

Princeton University
Digital World of Art History
July 13, 2012

Art Clouds: Reminiscences and Prospects for the Future

A talk by Eleanor E. Fink

My gratitude to Colum and the Index of Christian Art for inviting me here today. It's an honor to share my thoughts past, present, and future about a concept that I have been talking about for years that is now becoming more tangible through the Linked Open Data Cloud – *that is being able to search across institutions and around the globe for art information as if it were all contained in one space or on the same platform. In essence as if it were a virtual database.* The concept seems most fitting based on what I have heard today about the Princeton Humanities Project and of course the collaborative work surrounding the Index of Christian Art and Morgan Library, the Avery and many of the others projects presented.

Much has been written about how new information and communication technologies are radically transforming how we live and work and more precisely how individuals interact with content. These new technologies in combination with an ever-expanding array of applications, social networking platforms and tools enabling users to create their own content, have the potential to dramatically impact how we *experience content*, and most importantly, *how scholarly research is conducted!*

What I am going to present today, may already be known to you. But it is relatively new and few cultural institutions have taken the plunge. The exception is a project in Europe called Europeana – a vast digital library of museum, archives, and library information. At the moment the Amsterdam and British Museums have converted collection records to Linked Open Data. Europeana plans to apply LOD across its vast digital library. In the US few art museums have converted their data records to LOD. Yale Center for British Art is in the process and soon the Smithsonian's American Art Museum will convert 44,000 of its object records. The Yale application should be of particular interest to Princeton because it is a cross organization data harmonization and access project using a linked data decentralization approach. Some of the challenges faced by Yale – namely integrating their knowledge with other organizations, are precisely what the Smithsonian project hopes to address.

As is always the case with new technology, the cultural stratum is not as fast to react and adapt. I recall a talk we invited Stewart Brand to give when I was the director of the Getty Information Institute. Stewart Brand is the president of the Long Now Foundation. His talk was on the pyramid of global change and in particular how different strata respond to technological changes.

Cultural institutions were on the bottom of the pyramid - meaning often the last to adopt new technology. Perhaps with social media as a level playing field that pyramid might look a little different today, but not significantly. So when it comes to LOD, we find that publishing houses and news media like the New York Times, Reuters, and BBC are leading the way as well as industry and government. Before I talk about the value of LOD and what it offers scholars, I would like to go back in time because it explains my excitement about the concept of *LOD Art Clouds*.

I began my professional career decades ago managing large research databases at the Smithsonian that were off line and stored on a central mainframe computer that served all 13 museums! It seems ridiculous now, but we entered data using key punch cards and later optical character recognition before we were able to enter data directly into our own computer. Of course this was long before email. We entered hundreds of records and then waited weeks to receive a printout of what we had been entered to check for errors.

I oversaw five art research databases comprising more than 500,000 individual records. That figure represented more data than what the Getty Art History Information Program comprised when I joined them in 1987. One of the Smithsonian projects was a nationwide inventory of American Paintings in public and private collections that our museum, the Smithsonian Museum of American Art undertook in preparation for the celebration of US bicentennial. Some of the additional databases comprised the pre-1877 Art Exhibition Catalogue index, the Peter A. Juley and Son Collection of historic photographic records, and later SOS – a nationwide inventory of American Sculpture located in parks and town squares in every corner of the United States.

Managing these projects at that time had a profound influence on my career. It was the beginning of the quest for standards and for helping researchers access art information. It was also the beginning of a vision that centered on being able to collaborate across institutions to produce standards and in turn also use those standards to interconnect art information and create virtual databases.

At some point I asked the Smithsonian Office of computer services to run a comprehensive artist index against the 5 databases. I discovered we were not consistent in entering names of artists, places, subjects, and other fields of information. I went to the library of congress to find out what standards they were using. I discovered they had a nationwide service called NACO – Name Authority Cooperative Program. It was a central name authority catalog built by librarians around the US trained in the system who could submit the name of an author not yet in NACO that was then vetted by NACO. I pondered whether art museums might do the same for artist names? Surely every museums and art research project would struggle as we had to be consistent and what if some day we could search across databases held by institutions around the world. If an artist's name was not the same in all cases or if there was no mapping of variant names, we would be hindered from finding all that we were searching for. I met with the head of NACO and asked why not include the names of artists and was told, who would be the

authority??? Fast forward several years and those of you familiar with the Getty vocabularies and Union List of Artist Names will know how we addressed NACO's concern.

So the Smithsonian research databases led me down a path searching for standards and I became the founder of what was called the Getty Vocabulary Coordination Group and later the director of the Getty Art History Information Program or Getty Information Institute (GII) as it was called when the Getty Center opened. The Smithsonian research projects began a quest and a vision inspired me – one day being able to search for art information seamlessly.

The Getty comprised several scholarly databases when I joined¹. The mandate of the Getty Art History Information Program (AHIP) established in 1983 was to explore ways of applying computer technology to improve scholarly access to and use of art-historical information. This was accomplished a number of ways such as building a critical mass of electronic information essential to art-historical research (2) identifying model approaches and acting as a catalyst among other institutions with similar interests to foster standards and procedures for sharing information on an international scale, (3) analyzing the needs of the art-historical community and tailor emerging technology to the requirements of humanistic research and (4) influencing the development of a technical structure through which others could contribute and retrieve data.

The Getty had the weight to convene top people and write white papers that got the attention of policy makers.

The model databases were

- The Avery Index to Architectural Periodicals (with Columbia University);
- The Bibliography of the History of Art (formerly RILA, with the Clark Art Institute in Williamstown, Massachusetts and the CNRS;
- The Census of Antique Art and Architecture Known to the Renaissance (with the Warburg Institute in London and the Bibliotheca Hertziana in Rome);
- The Witt Computer Index (with the Courtauld Institute in London); and
- The Provenance Index.

Like the Index of Christian Art, all of these efforts collected data from other institutions. But the collection was through paper documents and later CD rom.

In addition there were consortia such as Architectural Drawings Advisory Group (ADAG), later called the Foundation for Documents of Architecture.

These data collection projects, begun in an era of stand-alone computing, posed many of the challenges that remain fundamental to databases in the humanities.

¹ Eleanor Fink, [The Getty Information Institute: A Retrospective](http://www.dlib.org/dlib/march99/fink/03fink.html) (<http://www.dlib.org/dlib/march99/fink/03fink.html>)

First, their data collection involved original cataloging, abstracting, and recording from primary sources. This labor-intensive and highly specialized work needed to be performed by professionals with either advanced degrees or an in-depth knowledge of the subject matter.

Second, because unrecorded source materials are boundless and can absorb a limitless supply of resources, questions related to sustainability arise. Who is responsible for continued refreshing and updating of the data? How is the long-term maintenance to be funded? And what limited objectives need to be defined? *Sound familiar!* Most research database projects face some of the same challenges today.

Additionally, the complexity of art-historical information pushed computer technology to its most sophisticated limits. Because this type of information has inherent idiosyncrasies that cannot necessarily be reduced to equivalencies without losing critical data, rigid categories failed to adequately describe even what may appear to be an ordinary object. You know this well. For example, a "simple" architectural drawing may:

- involve several key people to be identified (client, architect, draftsman, etc.),
- be only one of several versions drawn over time,
- refer to a building that was never built or one that was subsequently reconstructed or destroyed,
- have belonged to a commissioned competition and later grouped with other drawings in a bound portfolio, or exist in multiple surrogate forms (i.e., blue-line prints, photographs), and so on.

These challenges were eventually addressed by building production platforms in-house. No commercial off-the-shelf software could handle the range and complexity of relationships these resources required. The Getty invested considerable resources in building these projects but could not search across them. The Getty Vocabulary Coordination Group was then created (1) to help establish standards that would ensure the internal management as the projects got larger and the number of contributions from partners increased and (2) to enhance the possibility of searching across databases.

Since research involves making connections and art information does not reside in one type of institution, be they libraries, museums, archives, or private collections, a holistic approach to developing standards involving many levels of practice and expertise was needed.

In the case of the Art & Architecture Thesaurus (AAT), GII's departure from the traditional approach to vocabulary "authorities" took advantage of the new flexibility offered by technology. The AAT proposes that there can be several acceptable terms for the same object, idea, or technique and links related terms so that the use of any one of them/ points to the same descriptive concept.

GII's Vocabulary Coordination Group (VCG), founded in 1987 developed two additional vocabularies: the Union List of Artists Names (ULAN) and the Getty Thesaurus of Geographic Names (TGN). It also developed two Meta data frameworks: Categories for the Description of Works of Art (CDWA) and Object ID.

Like the AAT, the vocabularies are based on the flexible principle of mapping together multiple name forms rather than providing just a "preferred name". They address international requirements with respect to incorporating vernacular terminology and have consequently become *de facto* standards in many museums and visual resource collections internationally. This flexible approach is significant in a field such as art history where, for example, two art historians might record an artist's name differently, each version being "correct" and with its own traditions of use. The result of mapping together variant names is a robust set of vocabularies that have proven as useful in locating and retrieving information as they have in providing guidance in cataloging: A search for Constantinople will locate records cataloged as Istanbul and Byzantium.

Then the Internet arrived and our vision sharpened! At the dawn of the internet and soon thereafter the web, navigating a solar system of art data hubs now could be envisioned. Indeed GII focused on this vision and the concept of universal access to art and cultural heritage information. Thus we produced some of the glue needed to manage information and navigate data. But in 1999 and still the case today what we have is a solar system with individual stars. We can reach them, but only then find out what they contain. We cannot as yet, seamlessly search across them. Museums have created isolated space stations!

The GII (again the former Art History Information Program) increasingly played a catalytic role and proposed that the cultural heritage sector develop data structures and vocabulary tools that assist in making connections across repositories of information. Well-designed structures and vocabulary tools would allow students and researchers to search an enormous variety of the world's electronic databases of texts and images as if all the information were contained in one place. GII promoted the concept of digital libraries or a virtual database and worked with many institutions to develop research databases, data structures, standards, guidelines, vocabularies, and demonstration projects. One such demonstration project was L.A. Culture Net, a digital gateway to the collections in Los Angeles that became functional when the Getty center opened. The Getty vocabularies were used on the front end to help navigate to specific items in a search.

GII produced a video called the Virtual Database - Art Information on the Networks. It centered on a scholar's journey for connections – *seamlessly* being able to bring up the many art objects whether sculptures, drawings, models, or paintings associated with Saint Peter's Basilica no matter where those objects currently reside: Metropolitan Museum of Art, Norton Simon Museum of Art, Windsor Castle, Belgrade museum, Berlin, Stockholm - in essence searching disparate databases as if the information were held in one place.

In addition to the concept of the virtual database, our work in building standards and model databases that were collaborative led to another thought in my mind that goes right back to the National Name Authority Cooperative Program (NACO) that allowed librarians across the US to *work together* to build the name authority. I called the concept “*Interworkability*.”

“Interworkability” was based on using Information Communications Technology and interfaces that allowed experts to collaborate and interactively build research resources like BHA and standards like the ULAN on a single platform. The Getty didn’t have to hire all the staff to do it. The resources could be developed by direct input as long as the interface allowed for review and vetting of new information. I viewed “interworkability” as a good solution to the administration issue about costs and sustainability. Unfortunately, it was not pursued and we have lost BHA. In an age of social networking the concept certainly appeared to have possibilities. Fortunately, when the Getty Information Institute was shuttered in 1998/99, the vocabularies were transferred to the Getty Research Institute.

Turning now back to Linked Data Art Clouds, perhaps you see the similarities. There are four key values to Linked Open Data.

First of all, in Linked Open Data *everyone* can access, reuse, enrich and share the data published in LOD format! It is somewhat like Wikipedia and it makes “interworkability” more feasible.

LOD also disambiguates your search. In other words it eliminates the noise you sometimes get in a Google search. What you are seeking is more precise because of the way you tag the information you are converting to LOD.

LOD connects you to all kinds of relevant information creating a playground for serendipity and it updates that information dynamically within the same web space.

In essence LOD is able to connect data from the Web that was not previously connected. For a very long time, information on the Web was mostly connected through hyperlinks that create connections on document level. Although these connections already provide a wealth of context by enabling one to click through to various resources, they are coarse-grained and do not express what type of connection there is between two pages, merely that there is one. Linked data aims to connect more fine-grained bits and pieces of data, information and knowledge because in converted LOD data, the relationship of information is made explicit².

Let’s take a look at the Europeana video demonstrating the value of linked open data when searching for the figure of the goddess “Venus.” <http://vimeo.com/36752317> The video is also available through youtube

² Sharing cultural heritage the linked open data way: why you should sign up (Museums and the Web 2012).Johan Oomen and Lotte Belice Baltussen, Netherlands Institute for Sound and Vision, The Netherlands with Marieke van Erp, VU University Amsterdam, The Netherlands

http://www.youtube.com/watch?v=uju4wT9uBIA&list=UU_Z8vHTU2qx4_qBIy9f_u0A&index=1&feature=plcp.

The Smithsonian's Museum of American Art (SAAM) is embarking on converting 44,000 records to LOD. The work is being handled by the Information Sciences Institute (ISI) at the University of Southern California. SAAM is undertaking this project for several reasons including that it will streamline many functions within the museum; it will lead to new applications particularly those that will allow visitors to experience content in new ways; and they will be able to discover what happens when scholars outside the museum contribute information. We hope several museums that have American Art will join in the effort and populate the LOD Cloud producing in effect an "Art Cloud". The project will serve as model that can provide LOD conversion guidelines and tools for other museums and art organizations. ISI's broad profile, tools, expertise and extensive work in the development of the modern internet structure and protocols, will help SAAM take a giant step into the LOD universe³. ISI has already developed tools that facilitate the translation of data to LOD and following the project intends to make the tools accessible to the art information community.

The concept of "interworkability" seems to work well in the case of Wikipedia. SAAM sees their project as a daring entry to a new way of handling art information and how scholarship is conducted. Will scholars collaborate, build bridges, and share their expertise? We intend to find out?

Thank you.

Eleanor E. Fink
eleanorfink@earthlink.net

³ ISI has a broad profile of expertise including e-commerce, cyber security, health informatics, satellite technology, sensor networks, and various forms of high performance computing.