Ecole du Louvre
ICHIM
Paris 03
www.ichim.org

Les institutions culturelles et le numérique
Cultural institutions and digital technology

École du Louvre
8 - 12 septembre 2003

DEVELOPING A KNOWLEDGE SITE IN DISTRIBUTED INFORMATION ENVIRONMENTS : USING A PHASED APPROACH OF HARVESTING, STANDARDIZING AND REPURPOSING

Dr. Ann Borda – Head of Collections Multimedia,
Science Museum, London, UK

Alpay Beler – IS Architect and Web Project Manager,
Science Museum, London, UK

« Acte publié avec le soutien de la Mission Recherche et Technologies
du Ministère de la Culture et de la Communication »

Abstract

The following paper describes an approach to creating a publicly accessible 'knowledge' site ('Science & Culture') by accessing, standardizing, and repurposing digital resources from across distributed environments.

The challenges which this approach encompassed was undertaken in two key phases:
implementing a harvesting model in order to overcome the challenge of retrieving and managing information from distributed source systems and
integrating and delivering this content to the web and to other channels.

Phase 1 involved the development of a batch content hub ('interim database') which was built using Dublin Core (DC) fields as the primary data structure and which served as the central container for export files originating from five separate source systems. The interim database also functioned to normalize data, generate automatic fields and to process data through specific tools.

In Phase 2 of the process, the data records held in the interim database were extracted as XML wrapped DC fields to the web content management system (CMS). The extracted metadata was subsequently integrated and managed for display, building searches and relational linking with other data objects, as well as for supporting personalization functionalities and user tools.

KEYWORDS: Harvesting; Dublin core; information seeking; resource discovery; web personalization; content management

INTRODUCTION:

Science and Culture is a large scale internet project to make publicly accessible 40,000 digitized images and accompanying text records, 10,000 library records, and 10,000 object records sourced from the Science Museum (including the Science and Society Picture Library) in London, the National Railway Museum in York and the National Museum of Photography, Film & Television in Bradford.

The USP of the website is not only to make this rich quantity of material accessible, but to contextualize it through intelligent display, searching (resource discovery) and relational linking. Complementary to this, the site aims to develop user-focused activities and personalization tools that are supported by these resources. [1]

Consequently, the need to effectively manage, standardize and to integrate data sourced from 5 databases in use at three distributed sites for delivery to a web platform (and interface) was central to the challenges facing the project implementation.

PHASE 1: HARVESTING

The main issue on accessing information from 5 systems was accompanied by several issues:

Different types of legacy systems

Different data standards and platforms in use

Diffuse coordination of 'repurposable' content

Mix of non-networked and networked systems

In particular due to the network issues involving the WAN set up and connectivity of the databases across the sites, it was decided not to go forward with Z39.50 or related interoperable processes (e.g. OAI Protocol for Metadata Harvesting) for retrieving content across the databases. An additional consideration was the time in configuration and upgrading of systems at each site to undertake this.

The challenge was further complicated by the roles of the source databases, each having their own distinct functions. For example, three of the databases are also self-contained image management systems and, thus, a component of the solution would require management of digitized image files referenced with related text records.

Table 1: Source systems overview

Location	Source Database	Function	Type of System
National Railway Museum (York)	iBase	Image management	Microsoft Data Engine (MSDE) v. 7.
National Museum of Photography, Film & Television (Bradford)	iBase	Image management	Microsoft Data Engine (MSDE) v. 7.
Science Museum – Science and Society Picture Library (London)	Capture	Picture library database	File Maker Pro + C++
Science Museum (London)	MultiMIMS Y 2000	Collections management	Oracle 8.0 running on a Microsoft Windows 2000 server.
Science Museum Library (London)	Unicorn	Library management	C/ISAM (Informix) with BRS-Search running on SunOS5.7 (Solaris 7).

After examination of the situation, it was recommended that an export-based harvesting model be used to create a central or ‘interim database’ (ie. Interim to the source systems and web content management system), whereby the five databases each have their own separate export routines and reporting tools. In this way, the technical staff at each site also had more autonomy and management of the model without overheads with system configurations and lengthy testing periods. However, it was agreed that future projects would look to a more streamlined and automated process based on the OAI model.

PHASE 1: INTERIM DATABASE

The development of an ‘interim database’ would act as a collection point for data from the various systems. It would not act as a cataloguing database, although a small amount of data cleaning might be permissible.

The primary purpose of the database was to feed the web site content management system [and other future in house applications] with meta-data and digital files.

The logical design would focus on:

Simplicity and efficiency of retrieval of information

Optimisation and consistency of information

Therefore, the database would be a separate entity from the web CMS due to the differences in functionality and purpose, namely to ‘normalize’ data, generate automatic fields and tags and to manage the data through specific tools.

As a result, the database was to be ‘fronted’ by a conversion program as part of the back-end in order to provide translation between the legacy format and the web formats, e.g. in the case of name inversion and/or date normalization.

Significantly, it was held that the database had the potential to develop into a digital media hub which could support future digital resource projects by providing the following benefits:

Conform to future standards (e.g. Dublin Core)

Repurposable – data to be held in an XML wrapper which allows it to be used in a variety of applications : ‘Create once and use many’

Customizable according to needs

Ability to share data across organizations and between institutions

Creating different channels

PHASE 1: DUBLIN CORE

From its early genesis, it was agreed that the database should be built using Dublin Core (DC) fields as the primary data structure and that the export files be compliant with these. This reasoning was also supported by the guidelines provided to the project by its funders – the New Opportunities Fund, which had formulated recommendations on technical implementation. [2]

Secondly, it was seen as highly desirable that all exported data would be more manageable in this form for delivery and repurposing for the web. It would further allow:
more 'value' in resource discovery and relational linking
in using structured data for the creation of personalized environments where digital objects could be profiled utilizing the metadata elements.

It then followed that one of the next major challenges was the mapping of fields to Dublin Core (DC) from across the five systems. Currently DC does not provide sufficient qualifiers to achieve a desirable level of specificity for all of its elements. For example, if there is more than one creator, there is no qualifier to indicate the specific type/role of each creator. Local qualifiers needed to be supplied to add the specificity that was required; however, it was also recognised that these qualifiers may be ignored by systems that do not know about their existence.

Thus, it was decided to extend the existing schema to simplify mapping and to retain a certain level of specificity. This was further required to manage the different record types originating from the various systems.

The mapped fields which would appear for web display were similarly guided by record type, although 'common' fields existed across the records such as 'Title' and 'Description'. An added factor was the need to take into account the user experience and to keep data display intuitive and clear across the resources.

A map of DC elements is provided below showing qualifiers for each record type:

Table 2: Field mapping to Dublin Core & Qualifiers (Mapping to Data Types):

Dark shading represents fields for web display; lighter shading indicates search fields (non-display)

DC ELEMENT	DC QUALIFIER	Image only	Image text	Object	Library
TYPE The nature or	Image				
	Text				

genre of the content of the resource.	Physical Object				
	Sound				
FORMAT The physical or digital manifestation of the resource.		Digital capture	text/xml	text/xml	text/xml
TITLE Typically, a Title will be a name by which the resource is formally known	Title				
	Object name			If no title	
DESCRIPTION An account of the content of the resource.					Rare
CREATOR An entity primarily responsible for making the content of the resource.	Maker				
	Photographer				
	Author				
	Artist				
CONTRIBUTOR An entity responsible for making contributions to the content of the resource.					

PUBLISHER An entity responsible for making the resource available					
DATE A date associated with an event in the life cycle of the resource.	Created				
	Modified				
	Used				
IDENTIFIER An unambiguous reference to the resource within a given context					
SOURCE A Reference to a resource from which the present resource is derived.					
RELATION A reference to a related resource.	Web		Link fields		
	Image		Link fields		
	Text		Link fields		
	Object		Link fields		
	IsPartOf				

	InJournal				Articles
SUBJECT The topic of the content of the resource.	SAC				
	Name				
	LCSH				
	Controlled				
	SHIC				
	SSPL				
COVERAGE The extent or scope of the content of the resource.	Spatial				
	Temporal				
RIGHTS Information about rights held in and over the resource.	Credit				
LANGUAGE A language of the intellectual content of the resource.	NOT USED but could be added automatically				

PHASE 1 : EXPORT FORMATS

A single format was not possible across the 5 source databases currently in use at the different institutions. Only three of the legacy systems could export using XML, and others using delimiters, with the UNICORN (library) records containing MARC tagged fields. Consequently, the conversion program of the interim DB required further tweaking to strip out MARC fields and to deal with the variations among the CSV exports (e.g. rogue characters, formatting, etc.). An XML DTD was created for those systems which could export into XML.

Table 3: Export formats of legacy systems

Location	Source Database	Export
National Railway Museum (York)	iBase	XML
National Museum of Photography, Film & Television (Bradford)	iBase	XML
Science Museum – Science and Society Picture Library (London)	Capture	CSV
Science Museum (London)	MultiMIMSY 2000	XML
Science Museum Library (London)	Unicorn	Tab delimited

Exports coming from the interim database into the web CMS are wrapped in XML and retaining the DC field structures. The DC elements encoded in the record use a shortened form (dc:title, dc:creator, dc:subject, dc:description, dc:publisher, dc:contributor, dc:date, dc:type, dc:format, dc:identifier, dc:source, dc:language, dc:relation, dc:coverage, and dc:rights). Qualifiers are added by inserting a full stop between (e.g. dc:type.image). Published guidelines for encoding simple, and qualified DC in XML were referred to in the process. [3]

The web CMS system itself is built in XML so both the interim database and CMS are compatible and there may be an opportunity to integrate the two systems over the course of the project. However, the primary concern is that XML-encoded data remains an agreed requirement for future data migration purposes, to make resources available to other metadata harvesting schemes such as the Open Archives Initiative (OAI). [4]

PHASE 1: CATALOGUING

Arguably one of the more difficult aspects of Phase 1 was ensuring that the data being harvested from each source database was consistent in style and content, and this led to a number of considerations in moving data from each legacy database to a 'common' database.

Each contributing database has its own rules of cataloguing and its own metadata standards.

Unicorn records conform to AACR2 and are coded using USMARC.
MultiMIMSY records conform to Spectrum (The UK Museum Documentation Standard) and include elements conforming to AACR2

Image cataloguing varied across the three image databases in use.

As a consequence, a standardized cataloguing system was decided upon at the outset of the project and a dedicated team outlined key areas [5]:
A list of data fields for export, their content and format along with DC mappings from the fields in each source database. It was noted that the same type of data can be formatted differently in each database, even though it is held in the same type of field.
Cataloguing terminology for consistent entry should be chosen and, where possible, loaded into the appropriate vocabulary control sections of each legacy system to make terms easier to reference.

Methods for text and multimedia file naming and archiving.

Once data standards and mappings were agreed upon, the data from each system was analysed to determine which normalisation issues needed to be addressed by the cataloguers and which by the conversion program. Normalisation would allow for another important level of standardization and consistency in formatting.

Normalisation was a particular issue with Names and it was necessary to be able to accommodate where possible, an authorised and normalised heading for the interim DB to manage. For example:

Authorised heading:	Watt, James, 1736-1819
Normalised heading:	James Watt (1736-1819)

There was no easy resolution, but it was decided that owing to the several thousand names which would be generated by the project, the Science Museum Library would be the ‘clearing house’ for name authorities. The Library has the expertise, access to an external authority file and the necessary reference material to back up the service.

Similarly, the Library provided guidance with Place Authorities, and the cataloguing teams were provided access to The Getty Thesaurus of Geographic Names (TGN) to use as standard practice, except where the preferred term is in the vernacular. [6]

PHASE 1: IMAGE MANAGEMENT

The management of the image files in relation to the text files was the second largest issue after the cataloguing agreements necessary for the standardization of the text records across the source systems. Three of the source systems (2 iBase and Capture systems) also had image management functions and their text records accompanied image files.

This introduced a range of issues beyond the scope of this paper, but the key was referencing the image files with standardized file names and creating relationships with the text records via dedicated data fields. It was important, therefore, that as part of the interim database build there was also an image management function so that the exported text records would retain a ‘mapped’ reference to relevant image files and this reference would be retained when exported into the web CMS.

The interim database itself was a simplified version of the iBase system (MSDE v.7) as it was already utilizing Dublin Core elements and had core image management functionalities, including the ability to process the large resolution master TIFF images to JPEGs for web delivery.

An overview of the full harvest model is illustrated below showing the interim database in relation to the source systems, web CMS and Storage Area Network (SAN) which holds the image files referenced by the interim database.

PHASE 2 : WEB CMS

The Web content management system, a COM database based on MS architecture and running on MS based server platforms, was specified to handle XML and DC elements, so was an ideal receiver of the digital records and metadata originating from the interim database.
The CMS was also key to managing the relationships among the record types which were encoded in the DC fields and for handling the web image references and files.

In particular, the management of different types of 'links' was critical for cross-collection references amongst the records, for searching, and for context building.

There were two ways data elements were related to one another in the source systems:
The like-subject links are implicit links that connected data elements based on their subject-field content, e.g. records with descriptors containing 'transport' would be linked.
The other are explicit references, which involves placing an explicit pointer in the database that links elements. In this case, they would not necessarily have to have the same subject; links would be created by the cataloguers (i.e. direct linking).

Thus, the DC element DC.RELATION was the key to direct linking between related object, image and library records. These links were entered primarily in the DC mapped fields in the source system and then consolidated in the interim database via purpose scripting so that the relationships across site resources could be made and extracted to the CMS.

These links are usually made to the unique identifiers of other records. Some direct relationships are expressed in the DC.Source element. However the converse relationship goes into the Relation element as shown in Figure 2.

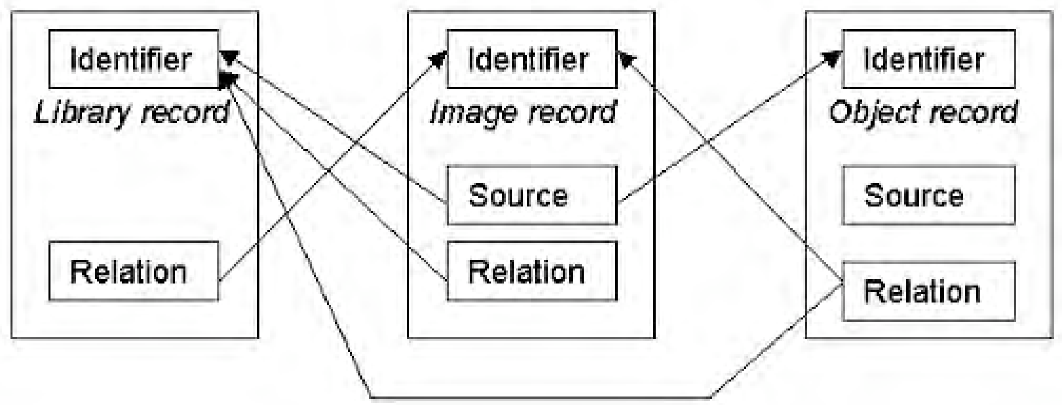

Fig. 2: Relational linking among record types

So, for example, a photograph by Treacy of the Mallard (steam locomotive) would have its own record, but would link to the object record for the Mallard by copying the object identifier number into the image's Source field. But the record for Mallard itself would include the identifier for the photograph (or any other image) in the Relation.Image field. Similarly a book record could have the image identifier in the Relation element linking it to an image from the book. But the image record could then be linked to the book record by using the Source element.

These links will help to guide the user to directly related resources.

Such relation links which could be made where there is a direct relationship, include:
An image from a book: Link book to image record(s) (use Relation.Image).

A book record with another book record (a translation or new edition) (Use Relation.Text).

An object record with image of the object (use Relation.Image).

An object or image record with a book or article describing the object or image (Use Relation.Text).

An object or image with a sound clip (Use Relation.Sound)

An image to another image (use Relation.Image), or an object record to another object record (use Relation.Object) where it is a model of an object (e.g. a model locomotive linked to a photograph of the real thing) or where there are other examples of the same object (e.g. separate photographs of Davy's Safety Lamp). Where intra-database links already exist (e.g. on MultiMIMSY) these should be mapped to this element and qualifier.

PHASE 2 : SUBJECT MAPPING

In an initial prototype stage of the website, it was indicated by test users that a more 'common sense' subject approach should be implemented to gather together and allow searching for resources.

To date, there had been no separate subject control because of the diversity among the source systems, e.g. LCSH, AAT, SHIC. Notwithstanding in Phase 2 of the project, the cataloguing teams based a subject schema on an existing in-house thesaurus based on the curatorial collection names and used by the Science & Society Picture Library (Science Museum). This schema was subsequently modified as a meta subject level for both web display, search and retrieval.

The subject categories became part of a Subject Authority managed in the interim DB and CMS so the source records would be mapped to this on export to the web. The subject category Broad2 is also encoded in DC across the record types and is the principal category for indexing resources for the site.

Table 4: Subjects : Broad 1 and Broad 2 categories (shortened version)

DC:Subject.Broad1	DC:Subject.Broad2
Entertainment & Media	Cinematography & Film
Entertainment & Media	Photography

Entertainment & Media	Television
Entertainment & Media	Visual Arts
Medicine & Health	Anaesthesiology
Medicine & Health	Anatomy & Pathology
Medicine & Health	Classical & Medieval Medicine
Medicine & Health	Clinical Diagnosis
Medicine & Health	Dentistry
Medicine & Health	Nursing & Hospital Care
Medicine & Health	Obstetrics, Gynaecology & Contraception
Medicine & Health	Orthopaedics
Medicine & Health	Psychology, Psychiatry & Anthropometry
Medicine & Health	Public Health & Hygiene
Medicine & Health	Radio Medicine & Body Imaging
Medicine & Health	Surgery
Natural World	Astronomy
Natural World	Environmental Science & Technology
Natural World	Geology & Geophysics
Natural World	Meteorology
Natural World	Natural History
Natural World	Oceanography
Science & Technology	Biology & Biotechnology
Science & Technology	Chemistry
Science & Technology	Computing & Data Processing
Science & Technology	Mathematics
Science & Technology	Physics
Science & Technology	Space Technology & Rocketry
Science & Technology	Telecommunications
Science & Technology	Weighing & Measuring
Society & Wars	Domestic Life & Household Management
Society & Wars	Ethnography, Customs & Beliefs
Society & Wars	Food & Drink
Society & Wars	People & Personal Life
Society & Wars	Sports & Pastimes

Society & Wars	Wars & Wartime
Trade & Industry	Agriculture & Fishing
Trade & Industry	Building Construction & Architecture
Trade & Industry	Fuel & Power
Trade & Industry	Glass Technology & Ceramics
Trade & Industry	Hand & Machine Tools
Trade & Industry	Mining & Ore Dressing
Trade & Industry	Plastics
Trade & Industry	Textile Industry
Transport	Aeronautics
Transport	Navigation
Transport	Railways
Transport	Road Transport
Transport	Water Transport

PHASE 2: RECORD DISPLAY

The metadata elements of all records are exported into the Web CMS. However as Table 2 (DC Mapping) indicates not all fields are displayed. The fields are divided into those that display and those for searching and administrative use (non-display). Those fields which display are common across the record types, such as 'title' and 'description'.

However, as discussed previously there are important DC Relation fields which provide the means for creating relational links ('connections'), and for resource discovery. In regard to the latter, advanced search options, in particular, provide the user with a means for richer searching across the metadata elements.

Fig. 3 A sample record showing display fields

The full XML/DC record which is exported to the CMS from the interim database and subsequently managed to support different functions (e.g. display, search, linking) is managed as both an 'entity' and as a set of metadata elements that can be accessed and processed separately. The Relation fields, for instance, provide the means by which cross-linking and 'reverse' linking to other data objects can occur. An example of reverse linking is the linking of an image and its associated record in a topical narrative (deep linking) and that link then appears in the record so that the user can follow other contexts in which the record can be found. Hence, the user can drill 'up' as well as 'down' through the content hierarchy of the site.

A sample XML/DC record that corresponds with the display record is shown below.

Table 5: XML/DC record

- <dcschema xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="D:\NOF\dcschema.xsd">

- <dc_record>

<dc_identifier>1980-108</dc_identifier>

- <dc_creator1>

<name>Coster, Salomon, d. 1659</name>

<role>maker</role>

</dc_creator1>

<dc_relation relationType="Object">T/1980-108</dc_relation>

<dc_relation relationType="Image">TIM100282</dc_relation>

<dc_relation relationType="Image">10326522</dc_relation>

<dc_subject subjectType="Controlled">pendulum clock</dc_subject>

<dc_type>physical object | text</dc_type>

<dc_format>text | xml</dc_format>

<dc_title>'Haagsche Klokje', pendulum clock, c. 1657.</dc_title>

<dc_description>This clock by Salomon Coster (d. 1659) of the Netherlands is one of the earliest pendulum clocks ever made. The Dutch scientist Christiaan Huygens (1629-1693) designed the first successful pendulum clock in 1656. He worked with Coster, an experienced clockmaker, to apply his new invention to commercial use. In 1657 Huygens had the patent protecting his invention assigned to Coster, but sadly Coster died suddenly in 1659 after producing only a few pendulum clocks. This is one of only seven pendulum clocks made by Coster known to have survived. The application of the pendulum clock to timekeeping during the scientific revolution of the 17th century was arguably the most fundamental advance in the history of time measurement.</dc_description>

<dc_date_created>c. 1657</dc_date_created>

<dc_subject_broad2>Time Measurement</dc_subject_broad2>

<dc_cover_spatial>World > Europe > Netherlands > The Hague</dc_cover_spatial>

<dc_language>en</dc_language>

<audit_ok_for_nof>true</audit_ok_for_nof>

</dc_record>

PHASE 2: PERSONALIZED ENVIRONMENTS

The creation of a personalized environment is one of the imperative aims of Science & Culture, allowing users to custom-build their own user experience by adding resources to a 'clipboard', save tours, create their own links and self-market the site by passing on e-cards, for example.

It also follows on evaluation findings undertaken by the Visitor Research department of the Science Museum whereby the user is no longer simply interested in playing a passive role in information seeking.

To accomplish this, however, it is essential that the metadata elements of the records are consistent, accessible and manageable. Not least the proposed personalization services for the Science & Culture site are to be mainly supported by the image and data records.

Some of the proposed personalization services on the Science & Culture site are as follows:

A number of pages giving a 'personal space' where users' preferences are used to filter / drive content, namely:

My links: The user will seamlessly be able to add a copy of a link to their personal links area. Within their "my links" section they will be able to modify and append a brief description to the link. They will be able to generate a text file of all their links and be able to send as e-mail or download it as a text file.

My search: The user will be able to save search queries seamlessly to their "my search" area. The user will have an option to delete or modify searches from their personalized area.
In addition, the users will be able to search directly from their own personalized search area.

My 'lightbox' facility, where users can view images that they have saved

My 'Gallery' facility, where users transfer select images from their lightbox and create a mini gallery in the form of a slide show, with annotations of their choice to display sets of images, or create stories. This creative building process provides an individualized means of making learning resources of direct significance to the user. The users will also have the ability to e-mail their galleries to friends or to the Science Museum to enter the monthly 'Best Gallery' competition.

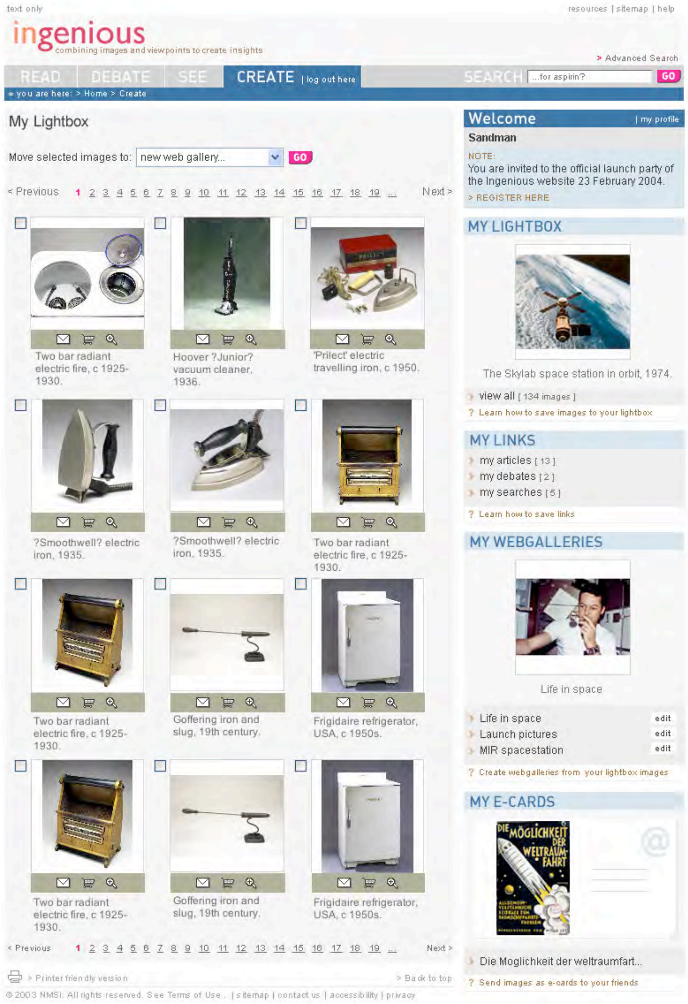

Fig. 4 A screenshot of the 'my lightbox' feature

Importantly this user-focused vision through the use of personalized services not only enhances the richness and interactivity of the site, but are key to longer term sustainability. There are three main areas of strategy through the use of personalization which can underpin a sustainable site; namely

(1) perpetuation of new content; (2) community development; and (3) e-commerce opportunities.

Perpetuation of new content: The users will be asked to inform the museum curators their views and contributions to written articles. The users will be able participate in debates. The users active involvement in content building via the use of My Gallery functionality will aid the site developers in ascertaining a good idea of how the image collections are utilized.

Community development: Phase 2 of our site development intends to build on the personalization and take it one level up to community building; that is to build channels where the users can subscribe and be able to share opinions, images, articles within their chosen channel. The intention is to be proactive with the site content and let the user know periodically what new developments have taken place: what categories of new images and records have been digitized and have been added to our collections; what new topics have been added; interesting debates that have taken place, etc.

E-commerce: This will provide the users the opportunity to buy images in a variety of formats. This can be in the form of prints, film, or digital. There are also plans to build curriculum based e-learning material, which may have commercial value. In addition to this, via an advanced form of personalization, it is being considered to give school teachers the functionalities necessary to create online learning and teaching materials, as well as provide the course tutors and teachers tools to monitor student progress and be able to track their work online.

Use of data mining will be crucial in actively monitoring site usage, in the way functionalities are utilized by the users, the level interaction attained in different site areas, plus the range and sophistication utilized for search. Data mining on the personalization functionalities will be additionally very important in making a decision on how far to take the site forward with more advanced forms of functionalities as mentioned above.

Complementary to this and most importantly, it is the aim of the site to form the basis of a powerful, expandable digital resource from which several ancillary products can be derived, such as CD-ROMs, PDA (personal digital assistant) resources and other multi-channel products that a user or groups of users can readily access and assist in their customization process.

Figure 5: Reversioned web page for PDA

CONCLUSION

In conclusion, the Science and Culture project is an ambitious one, with content rich collections distributed across diverse information environments. There are, consequently, a number of challenges in accessing and managing these resources, not least organizational ones!
Notwithstanding, the approach undertaken in the two phases described in the paper has proven to be a successful one.

At the outset, it was realized the importance of standardization and this has been central to the export harvesting model through to the implementation of the website itself. It was also a consideration to ensure that the source systems benefited from the process and that exported resources were not significantly changed or enhanced during their migration in order that internal use and access could rely on the same quality of information.

As a result, there is a sustainable mechanism in place that future projects can readily build on, as well as a wealth of repurposable content in XML/DC format for other applications and channels to utilize. As illustrated, such content can support advanced resource discovery and web functions like personalization tools which can only be effective with consistent, standardized and accessible resources in place. These are the main steps toward the development of a true 'knowledge' site.

REFERENCES

1.Borda, A. and Bud, R. (2002). Engaging with Science & Culture: Major missions across cyberspace to share good history. Conference Proceedings EVA 2002, London (July 2002), s1-8.

2.New Opportunities Fund. (February 2003). Nof-digitise technical standards and guidelines. (http://www.peoplesnetwork.gov.uk/content/technical.asp)
Last updated February 2003; last consulted July 2003.

3.Powell, Andy. (2002) Guidelines for Implementing Dublin Core in XML. http://dublincore.org/documents/dc-xml-guidelines/. Last updated April 2002. Last consulted July 2003.

4.OAI Executive Committee.(2002). The Open Archives Initiative Protocol for Metadata Harvesting. http://www.openarchives.org/OAI/openarchivesprotocol.html
Last updated June 2002; last consulted July 2003

5.Wyatt, Nick et al. (2002). Science and Culture: NOF Cataloguing Guidelines. [London: Science Museum], 2002.

6. J. Paul Getty Trust. (2000). Getty Thesaurus of Geographic Names. http://www.getty.edu/research/tools/vocabulary/tgn/
Last updated 2000; last consulted July 2003