Beyond Borders: The Humanities in the Digital Age

James Cuno, President and CEO, J. Paul Getty Trust

October 23, 2014

Keynote Address at *Museum Computing: An Approach to Bridging Cultures, Communities and Science*, the 2014 Pacific Neighborhood Consortium (PNC) Annual Conference and Joint Meetings at the National Palace Museum in Taipei, Taiwan

Thank you very much for inviting me to be a small part of this important conversation. I am painfully aware that I am not a techy, not an "in-the-trenches" kind of guy, who can see the possibilities and limitations of the digital world from the bottom up and inside out. I am a user and an enabler. I ask dumb questions—even *painfully* obviously dumb questions— that might lead to new developments that users like me can use to advance our field. And let's be honest, most of us—not you, of course, but the *us* that comprises the majority of users of the new technologies—encourage many of the new developments to come into being precisely because we don't know the limits of what is possible. All we can do is ask "why not?," which very well might be the most important question of all.

I want to frame my remarks this morning by emphasizing the "beyond borders" in my title. The World Wide Web was developed and given that name—*world wide*—because it aspired to go beyond borders, to be accessible independent of political borders, and even, as it gathered momentum, to work against those borders, to be freely accessible to everyone with access to the internet, which today means about 40% of the world's population, a number that is steadily increasing, especially in the developing world.

In the field in which I work—broadly, the visual arts and cultural heritage—this has provoked us to think of ways to dismantle the false architecture of nationalism by which our field has been organized: French art as distinct from Chinese art, for example, each having qualities and characteristics that are naturally and essentially theirs and that can't be shared between them or among any other regional cultures. Cultures have as many similar qualities and characteristics as they have different ones. And to the extent that they have different ones, it is not because it is by nature that they do but because of the limitations of their makers' experience of the world. Artists make art provoked by their experience of the world. The more of the world they experience, the more they will make works of art in dialogue with the large, multiethnic, multilingual, and multipolar world that is the world in which we live. And with more and more people crossing borders and living in foreign lands and with access to the internet, that dialogue will be rich and diverse and beyond the borders of modern nation states.

I want to focus on three aspects of the new technology as it pertains to my field:

- **access to intellectual and cultural assets**, including big and deep data;
- **collaboration** within interoperable environments; and
- new ways of **analyzing** that data and **publishing** the results.

These, I believe, are the aspects of the new technology that will shape our field in the coming years.

Making Resources Findable

We at the Getty are aggressively digitizing our intellectual assets: works of art in our Museum's collections, archival materials and books in our Research Institute's library, and implementing the Open Content Program, which provides free access to and unrestricted use of high-resolution images to which we hold the all rights. Anyone with Internet access can search for Open Content images and associated cataloguing information from the Getty Museum and the Getty Research Institute via the Getty Search Gateway (http://search.getty.edu) or through the "Search Tools & Databases" section of the Getty Research Institute's website.

That said, an image without authoritative data is virtually useless for research purposes. Via Getty Search Gateway, users can obtain detailed descriptive data about, for example, van Gogh's painting *Irises* in the Getty Museum. At first thumbnail (or limited) information, but then, more information, including the object's catalogue number and dimensions, where the painting was signed by the artist, and a high-resolution image of the painting itself; freely downloadable, with no restrictions on their use.

One can get similar access to books in the Getty Research Institute's library, and not only their cataloguing information but also their content. And now, through the Getty Research Portal (http://portal.getty.edu) or through the "Search Tools & Databases" section of the Getty Research Institute's website, readers can search digitized holdings of some of the world's leading art libraries through a single point of entry. All records in the Portal contain links to full digitized copies of those books, which can be viewed and downloaded free of charge by all users. As of early this month, there were more than 44,000 volumes available from the Portal, contributed by dozens of international libraries.

For example, I do a simple keyword search on "Taipei" and I get, among other items, a digitized version of the papers of a symposium on the treasures in the National Palace Museum; the digitized publication is keyword-searchable and downloadable in its entirety from the library of the Metropolitan Museum of Art, one of the first libraries to contribute to the Portal. We are constantly seeking new contributors to the portal, and in particular we hope to increase the number of holdings relating to non-Western art, architecture, and material culture.

Research Across Languages

But of course the more international these resources become, the more we need assistance in searching terms in multiple languages. For decades, the Getty Research Institute has been building powerful multilingual electronic thesauri designed to facilitate cataloguing and documentation; to serve as lookup tools or knowledge bases; and to function as powerful search-and-retrieval tools. Now, with the advent of Linked Open Data, we are able to freely share these tools with the international cultural heritage community so that the information and semantic links in our vocabularies (http://vocab.getty.edu) can be exploited in a variety of ways by institutions around the globe.

Linked Open Data is a way of expressing data "in the cloud" that retains the specific meanings and relationships between terms, names, and concepts. From the beginning, we made our vocabulary databases available free of charge on the Web, where users can look up terms,

browse the hierarchies, and cut and paste information. We have also licensed the datasets to hundreds of cultural organizations and software vendors around the world.

Last year, in the spirit of the Getty's new Open Content Program, we began the arduous task of converting our electronic thesauri into Linked Open Data. All of the data in our *Art & Architecture Thesaurus* and *Thesaurus of Geographic Names* are now available as LOD, free of charge under an Open Data Commons license. Our other two vocabularies, the *Union List of Artist Names* and *Cultural Objects Name Authority*, are slated for release as Linked Open Data during the next year.

Meaningful relationships among Linked Open Data from various museums and libraries can be made automatically via computer programs in the LOD "cloud"; this kind of meaningful linking can be greatly enhanced with the information and relationships that are encoded in the Getty's electronic thesauri.

Recently Europeana (http://www.europeana.eu), a huge portal of information and images from more than 2,000 cultural heritage institutions in Europe, announced that it will incorporate the Linked Open Data version of the Getty's *Art & Architecture Thesaurus* to enhance searching and browsing their large, varied, multilingual data repository.

One reason why the Europeana consortium decided to incorporate the Linked Open Data version of the Getty's *Art & Architecture Thesaurus* is that it contains a great deal of multilingual terminology that can significantly enhance access to a data repository with records in various languages.

Working with several international partners, including the Academia Sinica here in Taipei (http://www.sinica.edu.tw), we are striving to make the *Art & Architecture Thesaurus* increasingly multilingual and multicultural. For example, the *Art & Architecture Thesaurus* record for "marble" includes variants in French, Italian, Spanish, German, Dutch, and Chinese. Multilingual tools like this can help us to bridge national, cultural, and linguistic boundaries in ways that were unimaginable even a decade ago, thus opening up the transnational, cross-cultural perspectives that I mentioned at the beginning of my talk.

This is an example of how at the Getty we are not only providing free acess to our intellectual and cultural assests but have developed the tools to link with the assets of other insitutions to build big and deep data sets that can be analyzed. Data sets on their own, however deep, are limited by the means we have of analyzing them. Breaking down the linguistic barriers to access is a crucial first step in advancing research in our field. But of course, it's not enough to build and analyze data. One needs to broadly disseminate the results of the analyses.

Online Scholarly Publishing

To this end, a few years ago, the Getty Foundation funded a initiative to develop new kinds of online scholarly catalogues of the works of art in museum collections. These were the initial participating institutions, all of whom have now released their first born-digital, online catalogues. And a number of them have published or are preparing to publish their second

OSCI (Online Scholarly Catalogue Initiative) catalogues with plans of continuing with such publications indefinitely.

During the course of the OSCI project, the participating museums came to realize that no single system yet existed to handle the range of requirements indicated here on this slide and so they developed a variety of innovative solutions. We placed no restrictions on them. We wanted to encourage innovation, not dictate solutions. And above all, we wanted the participating institutions to work together to share their ongoing developments.

They thus came up with three different technical solutions: the OSCI Toolkit, developed for the Art Institute of Chicago by the IMA Lab, produces on online catalogue that look more like a book and are accessible on a variety of electronic devices; a second approach chosen by San Francisco Museum of Modern Art makes use of existing Web content that can also be enhanced in order to enrich the online catalogue; and a third solution, which links more or less directly to museum collection management systems, using existing data that has already been compiled by the museums on their holdings; the Seattle Art Museum adopted this approach. Information about the published catalogues and the initiative is available on the Getty Foundation website (http://www.getty.edu/foundation).

Digital Publications

We at the Getty are deeply committed to digital publications. We have begun to make full digital copies of publications from our backlist available free for download via Getty Publications' Virtual Library on our website; to date we have some 277 publications available for download. Also available are the Getty Research Institute's "Introduction to" series, which deals with issues related to technology and cultural heritage, and more than one hundred publications produced by the Getty Conservation Institute, including scientific research papers, conference proceedings, case studies, project reports and bibliographies, as well as the *Abstracts of International Conservation Literature*.

Most important in these research and publishing developments are efforts to exploit the electronic environment to both conduct collaborative research projects, and publish the results of that research. This allows museums and libaries to make rare and unique materials, which formerly could only be consulted in person, available to anyone with access to the Internet.

The first digital critical facsimile edition produced by the Getty Research Institute is slated for publication online, free of charge, later this year. The research that supported this project was produced in the Getty Scholars' Workspace, a digital environment that the Research Institute custom-built to facilitate collaborative research and knowledge sharing among scholars who are geographically separated.

Digital Humanities for Conservation

Recent technological developments have encouraged their use in another kind of publication, one devised as a way to inventory and manage heritage sites in real time. Identifying an essential need among many of the world's cultural heritage organizations, the Getty Conservation Institute partnered with World Monument Fund to develop Arches, which is a

comprehensive software platform especially designed to help inventory, manage, and protect the world's immovable cultural heritage. Using the latest semantic technologies, the system is open source and is freely available for cultural heritage organizations to independently implement and, if necessary, customize. Arches incorporates a broad range of international standards, which promote sharing and longevity of data regardless of inevitable technological advances. Arches includes the ability to create and manage complex controlled vocabularies, such as the Getty's thesauri, which will ensure valid data entry and aid in searching.

Many more details on the project are available on the Arches website (http://www.archesproject.org). The latest round of major development will be completed in January 2014 with the launch of version 3.0, which will be immediately available to download and implement. Arches is currently being used by a number of international cultural organizations—for example, the University of Applied Sciences, Mainz, Germany, for a project that maps and manages a dataset of several thousand archaeological findspots and the city of Los Angeles to manage its extensive data from SurveyLA. It is also being assessed by English Heritage and Parks Canada for national implementations.

Supporting Digital Thinking

Over the past few years, formal studies and informal conversations have revealed that scholars in the humanities are less inclined to exploit the power of available digital tools to research and publish in their fields.

Last year the Getty Foundation interviewed leading art and architectural histiorians in the United States and learned that one of the major obstacles discouraging them from using available digital tools and building new ones is their lack of familiarity with available tools and their applicability to art and architectural historical questions.

So, this past summer the Foundation began to sponsor week-long, intensive workshops to inroduce art and architectural historians to the issues, tools, and opportunities offered by information technology. The first three were held in Cambridge, Massachusetts; Washington, D.C.; and Los Angeles, at Harvard, George Mason University, and UCLA, respectively, three universities with robust digitial humanities centers.

Three hundred and fifteen scholars applied to the workshops and 56 were accepted. The Getty Foundation is now exploring options for expanding the program to more universities within the United States and of course to other universities in other countries. Introducing art and architectural historians to the potential of information technology holds the greatest promise for our field. And the Getty means to play a leadership role in this endeavor.

Visualizing Networks

Let me close with some thoughts about current projects in the digital humanities. I am especially interested in those projects that are engaged with analyses of deep data sets. Given the relative sophistication of optical character recognition, it's no surprise to see that many of those more interesting projects employ visualization programs to map relations among data.

You probably all know the "The Stanford Geospatial Network Model of the Roman World," or ORBIS (http://orbis.stanford.edu/).

"By simulating movement along the principal routes of the Roman road network, the main navigable rivers, and hundreds of sea routes in the Mediterranean, Black Sea and coastal Atlantic," as the ORBIS home page describes it, "this interactive model reconstructs the duration and financial cost of travel in antiquity," expressing Roman communication costs in terms of both time and expense, and revealing "the true shape of the Roman world and provides a unique resource for our understanding of premodern history."

All one does is enter starting and culminating points, and whether one wants to travel the fastest or safest routes and ORBIS maps it for you. For example, I chose to map the fastest route between Rome and London as it would have been over the summer months. By simply pushing a button, I got this route over road, river, and coastal and open seas, and learned that the trip would have taken 27 days, covering 1967 kilometers, and resulted in a cost per kilogram of wheat at 6.85 *denarii* by donkey, 7.87 *denarii* by wagon. If I had chosen the cheapest route, the trip would have taken 39.2 days, covering 5409 kilometers, with a cost per kilogram of wheat at 3.98 *denarii* by donkey and 3.98 by wagon. More sophisticated searches result in more detailed information, of valuable use by scholars, teachers, and students.

You may also know of Stanford's "Mapping the Republic of Letters" project (http://republicofletters.stanford.edu/). This site explores networks of correspondence that stretched across countries and continents, the social networks created by scientific academies, and the physical networks brought about by travel. These networks facilitated the dissemination and criticism of ideas, the spread of political news, as well as the circulation of people and objects. "Mapping the Republic of Letters" uses sophisticated, interactive visualization tools to create a repository for metadata on early modern scholarship.

One case study draws on data from the catalogue of the Bibliothèque National de France to show a visualization of places where Voltaire's works were published. On one map, I can show the true stated locations where Voltaire's works were published (some 1194 of them), with an indication of comparative numbers. The highest concentration was in Paris (some 333 editions), but he was also published in Istanbul, with five editions—one more than Stockholm. But I can also explore 162 *falsely* stated places of publication to give a more complete picture of the geography of Voltaire's publications. And then I can map onto these falsely stated places of publication a map of the *actual* places where Voltaire's works were published and their relative numbers.

Colleagues at the Getty have developed similar projects using the Getty Provenance Index®, a very large set of databases containing more than a million records relating to inventories, sales, and auctions of paintings in Western Europe and North America from the 15th through the 20th centuries.

This data lends itself to visualizations focusing on cross-border traffic and its international agents. One study produced a network diagram showing the major centers of the secondary art market during the first two decades of the 19th century. They were, as you may have surmised, London, Paris, Amsterdam, and Antwerp.

Based on 230,000 records, the network diagram maps individuals such as buyers, sellers, auctioneers, and experts who were active at multiple places in Europe between 1801 and 1820. The diagram uses circles to represent individuals, and squares to represent places. Node and link colors indicate the prevalence of selling (blue), brokering (gray), and buying (orange). The size of the circles and squares as well as the thickness of the lines connecting them is dependent on the total activity and number of transactions over the 20-year time frame.

Another network visualization, this one based on the records of the legendary art dealers Goupil & Cie. as recorded in the Getty Provenance Index, reveals that each branch of the Goupil enterprise was a highly robust entity in its own right, with a small number of artists and patrons shared between branches (the red dots in between the four gallery nodes).

Deep Data and Humanistic Inquiry

But there is more than this that we should be doing. With the rapidly developing and rich potential of image recognition, we should be able to get past character recognition to explore a range of visual data, from the incidence of certain colors, lines, forms, or compositional structures in prints, drawings, or paintings since, for example, the 15th century in Europe, to a comparative study of the appearance of specific formal structures in sculptures from east and south Asia through Europe, Africa, and the Americas.

More and more, museums, auction houses, and dealers are digitizing works of art in their collections. If there were a tool that would allow us access to, for example, all 17th-century Dutch paintings in public collections that have passed through auction houses and dealers in the last ten or more years, we'd have access to a rich and perhaps even deep data set ripe for various analyses. Already there is the Cranach Digital Archive, an interdisciplinary collaborative research resource providing access to art historical, technical, and conservation information on more than 1,100 paintings by Lucas Cranach (http://www.lucascranach.org), including almost 10,400 images and 790 PDF documents from 146 contributing institutions, as well as 150 digitized and transcribed archival documents and 2,750 literature references. In one place one can access Cranach's paintings, sorted by attribution, dating, collection, function, form, etc., pull up a single painting, such as the central panel from an altarpiece triptych, explore its underdrawing detail by detail, and then call up its cataloguing information.

In every respect, there's a lot of information here. But what I would like to see us develop are the data and analytical programs that would allow us to do the kind of analyses that the Stanford Literary Lab has done with literary data. One such project is called "Style at the Scale of the Sentence." It is available as a downloadable PDF at their website (http://litlab.stanford.edu).

Researchers took their data from the Chadwyck-Healey nineteenth-century database of some 250 well-marked British novels and separated the sentences into three types: those containing dialogue, those containing a mixture of narrative and dialogue, and those containing only narrative. They then concentrated on sentence types based on clause combinations—sentences consisting of a single independent clause, sentences consisting of a dependent clause followed by an independent clause, and three other sentence types. Without going into great detail, let

me just say that they found that these five sentence types account for 40% of all narrative sentences and about 65% of sentences between ten and twenty words.

They then concentrated on certain sequencing structures in these sentences and found that different authors favored different kinds sequencing, as represented in this graph of sentence types used by seven different English language authors, from Radcliffe to Hemingway. They further broke down the data by looking at "most distinctive words" used by these authors. They also charted the use of different verbs in different types of sentences. I needn't go into it any further—suffice it to say that they continued to break things down and reconstruct them in ever-more-granular sets. In the end, this tested their very balance between the "digital" and the "humanities." The more and more "digital" they went, the more they needed the "humanities" to make sense of their findings. Digital data mining alone, they discovered, would have been, in their word, "blind." But without the empirical content uncovered by data mining, the concepts that framed their conclusions would have been empty. Only from the encounter between deep data and humanistic inquiry, in their words, "did critical knowledge arise."

This is the kind of analysis—confronting data and humanistic inquiry—that I hope we will soon be able to do on works of visual art. I hope we will soon be able to break them up into very tiny bits and draw conclusions abut them in ways we aren't now able, and not just in the service of arbitrary distinctions (French art or Chinese art, for example), but in the service of studies that cross nations, continents, *and* centuries.

Thank you for letting me take so much of your time this morning to reflect with you on the current state and future potential of digital applications in the advancement of our knowledge and management of the world's artistic legacy beyond borders. That is, I am certain, the promise of the digital humanities.