Digitization Does Not Equal Access: Challenges in Creating and Disseminating Cultural Information in the Age of the Internet

Murtha Baca
Head, Digital Art History Program, Getty Research Institute (GRI)

CIDOC Conference, Dresden, September 8, 2014

Key Concepts

- OPEN ACCESS, OPEN CONTENT
- Democratization of information and knowledge
- Cross-cultural, multicultural dialogue
- Changing behaviors and "venues" (mobile devices, social media)
- Who are our users?

Key Concepts
continued

- ❑ Documents vs. data
- ❑ Structured, machine-processable data
- ❑ The Semantic Web, & Linked Open Data (LOD)
- ❑ The importance of language

What is our mission as cultural institutions?

- ❑ to collect and preserve objects, artifacts, the built environment, etc.
- ❑ to foster education and research
- ❑ to "build bridges" between people and collections, and between different cultures
- ❑ to make our collections, our data, and our research widely available—both on site and on line

Digitization does not equal access!

- Digitized collections without adequate documentation are in essence un-manageable, un-preservable, difficult to transfer, share, and transmit, and often inaccessible.
- Even well-documented collections can remain inaccessible due to technical and linguistic barriers.

Barriers to Access

- diminishing resources (human, technical, financial)
- institutional inertia and inefficiencies
- lack of descriptive metadata/cataloguing of collections
- language!
- the "Deep Web"
- the "Google/Wikipedia factor"

Ways to Enhance Access

- ❑ by documenting our collections!
- ❑ by using appropriate data standards and controlled vocabularies
- ❑ by employing new strategies for metadata creation, shared workflow, etc. (also, sharing data and tools among institutions)
- ❑ by generating and analyzing user metrics

Ways to Enhance Access (2)

- ❑ by contributing to union catalogues and other "federated" resources (e.g. Europeana, TELDAP union catalogue, SARI, ARTstor, OCLC WorldCat and Art Discovery Catalogue, etc.)
- ❑ by making collections metadata and related digital assets "harvestable" via protocols such as OAI/PMH, and/or available as Linked Open Data (LOD)

Ways to Enhance Access (3)

- by building and implementing multilingual tools
- by implementing strategies for greater Web accessibility:
  - Web page metadata
  - optimizing for Google and other commercial search engines
- by exploiting popular resources like Wikipedia, Facebook, Twitter, YouTube, etc.
- All of this is made possible by good, standards-based documentation!

Tools for Documentation Professionals: Conceptual Models

CIDOC Conceptual Reference Model

❑ The CIDOC CRM is a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

❑ It is also a tool to promote a shared understanding of cultural heritage information.

http://cidoc.ics.forth.gr/index.html

CDWA/CCO Entity-Relationship Model

Entity Relationship Diagram for CCO
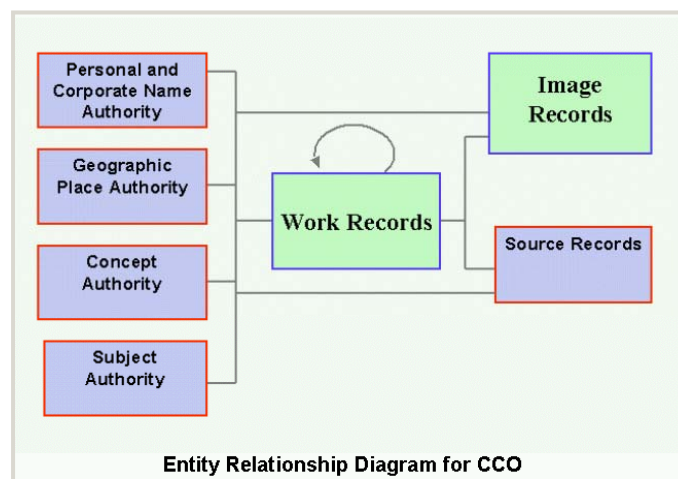
http://www.getty.edu/research/publications/electronic_publications/cdwa/entity.html

Tools for Documentation Professionals: data structures/metadata schemas

CIDOC
COMITÉ INTERNATIONAL
POUR LA DOCUMENTATION

"LIDO" (Lightweight Information Describing Objects)

LIDO is a simple XML schema — a metadata "container" for expressing, sharing, and harvesting core museum information.

http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/

LIDO
Lightweight Information Describing Objects

Tools for documentation professionals: data value standards (=vocabularies)

Example of a "cross-cultural" controlled vocabulary: the *Art & Architecture Thesaurus* (AAT)

The *Art & Architecture Thesaurus* (AAT)®

THE AAT IS A STRUCTURED VOCABULARY OF MORE THAN 36,000 CONCEPTS, INCLUDING MORE THAN 245,000 TERMS WITH DESCRIPTIONS, BIBLIOGRAPHIC CITATIONS, AND OTHER INFORMATION RELATING TO FINE ART, ARCHITECTURE, DECORATIVE ARTS, ARCHIVAL MATERIALS, AND MATERIAL CULTURE. ITS DATA MODEL IS DESIGNED FOR MULTILINGUALITY.

http://www.getty.edu/research/conducting_research/vocabularies/aat/

---

still lifes
still life
still-lifes
still lives
nature morte
natura morta
stilleven
Stilleben
naturaleza muerta
nature reposée *(early 18th-century French term)*

AAT "still life" record on the Web

Many AAT records include terms in other languages, as well as historical variants

AAT record on the Web for "watermarks," with multilingual & historical variants

Art & Architecture Thesaurus: Multilingual Initiatives

- ❑ Spanish: Centro de Documentación de Bienes Patrimoniales (CDBP) Santiago de Chile)
- ❑ Dutch: Netherlands Institute for Art History (RKD)
- ❑ Chinese: Taiwan e-Learning and Digital Archives Program (TELDAP, Academia Sinica)
- ❑ German: State Museums of Berlin/Institute for Museum Research)

A multilingual vocabulary for subject access to visual material: the ICONCLASS system

http://www.iconclass.nl/
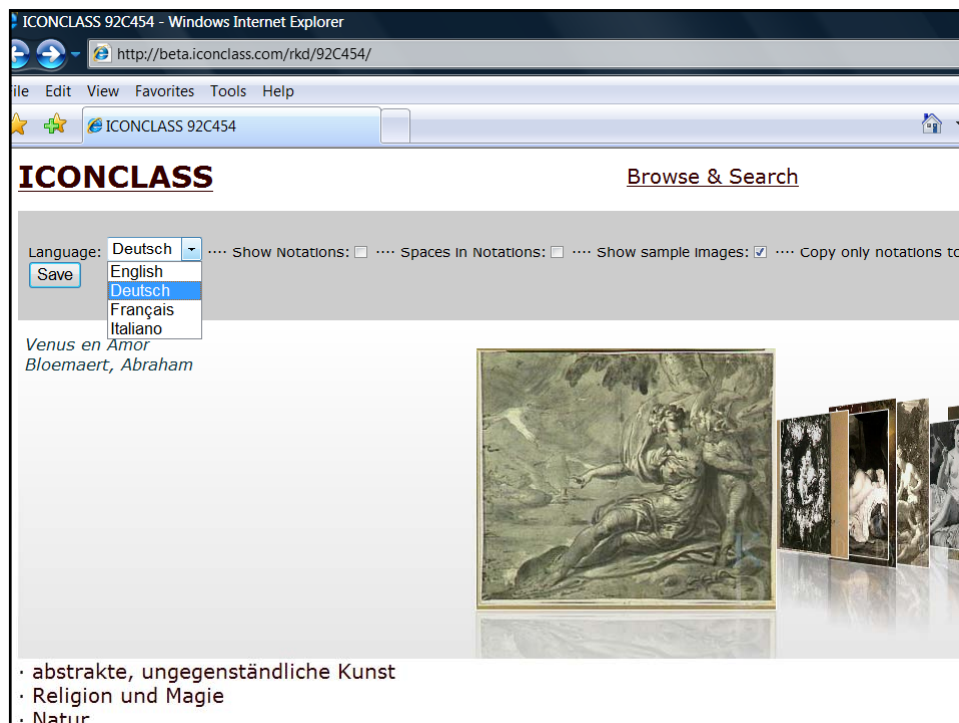
Example of a "regional" controlled vocabulary:
the *Tesauro Regional Patrimonial* (TRP)

www.tesauroregional.cl/

The TRP contains terminology relating to pre-Columbian cultures primarily in the central Andes and Mesoamerica. Terms are in Spanish and regional vernacular languages.

ISSUES in Exposing Cultural Collections to the Web

Web Search Engines

❑ What do search engines do? — They "index" the Web.

❑ Web search engines index HTML Web pages (static and dynamic) versus content from proprietary databases, which is mostly impenetrable by commercial search engines like Google.

"Visible Web" versus "Deep Web"

❏ The Visible Web is what you see in the results pages from general Web search engines & subject directories (static HTML pages) — it is what is "penetrable" by Google and other search engines.

❏ The Invisible or Deep Web consists of data from dynamically searchable databases that cannot be indexed by search engines, because the Web pages resulting from searches are not "stored" anywhere.

Questions to ask ...

▪ Are your collections "reachable" by commercial search engines?

▪ If yes, how will you "contextualize" individual collection objects?

▪ If not, what is your strategy to lead Web users to your search page?

The "Google Factor"

Facing the facts

□ Google is the starting point for most users' (both experts and non-experts) searches.

□ "If you can't be found from Google, you don't exist."

□ Is English is the *lingua franca* of the Web?

What Google "looks at" regularly

- ✓ Title tag

- ✓ Text on the Web page

- ✓ Referring links

What Google "looks at" only occasionally

- ✓ Keywords metatag

- ✓ Description metatag

What Google "ignores"

✓ Graphics, flash animation
— anything it cannot index!

A Collections Access "Checklist": Fundamentals

❑ Create standards-based <u>core</u> documentation for collection objects.

❑ Broaden user group by creating multilingual access points.

❑ Expose collection information to search engines, and/or make it easy for users to find your collections search page.

Collections Access "Checklist" *continued*: exploring new avenues

- ❑ Make collection metadata & images "harvestable" and/or available as LOD.
- ❑ Create/and or enhance carefully selected Wikipedia entries, with links to your site.
- ❑ Explore the use of social networking tools like Flickr, YouTube, and Twitter.
- ❑ Only do this after you've created standards-based core documentation for your collections!

CIDOC
COMITÉ INTERNATIONAL
POUR LA DOCUMENTATION

Some Recent Developments

- ❑ Online Scholarly Museum Catalogues
- ❑ Collaborative Digital Workspaces for Conducting and Publishing Research
- ❑ Linked Open Data (LOD), the "Semantic Web," and "Virtual Authority Files"

Online Scholarly Catalogue Initiative (OSCI)

Sponsored by the Getty Foundation

OSCI Partners

Art Institute of Chicago
Freer | Sackler Gallery of the Smithsonian
Los Angeles County Museum of Art
National Gallery of Art, Washington D.C.
San Francisco Museum of Modern Art
Seattle Art Museum
Tate
Walker Art Center

Why do online museum catalogues?

- **Interactive & easily updated**
- **More comparative images**
- **Enhanced image viewing: zoom technologies, side-by-side comparison, etc.**
- **Multimedia capability (audio, video)**
- **Archival documentation**
- **Conservation documentation and analysis**
- **Flexible searching**
- **Users can make their own collections**
- **Annotations and comment functions**
- **Citation tools**
- **Exportable to a variety of devices and formats**
- **Ability to reach expanded audiences**

Three Technical Solutions

1. OSCI Toolkit
    Micro website that has the appearance of an ebook
    (Art Institute of Chicago)

2. Web content management system
    integrates the catalogue with existing & enhanced web content
    (San Francisco Museum of Modern Art)

3. Collections management system (TMS/eMuseum)
    catalogue is assembled from enhanced online collection pages
    (Seattle Art Museum)

OSCI Catalogues

**To browse the published catalogues
and
learn more about the initiative
go to:**

https://www.getty.edu/foundation/initiatives/current/osci/

Developing tools, methods, and standards for teaching, conducting, and publishing humanities research in digital form:

GETTY SCHOLARS' WORKSPACE
™

WHAT IS
THE GETTY SCHOLARS' WORKSPACE?

An online collaborative venue with a toolset specifically designed for teaching, conducting and publishing art-historical research -- a "back end" production environment, *not* a website.

WHY WORK DIGITALLY?

to enable access and information sharing versus hoarding—part of the Open Content philosophy

to capture scholarly dialogue, multiple perspectives

How might the Scholars' Workspace "change the game"?

By causing a significant *paradigm shift*

- ***Break* with the single authorial model**

  Capture conversations, debates, different interpretations and perspectives

  Enable production of new knowledge, visualizations, and new types of publications

- ***Transform* the way that art history is taught (proposed Getty Digital Seminar)**

- ***Validate* digital publications—change the "reward system" of academic art history, and re-define the concept of a "publication"**

FEATURES | Image integration, cropping, comparison

Kirchner Light Table   GETTY SCHOLARS' WORKSPACE

MAKE COMPARISON

Recent Images

CROP   CROP   REMOVE   CROP   REMOVE

FEATURES | Text annotation

In Corinth's painting, however, Death does not serve as a muse. There is no relation to art, neither regarding the artist, not the skeleton, which hangs from an iron hook attached to its head. It is displayed in an almost scientific way, in an arrangement, in which it could serve for anatomical study purposes. It is an object of utility rather than the creative inspiration that Boecklin's or Thoma's skeletons are. Contrary to Boecklin's or Thoma's works, there is no movement in the painting, the skeleton and the artist seem stiff, Corinth himself

There is not enough evidence in this paragraph to support your claims

tscutt - 11/25/13 11:30 AM

Exhibition Catalogs

I disagree.

tscutt - 1/15/14 9:58 AM

GETTY SCHOLARS' WORKSPACE

FEATURES | Image annotation

is this a human figure?

AFoerschner - 10/10/13 1:55 PM

What is this?

FAlbrezzi - 11/21/13 3:29 PM

Arnold Böcklin, Self-Portrait with Death Playing the Fiddle, 1872 (crop)

GETTY SCHOLARS' WORKSPACE

FEATURES | Document comparison

GETTY SCHOLARS' WORKSPACE

FEATURES | Bibliography building and integration

GETTY SCHOLARS' WORKSPACE

FEATURES | Essay authoring, with image integration

GETTY SCHOLARS' WORKSPACE

FEATURES | Item-level Metadata, with links to authoritative external resources

GETTY SCHOLARS' WORKSPACE

FEATURES | Future developments: timeline builder

GETTY SCHOLARS' WORKSPACE

FEATURES | Future developments: plan exhibitions, re-create historic exhibitions/displays

GETTY SCHOLARS' WORKSPACE
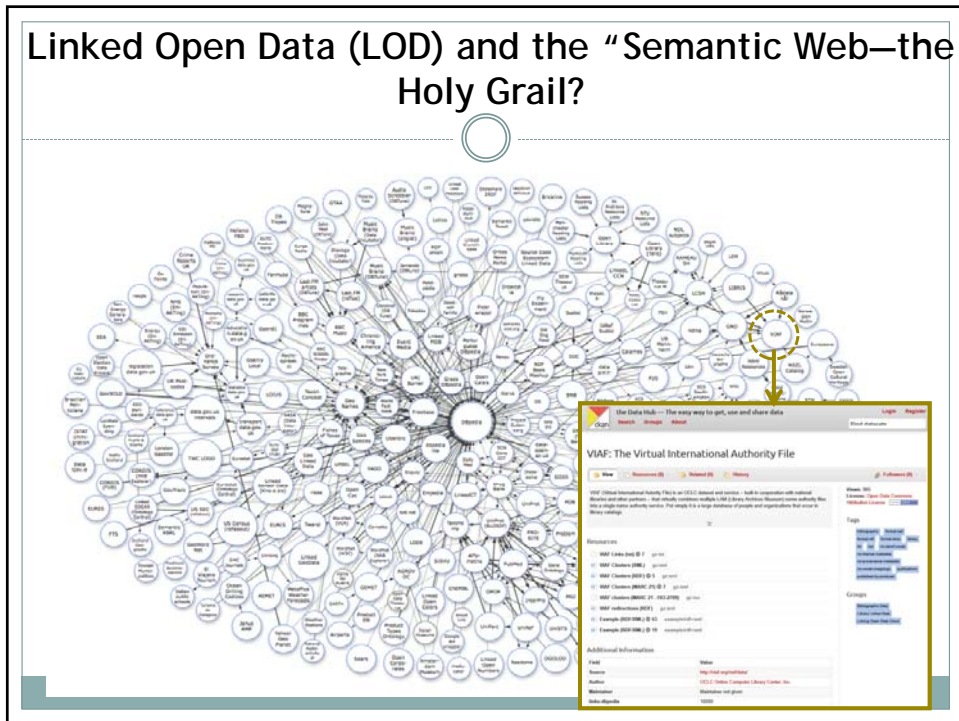
What Comes Next?

Linked Open Data (LOD) and the "Semantic Web—the Holy Grail?

Good documentation & carefully thought-out dissemination strategies can help us avoid chaos and reach more users!