UC Berkeley

UC Berkeley Previously Published Works

Title

Capture and Release: Access to Library, Archive, and Museum Collection Information

Permalink

https://escholarship.org/uc/item/5ch6m9x7

Author Elings, Mary

Publication Date

2012-09-01

SESSION 4 ACCESS TO COLLECTIONS

Capture and Release: Access to Library, Archive, and Museum Collection Information

Mary W. Elings

Introduction

In thinking about access to collection information in libraries, archives, and museums, clearly we all suffer from the same concerns about making our collections available to the broadest audiences. We all share the desire to bring all of our collection information together in one unified system and make that system easily accessible to users. Whether it is a library or an archive in a museum, or a museum in a library or an archive, the challenges remain the same.

With respet to access and our goals of getting collection information out to users, both internal and external, the visual metaphor of a dandelion spreading its seeds comes to my mind. In this metaphor, the seeds of the dandelion represent bits of information that spread from our information and access systems. As the seeds spread, new knowledge can occur where the seeds touch down, where they then grow and spread more knowledge over time. In this metaphor, the dandelion represents the collections information we capture as institutions, that we can then share with or release to, others, propagating knowledge, interest, and potentially new scholarship in the process. Just like the dandelion, knowledge—if we share it—can be a beautiful thing, with a tenacious will to propagate.

Mission and Goals

In a discussion of access, it is important to look at our missions and goals. Our primary mission as libraries, archives, and museums is to provide access to collections and collection information. We provide this service in various ways: through exhibitions, programs, websites, and catalogs, and in person. In the present era, as discussed in the opening session of this symposium, our missions are changing from being primarily about education—being the voice of authority—to a more interactive relationship with our audiences, promoting exchange and sharing voices. A quick look at several library, archive, and museum mission statements currently online shows that they present their goals and aims in terms of letting "knowledge grow from more to more" or "to stimulate appreciation" and "advance knowledge," "to increase and diffuse knowledge," and to "support teaching, learning, [and] research."¹

With words like "stimulate," "promote," and "support," our roles are becoming increasingly about facilitating knowledge building and encouraging intellectual activity at whatever level it may occur around our collections. And with this comes new scholarship and new forms of scholarship, which we may also want to collect and provide access to, as well as curate and interpret.

In addition to all this, we are also responsible for the proper care, preservation, and management of collections. This is done largely through intellectual control, reflected through our collection information management practices. These practices are well codified, but they are also changing as our roles and information technologies change, which is the focus of this paper.

Strategies

When collections moved from private cabinets of curiosity to public entities, we began to develop organizational and classification systems to manage and provide access to them, which was the beginning of the divisions we now see in the libraries, archives, and museums we know today.

Since then, we have developed professional practices around the collections and information we manage. We have created information standards for our collections, including data structures like MAchine-Readable Cataloging (MARC), Categories for the Description of Works of Art (CDWA), and Encoded Archival Description (EAD), and content standards like Describing Archives: A Content Standard (DA:CS) for archives, Cataloging Cultural Objects (CCO) for museums and visual resources, and now Resource Description and Access (RDA) for libraries. We use controlled vocabularies like the Library of Congress vocabularies, the Getty vocabularies, and various term authorities. We have specialized XML formats and protocols for sharing our records, as well as high-level data frameworks or conceptual reference models like FRBR and CIDOC-CRM to help us model our data.

Libraries, archives, and museums were early adopters of technology to aid us in this work, transferring classification and organizational constructs to the computer, and then to the Web. We started putting our collection information up as records online and then began to digitize some of our holdings. We first



posted them as static HTML pages, then as marked-up pages, then as database-driven websites, then as XML-driven websites. When Web 2.0 tools came along, we started putting digital collections and collection information into interactive venues to improve access and engage users. We have used blogs, tagging, commenting, Facebook, Flickr, Twitter, Tumblr, YouTube, Pinterest, Wikipedia, and Wikimedia to put our content where our audiences are working and playing. We continue to experiment with new technologies to improve searching, share more data, and find better ways to provide access to information and engage our users.

We have also collaborated with our fellow information professionals, inside and outside our institutions and across our professional divisions. We have developed consortiums and partnerships to take advantage of the skills and expertise of those around us. We have collaborated with users in social networking venues, letting them help us by tagging or commenting on our collections and letting them begin to share in this role of "voice" in limited ways. These strategies have been largely successful and have helped us move forward in meeting our missions, but we still have work to do.

What Did and Did Not Work

Our professional practices have resulted in sometimes rigid silos of practice and content that don't always play well together physically or virtually.² We have often tried to integrate library, archive, and museum collection information in institutions into single-access systems with the goals of improving access, increasing efficiency, and reducing costs. Unfortunately, either the technology or the cultures were not ready for the change, and our differences in professional standards and practices have proven challenging.³ True data unity within our institutions has proven elusive, and our systems still remain largely fragmented. The Bancroft Library, as part of the University of California Berkeley Library, maintains its collection information in a campus-wide integrated library system (ILS), yet we also have a content-management system for our digital collections, and a separate system for locally managing our archival information not handled by the ILS. The Bancroft's newest research center, The Magnes Collection of Jewish Art and Life (for all intents and purposes a museum within a library) has a single library-archive-museum system that accommodates the diverse holdings; however, it remains a single system within a larger institution that uses several other systems. While the Magnes as a center maintains a single system, the layers of our complex organization make integration impractical, if not impossible. Our institution, like many others, is an example of a large interdisciplinary organization that is integrated physically, yet our access systems are not yet fully integrated.

Aggregation, which allows us to pull our information records out of local systems and share them either across an organization or with a content aggregator, has been more successful than single-system integration. Aggregation leaves data in the library-, archive-, or museum-specific systems-which can be necessary to serve internal needs-yet accomplishes the goal of unifying data in one place to serve access needs across an institution and for users outside the institution.

Institutional-level efforts like the Smithsonian Collections Search Center, which pulls data from across multiple Smithsonian libraries, archives, and museums into a single search interface, have been successful for internal staff and external audiences.⁴ Similarly, the large scale cross-community effort Europeana, a European-Union-funded aggregation, successfully pulls collection information and content from 1,500 libraries, archives, and museums in twenty-seven countries. Providing access to 2.4 million items, this aggregation has succeeded in doing what many complex institutions hope to do, but it has done so on a massive scale.

Despite these successful examples, aggregation in general is difficult to achieve with current tools and has significant associated costs. The work required to bring disparately structured records together in a single system is resource intensive, and, as a result, funding such efforts has been difficult to sustain.⁵

Reaching outside our professional communities through partnerships to help us improve access has proven helpful in some cases. For example, we never would have been able to digitize and provide access to the quantity of books that Google and the Open Content Alliance did in such a short time. More recent projects like The Google Art Project and The Google Cultural Institute are also helping us get visibility for our collections and collections information at the network level that we could not achieve on our own.6 These non-library-archivemuseum collaborators can bring technology muscle to the table that we simply do not have, but these limited partnerships will get only some of our information out there and for only so long.⁷ As memory and information institutions, we need to look for solutions that will deliver all of our information to the network level and serve us over the long term.

What Do We Want and How Will We Get There

Though we have long-established, well codified professional practices, rich data standards, expertise with technology, and a willingness to work together, we still have trouble sharing data to expose our collection information directly on the Web, makacross our institutions and our communities, and at the network ing it more accessible and useable than our tidy records tucked level. What we want is access that crosses all of these lines, from away in our local systems. The fact that we have well structured local to the broadest possible reach. data, however, is a really good start. Libraries, archives, and mu-If I had to assign a title to 2012 in our communities, it seums are approaching linked data with a long history of managwould be the year of linked data in libraries, archives, and ing information in structured ways and a lot of good source data museums. Jon Voss, a gentleman whom I will call a linked-opento move this forward. This is simply the "next big thing" in our data evangelist, was a plenary speaker at several library, archive, information management evolution.

and museum conferences in 2012.8 He also organized the successful LODLAM Summit in 2011, and another is scheduled for 2013.9

Linked data is not completely new, having already arrived on the scene in 2006. It is being implemented in Europe and the United States in domains such as the sciences, healthcare, publishing, government, and multimedia. Tim Berners-Lee (credited with founding the World Wide Web) has often discussed it, but it took Jon Voss to make it relevant to libraries, archives, and museums.¹⁰ And we seem to have gotten the message. The promise of linked data is that it may be able to solve some of the problems we have had in the past with uniting our data and making it available in various contexts, from our local systems to global systems.

What Linked Data Does

Linked data links information resources together out on the Web. In the past we linked together webpages, and these webpages might have links on them, and perhaps some images, captions, and other content. While these pages are generally easy for humans to comprehend, they are not always understandable to computers. The relationship of a particular link or image to the text or a caption may not be explicit, so a machine cannot process or understand it. Linked data makes those relationships explicit. Linked data uses web technologies to express these relationships, which in turn makes them machine understandable and actionable.

If you think of thousands of data points—like the seeds of a dandelion—living out on the Web, each of those points can be our queries based on the information we make available. linked together, creating a web of data. We can link bits of our For example, take the painting by Johannes Vermeer titled data together and publish them on the Web where others can Woman with a Lute, held by the Metropolitan Museum of Art. then use them or build access on top of them. In basic terms, all The painting has several triples or relationships that can be exthe information on the Internet becomes a web of linked data pressed. Woman with a Lute has type "painting." It has artist "Vermeer." It has city "New York." It has museum "The Metpoints expressed in the form of simple relationships. Libraries, archives, and museums have a massive amount of ropolitan Museum of Art." Finally, the painting has depiction in well structured, standards-based records or "metadata" about a thumbnail online.

the collections, but this information does not get exposed or tra-On the Web, we can see this example as linked data in DBversed in the ways that linked data can. Linked data will help us pedia, which is a linked dataset exposing data from Wikipedia

How Linked Data Works

The concept of linked data is that one thing is related to another thing through a simple relationship, and these relationships are the basis for linking data to other data. These relationships are expressed using the Resource Description Framework (RDF).¹¹ The relationship is called an RDF Triple, because it has three parts: there is a thing, which has a property, and that property has a value. An example of this would be the book War and Peace (the thing), which has an author (the property), and that is Leo Tolstoy (the value). This relationship is also referred to as a subject-predicate-object relationship. The relationship in RDF is expressed using an XML syntax to create the triple. Linkages between data are made through these sorts of triples.

Each part of the triple has its own unique identifier, called a uniform resource identifier (URI). This unique identifier points to a location on the Web. The links from one URI to another use the Hypertext Transfer Protocol (HTTP), so each part of the triple has an active Web address. For example, the book War and Peace has a unique identifier: http://www.worldcat.org/ oclc/918470. The property for author has a unique identifier: http://dbpedia.org/ontology/author. The person Leo Tolstoy has a unique identifier: http://viaf.org/viaf/96987389/.

The pages to which these URIs point are not necessarily user friendly. Linked data is not meant to be presented in a pretty webpage, but rather as a computer-friendly page, meant to give the computer the information it needs to do its job. Computers can read and understand this information and respond to

and other sources. In the record for Woman with a Lute in DBpedia (http://dbpedia.org/page/Woman_with_a_Lute), we can see that the painting has a number of triples: there are several abstracts, a link to a thumbnail, the artist, the city, and the museum, among other relationships. By selecting the artist link "Johannes_Vermeer," users are taken to the DBpedia page for Johannes Vermeer where they see information on the artist and much linked data about him, including his birth date, birth place, by whom he was influenced, who trained him, and, further down the page, other works to which Vermeer has an "artist of" relationship.

Almost all of the data in DBpedia came from Wikipedia. Imagine if that information could be augmented by collection information from the museums that own these works? In theory, we have the best, most accurate data on our collections tucked away in our information systems. Consider what the record in DBpedia might look like with that data added to it and links made to abstracts and images held by the Metropolitan Museum of Art? Consider the relevance of that information to a user: priceless.

And the Point Is

The point of all this is that this linked data is already being put out on the Web. As cultural information professionals and content providers, we want to start to build services on top of linked data whether we put it out on the Web or use it to bring together our own collection information and digital resources on our own site. There are several examples of people doing this. Jon Voss has launched a project to pull together Civil-Warrelated data to make connections between events, people, and content that have not been made before.¹² The goal will be to develop new resources like Conflict History, a site that plots historic conflicts on Google Maps using data from Freebase, a massive open dataset.¹³ The Vistory mobile application pulls videos from the Dutch Open Cultuur Datasets provided by the Netherlands Institute for Sound and Vision to allow users to view historic videos near their location. It then lets them take a photograph with their mobile camera overlaid with a video frame to create a new visual resource.14 Europeana is also using linked data in their semantic search prototype, which helps to disambiguate and contextualize search terms. The semantic search prototype helps users by differentiating, for example, between "Paris" the place and "Paris" the person, and segments search results by these and other related entities.¹⁵ The linked data allows the search to associate entities and put them into context in the search and the results.

These examples from the cultural heritage community are

just a few of the projects currently under way, and new ones are coming online regularly.¹⁶ The point is that linked collection information data can be used by an institution or, if open, by others. We can build curated sites or Web portals, create our own integrated collection searches, let others create new applications on top of our data, let it be linked to other data, and/or surface our collection information at the network level in semantic search engines. The possibilities are exciting.

Benefits

Linked data can give us more flexible data than ever before because it is atomized. Think of the data as building blocks that you can connect, build into something, reconnect, and build something else. It makes our data more agile and easier to use or more intelligently mash up into different applications. Europeana has released its aggregated data as linked data as part of its strategy to make its data more flexible and interoperable on the Web on a massive scale Some institutions are working on using linked data triples to deconstruct and then reconstruct their structured records on the fly.

The broader benefit to linked data is that it will make our information more findable and ultimately more accessible using new semantic Web tools. It will be bringing together not only our data but data from across the Web, which has the potential to facilitate greater discovery and intellectual activity. This will be even further facilitated if we can make that data open.

Not all data is linked and not all linked data is open, so there is a movement to make all library, archive, and museum metadata (not necessarily all digital content) available as Linked Open Data. The benefits of putting our data out on the Web with an open license means that it is much clearer to the users of that data what they can do with it.

As our missions change to make us more outward-facing, collaborative partners with our audiences, as was discussed in the opening session of this symposium, we should be opening up our data. In this way, those who have interest in our content will have the freedom to innovate, create, and experiment with it in ways we either cannot because of a lack of resources and staff, or have not even considered possible. Making our linked data open can encourage more use and reuse, as well as applications being built on top of our data, so we can finally have our collection information truly playing at the network level. As we continue to make new digital research collections available and link this open data and content on the Web, we will be able to stimulate, promote, and support new forms of research, and ultimately new forms of scholarship.

Getting Started

How do we begin moving in this direction? The implementation of new standards like Resource Description and Access (RDA) in the library field and the use of Lightweight Information Describing Objects (LIDO) in the museum field are already moving us in the direction of linked data principles.

RDA, which is built on the theoretical framework expressed in the Functional Requirements for Bibliographic Records (FRBR) conceptual model, was created with linked data in mind, so it is already positioned for this change. With the additional movement toward a new bibliographic framework to own institutions and at the network level. ultimately replace MARC, the wheels are already in motion for We are already moving in this direction and have the well structured data we need to be slightly ahead of the game. The libraries and archives to move more and more into the linkeddata realm. LIDO, a harvesting schema for museum description potential benefits may be even more interesting and gamechanging than outlined here, and I wanted to close with another designed on semantic principles, was developed as an application of The CIDOC Conceptual Reference Model (CIDOC-CRM), image as a vision of where we might be headed. A "digital dandelion" from the San Diego Super Computer Center that a semantic framework for cultural heritage information.¹⁷ Like RDA for libraries and archives, LIDO will move museum data shows a linked data graph of the Internet.²¹ The future of access is here, in the linked environment. in the linked-data direction.

The more data we put into this environment, the more vis-Several library, archive, and museum entities, such as the Library of Congress, the National Archives, the Rijksmuseum, ible we will become, and the more our data will be used. Some of the work we want to accomplish with our collections cannot OCLC, the National Library of Spain (BNE), the British Musebe done by us alone; we will need partners and innovators who um, and Europeana, are now exposing their data as linked data. will do this work with us, alongside us, or on their own. We need More are forthcoming. The Web is already transitioning to linked data, so we need to take the risk of letting them in, to be facilitators in knowledge to get started. In very broad strokes, there are a few basic steps building and collaborative scholarship.

to putting your collection information out as linked data:

- Create a namespace for your URIs
- Develop a data model
- Convert structured data to RDF triples
- Publish your graph to the Web
- Start linking

This is, of course, easier said than done, and it requires some technical expertise and programming skills to jump into this. There are a number of useful resources about doing this work, including the Open Data Handbook and the book Linked Data.¹⁸ Whether your institution takes tiny steps or big ones towards this goal, start looking into it. You will see that it is already out there and growing.

Google is already using linked data in its search results and its global knowledge graph, as is Facebook in its open graph.¹⁹ If **Additional Resources** we look at the Linked Open Data graph from 2009 as compared W3C Linked Data to 2011, it has clearly grown in size.²⁰ To date, libraries, archives, http://www.w3.org/standards/semanticweb/data and museums are not well represented in this graph, but that is

changing. New datasets are starting to appear in the cultural sector and will continue to add to the web of data.

The Promise

The promise of linked data is that we will be able to better meet our missions and goals. We will be able to provide improved access to collections and information, facilitate new knowledge building and collaborative scholarship, and provide for more efficient management of our collections information using new tools and methods for creating and sharing our data across our

We need to capture the data for our own purposes but also release it to encourage the growth of new knowledge and new forms of scholarship. Perhaps one of the roles for the twentyfirst-century library, archive, and museum is to support this sort of work, just as we support living artists, authors, and other creators now. Part of our work could be to collect, curate, and preserve the best of the new types of cultural and scholarly resources built on this growing web of data. These new cultural and scholarly resources will still need a place where others can experience them, understand them, and discover them. We should be at both ends of the process, to support the creation and innovation, and to provide interpretation and stewardship into the future. As libraries, archives, and museums with long-established, well codified professional practices, rich data standards, expertise with technology, and a willingness to work together, we are in an ideal position to do this work together.



Library of Congress Linked Data Service http://id.loc.gov/

OCLC Linked Data Video http://youtu.be/fWfEYcnk8Z8

Europeana

http://pro.europeana.eu/linked-open-data Fun Video: http://vimeo.com/36752317

British Museum Semantic Web Collection Online http://collection.britishmuseum.org/

Notes

1. Excerpted from The University of Chicago Library (http://www.lib.uchicago.edu/e/about/mvv.html), The Metropolitan Museum of Art (http://www.metmuseum.org/about-the-museum/mission-statement), The Smithsonian (http://www. si.edu/About/Mission), The Library of Congress (http://www.loc.gov/about/mission. html), and The University of California Libraries (http://libraries.universityofcalifornia.edu/about/vision-and-priorities).

2. Mary Elings and Günter Waibel, "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums," First Monday 12.3 (March 2007), http://firstmonday.org/article/view/1628/1543.

3. Leah Prescott and Ricky Erway, Single Search: The Quest for the Holy Grail (Dublin, Ohio: OCLC Research, 2011), http://www.oclc.org/research/publications/ library/2011/2011-17.pdf.

 "2010 Best Archives on the Web Awards, Best Re-purposing of Descriptive Data Category," Archives Next, http://www.archivesnext.com/?p=1512.

Elings and Waibel, "Metadata for All." See examples of content aggregations in introduction.

6. The Bancroft Library (in partnership with the California Digital Library) is participating in this newly launched project: http://www.google.com/culturalinstitute.

7. Stephen Shankland, "Google Scraps Newspaper-Scanning Project," CNET, May 20, 2011, http://news.cnet.com/8301-30685_3-20064593-264.html; Jessica Mintz, "Microsoft Ends Last Chapter of Book Scanning," MSNBC Internet, http://www. nbcnews.com/id/24819964/ns/technology_and_science-internet/t/microsoft-endslast-chapter-book-scanning/.

8. Plenary, Rare Books and Manuscripts Section (RBMS) Preconference, American Library Association, June 21, 2012, San Diego, CA; Plenary, Society of American Archivists (SAA) Annual Meeting, August 6, 2012, San Diego, CA; Speaker, Museums and the Web 2012 Conference, April 11-14, 2012, San Diego, CA.

9. LODLAM: Linked Open Data in Libraries, Archives, and Museums, http://lodlam.net/.

10. Jon Voss, "An Introduction to Linked Open Data in Libraries, Archives & Museums," Smithsonian Videos, http://youtu.be/vUBTd8WZZ5A; Jon Voss, Society of American Archivists 2012 Keynote, http://youtu.be/gyR59nFFMRg; Jon Voss, RBMS Plenary: "Discovery: Linked Open Data," http://www.slideshare.net/jonvoss/ rbms-lodlam-presentation.

11. "Resource Description Framework (RDF)," W3C, 2004-02-10, http://www. w3.org/RDF/.

12. Civil War Data 150, http://www.civilwardata150.net/.

Conflict History, http://conflicthistory.com.
Vistory, http://www.vistory.nl/what-is-vistory.shtml.

15. Europeana Semantic Searching Prototype, http://eculture.cs.vu.nl/euro-

peana/session/search. 16. Look for updates on projects, venues, and discussions on the Google

Group Linked Open Data in Libraries, Archives, & Museums, http://groups.google. com/group/lod-lam.

17. Lightweight Information Describing Objects (LIDO), http://network.icom. museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/; The CIDOC Conceptual Reference Model (CRM), http://www.cidoc-crm.org/.

18. The Open Data Handbook, Open Knowledge Foundation, 2010-2012, http://opendatahandbook.org/en/; Tom Heath and Christian Bizer, Linked Data: Evolving the Web into a Global Data Space, 1st ed., Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136 (Morgan & Claypool, 2011), http://linkeddatabook.com/editions/1.0/.

19. Thomas Steiner and Stefan Mirea, "SEKI@home, or Crowdsourcing an Open Knowledge Graph," Fall 2012, http://openknowledgegraph.org/; Margaet Heller, "Real World Semantic Web?: Facebook's Open Graph Protocol," ACRL Tech-Connect Blog, May 9, 2012, http://acrl.ala.org/techconnect/?p=855.

20. Richard Cyganiak and Anja Jentzsch, Linking Open Data cloud diagram, http://lod-cloud.net/.

21. "Digital Dandelions: The Flowering of Network Research," http://www.jacobsschool.ucsd.edu/news/news_releases/release.sfe?id=685.

SESSION 4 ACCESS TO COLLECTIONS

Seb Chan

I will be discussing the journey that I have taken for the last decade and the way I see metadata as becoming a cultural source code. I am currently working at the Cooper-Hewitt in New York, which is part of the Smithsonian. The museum is closed for renovations and will reopen in 2014, which is giving us a great opportunity to rethink everything about the museum from the building up. I came to the Cooper-Hewitt from Sydney to change the digital layer of the museum and to build digital right into the building itself.

I previously worked at the Powerhouse Museum in Sydney, a large design and science museum. I am not one of you. I do not come from the library world or from the art world. I come from the Internet, but also from a background in social policy and sociology.

Nick Poole, the CEO of the Collections Trust, stated earlier last year that museums have really changed, and the core functional model of the museum has expanded to incorporate publishing and broadcast as well as the physical institutional experience.

When I started at the Powerhouse Museum in 2001, we were all talking about the virtual museum. I shudder when I hear people talking about this still because this is such an old idea. We talked about being an encyclopedic resource. We talked about being authoritative. This is because we thought we were in control of everything. We thought that the Internet had security guards like the security guards who stop people and search their bags in the foyer or prevent photography in the galleries.

When I left the Powerhouse in 2011, the 1950s notion of the museum without walls had become a reality, and museums started to take on the role of a library, becoming a data provider. We realized that our authority was contextual, and we also realized that as we became more visitor-centric, the visitor was in charge.

caiuw. on. ords 2.3 rail/ ve gaisti-ET, a. w. sety sety

rogle

m. o/; 12, ta: Se-