www.ichim.org

Les institutions culturelles et le numérique
Cultural institutions and digital technology

École du Louvre
8 - 12 septembre 2003

GETTING COLLECTIONS INFORMATION TO NEW AUDIENCES

Ian Croxford - Victoria and Albert Museum, UK

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

Abstract

In common with many other large museums, the Victoria and Albert Museum has a variety of information systems aiding the management of its collections. The appetite of new audiences for interactive displays in galleries, access to collections on the web and via digital partnerships puts new demands on this pool of information and the systems and processes that maintain it. Generally the object information was not created with public access as a primary purpose. Scholarship may have changed in the 150 years the museum has been collecting objects. Much of the contextual information that the public require, which is second nature to the professional curator, is lacking. Simple things such as "what was it used for?" are not recorded. What everyone wants is high-quality high-resolution images, which create issues of copyright, and consume gargantuan amounts of disk space and network bandwidth.

The V&A has piloted several approaches:

Development of new databases for the British Galleries project, which met critical acclaim, yet consumed large resources of professional time in regenerating the information from scratch. These have been difficult to re-purpose for the web, and are not seen as being sustainable in the long term. Duplication of data leads to duplication of effort and the risks of inaccuracy and inconsistency.

Portals based on Z39.50 technology, and now XML can provide joined-up access for museum or library professionals, provided the operational systems fully support them and the skills-set and knowledge exist, but are not suitable for the general public and give concerns for security and performance.

Participating in Consortia such as SCRAN and AMICO have allowed us to get object data and Images onto the web without investing in the infrastructure, but at a high cost of manually repurposing data and images.

Duplication of a cut down version of the main Collections Information System and Photo Catalogue with a web front end have allowed us to put 14000 selected images and object information on the web or the first time, but this is seen as an interim solution that

is not easy to scale up to representative coverage of the whole museum collection.

Based on this experience we have a good understanding of the technical, operational and human issues involved. Our current project is the creation of a digital repository that will "harvest" information from the operational collections management systems, in a normalised form, combine this with contextual information from the current gallery projects, and make this information available 24 by 7 for publishing via the various channels:

Interactives in the Galleries.

In "identical" form via the web as part on the Online Museum.

Via the Collections Information Gateway – a new service enabling the public to get consistent information on the V&A's collections in person, by telephone, letter, email, the web and any other means of communication that may become appropriate.

Through the proposed wayfinding system - part of the Future Plan for the museum.

Through new applications using handhelds or other devices once they've proved their value and applicability.

The main object of the exercise is to come up with a process that is sustainable in costs of technology, support and most importantly professional time.

Our mantra is "write once, use many".

Keywords: Collections Management, Interactives, Web Publishing, Digital Asset Management, Databases, Warehousing, Portals

Introduction

The Victoria and Albert Museum is now embarking on an ambitious Future Plan which includes re-organisation and redisplay many of its collections in a way which is more

understandable and attractive to 21st century audiences. The first phase of this plan was the redisplay of objects to form the British Galleries 1500 –1900. This project which opened in November 2001, achieved critical acclaim ad let to the V&A being named European Museum of the year in 2003.

Integral to all of these Future Plan gallery projects is the use of multimedia to enhance the display and interpretation of the objects, and a study area allowing access to a database giving a full range of information about the objects in the Galleries.

This led to the immediate problem which is the subject of this paper. How to get access to the vast range of knowledge and scholarship that the Museum has built up on these objects in 150 years and present it to the various audiences in an appropriate way that will keep them informed, educated entertained and engaged?

Problems

Several areas of difficulty were identified :
Diversity of sources and systems
Context
Language
Learning Styles
Photographs
Sources and systems

In the mid 90s the V&A embarked upon a major Information systems project to acquire a new collections management system. After an extensive detailed and thorough exercise to define the operational requirement and a full tendering exercise. System Simulation Ltd. (www.ssl.co.uk), won the bid and developed the V&A Collections Information System, CIS which they then marketed as MUSIMs. This is a very sophisticated system which ensures the accurate and precise recording of information by the use of authority files. If all objects were catalogued in CIS it would be relatively simple to extract data from it for use in Gallery Interactives and on the web. However many collections are still catalogued

in the databases, Modes and Filemaker which were in use before CIS was developed, Loading these into CIS is proving costly and time consuming as:

Not all mandatory fields for CIS have been catalogued in some collections.

Matching data against the authority files can be laborious iterative process as new authority records have to be created manually.

There is a reluctance to just load old data, and the natural perfectionism of many curatorial staff means that scholarship often has to be checked and re-researched.

With many other priorities on time and resources, this task is often neglected until either a gallery project, or a directive makes it a priority.

One of the major collections in the V&A is the National Art Library, who's objects are books, and for whom a more traditionally library system, Horizon from Dynix Systems (www.dynix.com) was seen to be more appropriate. The National Art Library (www.nal.org.uk) have been successful in getting funding for electronically cataloguing their entire collection. Thus the NAL collection in Horizon is better catalogued electronically than any of the other collections, but the database structure is very different. New galleries such as the British Galleries use objects from all collections including books from the National Art Library.

However the biggest Information system for object information is still paper. The Victoria and Albert museum has been collecting for 150 years and it is only in the last 25 years that computers have been available for cataloguing of inventory of objects. Thus less than 20% of the total collection is catalogued in any electronic system. The vast majority of information is still held in paper registered files, many of which were handwritten. 19th century copperplate handwriting is still likely to defeat OCR technology. The best hope for back converting this data at a reasonable looks to be a combination of scanning, and typing in the data using third world labour, but at the moment the cost of doing so and the concerns about the quality of the information means that this project is still on the too difficult list!

Context

Typical catalogue entries in a collections management system contain plenty of detail about the object, what its made of who the artist was, where was made, sometimes a lot of information about its history and provenance. What tends to be lacking, since it is obvious to any serious art historian or curator, and yet what our audience research shows is the most frequent thing the public wants to know is "what is it?" "what was it used for?" and "how was it made?".

This basic contextual information is less important when dealing with things for which the audience is familiar, e.g. contemporary objects, items in regular use such a tea pots, or paintings or other purely decorative and ornamental objects where anyone can understand or have an equal opinion of its purpose. Where we particularly have a problem, is where the postmodern audience no longer has any understanding of the social, political or religious culture of the society and period when the objects were created, or even when collected.

For the British Galleries it was relatively easy, as most of our target audience have some knowledge of the period of British History, and many of the objects would be familiar to a modern audience. For our next major project which is the redisplay of our medieval and renaissance collections the task is much harder. The medieval period is not well understood, and where it does appear in popular culture, the dark ages leads to the worst misconceptions. Medieval art is considered by some to be an oxymoron. Much of our medieval and renaissance art is religious in origin. Our challenge is to give the audience an insight into the world view of the society and people who created them in order to understand the object better, and be entertained, educated and engaged. Our existing data is sadly lacking in this direction.

Also for a museum with a long history we have the problem that society's views have changed in the period that the museum has existed. Some galleries such as the cast courts or the church plate still reflect the ideas and values of the Victorians who created them. It was assumed at the time that every educated person would understand classical mythology and be familiar with Church ritual. In the 21st century this is not the case, and

we run the risk of excluding those who lack the educational background and cultural understanding. Unless we wish to only cater for the upper middle aged, upper middle classes then we need to do more to explain, and it is an opportunity for new media to do so, providing we can get the information into the databases and link it to the objects in a meaningful way.

Language

Catalogued entries are generally written by curators for other curators to read. They are highly researched, accurate often detailed descriptions. However they are not written in a form that is easy to read, and often miss out on elemental things like punctuation. For use as label text, or in a system for public access they are often considered useless, as they do not meet the quality control of those who protect the museums brand and reputation. We thus have the awful problem that many items are rewritten several times, first by a curator for the facts, secondly by an educator who writes for our target audience educated 12 year olds, and thirdly by an editor who ensures consistency of style.

This is acceptable for new galleries where there is a lot of funding available, but cannot be sustained for the whole museum. Therefore two approaches have to be adopted.

Firstly we have to train curatorial staff to write the original text in a form which is acceptable for public consumption. Generally there is now wide acceptance that the world wide web is her to stay and we now have a management which recognises public access to our collections and encourages staff to get the information right first time so that more object information can be published.

Secondly we need to recognise that the vast amount of data we have but which does not meet the quality standards is still of great interest to the worldwide research community, and we should find a way of publishing it without re-writing and re-researching every entry. For several years the IT enthusiasts have been proposing that we should open up “the vault” a web site of data that is taken on an as is basis, without the V&A brand of quality of research or presentation. However this is still to become museum policy, and

the vast majority of the V&A's collection will continue to be unavailable to virtual visitors.

Finally our other great embarrassment is that our collections information is only available in English. We have looked at automatic translation, but it makes a pretty poor job of descriptions of art objects. The only solution that is considered good enough for the V&A brand is to get native speakers to translate. For the vast majority of our collections this is too expensive, and it has only been done for those collections which are considered ethnically important.

Learning styles

Our collections information is organised for serious researchers and academics who like nothing batter than lots of text that they can wade through at their own pace. Our audiences have different styles. Many like to passively sit in front of videos that linearly take them through the subject. Some like to channel hop. Some like to mine deeper and deeper, whilst others have more active styles and like to participate and interact. We have to cater for all styles and provide interfaces and pathways through our information to meet all needs.

Photographs

What everyone wants is to visualise the object. Ideally they actually come to the museum and see the objects for themselves. But in the new virtual world we have to cater for the audience who cant do this. Also for many objects there is problems of access. We can’t let people see the underside or the back, or much of the fine detail unless we use the photographs. Often things are revealed during conservation which are of great interest. In a big museum such as the V&A people will use object databases to plan they visits and need to see what objects are worth walking hundreds of metres to see.

In short everyone wants to see the picture as well as the caption and contextual data. Frequently they want to see different views at different resolutions. This leads to two unique problems. Firstly the photographic images and increasingly video and audio consume orders of magnitude more storage and bandwidth that the textual databases. Secondly, in the V&A and other museums they are held in different systems, often managed by different staff. Bringing the two together has been a major priority in the last year. We have managed to get our photo catalogue and our Collections information into the same data base, however all the records still need to be linked by hand and this is taking some time.

Photos have similar quality issues to data. Many photos are taken for conservation or inventory reasons are not publication quality. Many photos that are publication quality still need to be cropped, straightened or have colour bars removed before being given to the public. Of course the majority of photos are still on old analogue media. There is still some resistance to publishing photos on the web as it may undermine our copyright and photo library income, but most have come to the confusion that our images have already been pirated on the web from scans of books, so we'd be better giving the real thing (but only up to 768 by 768)!

Solutions

The British Galleries Database

For the British Galleries project it was decided to create a new database for the project rather than modify existing systems. The purposes of this database were three-fold. Firstly it was used to project manage the project, tracking all the objects in the galleries as they progressed through selection, conservation until final installation in the gallery. Secondly it was used to manage all the text for gallery labels for objects, interpretive panels and the main gallery themes. Thirdly this information was used to feed the British Galleries Online interactive that was developed by Oyster Partners along with the other award winning interactives in the British galleries. For more details see BROD(2002).

The main advantage of this approach was that it was successful, and the gallery opened on time to critical acclaim.

The disadvantage was that it was labour intensive, as all the information was re-created from scratch and re-worked extensively. A new database island of information was created that was separate from the main collections systems, only the inventory information on location of objects being linked, and even that was a manually intensive process.

The worst limitation was that the British Galleries Online interactive that was written in JAVA for standard windows NT, Internet explorer could not be used as part of the online museum, the main museum web site that was relaunched at the same time, November 2001 as the Galleries were opened. The problems were simple, insoluble, but could easily have been avoided with hindsight.

Firstly none of the images of loaned objects had been copyright cleared for use on the web.

Secondly non-standard fonts had been used by the designers, and any attempt to map them onto standard fonts results in a total mess.

Thirdly the layouts were designed only to work on a 1024/768 screen in kiosk mode.

Finally, in the galleries all PCs have 100Mb/s access to the powerful dedicated server. With lower speeds some of the features such as magnifying the photographs become unworkable.

Because we could not get the BG information on the web, we had to look to third party help, and so we had to export a subset of our data to SCRAN,(www.scran.ac.uk). This got our data online ,a kept the Director and Trustees happy, but was once again a manually intensive process, and is frozen in times as the SCRAN database is not updated as objects move in and out of the gallery.

Later we also joined AMICO (www.amico.org) and the BG data was used to supply our submission, again with a lot of manual manipulation of data to get it into the right form.

Despite the project being finished the British Galleries database is still being maintained, as objects are occasionally changed in the Gallery. Also some effort is being made to complete the data that was inevitably missed in the effort to complete the project. In particular some Artists Biographies are still outstanding nearly 2 years after the gallery opened.

As the system was written specifically for the British Galleries, and hard links to the Conservation system meant that effectively conservation work on non BG objects could not be tracked and managed, we had to hastily develop a more generic system PROJEX to replace the British Galleries Database for use with other smaller projects. This is a temporary system with limited functionality that is due to be replaced by the Gallery content database (see below). In the meantime PROJEX and BGD still have to be maintained, and the data in these systems is separate from the main Collections Information System.

Improvements to CIS – Access to Images

Having got the British Galleries finished, and finding ourselves with a situation that was not ideal, and not sustainable in the long term, we set about rectifying matters.

The immediate priority was to feed the web with high quality images of objects with good catalogue records that was better than that available through SCRAN or AMICO and represented a wider range of the V&A collections than just the British Galleries. We were committed to getting 10,000 object images and text on the web by April 2003. This replaced an earlier “Images Online” which had 2000 low resolution images, and text from yet another stand-alone system.

Having learnt the costly mistake of trying to keep data in lots of separate stand alone systems, the decision was taken to use the main Collections Information system as the

source, link it to the photo catalogue by merging the databases and back load all the data from British Galleries and Images online. This project had a very tight timescale, required a phenomenal effort from IS staff, SSL the software suppliers, Collections, Photographic and Curatorial staff as each object description and photograph needed quality control and approval, and often re-working. The result was achieved on-time, under budget and is something we're very proud of. You can see it at http://images.vam.ac.uk. Now the system is being driven from the main Collections Information System, more data can be published simple by ticking the approval flag, and thus the target is to get another 10,000 objects live this year. This should be a comparatively painless exercise as whenever an object goes through the "gateway" i.e. is photographed, has conservation work, is loaned out, or returns then the records should be updated to the level suitable for approval for public access.

Thus the ideal option for any museum that can achieve it is to have a single Collections Management system which is also suitable for gallery projects, and public access via interactives and the web, and to have the procedures in place to keep the data fully up to date. Unfortunately a museum with the size and history of the V&A just can't achieve it. I look forward to meeting anyone from a similar organisation who has.

Using the main collections management system for live public access has three basic problems.

The first is one of security. The system contains information which is highly confidential such as donor details, and details concerning the physical security of the objects.

The second is one of performance. Our internal systems are sized for internal usage which is reasonably predictable, as our own staff, volunteers, students and associates are limited in number. The web in particular is subject to the whims of success. As we discovered when we launched the Art Deco site on the web earlier this year and became Internet magazine website of the week, visitors increased ten fold and brought the server to its knees. It is not acceptable to have main operational systems brought down in this way.

Thirdly there is vast amounts of information in the collections systems which we don't actually want to bring to the new audiences at present as it just isn't good enough.

The answer to all these problems is to clone the system and put the second system outside the firewall. A regular filtration or harvesting process selects that data that is suitable for publication and copies it to the public server, which can often be treated as sacrificial, as it is regularly refreshed from the live system. This is how our current Access to Images works.

An added advantage of this is that the public server can be hosted externally if internet bandwidth or reliability are a problem, or even duplicated for additional performance and security.

In fact we use an additional staging server so that updates can be verified before release to the public.

Portals and warehouses

At present for the V&A a single source system with all the functionality for all purposes is not an option. Therefore we need a publishing process which brings the data from all the source operational systems and brings it to the target audiences.

There are two competing middleware technologies that we are considering.

Data Warehouse

The traditional 1990s method of bringing data from disjoint legacy systems is to build a huge database using a RDBMS, defining a schema which is common to the source systems and developing update process to harvest the data from the source systems, and provide reverse update processes. Typically one does not need to harvest all the attributes, and in the V&A's case we have identified some 20 fields as being those of common interest to all systems. Data warehouses can be developed quickly and cheaply as the warehouse does not require the complex functionality for user updating, with such

features as locking, 2 phase commit etc. The drawback is that it is yet another database to maintain, and there is the risk that data can get out of step and inconsistent.

Portals

Since all of the operational systems as XML compliant, or due to be in the next major release. In theory all we need to do is define an XML schema instead of the RDBMS schema, and use a XML message broker instead of the database. The advantages are that the data is held in the source systems, and is therefore current, and does not require additional storage. However the source systems are then at the mercy of demand as above, and if the source systems are taken down for maintenance then the data is unavailable, which is not acceptable for 24 by 7 web applications. The answer to this is to cache the data in the portal. We then have essentially the same logical solution as the warehouse only using an XML database, rather than a SQL RDBMS.

Illustration

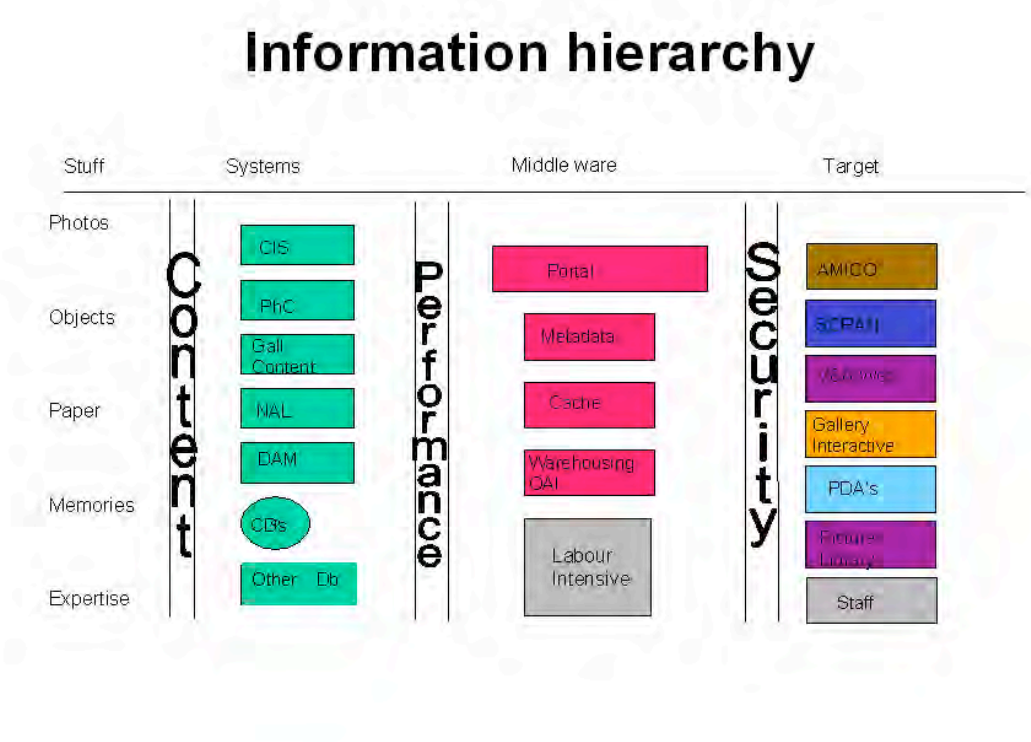

Figure1: A schematic of the Information Hierarchy

Gallery Content Database

The most important IT development for 2003/4 will be a Gallery Content Database that will meet the bulk of the needs of the current and future Future Plan projects, that builds on the excellent work and systems developed for the British Galleries, and obviates the need for future one off developments, and reduces the maintenance effort.

There are three main components:

A project management database to enable the Project team to assemble and manage the objects and other information panels, stories, videos, packs etc necessary to create the Gallery. This would take the existing British Galleries database, and its generic derivative PROJEX as the prototype and develop further functionality on top of it whilst making the code more sustainable.

A gallery information system to inform members of the public via kiosks, and any other suitable access media about the contents of the Gallery. The functionality will be similar to British Galleries Online, but a key requirement will be for the system to generate automatically from the main database, with an absolute minimum of reworking, and the ability to cope with multiple access channels must be built in from the start.

An information system to inform members of the public via the web, or from other locations not immediately in the Gallery. This is similar to the system above, but has to allow for the fact that the user cannot easily get to see the original objects.

Essentially 2) and 3) are the same with the vast majority of code and all of the content being common. The only differences are presentational, such as the use of big buttons for touch screens, different clients within the museum and some variation in audiences.

Principles

It is vital that data is handled only once. So that when an object is put in the gallery, or the text on a panel is changed it should be entered into the common database, and be available to each system. Where data is controlled in a source system such as CIS for location, then this data should be fed automatically. A key feature of this project will be to map where data is mastered, and where it is derivative. If data is entered into the Gallery Content

database it should be fed back into the master system at the earliest opportunity. One database should be used for all galleries and projects. The data should be available in an open manner for future projects. The master database must be 100% available, and be self-consistent and not dependent on other operational systems. The Gallery Content Database is a publishing mechanism, taking data from source systems and making it available via a wide variety of delivery channels.

Why this is necessary:

All future Future Plan projects will need systems that provide this functionality. It will be too expensive to develop one-off systems similar to those developed for the British Galleries for each project, and far too expensive to maintain multiple systems

Other benefits

Once there is a critical mass of content in the database, which should be achieved when the Medieval and renaissance Galleries are opened, then the data can be used for other purposes:

Access to Images developments – The object content developed for the new Gallery projects will be able to be used in Access to Images without extra effort. We could if we wish use some of the additional contextual information with very little additional development.

Partnerships such as SCRAN and AMICO can use the data from this database directly.

New Portals or virtual galleries can be easily created using the database and re-using much of the data. This will allow different groupings of objects to the physical galleries.

The Information Gateway will be able to use the data directly.

New multimedia systems using hand held computers and radio networks would be able to be created using the existing content.

A future orientation system enabling visitors to navigate their way around the museum in order to locate specific objects, panels etc will rely on the content held in this database for locating objects.

Content

There are three main types of content in this database.

Catalogue and Inventory Information

For museum objects this will come directly from CIS, which is the primary source of object information. A facility has already been developed by ISSD and SSL together to "harvest" the information from CIS and keep Oracle databases updated. This runs as a background task and will ensure that the Gallery Content Database is always up to date with CIS to within a few minutes. This is in line with the Open Archives Initiatives standard for synchronising different systems. For loaned objects the data can be input directly. Object information should be catalogued in CIS. If for some reason it is entered directly into the Gallery Content database, then it should be loaded back into CIS as soon as possible.

Images

It was originally the intention that Images would be held within the Gallery content database as the primary source. As images are needed at a variety of different resolutions and formats for published catalogues, panels, interactive displays, web delivery, thumbnails etc. The system would have to provide facilities for manipulation of images. In terms of complexity this facility would exceed all the other functions put together in cost of development and maintenance. Therefore it is proposed that a Digital Asset management system be procured to handle all the images and other digital assets, such as videos, audios and 3D models that are required by the gallery content database. Thus the Gallery content database will only need to hold the references to these images in the Digital Asset management system, and copies of those images need by the Gallery Interactives and Web version for efficiency. (See below for further discussion on the Digital Asset Management system).

Contextual Information

The database will also hold the other interpretative items such as panels, stories, videos, themes etc. necessary to produce the content for any Gallery. This is likely to be limited initially to the functionality of BGD and PROJEX, but is likely to grow considerably in future developments, when funding is available and we learn with experience form each Future Plan project.

Options

We are keeping the specification deliberately open, concentrating on functionality so that we can compare all the available technological options, and not risk upsetting any vested interest groups.

The Gallery content database could be:

A development of one of our existing systems, such as the current CIS or BGO.

A new data warehouse using Rdbms technology (Oracle and/or MySQL)

A pure XML based portal.

A development based on specialised software designed for the purpose such as Cognitive Applications COBOAT.

Perhaps more importantly, the specification exercise we are now going though is mapping the processes involved in publishing collections information and is identifying where in the complex interrelated V&A systems data is mastered and where subsidiary copies are held, and what sort of currency of updates is required.

Digital Asset Management

The items most remote audiences are interested in are high quality photographs of the objects. In terms of Gigabytes of storage used, Digital Images current consume more space on V&A servers than all the other items combined. There are now 1700 CDs of

high quality digital images (30MB TIFF) that cannot be loaded on the servers as we haven't the space. Still the bulk of our photographs are held on traditional photographic media, and we have a large room full of Umatic video tapes of broadcast quality video that has been shot at V&A events over the last 20 years. We are starting to do 3d Imaging of our objects.

We need to manage these digital assets, come up with a preservation strategy, and make as much as possible available in the Galleries and via the web, as well as to our own staff who are crying out for access to good photographs.

A particular issue in the last two years has been the manual effort involved in manipulating images, resizing, cropping, colour balancing and watermarking to make images available on our own web site and through partnerships such as SCRAN and AMICO.

At the same time our Picture Library V&A Images is developing its own web presence and needs to be able to deliver pictures electronically to customers.

We clearly need an off the shelf Digital Asset Management System to do al this, and work alongside the Gallery Content Database. As our next door neighbours, the Natural History museum have exactly the same issue, we have decided to pool resources and progress a joint project to acquire a system, and we have put a bid into the UK Treasury Invest to Save programme to try to obtain funding for it.

The Future

Once we have a stable platform to manage the information relevant to our audiences in the Galleries an online there are two obvious developments.

Data loading

The first is to try to hit the 80%+ of collections information that isn't even computerised. Some of this will be picked up as the Future Plan re-orders and re-displays galleries. But we are now making funding bids for back conversion of old data. Some like our Prints drawings and Paintings catalogues are known to be in good condition and could be

handled relatively easily. Others will need either researching from first principles to bring the data up to standard, or the decision to release poor quality information on an "as is" basis to an alternatively area of the web site that is referred to as "The Vault". I would welcome feedback from others on this issue, of whether it is better to enforce quality in the pursuit of excellence, or whether freedom of access to information should be the overriding concern.

New media

The second area is to make different access channels available. We are avidly watching the progress on the use of handhelds within other museums and Galleries, and still hope that we can make them available when the next major gallery, Medieval and Renaissance Europe opens in 2006/7.

Information gateway

We are establishing a central point of contact for finding out about all information in our collections. This is adopting a more user orientated approach, than our traditional collection and object based approach to giving information, since experience shows that the public often ask us about object we haven't got (usually those in the British Museum), or the favourite question "When is Queen Victoria's birthday?" Our existing systems cannot answer either of these since neither the Elgin marbles or Queen Victoria are actually objects in our collections.

We are also considering whether collections information can be incorporated into a new wayfinding system, since finding the gallery where the objects we have got are can often be difficult for members of the public, and even our own staff at times.

References

Brod, N. (2002) Diving in at the deep end: the British Galleries at the V&A. Museums and the web 2002 Proceedings
http://www.archimuse.com/mw2002/papers/brod/brod.html