

Frances Lloyd-Baynes
Documentation Manager and Head of Records Section
Victoria and Albert Museum
Cromwell Road
South Kensington
London SW7 2RL
United Kingdom
Phone: +44 (20) 7942 2626
Fax: +44 (20) 7942 2627
f.lloydbaynes@vam.ac.uk

Integrating for Access: Delivering Aggregated Collections Information at the V&A

Summary

True accessibility to a museum's collections in the 21st century must involve electronic access to its documentation assets. Delivering this requires museums to identify and negotiate end user requirements, to re-purpose material, to maximise limited resources for content creation, and to develop the technology for capture and delivery. Where information lies in disparate systems, ways must be found of joining information sources to both maximise and simplify access to content. The V&A's experiences undertaking the Core Systems Integration Project, which initially aims to aggregate and deliver museum object, bibliographic and archival data, provide a useful example for others engaging with these issues.

Objectives

This paper endeavours to

- describe the V&A's Core Systems Integration Project ;
- explore the technical and content-related issues of delivering via a single interface information drawn from varied systems and created with different underlying approaches;
- offer an example for organisations undertaking similar initiatives by describing the issues faced and solutions being implemented by the V&A.

Introduction & background

The Victoria and Albert Museum (V&A) is considered one of the world's greatest museum of art and design. The purpose of the V&A is to enable everyone to enjoy its collections and explore the cultures that created them; and to inspire those who shape contemporary design. The Museum's holdings of approximately 4 million objects represent over 3000 years worth of artefacts from many of the world's richest cultures and include books, ceramics, fashion, furniture, glass, metalwork, paintings, photographs, prints, sculpture, and textiles. All our efforts are focused upon a central purpose: the increased

use of our displays, collections and expertise as resources for learning, creativity and enjoyment by audiences within and beyond the United Kingdom.

The V&A in South Kensington is part of a family of London museums including The Museum of Childhood in Bethnal Green and The Theatre Museum in Covent Garden. The Museum of Childhood is home to one of the world's largest and oldest collections of toys and childhood artefacts. The Theatre Museum is the National Museum of the Performing Arts. Situated in the heart of London's theatreland, it houses the world's leading collection of material relating to the British stage; all the live performing arts are represented. [Visit the V&A's website at www.vam.ac.uk for further information.]

The V&A collection incorporates the holdings of the National Art Library as well as several archives: The Archive of Art & Design, The Beatrix Potter Collections, the V&A Archive and the Theatre Museum Archive. The National Art Library operates as a major public reference library. Its strength lies in the range and depth of its holdings of documentary material concerning the fine and decorative arts of many countries and periods. It is also the curatorial department for the art, craft and design of the book. The archival collections range from the archives of those involved in the art and design process, particularly 20th-century British design (The Archive of Art & Design), to the correspondence, art work, photographs, literary manuscripts and other memorabilia of the author Beatrix Potter (the Beatrix Potter Collections), and to detailed records of the acquisition and provenance of the Museum's objects, its history and the architectural history of the South Kensington site (The V&A Archive). The Theatre Museum Archive is the main repository for documentation of contemporary performance including the programmes that the Museum systematically collects from over 200 British venues. In all, the collections of the NAL and Museum archives are estimated at approximately 1.5 million items. In addition to this bibliographic material, the Museum maintains a growing picture library of over 150,000 images, both digital and analogue.

In a museum the size of the V&A (800+ staff), a high degree of specialisation of roles is an inherent part of the organisational structure and culture. Four major curatorial departments are responsible for the physical management of the museum (i.e. 'non-bibliographic') objects, while librarians and archivists manage the bibliographic material. Conservators preserve the objects; Registrars manage object loans; Picture Library and Photo Studio staff manage the creation and use of images. A wide range of less museum-specific departments manages finance, human resources, public affairs, visitor services, and other organisational activities. The need to automate the work of such varied areas has led to a proliferation of systems, developed independently to meet individual department requirements. There is no single software infrastructure linking these systems and duplication of content is inevitable. Museum users wishing to understand the life-cycle of a collection object must refer to multiple source systems, though few users would have

access to all of them. In 2005 the Museum initiated the Core Systems Integration Project (CSIP) in an effort to address this situation.

Core Systems Integration Project

Users access the V&A's electronic information via gallery multimedia on-site, via the web, or through mediated enquiries. Data delivery has been achieved to date in a piecemeal fashion. The NAL's bibliographic records, kept in the MARC format, are made publicly available through its database's proprietary publishing solution, effectively a Web-enabled portal into the source database, as well as a Z39.50 server. Since 2003, 'Access to Images' has established a publishing mechanism to the Web for Museum object and image data drawn from their respective source systems. Stand-alone gallery multimedia have been developed, such as the British Galleries Online database, but these are neither connected to the source systems nor directly publishable to the Web. Access to the complete electronic archive files is available to users via collection-level descriptions in the NAL's system, supported by links to XML files. The XML file content itself is not searchable. Those interested in object data would benefit from a single point of access to all the object-related material. The Museum would likewise benefit from a unified facility to 'publish' this information to the Web, the galleries, etc.

To satisfy our users' access requirements, we must provide information drawn from the separate systems, collected and presented as though it was from a single source. However, the Museum's current combination of systems and associated publishing processes are isolated from one another. We therefore require some means by which the data can be aggregated before publishing. The Core Systems Integration Project aims to achieve this. The complete range of project deliverables will include

- Gallery Services application (to be available at the Information Desks and accessed by staff answering visitor enquiries; later opening up access to staff around the Museum, then eventually to the public)
- PROMIS (a replacement project management system, linking to source object systems)
- Publishing Process (a multimedia standard template available for generating object databases/multimedia, with the Web and the galleries as key output channels)
- 'Virtual Repository' (a central data 'switch' or 'repository' through which the various source system are able to 'relate' to each other, and to output information to various applications)
- Data Mastering Protocol (a policy statement that would clearly identify where data items are to be mastered and how the various data items in the Museum relate to one another).

The CSIP will initially focus on making the collections information available to a wider audience. Three database systems hold the primary information

relating to the Museum's collections of objects, bibliographic material and digital images: the Collections Information System (CIS), the NAL's Dynix system, and the Photo Catalogue (now linked with CIS), respectively. The archival finding aids have no database, but are stored in a variety of XML files in EAD (Encoded Archival Description) format (though collection-level records are available via Dynix). These 'source systems' will be the first focal point of CSIP.

During 2003-2004, a series of groups began meeting to consider the current lack of infrastructure and connectivity. The Museum's Information Systems Services Department (ISSD) took the lead, bringing in an external consultant to advise on the best way to achieve system integration. A group of Museum staff from the areas that would be affected by the CSIP were brought together to discuss and agree the project remit. Interviews of interested parties (such as Gallery Services staff) contributed to the dialogue. During the consultation period, the following closely related problems became apparent:

- It is not possible to query the source systems with a single query.
- It is not possible to link data from CIS or Dynix with Web pages published and managed by the Museum's web Content Management System (CMS).
- There is no dynamic link between Dynix and the Photo Catalogue.
- The existing, generic multimedia system is isolated from the source systems.

It also became clear that, in the past, applications have been developed that duplicated data from the source systems into their own local storage. The problem with this approach is that any changes to information stored in the source systems are not replicated in the application, and worse, any definitive information changed in the application is not available to the primary mastering system. The idea behind CSIP is to develop system architecture whereby applications can access object information via a Virtual Repository rather than mastering object data locally. Ensuring that the source systems (CIS/Photo Catalogue, Dynix, EAD XML files) are the only places where object data can be mastered will ensure that the latest information regarding an object is always available. Integrating the Museum's core systems will also remove the dependencies on manual data manipulation tasks inherent in current practice, thus improving efficiency and accuracy of data delivery. These problems must be addressed in order for CSIP to succeed.

The project consultant recommended in his report that the Museum would achieve the best results for CSIP working with its current system suppliers. If the suppliers were interested in the project, they would be able to deliver integration better than someone less familiar with the existing systems would. As Dynix is presently being replaced by Horizon software and the Horizon suppliers were not ready to undertake the proposed CSIP approach, the Museum turned to System Simulation Ltd (SSL), London, suppliers of the CIS/Photo Catalogue system. SSL proved interested in the project and delivered their thoughts on how CSIP might be achieved in August 2004. By

October 2004 the Museum had decided to go ahead with SSL. After several months spent refining the proposal, the contract was signed and in February 2005 work on CSIP began in earnest. A selection of V&A staff responsible for collections information management and systems within the Museum are now working with SSL to deliver the CSIP outputs. [A list of CSIP working group members appears at the end of this document.]

Virtual Repository

The first step in delivering systems integration to the V&A is the creation of the Virtual Repository (VR), which will enable access to the Museum's core data through a standard interface, allowing the heterogeneous systems to work together as a single web of computation. The Virtual Repository will be based on SSL's Index+ Bridge, interfaced to the source systems (CIS/Photo Catalogue, Dynix, Archive XML files). [Note: Index+ is SSL's proprietary software.] The VR will provide its services using the SOAP (Standard Object Access Protocol) standard. A common data model, developed by V&A staff, will support data aggregation. Data will be either harvested (archival material) or retrieved from the source systems using the Z39.50 protocol (CIS/Photo Catalogue, Dynix), though other search/retrieval protocols (e.g. SQL) could equally be employed. Should the performance of these protocols become an issue, the Museum can move to harvesting all of the data.

The Virtual Repository services are designed to support Application Layers, which utilise the VR and deliver services to end users [either individuals or systems]. Two Application Layers will be developed in the first CSIP development phase:

- A layer to support the Gallery Services application
- An image server for NAL's Dynix/Horizon database.

The Gallery Services layer prototypes a website that will demonstrate how the relevant VR services are used in a website context. The 'Images for NAL' layer is a system interface that provides images in a way that can be used by the NAL Web OPAC (Online Public Access Catalogue). The Virtual Repository is intended to simplify the interface to the Museum source systems, enabling further client applications to be developed more rapidly and cost effectively in future. The first two Application Layers will provide a proof of concept for the Virtual Repository.

In order to deliver aggregated data, the Virtual Repository requires a Common Data Model (CDM) for mapping the source systems. The CDM must be sufficiently detailed to support the first application layers (Gallery Services and Images for NAL), as well as any additional applications we hope to develop in the near future. The data model must be in place before the application layers can be implemented. In practise, it is being developed in parallel with the first two applications.

Gallery Services application

Gallery Services, comprising over 100 Gallery Assistants, Supervisors and Managers plus around 90 volunteers, is the Museum section tasked with delivering a quality visitor service across the main site front-of-house areas. These staff will be the primary end users of the application, though the V&A plans to make it available to all Museum staff as well as to the public (for both on-site and Internet queries) in future. The Theatre Museum Front of House team also plans to utilise the system. The objectives of the Gallery Services application are to

- meet the needs of Gallery Services staff , and
- open up access to information regarding the Museum's bibliographic and non-bibliographic objects through a single interface.

As noted, the Gallery Services application will also provide a proof of concept for the Virtual Repository, from which it will source its data.

V&A Gallery Services staff have communicated their requirements for the application through a series of interviews. Their key need centres on the provision of access to information on objects (i.e. 'non- bibliographic' material). Access to the objects' current location is vital, enabling staff to offer visitors guidance on items on display, and those available through the NAL, the archives or via appointment with the curatorial staff. [Note: few NAL and archival objects are on display throughout the Museum at any one time, but are generally available on request.] Gallery Services staff also need to access object details through 'surrounding information' such as peoples' names (e.g. artist/maker; people depicted), periods, places, dates, materials, etc. Frequently these requirements will overlap, e.g. visitor queries on 'all the Art Deco silver items currently on display' (= style/period + material + location). The information is needed quickly as during the Museum's busy periods the amount of time available to spend with each visitor is necessarily limited (although the ability to access more detailed information as required is considered desirable). The need to answer queries quickly also constrains the volume of material that can be returned via search queries and the level of detail a staff member can reasonably sift through. Illustrations (i.e. digital images of the items) are extremely valuable in helping visitors identify objects of interest, particularly when objects are not on display.

Although similar information on bibliographic material is also required, this need is less pressing for the V&A Gallery Services staff (who largely answer queries on display items). The requirements of the Theatre Museum in this regard are quite different, however. Their primary collections are documented on both CIS and Dynix (with the NAL's material). Thus, any search potentially covering these collection objects must focus equally on both systems. (The majority of electronic archival records that the Theatre Museum has appear on Dynix as collection-level descriptions. Very little is as yet available in EAD XML files.)

Future plans for the Gallery Services application include opening up access to staff around the Museum, then eventually to the public. While the simplicity of the initial interface would suit a general public user, it will not provide sufficient detail or depth for a more experienced user or specialist query. The project team is grappling with the level of detail to make available at the outset to end users and the point at which a user must switch to the source system for more specific querying. (Of course, at this point users would lose the cross-system search capability.) The decision taken in this will affect the level of detail we provide for the Common Data Model. Too limited an approach now will restrict the detail that can be provided in future applications. Under the current proposal, changes to the Common Data Model are not envisaged as a common or regular occurrence. While it will be possible to modify the data model, it would entail changes to all the applications drawing content from the Virtual Repository as well as to the data gathering done from the source systems. The V&A must, therefore, develop a data model now that will support the anticipated range of applications, such as PROMIS and a loans management module.

While we plan to map the various data initially to the Dublin Core standard, the Common Data Model creation is beginning with a focus on data access points rather than a survey of data fields. The access points provided to application users will differ from the data retrieved and presented, though to what extent is still to be defined. The simplicity of the user interface requested indicates that Gallery Services users do not want to be burdened with a complex search screen. However, a fairly simple approach to searching/retrieval would prove necessary anyway, both because of the varied structures of the source systems and the different granularity of their content. For example, CIS captures people's names and makes explicit their roles/relationships with an object, either in an associated field (e.g. Artist/maker + Role) or via a distinct name field (e.g. 'Donor'; 'Associated person'; 'Person depicted'). Each of these fields is indexed individually and authority controlled so access is possible at each data point. In the NAL system roles and relationships of names are distinguished at the record level, but are currently only accessible through a single name index where authors, subjects, publishers, donors, etc., are indexed together. The user is left to distinguish the role/relationship with a bibliographic object by reading through the record content and cannot refine the data selection through searching. The Archival data will be indexed as it is harvested, based on the EAD tags in the XML files. However, its approach to names will also result in broader groupings than available in CIS.

Issues

The Core Systems Integration Project faces a wide variety of issues such as the disparate granularity mentioned above, which need addressing. Below is a list of some further issues with which we will be grappling in the coming months.

- The hierarchical nature of the archival findings aids. Unlike the data held in CIS and the NAL systems, the archival files hold data in a hierarchical format [fonds, sub-fonds, series, etc.]. Each component is described in such a way as to minimise duplication of data from above or below it in the hierarchy; but the consequence is that any descriptive component retrieved and presented out of its hierarchical context will not make sense.
- Item level vs. collection-level descriptions. When searching across the systems, users will be faced with results drawn from a search against individual items (CIS; NAL data) and collections (archival files; NAL). Making the results understandable for the end user will be a challenge.
- Authority-use variation. The CIS relies heavily on authorities (thesauri, name files, term lists) to control and enhance access to its data. These authorities (including the Getty's Art & Architecture Thesaurus) allow cataloguers to document objects very specifically, while the authority provides broad access searches (e.g. 'London' is catalogued; search for 'United Kingdom' finds this data element via hierarchical Place thesaurus). The NAL maintains authority control over primary access points and makes use of a number of thesauri and term lists, including the Art & Architecture Thesaurus and Library of Congress Subject Headings. However, as the hierarchical relationships between terms are not embedded in the architecture of the system, optimum retrieval depends on cross-references within indexes and user navigation.
- System structure vs. User expectations: The Gallery Services interface must provide inexperienced computer users/non-information specialists with consistent, meaningful results to searches. The varied structures of the systems and approaches to data capture mean that a single search (e.g. for items of a given date) will pose three very different queries to the source systems and retrieve quite different results. Again, the challenge rests in making the search results understandable for the user.
- Appropriate mapping model for CSIP: While the V&A are initially mapping to Dublin Core, we expect this standard to be too simple to meet our needs. We are considering use of the CIDOC Conceptual Reference Model (CRM). One benefit of using CRM would be that a map of the V&A's object data already exists (created for the EU-funded SCULPTUER project). Likewise, mappings exist to the EAD and the Functional Requirements for Bibliographic Records (FRBR) Library standard, which can then be mapped to MARC. SSL plans to discuss the CRM implementation with SCULTEUR partners IAM (University of South Hampton) to learn from their experiences.

Project Status: March 2005

At the time of writing, the first phase of CSIP development has been underway for 2 1/2 months. V&A staff are working in parallel both to define the Gallery Services application structure and interface, and to formulate a Common Data Model for use in the Virtual Repository. The main issues that have arisen during these processes are reflected in this paper. Plans to deliver the prototype Gallery Services application are on target for summer 2005 delivery. An update of the project highlighting the V&A's agreed approach to handling the initial CSIP implementation will be made in May at the CIDOC 2005 conference in Zagreb, Croatia.

Core Systems Integration Project working group

Doug Dodds, Head of Central Services, Word & Image Department

Frances Lloyd-Baynes, Documentation Manager

Christopher Marsden, Word & Image Department

Alan Seal, Head of Records & Collections Services

Judith Smart, Systems Manager, Word & Image Department

Bryan Solomon, Development Project Manager, Information Systems
Services Department (Chair)

Seonaid Woodburn, Database Administrator, Records & Collections Services

Documents and sites referenced

Material for this paper was largely drawn from the following documents:

- Core Systems Integration Project: Requirements. Version 1.01. Victoria and Albert Museum, 2004.
- V&A Core Systems Integration Project: Phase 1 Proposal. Version 1.1. System Simulation Ltd., 11 October 2004.

The following websites were referenced:

- Victoria and Albert Museum www.vam.ac.uk accessed 18/03/2005
- The Apache Software Foundation <http://ws.apache.org/SOAP/> accessed 20/03/2005
- National Information Standards Organisation www.niso.org.z359.50/z3950.html accessed 20/03/2005
- International Organization for Standardization <http://www.iso.org> accessed 20/03/2005