



EnrichUK

Pete Dowdell and Ann Chapman, UKOLN

EnrichUK.net

EnrichUK.net [1] is the showcase for the websites of all 150 projects funded as part of the £50 million New Opportunities Fund (NOF) Digitisation of Learning Materials Programme (or *nof-digitise*). The portal provides a search facility across a database of collection level descriptions for each site and, in many cases, for collections within sites, thus providing access to material across the whole of the *nof-digitise* programme in ways that would not be possible from individual project websites.

Projects funded under the programme range from national centres of excellence to regional 'sense of place' groups, along with a variety of voluntary and community organisations and small local museums.

The material accessible through EnrichUK.net is targeted particularly to users of public libraries through the People's Network and to schools through the National Grid for Learning. The target user base is therefore very wide, encompassing people of all ages, in both rural and urban communities, and with a huge range of interests.

EnrichUK.net provides access to material that someone might otherwise find it difficult to find and use due to a number of reasons. The original material of interest might:

- Not be widely known as being available.
- Form part of a network of resources on a topic, and finding one resource on a topic may not provide any links to other related resources.
- Be held at a site at some distance from the user's home.
- Be held at a number of sites, requiring several visits to be made.
- Be in a physical condition that means that direct handling of the original material is not possible.

Background

The *nof-digitise* programme was launched in August 1999 to put information and resources which support learning into digitised form. It has created innovative on-line learning resources for the potential benefit of every citizen in the UK. The programme brought together a wide range of partnerships and organisations and was designed to support lifelong learning, through the digitisation of material in three broad thematic areas:

Citizenship in a modern state

This first area covers information that helps people to access and use public services. It includes information on people's rights and obligations and also on how to be effective in dealing with public-sector organisations.

Cultural enrichment

The second area focuses on material that reflects the heritage of a community, region or a country, or embodies an aspect of identity. It includes many different forms of art in the areas of fine or performing arts, design and media.

Reskilling the nation

The third area includes material aimed at helping people to enhance basic literacy skills to improve understanding of other forms of information such as science, health and IT.

Who took part?

A total of 500 organisations grouped into 37 consortia and 34 individual projects covering a broad swathe of topics contributed to the programme. Content digitised under the programme includes newsreel footage from British Pathe, early Ordnance Survey maps of the UK held by the British Library and others, images of museum objects, Citizen's Advice information, newspapers, music and more.

The *nof-digitise* programme has created a huge amount of digital material, with one million pages of text, 400,000 images, thousands of film and video clips and over a thousand learning journeys. Details of the funded projects are available on the *nof-digitise* site [2].

Objectives

Creating this large array of digital resources was only the first step of the process; the resources then needed to be made accessible. Simply bringing the resources together as a portal was not enough on its own; it required a search facility that was both easy to use and could search across the large quantity and range of the digitised resources.

The solution was to create a database of descriptions of the various collections of digitised resources. Since the descriptions should conform to an existing standard if one was available and the RSLP Collection Description Schema was emerging as the 'de facto' standard for this type of description, it was decided to use the RSLP schema as the basis of the collection description database.

The database would be searchable and successful searches would display collection details, together with links to the website of the resources found. It was also important that the user interface should provide easy to use browsing and searching and the description contributor interface should also be a simple process.

The objectives were therefore to:

- Review the RSLP Schema to assess its applicability to the EnrichUK.net collections.

- To identify missing and redundant data elements, and potentially unhelpful elements labels.
- To define a set of terms to support topic, geographical and time period searching.
- To set up a database to hold the collection description records.
- To set up a user interface for browsing and searching.
- To set up a contributor interface for data entry.

What we did

The EnrichUK.net collection description model takes a simple approach with three main descriptions: project, collection and agent.

For collection level descriptions, the EnrichUK.net project team decided to create their own schema, based on the RSLP schema. By using the RSLP schema as the base, the collection descriptions could be searched alongside other collection databases and the records created could potentially be shared and exchanged.

However, a review indicated that the RSLP schema was not a perfect fit and required some customisation to fully satisfy the objectives of EnrichUK.net.

Step 1: Defining the data

The first step was to identify the data that EnrichUK.net wanted to hold about collections. The data required relates to five groups of elements.

Descriptive elements: the elements required here were the title of the collection, the description (containing a brief text description of the collection) and the name of the project creating the content of the collection.

Indexed searching: EnrichUK.net was designed to allow searching across subjects covered by the programme. As well as topic areas, this includes indication of content by both geographical area and time period, and the language of textual items held. In addition, the names of people associated with specific collections were also to be included.

Related resources: For each project that carried out digitisation work for a collection or collections, data was also to be included. Details to be held were project title, description, launch date, website, and a list of all collections digitised as part of the project; this data would be publicly displayed. Some contact details were also held but not made publicly visible.

Access to the collection: All collections are openly accessible via websites, so a URL link to the relevant website was required.

Housekeeping: EnrichUK.net needed to hold some details for administrative purposes. It was decided to use an Agent record to hold name, organisation, telephone numbers and email details. This data would not be publicly visible. The project records also hold the same 'not publicly visible' contact details.

Step 2: Identifying a schema

Having defined the required data, it was decided to use the RSLP collection level description (CLD) schema as the basis for an EnrichUK.net collection-level schema. While the RSLP CLD Schema element set covered much of the data some problems were identified. These were:

- elements not required
- element label terminology
- missing elements
- elements requiring qualification

Step 3: Elements not required

On review, it was decided that a number of RSLP CLD elements were not required for EnrichUK.net and these were omitted from the schema.

- Identifier – no identifiers are assigned to the collections being described so there was no requirement for this element.
- Strength – this element holds a free text or formalised indication of the strengths of the collection. Since each collection has a particular focus, use of the subject elements is sufficient.
- Physical characteristics – not required since all collections hold only digital files.
- Access control – not required since all collections are freely accessible via websites.
- Accrual status – not relevant to these collections as the digitisation is a one-off programme and further materials are not being acquired.
- Legal status – not relevant to these collections.
- Custodial history – not relevant to these collections.
- Note – no requirement for this field.

Step 4: Element labels

It was decided to change the labels for some elements to ones that EnrichUK.net users would find more understandable.

- **Website** – replaces Location – this element contains the URL for the collection website
- **Topic** replaces Concept: – this element contains terms from an EnrichUK.net customised two level list of subject terms
e.g. Working life - textiles
- **Related People** replaces Names – these are entered in direct name order.
e.g. Robert Witt, Lord Lee
- **Region** replaces *Coverage:geographic* – this element contains terms taken from an EnrichUK customised list of regions
e.g. East Midlands – Derby City
- **Date** replaces *Coverage:temporal* – this element contains historical period covered by the contents of the collection, shown by date ranges
e.g. 1589 AD – 2003 AD

Step 5: Missing elements

With the required RSLP elements defined, some data still remained to be accommodated. Data about owners and administrators of collections is held in RSLP agent elements but EnrichUK.net also wanted to hold details about the projects.

For project details, therefore, a project record was created, with appropriate elements. Project records are linked to the relevant collection record. As the project is often the owner / collector / administrator of the collection, agent records are derived from the project records.

- Agent record – this record has elements for name, organisation, telephone, email and address details; these records are not for public display.
- Project record – this record has elements for title, description, launch date, and website which are for public display. There are further elements for organisation, telephone, email and address details which are not for public display.

Step 6: Defining customised sets of terms

EnrichUK.net decided to use customised sets of terms for elements Topics and Regions. For both elements a two-level set of terms was defined. These were intended to be easily understood by users, to reflect their interests, and – most importantly – the broad subject areas covered by the *nof-digitise* projects themselves.

For **Topics** a two-level set of terms was created. The first level contains thirteen headings:

- Arts & Culture
- Environment
- Health & Lifestyle
- Historical Periods
- Learning & Skills
- People & Community
- Politics & Government
- Our Past
- Our World
- Science, Nature & Invention
- Sport & Leisure
- Working Life
- World Cultures.

All headings have a second tier of headings except Health & Lifestyle which has no lower level headings. Thus Our World has second level headings of:

- Coastline
- Geography
- Bio-diversity
- Rivers & waterways
- Landscape.

For Regions again there is a two-level set of terms. All terms at the first level are for the UK except for the single term Rest of the World. Just two terms, All Great Britain and England, do not have terms at the second level. In effect, second level headings for All Great Britain are the remaining terms at the first level. Similarly, second level terms for England (i.e. English regions) are the remaining terms at the first level. Level two terms for

Rest of the World are restricted to names of the five continents: Africa, Asia, North America, South America and Europe.

Step 7: Database platform

Database

The tables containing collection descriptions, agent records and associated administrative data are stored in a MySQL database. [3] This was chosen because it is available for this kind of application, its good integration with the PHP language and its good performance on look-up (read-only) queries.

Server-side logic

PHP [4] was used to provide all server side processing needs. PHP was chosen as it is a free open-source language, widely supported and well integrated with the chosen database and web server.

Server infrastructure

Apache [5] was chosen as the web server application. This integrates well with all the other technology choices and is the industry leader in web hosting. The hardware is housed at University of Manchester MIMAS [6] hosting facility using Sun Solaris rack computers.

The EnrichUK.net project was managed by Pete Dowdell at UKOLN with additional support from Andy Powell, Pete Johnston and Pete Cliff. Deliverables included the full functionality of the public side of the site plus a secure project area for the addition and editing of collection records on the system and a fully featured administration toolkit allowing NOF to control the site and the database.

UKOLN has ongoing responsibility for maintenance and administration of the site.

Lessons learnt

Data entry

The creation of an input form for collection records is complex, as it requires many multiple entry fields – subject, names, languages, for example. This needs to be carefully considered when specifying a system as this type of functionality can lead to complex issues of maintaining state within a web data entry process, and directly informs decisions taken for the design of the database (multiple entries are usually handled by creating separate tables for these fields and referencing to the collection record; an alternative is to use data separator tokens within a single data field in the database).

Controlled lists

Creation of customised controlled lists for geographic and subject classification obviously creates a problem when cross-searching other databases of collection descriptions, which use other sets of terms. Projects implementing data systems using customised lists should be aware of the issues and allocate sufficient resources to ensure that these classifications are both tailored to the application and easily repurposed to mesh with other standard classification schema. This may require the

