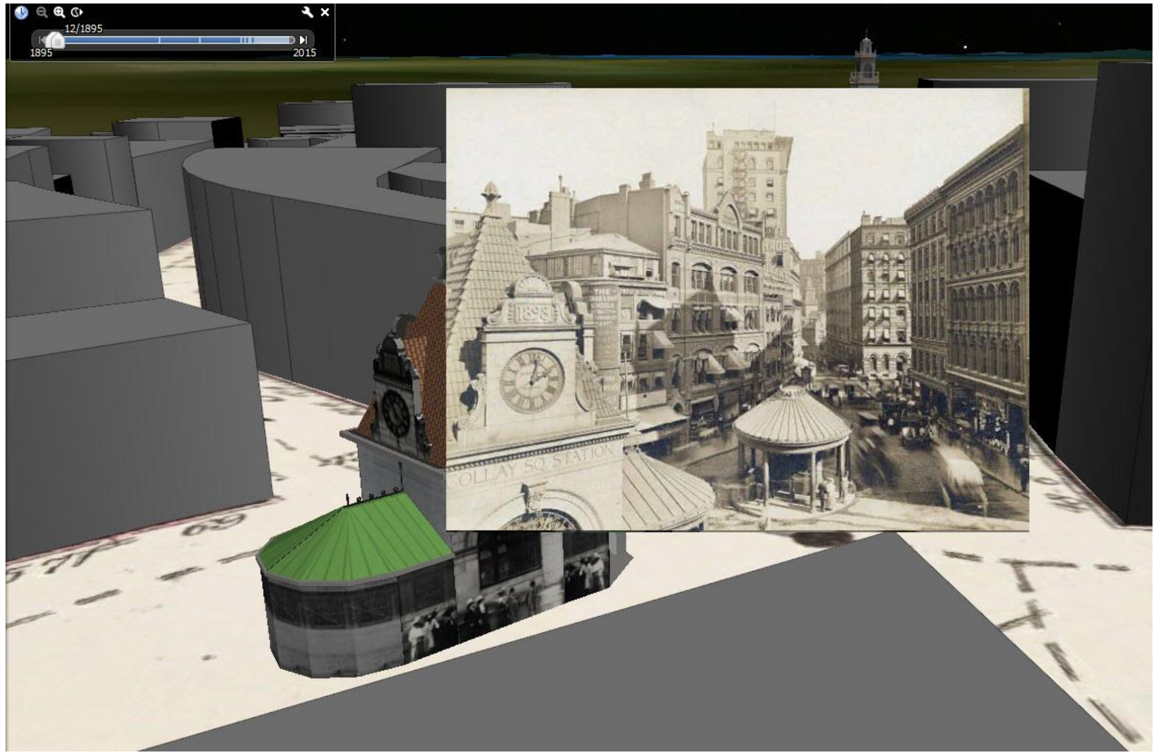
Information Stewardship

for Cultural Heritage Preservation:

Tools, Standards, Strategies



Prepared for:

University Partnership between National College of Art Rawalpindi, Pakistan, and Boston Architectural College, United States, in the field of Architecture and Heritage Conservation and Management

Funded by State Department grant award
#SPK33013CA091

By: Paul B. Cote **March 29, 2016**

About the Project

This handbook is intended to reach people who are involved n the preservation and study of cultural heritage – particularly with respect to the physical embodiment and conception of places as they relate to cultural and historical context.

We hope that the book provides a leg-up and an opportunity to discuss and explore technologies that permit disparate researchers and institutions to collaborate in a description and discussion of places in context.

It will be helpful for an executive director who wants to treat each page as a 15 minute executive summary with some examples to click on. At this level, an executive director might gain an appreciation for changes that are in progress and broad implications of an institutional and inter-institutional nature.

Or the pages of this book might be used as readings and demonstrations that would support a seminar or a sequence of practical workshops for students involved in a research project or design studio that has one of its aims to showcase new modes of information stewardship, heritage documentation, and of place-based scholarship and discourse.

The topics have been arranged in a progression that leads from big inter-institutional patterns of the web and federations of libraries; to a discussion of a few basic principles of metadata and other standards that put order into and that reveal order in formerly disconnected observations, ideas and documents. Standards also turn out to be a critical aspect of how organizations and individuals ensure that their work becomes heritage as opposed to road-kill on the information superhighway. The next several topics are concerned with applications – with a demonstration of a specific tool. Each of these has been selected to demonstrate how individual researchers participate in the larger patterns introduced in the earlier topics. The final topic looks at the trajectories that might be expected in the near and ten-year horizon.

This handbook makes frequent reference to web sites that demonstrate applications or standards or deeper reading. So it has been designed to be explored as a digital document that is connected to the internet. It is inevitable that some of the links will need to be updated periodically. Updates will be posted www.PlaneTable.org/handbook.

Acknowledgements

This document was prepared thanks to collaboration between the Boston Architectural College and the National College of the Art, in Pakistan, sponsored by the United States Department of State. The objectives for this partnership include: curriculum development with emphasis on heritage conservation; and management and preservation and sharing of cultural heritage information.

This booklet and associated teaching materials were conceived as an augmentation to Johanna Rowley's course on Heritage resources documentation presented at the Boston Architectural College in the Fall of 2015. It is intended to provide preservation professionals and students with a conceptual and technical experience to develop digital collections focused on conservation of architectural heritage. The following individuals deserve credit for envisioning the opportunity to develop this handbook and for making its development possible:

From the National College of Art, Rawalpindi campus: Dr. Nadeem Omar Tarar, Director, Ar. Ali Ahmed Shah, Associate Professor & Head, Department of Architecture, and Ar. Khalid Ibrahim, Assistant Professor and Point of Contact for BAC/NCA University Partnership. From the Lahore campus: Nadeem Hasan Khan, Registrar NCA Pakistan, Point of Contact for BAC/NCA University Partnership.

From the Boston Architectural College, in the United States: Don Hunsicker, Dean and Faculty, School of Design Studies; Johanna Rowley, Director of Practice Assessments, Dragana Kovacevic, Project Manager, and Aidan Ackerman, Director of Design Computing. Special thanks to Eleni Glekas, Director of Historic Preservation for reviewing the work and making many useful suggestions.

About The Author

Paul Cote is a lecturer in Landscape Architecture at the Boston Architectural College, specializing in Geographic Information Systems (GIS). Paul was lecturer in Landscape Architecture and Urban Planning and Design at the Harvard University Graduate School of Design for 19 years. He has served on the advisory committees of the Harvard Geospatial Library and the Norman Leventhal Map Center at the Boston Public Library and on the editorial board of the Journal for Geography and Map Libraries.

Title: Information Stewardship for Cultural Preservation: Tools, Standards and Strategies

Author: Paul B. Cote

March 16, 2016

Contents

Preserving Knowledge of Places

Preservation of cultural heritage involves understanding situations as relationships between culture, physical artifacts and places. Cultural heritage preservation is concerned with how situations change over time. Change is sometimes gradual and other times catastrophic; but change is inevitable. Theft, war, and decay cause the materials and the meaningful arrangement of cultural artifacts to disintegrate and disappear from their context.

Research and Documentation

Research and documentation play critical roles in the study and conservation of architectural heritage. Research requires making sense of documentation created in the past and creating new documentation concerning sites and their context as they currently exist.

Types of Documentation: Images, Descriptions, Drawings, Maps and Plans, Models, Policy Documents perform the following indispensable roles:

- Provide a record of the status of a place at a certain time.
- Establish an inventory and assessment of the value of resources and their level of threat.
- Are necessary for planning and prioritization of preservation, maintenance and monitoring of places.

Information Stewardship

In cases where cultural resources are altered, taken, destroyed or simply succumb to decay, information is all that remains. If architectural preservation is solely concerned with stabilizing and conserving physical artifacts in their context, then collecting documentation may be viewed only as a means rather than an essential product of the preservation enterprise.. If the documentation that has been collected and developed in the course of preservation projects is not discoverable and accessible to future scholars and preservationists, then the memory and significance of that situation will be lost forever.

Therefore, those concerned with the preservation of cultural heritage in the built environment are also concerned with collecting, developing, preserving, and sharing of information about the situations and their spatial, cultural and historical context.

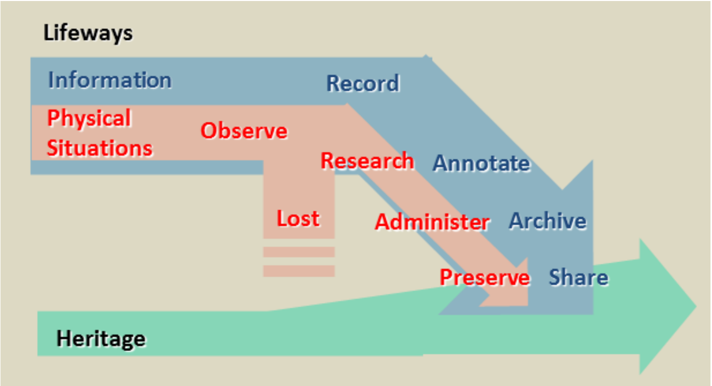


Figure 1: Heritage Information and Physical Situations

Technology for Research and Archiving

Until recently, the preservation of information consisted in developing a filing system for photographs and maps and written material which would be organized in boxes, folders, or cabinets. At some stage in the lifecycle of a project or of a researcher, some portion of these files might be formally accessioned into an institutional archive where it might one day be discoverable by future scholars or conservation professionals.

The past several years have brought an explosion in technology for digitizing, cataloging, discovering, and for making sense of documents. Specialists in archiving and cataloging are more important than ever. At the same time many investigators and institution face a gap terms of understanding the means by which digital archives are built, engaged with, and preserved.

Crossing the gap requires effort and institutional change, new roles and responsibilities, and new costs. There are reasons to go through this trouble. The first argument is opportunity or obligation to preserve and disseminate materials related to cultural sites with the public. Second, because new technologies for preserving and sharing documentation permit new modes of scholarship – for connecting information and knowledge together. To better understand heritage and to more effectively prioritize, plan and preserve situations and the record of their existence.

Remembering and Understanding

Figure 1 diagrams the lifecycle of heritage, which consists of the situations either preserved either in the physical form or as annotated, archived observations. Improved understanding and tools for information stewardship will increase our ability to remember important ways that configurations of the built environment are connected to their spatial, cultural and historical context.

With the revolution in more accessible tools and networks, the ordinary researcher and scholar play more active roles in the process of information stewardship. To fulfil this imperative, leaders in heritage preservation must have an understanding of the lifecycle of information within their institutions: from research and scholarship, through institutional curation community engagement, and accessioning of archives for permanent management in centralized libraries.

Information Lifecycle in Cultural Heritage

Cultural heritage specialists may see their activities as oriented toward preservation of artifacts in situ. Their investigations constitute an information engine that collects, produces, refines and shares information about cultural heritage. To understand and evaluate this aspect of the study of cultural heritage it is helpful to have a conceptual overview of the lifecycle of information in a preservation enterprise.

A strategy for developing and preserving information can be broken down into four phases or institutional scopes which will be explored in more detail through exemplary projects introduced in the next topic and in the discussions of applications throughout the rest of this handbook.

1. Project-Based **Research and Information Development;**
2. Institutional **Curation and Asset Management**;
3. **Sharing** with Communities of Specialists and the General Public
4. **Preservation for Posterity**

Although these phases may be seen as a progression in terms of development, refinement and preservation of information, how these phases also fit together in a cyclical fashion, as depicted by the diagram in Figure 2, information products of one project become resources for future projects within the institution. Hopefully after the institution itself has disbanded the resources will be available for successive studies.

The topics that follow in this handbook have been selected to provide an overview of the new roles that digital information and associated tools, standards and strategies will make cultural institutions more effective at preserving heritage

Research Strategies and Tools

In the investigative phase of a project, several types of tools that come into play:

Research Tools and Document Management Tools introduce information stewardship practices into the research process providing a means of developing a repository of relevant documentation for a project. These resources are then more accessible to **Synthetic Tools** like catalogs, tags, search tools and georeferencing help bring order to previously unrelated documents -- uncovering new connections between artifacts and their cultural context.

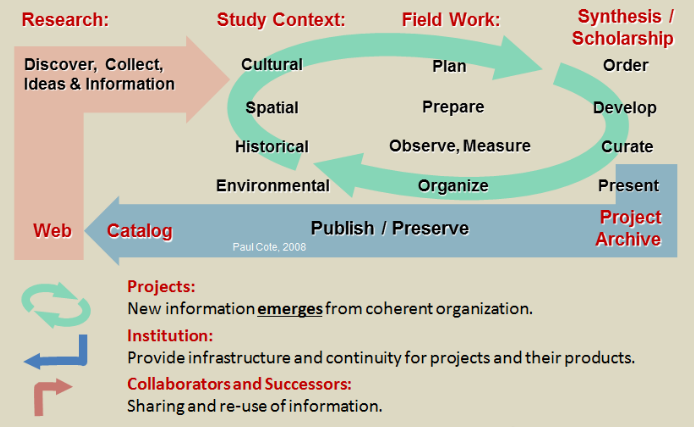


Figure 2: Lifecycle of Information in Cultural Institutions

Tools for Field Research make teams more effective in making and annotating the observations that an organization leaves for its successors. Ultimately, findings and recommendations may be developed and issued with **Exhibition Tools** that bring important aspects of heritage to light and generate valuable discussions among interested communities. These families of tools will be discussed and demonstrated in future topics of this handbook.

Institutional Strategies

Within a cultural heritage institution, projects may end or go dormant, while the information that has been gathered may be maintained to be picked up later by future projects or preserved within an institutional archive. With this in mind, the institution should have a strategy for information stewardship that aims to make sure that as projects end, the information that was developed can be preserved, discovered and used again. From this perspective, project document management tools provide a bridge between ephemeral projects and the continuing interests of the institution. Curation comes into play as specific assets are chosen to be preserved for future use and some subset of these is chosen to be shared with specialist communities or the general public. Therefore, institutional directors should be concerned that the document management tools used in projects are compatible with institutional strategies for preserving and sharing information.

Community Strategies and Standards

To recap the previous two headings, ne institutional strategies for information stewardship involve creating an infrastructure and capability for stewardship and more effective development of information throughout its lifecycle. This infrastructure involves critical hand-offs of information between systems and documents. All of this is made possible by standards.

Standards such as those that make the web work, or that allow libraries to share catalogs, or that embed essential information into photographs have become so successful that they may be practically invisible - -and yet they are at the heart of institutional strategies for information stewardship. So it will not be difficult to find the recurring theme of standards in the topics that follow.

Exemplary Cases

To discuss practices and trends in preserving heritage information it will be helpful to introduce some cases that illustrate trends in information stewardship in cultural heritage preservation. These cases will be considered in more detail as this handbook expands on specific tools, strategies and standards.

David Kruh and the Scollay Square Project

Scollay Square is a famous former neighborhood in Boston. Celebrated by many as the heart of the city from the mid-19th to the mid-20th centuries. Through neglect or evolution, Scollay Square went into decline and was abruptly demolished in an act of Urban Renewal in 1962.

There is a great deal of photographic and cartographic documentation of Scollay Square. Much of this documentation has been compiled by Mr. David Kruh along with primary research and oral history from people who worked and played in the Square.

Enter the term "Scollay Square" into the Google web search engine and you will discover Mr. Kruh's History of Scollay Square web site. Library catalog search will turn up David Kruh's two volumes on the square, Scollay Square, Published by the Images of America Series; and Always something doing : Boston's infamous Scollay Square, Published by Northeastern University Press.

Kruh's Scollay Square project will serve as our example of an individual project related to cultural heritage preservation. Mr. Kruh's web site is simple and effective means of informing the public of this important chapter in the heritage of Boston.

Cleveland Historical

Cleveland Historical is a project of the Center for Public History + Digital Humanities at Cleveland State University. From their home page: "Cleveland Historical lets you explore the people, places, and moments that have shaped the city's history. Learn about the region through layered, map-based, multimedia presentations, use social media to share your stories, and experience curated historical tours of Northeast Ohio.

"The Center for Public History + Digital Humanities seeks to transform how we teach, learn, and study history. In partnership with scholars, students, cultural organizations, and communities, CPHDH has developed state-of-the-art public history exhibits, interpretive research projects, teaching initiatives, and digital resources.

ClevelandHistorical demonstrates the use of Omeka, an open-source platform for document management, archiving and ecxhibit-building tool. may be explored through a mobile application or on a stationary web browser.

This exemplary project also demonstrates how multiple institutions, individual scholars and other community members can put their resources and knowledge to piece together important stories about Northern Ohio.

The Digital Public Library of America

Digital Public Library of America is not a physical library. Rather, DPLA is part of community of library experts, including their European counterpart, Europana, and OCLC WorldCat who develop open standards and tools that allow specialized collections to share their catalogs so that they all may be searched from one central clearinghouse. These union catalogs are essential for researchers who can discover information in places that they would otherwise know nothing about.

DPLA is also a federation of regional aggregators similar to ACIC or Digital Commonwealth. The collaboration among librarians on common standards for cataloging and sharing digital information resources multiplies the efforts of individual cultural institutions by providing a mechanism for making their resources discoverable, and in many cases for taking on the responsibility for long-term preservation and stewardship of digital assets.

The Google Cultural Institute

You may think of the world of cultural heritage on the web as a tree, where projects like Kruh's are the soil, ACIC are the roots, federated libraries, the branches, standards are the trunk. In a tree like this, **Digital Humanities** projects like The Google Culture Institute (GCI) are among the fruits.

GCI provides an authoring environment that people can use to develop and share ideas about cultural heritage by using digital artifacts that are accessed form a multitude of collections, like CornishMemory.com or DPLA.

The Getty Research Institute Vocabularies

To continue the tree metaphor begun above, if preservation strategies involve cultivating information for scholarship, then projects like the Getty Vocabularies represent a growing understanding of ecological activities that make the soil alive and recycle fruit into living scholarship that eventually enriches the soil. Based on the fundamental architecture envisioned by the World Wide Web Consortium, The Getty vocabularies illustrate the potential to bring new order to knowledge through concepts such as Linked Data and the Semantic Web.

Architecture of the World Wide Web

We are all familiar with the World Wide Web (WWW), but it will be worthwhile to label a few of the fundamental concepts that make the web such a useful and resilient tool for collaborating in research.

The topics that follow in this handbook will explore several tools and strategies for organizing and collaborating in research. Each of these will demonstrate how standards are evolving to expand the amount of content and the richness of opportunities for connecting resources together. Understanding the architecture of the web will help us to evaluate alternative strategies for making sure that the content we develop stays safe and accessible despite a technological environment that is marked by competitiveness, rapid change and obsolescence.

Resiliency: Stable Web Standards

Stable, open standards have made the web a widely used and resilient platform for sharing information. Web standards are developed and governed by the World Wide Web Consortium (W3C). The W3C convenes experts from industry and government and the user community to establish goals, and to develop principles and standards for encoding and exchanging information on the web.

The not-for-profit model for conception and governance of web standards has created an environment where individuals and institutions can develop content and applications without concern that their efforts will be made obsolete due to profit-driven decisions by software makers.

The diagram in the middle of this page depicts the basic architecture of the World Wide Web, and the standards for encoding, publishing and referencing information on the web.

W3C standards for encoding documents (HTML) and protocols for requesting and delivering them (HTTP) also have led to the commodification of free, open source software components like **standardized web browsers, web servers and authoring tools for web content**.

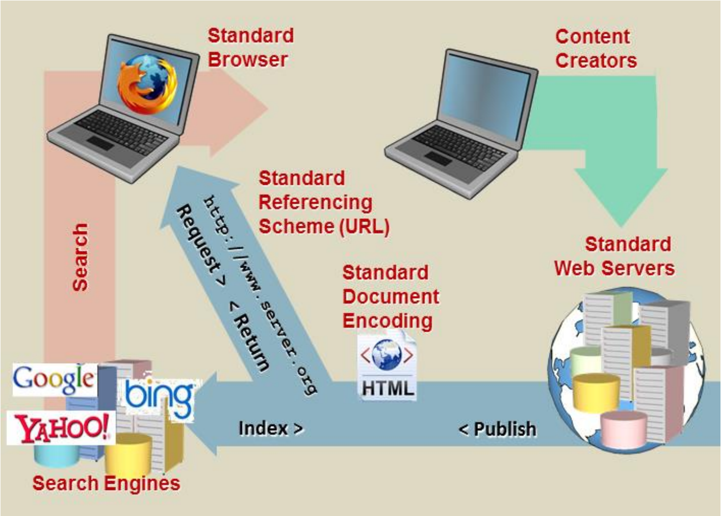


Figure 3: Architecture of the web.

A Simple Web Example

David Kruh's Scollay Square blog provides an example of the power of hypertext and the accessibility resiliency of the web. Mr. Kruh has been able to put together a rich collection of images and text from many sources. His blog includes links to many related resources on the web. Mr. Kruh has hosted his web page on a web server, which he has apparently developed in association with some of his other interests.

Web Discovery Patterns

The ability to easily create and publish and link together content is only part of the story. The third aspect of the architecture of the web are the patterns for discovering content as described in the visionary Web Services Architecture document developed by the W3C.

Indexing Search Engines

Google search is an example of an **Indexing Search Engine**. David Kruh's blog is easy to find on the web because Google has discovered it, and evaluated it as a rich resource regarding Scollay Square.

Federated Search Agents

Digital Public Library of America is an example of a **Federated Search Agent**. Indexing search engines like Google's employ a full-text approach to cataloging, federated search tools make use standards for tagging documents to develop catalogs that may be searched more systematically, by Author, Subject and Location, for example. These tools permit a much more systematic and predictable environment for discovery. Federated search engines will be covered in more depth in a future topic.

Aggregation Strategies

Both of the web discovery patterns discussed above, Indexing and Federated rely on a network of independent servers for the storage and delivery of the actual objects form the digital archive that is hosted by an independent collection. The federated catalog is an intermediary that provides references to these resources once they are discovered.

A second type of federation strategy involves duplicating the content and catalog information so that the discovery and the retrieval are both handled by a central archive. Aggregation strategies allow small collections to use a simple toolkit for their archiving needs, and let larger archiving hubs be concerned with the complications of wide distribution and very long term preservation. This strategy will be discussed in the federated Libraries topic later in this document.

Web Based Research Tools

Research and scholarship are always collaborative. A project begins with an exploration prior work left by previous projects. It proceeds through organizing and revising ideas, and recording new observations. If a project is successful, the record of contributing sources is preserved as references; and new information, ideas, policy and questions are published and extend the web of knowledge related to the subject.

Zotero: A Collaborative Cataloging Tool

An exploration of the research tool, Zotero will shine a light into the way that internet standards augment the processes and effectiveness of collaborative research.

Zotero is a free and open-source research tool that does three things:

1. **Makes it easy to compile bibliographic information about sources.**
2. **Creates a catalog of references that may be tagged, searched and re-ordered in useful ways.**
3. **Facilitates sharing and re-use of bibliographic information between computers and collaborators through the Internet.**

These capabilities make Zotero a worthwhile tool for solitary scholars. The benefit increases geometrically in collaborative projects. Zotero is particularly useful for keeping an annotated record of resources discovered on the web. Bibliographic records may also be created from scratch should references be encountered in print documents.

Demonstration: Using Zotero

To follow this example, you can install Zotero as an extension to the open-source Firefox web browser. Having discovered David Kruh's History of Scollay Square Blog, we might evaluate it as a record of research and an inventory of resources. Some of the photographs are associated with references to the archives where the originals may be found. Mr. Kruh makes good use of the capacity of HTML to include links to other interesting web sites. Some of these links are broken as the referenced resources have either been moved or have disappeared from the web.

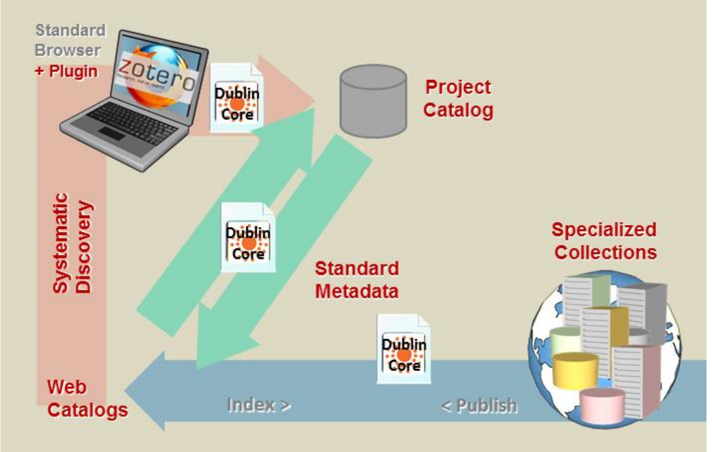


Figure 4: Zotero exploits metadata embeded in search results.

Mr. Kruh's web site has a wealth of information that we want to file away for future reference. So we push the Zotero button in the browser. Zotero creates a bibliographic record that includes the name of the blog, and the title of the page and the date that I accessed it. I can now fill in the author name and other tags. At the same time, Zotero saves a complete copy of the page and all of its images and references for safe-keeping in case the original site disappears from the web.

Facilitating Research and Documentation

As I continue my research, Zotero saves records to my private catalog on Zotero.net, which will be available on any computer I may be using on the internet. I can share selected folders on my catalog with my collaborators so that we can immediately share our discoveries. The network-based catalog multiplies the value of the individual research efforts of collaborators.

When I use or refer to one of the resources that I have discovered, and need to include a formal bibliographic reference, right-click the entry in my Zotero catalog and choose "Create Bibliography from Item."

Embedded Metadata

Next, I will go the federated library search tool, OCLC WorldCAT to see if I can find any books about Scollay Square. Here I find a reference Mr. Kruh's books including Always Something Doing. When I press the Zotero button on my browser and choose *Save to Zotero using COINS*, Zotero saves a much more complete bibliographic record that includes not only the Author name, but also the Publisher and year of publication. Zotero is able to extract this information from structured metadata (data about the data) that has been purposefully embedded in the catalog pages published by WorldCat and other libraries.. The interoperability of library catalogs for personal and project-based research tools is an important aspect of the new modes of scholarship that are enabled when institutions employ community standards for archiving and cataloging.

More about COINS and other key metadata tools, standards and strategies are discussed in the next topic.

Metadata: What is It?

The previous topic on Zotero introduced the idea that a researcher can build a shared digital catalog of resources related to a project. Tools like Zotero introduce a new epoch in scholarship. Formerly, the development of catalogs was handled by library specialists. Now, researchers can build and manage their own catalogs. Thanks to web standards for metadata, a tool like Zotero makes this momentous shift in capabilities almost effortless.

The idea that researchers may be involved in developing catalogs can have a large impact on the lifecycle of information in preservation studies. To understand and exploit these impacts it will be helpful to understand a few of the fundamental concepts, terms and trends related to metadata.

Classic Bibliographic Records

Metadata is descriptive information about documents. In elementary scholarship we are all exposed to metadata as a bibliographic entry that provides a reference for a document that includes Title, Author, Publisher and Date and Place of Publication. Bibliographic references included in a document connect that document to the overall web of scholarship.

Digital Metadata Standards

Federated search strategies and the architecture that allows Zotero to harvest catalog records are made possible by new standards for metadata content and encoding. Standards for metadata content provide, in conceptual terms, what aspects of a document ought to be described. An encoding standard describes the way that the metadata can be encoded so that it can be written by one tool and used by another.

Metadata Content Standards

Specialized and activities have specific needs for describing documents. For example, the International Press Telecommunications Council (IPTC) has described a metadata content standard for describing electronic photographs used in a journalistic context.

In 1995, the Online Computer Library Center, a digital library interest group, hosted a meeting at their headquarters in Dublin Ohio to discuss an architecture for sharing metadata on the internet. This architecture described a core content standard or metadata that would apply to all sorts of documents. This standard became known as the Dublin Core metadata standard.

Abstract Model

You can think of the Dublin Core first as an abstract model that defines what a document – or resource -- is, and how it is related to its descriptive information. This is not as straightforward as it sounds because in the digital world, a document may have a paper original and also digital surrogates that vary in quality.

Terms, or Tags

The aspect of metadata that most people interact with in catalogs is s set of tags, or terms, that describe aspects of documents. These tags include the classic bibliographic terms: Title, Author, Publisher, Publication Date, etc. And many more. Each term has a definition that can be used to determine whether or not a given value is appropriate for that tag. Most cataloging systems do not require all of the Dublin Core tags to be filled out.

Application Profiles

The Dublin Core model also established a means of extending the list of core tags or for adding fields to suit the additional descriptive requirements of specialized content domains. These extensions are known as Application Profiles. For example, the IPTC has recast its content standard for photographic metadata as a special application profile of Dublin Core. The DPLA application profile extends Dublin Core with several specific concepts and tags.

Metadata Encoding Standards

While content standards for metadata make it possible for a variety of institutions to coordinate their documentation, creation of shared catalogs and the metadata harvesting capability of tools like Zotero also requires a predictable standard for encoding metadata. We saw this in the previous topic when we instructed Zotero to gather metadata from the WorldCat search return page for David Kruh's book. In this case, the WorldCat encoded Dublin Core metadata into their web page using the encoding standard, Context Objects in Spans, which is a means of embedding metadata in HTML documents.

Later in this handbook we will look at a few other very useful encoding standards for metadata: The Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) which is used by DPLA and Europeana for harvesting catalog records for their union catalog. The Extensible Metadata Platform (XMP), is an international standard for embedding metadata into a variety of digital documents.

A recurring theme In all of the topics in this handbook will be the utility of keeping digital documents associated with metadata. Open standards for metadata content, encoding and embedding helps us to keep track of the provenance of information resources and how they are connected to their greater cultural, temporal and geographic context.

Library Web: Federated Libraries

Everyone who has used the web is familiar with search engines like Google. According to the W3C Reference Model for Web Services, discussed in a previous topic of this handbook, the Google search engine is an **indexing discovery agent.** Google is omnivorous. It attempts to read and record the frequency of almost every word. This sort of index of the entire web is a useful, yet very general means of discovery.

Specialized professional communities (like cultural heritage specialists) have created search engines that are more articulate than Google for discovering and linking together information about particular sorts of things. Specialized library search engines create union catalogs by harvesting and indexing standardized metadata from their network of independent collections. In the terminology of the W3C Web Services Architecture these specialized search engines are called **Federated Search Agents**.

Union Catalogs

Some of the more successful federated search agents are maintained by coordinated federations of libraries and cultural heritage institutions: Digital Pubic Library of America and WorldCat, Europeana, OpenGeoPortal, and ArchiveGrid.

These institutions maintain a union catalog in which each record has a link to the resources which are accessed from the publishing collection. The federated union catalog presents a single search interface for all of the holdings of the federation with a single search. Considering the thousands of libraries and other cultural institutions involved, participation in library federations is necessary if an institution is making a sincere effort to make their holdings discoverable.

Federated union catalogs are centers for innovation in search and discovery tools, thus they consolidate expensive research and development and allow individual collections to focus on collecting and preservation of resources. Some of the innovative searching capabilities that will be illustrated in future topics in this handbook include: map search, faceted search and connecting resources with semantically linked data. Many of these advanced search methods allow investigators to use patterns in the ways that resources are tagged – (for example geographic and temporal patterns) to understand how resources and ideas are related. The power in this pattern-making capability is strengthened by the quantity an diversity of resources that may be searched through the union catalog.

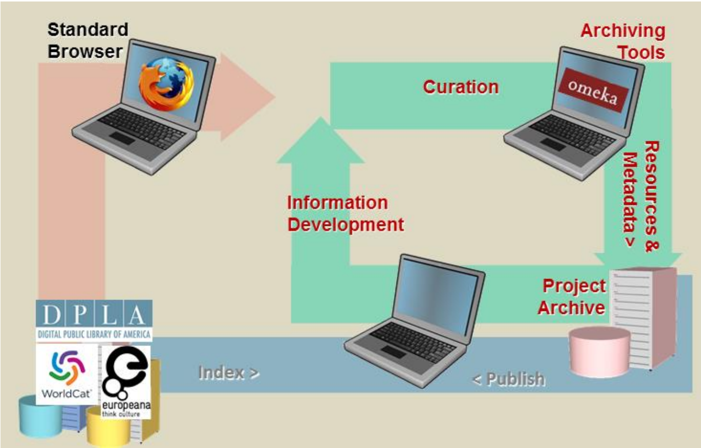


Figure 5: Federated catalogs link specialized collections together.

Registries and Stable Web References

Another advantage that union catalogs provide is that they provide a stable means of referencing resources even if the source institutions are not able to maintain their internet addressing schemes indefinitely.

Metadata Publishing and Harvesting

Rather than trying to index everything on the internet, federated catalogs depend on participating institutions to deliberately publish their metadata, which is then harvested by the federated cataloging system. Users search and discover resources and then find the digitized resources themselves from the originating institution.

Open Archive Initiative OAI-PMH

One successful protocol for publishing and harvesting metadata has been devised by the Open Archive Initiative protocol for Metadata Harvesting, (OAI-PMH)

Aggregators

The W3C reference model describes several other types of discovery agents. Aggregators are another type. These are agents that do not merely create catalogs of metadata, but also collect and preserve digital assets themselves.

Digital Public Library of America uses a federated strategy involving 50 state-based aggregators or Hubs. While the DPLA itself is only a catalog, the hubs build actual collections of digital artifacts – which in tern may be surrogates for the physical artifacts and documents that might be held in specialized collections throughout the state. This hub strategy relieves smaller institutions of the difficult technical task of digitizing and maintaining digital repositories that must meet the demands of very long term preservation of artifacts in the face of the evolution and obsolescence of data formats, data corruption and natural disasters.

Wikimedia Commons and History Pin (to be discussed in a future topic) and the commercial web sites, Flickr and Google Cultural Institute are also aggregators. These sites may play a role in making resources more broadly accessible.

Institutional Stewardship Strategies

The web and federated library search are a boon for historians, preservationists and researchers. While we appreciate the easiness of discovery and accessibility, connecting the holdings of institutions and research projects to this world requires that leaders of diverse institutions take strategic initiative.

Planning and Costs

This initiative will probably require that members of the organization learn new things and take on new responsibilities to plan and implement new practices of information stewardship, and there will be additional ongoing costs as well. A piecemeal approach to cultivating an institution's information resources is likely to fail. Success will involve planning from end to end of the information lifecycle within the institution.

Benefits of an Information Stewardship Strategy

Many institutions involved with cultural heritage preservation have a mandate to preserve information that they create. Professionals in this field collect and create new observations and documentation through their projects that preserve aspects of heritage, even when the physical resources are destroyed or removed from their context. In this sense, it may be considered an imperative of the preservation enterprise to participate in federated networks for information preservation.

An institutional strategy for information stewardship will bring benefits easier handling and accessibility of information within the institution, which will bring greater effectiveness in investigation and planning of projects. The institution will also see improvements in its ability to communicate with the public and benefactors about the value of its work.

Lifecycle Planning

The stewardship strategy will consider the ways that information is gathered and created within the institution and how it is used in research, planning and exhibits. Each of these production tasks will benefit from a holistic information stewardship plan.

Figure 6: Lifecycle of Information in Cultural Institutions

Perhaps the most important part of the information lifecycle is consideration for how the institution plans provide for the care of its information assets in a very long term (perpetual, if possible) time frame. It may be most effective to consider this very long term aspect of information stewardship first.

The long-term aspect of stewardship will probably involve partnerships with federated libraries. In the United States, this might be the regional hub for DPLA. In Europe, Europeana, or WorldCat ArchiveGrid,. Considering the end-point will establish the sorts of metadata standards that will be desirable to cultivate at earlier lifecycle stages of information within the institution.

Asset Management / Repository Strategy

Near the core of the institutional stewardship strategy, there will be a need for tools for creating digital information assets and associated metadata. These tools are sometimes known as **Digital Asset Management Systems** or **Repositories**. Their basic functions include the following:

- **Author and manage metadata assigned digital and non-digital artifacts;**
- **Present a searchable catalog;**
- **Provide network access to digital documents accessible with varying levels of access.**
- **Exchange metadata and digital objects with archival systems including federated libraries and aggregators.**

In the next topic we will explore a system that meets these goals.

Evaluating an Archival Strategy

Given that important heritage information is collected in a wide variety of smaller cultural institutions whose leaders are neither experts in information technology nor library science, there is a need for published norms for information stewardship that can be used to evaluate local information stewardship processes. The Society of American Archivists provides one set of norms for evaluating an institutional Archiving Strategy. This standard defines the fundamental concepts and concerns involved in archival systems. A key concept to consider is that the preservation of cultural heritage requires that the different types of documentation gathered or created by an institution should out-live the institution itself. Therefore the approach to preserving and sharing information resources is a responsibility of cultural heritage and humanities experts and cannot be completely delegated to staff.

The Archivist Role

It cannot go without saying that organizing and continued preservation of information resources is an ambition that cannot be left to chance or to junior staff. If regular procedures for information stewardship are not set out and supervised on an on-going basis by an archivist, then one can be almost certain that the information legacy of projects and investigations will be unrecoverable after a project concludes. The following is a short list of responsibilities for the archivist role in a project or institution.

- **Establish standards for metadata and controlled vocabularies.**
- **Provide easy-to-use repository systems and training to investigators**
- **Make backups, and test recoverability on a regular schedule.**
- **Promote selected information assets from individual investigations to institutional archive.**
- **Maintain relationship with an independent aggregating archive or depository.**

An institution that has an ambition to investigate and preserve the memory of cultural heritage that is not paying active attention to these aspects of the archivists role, will likely see all traces of its efforts be completely forgotten in time.

Open Archives: Preventing Lock-In

Eventually all systems become obsolete. An important role for the archivist is to make sure that the information assets of projects and of the institution may be replicated in a repository system other than the one that is in use at the time. The capability to migrate the digital assets to another system is a matter of understanding how to systematically address each item in the repository and to create a dump that maintains all of the critical information and inter-relationships between every element in the archive, such that each piece and every relationship can be reconstituted in a new system.

A technical viewpoint on archival strategies is provided by the Reference Model for Open Archive Information Systems (OAIS), an international standard. The OAIS establishes a conceptual model for an archival system that is capable of migrating its assets in the manner described above. The same standard may also be applied to a system that is capable of being the destination for such a migration. Theoretically, the OAIS provides a certifiable benchmark for establishing whether a given system can be depended to preserve institutional memory in spite of the continuing march of technological obsolescence.

Replication and Aggregators: Lots of Copies Keeps Stuff Safe

The catch-phrase, "Lots of copies keeps stuff safe." is the name of a project originating in the Stanford University Libraries. The LoCKSS project is oriented toward replicating and keeping local copies academic journals at subscribing institutions. The goal is to assure the preservation of academic knowledge in the event that the publisher folds, goes off-line or simply cancels subscriptions. LoCKSS is a restatement of the principle against keeping all of your eggs in one basket.

Although LoCKSS is a project of its own, the general concept can be applied to cultural heritage preservation in general by actively copying assets and associated metadata to distributed hosts. One strategy for this would include publishing selected assets to aggregating libraries. In the case of images, it seems that some libraries use Flickr for this role. One could also see a similar role for Wikimedia Commons.

Embedded Metadata and Replication

The hope expressed by the Lots of Copies Keeps Stuff Safe principle is that the best way to keep cultural heritage alive is by distributing it. One of the difficulties of this – as expressed in the Reference Model for Open Archival Systems – is keeping metadata and digital documents together. David Kruh's History of Scollay Square blog illustrates this problem. Mr Kruh has invested effort in collecting images and annotating these with contextual information. Unfortunately, for many of the images included in the blog, information about provenance has been lost.

Technology for embedding metadata into documents, including the Extensible Metadata Platform (XMP) was discussed in the earlier topic on Metadata. In the future, if the principles of Open Archival Systems included the embedding of metadata into documents themselves, this could assure that when images were copied and re-used, the cultural associations of these images will be preserved. The topic, Managing Geotagged Photographs, later in this handbook will explore tools and strategies for creating and managing embedded metadata.

Omeka: A Simple Repository Solution

The previous topic discussed the role of a repository system: to provide a resilient container for information assets and associated metadata. This topic will explore an example of such a tool. [Omeka](#) is a web-based platform for developing an archive of well-documented information resources. Omeka has many plugins that make it relatively easy to exploit these assets in exhibits and other public engagement applications. Omeka is being actively developed as an open source project of the Roy Rosenzweig Center for History and New Media at George Mason University -- the same group that created and maintains Zotero.

An Archival Repository

Resources in the Omeka archive are known as **Items**. Any item may be assigned a level of access at different stages of its lifecycle. Project researchers can develop resources with metadata that begin as provisional private resources for a specific project. Subsequently, these resources and their metadata may be vetted by a curator and made accessible to the public via the Omeka search or browse tools. Items may be come parts of thematic collections or narrative exhibits.

Metadata Authoring Environment

Adding an item to Omeka, is a relatively simple process of uploading it and filling out a Dublin Core Metadata form. Specific metadata requirements for items of different types can be set by the super user. Once created, an item is available for searching, or browsing or inclusion in exhibits. Any time the item is returned by Omeka, it is always associated with its metadata.

The Georeferencing Plugin

There are many plugins that can be added to Omeka including a tool that permits users to precisely geotag the location of an item's subject. Geotagged items may then be searched and discovered by a map search.

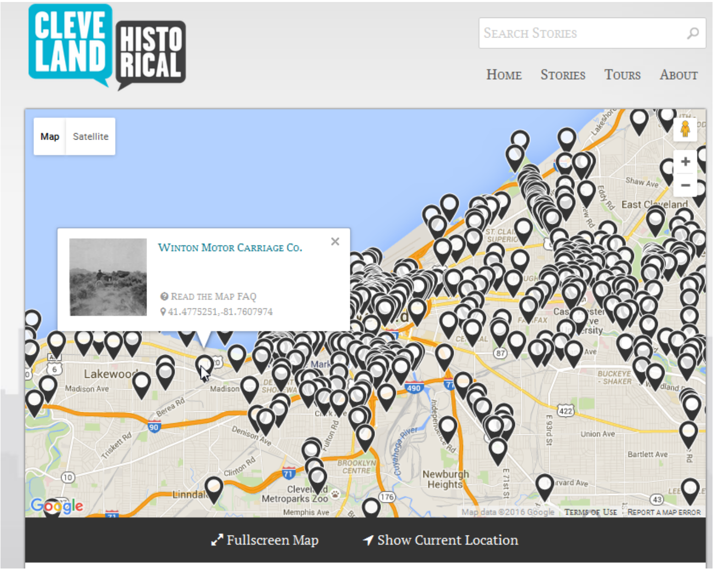


Figure 7: Omeka provides many ways of creating and preserving order in disparate documents.

Participating in Federated Libraries

Omeka is capable of publishing metadata to federated catalogs or formatting packages of resources and metadata to send to aggregate collections using the standard OAI Protocol for Metadata Harvesting discussed previously in the topic on federated libraries.

Community Engagement

Administrators may also enable commenting capabilities to the public or to selected communities of registered users. This can be a way to engage the public in helping to identify the places people and things depicted in documents. [ClevelandHistorical.com](#) is an example of an Omeka project that supports community participation.

Digital Humanities

Omeka is an amazing Swiss-army knife for cultural heritage preservation. In addition to providing an easy-to-use repository system, Omeka is also a platform for creating exhibits – as we will discuss in a future topic of this handbook.

Setting Up an Omeka Database

Omeka is free to install on any computer running Linux and the basic web server software (Apache, MySQL, PHP). The most straightforward way to get an instance of Omeka running is through a hosting service like [Omeka.net.](#) Or one can install Omeka on a server that has been set up through a service provider such as Amazon Web Services. Please refer to the appendix for more information.

One or more people are designated as Super Users who can then create accounts with various levels of access: contributors can create new items and add and edit metadata. Researchers can browse and access all items, and the public is able to access the archive over the web and search and discover items that have been designated as Public.

Alternatives to Omeka

Omeka has advantages of being relatively easy to install and maintain. But this topic should not be seen as recommending Omeka for every institution. Omeka has several competitors in terms of project-based asset management (one example is [ResourceSpace](#)) to institutional repository systems (one example being [Fedora Commons](#).) Choosing tools to facilitate work and to manage assets in a particular setting will require a careful assessment of the institution's needs and capacities.

Web Based Scholarship and Exhibits

So far, the topics in this handbook have focused on the means of attaching metadata to objects for the purposes of archival preservation, discovery and sharing. For organizations trusted with cultural heritage preservation these virtues may be considered necessities. The next several topics are going to look at new modes of scholarship that demonstrate the benefits of information stewardship toward better understanding of heritage as the connections between and among particular information resources.

Information Ecology

Scholarship involves putting ideas and prior documentation into meaningful order. Maintaining proper references in publications has always been a tedious chore for writers and exhibit curators. So it happens that digital archiving, web-based catalogs, and embedded metadata (as demonstrated with Zotero earlier) are creating a new culture or ecology of information, where resources and references are more positively attached. This association of a document with descriptive metadata also serves to connect the artifact with its cultural, temporal and geographical context.

Hypertext Exhibits

One of the simplest and yet revolutionary aspects of the World Wide Web is simple means of encoding documents with text and images that link to each other through Uniform Resource Locators commonly referred to as URLs. Documents that link to eachother with URLs are known as Hypertext. Thanks to the international open standard for encoding hypertext (HTML) practically every computer and smartphone has a browser that can render hypertext and follow references from one document to another.

For a basic hypertext exhibit, let's consider David Kruh's [Scollay Square History](#) web pages. There is no questioning the value of this resource that is and so full of information and so easy to access. Mr. Kruh has gone to a lot of effort to obtain old photographs and narrative history. Yet, there is some room for improvement in the lack of attribution or bibliographic references that might help us to find the original source documents if we were interested. To create a site that has all of the proper attribution may be as much as five times as difficult as getting the most interesting and valuable information into place. The added value of thorough attribution may not be seen worth the effort, except perhaps in peer-reviewed scholarship. What scholars like Mr. Kruh need is a hypertext authoring toolkit that reduces the drudgery of maintaining these references.

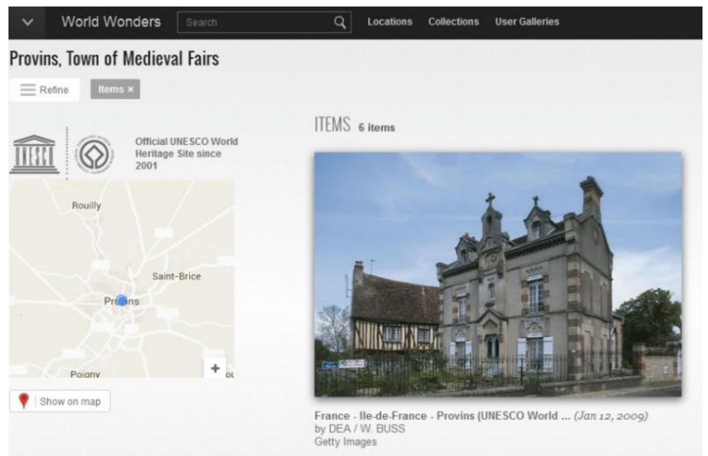


Figure 8: An exhibit item from Google Cultural Institute

Next-Generation Web Exhibits

At least a small aspect of scholarship is about putting prior work and new ideas into meaningful order. In web-based scholarships many ways of ordering information may be employed within a single document or exhibit..

Google Cultural Institute

To quote from the web site of the Google Cultural Institute (GCI): "For hundreds of years, cultural institutions have collected and safeguarded our history and heritage. Powerful technologies can amplify this mission, while preserving these artifacts for a worldwide audience today and tomorrow. "

Spend some time exploring a few of the exhibits created with the on-line tools accessible to anyone through the GCI. It is folly to try to explain these when you could simply scroll and click through a few of them yourself.

A few common patterns are worth discussing in this context. Taking the World Wonders exhibit as an Example: we can appreciate the way that the map makes use of georeferencing tags in the metadata to associate exhibit items with each-other or to other places that may be of interest. Second, for any resource that is on the screen, you can click the Details Tab to vreveal metadata. The association of subjects with panoramic photography collected by the CGI team is also remarkable.

You can learn more about how this the GCI Exhibits tool works by registering and creating an exhibit of your own. Once inside, you will see that the exhibit-building interface makes it very easy to capture digital artifacts from archives all over the world and to add your own commentary an associations with other artifacts.

Omeka Exhibits

Omeka also has a very nice exhibit-making capability that you can explore on the Omeka Showcase. Digital Public Library of America also has an extensive collection of Omeka exhibits.

Linked Data and the Semantic Web

Catalogs provide machinery for searching and associating information resources. A catalog search may reveal order in a set of otherwise independent items that happen to share common tags. This effect is enhanced by tagging conventions known as controlled vocabularies. Controlled vocabularies like the Library of Congress Subject Headings have been in use for over 100 years.

Controlled Vocabularies

The simplest way that independent items may be linked though controlled vocabularies is when they share an identical keyword, such as Author Name. But controlled vocabularies may potentially link items when keywords are synonymous – for example one document referring to Samuel Clemens and another Mark Twain. Controlled vocabularies may also serve to link cataloged items through taxonomic relationships for example one item tagged as Porcelain may relate to another item tagged as Stoneware, since each of these material keywords is related to a parent class: Ceramic.

The employment and expansion of this practice has multiplied since the advent of digital libraries. It is important for cultural heritage professionals to understand this phenomena for two reasons: first, so that they can use controlled vocabularies to make their information more discoverable and connected, and second so that they may recognize an opportunity to participate in the development of vocabularies and taxonomies and similar resources that may be their particular area of expertise.

Faceted Search

In our exploration of Digital Public Library of America, we observed the utility of facets, displayed along the left side of the search return page. The facets allow you to refine a search by showing a listing of the Subjects and Authors that occur within the list of items returned by our original search. These connections of one resource to another may take our discovery process into entirely new and useful directions. For this to work, people involved with cataloging must understand and use controlled vocabularies.

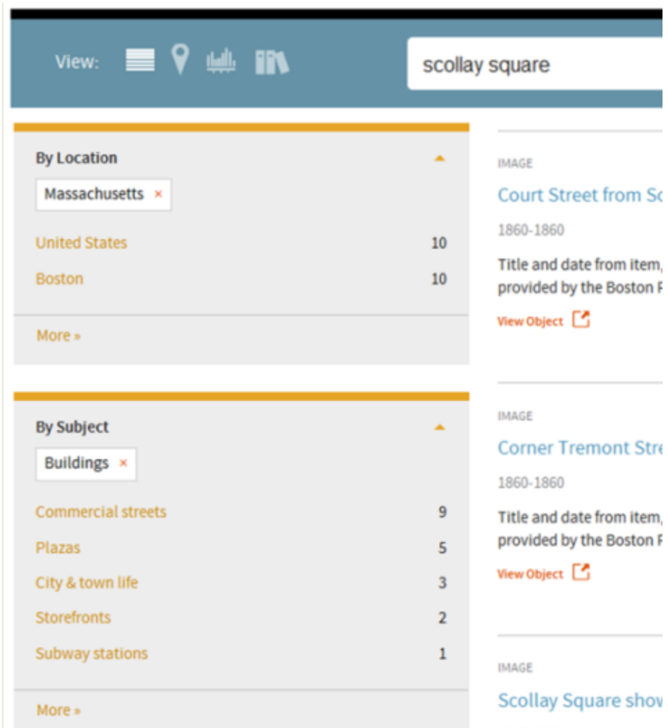


Figure 9: Faceted references in DPLA.

Getty Vocabularies

To get a look at a sliver of this world of naming authorities, we will explore the Getty Library Vocabularies for Art and Architecture. An initial search for Buckminster Fuller as an architect, reveals his authoritative name record, but also links Mr. Fuller with the name of his firm, academic history and collaborators.

Logical Referencing Systems: Ontologies

The Getty Vocabularies for Art and Architecture demonstrate how logical systems for referencing and connecting information allow order to emerge in collections of millions of things that are cataloged by diverse and distributed individuals who are not necessarily collaborating on purpose, but whose behavior constitutes a self-organizing collective intelligence.

These systems are known as **Linked Data**. Often they are similar to taxonomies or hierarchically organized thesauri. Sometimes they are more complicated. Notice how the entry for Buckminster Fuller in the Getty Union List of Artists Names links him to a rich tree of biographical connections.

The Semantic Web

The Getty Vocabularies demonstrate how diverse knowledge may be accumulated within a framework or network that can be used to automatically place new information into context. This idea is one of the foundational capabilities envisioned by the World Wide Web Consortium that is only slowly being realized.

The conceptual model and the encodings and protocols for the semantic web have reached a provisional state of maturity at this writing, and projects like the Getty Vocabularies and DB-Pedia are proving the capabilities. One of the challenges of bringing the Semantic Web to fruition will be in adjusting the ancient culture of scholarship to promote sharing of knowledge by heretofore independent an insular scholars. For a model of how this could work it is useful to explore the outreach efforts of the Getty Vocabularies project.

Given that each cultural heritage institution has some facet of the world that is its specialty, the leaders of these institutions should have a plan for seeing that their expertise will find its place in the semantic web.

Maps on the Web

The first topic in this handbook introduced an idea that cultural heritage preservation is connected with the understanding of situations. Situations involve context: artistic trends, historic political drama, the interaction of personalities, cultures and places. Geography and spatial circumstances are among the inter-tangled aspects of situations as they relate to cultural heritage. So maps, old and new, are very useful information resources for preservation studies.

Exploring the ways that maps are being organized and accessed on the web illuminates more exciting trends for connecting resources introduced in previous topics: Federated Libraries, Web Standards, Metadata, Discovery Systems, Georeferencing and Linked Data.

Specialized Collections and Federations

Geography libraries and map collections have long been centers of specialization in libraries. This may owe to special storage and cataloging challenges that maps present. Thanks to large format scanners and networked, federated systems for sharing and using them, historic maps are now among the most exciting resources for digital humanities applications.

In the context of our discussion of georeferencing and linked data in previous topics, maps must have tremendous utility for helping people link pieces of information together by providing an information framework of contemporary or historical landscapes. We have already observed how contemporary base maps and aerial photographs provide a powerful and intuitive means for exploring and discovering data (in DPLA and ClevelandHistory.org) One thing that we have not seen yet is the added dimension added to this understanding when the base map is a scanned historic map.

Demonstration: Open Geoportal

Open GeoPortal (OGP) is a federated library project, like WorldCat or digital Public Library of America. Several geography and map libraries and a few government and non-governmental agencies have indexed their metadata and shared it with the OGP consortium to provide a one-stop discovery tool for geographic data.

To demonstrate OGP, first use your middle mouse button to man and zoom the map to the Boston area. Use the Advanced Search options to take a look at the list of organizations contributing their metadata. Use the Data Type pulldown to limit the search to Scanned Maps. Now as you zoom in, you will see that the list of returned maps changes to list only the maps whose coverage is contained within the window.

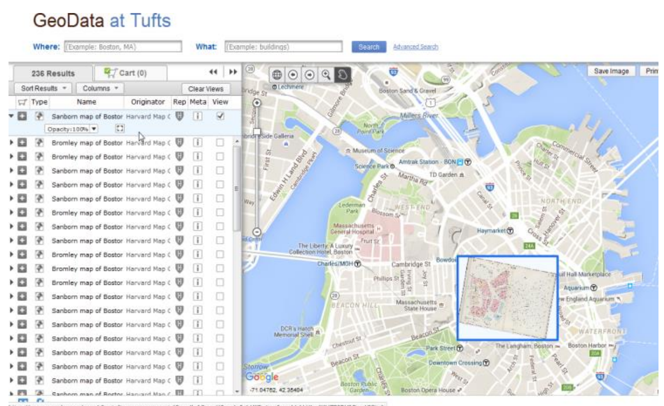


Figure 10: Bounding-box search and map preview in Open Geoportal

Bounding Box Search

One thing to note here is that the metadata for maps contains not merely one pair of coordinates to designate a point, but two pairs, to designate a rectangular domain of space. If you hover over the listing for one of the maps in the left panel, you will see its bounding rectangle on the screen!

Georeferenced Maps

Use Harvard's instance of OGP to find a scanned map that covers a small area in Boston, and click the preview check box. Notice that the map shows up in the proper alignment on the reference base map! This occurs because of an open, community-governed web protocol for displaying geographically referenced maps called Web Map Services, which we will discuss in its own topic.

Other Sources of Old Maps on the Web

Open GeoPortal is a federation of several map collections. There are a few other web-based search tools for old maps that are worth taking a look at:

- The David Rumsey Collection
- Old Maps Online

Maps as Linked Data

Our discussion of linked data and the semantic web in the previous topic explored an example of how biographical information about artists and their relations to projects and time periods creates a logical referencing system that discovers order in independent ideas or particles of knowledge. When new information is connected to these semantic networks, new information emerges about cultural context.

The conceptual taxonomies and hierarchies involved in projects like the Getty vocabularies may be relatively abstract. The ontology of space and time as a means of organizing cultural information is a demonstration of complex logical referencing that most people understand intuitively.

Geographically Linked Data

Geography as a Linked Data Ontology

It is fun to think about maps, geographical coordinates and temporal references form an ancient approach to linked data. A map is a logical/meaningful framework. Spots that are close to each-other on the planet and existing at the same time are related by physical coincidence and probably other cultural relations..

So items in separate repositories that have been tagged with simple dates and geographic locations – even place names --, can fall into place and lead to one-another's discovery through any number of search approaches.

Demonstration: Historypin

There is hardly a better example of Geography as Linked Data as [HistoryPin](), a product of a social networking startup named [ShiftDesign](). Per their own web site: "Historypin is a way for people to come together to share and celebrate local history. It consists of a shared archive, a mutually supportive community and a collaborative approach to engagement with local history."

To demonstrate the extraordinary ability of Historypin to contextualize information you might go to the map and zoom in to Boston. You will see a display of bubbles indicating clusters of resources. Refocusing the map adjusts a geographic filter for the resources presented in the right-hand panel of the screen. You may narrow your search to a particular historical epoch by shifting or narrowing the brackets on the time slider at the right side of the screen.

Interoperability with Google StreetView

A most interesting aspect of Historypin is that it is possible to tag photographs not only with their location, but also with the direction of their point of view. If you enter the search term, "[61 Arborway, Jamaica Plain]()" HistoryPin will return a photograph of a house that has been tagged with very specific metadata regarding the location of the camera, the direction that the camera was facing and even the focal length of the lens.

You should now see an historic image superimposed on a modern perspective image from an independent collection of images captured by Google's mobile image collection trucks. By dragging the transparency slider at the top of the image you can see how the photograph aligns with the current scene.

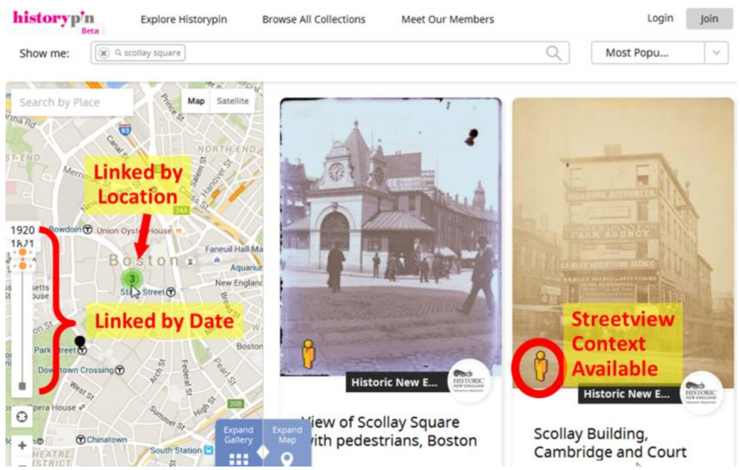


Figure 11: Historypin linking historic images to Google's Street View

Google Street View API

Google's vast collection of Street View images are captured by a special truck that has driven practically every street in the world capturing several frames per second in several directions at the same time. Each frame is tagged with the coordinates of the truck's location when the image was captured, along with information about the three-dimensional direction the camera was pointed and other parameters regarding the optical properties of the camera. Any photo can be accessed via a URL that is encoded with the a place and a direction. For more information on how this works, you can visit Google's documentation on how to retrieve images using the [Streetview Application Programmer Interface (API).]()

There are several interesting discussion questions we could pose that may highlight the vast opportunities of linked data:

- How is the addressing system of latitude, longitude, and three dimensional direction similar to URLs?
- How is it similar to the Getty taxonomy of artists' names?
- Wouldn't it be cool if HistoryPin allowed users to contribute georeferenced historic maps? Or if HistoryPin made use of the georeferenced maps shared by members of the Open Geoportal federation?
- Won't this be a very interesting application when the Google Street View images are 500 years old?
- Do you believe that this privately run application and its repository will be accessible 500 years from now?

Four-Dimensionally Linked Data

Cultural heritage is to a large degree a matter of understanding historical information in context. The potential of the sort of three dimensional referencing demonstrated by the Google StrEet View API is remarkable in that a new understanding of historical context may be discovered in photographs that were taken in a certain place in time – even if the objects captured in the phot were never listed among the keywords or descriptions attached to the metadata for those photos.

Georeferenced Maps and Web Map Services

The prior topic, Maps on the Web, we discovered an old map of Scollay Square through the Open GeoPortal. OGP was able to preview in near perfect registration over a modern map – right on our computer screen. This ability to juxtapose related resources is made possible by an open web protocol named Web Map Services (WMS). This protocol for sharing georeferenced maps is a standard governed by the Open GeoSpatial Consortium (OGC).

Georeferencing

There are several pieces of the problem of making maps overlay with each-other on the web:

- First, a person must assign latitude and longitude coordinates with specific places in the scanned map image.
- Next, software is used to transform the image to a specific map projection. This process of encoding an image with earth coordinates is called **Georeferencing**.
- Next, the image must be placed on a web server that is able to take a reference to a specific map and area of the planet, and then return he appropriate area of the map at the correct resolution. This is the **Web Map Server**.
- Finally, a web developer must be able to create web pages that will formulate the right sort of request to a map server when a user pans and zooms a map widget on any browser. These components are known as **Javascript Libraries** and the **WMS Protocol**.

Miracle of WMS: Stable Open Standards

From a biological and sociological point of view it is amazing that human beings are able to collaborate to such a degree. This is testimony to the world of web standards governed by the World Wide Web Consortium and the International Standards Organization (ISO). These agencies are non-profit consortia that involve stakeholders in the development and governance of standards. It is due to this collaborative governance that a broad community of contributors feels confident enough to invest time and effort to add meaning to content by means of georeferencing and to build web applications that use these maps.

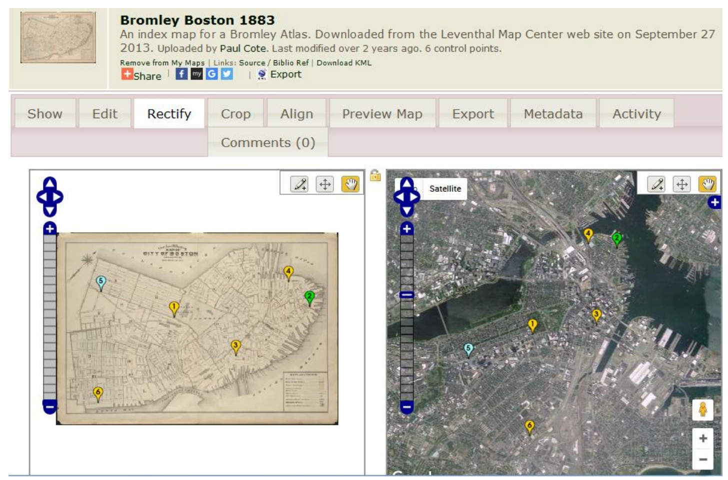


Figure 12: The MapWarper tool.

Web Based Georeferencing Tools

Ordinarily, the process of georeferencing and transforming scanned maps is carried out using a desktop Geographic Information System like ArcGIS or QGIS. There are a couple of exciting web sites that allow georeferencing and WMS hosting through easy-to use web interfaces. The David Rumsey Map Collection is one example. For my Scollay Square project I used Harvard's World Map Map Warper. As illustrated in the center of this page. MapWarper is an open source project originally developed by Tim Waters for the New York Public Library and Topomancy LLC. As a standards-based open source tool it has been incorporated into many useful web sites.

Harvard World Map / Map Warper

The map warper interface allows any user to register and upload their own scanned map. Then, control points are chosen from the map and associated with points on a contemporary base map or aerial photograph as shown in the figure on the left. When four or five control points have been added, the registration of the historic map can be checked and control points adjusted as necessary. When the user is satisfied, the georeferencing can be concluded, and the historic map will be available as a Web Map Service that can be used in digital humanities applications such as Omeka Neatline (which will be explored in the next topic.)

Web Map Tile Services

The web server extensions that enable classic Web Map Services to work require a higher level of hosting support than entry-level applications like Omeka. and demand an extraordinary amount of system resources. This means that web map services are more difficult to host than a simple Omeka server.

A relatively new standard for sharing georeferenced maps employs a clever means of breaking maps into according to a standard geographical grid. These tile sets may be requested and delivered using a protocol very similar to WMS, but handled from a standard web / PHP server. This protocol is known as Web Map Tile Service. The WMTS standard is also governed by the OGC. Hopefully this means of sharing georeferenced maps will soon be built into tools like Omeka and Historypin.

Geo-Temporal Exhibits: Omeka Neatline

The multi-purpose one-stop digital humanities suite, Omeka, has been featured in three or four topics in this handbook already: as an archival repository, searchable discovery tool, crowd-sourcing application, and exhibit builder/publisher. This topic will explore yet another remarkable capability of Omeka that employs georeferenced historic maps and temporal references.

Open Source Ethos and Resiliency

One reason for the success of Omeka is its open-source license and its community-oriented development process. The open-source ethos is a major factor in the success of the World Wide Web. Open source software fulfils the same role as open standards in terms of encouraging people to create and share content without fear that their investments will be made obsolete by the competitive software industry.

The open source movement has been successful because the price (free) makes the tools popular, and the ability to make modifications allows innovation to come from a multitude of places.

Omeka Neatline Extension

The Omeka Neatline Extension is a case in point. Neatline is a product of the [Scholars' Lab at University of Virginia](). It is a plugin for Omeka that provides tools for creating exhibits that link Omeka items and georeferenced maps with an interactive time-slider. A variety of Neatline exhibits may be explored at the [Neatline Demonstration Page]().

As with ordinary Omeka exhibits, you can include ordered lists of items with associated with narrative text. Each item can be linked to a graphical annotation on a map. Clicking the item on the map scrolls the narrative to the related item and slides the timeline to the appropriate historical period. Clicking on the item in the narrative on the time-line displays the map that is appropriate for the time-span of the item and zooms to the appropriate place on the map. Any number of historic maps may be incorporated, as well as very detailed contemporary aerial photography (thanks to the [Google Maps API]().)

The Simile Time Line

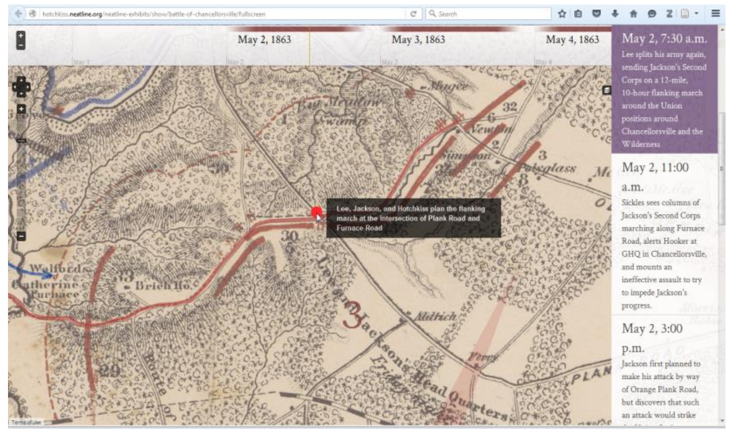


Figure 13: An exhibit from the Neatline demonstrations page.

The Time-Slider function in Neatline is the product of another open source project, [Simile](). The Simile time line is a program module that makes it easy to visualize temporally referenced data. Recalling our discussion of the Semantic Web in an earlier topic, independent particles of information assemble themselves into meaningful order when placed into context in logical referencing systems.

The Simile widget provides web developers with a one-dimensional referencing tool that is ideal for understanding the ordering of events in time. While many of the logical referencing systems being leveraged in the world of linked data, time is the one that requires the least amount of explaining. And yet, practically every information artifact ever cataloged is tagged with at least one temporal reference.

Digital Humanities Showcase

The Neatline plugin for Omeka is another brilliant example of an authoring tool for rich digital humanities projects. It shows potential for semantic web/linked data projects, although, other than the maps, all of the data are local within the Omeka repository.

Neatline Dependence on WMS

One drawback of the Omeka Neatline extension is its dependence on Web Map Services – which the Omeka toolkit does not provide. Hosting a WMS is not nearly as simple as hosting an Omeka archive on your own server or with a hosting service. In the previous topic, we demonstrated how one may georeferenced maps and host them as WMS on sites like [Harvard World Map Warper]() or the [David Rumsey Map Collection](). But depending on free services like these for critical aspects of an exhibit is not a safe strategy for anything but a demonstration.

Open Source and Open Standards to the Rescue?

There is a hope that Omeka and the Neatline extension may support the simpler [Web Map Tile Service]() in the near future. In fact, if a person really wanted this to happen, it would be possible to check out the source code for Omeka and create a plugin for incorporating the open source [Web Map Tile Server script by Klokan Technology](). This map server utilizes the ubiquitous PHP capability built into most hosted web servers. Then Omeka could host georeferenced maps and the Omeka Neatline could be altered to use these maps as overlays.

All of this is feasible, thanks to the all of the sharing of open-source tools made possible by web standards.

Earth Browsers

The topics in this handbook have discussed various means of taking care of information about places so that independent documents can be discovered by search queries, or interactive discovery interfaces like HistoryPin or exhibits that put references to archival material into narrative order.

New understanding emerges when we put independent resources into logical order, or when resources order themselves because of the tags that have been assigned to them. HistoryPin and Omeka Neatline demonstrate how the logical referencing systems of time and space serve a practical ordering function.

Earth Browsers exploit yet another dimension of contextualization by framing maps, perspective images and three-dimensional models in their three dimensional fullness from any position on the ground or in the air.

Google Earth as a Browser for the Geo Web

Google Earth is the most well-known immersive 3D web browser. To begin an exploration of Google Earth, you might enter a location in the search window. Try "*Colosseum, Rome*" You may expand the Layers panel at the bottom left to observe a deep collection of three-dimensionally referenced content that Google has curated to demonstrate the geographic web. This collection includes thousands of photographs of the area contributed by tourists, Google's own street view photography, high-resolution panoramas, a world-wide terrain model and three dimensional models of all of the buildings in Rome and other major cities around the world..

It may be a challenge at first to navigate in the three dimensional world of Google Earth. The control at the upper right corner of the window or your mouse's scroll wheel may be used to view content form any angle. Once you master this, you can click the Clock button on the tool bar to expose the time slider which will filter content based on temporal tags. At first this feature allows you to choose among epochs in Google's collection of historic aerial photos and satellite imagery.

Explore GeoWeb Content

Tilt your view into perspective and get near the ground and you will see that there is also a world-wide terrain model. Turn on 3D Buildings in the Layers panel and explore.

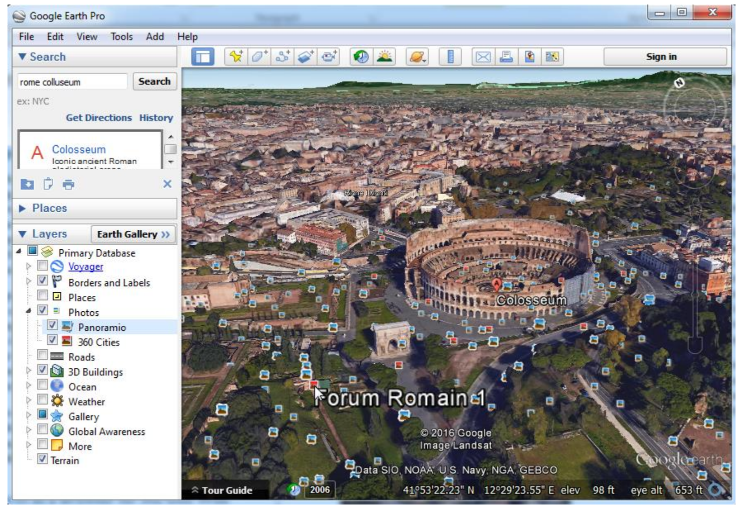


Figure 14: Google Earth arranges web resources in four-dimensional context

Google has included a few of the map services from the David Rumsey Map Collection, which you can find under the Gallery sub-section of the Layers Panel. You may need to zoom back to a high perspective on greater Rome to see the spots indicating Rumsey maps. When you find one, click it to see some metadata and to drape the map service onto the terrain model..

The Layers panel has hundreds of geographically located and 3D-enabled documents that you can explore. For a more comprehensive demonstration of Google Earth and its curated web content visit the Google Earth Showcase.

In a future tutorial we will explore the three dimensional extensions to web standards that allow institutions and individuals to reference information so that it may be discovered and explored in earth browsers.

A Unique Proprietary Browser

One major problem with Google Earth is that it is a unique and proprietary tool. When you consider this in terms of the factors that have made the web successful, you might understand why people may be reluctant to invest lots of money developing content that can only be appreciated by means of a browser that may be discontinued by Google at any moment. Until recently, Google supported a plugin that created a 3D Google Earth window in any ordinary web browser. But this was officially de-supported in December of 2015, an event which caused great distress to many members of the geo humanities community.

Cesium: An Open Source Earth for Standard Browsers

An open source earth browsing tool named Cesium now has promise to make Earth-Browsing a free commodity in any browser. But it is not yet nearly as functional or content rich as Google Earth. This is truly a critical frontier for Geography and for understanding cultural heritage in context.

Google Earth as an Authoring Tool

In an earlier topic we explored Earth Browsers that help us to organize and visualize three-dimensional relationships among geographically referenced resources. Google Earth is one such Earth Browser, and another is Cesium – which is an open source earth browser that will run in any ordinary web browser.

The capability that these browsers have to put information together in three dimensions is a striking demonstration of how new information emerges when diverse historical documents like maps and photographs are put together in the logical dimensions of space and time.

In this topic we explore how Google Earth can be used as an authoring tool that provides us with the following capabilities that you can explore through Google's On-Line Help:

- Georeference maps and plans
- Create simple three dimensional models
- Use perspective photographs to understand spatial relationships
- Integrate detailed 3D models
- Link model resources to hypertext and the greater web.
- Choreograph camera movements (tours) and save them to video.
- Save all of the above in an open, community governed encoding format, KML.

These capabilities enable authors to add the necessary four-dimensional referencing tags to their own photos and maps. Integrating fragmentary two-dimensional evidence to develop, present and share a coherent three dimensional understanding of a historic site.

Emergent Information

The most interesting effect that is illustrated by this multidimensional synthesis is that when you place the fragments into relation with each other, and are able to move around to view the assembled pieces from different points of view, new information emerges about the situation. The new emergent information is not a part of any of the pieces. The value of the composition far exceeds the sum of the pieces.

Demonstration: Google Earth Authoring

Download the example Scollay Square KMZ file and open it in Google Earth. Explore how this composition of historic maps, photographs and 3D models brings to light the relationship of this historic situation with the contemporary one in Government Center.

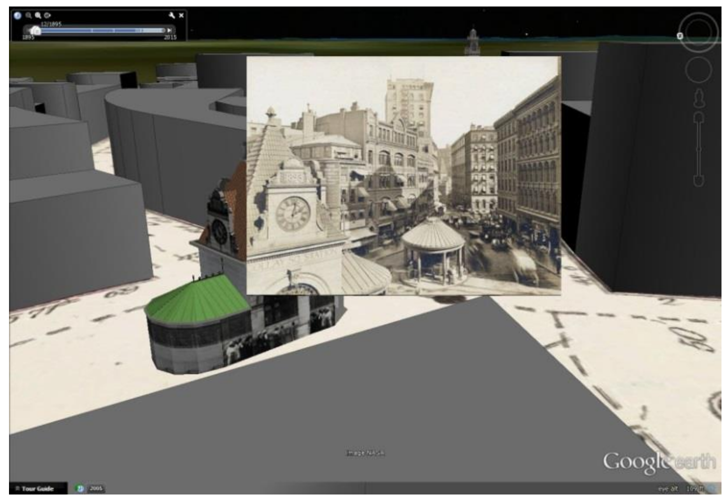


Figure 15: Google Earth georeferences images and models in four dimensions

Our Google Earth project for Scollay Square begins with a view of Boston's Government Center plaza as it exists today, illustrated by Google Earth's web based 3D model. Turning off the modern 3D Buildings layer, we can turn on our scanned 1905 map (obtained from Open GeoPortal). This map records the footprints and the heights of the buildings. The Sanborn map indicates the presence, in 1905, of photography studios on the west side of Tremont Row.

We have used the polygon drawing tools to trance the buildings from the map, and add heights to them to extrude a rough 3d model of 1905 building massing. Slide the time-slider at the top of the window to see several epochs of buildings.

We have a photograph of Scollay Square taken from an up-stairs window, we know not where exactly. The photo includes a portion of the old Scollay Square subway kiosk, whose footprint is not included on our map.

We now use our 1905 building models created from the map, to place the photo into three dimensional perspective. The photo, now oriented logically with the rest of the model, seems to have been taken from the second story window of the photography studio. Placing the photo into perspective shows us exactly where the subway kiosk was situated. We can use this information to place a three dimensional model of the kiosk into the correct location in Google Earth. The construction of the kiosk model from old photos will be discussed in the next topic.

KML: Hypertext with Space and Time

The most lasting and important aspect of Google Earth is the encoding format, KML -- an open standard governed by the Open Geospatial Consortium. KML is very much like HTML as a means of putting images and text and other media together in meaningful ways. The beauty of KML is that in addition to using URLs to reference information it adds a powerful set of tags for referencing objects and viewpoints within the logical domains of space and time!

Preservation and Exchange of 3D Models

Our exploration of Google Earth as an authoring tool for the geo-web demonstrated a few ways to use historic images to reconstruct and to develop new information about the cultural context of situations. Google Earth is 3D modeling tool of very broad geographic scope. But Google Earth is not a very good tool for creating detailed three dimensional models of buildings or monuments.

A Community of Collaborating Tools

One can't expect one tool to be able to do everything. The job of putting diverse information together will always require a variety of tools that fill specialized niches for encoding specific types of understanding as data.

Standard exchange formats permit information encoded by specialized tools to be handed off to other tools. As we have seen in other domains of internet-based information sharing, three-dimensional modeling is another domain that proves the importance of stable open standards for encoding and exchange.

COLLADA: Encoding for 3D Models

Like HTML has created a web of hypertext, COLLADA performs a similar role for encoding and exchanging detailed three-dimensional models. The story of COLLADA begins in the game development studios of Sony Corporation. Designers at Sony, who produced the landscapes, buildings, and characters for video games, depend on variety of tools for modeling pieces of these environments. Many of these tools are licensed by competing software companies. Inability to exchange information among the tools was a problem that cost valuable time and money.

Sony decided to build a common language that all of the different tools export and import 3D models with realistic textures. This is how the COLLADA (Collaborative Design Activity) format for 3D model exchange was born.

To promote COLLADA as a safe and stable means of preserving and sharing information about 3D models, Sony handed the COLLADA standard over to the Khronos Group, a non-profit consortium of stake-holders with interests in 3D modeling.

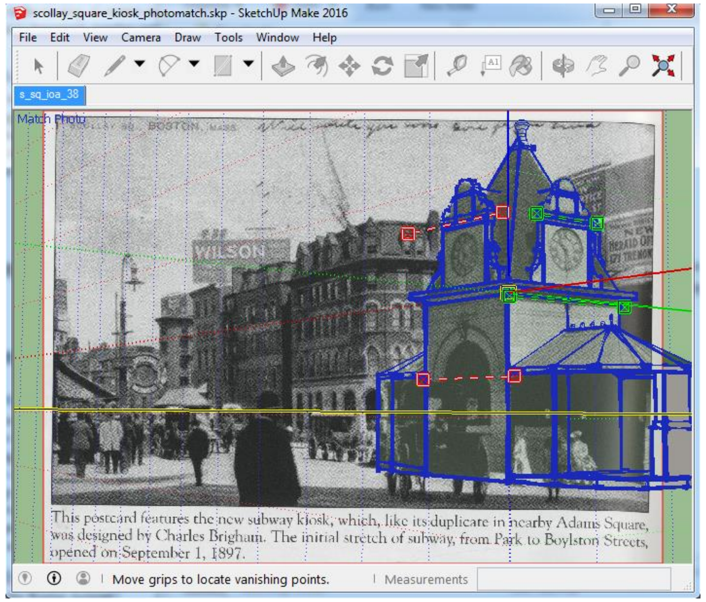


Figure 16: Sketchup uses historic photos to create a 3-D model

SketchUp

Most 3d modeling tools now export and import models in the COLLADA format. SketchUp is the most interesting of these from the perspective of modeling buildings in geographic context. SketchUp is free to install and use for educational or personal use.

Georeferencing in SketchUp

One of the amazing things about SketchUp is the way that it facilitates the georeferencing of models. A model can be started with the Geo-Location feature, which invites the user to drag a box on a web map, to obtain a detailed aerial photograph and terrain model to use as a base for locating your model. Your model is now georeferenced. As a result, the Google Street View API may be used to drape Street View photography onto vertical walls of buildings. When the model is finished, it can be exported as COLLADA or as a zipped KML (KMZ) file that will locate itself properly in Google Earth.

Photogrammetry with SketchUp

The Photo Match feature is one of the most interesting aspects of SketchUp from a historical preservation point of view. The image in the center of this page shows how an historical photograph opened as a Photo Match layer can be used as source material for three dimensional models.

The Photo Match tool de-codes the perspective in the photo so that all of the edges can be used as guides for blocking out and scaling the building massing. The resulting model has the photo mapped to its faces. These image textures aid in the articulation of doors, windows and roof details. Download the Scollay Square Kiosk Photomatch example .

The Sketchup 3D Warehouse

The SketchUp 3D Warehouse is a cloud-based archive of 3D models. This is yet another way that the creators and patrons of Sketchp have found to facilitate the modeling of places. There are so many models that can be found here including most of the famous landmark buildings in the world. It is not uncommon to start a model of a site, and check in the 3D Warehouse and find that some of the key buildings that you need have already been modeled and shared by others.

Field Work

No handbook on information stewardship for historical preservation would be complete without a discussion of collecting data in the field and strategies for managing and archiving these data. One never knows if today's photographs will be priceless historic documents 100 years hence!

Planning

This topic considers a special case of field work that involves a team of collaborators in the field to collect photographs over a broad complex of sites. In such a scenario it is useful to plan ahead to make sure certain information is collected – especially if it will be difficult to return to the area for a repeat visit.

Planning with Historic Maps

Beginning by organizing contemporary and historic maps and perspective drawings and photography in Google Earth may highlight particular places and camera angles that ought to be covered to understand the contemporary context of the historic situations. Traces of historic sites are sometimes concealed by contemporary buildings and vegetation. Laying historic maps and plans on top of contemporary base maps reveals key corners that ought to be looked for and possibly excavated during a visit.

Google Earth in the Field

Google Earth may be brought into the field on a GPS-enabled tablet. No internet connection is required. With maps and models prepared in advance, Google Earth can be very helpful means to orient yourself with respect to the historic situation that is being studied while you are in the field.

Currently, the Google Earth Application for Android does not support georeferenced images or models. Bringing these functions into the field will require a Windows computer or tablet or an Apple Laptop.

Android Tablets

Android tablets are an inexpensive multipurpose device for bringing georefrerenced information into the field and for collecting georeferenced photography and video. There are a few Android applications that do support overlay of georeferenced maps with GPS location in the field, including [Maverick Maps] and [QGIS for Android].

Figure 17: Google Earth may be brought into the field on a mobile tablet.

Geo-Tagging Photos

Sending a group of people into the field with cameras can yield such a large number of photographs that it can be difficult to sort through them all. Automatically tagging the photos with their location can make it much easier to find photographs that are related to a specific location.

Part of the preparation for field trips should involve taking stock of the global positioning systems (GPS) capabilities of members of your party. Many cameras and practically all mobile phones have the capability to embed GPS coordinates into photographs. GPS apps function even where no wireless service is available. It is a very good idea to have the members of your party test this capability before the field trip – since troubleshooting on site can be time consuming.

If members of your party have digital cameras that are not GPS enabled, it is possible to automatically tag their photos if the camera's clock is synchronized with a GPS on their mobile phone. A free GIS App for i-phone or Android can be set to log location every 30 seconds. Photos can be georeferenced later by syncing their time-stamp with the nearest pint in the GPS log using a tool called Geosetter as discussed in the next topic.

Managing GeoTagged Photos

It is often said that our age is marked by information anxiety. We have too much information; too much of it is redundant; and we have too few means of sorting it out. In his excellent book on information architecture, Ambient Finadability, Peter Morvile, explains how new tools are helping to bring information into order. No situation illustrates this more poignantly that that of the management of field-collected photographs.

When you send a studio course of fourteen students into the field to make observations you can end up with a serious clog of information. Without a doubt there is much treasure to be found among photographs collected by architecture students focused on local character and change. But there will be thousands of often redundant photos ordered in their original data folders according to camera and date. It is practically hopeless almost to try to find anything specific.

With proper planning and preparation of equipment (discussed in the previous topic) terabytes of photos collected by teams of people can be turned in to well-ordered catalogs that are searchable according to subject or searched according to map location. This sort of organization requires a bit of preparation before the trip and a coordinated tagging session afterwards. It helps to have at least one member of the party to help to coordinate these activities.

Geosetter

Geosetter is a free tool created by Friedrich Schmidt may be one of the most interesting illustrations of the convergence of georeferencing, the semantic web and linked data, and mobile applications.

Demonstration: The Mumbai Studio Case

Fourteen students from the Harvard Graduate School of Design arrived at Mumbai for a week of field research. Before the trip the students attended a one hour meeting to outline the strategy for collecting georeferenced photographs. On-site, the students divided into four parties to photograph sites selected by their Studio critic. Some of the students were using GPS enabled cameras; others were using digital cameras without GPS. These photographers set their smart phones to log their GPS location every 30 seconds.

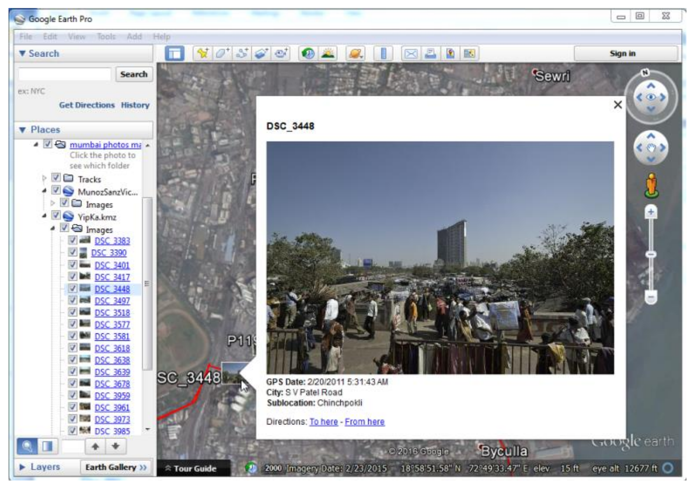


Figure 18: Geosetter creates a catalog of field images that may be browsed n Google Earth

Returning from the studio, the students placed their photographs in a shared directory and used Geosetter to tag and map the images. The first stage or processing with Geosetter is automatic. Photographs that were not georeferenced by GPS-enabled cameras are tagged with their location by correlating their time stamps in the photo with the GPS log from the photographer's smart phone. The result of the georeferencing phase is that Geosetter creates a KMZ file that maps each of the images in Google Earth.

The second stage of automated processing, Geosetter uses the coordinates embedded in the photo to lookup geographically linked data from the GeoNames to associate each photo with corresponding geographic names: Country, Province, City and Neighborhood where the photo was taken.

Now that each folder of images has a map index, it is possible for all of the members of the studio to find photographs associated with a place. Then the participants can tag the photographs based on subject themes that may be useful for sorting and discovery of images later.

Embedded Metadata

Al of the metadata that is developed with Geosetter is embedded directly into the image files using the Extensible Metadata Platfotm (XMP). Embedding metadata this way assures that no matter where the file goes, its metadata will always go with it.

Cloud based aggregators of photographs, such as Flickr or Google's Photo Organizer will automatically create catalog records from these embedded metadata. Subsequently, when the photos are downloaded and used by others, they will always remember their provenance.

Alternative Methods of Collaborative Tagging

Geosetter is a lovely tool for demonstrating the power of the semantic web. Similar collaborative photo-tagging projects can also be carried out using Flickr or Google Photo. There can be advantages to managing these projects on the cloud, although one always wants to keep their own archive of the enriched photographs.

Conclusion: Considering the Future

The world of information stewardship is marked by very active development. And yet the essential responsibilities and strategies are well defined. Enterprises geared toward preservation of cultural heritage have a responsibility to preserve information and to make it accessible to communities of scholars and the public and ultimately to our successors in future generations.

Cultural heritage is information in context. Our exploration of the various forms of referencing that are emerging in library catalogs and web applications like DPLA, OpenGeoPortal, HistoryPin, Google Cultural Institute, Omeka, and Google Earth all point to an expanding and compelling capability to place information resources into logical order, spatially and temporally and in terms of many thematic ontologies. The understanding of relationships that emerges when documentation is understood in context is often much more that the information content of the individual documents.

Information Stewardship in Cultural Heritage Preservation involves not only preserving documentation pertaining to historic sites or artifacts, but also understanding how the documents may be discovered in catalogs and used in applications that place particular pieces of information into cultural context.

Toward Self-Organizing Resources

Each of the topics in this handbook explores an aspect of the broader trend to more tightly link standardized metadata to information resources through their lifecycle. This common trend is leading to information resources that order themselves. Zotero provides an example of how the act of saving references can lead to the automatic emergence of a catalog. For example digital cameras now embed metadata into photographs that can record the time and place of capture. Now that cell phones have on-board compasses and accelerometers and altimeters there is no reason that embedded metadata should not have all of the tags necessary to place that photograph into proper four dimensional orientation within an earth browser like Google Earth and to orient itself appropriately with other perspective images through protocols similar to the Google StreetView API.

This trend toward self-organizing resources should become built into our authoring and composition tools as the employment of embedded metadata encodings like XMP becomes an expectation for authoring and composition tool.

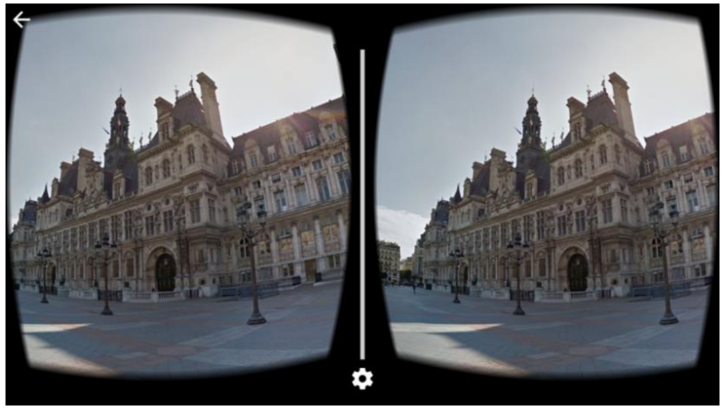


Figure 19: A 3D Immersive scene viewed on a mobile phone using Google Cardboard

Augmented Reality

At the time of this writing, there is a lot of interest in immersive augmented reality applications, like [Oculus Rift](), and [Google Cardboard.]() The demonstrations of these tools mainly feature fictitious environments. Google has taken advantage of its giant collection of georeferenced perspective images collected by its Street View platform to create an world-wide immersive 3D model. This model, along with the 3d models in Google Earth provide a framework for creating mobile tools for using historic documentation to understand places as they once were.

Hopefully the understanding of standards, tools and strategies for cultivating the connectedness of information covered in this handbook provide an understanding of how thoughtful stewardship of information will foster this capability to overlay our historical understanding of places onto the world itself and multiply the value of our efforts to preserve cultural heritage.

Collaboration of specialized communities through open standards and Open Source Software

Another recurring theme that emerges from this survey of information stewardship technology and trends is that new ways of collaborating are allowing specialists focused in particular areas of heritage conservation to focus on their work and make their knowledge part of a world-wide network of resources including traditional forms of documentation, but also by systematically encoding their knowledge about specific cultural artifacts or sites, buildrrs, artisans and schools and how they are related as linked data.

This ability to specialize and collaborate also involves projects to develop and extend open-source tools for scholarship and community-governed open standards for metadata content and encoding.

To users of the web, this rich ecosystem of inter-related knowledge, tools and standards may be nearly invisible while people will be using it and expecting things to work. It is hoped that the topics of this handbook have illuminated some of what is necessary behind the scenes to being these opportunities to fruition.

Appendix: Omeka Hosting

Omeka Strategies

Omeka is a very nice tool for teaching since it demonstrates a rich set of functions for learning about metadata, archive management and digital humanities. Omeka is a useful tool for a small project with a dedicated researcher who is comfortable with web technology. But Omeka may not be the right choice as a complete archive repository solution for an institution. Institutional strategies for information stewardship are discussed in depth in one of the earlier topics in this handbook.

Omeka Neatline and Web Map Services

Our pilot course at the Boston Architectural College (BAC) made use of the Neatline extension which provides a capability for creating exhibits that link documents and narrative text with timelines and georeferenced maps. The georeferenced maps used in Neatline exhibits must be served by a Web Map Server (WMS). Omeka itself does not include a WMS capability. In our pilot we used Harvard World Map to georeferenced maps and to host them as WMS.

After some time fiddling with Neatline, and experiencing problems with the server capability, my conclusion is that it will be easiest to demonstrate Omeka Neatline exhibits from the Omeka Neatline demonstration site, and use Google Earth for the workshop on creating geo-temporal exhibits..

Omeka Plugins

The basic Omeka installation provides the tools for archiving documents of many types, creating metadata for items, searching for items in the archive and creating exhibits linking images and other documents with narrative text. Omeka uses plugins to extend the basic functionality. The Geolocation plugin is an important one which allows resources to be tagged with latitude and longitude coordinates and discovered with map searches.

Hosting Through Omeka.Net

Omeka has its own hosting service, Omeka.net. Unfortunately none of their plans includes the Neatline Extension. But for $99 per year, you could have an Omeka installation with geolocation.

Self-Hosting on Amazon Cloud

For our pilot project at the BAC, we made use of a self-hosted Omeka installation that is installed on a server on the Amazon Cloud. We use a low-volume server (aka Small Instance) which costs about $40 per month. This hosting method is cheap but requires a person who knows a bit about setting up a linux server.