the conference *Digitising Photographic Collections*, held at the Rochester Institute of Technology in 1997, James Reilly stated that in all forms of imaging there is a spectrum of choice, and that we cannot rely on standards to direct institutional purposes. Rather, he insisted that our purposes must drive technical choices. The corollary to Mr. Reilly's observation is that in all types of institutions there is a spectrum of purposes. The goal of this paper is to promote discussions that will lead us to articulate common objectives related to digital image capture.

As a preservation librarian, I will admit to a bias in providing this summary of imaging practice in U.S. libraries and archives. It is not, as you may think, to suggest that preservation-quality products ("boutique images" as Michael Lesk has dryly observed) are the only ones worth making, or that investments in the highest quality you can afford will pay off in the long run. Instead, my training and my background lead me to advocate controlling the process as a common objective, regardless of the nature of the source materials, user needs, or image quality requirements. We are talking, after all, about best practice, not best products.

What strikes me about the language of our profession is the degree to which we talk about process rather than product ("digital imaging," "digitising," "microfilming," "photocopying," or even the catch-all "reformatting" are familiar terms). Perhaps we take for granted the permanence and quality of the products we create routinely in brittle book programs. Or is it possible that we focus on process for two reasons? First, no matter which technology is used we know how many things can go wrong from the source material to the copy; and, second, we know that in all cases we must give up something from the original (that to balance quality and cost we must sustain some loss of information). For these reasons, we organise reformatting activities into a series of controlled workflows. I think we should sustain, if not strengthen, this approach to best practice as we establish new guidelines for new technologies. Systems measurement, thorough inspection of pictorial reproduction (image quality) and completeness, the testing and certification of media stability, the training of technicians, and the systematic gathering and distribution of documentation arguably serve all scanning objectives equally well.

This working group, *Guidelines for Digital Imaging*, has decided to review five topics related to digital image capture, with the objective of identifying the points in the reformatting workflow that would most benefit from the development of guidelines. They are materials handling, systems quality, digital master quality, derivatives, and file naming. Each of the five sections includes introductory comments, a list of issues, a review of technology, and a summary of practice to gauge where perspectives are similar and where they vary. The sections conclude with lists of questions to promote further discussion.

In one of our conversations about this conference, my colleague Jane Williams used the wonderfully apt phrase "fit for purpose" as an umbrella term to put image capture guidelines into a usable context. This phrase nicely encapsulates James Reilly's adage that there are spectrums of choice in imaging technology and techniques, and spectrums of purpose.

Naturally, guides to practice should not be so rigid that they apply only to narrow bands of the spectrum. We have the best chance of being inclusive if we can identify the boundaries and work between them.

### 1 Handling

Physical handling is one of the most destructive things that can happen to a fragile object. One of the best ways to preserve it is to limit physical access to it. This is a very strong case for creating a digital library of such objects.

Peter Noerr, *The Digital Library Tool Kit*, Sun Microsystems, Inc., April 1998, p. 21. Available at <http://www.sun.com/products-n-solutions/edu/libraries/digitaltoolkit.html>

The potential preservation benefits of digitising valuable, unique, or fragile materials are so well understood that they are now stated almost as a matter of course among the rationales for scanning historic collections. What is pointed out too rarely, however, is that digital images do not make themselves. The logic that is offered to invest in digital image quality (namely, that reducing use of materials enhances their preservation) also applies to the digital conversion process. Physical handling threatens historic collections. When considering handling at the level of the collection rather than the single item, it is not unreasonable to argue that a group of materials may never be handled to such an extent again during its lifetime.

Handling is sometimes required twice in the reformatting workflow, once for processing and cataloguing, and again for imaging. Although we can presume that scanning technology will evolve, handling is the one area of digital capture where we should not expect industry to develop acceptable products without significant participation from our community. We are not a large enough market to entice manufacturers to develop scanners that "fit" non-standard materials, and it may take some time to convince engineers, for example, that opening bindings fully to 180° is often a bad idea, whether the bound material is imaged face up or face down.

Reviewing library imaging practices to date, it is therefore not surprising to find that compromises have been made to get the job done. In some cases, handling practices have been less than ideal; in others, image quality was lowered or costs were raised in order to protect source materials.

For example:

- disbinding demonstrates a willingness to sacrifice the source material in order to meet quality and/or production goals;
- creating photo intermediates for scanning (e.g., Kodak Photo CD projects) ( particularly in cases where photography has been heavily production oriented) represents a willingness to sacrifice digital quality in order to meet preservation and production goals;
- using 1-bit overhead scanners to digitise bound historic materials demonstrates a willingness to sacrifice digital quality and production goals to meet (or approach) preservation goals.

#### A. Issues

- disposition policies
- (pre- or post-scanning) conservation treatment implications
- handling policies (including exposure to light, heat, humidity, etc.)
- digital image quality requirements
- cost implications

- the role of the original
- the movement of more special collections materials into the reformatting workflow

### B. Technology

Risks to materials are generally lower in face-up reformatting workflows, so digital cameras and overhead scanners are of particular interest to the library, archives, and museum communities. Within the past year or two, imaging experts appear to be approaching consensus that direct digital photography offers exceptional capabilities for image quality (in some cases even exceeding the reproduction capabilities of film ) and that it may not pose undue risk to historic photographs and other works of art. (Colet, D'Amato, Ester, Frey, Mintzer) Lighting and its effects regularly emerge as issues that concern curators. One wishes for a definitive study about the effects of lighting (at which cumulative point does light damage a given object?) and the actual, rather than relative, harmful effects of the types of light used with digital cameras and flatbed scanners.

Pending the outcomes of more research and discussion in this area, we should follow with interest a number of improvements in digital cameras. New area array cameras such as the Leaf DCB support the short exposures of flash photography, although their resolution (and potential for scanning oversize materials) is much lower than the longer exposure linear array systems such as the Phase One. The Marc II system now being used in the Library of Congress Prints & Photographs Division is one of the newer cameras that combines the advantages of the area and linear array cameras by stitching together multiple exposures to create large files (up to 11K x 11K) at relatively high speeds.

All this is to say that face-up, "contact-free scanning" is becoming affordable. Some institutions are reaching the conclusion that the costs of producing reasonably high-quality digital photographs are comparable to the combined costs of producing high-quality 35mm photo intermediates and production Kodak Photo CD scans. For digital access projects, this raises the question about the preservation value of 35mm film. Should guidelines for digital image capture recommend that direct digital capture is preferred over the photo intermediate-to-scan approach?

For non-continuous tone materials, particularly the "mixed" format of the prototypical brittle book, face-up (overhead) scanning is something of a mixed blessing. The cradle design of the Minolta PS3000 scanner, for example, presumes that one always wants to open materials fully to 180°, and then to press down on the pages to hold the item flat. Until very recently, the ImageAccess BookEye scanner did not have any type of binding support integrated with its system, so to our community it represents something of a small victory to see that this vendor has recently upgraded their system to include an adjustable cradle. Like the Minolta, however, this 1-bit scanning system has been optimised for modern materials. Getting legible images from aged low-contrast paper with peaks and valleys across the surface requires a high number of rescans with 1-bit scanners, regardless of their optical resolution.

In two exceptional cases (Zeuschel and Xerox PARC), industry is working to develop a digital camera/integrated cradle optimised to rare books. Unfortunately, these scanners are among the most expensive; and the Xerox book scanner is still in development. The more general trend in rare book scanning appears to be the use of a custom cradle with a standard digital camera. These cradles are almost always designed by conservators. Notable prototype systems have been used in several library scanning projects. (Mayer, Mintzer, Riser, University of Oxford) Collaboration in this area will be instrumental in developing systems that balance the needs of handling and production.

### C. Summary of Practice

As might be expected, handling practices vary widely, according to the value and condition of the source materials, the preservation objective (to represent or to replace), cost, willingness to customise equipment, and, perhaps most importantly, the level of participation by conservators.

#### Bound materials

##### disbinding

- duplicate volumes selected for JSTOR; brittle volumes (books and journals) at Cornell University and the University of Michigan
- intact, open to 180°, face down (with pressure)
- Harvard law volumes for Studies in Scarlet (sewing threads cut in some cases)
- intact, open to 180°, face up in cradle (with varying pressure on pages)
- University of Virginia (approx. 70% of Early American Fiction); Harvard law volumes; Library of Congress law volumes; Internet Library of Early Journals
- intact, open to 120-150°, face up in cradle (camera/overhead scanner)
- University of Virginia (approx. 30% of Early American Fiction; see, Riser); Harvard law volumes (approx. 20%); Library of Congress law volumes; Vatican Library incunabula (Mintzer)
- intact, open to 90°, face-up in wedge-shaped cradle, glass prism lowered
- Zeuschel camera system; prototype book scanner at Xerox PARC

#### Photographs (excepting roll film and fiche)

- higher-end labs maintain numerous controls: dust and air filtration, climate control, electrostatic cleaning, requirement to wear gloves, heat diffusion and UV filtration for lighting (Metropolitan Museum of Art, MoMA, LC Prints & Photographs/JJT, Inc.)
- mylar sleeves sometimes used, as mandated by curatorial/conservation review during preparation (NARA, Library of Congress NDLP)

#### 35mm roll film (negative, slide film, microfilm)

- autofeed scanners widely used (e.g., Mekel film scanner, Kodak's former PIW); practice very consistent with this format, although it has not yet been established conclusively whether print master microfilm can/will be scratched by high-speed film scanners; it is also worth noting that when metadata is keyed from the film (rather than from digital images) the film is handled twice: first in a microfilm reader, then in a film scanner

#### unbound single sheets (manuscripts, prints)

- most often scanned directly on flatbed scanner, sometimes photographed first to 35mm film

### D. Discussion Questions

- Should digital image capture guidelines focus exclusively on issues associated with the digital images, or should they address both the digital images and the source materials that are handled to create them?
- Should handling guidelines be written to accommodate a range of disposition policies? Or should we presume that all source materials must be saved?
- For any class of source material, what is the optimal practice for handling?
- When is it acceptable to follow less than best practice?
- Under what circumstances are auto feeders for paper or film acceptable?
- Same for disbinding?
- Should guidelines distinguish between contact (with rollers, vacuum feeders, glass) and contact-free digitising?

- Should this section of image capture guidelines/recommendations be written in "if, then" fashion? (e.g., if source materials to be retained are altered in scanning, then they must be repaired or rehoused?)
- Should image capture guidelines provide instructions regarding the proper ways to stabilise and/or support materials during digital image capture?
- What information about the source material must be documented as administrative metadata? (This issue is particularly important in reformatting projects where film intermediates are scanned.)
- Will we be able to implement guidelines without also offering hands-on education and training to many imaging service bureaus?

## 2 System Performance<sup>1</sup>

There are several incentives to define technical guidelines for system components (scanners, software, monitors, printers): to be able to provide specific contractual language regarding the baseline performance for equipment; to make informed comparisons of products and services; and to document whether an imaging system is operating consistently at optimal levels.

While it is true that clearly written technical literature helps us become educated about the key components of scanners and their relationship to image quality, imaging experts remind us that judgements about quality are ultimately subjective. Producing a good image results from knowing what a system can do and what the observer wants to see. Selecting and evaluating equipment, however, can be based upon objective measures. Guidelines and tools in this area of image capture give us control over the system by allowing us to measure the quality of the signal, rather than the quality of the image from a given scanner.

Engineers and image scientists report that the four most important system characteristics to evaluate and monitor are noise, detail, tone, and colour reproduction, (D'Amato, Gann, Reilly). The first two are critical for 1-bit scanning; all pertain to colour imaging. One of the more appealing aspects of digital technology is that it makes it possible to conduct objective assessments of system performance in these key areas. With the proper tools, precise measurements can be obtained for each of these values. Moving from theory to practice, however, will require the development of easy-to-use targets and associated software, likely to be available in months rather than years, but not here today. Photographic & Imaging Manufacturers Association (PIMA, Reilly, Williams).

Where objective measures are in use to document system performance (LC, Smithsonian, NARA), there has been resident photographic expertise, training from imaging experts, or both. Subjective techniques have been used in other in-house projects, where practice generally falls into two categories: evaluating images on screen or in print. Imaging service bureaus, as might be expected, have various levels of expertise and experience. The better ones are advising the libraries and museums how to set up their systems.

When subjective methods are used, experts unanimously agree that system calibration is essential. It is something that we must learn to do if we are to represent our digital products as being of high quality.

---

<sup>1</sup> in this context, "system" refers to the combined performance of scanning hardware, software, file format, and compression algorithm used to save images to disk.

**A. Issues**

- obtaining a good signal
- defining top capability
- ensuring consistency of performance
- certification of master (for legal deposit, preservation, etc.)
- targets and their use
- subjective versus objective methods of quality control

**B. Technology**

- noise (uniformity of response)
- detail reproduction (spatial sampling frequency and MTF resolution)
- tone reproduction (linear grayscale and spectral response function)
- colour reproduction (linear colour and spectral response functions)

From the librarians' perspective, the state of technology may be summarised as the need to integrate technical concepts, tools, and training into the digital image capture workflow if we are to move objective quality assessment from theory to practice. As illustrated by the examples in the Appendix, technical specifications must not be taken at face value. Forecasting image quality with a single specification (such as dpi or bit depth) is particularly risky.

We know that it is important to test and monitor equipment, but why not just look at scanned targets on screen or in print? One of the limitations to this subjective approach is that even with calibrated systems it is not easy to determine where quality loss occurs from one component to another. We also cannot infer from a high-contrast target how a scanner will perform with low contrast material, or what its capabilities might be to capture shadow detail.

Targets designed to measure Modulation Transfer Function (MTF) make it possible to assess detail reproduction across a range of tones. Experts explain that the MTF test reveals whether a scanner's sampling rate on output is actually less than its specification (i.e., the optical resolution or input dpi). Don Williams, an image scientist at Eastman Kodak Company, explains that although it is reasonable to assume a 400 dpi scanner will sample a document every 1/400", it is not safe to assume that the scanner will actually resolve details this small. (Williams) In fact, it is possible that one scanner with a better signal-to-noise ratio will resolve the same detail at a lower sampling rate. In other words, the potential payoff in implementing the use of MTF targets and analysis is to produce higher quality in smaller file sizes. When comparing systems and services, one other advantage to measuring MTF is that a scanner's exposure setting has no effect on the accuracy of the reading. (Gann)

Traditional photographic targets (Macbeth ColorChecker, the Kodak Q60 targets, Kodak Q13 and Q14 grey scales and colour patches,) are being used to measure noise, tone reproduction, and colour reproduction, although in the long run these may prove to be less than ideal for objective measurement of digital systems. Perhaps with some beta testing and funding from our community, organisations such as PIMA or the Image Permanence Institute will be able to manufacture a single integrated target for system evaluation. Targets such as the RIT Alphanumeric Test Object to measure 1-bit systems may not be sophisticated enough to determine how well a scanner will represent details in low-contrast or illustrated documents. These 1-bit targets can, however, be used to compare systems, particularly the thresholding capabilities of different scanners which have the same optical resolution.



### C. Summary of practice

As with handling, practice varies widely. Several institutions have set up in-house imaging facilities, ranging from document scanning labs to fully configured digital photography studios. Some institutions regularly use targets, some do not. Scans of technical targets could be saved as administrative metadata to accompany digital objects, but this practice is rarely followed. Grayscale and colour bars are sometimes included with digitised photographs (see, for example, the imaging recommendations in the MOA II White Paper) but there is a difference between using targets to document the image and using them to document the scanner.

Several of the leading imaging service bureaus in the United States have established, refined, and documented the techniques they use to monitor scanning systems. (Ester, Preservation Resources) Collaboration among experienced practitioners, imaging scientists, professional photographers, and "digital librarians" will likely be the successful formula to document, then implement best practices in the appropriate use of targets for all types of scanners.

Poor environments can and will degrade system performance and image quality. Scanning of Caribbean Newspapers was actually suspended at the University of Florida during building renovations because vibration and dust made system performance unreliable. When environmental control is feasible, specifications range from the wall sockets (using line conditioners to control voltage) to the ceilings (requirements for paint colour and indirect lighting to create calibrated viewing environments), and from the HVAC system to the furniture. Controlled-environment imaging labs have been created at the Denver Public Library, the National Archives and Records Administration, the Museum of Modern Art, the Library of Congress, and selected service bureaus in the U.S. Specifications for environmental control should also be included in any guidelines for systems quality.

### D. Discussion Questions

- What are the most important measures of system quality?
- Are objective measures preferred to subjective measures? Should both be used?
- What constitutes optimum practice in this area, and when is less than best practice acceptable?
- When, for example, is it acceptable to use/document the manufacturer's specifications (of dpi, bit depth, density, etc.) as a statement of acceptable system quality?
- If targets and tools are recommended, what specific measurements should be taken, and where should the numeric data be recorded?
- The Museum of Modern Art, for example, specifies that density measurement numbers be recorded with a calibrated grayscale

### 3. Quality of Digital Masters

Creating digital images worth keeping is the heart of the matter in proposing best practices for digital image capture. Preservation librarians sometimes make the case that increasing quality will extend the usable life of a digital collection. Associating longevity with image quality is arguably a matter of speculation, but there are limits to this logic. At some point, one reaches the maximum lifespan. Another, perhaps more cynical, point of view is that only images that are used have a chance of long-term survival.

At the Rochester conference on Digitising Photographic Collections, James Reilly stated that it is "utterly crazy to think that one image will serve all purposes." Other imaging experts assume that there will be only one chance to scan materials, so one should attempt to capture all of the information content in the original to create a "rich" master that will have great flexibility and lasting value. The two points of view are not diametrically opposed, but their

implementation logically suggests different concepts of "best practice." With proper allowances for cost differences in each approach, perhaps two sets of guidelines could be developed.

In this arena of digital image capture, guidelines should address every component of the digital image with an eye toward ensuring openness and flexibility (and therefore longevity). Open formats are those which are widely supported by software that offer both capabilities to read from and write to a specified format. It is up to us to decide whether or not a format must be supported by a standard in order to be considered open.

The guidelines should also provide clear statements of the value judgements that are at work in defining pictorial quality in a master file. Digital masters, we must be reminded, are rarely delivered to the user. Assessing the quality of master images quickly teaches the lesson that "digital image" is an oxymoron. A viewable image is a product of the encoded file and the characteristics of a given viewing device. To define pictorial quality objectively, we need to respond to the technical challenge of eliminating the monitor or printer from quality assessment. With targets and associated software, this can be done.

#### **A. Issues**

- producing a good scan
- reproduction quality
- defining sustainable loss
- functional quality
- capability to produce many types of derivatives
- file size
- reversibility
- keeping an image you can always go back to; no irreversible transformations, such as sharpening, in master
- persistence/durability
- stability and flexibility

#### **B. Technology**

There are a number of technical issues in this category. Quality and longevity goals are paramount, but it is also important to account for the tradeoffs between quality and cost. Determining how best to stitch the following pieces together to create good master images can be somewhat tricky. Here are some of the separate issues that have a bearing on the quality of the master image:

- file format, including version, byte order, and headers
- compression
- colour space
- bit depth
- settings for black point, white point, gamma
- use of targets
- spatial resolution
- number and degree of image enhancements (e.g., sharpening)
- administrative metadata to reside within the image
- all the "negatives" to avoid (moiré, clipping, cropping, skew, etc.)
- lighting (digital cameras)

To view these technical issues from the management perspective is to see a long list of decisions that must be made. But if we do not make them, someone else will on our behalf. To control this part of the digital image capture process, we must become technically literate,

and even more importantly, we must have the vocabulary to describe what we want. Just as rules for metadata syntax ensure reliable searching, rules for image quality syntax will help to ensure consistencies in image quality. As an international group, can we reach consensus on this issue?

One industry trend is giving us more latitude to configure scanners to meet guidelines for file format, compression, and image enhancement. With more scanners being built to either the ISIS or TWAIN standards, scanning software and hardware can be evaluated separately. Using the objective metrics described in the previous section, we might configure a scanning system by selecting the hardware based upon engineering, image quality, and price, then select (or write) the scanning software based upon file format specifications, desired enhancements, and ease of use.

### C. Summary of Practice

In surveying the guidelines that have been published, it is important to assess first whether scanning guidelines were related to project goals of any kind, or were selected because "Institution X" had previously used that standard. Second, it is important not to compare apples to oranges. Differences in source material must be taken into account, of course, but it is also essential to consider the philosophy that informs the specifications.

#### C.1. Quality

Broadly speaking, there are two schools of thought regarding image quality:

One approach is to attempt to capture a "faithful reproduction" of the original material. Like microfilm guidelines for image quality, this specification allows certain losses to be sustained from the original in order to meet comparatively high production goals. Decisions about image capture in these projects are principally guided by the content of the source materials (which presents its own challenges in the hybrid projects to produce both preservation microfilm and digital images).

**example:** brittle books scanning projects at Cornell University; specification: 600 dpi 1-bit Group IV TIFF images, with use of Xerox resampling/descreening algorithms for halftones.

The other approach is to create masters capable of producing images that are "optimised for presentation" and performance. Decisions about image capture in these projects are principally guided by user preferences, by current capabilities of technology, or both.

**example 1,** image capture guideline driven by scanning technology: Caribbean Newspapers Project at the University of Florida to scan 35mm microfilm, specification: 400 dpi 1-bit Group IV TIFF images, because it was the highest resolution offered by the microfilm scanner that met their performance requirements (budget and quality); to produce legible master images, the microfilm images could be enlarged to one-half the size of the original newspaper.

**example 2,** image capture recommendation driven by printing technology: recommendation by Mitretek Systems, Inc. and Allen Press to the Smithsonian Institution National Museum of Natural History (NMNH) to scan illustrations at twice the anticipated halftone line frequency (e.g., scan at 600 dpi for printing at the highest line frequency of 300 lpi); specification: 600 dpi 24 bits per pixel: quality deemed comparable to 4 x 5 colour film (D'Amato).

Photographs belong in a slightly more complex category, which combines elements from the two perspectives of image quality summarized above. At Corbis, they have established four

categories of image reproduction, and four sets of corresponding guidelines for digital image capture (Süsstrunk):

- the photographic image as rendered (match appearance of original);
- the photographer's intent is rendered (enhancements are applied to scans of photos that were not exposed or processed correctly);
- the original appearance of the photograph is rendered (i.e., digital photo restoration);
- or the original scene is rendered (film biases, rather than flaws, are subtracted from photo intermediates. Sabine Süstrunk of Corbis observes that "this gets one into the business of creating a colour appearance model"). Michael Ester at Luna Imaging, refers to this approach as "matching to the scene" in order to avoid the mistake of enshrining the flaws of film.

### C.2. File Format

Whether the scanning objectives are to replace or to represent source materials, the choice of file format, compression, colour representation, and file-naming convention should, in the words of George Farr at the National Endowment for the Humanities, "close no doors."

The most popular and open file format for master images is TIFF, Intel byte order, version 5.0 or above. For 1-bit images, Group IV compression is widely used. Grayscale and colour TIFF images are often stored in uncompressed form. Other formats are also being accepted by digital repositories. Columbia University's guidelines include Photo CD as a master image format, and the Library of Congress guideline for illustrations is PCX (Xerox 5200 scanner to produce diffuse dithered images of printed halftones). In their evaluation of approaches to create high-quality digital reproductions of complex illustrations, the Smithsonian Institution NMNH preferred JPEG (Baseline Sequential compression) as the format for digital masters. At approximately 10:1 for colour images and 5:1 for grayscale, images met the NMNH goals for both (faithful reproduction) image quality and file size. (D'Amato) The National Library of the Czech Republic, also setting the standard of faithful reproduction, adopted JPEG with 3:1 compression as the master file format for manuscript scanning.

The conservative approach to creating digital masters favours using no compression. At the other end of the spectrum, one might choose to save images in a proprietary form of wavelet compression. As digital collections increase in size, this decision has important economic ramifications. Even with a repository of 1-bit images, JSTOR was able to save considerable sums of money annually by moving their digital masters from Group IV to Cartesian Perceptual Compression (CPC), which is more efficient. The logic behind the decision to avoid all types of compression is to minimise loss when things go wrong. When a bit flips in an uncompressed image, it produces one dead pixel, but with certain types of compression much larger portions of an image can be lost. In practice, however, this rarely occurs, particularly if digital masters are stored in a system that is programmed for automatic error detection and analysis. Guidelines for compression and file formats should appropriately weigh both (preservation) risks and (fiscal) rewards.

### C.3 Colour Space

Decisions about file format and colour space sometimes go hand-in-hand. Kodak's PCD and Flashpix formats, for example, restrict colour interchange to YCC and NIF RGB. In the past year, industry consortia have adopted sRGB as a standard for colour encoding, in part to improve colour matching between scanners and peripherals, but mostly in pursuit of the goal to represent colour consistently on the Internet. The relative advantages and disadvantages of file format/colour encoding combinations should be fully explored in digital image capture guidelines, particularly for the digital master. Is one across-the-board recommendation

viable for all materials? This decision has ramifications in the choice of scanning software, the scripts that will be written to create derivatives, and migration schedules.

#### **C.4 Bit Depth**

Guidelines for bit depth have traditionally been specified in one of three categories: 1-bit, 8 bit, and 24 bit, even though scanners sample as many as 14 bits per pixel. Dr. Franziska Frey and others recommend capturing a minimum of 12 bits per pixel to capture the dynamic range of photographic prints. (Reilly/Frey) Newer scanning software and image processing programs (e.g., Photoshop 5) accommodate up to 16 bits per pixel. Again, scanning guidelines should consider the tradeoffs between file size and information loss in this category.

#### **C.5 Tone Distribution**

In addition to making a decision about bit depth for tone reproduction, project managers have instituted practices to control tone distribution. The Library of Congress RFP for pictorial materials (97-9), and Steven Puglia and Barry Roginski's guidelines (NARA) provide detailed overviews of this technique, which depends on the use of targets. I am slow to appreciate fully whether this practice alters the appearance of some images — especially the high-key and low-key originals (to the point that they “look wrong” when compared to the original. Decisions about black point, white point and gamma presumably have a similar effect. (Corbis, for example, uses a gamma setting of 1.6 to produce a neutral, relatively flat image.) Guidelines about tone distribution and representation in digital masters can be stated in the negative (avoid clipping ) but, in practice, how does one determine if the master is right? Should quality control be restricted to evaluating histograms, or should the image be evaluated on a calibrated monitor or printer? Or both?

At the risk of sounding naive to professional photographers in the audience, I will ask the question, “Is it possible that in applying these techniques to regularise tone distribution that a ‘good scan’ would not necessarily look the same as a ‘pleasing image’?” Image capture guidelines must be explicit on this point so that quality control procedures will be consistent with the desired outcome for the digital master.

#### **C.6 Targets**

The decision about whether or not to scan gray scale and colour wedge targets with the master image raises other questions of documentation. Should dimension scales also be specified? If grayscales and colour wedges are photographed with the source material, should their digital values be recorded as administrative metadata? Is there software that can do this automatically? Preservation metadata is a topic that will be addressed separately during this conference, but image capture guidelines must specify not only what metadata are essential to preserve image quality from generation to generation, but also where this metadata should be located. There are three options: within the file header, outside the header, or both. Guidance from imaging experts should be solicited to advise us of the pros and cons of various formats to record this information in file headers (e.g., TIFF EP will purportedly accommodate additional data about the signal of digital cameras).

#### **C.7 Resolution/Sampling**

Given what imaging experts have been saying about MTF, it is with some reservation that the following numbers of sampling rate (i.e., input resolution) are offered to define a spectrum of image quality. Nevertheless, resolution is one of the most important decisions to make about image capture as it has major ramifications in file size.

Depending upon the format of the source material, a resolution specification (e.g., 600 dpi) can be tantamount to saying, "there is only one scanner you can use." In Yale University's Project Open Book, for example, no microfilm scanner met their specification for 600 dpi 1-bit scanning, so they created one by investing in custom software.

Specifications for resolution vary. In some cases, a minimum file size is specified. In others, a single dpi resolution is used in order to ensure that enough pixels are in the master image to create a specified output, such as a full-screen image. (To a certain extent, this was the logic behind the 18MB Kodak Photo CD file format, which ensured that the 16 Base image could produce a quality 8 x 10 print.)

Examples of resolution specifications include:

- 5,000 pixels for photographs (LC)
  - 45-60MB (in order to be able to print 8 x 10 duotone or tritone reproductions) (MoMA)
  - 300-500 dpi 24-bit for early American fiction (text) (Univ. of Virginia)
- One needs to develop a translation matrix to compare these numbers. (Which of the above, for example, represents the highest dpi?) When evaluating these specifications, one should segregate them according to the quality objective they were intended to support. To achieve fidelity (and if so, at 1:1 with respect to the original)? Or to achieve a target level of quality for a target output device? Either could be interpreted as meeting a preservation objective. A comparison of guidelines from several projects shows that practices vary widely:
- machine-printed text: ranges from 300 dpi 1-bit (NARA) to 600 dpi 24-bit (CMU "Antique Books") as specification to scan black and white originals
  - manuscripts and illustrations: ranges from 100 dpi 4-bit (Archivo de Indias) to 600 dpi 24-bit (University of Kentucky); specialised material, such as papyrus, has also been scanned at up to 600 dpi 24-bit, although that may have been to support 10:1 on-screen magnification rather than the setting needed to reproduce the details of the original fragments
  - photographs: widest range of resolution, where dpi resolution can vary significantly in a single project. The NDLP Pictorial Materials specification (RFP 97-9) is a good case in point. The 5,000 pixel specification for resolution translates to 1,667 dpi for the smallest Civil War photographs and to 385 dpi for the larger ones

### C.8 Intent/Documentation

A final point to consider about the digital master, particularly for visual collections, is the issue of photographer's intent. In every reformatting project, we will apply our biases (sometimes under the guise of "best practice") in the digital capture process. If we want our future colleagues who will manage these digital collections to preserve our vision of the "right image," we must find a way to document what effectively is the "copy photographer's intent." All of the specifications reviewed above, when used in combination, are directed to serve a specific purpose (either fidelity or a preferred representation. I would suggest that this is where our guidelines must stake one or more claims about the importance of documentation. Because of the great flexibility inherent to digital images, as well as the premise that digital masters will be carried forward to subsequent generations of display and print technology, it is important to document what we wanted the digital master to do when we first copied the original analogue print or film. We should indicate, for example, which medium conveys the message ( print, screen, or film?)

Examples of copy photographer's intent include:

- to be able to output the digital master to the same analogue medium as the original in order to simulate as many characteristics as possible of the source material (D'Amato)
- to create an image that partially fills the screen, fills the screen, can be magnified, or all three (LC Ameritech guidelines for photography); consider, however, the usable lifespan of such an image ( what happens when screen resolutions double, then double again?)
- to create a master that can easily be processed to create full text

Many more examples can be given, but this short list suffices to reiterate James Reilly's point that one image (digital master) will not always serve all purposes. We might want to consider the question, "Which purposes earn the designation of 'preservation quality'?"

### C.9 Quality Control

Since digital master creation has been informed both by the "fidelity" and the "presentation" objectives, it follows that quality control procedures also fall into two categories:

- in fidelity workflows, quality control tends to be subjective, where digital masters, output to the screen or to print, are compared to originals (sometimes with magnification); "matching" is the operative word in quality assessment; as noted in the previous section, equipment calibration is essential in this approach
- in presentation workflows, image attributes are measured against target values for tone, file header data, accuracy of file name, placement of targets, completeness, and degree of skew; "within range" is the operative concept

### D. Discussion Questions

- Are good scans and good images necessarily the same thing?
- What does "preservation quality" mean? Does the concept of "sustainable loss" apply to digital imaging as it does to traditional photographic methods? More specifically, are comparisons to photographs, photocopies, or microfilm apt?
- What metric(s) do we use to judge the quality of digital masters?
- If subjective (e.g., going print-to-print), when is it acceptable to use uncalibrated devices?
- If objective, which targets and software now exist to make practice viable?
- Should we adopt the Quality Index as a method to define and describe image quality?
- If so, will it mimic the practice for microfilm, or will it refer instead to technical parameters?
- When should targets be used?
- For any class of material or digital image, what are the optimal specifications for pictorial reproduction? Attributes include colour, tone, and detail reproduction; dimensions, cropping, and skew.
- Other than pictorial reproduction, what other attributes of the digital master should be addressed in these guidelines? Attributes include file format, compression, colour encoding method, tone distribution, and file header information.
- Will our guidelines be fixed and optimised for generality, or flexible to accommodate specialised technology?
- LC precedent of accepting PCX files, for illustrations, in addition to TIFF Group IV for black and white text
- Forthcoming grayscale upgrade to Minolta PS3000 scanner raises same issue ( will 10:1 8-bit JPEGs be designated as acceptable preservation masters?)
- In hybrid approaches where film is scanned in the place of the original source (e.g., Photo CD projects), is it acceptable to issue scanning guidelines without addressing/prescribing the preferred techniques for photography?

Ester report helpful, but key questions remain about control of lighting and choice of film. The George Eastman House, for example, elected to use 35mm MP Film 5032 fine grain

release positive (motion picture) film masked with blue acetate film to produce an optimum image for scanning.

#### 4. Quality of Derivatives

All of the care and attention given to image quality prior to this stage can be vitiated by poor decisions when creating derivatives from the master images. Were it not for limitations of network bandwidth and the limited choice of file formats supported natively by web browsers, guidelines for this stage of production would be focused exclusively on the goal of producing good (i.e., "pleasing") images. For image presentation, knowledge of audience is critical, as is the control over interface design. In current networked applications, image quality and file transfer speed are in direct competition unless proprietary compression schemes are used. (Some wavelet formats, for example, achieve high compression ratios and can decompress quickly at the client.) For this reason, perhaps we should reflect on whether a universal standard for "open, non-proprietary formats" best meets preservation and access objectives. This principle certainly makes sense for master images stored in reserve in a repository, but also for images that are to be used?

This consortial, international conference provides an ideal opportunity to test whether it will be practical, or possible, to develop general guidelines in this area. They may, in the end, be defined by the policies of each holding repository, which account for the needs of their community, and which depend upon the depth of technical and administrative infrastructure.

Setting aside the question of service models and the user interface, we can address many of the same issues of pictorial and functional intent that relate to the digital master. We might want to consider if the rules/recommendations should distinguish between print and on-screen derivatives. How much skew is acceptable? Should images, particularly those scanned from microfilm, be cropped? Is zooming necessary? Should dark, low contrast originals be delivered as "authentic" or as "easy-to-read" copies? Should targets be retained with pictorial images?

##### A. Issues

- producing good images from good scans
- producing good images from bad scans
- batch production
- consistency across many devices
- speed and legibility tradeoffs
- users' needs and preferences

##### B. Technology

The operative rule is that whatever was not done during scanning to create a pleasing image must be done at this stage. In a number of photograph scanning projects, this task requires the judgement of a human observer with a trained eye. (Bancroft Library) Digital photography workflows raise interesting questions about derivatives. If "raw scans" are sent to another workstation for tone and colour correction, then saved as high-resolution files, is the second image the digital master or the first derivative? Should this be documented in any way?

Image processing software is widely available, and each software upgrade seems to make it easier to execute scripts for batch production. As is the case with scanning, however, there are significant differences between creating digital images and creating good digital images. Where the scanning objective is to create images worth keeping, it seems that the goal here is to execute a program that will automatically generate images worth distributing.



Questions about parent-child image relationships (if such a concept is relevant in an image database) must be answered at this stage of the image capture workflow. Scanners will soon offer the capability of simultaneous output of grayscale and bitonal images. The image processing boards that make this possible to create multiple images during scanning could also be installed at the server to generate multiple images on the fly. High-speed, affordable grayscale scanning will create, for example, the technical challenge of making good 1-bit images for printing.

With 1-bit digital masters, the opposite challenge applies: the need to create grayscale derivatives for on-screen display. Programmers at the University of Michigan met this challenge several years ago by writing the (publicly available) TIF2GIF utility that is an important part of the JSTOR and University of Michigan Digital Library infrastructures. In these applications, grayscale GIF images are created on the fly, "just in time" from 1-bit digital masters. (Price-Wilkin)

Collaboration between library practitioners and image scientists could be extremely fruitful in working to develop production tools optimised to meet digital image capture specifications. Anyone can purchase Debabelizer, Photoshop, or Image Alchemy, but how many librarians and computer programmers have the skills to apply filters in the correct order in a batch script, or to make use of Photoshop's "predefined mathematical operation known as convolution" to make custom filters?

#### **B.1 Need for Production Tools**

In this area, I believe we should not settle for guidelines, but work together to develop the tools that can create derivatives with no distinguishable pictorial loss, or derivatives optimised to the presentation of tone and colour on a target device (such as a 1.8 gamma 800 x 600 monitor).

#### **C. Summary of Practice**

Practices vary according to the quality of digital masters and the target audience(s). At one extreme (of production and quality), 90% of the Corbis images are edited manually for tone and colour reproduction. (Süsstrunk) At the other end of the spectrum, digital masters are downsampled without enhancements and converted to JPEG or GIF images. Where one expert favours JPEG for illustrations of natural scenes (D'Amato), another favours GIF for scanned photographs. (Puglia) GIF compression is lossless, but its palette is limited to 8-bits.

The Columbia University guidelines recommend JPEG compression at a "Quality Level of 50." The Library of Congress recommends 15:1 compression for 24-bit images and 10:1 compression for 8-bit images. There are practical limitations to these recommendations, as no software exists to specify a set ratio of JPEG compression. These numbers, therefore, should be viewed as rules of thumb rather than prescriptive guidelines.

Target sizes for thumbnail images range from 15-20KB (NARA, Univ. Virginia) to 55KB (CA Heritage); the Library of Congress sets 150 pixels as the maximum image dimension. For full-screen "reference" images, sizes increase to 100-200KB (Univ. of Virginia, Bancroft Library), and to 640 pixels at the Library of Congress. This spectrum could be subdivided further with numbers from other projects, but the question of greatest relevance in interpreting practices and specifications at these and other institutions is this: Are users satisfied with the images?

How can we answer this important question? The project teams in the Museum Educational Site Licensing Project, which made 9,000 images available for study and teaching in seven universities, conducted user evaluations and found that . . . opinions varied. The following

excerpts from their final report send a clear message. The problem with practices to date is that derivative images have been made for computers rather than for people:

*For now, most designers of delivery systems select standard image sizes for their derivatives based primarily on dimensions of display devices . . . trying to strive for some balance between dimensions, quality and file size/compression. . . . [T]here is little we can say conclusively about the relationship between image production specifications and user satisfaction. . . . so many variables are involved that it is impossible to draw conclusions or make clear recommendations.*

It is unclear whether articulating any absolute guidelines for producing digital images is possible. What is possible is the articulation of a sound project planning framework and guidelines developed for particular types of originals and surrogates, digitised for specific uses and users. (Stephenson, p. 60)

Perhaps the guidelines for this area of digital image capture will be addressed indirectly by the Working Group on Selection. In the process of determining what content should be digitised, it may be possible to specify what the content is supposed to do in electronic form. If derivatives are to be optimised to uses and users, we need some way of finding out what our audiences want.

#### **D. Discussion Questions**

- Can guidelines be written to ensure consistencies in display across a wide range of computer monitors and/or printers?
- Are there ways to control information loss when conducting image transformations?
- Is it better to apply enhancements at this stage or during scanning?
- What are the effective filters to use in producing "pleasing images?"
- Is it possible to document image transformations, particularly when they result in information loss? Is this documentation required?

### **5 File Organisation**

Much has been said and written about digital object identifiers, but relatively little appears on the subject of digital file identifiers. Every file must have a name, of course, and rules about these names can be significant in a project workflow. Image capture guidelines should address the implications for cost and for quality (as it relates to functionality in the database) related to file naming practices. (Note: the administrative metadata issues which also relate to file names and to file headers are to be addressed in another session at this conference.)

#### **A. Issue**

- navigation within a digital object

#### **B. Technology**

Some scanning drivers offer the capability of programming a file naming scheme, with wildcards, for a designated batch of materials. As noted below, what they do not do is make exceptions for anomalies. The best technology for this task might be no technology at all. It may be worthwhile to explore alternative database models that do not depend upon meaningful file names for navigation.

#### **C. Summary of Practice**

With a file naming specification, the first thing a project manager would do is determine whether the names could be generated automatically by the scanning software. If not, then the names have to be changed in a post-scanning operation. If there is a logical order to an

object, then the filenames often have to be named so they can be sorted into sequential order. In some cases, guidelines also stipulate that page numbers also be embedded in the file name. Due to the irregularity of pagination for historic materials, it is impossible to program a file naming utility (such as TiffView Tools) to incorporate this information. Some tools are available to display page images next to a column of file names to make it easier to key the data, but the challenge remains to reduce the manual activity of this operation, which can be considerable.

In some cases, file names also contain feature codes to designate image type, or even image quality. Library of Congress NDLP guidelines include one-letter codes at the end of the file name for selected images: u = archival, r = ref, and t = thumbnail; the code "v" is used in some projects to denote images of very high quality. Other projects embed ownership information in file names, although when this administrative metadata resides in each image, it is more often located in the file header.

Structural metadata can be created early or late in the imaging workflow. Data should be evaluated from a number of comparable projects to determine when this activity is most efficient and cost-effective. Until tools can be developed to automate fully some of these procedures, image capture guidelines should mediate between functional considerations and cost effectiveness. In this case, file naming represents both digital image quality and database design quality. It is important to confirm that we are establishing models that will scale and will persist.

#### **D. Discussion Questions**

- Other than uniqueness, what rules apply to directory and image file names?
- Should page numbers or feature codes or both be embedded in a file name?
  - If not, what are the costs and benefits of the alternative(s)?

#### **6. Conclusion**

The development of digital imaging guidelines that can serve the varied needs of a range of cultural institutions promises to be a huge undertaking. Keeping them up to date will be an even bigger challenge if the goals are to keep pace with technology and to respond to users' expectations of image databases.

We can accomplish much, however, if we follow the lead of our sponsoring organisations. As noted in the introduction to this joint conference, RLG and NPO "act on behalf of their constituencies to establish uniform best practices and to disseminate widely the results of consensus-based working groups." We can establish consensus-based guidelines, if we focus on best practices, rather than best products. Despite changes in technology (even the arrival of whatever will replace raster images as the best format for copying and distributing materials in our collections) we can rely upon a stable approach to managing the process of converting materials into electronic form. Controlling the digital capture workflow does not guarantee that things will go right, but making reasonable efforts to control what we can will put us in the best position to ensure that high production will have the potential to yield high quality.

## **BIBLIOGRAPHY**

### **Handling**

"Digitising the Primary Source Material at the English Faculty Library, University of Oxford." Available at <http://info.ox.ac.uk/jtap/reports/digit/>

Library of Congress National Digital Library Program, RFP96-18 for Digital Images from Original Documents, Sections C.4 and C.6.2. Available at <http://lcweb2.loc.gov/ammem/prpsal/ctoca.html>

Mayer, Manfred. "Scanners," message posted to Conservation DistList, 9 December 1997. Available at <http://palimpsest.stanford.edu/byform/mailling-lists/cdl/1997/1549.html>

#### Examples of Cradles

ImageAccess Home Page, click on "New Features" hyperlink beneath the BookEye Scanner. Available at <http://www.imageaccess.com/>

Minolta PS3000 scanner, available at <http://www.minolta.de/beo/office/produkte/scanner>  
Also see, Peter Leggate, "Internet Library of Early Journals Annual Report (August 1997)," scroll to the section, "Scanning with the Minolta PS3000 face-up book scanner." Available at <http://www.bodley.ox.ac.uk/ilej/papers/ar1997.htm>

Mintzer, Fred C., et al. "Toward on-line, world-wide access to Vatican Library materials," IBM Journal of Research and Development, Vol. 40, No. 2 - Services, Applications, and Solutions, 1996, see Figure 3. Available at <http://www.almaden.ibm.com/journal/rd/mintz/mintzer.html>

Riser, John. "Hand-crafted cherry book cradles." Available at <http://www.lib.virginia.edu/speccol/scdc/cradle.html>  
See also, available at <http://www.lib.virginia.edu/speccol/scdc/cradle2.html>

#### Systems: Evaluating Quality

Colet, Linda Serenson, Kate Keller, and Erik Landsberg. Digitising Photographic Collections: A Case Study at the Museum of Modern Art, NY, Presented at the Electronic Imaging and the Visual Arts Conference, The Louvre Museum: Paris, September 2, 1997. See, especially, the section, "Evaluating Digital Equipment."

Gann, Robert, Ph.D., Judith Goode, and Rod Wadas. Reviewing and Testing Desktop Scanners. Second edition. Hewlett-Packard Company, Learning Products Department, 700 71st Avenue, Greeley, CO 80634, 5963-5550E, Second Edition, January 1995.

PIMA (Photographic & Imaging Manufacturers Association, Inc.) / IT10 Technical Committee on Electronic Still Picture Imaging home page. Follow various links to track progress in developing standards and tools to measure noise, resolution, tone, and colour reproduction. <http://www.pima.net/it10a.htm>

Reilly, James and Franziska Frey. Recommendations for the Evaluation of Digital Images Produced from Photographic, Micrographic, and Various Paper Formats, May 1996. Available at <http://lcweb2.loc.gov/ammem/ipirpt.html>

Williams, Don. "What is an MTF ... and Why Should You Care?" RLG DigiNews, vol. 2, no. 1, February 15, 1998. Available at <http://www.rlg.org/preserv/diginews/diginews21.html>

#### Digital Masters: Quality Issues

Each of the following publications presents a clear rationale for the scanning guidelines that have been advocated. In some cases, the authors relate scanning decisions to the characteristics of the source material; in others, to meet baseline output characteristics of digital images, prints, or film.

Agfa Educational Publications. An Introduction to Digital Scanning, Digital Color Prepress Series, Vol. 4. Available at <http://www.agfahome.com/publications/dcp4page.html>

D'Amato, Donald, Ph.D., and Rex C. Klopfenstein, Mitretek Systems, Inc.  
"Requirements and Options for the Digitisation of the Illustration Collections of the National Museum of Natural History," March 1996.  
Available at <http://www.nmnh.si.edu/cris/techrpts/imagopts/index.html>

Ester, Michael. Digital Image Collections: Issues and Practice. Washington, DC: Commission on Preservation and Access, 1996.

Kenney, Anne R. "Digital-to-Microfilm Conversion: An Interim Preservation Solution," *Library Resources & Technical Services* 37 (October 1993) pp. 380-401; Erratum 38 (January 1994) pp. 87-95.

Neff, Raymond K., "A New Consortial Model for Building Digital Libraries," Session #7 Multi-Institutional Cooperation, Scholarly Communication and Technology: Conference Organised by The Andrew W. Mellon Foundation at Emory University April 24-25, 1997. See, "Appendix E: Technical Justification for A Digitisation Standard for the Consortium." Available at <http://www.arl.org/scomm/scat/neff.html>

Puglia, Steven and Barry Roginski. NARA Guidelines for Digitising Archival Materials for Electronic Access, January 1998. Available at <http://www.nara.gov/nara/vision/eap/eapspec.html>

#### **Derivatives: Quality and Production Issues**

Bancroft Library, "Digitising the Collection: Image Capture," California Heritage Collection, University of California, Berkeley. Available at <http://sunsite.berkeley.edu/CalHeritage/image.html>

Fleischhauer, Carl. "Digital Formats for Content Reproductions." National Digital Library Program, Library of Congress, July 13, 1998. See Section IV, and note descriptions of contrast stretching, diffuse dithering, and other enhancements. Available at <http://lcweb2.loc.gov/ammem/formats.html#IV>

Price-Wilkin, John. "Just-in-time Conversion, Just-in-case Collections: Effectively leveraging rich document formats for the WWW," *D-Lib Magazine*, May 1997. Available at <http://www.dlib.org/dlib/may97/michigan/05pricewilkin.html>

Stephenson, Christie and Patricia McClung, eds. *Delivering Digital Images, Cultural Heritage Resources for Education: The Museum Educational Site Licensing Project, Volume 1*. Los Angeles, California: The Getty Information Institute, 1998.

Süsstrunk, Sabine. "Imaging Production Systems at Corbis Corporation," *RLG DigiNews*, vol. 2, no. 1 (February 15, 1998). Available at <http://www.rlg.org/preserv/diginews/diginews2-4.html#technical>

University of Virginia Library Electronic Text Center, "Sample Scans: The Electronic Archive of Early American Fiction," scroll to "File size and image quality comparisons." Available at <http://etext.lib.virginia.edu/projects/scantest.html>

**File Organization: File Names as Structural Metadata**

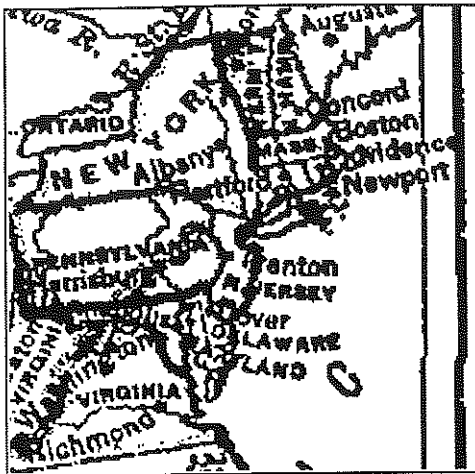
Library of Congress Internal Documentation, "Turning pages within in a digital reproduction," May 4, 1998. Available at <http://memory.loc.gov/ammem/award/docs/page-turning.html>

"Naming and Linking Strategy," RLG Digital Collections Project: Studies in Scarlet: Marriage, Women, and the Law, 1815-1914. Available at <http://lyra.rlg.org/scarlet/name.html>

Preservation Resources, Scanning Questionnaire, 1998. p. 4; quality control documentation. Available at <http://www.oclc.org/oclc/presres/scanning/scanquestion.pdf> and <http://www.oclc.org/presres/scanning/qa.htm>

Seaman, David. "Guidelines for SGML Text Mark-up at the Electronic Text Center." See section, "Specific Procedures for Adding Image Headers," which outlines batch methods to populate file headers with metadata. Available at <http://etext.lib.virginia.edu/tei/uvatei11.html>

Technical specifications alone not accurate predictors of image quality:



300dpi 1-bit TIFF image, Scanner A



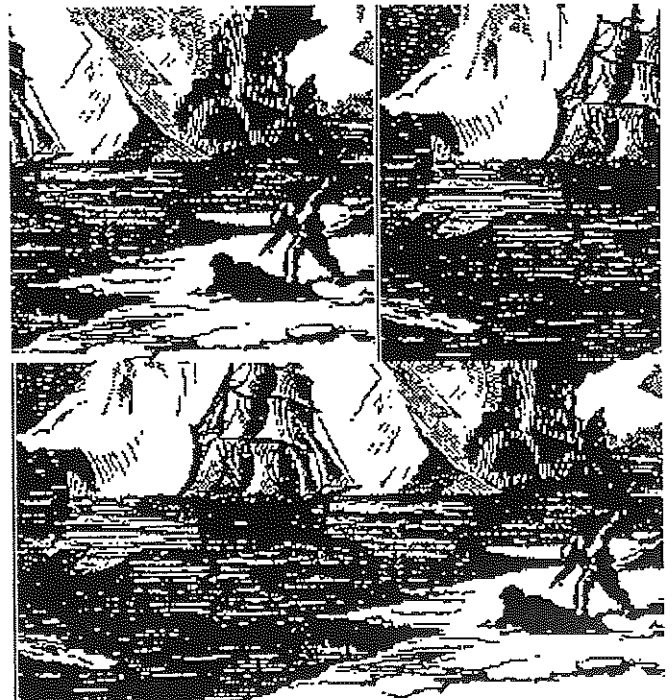
300 dpi 1-bit TIFF, Scanner B

300 dpi, Scanner A

300 dpi, Scanner B



ICEBERGS NEAR GREENLAND,



ICEBERGS NEAR GREENLAND.

Note: The images above were created by the same service bureau from the same reel of first-generation (camera negative) black and white 35mm microfilm created in 1998 at Harvard University. These examples are provided to demonstrate that standardised measurement tools (a

scanning target and associated software) for objective measurements of output image quality would be tremendously useful to access and compare scanners with similar technical specifications. In this case, the linear array of the CCD of Scanner B was slightly higher than Scanner A, yet Scanner A is far more expensive and has more sophisticated software. Even more dramatic quality differences in the detail reproduction capabilities of similar materials have been noted in other Harvard projects where different scanners of the same optical resolution (400 dpi) were used. A target that would permit objective measurements of noise, spatial resolution, and image processing combined would presumably help to make "digital benchmarking" (currently defined as relating dpi to detail to predict quality) a more viable technique for determining the minimum resolution needed from a given scanner to meet quality requirements.

For a dissenting opinion, see the message posted by Dr. Larry J. Schaaf to the PHOTOHST list on October 8, 1997. Dr. Schaaf concludes his letter by saying, ". . . direct digital scans are far too costly in potential damage to the original and in long-term viability. At least for some years, I think that scanning from intermediate transparencies is the only responsible and sane course of action."

See, for example, the most recent contribution to this discussion, in which the author measures light levels from a number of flatbed scanners and assesses risk against the total estimated life of an object rather than stating a threshold number of lux hours at which damage occurs. Timothy Vitale, "Light Levels Used in Modern Flatbed Scanners," RLG DigiNews, Vol. 2, No. 5, October 15, 1998. Available at <http://lyra.rlg.org/preserv/diginews/diginews2-5.html#technical>

One of the Library of Congress contractors, Systems Integration Group, manufactured a plexiglass support to make the scanner conform to the contract specifications (RFP96-18). At the American Library Association Annual Meeting this summer, the Minolta Corporation demonstrated a new product that included a wire-frame removable support that could be used with its cradle. They have not yet made it available in the United States as an add-on component.

Cornell University's Digital-to-COM Project incorporated a scanned technical target with each "digital book." These targets are not included with the on-screen versions of the books, but they were output to microfilm to document the quality of the scanner that was used to digitize the original printed materials. For more information see, Anne R. Kenney, Digital to Microfilm Conversion: A Demonstration Project 1994-1996, Final Report to the National Endowment for the Humanities, PS-20781-94. Available at <http://www.library.cornell.edu/preservation/com/comfin.html>

The Making of America II Testbed Project White Paper, Version 2.0 (September 15, 1998). Available at <http://sunsite.Berkeley.EDU/moa2/>

Stanislav Psohlavec, "Digitisation of Manuscripts in the National Library of the Czech Republic," *Microform & Imaging Review*, Vol. 27, No. 1, Winter 1998, 21-26.

See, for example, Marthyn Borghuis, Hans Brinckman, et al., TULIP Final Report (Elsevier Science, 1996), "Section II 1.4.5 Checksums". The authors note that on average, they encountered one incorrect bit (either during FTP of images or when writing them to CD) ("resulting in a fully incorrect image" per approximately 20,000 images, or one wrong bit per 1.6GB. Checksums were introduced to monitor image quality and "to safeguard the integrity" of the digital collections. Available at <http://www.elsevier.nl/homepage/about/resproj/trmenu.htm>

A minimum of 10 bits per pixel of noise free data is required to achieve a density of 3.0.



## Guidelines for Image Capture: A UK Perspective

### Dr Jane Williams

#### Technical Advisory Service for Images (TASi)

#### Introduction

This paper gives a UK perspective on image capture activities and provides recommendations for the production of guidelines for image capture. The paper is based on the outcomes of a workshop held by the Technical Advisory Service for Images<sup>(1)</sup>, the Electronic Libraries Programme (eLib)<sup>(2)</sup> and the Non-Formula Funding for the Humanities Programme<sup>(3)</sup> (all of which are funded by the Joint Information Systems Committee – JISC<sup>(4)</sup>) and other work carried out by TASi.

The author wishes to acknowledge the following people: Alan Lock, Karla Youngs, Oliver Moss, from the Technical Advisory Service for Images, University of Bristol and members of the image capture workshop discussion group: Dr David Saunders, National Gallery; Dr David Cooper, University of Oxford; Bill Jupp, University of Leeds; Alan Lock, University of Bristol. This paper draws heavily on reports written by the aforementioned, many of which are available from <http://www.TASi.ac.uk/>.

#### Technical Advisory Service for Images: Overall aims for Guidelines on Digital Imaging

The Technical Advisory Service for Images is a national service for higher education whose remit is “to advise and support the UK academic community on the digital creation, storage and delivery of image related information”. TASi was established partly in recognition of the proliferation of valuable, yet highly diverse resources that were accessible via networks and in the realisation that if they were to be exploited to their full potential and be available for years to come such resources needed to be:

- easy to search
- properly organised
- and of high quality.

It is this last issue, quality, that really encompasses the topic of this paper, that is, digital image capture.

The objectives of TASi are two-fold: first, we encourage the sharing and promotion of technical expertise and standards; that is we actively work with other service providers, initiatives and projects both within and outside UK HE to understand and reach a common consensus on best practice. Second, we disseminate this information to our clients (information professionals, librarians, teachers and researchers who are thinking of or actively engaged in digitising their collections) in the form of reports, case-studies, training events, etc. to ensure the creation of high quality digital image resources.

Our overall aim therefore is to educate and help our clients make informed decisions concerning all the issues surrounding the creation of digital image libraries.

Before taking a look at the specific issues of image capture and the factors affecting quality I would like to make two overall recommendations for any guidelines arising from this conference.

**Recommendation 1:** Education not prescription - it is imperative that we develop a global common framework to act as a guide to help make informed decisions based on local needs and factors.

**Recommendation 2:** Decisions made in one area will effect those made in another and we must educate people in the consequences of their choice.

So I would like to confirm the message from Anne Kenney, our Keynote Speaker, that it is guidance and not guidelines that we should be creating and I was pleased to hear other speakers from the conference refer to and confirm these recommendations. One of the most common enquiries received by TASi is “I have 5,000 photographs, which scanner should I

buy?" Our reply is a series of questions: "What are you trying to achieve and what are your objectives?" "What types and sizes of photographs do you have?" "Are they all black and white, colour, sepia or a mixture?" "What condition are they in?" "Do you plan to digitise the whole collection?", etc. There are many other questions too, such as whether they hold copyright or have permission to digitise, do they have the technical infrastructure and institutional support to embark on such a project, etc. The answers to all these questions will determine how they proceed with image capture and more importantly whether they proceed at all with the project. Many of our clients decide they are not ready to take on such a project and decide wisely to spend the next 12-18 months planning appropriately for when they can.

### **Joint TASI-eLib-NFF Digital Library Workshop**

The remainder of this paper focuses on the subject of digital image capture and presents some of the outcomes of a workshop held jointly by TASI and the eLib and NFF programmes. This workshop held in June 1998 aimed to bring together all the current JISC funded digital image capture projects to discuss current practice, decide best practice and make recommendations for future work - that is, what worked, what do we know, what didn't work and where do we need further guidance and solutions. A total of fifteen projects representing 35 collections of every conceivable type of material: Victorian journals, grey literature diaries, ballads, manuscripts, photographic prints and transparencies, negatives (cellulose nitrate negatives), postcards, posters, banners, art originals, music scores, rocks and even a Sabre toothed cat skeleton were present.

Other digital imaging initiatives from outside HE were also invited and representatives from SCRAN (Scottish Cultural Resources Access Network)<sup>(5)</sup>, the National Gallery (MARC/WISEUM project)<sup>(6)</sup>, the Natural History Museum (Cook's First Voyage)<sup>(7)</sup> and the Heritage Lottery Fund<sup>(8)</sup>, Research Libraries Group<sup>(9)</sup> and British Library<sup>(10)</sup> digitisation initiatives. Five areas for discussion were selected: criteria for selection, digital image capture, managing access, planning and managing the digitisation chain and digital preservation. Prior to the workshop a survey was sent to each project participating to document each project and form a set of case studies and to determine common issues. Details of examples presented in this paper and other case studies are available from the TASI WWW site (<http://www.TASI.ac.uk/building/building2.html>). The following presents the outcomes of the digital image capture group.

### **Outcomes of the Digital Image Capture Group and Recommendations for Digital Image Capture**

The first question the digital image capture group explored was "Why digitise, what is the purpose of digitising?" The majority of projects surveyed said digitisation was carried out for the dual purposes of providing both local and remote access to their collections and in doing so provided a means of preservation of the original by reducing handling of the materials. The desire to digitise is driven by a combination of the two. For some collections, for example film stock on a nitrate base, the preservation aspect is critical as the original material is deteriorating rapidly. In extreme examples, the digitisation process may be a last chance to preserve intellectual content and could destroy the source material. Whether the resource is unique will also need to be considered.

Image capture needs to be tailored to suit the purpose of the digitisation project, but, as far as possible, should also be influenced by potential future uses of the digitised material, some of which may not be foreseen at the time of digitisation. As a general policy, digitisation should not result in significant information loss.

**Recommendation 3:** Digital image capture should not result in significant information loss. But what do we mean by significant information loss? How do we measure this?

To define "significance" in this context remains difficult. Significance may be affected by economic restrictions as well as intellectual considerations. Under certain circumstances

only the textual information that can be extracted from a document (for example by OCR) may be considered significant. However, this may not suit the objectives of some users and might preclude further analyses of the digital images from a different perspective. Ideally we should provide one scan to fit all purposes. But is this really possible? Are we able to achieve this with current technology? I will return to this issue later in this paper.

The digital image capture group presented their main findings under three headings:

1. Imaging methodology
2. Formats
3. Image Quality

### **Imaging Methodology**

Returning to the most frequent request from TASi: "What scanner should I use?" the answer is "Well it depends...". The choice of imaging methodology will be determined by a number of opponent factors:

- Higher quality generally increases the cost as the equipment is more expensive
- An increase in the speed of the workflow (i.e. processes are sped up) tends to reduce costs, but will limit the quality of data
- The amount of damage that can be caused will determine how the material is digitised
- Conservation considerations - if materials need conserving before they can be digitised this adds time and hence increases costs.

Any method adopted will depend also on the type of information to be captured coupled with the degree of information loss that can be tolerated and the physical characteristics of the object - can the object be touched, placed on a glass sheet, bent, opened (in the case of a bound volume), etc? If the preservation is not important can the object be damaged or destroyed in order to simplify or speed up the digitisation process? In the case of bound books this so called destructive scanning where the binding is removed improves the quality of the digital image but the original integrity of the book is destroyed. If there are multiple copies of the volume then there may be just reasons for following the path of destructive scanning. The group also requested guidance in the area of costs for scanning different types of documents and image based materials.

**Recommendation 4:** The final decision should be made by those in charge of preservation or the custodians of the collections and not those in charge of technical issues.

**Recommendation 5:** Formulae for a matrix of options should be developed as a guide for dealing with different classes of material including photographic guidelines and rough bands for costs.

**Recommendation 6:** For mixed collections it should not be assumed that one imaging system be used for all objects.

To re-iterate a second point made by Anne Kenney in her keynote paper, 'one size does not fit all'. There may even be two projects with similar types of materials but factors governing the originals may determine the selection of different image capture set-ups.

### **Formats**

Once again the choice of format will depend on a number of things: the nature of the image, its intended purpose and the amount of information that can be lost. The amount of information loss that can be sustained will depend also on whether you are dealing with the digital master file or a surrogate copy which may have been manipulated. A number of surrogate copies with different resolutions and colour depths may be derived from the digital master to suit a variety of different purposes. Each purpose will determine a different choice of file format. Information about image file formats is widely available and this is one area where a technical understanding of what the different image file formats can offer is able to provide us with more specific guidelines. TASi divides these into four different categories depending on what the project is trying to achieve:

1. Acquisition
2. Archival
3. Editing/Manipulation - working copy
4. Delivery and Presentation - terminal format

#### *Acquisition*

This represents the first and most important step in the process. Errors made here will probably necessitate returning to the original to re-capture. It is unlikely that the resources will be available to scan again a few years down the line and therefore it is desirable to obtain the highest quality scan possible at this stage within current technological limits and conservation considerations. The requirements for file formats chosen for acquisition should:

- Be lossless so as to maintain the highest fidelity
- Retain the best colour/greyscale information possible - at least 24 bit colour/8 bit grey
- Retain any acquisition device gamma/colour calibration information
- **Recommendation 7:** PNG and SPIFF meet a lot of the requirements for a modern raster image format. Both:
  - Are open formats
  - Support metadata (as text, some fields pre-defined)
  - Support lossless compression (SPIFF supports lossy as well)

At the time of writing support for PNG was patchy but increasing, SPIFF has yet to 'take off' but most JFIF readers (i.e. software able to read JPEG files) should be able to import the file.

**Recommendation 8:** Support the uptake of PNG and SPIFF; until support for both these formats increases TIF is probably the 'safer' format.

#### *Archival*

A copy of the digitised documents should be archived at the highest resolution as is practical (if possible at acquisition resolution) in keeping with the storage facilities available. With mass storage decreasing rapidly in costs and with more choices available (e.g. CD-ROM; tape; DVD-ROM; removable cartridges; optical discs) there are options to suit most budgets. The same requirements appear again. In addition the format chosen should also:

- Be a standard one - proprietary formats may not be readable in 5 or 10 years time
- Preferably hold metadata associated with the image e.g. cataloguing; copyright information - the data will then move with the image: retrieval is easier; lost or broken databases become less problematic, etc.

**Recommendation 9:** TIF, PNG and SPIFF as above

#### *Editing/Manipulation - working copy*

For work in progress the proprietary formats of graphics programs can be very useful. Editing can be a time consuming process and formats that can save extra information (e.g. masks, channels and undo features) mean that subsequent editing can resume from where the last session finished without having to recreate the masks etc. Unfortunately this conflicts with the archival requirements above (standard files, metadata) and this is where post-edit archiving has an advantage.

If the images require a lot of manipulation then ideally manipulation would take place prior to archiving so that complex and time-consuming editing does not need to be repeated. However, ensure you make a copy of the image before manipulating such that if the editing process goes too far, i.e. information is irretrievably lost, you can return to your copy.

Whether one should archive an image and then carry out any manipulation or vice versa is debatable. If it is intended to edit the digitised image before archiving, the editing process must not remove significant information - re-editing may well be easier than re-digitising. If you are post-editing and storage space is not a problem then consider archiving the manipulated files as well.

Formats should be chosen on the amount of editing required and whether it is pre- or post-archival.

#### *Delivery and Presentation - terminal formats*

These are referred to as terminal formats because it is likely that to produce a further surrogate for a given purpose you will need to return to the archival or editing/manipulation copy. It is not recommended to convert further a delivery or presentation surrogate to suit a different purpose as by this stage significant information may have been lost. This stage is the last in the image manipulation chain and you may have more than one end point e.g. you may require small thumbnails of images for browsing purposes and images to be used for presentation purposes.

Questions that one should ask when considering what file format to choose include:

- What device is the image being delivered to (i.e. what are the needs of our audience) - screen, printer, slide?
- What are the capabilities of the receiving device? How many colours can it handle? What resolution?
- What is the nature of the image (photorealistic e.g. a scene, or presentation e.g. a graphic or pie chart)
- How am I going to deliver it? CDROM, tape, internet (modem or LAN/WAN connection?), cartridge?
- Do I know what use the image is going to be put to after delivery?
- 

With so many considerations, combined with the proliferation of file formats, it is little wonder that this is a subject for continued confusion and debate. With this in mind the following are more in the form of guidance rather than guidelines.

- Printed output: talk to the person doing the printing. This becomes even more important if it is a commercial printer (mistakes are costly). The majority of printers will probably require something along the lines of a TIFF/EPS file in CMYK format and you may need to ask for their printer profile.
- Screen - issues requiring consideration:
  - different computers may have considerable differences in their screen gamma values which can drastically alter the appearance of images
  - has the monitor been set up properly i.e. position, brightness, contrast, colour temperature etc. adjusted for local viewing conditions - these simple optimisations are rarely done and users need to be educated on how to calibrate monitors to achieve optimum viewing conditions
  - restricted colour depth and resolution really make themselves felt - it is surprising the first time that someone sees, on a 256 colour display, a 'higher quality' image (24bit TIF say) looking worse than a 'lower quality' one (e.g. 256 colour GIF)
  - is it a VDU, TFT panel or projection system?

If the delivery is going to be made via a network and with all these parameters to consider, there is a strong case to be made for using a duplicate of the file used for archival storage and performing an on-the-fly conversion to deliver a file that is in the resolution, colour depth and format that the user requires.

**Recommendation 10:** for photorealistic images where bandwidth or storage space are a problem then consider JPEG (at least for preview files), PNG (8 bit) or GIF. For JPEG the degree of compression used will depend on the amount of acceptable degradation. If bandwidth is not an issue then consider using TIF or PNG (24/48 bit). For presentation type images the choice is generally simpler - use PNG (8 bit) or GIF. If SPIFF support becomes more widespread then it can be used for both types of image under most conditions - choose the settings that suit the circumstances.

The projects surveyed revealed general consensus in using uncompressed TIF as an archival format and delivering via the Internet as GIF or JPEG. Some projects are using PhotoCD as an archival format and depending on the type of content and how important colour information is problems may be encountered further down the line when migrating to different formats as part of a digital preservation strategy.

#### *Issues for Digital Preservation*

Migration of master files to other formats should form part of any digital preservation strategy. We are witnessing the emergence of more advanced image file formats all the time (e.g. PNG, FlashPix). When migrating between formats the following should be considered:

- How much data will be lost when converting between different formats? And between different compression algorithms particularly if they are both lossy?
- How is the colour effected when changing to file formats which support different colour spaces?
- Do the metadata tags remain intact between formats?

It is impossible to know what to do when we do not even know what the image file format in 5 years time will be. As yet there is no ideal image file format for archival storage and the community should be doing all it can to influence new and emerging formats.

Some general recommendations then for guidance and guidelines for image file formats:

**Recommendations 11:** Avoid proprietary formats: use open and currently stable formats

**Recommendation 12:** A practical investigation involving subjective and objective measurements into format interchange and data loss should be carried out

**Recommendation 13:** A record of format interchange should be kept as part of digital preservation metadata and those charged with a digital preservation brief should acquire an understanding of the implications

**Recommendation 14:** New and emerging standards should be tracked and influenced by major digital image capture initiatives and bodies

#### **Quality Control**

Once again we are concerned with the minimisation of information loss. The quality of image required for different applications will be determined by the individual custodians and project co-ordinators, taking into account the type of material to be captured, economic and time constraints and the use to which the digital data will be put.

Image quality standards are primarily concerned with minimising significant information loss and maintaining consistency between sources, devices, operators, sites and so on – this applies throughout the digitisation chain and extends beyond the period of the creation of the archive and becomes a maintenance issue.

Image quality can be divided into two main types: aesthetic and technical, the former being very much a subjective matter. The technical quality of an image can be broken down into a number of topics including:

- Resolution (linear/area)
- Bit depth
- Colour/tonal accuracy

It cannot be emphasised strongly enough that any computer used for quality assessment/assurance (QA) of images must be properly set-up and calibrated – this includes

any software used for QA and the room in which the QA will be carried out. A simple checklist can be found at <http://www.TASi.ac.uk/building/qa1.html>

It is also worth noting that most image based sites/archives assume that everyone has normal colour vision, approximately 10% of males (and a lower percentage of females) have anomalous colour vision – you should be aware of this in your QA process.

### **Resolution**

The term resolution is used to refer to image size (in pixels), the resolution (in dots per inch, dpi) of the image acquisition device (e.g. scanner) or the dpi information within an image file that will be used for scaling printed output.

Depending on what image capture device you use (a scanner or digital camera) you may have control over the resolution at which you scan. In the case of digital cameras, the resultant image size is fixed or one of a few fixed options and there is virtually no control over resolution. With scanners, the resolution at which you capture can be varied but it is the optical resolution which then becomes the limiting factor and the image size will be dependent on this variable and the area to be scanned.

Even then, things are not always as straightforward as they might at first appear: two flatbed scanners with apparently identical resolution capabilities may exhibit a noticeable difference in resolving power on real scans. It is highly likely that the worst one will have an optical system that does not do justice to the CCD (charge coupled device, the transducer that converts the light into electrical signals).

How, therefore, can we obtain objective figures for resolution? For digital cameras the lens testing targets that have been used for many years in the photographic world should provide suitable results/comparisons. For scanners it is relatively early days, scaled down high quality prints of the above targets are one possibility, another is to use dedicated test charts such as the AIIM or IEEE std 167A-1987 fax test chart. Refer to the paper given by Stephen Chapman at this conference, for further information on technical targets.

### **Bit depth**

Important factors to consider from the image quality viewpoint are 1) the bit-depths retrievable from the system (as a whole) and 2) the amount of noise.

Using scanners as an example of the first point, the specification may say the scanner has 36-bit sampling (i.e. 12 bits per RGB), but the scanner software may not allow the retrieval of anything more than 24 bits (8 bits per RGB). Additionally, the software may not give you much control over how the 36 bits are compressed (downsampled) into 24 bits. If the originals have a wide dynamic range and maintaining this is important to your imaging requirements then you should take this into consideration – obviously you will also need to use file formats and software that support these higher bit depths.

Noise (random signal variation) exists in every capture device, the amount of noise is one of the differentiating factors between ‘professional’ and ‘consumer’ devices. Nearly all capture devices (digital cameras, scanners etc.) tend to exhibit more noise at low illumination levels. For digital cameras this means that the external lighting can play an important part (try not to work in dim conditions), whereas scanners usually use their own internal light source and normally this is not under user control. In both cases this translates into more noise in the darker (shadow) parts of the image (note that for negatives this equates to highlights once the image has been converted to a positive).

Obtaining noise measurements can be difficult – one possibility is to look at the variation within blocks of known even colour, the histogram tool(s) provided in most image manipulation software may give you sufficient information depending on your needs. If accurate noise measurements/processing are vital then consider using image processing software (such as MIPS, TNT/TNTLite and Noesis Visilog).

### **Colour and tonal accuracy**

These have been under consideration for many years by a relatively small number of people working in digital imaging but it is only recently that their importance has gained recognition by a wider audience. Interest in these areas is likely to increase greatly in the near future.

The importance of colour and tonal accuracy will vary according to circumstances, e.g. colour accuracy is irrelevant in monochrome radiography but tonal accuracy can be very difficult to maintain (due to the wide dynamic range in most radiographs). Surprisingly, colour accuracy when digitising old photographs can be problematic, this is partly due to them rarely being black and white but more often of a sepia or similar tint and it is these tints which can be hard to capture accurately. If these tints (and ageing discolourations) are not important to your archive then the problem is solely in accurate tonal reproduction. For images which have a diagnostic content, colour accuracy can be of paramount importance.

One of the difficulties in accurate colour rendition is the number of devices that may be present in the digitisation/editing/viewing chain: scanners, digital cameras, monitors, printers and so on. Each of these devices will have their own colour response, for instance neither monitors nor printers are linear devices (e.g. a 20% increase in the blue signal, say, does not necessarily give corresponding increase in the output), additionally the gamut (the range of colours a device can reproduce) varies according to device type (e.g. monitors can display colours that printers cannot reproduce and vice versa). Consider also that each model of a particular device is likely to have a different response to another model, e.g. two different models of scanner should not be relied on to give identical scans from the same source material.

Even when an image has been digitised there are problems when moving from one colour space to another, for example: when going from RGB (for monitors, say) to CMYK (for printing) colours can be mis-mapped due to the inability of the printing device to handle a particular colour (i.e. the colour is 'out of gamut'), if the resultant CMYK file is now converted back to RGB there is no guarantee (in fact it is highly unlikely), that the original (mis-mapped) colours will be restored accurately. Also do not assume that the RGB file from a scanner is working in the same RGB colour space as your monitor.

**The moral here is to keep the archival image in one colour space. We suggest keeping to colour space of the capture device (normally RGB), converting only as and when needed and when the colour space of the destination device is known. This last point should probably, with correct guidance, be left to the end-user of the image.**

### **Colour Management Systems**

It is to tackle just these sorts of problems that colour management systems (CMS) have been designed.

The primary function of a CMS is to 'close the loop' between the original and output images, such that the output image matches the original as closely as possible. They can range from low-cost software-only applications to full-blown, hardware-based device profiling systems. For further information see <http://www.TASi.ac.uk/building/qa1.html>. Note that if you are having your images digitised elsewhere you should ask them what CMS they have in operation and request device information where appropriate. For example most PhotoCD bureaux should offer this as standard. We strongly suggest that you avoid those who say, "What is a colour management system?"

### **So how does a CMS work?**

For a basic system you may simply be asked to make adjustments (to software settings) based on visual comparisons with a test target (which may be on or off screen). More complete/complex systems will require that a high quality test target is digitised (the nature of the target depending on the device being calibrated), the subsequent output (screen and/or monitor) is checked by another device (photometer, densitometer, spectrometer) that feeds



its readings into the software that then works out device profiles and applies compensation curves as and when necessary.

The simpler applications are more often system colour *calibration* tools than colour *management* systems (see box above for a basic monitor set-up procedure).

For a more detailed overview of colour management history and process a visit to the ICC (International Colour Consortium) website is worth a visit (<http://www.color.org/>). ICC was formed to tackle the problem of cross-platform and vendor specific colour management. The method used is to write the colour characteristics of a device (e.g. scanner, monitor, printer) to a profile which can then be used to translate colour information accurately to another ICC characterised device.

### **Why should I use colour management?**

If you are going to the trouble of creating an image archive you, presumably, want others to see (and perhaps print) those images and, also presumably, at their best. Without colour management this is unlikely to happen, differences in computer display capabilities and set-ups will lead to inconsistent reproduction – and this is on the same platform. If we bring in cross-platform issues then the problems increase enormously.

The lowest common denominator approach, i.e. using a browser-safe 216 colour palette (that is, the palettes that the common web browsers use when working on 256 colour displays), might work for some images for on-screen use but for high quality images (whether for use on-screen, printed or both) this approach is usually not good enough.

If you are charging for the use of your images then the issue becomes more problematic: the client sees an image on your site, decides it is just what s/he needs and orders the high resolution version, when s/he gets it the colours do not appear as they did on-line – if you are not able to demonstrate that the problems are not at your end then that's one person who won't be coming back.

In America law suits have arisen out of colour matching problems with goods purchased over the Internet.

**Recommendation 15:** Raise current standard of practice in assessing image quality; review and assess emerging technologies

**Recommendation 16:** Quality assessment should be carried out with the original material available

**Recommendation 17:** Need a best-practice approach to image quality; common standards may become more important in the future

Once image libraries begin to co-operate and share common indices, making research across related, but geographically separated collections a real and interesting possibility, common standards become more important. The danger, however, is that the standard adopted is the lowest common denominator.

**Recommendation 18:** Calibration of equipment (hardware and software) and room/studio conditions are of utmost importance; do not compensate for equipment not set up properly; end users should be educated in setting up display devices; checklists should be developed

**Recommendation 19:** Keep archival images in the colour space of the capture device and convert only as needed

**Recommendation 20:** Produce guidelines for objective measurements

### **File Naming**

Although this paper focuses on image quality and how it is effected by factors governing choice of image capture methodology, file formats and other factors, the overall quality of the digital collection also depends on how it is organised and how easy it is to search - a strategy for naming files will help achieve overall quality.

As pointed out in Kenney and Chapman (1996)(11) the organisation of scanned images to enable users to make the most of them involves three closely-tied strategies:

- **Organisation** of the scanned image files into a disc hierarchy (a system of directory, sub-directories and files) that logically maps to the physical organisation of the images/visual materials.
- **Naming** of the scanned image files in a strictly controlled manner that reflects their logical relationships
- **Describing** the scanned images files using metadata files linked to the images via the filename.

Each image should have a unique identifier which should not include transient information. Since the descriptive metadata and the image itself are not likely to be located together it is important to develop a clear strategy for resource identification. This means, for example, the allocation of a unique identifying number or code to every image - images can then be searched using that number and descriptive metadata can identify the resource and where copies of the various surrogates are held. Identifiers should not hold transient information. For example, an image of a dog within the Bristol Biomedical Image Archive has:

- An image id of 20437
- Its location of the WWW is <http://images.net/PNG/20347.jpg>
- Its location in the local file tree is `c:\image\batch1\smith\20437.png`

Image 20437 will always be found and details of where the various copies are held. The locations can be changed without losing information pertaining to the image. Surrogates may have letters as prefixes or suffixes to denote various resolution and colour depths.

**Recommendation 21:** There should be at any point in time a clear and well documented mechanism for deriving the location of an image from its identifier and other relevant metadata.

## Conclusions

In conclusion and to answer the question posed to the conference speakers "Are we in a position to be able to put together guidelines (or provide guidance) for imaging?" the answer is Yes but this must be done collaboratively and in partnership with those actively involved in such activity. However, guidance is not, on its own, enough and one definite overall recommendation from the TASI-eLib-NFF digital image library workshop was the need to provide hands-on workshops to demonstrate some of the issues raised in this and other papers.

**Recommendation 22:** Those responsible for decisions on selection, methodology, image quality and all aspects of the digitisation chain need to be educated in the consequences of their choices.

There is still a lot we do not know. We are not able to guess accurately how digital image capture technologies will develop. It is not feasible to produce one scan for all purposes for all time but we are currently able to digitise materials to a standard to achieve some of our requirements in the short-term (5-20 years). We are certainly not at a stage where we can digitise for preservation and throw away our analogue counterparts. We do not yet know the costs of digital preservation and migration. The actual digitisation is a crucial step but it is only one activity in the whole of the digitisation chain. Other issues such as cataloguing and the creation of the information pertaining to the progression of an image through the digitisation chain, copyright clearance, etc. are more costly and time-consuming and are the real activities that are likely to be done once and once only. In some cases, digitisation is the only option to preserve materials for future generations and in others digitisation can only be carried out once due to handling constraints. In these cases we have to make the decision whether to proceed or not and if so, ensure we achieve the best possible quality given current technological constraints. In other cases, however, I would like to end by proposing that re-digitisation of materials or a selection should not be ruled out in years to come. We should

not necessarily approach projects on the assumption that we will only have one chance to digitise our materials. Given the rapid advances in digital technology, the costs of re-digitisation in years to come will represent a small part of the total cost of producing and maintaining our collections and their digital counterparts. Digital image capture, for those collections where direct handling is not too problematic, may become a repeated activity within the life span of a collection and an integral part of preservation activities.

## REFERENCES

- 1 The Technical Advisory Service for Images is a national service for higher education providing advice and consultancy on the digital capture, storage and delivery of image related information across all subjects. <http://www.TASi.ac.uk/>
- 2 The Electronic Libraries Programme. Contributing to the development of the digital library, available at <http://www.ukoln.ac.uk/services/elib/>
- 3 Specialised Research Collections in the Humanities (NFF). Increasing access to researchers through cataloguing, preservation, conservation, digitisation and publicity, available at <http://www.kcl.ac.uk/projects/srch/>
- 4 The Joint Information Systems Committee. Stimulating the cost-effective exploitation of information systems by providing a national academic network infrastructure, content, applications and tools and awareness and training, available at <http://www.jisc.ac.uk/>
- 5 Scottish Culture Resources Access Network (SCRAN). Building a fully searchable resource base of Scottish material, culture and human history, available at <http://www.scran.ac.uk/>
- 6.National Gallery, London, available at <http://www.nationalgallery.org.uk/>
- 7.National History Museum, available at <http://www.nhm.ac.uk>
- 8.Heritage Lottery Fund, available at <http://www.hlf.org.uk/>
- 9.Research Libraries Group. Improving access to information to support research and learning, available at <http://www.rlg.org/>
10. British Libraries Digitisation for Access Initiative. Contributing to the development of the digital library, available at [#http://www.bl.uk/services/ric/diglib/digilib.html #](http://www.bl.uk/services/ric/diglib/digilib.html)
- 11 Kenney, A. R. & Chapman, S. (1996). Digital Imaging for Libraries and Archives. Department of Preservation and Conservation, Cornell University Library, Ithaca, New York 14853. ISBN 0-916-58202-3.



## Issues and Approaches to Preservation Metadata

Michael Day

Research Officer, Metadata

UKOLN

### Abstract

The creation and use of metadata is likely to become an important part of all digital preservation strategies whether they are based on hardware and software conservation, emulation or migration. The UK Cedars project (CURL (Consortium of University Research Libraries) Exemplars in Digital Archiving) aims to promote awareness of the importance of digital preservation, to produce strategic frameworks for digital collection management policies and to promote methods appropriate for long-term preservation - including the creation of appropriate metadata. Preservation metadata is a specialised form of administrative metadata that can be used as a means of storing the technical information that supports the preservation of digital objects. In addition, it can be used to record migration and emulation strategies, to help ensure authenticity, to note rights management and collection management data and also will need to interact with resource discovery metadata. The Cedars project is attempting to investigate some of these issues and will provide some demonstrator systems to test them.

### 1. Introduction

#### 1.1 Metadata and digital preservation

Most current discussions of metadata in the library and information communities have centred on issues of resource description and discovery (e.g. Heery, Powell and Day 1997; Dempsey and Heery 1998). Metadata is commonly understood as an amplification of traditional bibliographic cataloguing practices in an electronic environment. Perhaps the most widely known international metadata standard is the Dublin Core, an initiative that has a deliberate focus on simple resource discovery (e.g. Weibel and Hakala 1998). These are important issues. It is becoming increasingly recognised, however, that metadata has other important roles in the wider task of managing digital resources. For example, publishers and other rights owners are beginning to investigate the uses of metadata with regard to rights management (Rust 1998).

The Library of Congress Making of America II project has identified a threefold division of metadata (Making of America II 1998):

0. Descriptive metadata - primarily used for resource discovery. Formats currently used include the MARC formats, Encoded Archival Description (EAD) and Dublin Core.

1. Structural metadata - data that a system can use to help present a particular digital object to a user.

2. Administrative metadata - data that allows the management of a digital collection.

Preservation is essentially about management. In this scheme, preservation metadata (as with rights metadata) is a specialised form of administrative metadata.

#### 1.2 The Cedars project

The Cedars project is funded by the Joint Information Systems Committee (JISC) of the UK higher education funding councils under Phase III of its Electronic Libraries (eLib) Programme. The project is administered through the Consortium of University Research Libraries (CURL) with lead sites based at the Universities of Cambridge, Leeds and Oxford.

Cedars is a project that aims to address strategic, methodological and practical issues relating to digital preservation (Day 1998a). A key outcome of the project will be to improve awareness of digital preservation issues, especially within the UK higher education sector. Attempts will be made to identify and disseminate:

3. Strategies for collection management;
4. Strategies for long-term preservation.

These strategies will need to be appropriate to a variety of resources in library collections. The project will also include the development of demonstrators to test the technical and organisational feasibility of the chosen preservation strategies. One strand of this work relates to the identification of preservation metadata and a metadata implementation that can be tested in the demonstrators.

### **1.3 Digital imaging technology and preservation**

The use of digital imaging technology by libraries, archives and museums is largely concerned with the creation of digital surrogates of analogue material and tends to be motivated by two interrelated concerns:

5. Improving access to information;
6. Preservation - ensuring continuing access to this information.

With analogue materials there is a potential conflict of interest between improving access and ensuring long-term preservation. The use of digital imaging technology to create surrogates of analogue material, however, broadens the preservation strategies that can be adopted by custodial organisations (Weber, H. and Dörr 1997). Once an information object has been digitally imaged - assuming that certain minimum quality standards apply - it would be possible to output the image on a preservation-quality computer output microfilm (COM) for long-term retention while at the same time maintaining digital versions of the same for access. In this way, custodial organisations can separate the medium used for preservation from the media used for production and use (Kenney and Conway 1994, p. 19).

Even so, it is likely that some custodial organisations will develop collection management strategies that recommend the long-term retention of some digitally imaged material. Once digitised images have been created and there is a specific requirement for their long-term preservation, much the same preservation considerations will apply to them as to information objects that are 'born digital'.

The Cedars project will investigate some of the issues of preserving this type of material. The project demonstrators will include digitally imaged information objects like the journals made available through the ILEJ (Internet Library of Early Journals) project (Jupp 1997) and digital images of manuscript fragments held in the Taylor-Schechter Genizah Collection at Cambridge University Library (Taylor-Schechter Genizah Research Unit 1998). The Cedars project scope, however, is mostly concerned with digital resources that are 'born digital'. With these, interpretation of the resource is often fixed to its existence as a digital object and a human-readable surrogate will not always be adequate to express this. This is why the digital preservation of information created digitally is extremely important.

## **2. Digital preservation and metadata**

### **2.1 Preservation metadata for digital preservation strategies**

Digital preservation is as much a strategic problem as a technical one. For this reason it is imperative that the strategic context for the creation and preservation of digital resources be taken into account. This process has been eased by the appearance of the UK Arts and

Humanities Data Service (AHDS) report that outlines a policy framework applicable to the three main stages in the life cycle of a digital resource: creation, management/preservation and use (Beagrie and Greenstein 1998). Solving the technical issues of digital preservation will be important but is essentially subordinate to these wider, strategic, considerations.

The main technical problems of digital preservation relate to inadequate media longevity, rapid hardware obsolescence and dependencies on particular software products. There are currently three main approaches to digital preservation (Ross 1997). The metadata issues raised will differ according to which particular strategy is adopted but it should be noted that metadata strategies have an important part to play in all three.

### **2.1.1 Technology preservation**

This approach proposes that digital data should be preserved on a stable medium (and 'refreshed' or copied to new media as necessary) and associated with preserved copies of the original application software, the operating system that this would normally run under and the relevant hardware platform. This strategy may have some value for particularly important (or historic) examples of software or hardware or could be useful for the museum community (Swade 1992) but in the long-term is likely to be expensive and impractical. Tony Hendley (1998, p. 17) comments that the technology preservation approach "cannot be regarded as viable for anything other than the short to medium term". He also comments that it could be used as "a relatively desperate measure in cases where valuable digital resources cannot be converted into hardware and/or software independent formats and migrated forwards".

### **2.1.2 Emulation**

The second main suggested approach to digital preservation is technology emulation. This strategy relies (as with technology preservation) on the preservation of the original data in its original format. Instead of preserving the host software and hardware, software engineers would build emulator programs that would mimic the behaviour of obsolete hardware platforms and emulate the relevant operating system (Rothenberg 1995). In practice, data could be encapsulated together with the application software used to create it and a description of the required hardware and software environment. To facilitate future use, Jeff Rothenberg (1996) suggests attaching 'annotation metadata' to the surface of each encapsulation which would both "explain how to decode the obsolete records contained inside the encapsulation and to provide whatever contextual information is desired about these records".

Emulation is an important strategy that has potential applications where the look and feel of an original digital resource is of importance but where it is not worth investing in expensive technology preservation. Hendley (1998, p. 18), however cautions against relying solely on this approach and comments that collection managers "would be depending on the technical ability of the software engineers to emulate a specific environment and sustain it".

A related approach is the Digital Rosetta Stone (DRS) model developed by Steven Robertson of the United States Air Force (Heminger and Robertson 1998). In this model, digital documents would be maintained in their original file formats. In conjunction with this, a 'metaknowledge archive' (MKA) would be created to store the "the vast amounts of knowledge needed to recover digital data from a superseded media and to reconstruct digital documents from their original formats" (Robertson 1996, p. 23). The MKA would be a collection of the knowledge and processes necessary to recover and reconstruct digital documents maintained in their original file formats. This data would be used to re-create (or emulate) the hardware and software necessary to recover data from obsolete storage media and reconstruct the digital documents. The DRS model, like technology preservation

strategies, might have an application as a backup strategy where other approaches have failed. As a long-term solution to digital preservation, however, it is likely to be expensive.

### 2.1.3 Migration

A third approach to digital preservation is the periodic migration of digital information from one hardware or software environment to another. The Task Force on the Archiving of Digital Information (1996) has produced a good (much cited) definition:

*Migration is the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology.*

The point of migration is to transfer to new formats while, where possible, preserving the integrity of the information. A digital archive could convert incoming digital objects into a small number of 'standard' formats. For example textual data could be stored in a relatively software independent format like ASCII, in widely used proprietary formats like the Portable Document Format (PDF) or in formats based on applications of SGML (Coleman and Willis 1997). Over time, data would be copied and refreshed as necessary and periodically migrated to new formats for use with new generations of hardware and software.

Metadata has an important role in any successful migration strategy. Such a strategy will depend upon metadata being created to record the migration history of a digital object. In addition there is a need for contextual information to be recorded (and preserved) so that a future user can understand the technological environment in which a particular digital object was created. David Bearman (1994, p. 302) says that "content, structure and context information must be linked to software functionality that preserves their executable connections or representations of their relations must enable humans to reconstruct the relations that pertained in the original software environment".

## 2.2 Metadata and authentication

In addition to the purely technical problems of digital preservation, there will be a need to address problems of what Peter Graham (1994) calls intellectual preservation. How will users know that the digital object that they retrieve is the one that they want? How will administrators of digital repositories know that their holdings have not been subject to unauthorised changes, either accidental or deliberate?

The use of persistent and unique digital identifiers has a potentially important role in this regard. New identifiers would need to be assigned each time a particular digital object is updated or migrated. Current digital identifier initiatives include the Uniform Resource Name (URN) which is being developed for the Internet community by working groups of the Internet Engineering Task Force (Sollins and Masinter 1994) and the Digital Object Identifier (DOI), an initiative of the Association of American Publishers (e.g. Bide 1998). Legacy identifiers will also continue to be used for some of the digital objects that will need preservation, so - for example - some publishers could assign International Standard Book Numbers (ISBNs) to CD-ROMs or generate Serial Item and Contribution Numbers (SICIs) for online journal articles. On the other hand, other items in the project scope, electronic ephemera like Web pages or example, are unlikely to be assigned digital identifiers except for Uniform Resource Locators (URLs).



An additional approach to ensuring the authenticity of a given digital object would be to use simple cryptographic techniques like the production of a validation key value or checksum for each resource in a digital archive. An authentication checksum could be computed from each resource in a digital archive and stored with the descriptive metadata. When a user, or the archive, wants to retrieve the resource at a later date this checksum could be computed again and compared with the checksum recorded in the metadata. If the two agree there can be confidence that the document retrieved is the one referred to by the descriptive metadata. This general approach has been adopted for use by the European Telematics for Libraries project BIBLINK (Peacock and Powell 1998). Other possible approaches to the problem could use other cryptographic techniques like digital signatures. The World Wide Web Consortium (W3C) Digital Signature Working Group (DSig), for example, has developed digital signatures - currently implemented in PICS (Platform for Internet Content Selection) technology - for making assertions about particular Web information resources.

It is worth noting in addition that archivists and records managers share these professional concerns with preserving the authenticity and integrity of digital objects (e.g. Duff 1995; Duranti and MacNeil 1995). The University of Pittsburgh Electronic Records Project, for example, has defined a metadata model for business-acceptable communications that emphasises the preservation of a record's 'evidentiality' (Bearman and Sochats 1996).

### **2.3. Metadata and rights management**

Solving rights management issues will be vital in any digital preservation programme. Typically, custodial organisations do not have physical custody of digital objects created or made available by other stakeholders (e.g. authors or publishers). Instead they will negotiate rights to this information for a specific period of time. Permissions to preserve digital information objects will also need to be negotiated with rights holders and any such agreement may, or may not, permit end user access. A digital archive will have to collect and store any relevant rights management information which could be stored as part of the descriptive metadata. This could also be used to manage access.

The Cedars project has a Content Issues Working Group that will negotiate with rights holders for the use of material in the demonstrators. This group will also examine broader issues relating to long term storage and access to copyright material and work closely with rights owners to recommend best practice.

### **2.4 Other metadata issues**

#### **2.4.1 Resource discovery**

There will be no point preserving large amounts of digital information unless there is some consideration of resource discovery issues. Information objects that have been digitally preserved will need descriptive metadata that can aid resource discovery or ideally that can interact with other resource discovery systems, including existing library catalogues.

Recommendations on relevant resource discovery formats (e.g. Dublin Core, MARC) and metadata frameworks like the Resource Description Format (RDF) will constitute an important part of Cedars work on metadata.

#### **2.4.2 Collection Management**

There will be no need to preserve all existing digital resources, as not all will be worthy of long-term preservation. The Cedars project is interested in helping to develop suitable collection management policies for research libraries. This work could build on work carried out on selection criteria for Internet subject gateways produced by the EU funded DESIRE

project (Hofman, et al. 1997). The existence, or otherwise, of appropriate metadata for preservation, resource discovery and other purposes will be essential to allow appropriate decisions to be made about what items need to be included in digital collections and how these should be administered.

### 2.4.3 Metadata management

Another important series of issues relate to the management and migration of any proposed preservation metadata system. For example, metadata can either be stored in a database and linked (in some way) to the original resource or embedded in (or otherwise directly associated with) the original resource. Resource discovery and rights management metadata could form part of a searchable database that gives access to digital objects, while metadata specifying the technical formats used, the migration strategies operated and a document's use history could be stored closer to the document itself. Over time, this metadata will itself have to be subject to migration and authentication strategies.

## 3. Current initiatives and data models

The Cedars Access Issues Working Group has produced a preliminary study of preservation metadata and the issues that surround it (Day 1998b). This study describes some digital preservation initiatives and models with relation to the Cedars project and will be used as a basis for the development of a preservation metadata implementation in the project. The remainder of this paper will describe some of the metadata approaches found in these initiatives.

### 3.1 The RLG Working Group on Preservation Issues of Metadata

The Research Libraries Group constituted a Working Group on the Preservation Issues of Metadata in May 1997 and its final report (RLG Working Group 1998) is perhaps the best current assessment of the preservation and metadata requirements of digital imaging technology. The working group limited itself to a consideration of the data elements that describe digital image files, arguing that other specialist groups could be constituted to analyse other formats when the need becomes more pressing. The group also examined two 'core' metadata formats, the Dublin Core and the Program for Co-operative Cataloging's USMARC-based core record standard, so that the group could specify the metadata elements extra to these core element lists that would be important to serve preservation needs. The sixteen metadata elements deemed crucial for the continued viability of a digital master file were:

Element Brief	Description
Date	Date file is created.
Transcriber	Name of agency (or individual) responsible for transcribing the metadata.
Producer	Agency (or individual) responsible for the physical creation of the file.
Capture Device	Make and model of digital camera or scanner.
Capture Details	1. Name of scanner software, version information, scanner settings, gamma correction, etc. 2. Digital camera lens type, focal length, light source type, etc.
Change History	A record of modifications made to the file.
Validation Key	A mechanism allowing one to verify that the electronically transmitted file is what it purports to be.
Encryption	The technique by which data is encryption before transmission.
Watermark	Indicates whether (or not) some bits in the file have been altered in order to create a digital fingerprint or similar.

Resolution	Resolution determined by pixel dimensions, pixels per inch or dots per inch.
Compression	Indicate whether (or not) file has been conversed.
Source	Physical characteristics of the source, etc.
Colour	Pixel depth.
Colour Management	Systems (if any) used to improve consistency of colour.
Colour Bar/Gray Scale	Indicates presence (or not) of either, with type.
Control Targets	Information about targets included in scanned file.

The RLG Working Group also published three potential implementations of these metadata elements for discussion and experimentation: firstly a hypothetical Dublin Core record, secondly a mapping of the elements to USMARC and thirdly a simple XML implementation. The RLG working group report gives a useful indication of some of the individual metadata elements that need to be captured to help ensure some degree of digital preservation. The report encourages institutions to implement the RLG element set and to share their efforts with the rest of the community.

Other relevant metadata elements for are identified in the Making of America II Testbed Project White Paper, which include elements recorded at the point of capture for a digital master images, context metadata and rights management information (Making of America II 1988).

### 3.2 Open Archival Information System (OAIS)

The Cedars study suggests that there might be some value in adopting (or adapting) relevant metadata models. The most important existing model is the Reference Model for an Open Archival Information System (OAIS) published by the Consultative Committee for Space Data Systems (CCSDS). OAIS is an ISO initiative (co-ordinated by the CCSDS) that defines a high-level reference model for archives originally concerned with the long-term preservation of digital information obtained from observations of terrestrial and space environments but which would be applicable to other long-term digital archives. An archive (in OAIS terms) consists of "an organisation of people and systems, that has accepted the responsibility to preserve information and make it available for one or more designated communities" (CCSDS 1998). The OAIS model has a 'taxonomy of archival information object classes' (CCSDS 1998, pp. 50-57) that includes:

#### Content Information

This is the information that is the primary object of preservation. This contains the primary Digital Object and Representation Information needed to transform this object into meaningful information.

#### Preservation Description Information

This would include any information necessary to adequately preserve the Content Information with which it: is associated. It includes:

**Reference Information** - (e.g. identifiers),

**Context Information** (e.g. subject classifications),

**Provenance Information** (e.g. copyright)

**Fixity Information** (that documents the authentication mechanisms).

### **Packaging Information**

The information that binds and relates the components of : a package into an identifiable entity on a specific media.

### **Descriptive Information**

The information that allows the creation of **Access Aids - Information**: to help locate, analyse, retrieve or order information from an OAIS.

This taxonomy includes (and refines) many of the metadata types discussed in the Cedars report. Any high-level architecture developed for Cedars will probably conform to the OAIS model.

### **3.3 National Library of Australia PANDORA logical data model**

A separate model is the 'logical data model' developed by the National Library of Australia for its Preserving and Accessing Networked Documentary Resources of Australia (PANDORA) project (National Library of Australia 1997). This model is based on an entity-relationship diagram that identifies the logical entities that need to be supported by the PANDORA system. The highest level entities are:

1. Identification
2. Selection and negotiation
3. Capture
4. Preservation
5. Rights Management and Access Control

Each of these is divided into further entities and each of these into metadata attributes. Preservation metadata is defined as "entities required to support the management of copies within the archive, including activities to ensure both the immediate and long term accessibility of the item". The entities include 'File' and 'File Type' (e.g. M/S Word, HTML, ASCII, JPEG, PDF, TIFF, etc.), 'Format' and 'Format Type' (e.g. Online, Diskette, CD-ROM, etc.). The notes on Format suggest that such information should be recorded at the selection stage as part of technical assessment. It also recommends that "a history trail is kept of the format of the copy at the time of archiving and any technical processing that has been conducted on the copy to ensure preservation and access".

A copy of a publication may be converted from one format to another to improve accessibility in the host environment or to help migrate whole categories of publication to a new technology base. Generally a conversion from one format to another will involve tangible formats (e.g., to transfer files from diskette to CD-R) but there may also be a requirement to convert data from a tangible to online format or vice versa. When a format is converted to another format type, a record will be maintained of the conversion process, with a link to the new format type (National Library of Australia 1997).

### **3.4 Resource Description Framework**

The Cedars project will not just be adopting (or adapting) a high-level data model like OAIS. It will attempt to develop demonstrators that will implement selected aspects of digital preservation including those related to metadata. The precise nature of the metadata implementation has yet to be decided by the project but the Resource Description Framework (RDF) being developed under the auspices of the World Wide Web Consortium (W3C) is of potential interest. RDF provides a data model for describing resources and proposes an Extensible Markup Language (XML) based syntax based on this data model (World Wide Web Consortium 1998). The need to aggregate multiple sets of metadata was noted at the second Dublin Core workshop and was the principle that underlay the

formulation of the Warwick Framework container architecture (Lagoze, Lynch and Daniel 1996; Weibel and Lagoze 1997). Similarly, RDF aims to facilitate modular interoperability among different metadata element sets by creating what Eric Miller (1998) calls "an infrastructure that will support the combination of distributed attribute registries". The modular principle of RDF means that Cedars-defined preservation metadata elements could be aggregated with metadata types defined for other purposes, e.g. Dublin Core for simple resource discovery or structured data about terms and conditions. This type of interoperability is likely to be a useful aspect of preservation metadata systems.

#### 4. Conclusions

The definition and implementation of preservation metadata systems is going to be an important part of the work of custodial organisations in the digital environment. Projects like Cedars are attempting to investigate some of the relevant issues and provide some demonstrator systems that can test them. Individuals and organisations interested in the long-term preservation of digital information need to note of preservation metadata issues. The future of our digital collections will depend, to some extent, on how carefully we respond to this challenge.

#### BIBLIOGRAPHY

Beagrie, N. and Greenstein, D., 1998, A Strategic Policy Framework for Creating and Preserving Digital Collections. London: Arts and Humanities Data Service, 14 July. Available at <http://ahds.ac.uk/manage/framework.htm>

Bearman, D., 1994, Electronic evidence: strategies for managing records in contemporary organisations. Pittsburgh, Penn: Archives and Museum Informatics.

Bearman, D. and Sochats, K., 1996, Metadata requirements for evidence. Pittsburgh, Penn.: University of Pittsburgh, School of Information Science. Available at <http://www.lis.pitt.edu/~nhprc/Bacartic.html>

Bide, M., 1998, In search of the Unicorn: the Digital Object Identifier from a user perspective, rev. ed. British National Bibliography Research Fund Report 89. London: Book Industry Communication

Coleman, J. and Willis, D., 1997, SGML as a framework for digital preservation and access. Washington, D.C.: Commission on Preservation and Access.

Consultative Committee for Space Data Systems, 1998, Reference Model for an Open Archival Information System (OAIS), ed. L. Reich and D. Sawyer. CCSDS 650.0-W-4.0. White Book, Issue 4, 17 September. Latest version available from: <http://ssaoo.gsfc.nasa.gov/nost/isoas/refmodel.html>

Day, M.W., 1998a, CEDARS, digital preservation and metadata. In: Sixth DELOS Workshop: Preservation of Digital Information, Tomar, Portugal, 17-19 June 1998. ERCIM-98-W003. Le Chesnay: European Research Consortium for Informatics and Mathematics, pp. 53-58. Available at <http://www.ercim.org/publication/ws-proceedings/DELOS6/>

Day, M.W., 1998b, Metadata for Preservation. CEDARS Project Document AIW01.

Dempsey, L. and Heery, R., 1998, Metadata: a current view of practice and issues. *Journal of Documentation*, 54 (2), pp. 145-172. Available at <http://www.ukoln.ac.uk/metadata/cedars/AIW01.html>

Duff, W., 1995, Ensuring the preservation of reliable evidence: a research project funded by the NHPRC. *Archivaria*, 42, pp. 28-45.

Duranti, L. and MacNeil, H., 1995, The protection of the integrity of electronic records: an overview of the UBC-MAS Research Project. *Archivaria*, 42, pp. 46-67.

Graham, P.S., 1994, Long-term intellectual preservation. In *Digital imaging technology for preservation*, ed. N.E. Elkington, Mountain View, Calif.: Research Libraries Group, pp. 41-57.

Heery, R., Powell, A. and Day, M., 1997, Metadata. *Library and Information Briefings*, 75. London: South Bank University, Library Information Technology Centre.

Heminger, A.R. and Robertson, S.B., 1998, Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents. In: *Sixth DELOS Workshop: Preservation of Digital Information*. ERCIM-98-W003. Le Chesnay: European Research Consortium for Informatics and Mathematics, 35-43. Available at <http://www.ercim.org/publication/ws-proceedings/DELOS6/>

Hendley, T., 1998, Comparison of methods & costs of digital preservation, *British Library Research and Innovation Report*, 106. London: British Library and Innovation Centre.

Hofman, P., Worsfold, E., Hiom, D., Day, M. and Oehler, A., 1997, Specification for resource description methods: 2, Selection criteria for quality controlled information gateways. *DESIRE: Development of a European Service for Information on Research and Education, Deliverable 3.2 (2)*. Available at <http://www.ukoln.ac.uk/metadata/desire/quality/>

Jupp, B., 1998, The Internet Library of Early Journals. *Aslib Proceedings*, 49 (6), 153-158.

Kenney, A. and Conway, P., 1994, From analog to digital: extending the preservation tool kit. In *Digital imaging technology for preservation*, ed. N.E. Elkington. Mountain View, Calif.: Research Libraries Group, pp. 11-24. Available at <http://www.bodley.ox.ac.uk/ilej/papers/paper01.htm>

Lagoze, C., Lynch, C.A. and Daniel, R., 1996, The Warwick Framework: a container architecture for aggregating sets of metadata. *Cornell Computer Science Technical Report TR96-1593*. Available at <http://cs-tr.cs.cornell.edu/Dienst/UI/1.0/Display/nctr1.cornell/TR96-1593/>

Making of America II, 1998, The Making of America II testbed project white paper. Version 1.03, March 16. Available at <http://cs-tr.cs.cornell.edu/Dienst/UI/1.0/Display/nctr1.cornell/TR96-1593/>

Miller, E., 1998, An introduction to the Resource Description Framework. *D-Lib Magazine*, May. Available at <http://www.dlib.org/dlib/may98/miller/05miller.htm>

National Library of Australia, 1997, PANDORA Logical Data Model, Version 2, 10 November. Available at <http://www.nla.gov.au/pandora/ldmv2.html>

Peacock, I. and Powell, A., 1998, BIBLINK. Checksum - an MD5 message digest for Web pages. *Ariadne (Web version)*, no. 17, September. Available at <http://www.ariadne.ac.uk/issueit/biblink/>

RLG Working Group on Preservation Issues of Metadata, 1998, Final report. Mountain View, Calif.: Research Libraries Group, May. Available at <http://www.rlg.org/preserv/presmeta.html>

Robertson, S.B., 1996, Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents. Thesis (MSc), Air Force Institute of Technology, Graduate School of Logistics and Acquisition Management. Available at <http://www.au.af.mil/au/database/research/ay1996/afitla/robersb.htm>

Ross, S., 1997, Consensus, communication and collaboration: fostering multidisciplinary co-operation in electronic records. In: Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996. INSAR: European Archives News, Supplement II. Luxembourg: Office for Official Publications of the European Communities, pp. 330-336.

Rothenberg, J., 1995, Ensuring the longevity of digital documents. *Scientific American*, 272 (1), pp. 24-29.

Rothenberg, J., 1996, Metadata to support data quality and longevity. Proceedings of the 1st IEEE Metadata Conference, NOAA Complex, Silver Spring, Md., 16-18 April. Available at <http://www.computer.org/conferen/meta96/rothenbergpaper/ieee.data-quality.html>

Rust, G., 1998, Metadata: the right approach. An integrated model for descriptive and rights metadata in e-commerce. *D-Lib Magazine*, July/August. Available at <http://www.dlib.org/dlib/july98/rust/07rust.htm>

Sollins, K. and Masinter, L., 1994, Functional Requirements for Uniform Resource Names. RFC 1737. Available at <http://ds.internic.net/rfc/rfc1737.txt>

Swade, D., 1992, The problems of software conservation. *History and Computing*, 4 (2). Available at [http://www.cs.man.ac.uk/ccs/simulate/sim home.htm](http://www.cs.man.ac.uk/ccs/simulate/sim%20home.htm)

Task Force on the Archiving of Digital Information, 1996, Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group. Washington, D.C.: Commission on Preservation and Access. Available at <http://www.rlg.org/ArchTF/>

Taylor-Schechter Genizah Research Unit, 1998, The Genizah On-Line Database (GOLD). Cambridge: Cambridge University Library. Available at <http://www.lib.cam.ac.uk/Taylor-Schechter/GOLD/>

Weber, H. and Dörr, M., 1997, Digitisation as a method of preservation? Amsterdam: European Commission on Preservation and Access.

Weibel, S. and Hakala, J., 1998, DC-5: The Helsinki Metadata Workshop: a report on the workshop and subsequent developments. *D-Lib Magazine*, February. Available at <http://www.dlib.org/dlib/february98/02weibel.htm>

Weibel, S.L. and Lagoze, C., 1997, An element set to support resource discovery: the state of the Dublin Core, January 1997. *International Journal on Digital Libraries*, 1(2), pp. 176-186.

World Wide Web Consortium, 1998, Resource Description Framework (RDF) model and syntax specification, eds. O. Lassila and R. Swick. W3C Working Draft. Available at <http://www.w3.org/TR/WD-rdf-syntax/>

## 6. Acknowledgements

UKOLN is funded by the British Library Research and Innovation Centre (BLRIC), the Joint Information Systems Committee (JISC) of the UK higher education councils, as well as by project funding from several sources. UKOLN also receives support from the University of Bath, where it is based. The views expressed in this paper do not necessarily reflect those of the Cedars project, UKOLN or their funding bodies.

The author would like to thank Kelly Russell (Cedars Project Manager) and Dr Mark Nicholls (Deputy Keeper of Manuscripts, Cambridge University Library) for comments on an earlier draft of this paper.

For more information on the Cedars project, please contact:

Kelly Russell,  
Cedars Project Manager,  
Edward Boyle Library,  
University of Leeds,  
Leeds,  
LS2 9JT.



**The Role of National Initiatives in Digital Preservation**  
**Margaret Hedstrom**  
**Associate Professor of Information, University of Michigan**

The digital world is often characterised as chaotic, and the terrain of digital preservation is no exception. In recent years the research community has witnessed a flurry of activity around digital archiving and digitisation for preservation and improved access to materials in traditional formats. What is less clear is whether the time, energy, and money invested in digital preservation initiatives will produce strategies that can be generalised and, at the same time, are flexible and robust enough to satisfy the specific local and institutional needs. This article uses digital preservation initiatives in the United Kingdom, Sweden, and Australia to illustrate different models for national leadership and co-ordination.

Digital preservation is a global problem in a world where global communications transcend national boundaries and where hardware and software are sold on an international market. Numerous reports emphasise the development and adoption of international standards as an underpinning of long-term digital preservation strategies. Sharing research, training, and best practices is also a good way to avoid reinventing the wheel and to spread the costs of research and development. Some may wonder whether national strategies are needed in this type of environment and, if so, what useful role they might play. The following three examples illustrate areas where co-ordination and resource sharing on a national level can mediate between the particulars of local situations and global initiatives without creating isolated islands of uncoordinated activities.

#### **Digital Programs in the UK**

Digital preservation activities in the UK are far too extensive even to list in a short article such as this. The National Preservation Office based at the British Library is developing a national strategy for digital preservation that includes assembling best practices; developing scaleable and costed models for converting, accessing, and preserving digital information; identifying areas needing research and evaluation; co-ordinating UK initiatives in these areas; exploring funding models; and mobilising resources. In the UK, sectoral co-ordination within higher education and in disciplinary areas will support development of a national strategy.

The CEDARS Project is a higher education initiative funded through the eLib Programme of the Joint Information Systems Committee (JISC). CEDARS stands for "CURL (Consortium of University Research Libraries) Exemplars in Digital Archiving." The project is sponsoring practical demonstration projects and strategic working groups on Access Issues, Content Issues, Digital Preservation, and Infrastructure and Integration. The project's Access Issues Working Group has produced a useful summary of models for preservation metadata.

In July 1998, the Arts and Humanities Data Service released "Strategic Policy Framework for Creating and Preserving Digital Collections." The report, prepared by Neil Beagrie and Daniel Greenstein, includes recommendations on digital preservation, standards, policy, and future research based on case studies of real life concerns of data bank managers, digitizers, funding agencies, institutional archives, academic data archives, and legal deposit libraries. This project, also funded by the JISC eLib Programme, is part of a series of studies guided by the NPO's Digital Archiving Working Group, composed of members from UK Higher Education Libraries, Data Centres and Services, the British Library, the National Preservation Office, the Research Libraries Group, and the Publishers' Association.

### **Sectoral Leadership and Coordination in Sweden**

Sweden, which does not have a national preservation office or a national preservation strategy, relies heavily on national institutions in the library, archives, and museum sectors for policy and guidance. The Royal Library is especially active in metadata standards and digital preservation strategies for electronic publications. The Royal Library is sponsoring a project to implement legal deposit legislation for Swedish electronic publications and Web pages. Under the rubric of "Kulturarw3", the Swedish Web Archiving Project is using robots to identify and download all Swedish Web pages. The project has the potential to develop methods that are scaleable, because they use software agents to identify, download, and classify Web pages while working within the focused domain of Swedish Web sites. In August 1998, the National Libraries of the Netherlands and Australia joined with the Swedish Royal Library to form a project called "PreWeb" to sponsor discussions and information sharing about preserving World Wide Web documents. The Swedish National Archives coordinates activities in the archival sector where there is an extensive project underway at SVAR, a technical and advisory unit of the National Archives, to digitize church registries and other genealogical information. There is less coordination among museums, although INSAM, headquartered at the Nordic Museum, is providing leadership in the development of common standards, policies, and training for museum databases and other technology applications.

Digital preservation issues are serving as a catalyst for joint activities and co-ordination among the archives, library, and museum communities in Sweden. In May and June 1998, this author taught a five-day digital preservation course in three different Swedish cities, organised by the ABM (Archives, Bibliotek, Museum) group. The workshop created opportunities for professionals from each of these sectors to meet and discuss digital preservation problems and to consider shared solutions across institutions that traditionally have been identified with the types of materials they collect. Nearly 100 archivists, librarians, and museum curators participated in the Spring training program, and many resolved to use the training as a basis for future collaboration.

### **Distributed Responsibilities in Australia**

Many professionals are looking to Australia for creative ideas and bold strategies for digital preservation and access. The Australians have attracted considerable attention in the archival community with their policy on electronic records management as a shared responsibility and the release of the first records management standard (AS 4390) in December 1996. Australian electronic records management strategies assume that digital archives will be distributed and that the creators of digital materials and the archives must work together to develop technologically feasible and cost effective means to preserve corporate memory in the digital age.

The National Library's PANDORA Project is capturing, archiving, and providing long-term access to significant Australian on-line publications. The Selection Committee on On-line Australian Publications (SCOAP) has developed a set of selection guidelines. PANDORA currently provides access to more than 1,000 Australian electronic journals, magazines, and ephemeral digital items. Like the policies being adopted for preservation of official electronic records, the Australian national strategy for long-term access to electronic publications assumes co-operative arrangements among creators, publishers, libraries, and archives. To facilitate co-ordination and reinforce the linkages between preservation and access, the National Library recently merged the former National Preservation Office, the Distributed National Collections Office, and its International Relations Office, and formed National Initiatives and Co-ordination. Like similar initiatives elsewhere, the strategies

include development of standards, use of metadata, establishing criteria and priorities, and distributing responsibility for preservation.

### **Conclusion**

Returning to the original question of whether there is a role for national preservation strategies in a global digital environment, these three examples demonstrate that national policies and strategies remain useful, but that each national strategy may well adopt quite different approaches for infrastructure development, coordination, and funding. Dividing responsibility for preservation and developing guidelines and criteria for selection are areas where coordination on a national level seems particularly pertinent. Australia, Sweden, and the UK all have legal deposit legislation, but each country approaches legal deposit of electronic publications differently. The Swedish strategy is designed to acquire digital copies of all Swedish electronic publications, including Web pages. The current emphasis is on capture because of legal questions about access and a sense of urgency about acquiring copies of electronic publications before they vanish. The Australians are pursuing a strategy based on distributing responsibility between the National Library for electronic publications of national significance and state and local authorities for significant publications in their regions. Both of these approaches have merit, but what is more significant is that they begin to demarcate areas of responsibility for preservation that allow other institutions to turn their attention to neglected issues. Variation in national legislation governing copyright, legal deposit, access, and privacy also requires specific national strategies that conform to current law and legal traditions.

Leadership from national institutions, advocacy organisations, and funding bodies can direct efforts toward solving classes of problems, as many of the JISC-funded eLib projects illustrate. Co-ordination is also an essential element in an environment where digital materials are distributed and responsibility is shared. These examples illustrate how co-ordination between sectors, such as the linkages between higher education and discipline-based data centres in the UK, or among libraries, archives, and museums in Sweden, might be accomplished on a national scale. Efforts to achieve this degree of co-ordination on an international level often become too unwieldy or too distant from many of the actors and stakeholders. There is plenty of research and development to be done on an international scale in the areas of standards, software and systems development, and interoperability, but these efforts should not negate the contributions that national initiatives and strategies can make to a global infrastructure for digital preservation. In spite of the variations from country to country, there is also ample opportunity for sharing lessons and deriving general principles for national preservation strategies. At the moment, there is some cross-fertilisation between projects, but there is a ways to go before the preservation community has models for national strategies that can be modified readily to meet local conditions.



**Canadian Overview**  
**Karen Turko**  
**Head of Preservation Services**  
**University of Toronto, Canada**

My aim today is to consider some of the themes that we have been hearing in the last few days in the light of a Canadian context.

**Selection for digital imaging**

In some senses we may have an easier time of this in Canada than is the case elsewhere partly because of the work of the Canadian Institute for Historical Microreproductions (CIHM). Under the aegis of CIHM, selection for preservation has already been attempted and any digitisation efforts could certainly build on this initiative.

CIHM was established in 1978 to locate early printed Canadian materials (books, annuals and periodicals), to preserve their content on microfilm, and to make the resulting collections available to libraries and archives in Canada and abroad.

*"Spanning three and a half centuries of Canadian development, the Early Canadian microform collection contains more than sixty-five thousand titles ... from North American native peoples' studies to religion and philosophy, from genealogy and local history to French Canadian language and literature. The Institute ... has located and preserved in microformat over 80,000 Canadian books, pamphlets and periodicals documenting Canada's printed record before 1920."*<sup>(1)</sup>

While the preservation of digital files is an overriding concern in Canada, there is a good chance that any digital files of printed Canadiana would have preservation microfiche already made of the original paper edition if the item in question fell within the purview of CIHM. This would not apply to collections which fall outside the above, however. There are many rich and varied collections in the country on other subjects. An interesting example in the context of a digitisation project following on from CIHM's work is the Early Canadiana Online/Notre memoire en ligne Project.

*"The Project's objectives are to: digitise 5,000 works from CIHM's major microfiche collection ... and to make them accessible on the Internet; develop a model for comparing the costs of storing and accessing works in both traditional media and digital format; and, evaluate the use and acceptability of various digital versions in comparison to the printed and microform versions of texts. The Project will also test the long-term economic viability of a co-operative digitisation operation modelled on CIHM. ... sponsored by the Mellon Foundation to better understand the costs and benefits of digital technologies and their potential impact on libraries and scholarly communication."*<sup>(2)</sup>

The existence of these two initiatives gives us a starting point in selection for a digital initiative of printed Canadiana at any rate. Another fledgling initiative comes from the Canadian Initiative on Digital Libraries<sup>(3)</sup> (mentioned below) which is putting forward a decentralised project of scanning local history materials to develop a multimedia national Digital Canadian Histories collection. Where more work is required is in assessing other kinds of collections in the country and attempting to prioritise or make a checklist of these materials for potential digital projects.

The Bibliotheque nationale du Quebec has digitized 360,000 pages of Quebecoise monographs and 20,000 documents, prints and posters and put these on the Web in PDF format.<sup>(4)</sup>

The University of Toronto Library has for several years been involved in a digital project in partnership with a Japanese Company, Maruzen International. This project is called UT Back in Print,<sup>(5)</sup> and it supplies paper copies of books which are out of print to customers in Japan. Orders for paper copies of the books are placed with University of Toronto Library by Maruzen, the University Press clears the copyright and the books are then scanned. At this point we have scanned about 6,000 books, mainly English language materials, and while the decision to scan is customer driven, many of the items fall within the period of paper embrittlement and within broad subject areas. The Library retains the electronic file to add to the growing digital library at the institution. This service is also available to other customers upon request.

The University of Toronto Library is also involved in several other digital projects. There is a Brittle Book Programme to scan items which are deemed to be in danger of deterioration, and which used to be microfilmed. Other digital projects include an investigation of map scanning, an electronic theses project in conjunction with the NLC and other Canadian libraries<sup>(6)</sup> and scanned documents for a website for the G8 meetings, which has now become the official website of the summits.<sup>(7)</sup>

### **Preparation of materials for digitisation**

In Canada we do not usually do destructive reprography i.e. books are not disbound or discarded after scanning or microfilming. This makes it rather more difficult and expensive when choosing and preparing materials for scanning but I do not see this policy being changed in the near future. For example, it was a considerable challenge to work with the scanning equipment we are using in the University of Toronto Back in Print Project to scan unbound books. This necessitated many months of experiment on the part of the Library staff and the vendor.

### **Digital Image Capture**

As part of the ECO Project I had occasion to put out an RFP for scanning the CIHM microfiche. After a fairly extensive testing procedure, I concluded that there was at that time no vendor in Canada who could scan to the standard required for the Project and the contract was awarded to Preservation Resources, Bethlehem, PA. The criteria used to evaluate proposals included resolution, metadata and communication. In all of these respects Preservation Resources was clearly far superior. I would say that we definitely have to work to bring Canadian vendors up to standard. This is a process that we have gone through before, for example, with microfilming. During a co-operative microfilming project from 1991-1994, again funded by the Mellon Foundation, we worked with vendors to encourage them to meet ANSI standards in microfilming and were successful with several. The same process now needs to be undertaken with scanning vendors. In this respect conferences like the one taking place now, which draw together experts from various countries to agree on best practices on such matters as digital image capture and the other subjects under discussion in the last few days, will be very useful in dealing with vendors. RLG has been very helpful in this regard in the past, notably in drawing up technical manuals which can be used in dealing with vendors, and also within our institutions in justifying policies and procedures.

For the Back in Print Project at the University of Toronto Library we embarked on an in-house scanning programme using Xerox XDOD equipment. The books are scanned at 600 dpi, 8 bits of greyscale. The TIFF files produced are proprietary and we are therefore experiencing the same problems other users of this equipment have in making these files available. We are hoping to add an overhead scanner which can also handle colour in the near future, so as to be able to scan more fragile books, but in general in this project we are

grappling with the same problems of access and copyright as everyone else involved in this field.

### Metadata

A fair amount of work has already been done in Canada in the area of metadata. The RLG Working Group on Preservation and Reformatting Information had two Canadian members. The report has been posted on the Web in Canada and many Canadian libraries have been considering it and commenting on the substance. In addition, under the aegis of a programme called "Schoolnet"<sup>(8)</sup> a metadata survey has been distributed. Schoolnet is a multimedia portrait of Canada produced by young people under contract to Industry Canada's Schoolnet Digital Collections. It displays over 100 collections from the holdings of Canadian archives, libraries, museums, associations, businesses, labour unions and other organisations. In the ECO project metadata is being taken from the MARC record. The National Library of Canada has been keeping abreast of international developments and emerging standards and there seem to be two points of interest, Dublin Core and GILS. The latter has been used in a pilot for a government wide metadata approach for the federal government in the Royal Commission on National Development in the Arts, Letters and Sciences 1949-1951. This is a large full text project of 15 briefs.<sup>(9)</sup> As well as the Dublin core and GILS approach, the NLC has been creating other MARC compliant metadata for use in other projects using their own data models. A challenge that we do have in Canada is the whole issue of bilingual metadata. This often involves a lot of work, often manual, in recreating key words on the input side, and also on the retrieval side we have to think of such issues as retrieving accented and non accented characters. When we use a French word with or without the accent, we have to ask if we are retrieving everything or only some of the data, for instance. Thus it is fair to say that there is an awareness of metadata concerns in the country and I feel this topic will be dealt with in the same manner in which we follow international cataloguing standards. Conferences such as this one certainly help in achieving consensus and moving efforts forward internationally. Perhaps this is another area where RLG can play a major role.

### Digital Preservation

Of course in this regard more can always be done and Canada's track record in the digital archiving arena is no better than that of any other country. Much data produced in the early days of electronic files is already lost, and there is no clear consensus as to just what has been produced within the country or indeed within an institution. An initiative has been launched recently, the Canadian Initiative on Digital Libraries, which may in part remedy some of these problems. "The Canadian Initiative on Digital Libraries" was established to:

- Formulate and implement strategies towards increased communication and education on digital library matters
- Promulgate digital library standards and best practices
- Define methods to improve coordination of activities and avoid duplication in the development of digital resources
- Establish strong relations with others in the information arena: creators, publishers, information technology industries, archives, museums, cultural agencies, government agencies at all levels, and individual and institutional users
- Raise awareness of Canadian digital library activities, both within and outside Canada.

One of the key areas of activity is

- Roles and responsibilities for long-term archiving<sup>(10)</sup>

The idea put forward in "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information"<sup>(11)</sup> that a national digital archive might be decentralised

might work very well in a geographically vast country like Canada with two official languages. One example of a decentralised digital archive might be at the University of Toronto Library, for instance, where we would be responsible for the archiving of not only the digital files we have created but also the 7600 electronic serial titles we have mounted in our digital library.<sup>(12)</sup> Other centres such as this could be established throughout the country. This is an idea that the Canadian Initiative on Digital Libraries is bound to consider, but the existence of the CIDL does not mean that our work is done. Much remains to be accomplished in this respect, and perhaps this area causes most concern because of issues of difficulty, leadership and expense.

In summary, in Canada we are working in every area covered by the conference and keeping abreast of and are involved in international efforts. In considering selection for digitization, a register of digitized files would be most helpful. Otherwise some coordinating efforts may come out of the CIDL effort. In preparation of materials for scanning, the culture in Canada would not tolerate other than a non destructive approach. In the area of digital image capture, historically organizations such as RLG have produced manuals and run training courses and conferences on such topics. As for metadata issues, I am confident that there will be full participation in international efforts and translation into Canadian norms. Digital preservation is perhaps the most problematic topic and needs much work and planning. Conferences such as this help greatly in drawing awareness to international efforts so that we do not waste time and resources in "reinventing the wheel" as cautioned by our keynote speaker. It is particularly important to not do this on an international level and not just in a North American context.

## REFERENCES

1. Canadian Institute for Historical Microreproductions, Home Page, available at <http://www.nlc-bnc.ca/cihm/>
2. Early Canadiana Online, available at <http://www.canadiana.org/>
3. Canadian Initiative on Digital Libraries, available at <http://www.nlc-bnc.ca/cidl/>
4. Bibliotheque nationale du Quebec, available at <http://www.bibliat.gouv.qc.ca>
5. University of Toronto Press, available at [www.utpress.utoronto.ca/backinprint.htm](http://www.utpress.utoronto.ca/backinprint.htm)
6. The Joint Electronic Thesis and Dissertation Project, available at <http://www.fis.utoronto.ca/etd/>
7. G8 Information Centre, available at <http://www.g7.utoronto.ca/>
8. Canada's SchoolNet, available at <http://www.schoolnet.ca/home/e/connect/>
9. National Library of Canada, Royal Commission on National Development in the Arts, Letters and Sciences 1949-51, available at <http://www.nlc-bnc.ca/massey/>
10. Canadian Initiative on Digital Libraries. A National Alliance, available at <http://www.nlc-bnc.ca/cidl>
11. Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. Washington DC: Commission on Preservation and Access, 1996.



12. University of Toronto, Contemporary Canadian Bibliographies, available at <http://www.library.utoronto.ca/www/news/frontpage/index.html>



**Closing International Reaction Panel Address: an Australian  
Perspective  
Colin Webb  
Manager, Information Preservation  
National Library of Australia**

**Introduction**

Here we are, almost at the end. Being invited to be part of this international reaction panel means I have felt obliged to listen to just about every word. I am glad to have had that focus, because I like what I have heard.

Of course, one of the risks of being almost the last speaker is that every gem of wisdom one might have to offer is likely to have been already said by someone else. So please bear with me if some of this sounds like the "me too" messages we all love on our email discussion lists.

Part of my role in this panel is to ask how productive we have been. I ask that question from the perspective of my own working environment, which is a national library that is part of a diverse and perhaps idiosyncratic library system, in a part of the world which is off the main cross-Atlantic tracks. It is an environment where digitisation practice and experience may be described as immature but not unsophisticated.

Anne Kenney said at the start that this was a great opportunity, because it was a gathering of many of the "stars", the key players, in preservation-related digital imaging. Being surrounded by such depth and breadth of expertise has been impressive, but I have been even more impressed by the evident willingness to listen, to look critically but generously at our individual and collective experiences, and to look for ways forward. It says a lot that this has been such a focused and hard-working meeting committed to doing its task well.

I am going to bring together some significant aspects of our experience in Australia with some of the messages I have heard over the past couple of days. I will be reflecting on why I like what I have heard, and why I think it is important. And I want to look at some issues for the way forward.

**So, what about our Australian experience?**

Tension of roles and diversity of needs - is prescription possible?

We have had a few exchanges here about our responsibilities. My view on this is that there *is* a tension between our roles as professionals within a broad community, and as professionals with responsibilities to serve a particular community's interests. But we do not need to be debilitated by it: medical professionals manage to worry about the health of the community while also focusing quite narrowly on the health of the individual sitting in the surgery chair. It is not beyond us to do likewise. We should not be ashamed of our affiliations and the responsibilities they imply - unless they are shameful. We need only be clear about whose particular axe we are grinding when we try to define what will work and what won't work.

I would imagine that as in Australia, most of us live reasonably comfortably with that tension in the preservation scene: most people have both an institutional focus and a wider professional interest.

Within national libraries we have to manage that kind of tension on a daily basis. In my environment, for example, we have our own important collections to look after, but we have a powerful leadership role in the diverse Australian documentary heritage field. We operate

with a strong system of state, university and public libraries, alongside a strong system of national, state and specialist archives.

When I say "strong" here, I am referring to a number of characteristics, often including strength of will and independence of spirit as much as strength of resources. Leadership in such an environment requires frequent negotiation and re-negotiation to achieve even collaborative goals we all agree on. It doesn't work for us to say: "you must do it this way". Even if we could say: "there is only one way to skin a cat", we could not say it and expect not to get scratched in the process!

So we are very aware that the different contexts in which we work engender different needs, and different perspectives. Prescription really doesn't work for us. People generally want something clear, but by way of advice that helps them deal with their own situational needs.

There will be some people in Australia disappointed that this conference did not issue a set of standards, but most will think it important that we recognised diversity of needs without saying: "anything you or your supplier feel like doing is okay". People are looking for are good decision-making tools. These are usually tools that lead to good decisions rather than tools that suggest you do not need to make decisions. So I guess this comes down to a strong vote for "guidance" rather than "guidelines", and for guidance that takes account of the tensions that genuinely exist in people's roles and in the diversity of needs in digitisation programs.

#### **Diversity of players**

A second characteristic of the Australian scene is the diversity of the people involved. Whatever is driving digitisation, we find our project teams include, for example, collection managers, IT visionaries, IT nuts and bolts people, designers, preservation managers, reformatting specialists, resource managers, cataloguers, reference librarians, library executives if the project is big enough. So it was important that we talked about building teams of people appropriate to the needs of the program.

This is a powerful message to take back, and it is a powerful message needing to be incorporated into any guidance we produce: we cannot afford to sit in our white lab coats and say: "this is preservation, We will do it"; or "this isn't preservation, we do not want to have anything to do with it"; or "this is the preservation bit, that's all I'm interested in"; or "this is basically about access, so you preservation people do not need to be involved".

The properly integrated nature of digitisation teams means that the guidance we give has to work for teams that may not be principally preservation driven but are certain to be multi-disciplinary, working with multiple objectives.

#### **Diversity of objectives**

I have talked about tensions, and I am particularly conscious of this in what I am going to say now in talking about objectives.

I want to start by giving a loud "me too" to what Seamus Ross has just said about project objectives and project management. While more than happy to enjoy the fruits of serendipity, experience has taught me the best way of getting what we want is to know what we want. Sometimes we do not go exactly where we want to, but a clear handle on objectives, with the occasional check on reality, is the best guarantee of keeping our sense of direction. Heaven knows we need that when we come to the crossroads and roundabouts of every project and every ongoing program!

We have recognised the reality that many - no, most - digitisation programs have diverse objectives: access, visibility, R&D, a dash of preservation. It is important for the real objectives to be teased out, declared, and thought through.

And of course we want to assert the need to be clear-minded about the preservation aspects. It will not be good enough to say generically that digitisation has preservation potential or preservation value, and leave it at that. We all know of proposals that speak vaguely of preservation outcomes as something that comes with the territory, or that do not mention preservation issues at all. If we are going to get preservation benefits, we have to decide how we are going to get them. For Australian preservation managers and digitisation managers, that message alone would be a good outcome.

But I find myself wondering whether our ability to provide guidance is constrained by the diversity of *our* objectives, and what I see as a collective ambivalence about the preservation role.

One abiding memory I will take from the past few days is of an argument that spilled across the tea break about minimum standards and what we should be allowed to call "preservation digitisation". For all the passion and certainty expressed on both sides of the argument, it seemed to encapsulate the trouble we still have, collectively, in defining the points of overlap or even contact between preservation and digitisation. The years of thought, experimentation, experience and analysis have sorted out many things, but still left us with this ambivalence.

Eventually it will be easier to provide guidance if we have sorted out where we are coming from. Are we trying to maximise a wonderful preservation tool, or trying to convince others with a different agenda to use this wonderful preservation tool? Or is it really a case of heading the Philistines off at the pass before they butcher our precious originals?

A great strength of our gathering is that we have been able to ask ourselves these questions and canvass these possibilities.

I would not want to make too much of this. We have lived with similar ambivalence and internal debate before on important preservation questions, and achieved generally good things. The road has been long, and the journey surprisingly uncompleted, on issues like standards for "permanent" paper, environmental conditions of storage, deacidification, exhibition exposure limits, acid free packaging. Maybe in all of these, as in digitisation, we seem to go round and round again because we are challenged to think about what is cost effective, rather than what is ideal, and having thought about it, we are challenged to prove it. This questioning is a positive step, even if it is a frustrating and at times confusing one.

I believe we can also live with this ambivalence regarding digitisation. In fact we *must* live with it because we are not going to resolve it overnight and yet we need to say useful and meaningful things now. For example, it is very important that we have asserted the need for preservation involvement in thinking through and managing digitisation projects, whatever their purpose.

But it pays us to be honest about where we stand. We will pay heavily for pretending to have resolved issues before we do resolve them. We are far better to admit the growing but still incomplete nature of our understanding than to claim things that will be misinterpreted and misunderstood. We are simply going to have to say on some issues that we do not know, or that we just haven't been able to agree.

As an additional twist to this "diversity" or confusion of objectives, we must also recognise that many preservation people feel poorly equipped to provide guidance. While we have talked about roles as enthusiastic proponents, cautious risk managers, or whistle-blowing referees, many of my colleagues in Australia have felt they could only choose between two other roles: sceptical spectator, or flat-earther hoping it will go away if they ignore it completely. One of the target audiences for any guidance we can give is our own profession.

### **Distance and markets**

A fourth characteristic of the antipodean scene is distance. We live in a large country with a fairly small population, which means small markets for the services we want. We simply do not have many players in the field.

The size of the marketplace produces some interesting phenomena - every one has everyone else over a barrel. We have found vendors willing to agree to poorly developed specifications in order to get work that will justify their investment, but we have also found ourselves being offered compromises all the way once the work begins. Often this has been positive, as our vendors challenge our sometimes naive specifications.

Again, this just reinforces the need for guidance that empowers people to make what can be quite difficult decisions involving trade-offs. It was important to hear Jane Williams talk about helping people recognise the likely consequences of their decisions, recognising risks and managing them consciously: "you can do this, and it may mean sacrificing that".

### **Communication - and re-inventing the wheel**

As well as giving us a small market, our large distances and small population mean that we can easily become isolated. The "tyranny of distance" is a cliché, but like most clichés it contains quite a lot of truth. For all our modern technology and travel, it is a long way to get that face to face contact we find so helpful. That probably makes it easy to misinterpret guidelines and standards as statements of absolutes, when with some more personal interaction it may have become clear that something else would have been more appropriate. This probably means that when we are looking for guidance we are looking for something more interactive and personal than a book of instructions.

This may in fact be an argument for doing something that looks very much like re-inventing the wheel. We recently published a book of preservation microfilming guidelines for Australia and New Zealand. There was no shortage of other people's manuals, standards and guidelines already on our shelves, but it was not a wasted effort. We needed something for our own conditions, and we needed something that would build ownership that was missing in our preservation community. These benefits more than made up for any inefficiencies of the re-invention process. It also gave us a chance to talk to each other and to assess what we were doing. In fact, it was one of the most useful experiences in my twenty-odd years of preservation work.

One doesn't have to re-invent the wheel, but it is a good idea to check the tread on your own tyres occasionally. Perhaps we need a model of guidance that encourages people to structure and process their own experience as well as receive the words of experts from far away.

### **The whole system**

The last characteristic I want to talk about is infrastructure building. By infrastructure I mean a range of things that make it easier, or even possible, for libraries to do their core business: in this context, managing and delivering digital information services.

It may be one of our obsessions; it is also one of our strengths. We give it a lot of attention because the National Library of Australia is expected to take a strong leadership role in making the library system work. We also focus on it because we want to maximise the value of anything we do - we want to provide frameworks in which our initiatives can work well. So we have consciously put more effort and resources into infrastructure building than into digitisation programs.

For us, infrastructure building includes creating collaborative databases that can be populated by digitisation projects. The most prominent examples have been a national Register of Archives and Manuscripts (RAAM), a database of finding aids, and a national pictorial access database we are currently exploring.<sup>(1)</sup>

It also includes tools for creating and managing metadata through our collaborative Metaweb<sup>(2)</sup> project, and information databases like our PADI website focusing on the preservation of access to digital information.<sup>(3)</sup>

Our most prominent focus has been on archiving and preservation of digital information, through our national PANDORA project, already mentioned by Michael Day and Margaret Hedstrom.<sup>(4)</sup> PANDORA represents a commitment to start, in the spirit of what I believe was Mark Twain's advice about getting ahead: the secret, he said, is to break our complex, overwhelming tasks into small manageable ones, and to start on the first one. PANDORA has been very much like that. After a period of wondering whether it was even possible to manage digital publications, we decided to look to our existing expertise and build on it. We went from selection guidelines, to a modest internal program, to a funded demonstrator project, to what is now an evolving national infrastructure for managing digital publications, involving a number of State libraries. We are starting to talk with university libraries about the role they could play in this archiving network.

This is part of our core business and we are determined to normalise it within our operations. While the PANDORA infrastructure is already quite impressive, it is incomplete. We are working right now to take major steps forward with it.

Firstly, we recognise that mainstreaming of PANDORA will be empty rhetoric if we do not embed it in an organisational and technical framework that the Library can sustain. That is why we have embarked on what we call our Digital Services Project (or DSP), which is an attempt to put together a system, or a system of systems, that will articulate with other people's systems as we manage our archived PANDORA content as well as the content from our digitisation projects, and our metadata for resource discovery and for collection management. This complex and challenging process has engaged the attention of people across the Library. An information paper should be going public for comment soon after I get back to Australia.<sup>(5)</sup> The DSP is proof of the National Library's commitment to developing an infrastructure in which our later digitisation efforts can have maximum usefulness.

This is really quite important when we are talking about guidance. Although there is a time and a place for experimental imaging projects and for building content, much of that will be either wasted, or at least more difficult to manage later on, if we are all digitising first and working out how to manage it later.

For PANDORA to become more than an interesting side-show, the second step we have to take goes beyond what I call the "archiving" arrangements that are a necessary starting point for preservation, where we "mark and park" the digital publications we mean to keep accessible. We have to make progress in addressing the software dependencies these things come with. Once again, the challenge here will include an adequate, useful definition of preservation: how much change will we accept, or be forced to accept, in the "essence" of the archived information?

We think the software dependencies will be a severe problem, especially with what we call physical format publications - CD-ROMs and floppy disks that come with a variety of complex proprietary dependencies and often with security devices meant to frustrate the very processes we need to undertake.

We have tried to establish our own agenda of research to find the best mixture of pathways through this preservation maze.<sup>(6)</sup> We will undoubtedly be leaning heavily on the experience of others. This conference has been a great opportunity to meet a number of people with experience or ideas about migration and emulation. Some have said it is easy, others that it is hard. We will be talking to you!

I want to give strong support to Margaret Hedstrom's call for effective indicators that tell us when intervention is needed, preferably before it is needed. At the level of the individual item this is fairly easy to achieve; across a whole collection of complex, heterogeneous

publications and other digital objects, possibly managed in a decentralised archive, it looks like quite a challenge.

The third direction we are taking with PANDORA and our digital management infrastructure is a communication one. We want to share as much information as we can. We use our website to tell the world what we are thinking, as well as what we are doing. We try to develop policies and strategies in consultation with our natural partners in Australia, the State libraries. We are doing the same thing internationally, again principally with our natural partners, other national libraries with similar interests, but basically with anyone who is interested, anyone who might be able to engage in useful conversation or work with us.<sup>(7)</sup>

We see all of these infrastructure developments as key parts of managing the digitisation process. It is hard to build them into guidelines, but at least we have recognised them as important issues. It is essential that any guidance talks about whole systems, not just resolution levels and formats. Again, this has been a real plus for this conference.

### Progress?

Two years ago I prepared a background paper on digitisation and management of digital information for a seminar on a National Preservation Strategy in Australia.<sup>(8)</sup> In that paper I tried to define the conditions that would be needed for digitisation to play a part in a national strategy. It is interesting to look back from the perspective of this conference, to see how far we have come in that time.

I said we would need to decide on the preservation role of digitisation - define what preservation objectives it was achieving, or could achieve, rather than just say: "we copied it, so we've dealt with the preservation problems". I think we have made a lot of progress on this issue, despite the ambivalence I have already mentioned. There is general agreement among us that digitisation is *not* a preservation panacea, but that in some cases it can have a preservation benefit if we put in effort to achieve it. There seems to be a growing understanding of that approach in institutions and in the community.

Secondly, I said we would need to find a lot of resources, and sort out how they should be allocated, given the likely conflicts between digitisation and other preservation responses to our deteriorating collections. This is being recognised in many institutions, although the pressure on resources remains. I think the conference has been helpful here in asserting that, at best, digitisation is one preservation tool among many and that it *is* expensive to make preservable digital copies and to make them accessible.

Thirdly, I said we would need to develop and bed down ways of managing digital copies - who owns them, who has access to them and under what conditions, and how we authenticate the copies. We are slowly getting there, partly through commercial pressure, partly political pressure to sort out laws that we might or might not like. In many countries, including Australia, it is too early to make any definitive statements or provide any definitive guidance. The risk of legal action makes this an area where we cannot afford for our guidance to be wrong.

Fourthly, I said we would need to improve the tools for decision-making in the digitisation process itself, to produce preservable copies worthy of preservation, while managing the significant risk of damage to originals being copied. Of course, this is the area most amenable to guidelines.

Fifthly, I said we would need to develop mechanisms for co-ordination of a national digitisation effort, involving decisions about roles, expectations, objectives, responsibilities. A couple of years ago the question was whether such co-ordination would be possible. Now in Australia there is great interest in such an approach, and many institutions feel they have a lot to gain by bringing their collections together for digital access. Whether that tends towards genuine preservation benefits, or develops into a pseudo-preservation smoke screen, may depend on the kind of messages that groups like this can make, and how clearly they are heard.



Finally, I said that we would need to ensure that the digital copies could be entered into an effective archiving and preservation system. As I have already indicated, I believe it is crucial for this issue to be addressed in any guidance we might put together. If you digitise, and want to maintain access, you *have* to plan for preservation action; you cannot say: "this project is only for access and has no preservation implications", if you mean access for more than a very few years.

So generally I would say we are doing quite well - and that this conference has moved us a few steps further - in meeting those requirements. But there is still some way to go, and I assume the work we do after this conference will make a very significant contribution.

Finally, I would like to comment briefly on the way forward - some reflections, decidedly unrigorous, on what our "guidance" might look like.

It seems to me we are talking about a growing body of information, ideas, advice and shared experience. This is hard to accommodate in the paradigms we have brought to standards and guidelines in the past. (Significantly, our Standards Association in Australia has recently been talking about moving away from published standards in some areas, to provide guidelines and advice. We are not the first people to recognise that the world is becoming too complex, and changing too fast, for a printed set of rules to mean very much.)

Perhaps we would be better served by something like a website rather than a book you might buy. We know from our PADI website that they take resources to develop and maintain, but if the content is good, the interface is good, and the owners and users are responsive, it is worth it.

We would be very interested in discussing the possibility of using our PADI site as a mirror site linked to RLG and the NPO to carry such a "guidance gateway".

All of us who have been involved with the preparation of standards and guidelines recognise that what we produce are a series of cause and effect statements. With digitisation some of those statements become very complex. For example, we might want to say:

- x is essential in all cases, or
- x is desirable in all cases and essential in some cases
- x is essential in some specific cases that we can predict but is definitely counter-indicated in other cases
- x is essential in some specific cases we can predict but may or may not be in other cases
- x is essential in some specific cases but we cannot predict which cases

... and so on.

The tool kit we offer has to include indicators for when one kind of decision is needed, and not another. That is a difficult task, but it is cause for hope that we are trying to offer guidance to humans capable of exercising judgment, rather than writing a decision-making program for a machine.

There will always be a tension between conciseness and the truth. That's why "it all depends" is both an honest and an intensely frustrating answer. When we are called on to advise on digitisation proposals, "it depends" is ultimately the most useful answer we can give, because it leads into the very process - the *only* process - that we have for helping people manage complexity. What we are looking for is a tool kit that helps us through that kind of process, and does so for all the different levels of expertise that we and our clients might bring to it.

Finally, one message we seem to be sharing here is that the apparently mad rush to digitise collections and to throw the originals away that may have been in the air a few years ago has dissipated. There is a widespread acceptance that digitisation is - or can be - a good thing to do, but that it is not the beginning and end of collection management or even of access. It is a tool to which we give value or not, by the things we do with it.

I believe this change has given us a little more room than we thought we had. We will never get our guidance completely, perfectly right, but we can afford to work on doing it well. While this conference could not be called a beginning, it assuredly is not the end of the

process. We have been tremendously productive. I look forward to sharing with you in the work that is still to be done.

## REFERENCES

- 1 National Library of Australia Information Preservation Section, 1998. Guidelines for preservation microfilming in Australia and New Zealand. National Library of Australia, Canberra.
2. Available at <http://www.nla.gov.au/raam/>
3. Available at <http://www.dstc.edu.au/RDU/MetaWeb/>
4. Available at <http://www.nla.gov.au/padi/>
5. Available at <http://www.nla.gov.au/pandora/>
6. Available at <http://www.nla.gov.au/dsp/>
7. Available at <http://www.nla.gov.au/policy/rsagenda.html>
8. Available at [http://www.nla.gov.au/nla/staffpaper/int\\_issu.html](http://www.nla.gov.au/nla/staffpaper/int_issu.html)
9. Available at <http://www.nla.gov.au/nla/staffpaper/cwebb3.html>

**Weaving the threads**  
**Peter Fox**  
**University Librarian**  
**University of Cambridge, United Kingdom**

During the course of the last two days we have heard a series of really excellent papers from speaker after speaker who are clearly experts in their field, whereas I am merely a library director - and by definition library directors do not know anything about anything! So you will be relieved to hear that I am not going to try and sum up each of the papers. We have been presented with a vast amount of information - some of it quite technical, and there would be little point going over that ground again.

At the beginning of the conference Anne Kenney stated that this was an unprecedented opportunity for action and there has certainly been a great deal of that. It seems to me that my job now is to try and identify those areas for future action - where do we go from here?

What I have found interesting is that the same topics have occurred again and again in different contexts and it seems to me there are two main threads or strands to the web I have been asked to weave.

The first is the issue of guidelines (or 'guidance' as we have been exhorted to prefer). What do we agree on? What do not we agree on? What do not we know? And how do we fill the gaps in our knowledge? Secondly, assuming that we can agree on standards and guidance, how do we get people to adopt them? In this there are two sub-strands: education and training, and the political dimension.

Mention has been made several times of the RLG and NPO guidelines for the preservation of microforms. I suspect that what many of us have been struggling towards is a set of guidelines similar to those but for digital imaging. In this, we should remember the warnings from several speakers, that digital is not the same as microfilm. It does not have the same level of stability, any decisions made at one point in the process will have some impact throughout, and, above all, to quote Anne Kenney, 'technology is not our friend'. Whatever guidelines we produce will have to be regularly revised.

We know that the hybrid approach (digitisation for access, microfilm or facsimile hard-copy for preservation) is a safe one and one that many of us feel comfortable with. But it cannot be a long-term solution and it certainly does not address the question that was touched on tangentially yesterday and in more detail by Margaret Hedstrom today, and that is the issue of information resources which are only electronic and have never been in analogue form. I do not believe it is an answer to say: 'let's digitise now and assume we may have to do it again in ten years' time'. The problem of brittle paper is undoubtedly a serious one, especially in North America, but the problem of preserving electronic information resources is even more urgent. And we will probably only have one chance to get it right. I know this conference is about digital imaging, but once analogue images have been digitised their preservation problems are essentially similar to those of electronic databases or publications that started life as CD-ROMs.

**So, where do we start?**

It is clear from our discussions that it is difficult, if not impossible, to create guidelines that suit all purposes. There is not necessarily only one right way. On the other hand, it is also clear that there are some areas where there is already a consensus on best practice, on areas

where it is possible to provide guidance - indeed where guidelines are already available - and where we can establish a common framework for making informed decisions.

It is also clear that there are areas needing further research, and many of these have been specifically identified in the reports from the small groups. There is also a need for documentation and guidance on what we already do know. I felt that the suggestions for a unified Web-site for guidance was a very valuable one; such a site could be mirrored by, for example, RLG, the NPO, and the National Library of Australia to give global coverage. Even at my very inexperienced level I have found that there is already too much information available and it is not always clear how far it is authoritative and reliable. This is an area of action for RLG and the NPO to take forward after the conference.

**Now I want to move from the technical issues to more general political ones.**

How do we get people to adopt any guidelines that we may produce? The first way of course is through publicity, and again it is the role of bodies like RLG and the NPO that is crucial. RLG has an international membership and is looked to by its members to provide this sort of leadership. Within the UK and Ireland, the NPO is well known and respected and plays a leading role in providing information about best practice. As part of this process, it has been agreed that the papers of this conference will be available shortly on the World-Wide Web and will also be published in paper form.

We have heard about the seminars at Cornell and you will have seen the blue flier in the conference pack for the summer school at Glasgow in July 1999 run by Seamus Ross. We need more of these, run by bodies such as these that know what they are doing.

Equally important is the need to influence the funding agencies and decision-makers. We must try and avoid the mistakes of the past. Over the last five years the Higher Education Funding Councils in the UK have funded a £50 million programme to support research collections in the humanities. A significant part of that funding has gone into preservation microfilming and it has not been a requirement that projects had to adopt Mellon standards. A number of us were concerned by this, and the NPO pressed the Funding Councils to agree to distribute information about the Mellon project. This was eventually done but it was only after the programme was well under way that such guidance was offered and attention drawn to standards and the role of the NPO, and by then adherence to the standards could only be a recommendation, not mandatory.

There will be a second programme of support for research libraries, albeit at a lower level and with different rules, but this will probably include digitisation projects. I am pleased to say that we have got the message across this time and it is likely that, before any projects of this type are funded, the NPO will be consulted about how far participants ought to adhere to whatever standards and guidelines are in place.

But higher education is only one sector. You heard yesterday of the report *New library: the people's network*, a programme in the UK for which the government has pledged several millions of pounds a year over the next four years for the digitisation of education and learning materials for public libraries. The report is all about access - what is worrying is that there is no mention of preservation in it and precious little about standards. Also in the UK, the Heritage Lottery Fund will undoubtedly be asked to provide grants for digitisation projects. At a European level, DG XIII of the European Commission is showing interest in this area and is carrying out an analysis of the existing guidelines to inform future work under the 5th Framework Programme.

So, on this side of the Atlantic there are a number of programmes and a significant amount of funding available. We have a responsibility to try and ensure it is spent in the most

effective way. I am less familiar with programmes in the US and elsewhere but the same cautionary note applies. If possible, bodies such as the National Endowment for the Humanities should be urged to adopt the same approach to any funding projects.

One specific action is already being planned. Many of you will be familiar with a series of reports based on the JISC/NPO studies on the long-term preservation of electronic materials, and co-ordinated by the Digital Archiving Working Party (DAWG), of which I am Chairman. These are detailed reports and in parts are quite technical. We have, therefore, commissioned a synthesis of around 80 pages which will be non-technical, aimed at the non-specialist, and will, we hope, be short enough for busy people to read.<sup>(1)</sup> We are planning to combine the launch of this, probably early in 1999, with a meeting for decision makers in the Higher Education Funding Councils, the Lottery, relevant government departments, the European Commission, etc., to try to impress upon them that decisions on what to digitise and how to do it must be based on best practice and take into account our knowledge of the sort of issues that have been discussed at this conference.

It was pointed out on Monday how truly international this conference is - there are representatives of 18 countries present. The problem of digital preservation is also a global one. I have stressed the role of RLG and the NPO but the participants here today are not just from North America and the British Isles. We need to ensure that we harness the best technical expertise from wherever it is on the globe and try to ensure that the money which is being spent on digital imaging - much of it from public funds - is spent in the most cost-effective way.

Those, then, are for me the two messages of the conference. This was an unprecedented opportunity for action. In this, I believe it has been extraordinarily successful and I feel we have made a great deal of progress. The success has been due in large measure to the planning and organisation by Vanessa Marshall of the NPO, and by Nancy Elkington and Robin Dale of RLG, but it was also due to your hard work and input. We must now ensure that we do not lose the momentum. From what you heard earlier this morning, it is clear that a lot of activity is about to take place, so my message to you all is that you need to produce guidance, you need to proselytise, but above all you need to go forth and digitise!

## REFERENCES

1. Feeney, Mary (Ed.) (1999) *Digital Culture: Maximising the Nation's Investment*. The National Preservation Office, The British Library, 96 Euston Road, London NW1 2DB. This publication is available free of charge from the above address.

## Appendix 1 Supporting Documents

In order to facilitate the objective of the Joint Conference - identifying consensus and creating guidelines for digital imaging - the speakers provided background documentation. All conference participants were encouraged to become familiar with the following documentation prior to the conference.

- Working Group on Selection

Selection Criteria, Guidelines, Decision-making Aids, available at  
<http://www.rlg.org/preserv/joint/selection.html>

- Working Group on Preparation

Survey from the Working Group on Preparation, available at  
<http://www.rlg.org/preserv/joint/imaging.html>

- Working Group on Digital Image Capture

Key Components to Quality (Guidelines), available at  
<http://www.rlg.org/preserv/joint/imaging.html>

- Metadata Information

"Metadata for Preservation" by Michael Day, available at  
<http://www.ukoln.ac.uk/metadata/cedars/AIWO1.html>

Making of America 2 (MOA2) White Paper, Version 2.0: Metadata Section, available at  
<http://sunsite.Berkeley.EDU/Moa2/wp-v2.html>

If you have any questions regarding the supporting documentation, please contact Robin Dale, RLG.

## Appendix 2

### Responses to preparation questionnaire

#### Physical Preparation

##### Points on which discussion group strongly agreed

- Microfilm model is useful
- Reliance on a digital copy - why is this inadequate if you have any other achievable option
- Emphasis on need for protection of originals - a strong responsibility - although it is valid to say it is not an issue in some cases
- Proper/safe handling measures - generally agreed that there is a need for this
- Value of proper planning
- The ability to easily to exclude costs of for example (re)cataloguing indexing - seen as normal part of repository function, therefore no need for special funding
- Strongly agree with need for safeguards to be incorporated into the planning process
- When devising best practise, be aware of necessity to include things that may seem obvious to those who have already done projects e.g. light levels
- Need steps to be spelt out for the uninitiated
- Very concerned about guillotining and use of flat-bed scanners
- Agree on the importance of temp/rh control through impact of lighting
- Retain control of materials and processes
- Post-scanning access
- Agree lighting - it should be appropriate. Need to use appropriate equipment for the type of material
- Need to assess all aspects of physical condition before embarking on project, including surface layer fragility from loose pigment to emulsion layers
- Agree with the need to match scanning equipment with type of material in a bid to minimise danger/damage to materials

##### Points on which discussion group strongly disagreed

- Didn't really disagree strongly with any of the statements
- The sweeping thumbs-down on guillotining (comment - remember, business environment binding can serve a regulatory/audit pedigree...)
- Book cradle - horizontal not suitable - many felt there was more room for its use than the speaker especially Zeutschel bookscanner cradle
- Simply not feasible for us to spend the time and money to allocate the space to reformat digitally and preserve the original
- Disbinding - disagreed

##### Areas that would benefit from further research?

- Lighting issues
- Effects on demand to handle originals post-scanning
- Cost-benefits analysis tools for really hard-nosed decisions
- Increase versus decrease in demand for original

- Light level efforts, photo stability, cataloguing, any difference available in production to speed up a project which allow it to become more cost-effective
- Effects on physical items that different scanning techniques have
- Need to develop a cheap book cradle that everyone can use
- Methods of scanning - more case studies
- More discussion on 'born digital'
- Information on outsourcing and collaborative efforts
- Relative costs of different parts of the digitisation process
- Assessment of damage
- How much use is material receiving - if it is being done for access and the originals are not receiving high use, why digitise?

### Editorial Preparation

#### Points on which discussion group strongly agreed

- Need to think process thoroughly from start - difficulties build very quickly
- Need to remember that target is to create a usable product
- Emphasis on editorial preparation
- With "apples and oranges" references regarding costs
- Need for careful preparation - can save time and money and should be included in guidelines
- With importance of identifying an item's structure properly prior to production
- Consider final results - unique ID e.g. URN (Uniform Resource Name)
- Need to structure work so can be managed in a production facility, usually involving students and clerks

#### Points on which discussion groups strongly disagreed

- "Reverse engineering not possible" - it is very possible
- Filming model is good but may be too simple

#### Areas that would benefit from further research

- Self referencing elements of structure
- Integration of metadata models so metadata we need for managing does not get lost in discovery metadata
- German model database for managing structure - guidance or an approved model for all
- Harvesting metadata...
- REAL costing models
- Curatorial responsibilities and technical responsibilities - divisions and overlaps
- How to determine that work has not been done elsewhere - seems like a fundamental preparation step
- Nuts and bolts - what Quality Control measures are being implemented?
- Comparison of different organisational structures i.e. the paradigm shift Ann talked of...has it been done?...does it work?



## Appendix 3 Participants List

Ms Valerie Adams  
Head of Preservation  
Public Record Office of Northern Ireland  
proni@nics.gov.uk

Mr Michael Alexander  
Systems Manager  
The British Library  
michael.alexander@bl.uk

Ms Frits Ankersmit  
Project Manager  
Utrecht University Library  
f.ankersmit@library.uu.nl

Ms Linda Armitage  
Researcher  
School of Information Management  
l.h.armitage@brighton.ac.uk

Mr Kevin Ashley  
Special Projects Manager  
ULCC  
k.ashley@ulcc.ac.uk

Ms Jackie Aspinall  
Sales & Marketing Director  
Riley Dunn & Wilson Limited  
jackie.aspinall@lineone.net

Mr Paul Ayris  
Director of Library Services  
Education and Information Support Division  
p.ayris@ucl.ac.uk

Ms Anne Barrett  
College Archivist  
Imperial College of Science, Technology and Medicine  
a.barrett@ic.ac.uk

Mr Neil Beagrie  
Collections and Standards Officer  
Arts & Humanities Data Service Executive  
neil.beagrie@ahds.ac.uk

Mr Reg Carr  
Director of University Library Services  
University of Oxford The Bodleian Library  
reg.carr@ulib.ox.ac.uk

Mr Stephen Chapman  
Preservation Librarian for Digital Initiatives  
Preservation Centre Harvard University Library  
stephen\_chapman@harvard.edu

Ms Alison Chilvers  
Research Student  
Department of Information & Library Studies  
a.h.chilvers@lboro.ac.uk

Mr Jo Clary  
Digital Resources Officer  
The Data Archive University of Essex  
jow@essex.ac.uk

Mr Andrew Colleran  
Director  
Quercus Information Limited  
andrew.colleran@quercus.co.uk

Mr David Cooper  
Project Officer  
OULAS  
david.cooper@las.ox.ac.uk

Ms Rosamund Cummings  
Records Manager  
University College London  
rhc@central-services.ucl.ac.uk

Ms Robin Dale  
Program Officer  
Research Libraries Group  
robin\_dale@notes.rlg.org

Mr Michael Day  
Research Officer (Metadata)  
UKOLN  
m.day@ukoln.ac.uk

Mr Mercedes Dexeus  
Directora del Departamento de Patrimonio Bibliografico  
Biblioteca Nacional de Espana  
m.dexeus@bne.es

Ms Nancy Elkington  
Program Officer, Research Libraries Group  
c/o BLPES, London School of Economics  
bl.nee@rlg.org

Mr Jose L Esteban  
Automation Coordinator  
Biblioteca Nacional de Espana  
jose.esteban@bne.es

Mr Christopher Fleet  
Assistant Map Curator  
National Library of Scotland  
c.fleet@nls.uk

Dr John Fletcher  
Head of Technical Services  
British Library  
john.fletcher@bl.uk

Ms Jose Fonseca  
Computer Technician Swedish Archives Information  
SVAR  
jose.fonseca@svar.ra.se

Mr Peter Fox  
University Librarian  
University of Cambridge Library  
pkf@ula.cam.ac.uk

Mr David French  
Conservator MSS Conservation Studio  
The British Library  
david.french@bl.uk

Mr Marc Fresko  
Principal  
Applerace Limited  
applerace@marc.easynet.co.uk

Ms Janet Gertz  
Director for Preservation, Columbia University  
101c Butler Library  
gertz@columbia.edu

Mr Adrian Glew  
Archive Curator  
adrian.glew@tate.org.uk

Ms Les Goodey  
Deputy Head of Photography and Digital Imaging  
Photography Department Cambridge University Library  
irg1000@cam.ac.uk

Ms Hazel Gott  
Events manager  
UKOLN  
h.a.gott@ukoln.ac.uk

Ms Margaret Graham  
Research and Development Manager  
Institute of Northumbria at Newcastle  
margaret.graham@unn.ac.uk

Mr Stephen Green  
Policy Adviser, Archives & Libraries  
Heritage Lottery Fund  
spgreen2@hotmail.com

Mr Francis Grew  
Curator  
Museum of London  
fgrew@museum-london.org.uk

Mr Macros Guillen  
Network Administrator  
English Heritage  
m.guillen@eng-h.gov.uk

Ms Ann-Chatrin Haggstrom  
Supervisor IT, Swedish Archives Information  
SVAR  
annchatrin.haggstrom@svar.ra.se

Mr Andrew Hampson  
BUILDER Project Manager  
Information Services, Main Library  
a.d.hampson@bham.ac.uk

Mr Doris Hayashikawa  
Coordinator  
Digital Research Library Center G20M Hillman Library  
dorish+@pitt.edu

Ms Margaret Hedstrom  
Associate Professor  
School of Information University of Michigan  
hedstrom@umich.edu

Mr Birgit Henriksen  
Head of Digitisation and Web Department  
The Royal Library, Denmark  
birgit@webadm.kb.dk

Mr Brian Hillyard  
Assistant Keeper  
National Library of Scotland  
b.hillyard@nls.uk

Mr Doug Hodges  
Database Policy and Planning Officer  
Information Resource Management  
doug.hodges@nlc-bnc.ca

Ms Alison Horsburgh  
Head of Preservation Services  
National Archives of Scotland  
ahorsbur@sro.gov.uk

Ms Anne R Kenney  
Associate Director  
Department of Preservation Cornell University  
ark3@cornell.edu

Mr Erich J Kesse  
Chair, Preservation Department  
George A Smathers Libraries University of Florida  
erich@mail.uflib.ufl.edu

Ms Tania Kettell  
Editor  
Primary Source Media  
tania.kettle@psmedia.co.uk

Ms Denis King  
Photographic Administrator  
The National Gallery Photographic Department  
denise.king@ng-london.org.uk

Mr Toby Kirtley  
Library Preservation Officer  
Preservation and Conservation Department  
trk@bodley.ox.ac.uk

Mr Colin Labrum  
Microfilm Specialist  
Kodak Limited  
colin\_labrum@knotes.kodak.com

Mr John Lauder  
Director  
Scottish Newspapers Microfilming Unit  
snmu@sol.co.uk

Mr Norbert Lossau  
Head of the Centre for Digitisation (GDZ)  
Niedersaechsische Staats- und Universitätsbibliothek  
lossau@mail.sub.uni-goettingen.de

Mr Craig Mauzy  
Staff Photographer Agora Excavations  
American School of Classical Studies  
agora@ath.forthnet.gr

Ms Marie Mauzy  
Photographer  
American school of Classical Studies  
mmauzy@ascsa.edu.gr

Ms Barbara McDonald  
Electronic Storage Project Manager  
Trinity College Dublin Library  
barbara.mcdonald@tcd.ie

Mr Ian McGowan  
Librarian  
National Library of Scotland  
i.mcgowan@nls.uk

Mr John McIntyre  
Head of Preservation  
National Library of Scotland  
j.mcintyre@nls.uk

Mr Paul McLaughlin  
Sub-Librarian Planning and Administration  
University of London Library  
pmclaughlin@ull.ac.uk

Mr Christopher Mills  
Head of Collections & Readers' Services  
The Natural History Museum  
c.mills@nhm.ac.uk

Ms Janice Mohlhenrich  
Coordinator Preservation Initiatives & Services  
Preservation Office Woodruff Library  
jmohlhe@emory.edu

Mr Jonas Mostrom  
Technician SVAR  
National Archives of Sweden  
jonas.mostrom@svar.ra.se

Mr Hans Mulder  
Curator  
Utrecht University Library  
h.mulder@library.uu.nl

Mr Mark Nicholls  
Deputy Keeper of Manuscripts  
Cambridge University Library  
mn.@ula.cam.ac.uk

Mr Raili Nugin  
Councillor of the Head of the Archive  
State Archive of Estonia  
nugin@netexpress.ee

Ms Siobhan O'Rafferty  
Librarian  
Royal Irish Academy  
s.orafferty@ria.ie

Mr Jonas Palm  
Head of Preservation  
The Royal Library, Denmark  
jp@kb.dk

Mr Joy R Paulson  
Head Reformatting & Replacement Preservation  
University of Michigan Library  
jpaulson@umich.edu

Mr Michael Pidd  
Technology Development Manager  
Humanities Research Institute  
m.pidd@sheffield.ac.uk

Mr John Ralph  
Head of Computing Services  
ULCC  
j.ralph@ulcc.ac.uk

Mr Seamus Ross  
Director, Humanities Computing & Information Management  
Faculty of Arts University of Glasgow  
seamusr@arts.gla.ac.uk

Ms Kelly Russell  
Cedars Project Manager  
Consortium of University Research Libraries  
k.l.russell@leeds.ac.uk

Mr Jan Sahlen  
Project Leader Swedish Archives Information  
SVAR  
jan.sahlen@svar.ra.se

Ms Jean Samuel  
Director  
Adata  
jean.samuel@btinternet.com

Ms Anne Scallan  
Project Manager  
The Wellcome Trust Medicine, Society and History Division  
a.scallan@wellcome.ac.uk

Ms Helen Shenton  
Deputy Director Preservation  
Collections & Preservation Department  
helen.shenton@bl.uk

Mr Bill Simpson  
Librarian  
Trinity College Dublin Library  
wsimpson@tcd.ie

Ms Abby Smith  
Program Officer  
Council on Library & Information Resources  
asmith@clir.org

Mr Stanley Smith  
Digital Imaging Specialist  
Experience Music Project  
stanleys@experience.org

Ms Emma Stewart  
Collection Development & Administration Manager  
London Metropolitan Archives  
lma@ms.corpoflondon.gov.uk

Ms Ann Swartzell  
Senior Preservation Program Officer  
Harvard University Library Preservation Center  
a\_swartzell@harvard.edu

Mr Simon Tanner  
Digitisation Consultant  
Higher Education Digitisation Service University of Hertfordshire  
s.g.tanner@herts.ac.uk



Mr Jonathan Tripp  
Technical Manager  
Kodak Limited Business Imaging Systems  
jon\_tripp@knotes.kodak.com

Ms Karen Turko  
Head, Preservation Services  
University of Toronto Library  
turko@library.utoronto.ca

Mr Joris Vader Linden  
Head, Studio for Micrographic and Digital Archiving  
Library of the Belgian Federal Parliament  
jen@village.uunet.be

Mr Robert van der Zwan  
Manager  
Interactive Open Learning Centre & Media Archive  
r.m.vanderzwan@open.ac.uk

Ms Julia Walworth  
Head, Special Collections/Rare Books  
University of London Library  
jwalworth@ull.ac.uk

Ms Anna Watson  
Senior Archivist  
Lancashire Record Office  
lancsrecords@treas.lancscc.gov.uk

Ms Colin Webb  
Manager  
Information Presevation  
cwebb@nla.gov.au

Ms Alison Webster  
Head of Preservation Copying  
Public Record Office  
alison.webster@pro.gov.uk

Mr David Weston  
Keeper of Special Collections  
Glasgow University Library  
d.weston@lib.gla.ac.uk

Mr Jacqui White  
Marketing Specialist  
Business Imaging Systems Deptment  
jacquiw@kodak.com



Appendix 4  
After Event report  
Joint RLG and NPO Conference on Guidelines for Digital Imaging  
Nancy Elkington Program Officer  
Research Libraries Group  
Neil Beagrie, Collections and Standards Officer  
Arts and Humanities Data Service Executive

Over ninety people from North America, Europe, and Australia convened at the University of Warwick (UK) September 28-30, 1998 for a first-of-its-kind working conference. The Research Libraries Group and the National Preservation Office (NPO) of the UK and Ireland joined forces to focus on creating an international consensus in the development of guidelines for digital imaging and digital preservation. Proceedings for the conference will be made available at the conference Web site (UK and Europe link) and in print form (from NPO) in the coming months. Updates that include summaries of plenary discussions, small group recommendations, and follow-up actions by RLG and the NPO will appear on the Web site in the coming weeks. Meanwhile, this abbreviated version is being made available as soon as possible after the event. Many thanks are due to speakers and delegates for making this a memorable and inestimably useful conference.

Fifteen speakers from the US, the UK, Canada, and Australia shared insights and challenges on selection issues, physical and editorial preparation requirements, image capture techniques and rationales, metadata approaches, and digital preservation imperatives. Delegates from over a dozen countries were able to air their opinions in small group discussion sessions and open question/answer periods amid several plenary sessions. And all participants took great advantage of the unique opportunity to exchange ideas and trade experiences less formally in hallways, lounges, and dining rooms over the three-day conference.

Anne R. Kenney (Cornell University), the conference keynote speaker, is one of the world's most experienced professionals currently engaged in exploring digital technology and seeking best methods for applying it to the needs of libraries and archives. In her remarks she acted as teacher and mentor, sharing the breadth of her experience in a series of ten guiding principles. Two of the most critical of these, at least in the context of the conference, were: "Don't reinvent the wheel" and "Don't reinvent an old wheel." Participants were admonished to avoid drawing too literal an analogy between preservation microfilming strategies and work flows and those that are relevant to digital conversion efforts. She further argued against the development of guidelines that are too rigid or too narrowly conceived; instead, she suggested that experts should be working toward agreements on providing guidance rather than guidelines. Kenney encouraged the group to undertake that which is doable, and to acknowledge the gap between the ideal and the possible. She also emphasised the relationships and inter-dependence of decisions taken during the selection, preparation, and capture in digitisation projects and the importance of a co-ordinated approach, citing the recent AHDS study and its advice on developing a strategic policy framework for creating and preserving digital collections. Finally, and most persuasively, she urged participants to recognise that sometimes "reasonable, well-informed people can disagree" on what constitutes best practice.

### **Selection**

Guidelines for Selection, the first of three focused sections combining formal presentations and small group discussions, was moderated by Julia Walworth (University of London

Library); the podium was shared by speakers Janet Gertz (Columbia University) and Paul Ayris (University College, London). Speakers identified examples of selection guidelines available on the Web, and summarised some of the common criteria used in digitisation projects, such as whether the original item is damaged or in danger, its in high demand, or has significant intellectual or artifactual value. It was stressed that most digitisation is access driven rather than preservation driven at present. Discussion raised the issue of whether preservation of digitised images would prove more costly than re-digitisation at some future date. It was generally concluded that this assumption may often be curtailed by the need to conserve the original objects and to invest in converting materials not yet in digital form.

The need for guidance rather guidelines proposed by Anne Kenney was widely endorsed. At the same time it was recognised that some education is needed to achieve and implement this: often in questions on digitisation a single solution or standard is sought. The answer to such questions invariably begins with "it all depends..." There is a learning curve on how to choose appropriate solutions and standards which are "fit for purpose," and that meet project and institutional needs and requirements.

Recommendations coming back to the audience from the small group discussion included:

1. No one set of guidelines can be prescribed that will at the same time be useful and usable by all types of institutions and repositories.
2. Institutions should be encouraged to formulate their own guidance documents that will help them reach decisions regarding digital conversion of collection materials.
3. There should be a unified Web site for information, case studies, and guidance on the process of selection to aid institutions wishing to expand their understanding of the issues and develop their own selection guidelines.
4. Selection guidance will only be effective within a wider institutional framework and consideration and development of institutional collection policies for digital materials.

### **Preparation**

Preparation, the second of three in-depth sessions, was chaired by Alison Horsburgh (National Archives of Scotland). Joining her to address issues of physical preparation was John McIntyre (National Library of Scotland); Ann Swartzell (Harvard University) addressed issues characterised as editorial preparation.

Papers presented were wide-ranging and detailed, drawing on the extensive experience of the speakers. Issues addressed included protecting the original item; assessments of condition and identifying vulnerable material; preservation/conservation records; procedures for safe handling; identifying suitable scanning methods; disbinding (when, if, and how!); scanning methods and their implications for both the originals and the quality of the digital image; use of intermediates and surrogates; types of lighting and their effects; and institutional policies for post-scanning access to the original items. The lessons learnt from preservation microfilming projects for preparation of materials and planning production workflow were also discussed, together with changes needed for digital production regimes. The importance of cataloguing and bibliographic control were also emphasised. The experience of a number of projects underscored the fact that preparation, cataloguing, and provision of metadata may be twice or three times the costs associated with image capture.

Recommendations coming back to the audience from the small group discussion included:

1. Consensus on the need for a range of guidance documents (some based on existing reports and published research) addressing: disbinding, lighting, book cradles, handling procedures, scanning methods, use of intermediates and surrogates, access policies to originals post-scanning, finding aids, file naming, and metadata.

2. Recommendations to address unmet needs in the areas of: costs (cost/benefit analysis models), staffing, training, physical work areas, terminology, digital preservation, and articulating the role of preservation and preservation-mindedness in digital imaging for access.

### **Image Capture**

Robin Dale (RLG) led the session on image capture, arguably the most complex of the three topics given extensive attention in the conference. Stephen Chapman (Harvard University) and Jane Williams (Technical Advisory Service on Images) explored a range of technical, intellectual, and educational issues. All the speakers emphasised the inter-relationships and decisions taken between image capture and the preceding stages of selection and preparation.

Stephen Chapman outlined the different trade-offs between image quality, cost, potential or acceptable damage to the original, use of intermediates, and the method of scanning employed. He also highlighted the paradox that while many digitisation projects are undertaken to reduce use of originals, the actual handling of the originals during the process of digitisation may be collectively higher than at any other period in the history of a collection. The conservation implications of digitisation itself need to be considered and the impact of digitisation on originals minimised. Digital cameras were discussed with the different requirements of and implications for getting a good signal, getting a good scan, and getting a good image. The importance of calibrating the overall system and components used for digitisation was emphasised together with the view that, in addition to guidance, there may be a need to identify at least minimum requirements for preservation quality imaging. Different aims and "capture levels" for image outputs were also discussed: the four image levels used by Corbis of rendered, faithful to photographer's intent, original photograph rendered (restored), and original film rendered (film bias corrected) were summarised.

Jane Williams outlined the work of the Technical Advisory Service for Images (*TASI*), a national service for the UK higher education community. Her experience again underlined the need for education rather than prescription, and for ensuring that digitizers understand the consequences of different choices they will face so that they can then make informed decisions. A wide-ranging overview with recommendations was presented of image capture, imaging methodologies, formats, acquisition, archiving, quality control, file naming, and delivery and presentation to users. The many clear recommendations included the suggestion that digitisation projects should consider an initial "dummy" run at the beginning of a project to allow for experimentation, calibration of systems, developing capture procedures and quality checks to reflect the fact that quality increases as scanning procedures are refined and experience developed.

Recommendations coming back to the audience from the small group discussion included:

1. General agreement on desirable minimum imaging characteristics of black and white (1-bit), gray-scale (more than 1-bit), and colour (24-bit and higher) images.
2. Need for a set of guidance documents to assist those with less experience in imaging programs and projects, including: definition of a "digital master," white paper on digital reproductions, guidance on quality control, principles for selecting colour space, and widely disseminated minimal attributes of file formats (as above).
3. The group suggested an action agenda (including the tasks listed in 2) that will be taken up by the conference organisers in the next few months.

### **Preservation Metadata and Digital Archiving**

Two conference sessions were linked and intended to give participants an overview of the issues and challenges inherent in addressing metadata and digital preservation needs. Michael

Day (UK Office for Library Networking) and Margaret Hedstrom (University of Michigan School of Information) gave excellent presentations on metadata and digital archiving, respectively.

Michael Day surveyed the array of metadata types and models put forward by a host of organisations with distinctly different purposes. He pointed participants to the work of three projects/organisations that are specifically addressing metadata requirements in the context of files that are intended to survive over the long term: RLG's Working Group on Issues of Preservation Metadata, the Making of America 2 (MOA2) White Paper, Version 2.0: Metadata Section, and the Consortium of University Research Libraries' CEDARS Project. Day concluded that there is a need for integration of reservation metadata models with other metadata forms, and assured the group that the CEDARS Project will, as one of its outcomes, develop guidance documents for others interested in metadata supporting digital preservation.

Margaret Hedstrom addressed digital preservation by encompassing the larger world of digital documents, whether they are "born" digital or arrive in that form after an analogue-to-digital conversion effort. Underscoring the disjuncture between implementations and best practices, she suggested that best practices will always be influenced heavily by the nature of the source documents, known and anticipated uses of the material, and by the institution's technical capacity. Hedstrom argued for the reliance on non-proprietary file formats and encoding schemes, cautioning that conversion into such formats must be balanced against a series of risk assessments (information loss, source integrity, cost of conversion). Further, she set out best thinking in storage and maintenance techniques, identified areas where we are currently lacking guidance and experience, and listed a range of techniques used to ensure access to digital materials into the future. She concluded by reporting on a survey of RLG members as to their digital preservation needs; a final report will be made available in the RLG Web site late in 1998. One outcome of the survey already identified in discussion with the audience was the absence of adequate institutional collection policies for electronic materials and the need to support and guide their development.

### Conclusions

An international reaction panel was formed in order both to reflect activities outside of the US and UK contexts and also to point conference organisers toward new ways of thinking about how to address the dozens of needs expressed over the course of the three days. Karen Turko (University of Toronto), Seamus Ross (University of Glasgow), and Colin Webb (National Library of Australia) shared the spotlight during this session.

The reaction panel provided an overview of current digitisation initiatives in Canada, the UK, and Australia. Karen Turko provided an overview of the Early Canadiana Online Project, University of Toronto Library Digital Initiatives, and the Canadian Initiative for Digital Libraries. Seamus Ross stressed the need for strategic vision to develop digital resources and their preservation, and referred to national initiatives within Higher Education in the UK such as the Arts and Humanities Data

Service (AHDS) and CURL Exemplars in Digital Archives (CEDARS). Drawing on his experience as Information Technology (IT) Adviser to the UK's Heritage Lottery Fund, he stressed the very different needs of funding agencies. Funders need to know risk assessment, how to identify best practice, and how to assess the public benefit of digitisation projects. The staff of digitisation projects must understand and be able to apply guidance and identify the best trade-off of different requirements and standards for their projects. Colin Webb spoke of the IT initiatives in Australia, particularly at the National Library and the National Archives, and their relationship with local and

regional institutions. He stressed the need to focus on the initial development of infrastructure and appropriate tools such as the customized harvesting tools for Web pages developed for the PANDORA Project. The importance of a sense of local ownership of policies and guidance was also emphasized - the benefits of learning from others and not re-inventing the wheel need to be tempered with "checking the wheels" ("kicking the tires") and ensuring commitment from the participants.

Peter Fox (Cambridge University Library) summarised the conference by harking back to Anne Kenney's keynote message that this is an unprecedented opportunity for action. He asserted that there were two main themes that we need to pay attention to: first, that there are several areas in which consensus and agreement is already possible, and second, that we need to think strategically of how best to persuade institutions to adopt guidelines and guidance provided. Fox encouraged RLG and NPO to work closely with others where needed expertise lies outside of their purview, to "harness expertise wherever it can be found."

**Finally, he urged all those in attendance: "Don't lose the momentum."**