

The Future of Museum Scholarly Catalogues

by Nik Honeysett

Museum Informatics: Something New, Something More

EDITOR'S SUMMARY

Museum scholarly catalogues have been the mainstay for delivering authoritative collection information, capturing rich illustrations and reflecting research. But rising costs and shrinking distributions are fueling a move from analog to digital. The Getty Foundation is exploring building a framework for seamlessly compiling and publishing catalogues directly online from a museum's collections management system. Such a shift is a sharp departure from tradition, with numerous cultural and curatorial implications requiring support from catalogue creators and consumers and involving new image licensing considerations as well as production contributors with new skill sets. Collections management software will need to support authoring and publishing needs. The Getty is using the new process to create both a print and an online version of a collection catalogue. In the process, it is reimagining the museum scholarly catalogue as it transitions from a product to a service.

KEYWORDS

museums
 museum informatics
 electronic publishing
 authoring software
 special collections
 museum collection management systems
 museum catalogs

Nik Honeysett is head of administration for the J. Paul Getty Museum. Prior to his current role, he managed the Getty's Web Group. Before moving to Los Angeles in 2000, he spent 12 years at Cognitive Applications, a United Kingdom-based consultancy building interactive kiosks, CD-ROMs and websites for museums and galleries. He can be reached at NHoneysett@getty.edu or through <http://www.getty.edu>.

The scholarly printed catalogue, critical to a collecting museum's mission, provides authoritative information about collection objects and is characterized by years of research and rich illustrations. Print catalogues form a building block of art history but are faced with increasing print costs, declining print runs and a cultural shift towards digital consumption embedded with rich engaging media. These challenges have been on the horizon for a number of years, and so in 2008 the Getty Foundation brought together nine institutions for a five-year initiative to explore the potential for scholarly collection catalogues transitioning to the online environment. Details of the Online Scholarly Catalogue Initiative (OSCI) can be found on the Getty Foundation's website (www.getty.edu/foundation/funding/access/current/osci_fact_sheet.html).

This initiative is not about simply creating a PDF as the conclusion of the editorial and design process. While PDF and print-on-demand versions are clear digital publishing options, this initiative is about creating an economic and sustainable framework to publish digital catalogues directly from data, taking advantage of the environment and options that the medium provides. With this initiative we wish to enhance the relationship between our catalogue's research and publication phases and establish a delivery environment that promises broad access, timely content and rich, potentially interactive, media. At its most fundamental, this change means publishing from our collections management system (CMS), a core information management application which is the central repository for information about works of art – everything our museum needs to document, catalogue and describe our objects.

Challenge

At first blush, one might assume that this transition from analog to digital is relatively simple. Museums are already in the business of creating and

publishing digital content. We have web staff and new media departments, and our traditional collections information departments are now being given pre- or post-fix descriptive tags denoting emerging new media options such as “digital collections information” or “collections information and access” departments. Surely, switching to digital is only a matter of merging the publication and new media departments or creating a workflow pipeline between the two, right? Unfortunately the solution is more complex, because this switch involves some cultural, business, financial and technological paradigm shifts.

The ultimate benefit to an online catalogue relies on taking advantage of the opportunities that the new medium provides. It requires making the editorial process easy, integrating with our existing architecture, having the appropriate resources to support it, ensuring that comprehensive access is provided and that what we create is better than what it replaces.

The tradition of a printed scholarly catalogue is long established, characterized by that visceral *thump* when it is dropped on a desk as the manifestation of a curator’s work. More importantly, curators are challenged with how online works that may not be obvious as their endeavors translate to the scholarly world. For example, the author’s identity is clear from the linear experience of reading a book: one sees the name every time the book is closed. But in the non-linear world of the online catalogue, where is the author’s name? Online publications appear much more the product of the institution than of the authors. Furthermore, how do scholars refer to them in their resumes? How do they send complimentary copies to fellow curators or trustees? How do they gauge acceptance of publications within the curatorial community when “copies sold” is no longer a measure? How do online publications support the road to tenure? These concerns are important cultural questions that, as a field, we have yet to properly answer.

There are further curatorial questions that determine how we architect our digital publication infrastructure and long-term implications for its maintenance and support, not only of the editorial system that we put in place but also of the delivery mechanism. Is a curator’s contribution to a digital publication ever finished? How often should curators update it? If they re-attribute a work of art, does that change immediately flow to the

publication? What functions should be made available with the publication to allow scholars to use it? The answers to these questions are in large part institutional policy questions but inform some fundamental requirements for the authoring, information management, data flow and delivery of our digital publishing environment.

The issue of curatorial buy-in is critical, since this transition involves communicating its value to two audiences, the creator and the consumer. In as much as Getty curators are helping to define how they create scholarly content going forward, we are also defining how it will be consumed. Identifying compelling reasons to transition, outside of economic and financial ones, will determine our success and this process is part of the broader issue of museums coming to terms with what it means to be a cultural organization in an online world, elegantly summed up by Sir Nicholas Serota, Tate director:

“...the future of the museum may be rooted in the buildings they occupy but it will address audiences across the world – a place where people across the world will have a conversation. Those institutions which take up this notion fastest and furthest will be the ones which have the authority in the future. ... the growing challenge is to ... encourage curatorial teams to work in the online world as much as they do in the galleries.” [1, 0:52:43]

This transition to digital publishing ultimately requires a fundamental rethinking of our scholarly catalogue production process and many of its traditional skill sets. Our current pilot project to produce an online publication has added a collections manager and a software programmer to the traditional mix of curator and editor plus a web developer and web designer and the services of a server administrator. We are still contemplating the maintenance and support issues for as long as we make it available.

A mainstay of scholarly catalogues has always been the use of comparative images, and their licensing is well established. However in this new paradigm, many institutions including ours are struggling with how the traditional and neatly packaged print-run license and periodic renewal for a popular catalogue, translates to a request for what may be unlimited access for an unlimited period of time. Many licensors (such as institutions and artists) are unwilling to commit to this open-ended license so our

requirement is to either choose comparative images that are in the public domain or implement some form of digital rights management mechanism that will alert us if an image is nearing its term limit.

Solution

To publish from data and create a rich online resource that elegantly supports long-form text and might include links, audio, video, animations and other rich media, requires a rich authoring environment and a rich publishing mechanism. While many museums are well versed in creating rich, educational interactive experiences within their galleries, these projects are often sub-contracted out to third-party vendors and are rarely properly integrated with the institution's CMS. There are exceptions, but for many CMS applications authoring and publishing was never a requirement, and vendors struggle to address these emerging requirements. Consequently, for us to publish from data requires that we integrate an authoring and publishing platform.

The Getty uses The Museum System (TMS – www.gallerysystems.com/tms) as our CMS, and our architecture additionally includes a digital asset management system (DAM – digitalmedia.opentext.com) and a web content management system (WCMS – promote.autonomy.com/components/pagenext.jsp?topic=PRODUCT::TEAMSITE), forming a best-of-breed approach to our information management. All three applications are enterprise-wide and loosely or tightly integrated through APIs or backend data and asset synchronization. As we planned our initial foray into the online scholarly world, we struggled with two major decisions inextricably linked to each other, the first concerning tools and processes, the second concerning content.

The first decision was which environment we should use to support this new production process. Our options ranged from using the authoring environment we have in our WCMS, deploying a new authoring and publishing framework such as Drupal, or implementing some other solution that more closely mirrors the existing print editorial process that our curators and editors are using. The second decision, about content, centers on whether we identify an existing print publication to turn into a digital publication or pick a project that has yet to start.

There was no easy or obvious answer to either, but our strategy has been guided by a number of factors:

- This transition does not require an immediate, all-encompassing solution;
- We need to experiment at all stages of the process with possible throwaway work;
- The initial publication must be big enough to test the process and functionality, yet small enough to complete;
- We are as non-disruptive as possible.

Our decision on the question of content was to pick a current project, and we chose a mid-stream catalogue project from our antiquities department on ancient ambers, with a plan to create a print version (shop.getty.edu/product/967.html) as well as an online version. (Check the Getty's digital publications at www.getty.edu/museum/publications/digital.html.) The decision on process and tools was much more complicated. Even though we are looking to pilot a framework for an ongoing program of digital publication, we have larger issues to consider.

Our CMS is central to how we intend to deliver our catalogue, but our strategy has to be broader and more comprehensive because we understand that, like other museums, we are becoming a mass medium institution, as Tate Modern director Chris Dercon has observed. [2]. Rather than retrofit a scholarly catalogue authoring and publishing tool around our CMS, we need a comprehensive solution to support publishing curatorial, educational and marketing content from a collection of data sources to a diverse range of formats, platforms and media channels. Our online scholarly catalogue is one of many publishing formats that would also include mobile apps, blogs, websites, RSS feeds, wall labels and others. The investment to date in our CMS, which is used campus-wide, precludes a replacement project, so our challenge is how we can augment its functionality, not only to provide rich authoring and publishing, but also to boost the complexity of the data relationships we can establish. Our solution to this dilemma is to construct a digital object repository (DOR) and, after reviewing various web-based publishing platforms including Drupal, we have chosen to augment the authoring and publishing production with the Django [3] web application framework. This Django instantiation will take input from our CMS via our DOR.

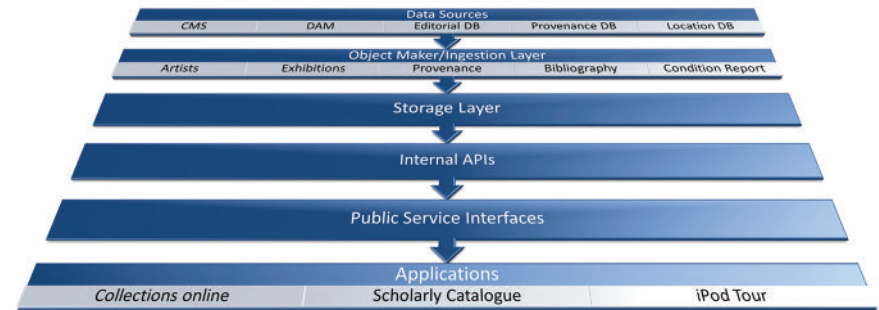
Our DOR implementation consists of a suite of services and serializations that create a sustainable and flexible way of extracting data from sources in defined data modules and deploying them to defined targets, possibly transformed in some way. It is constructed as a logical middle layer in our information architecture and allows us to create any number of object packages and establish any number of relationships between fielded data and packaged objects. It copies essential data from our CMS, DAM or any other data source, staging it as the latest information and formatted according to our business rules and style guidelines. This function will ensure that all the various applications have access to the same set of information in a more maintainable and sustainable way.

Django is an open-source web application framework based on the model-view-controller architecture, which provides a modular approach to application creation and delivery. With it, we can create a framework that is independent of the content and will provide a replicable model for further online catalogue publications. Django is built in Python and was originally developed in an online news environment, so it provides some intrinsic support for the demanding processes and workflows around content publishing.

While our ultimate goal is to tightly integrate our CMS, the Django publishing framework and our DOR into the editorial process, our immediate requirement is to publish a pilot catalogue online using an existing manuscript. Our ultimate workflow calls for a marked-up manuscript that would sit in a layer of our DOR, having drawn core information from our CMS, as input for our catalogue delivery tool. We intend to build support for the authoring process that would result in a marked-up manuscript at a later stage, but for our current purposes we shipped our manuscript to an XML mark-up service with a schema that we can ingest into our Django application. The ingestion process reads in the content, understands where images are based on the mark-up and creates the delivery publication.

Figure 1 conceptually summarizes the digital object repository stack. Services ingest relevant data from our varied sources such as the CMS or DAM to construct packages of objects; for example, metadata from images in our digital asset management system could combine with artist data from our CMS to form a package of data listing works by an artist. Internal APIs

FIGURE 1. Conceptual digital object repository stack



house the business logic of how these packages can be put together (for example, a set of artist packages are combined to describe the set of images and artists represented in an exhibition), and then public service interfaces allow applications access to that data wrapped in the appropriate format such as HTML or XML. Our Django framework sits across these layers.

Looking Ahead

Our ultimate goal is to ensure that we are creating a sustainable, reproducible mechanism as opposed to one-off publications for our catalogues and that this mechanism can sit alongside numerous other mechanisms for authoring and publishing content. To achieve this goal we need to be creating all our content, including scholarly prose, as data in an homogenous system rather than as piecemeal endeavors and discrete documents. However, we acknowledge that this time is one of transition and are cognizant of the needs of our curators and editors, who are most familiar with a word processing environment. While our initial pilot uses a document format as its input, the process is reproducible in that any required change to the content happens first within the document. The revised document is then re-ingested into the repository and application, a process which takes about 30 minutes. As we continue supporting documents as inputs to our process we will have the additional overhead of a document management task.

Over time, we will look to create a much more seamless environment and process to produce and update our catalogues. Any such transition into

HONEYSETT, continued

a fully data-driven environment will also need to implement an editorial review and signoff workflow to mirror the current processes we have for our print catalogues. We will also improve the functions that we intend to provide within the delivered catalogue itself for scholars in pursuit of their scholarly research activities. This task requires that our technology staff create elegant support tools and that our curators become more familiar and comfortable using applications to generate their research and write their content.

Conclusion

This paper reviews the Getty's attempt to answer the fundamental question: What is an online scholarly catalogue? We attempted to answer the question

at the start of the OSCI initiative, but the only way we can really answer it is through an iterative and experimental process, informed in many ways by this quote from the *Future of the Book* blog: "A book is no longer a physical object; it is not what it is, but how it works ... it is a way of communicating." [4]

As we make this transition, conversations that traditionally happened within the confines of a publications department are now happening across the entire institution and forcing a dialogue among groups that have never had to interact before. It is clear that our challenge is less about making the transition from analog to digital and more about the transition of a product to a service. ■

Resources Mentioned in the Article

- [1] MacGregor, N., & Serota, N. *LSE Arts and Thames and Hudson 60th anniversary discussion: The museum of the 21st century*. [video]. (July 7, 2009). Retrieved December 31, 2011, from www.youtube.com/watch?v=tVhXp9wU5sw.
- [2] Millard, C. (October 24, 2011). "It's about disruption": Tate Modern director Chris Dercon on the Art Institution of the Future. *ARTINFO International Edition*. Retrieved November 19, 2011, from www.artinfo.com/news/story/38902/its-about-disruption-tate-modern-director-chris-dercon-on-the-art-institution-of-the-future/.
- [3] For more information on the Django framework see www.djangoproject.com/ and [en.wikipedia.org/wiki/Django_\(web_framework\)](http://en.wikipedia.org/wiki/Django_(web_framework)). Both retrieved November 19, 2011.
- [4] Booknotes [blog post]. (January 28, 2011). *futureofthebook.com*. Retrieved December 23, 2011, from www.futureofthebook.com/2011/01/booknotes-74/.