

Built Works Registry Advisory Council Meeting Notes: 3 February 2012

Attendees:

BWR Advisory Council

Baca, Murtha
Bilson, Tom (by phone)
Blair, Sheila
Bloom, Jonathan M.
Chiku, Kakugyo S.
Cope, Aaron Straup
Farneth, David
Hagedorn-Saupe, Monika
Murray, Irena
Quigley, Sam
Saliga, Pauline
Simane, Jan
Sinha, Vandana
Stubbs, John
Waibel, Günter
Whiteside, Ann

BWR Project Team

Fabian, Carole Ann (Project Director)
Finer, Mary
Freedman, Thom
Kuan, Christine
Lam, Caspar
Ross, Jeff
Shulman, James (Project Director)
Wacker, Melanie
Wagner, Gretchen
Wees, Dustin
Ying, Bill

Welcome & introductions: (Carole Ann Fabian/Avery & James Shulman/ARTstor)

Introductions were followed by an update on project progress thus far.

BWR continuity & sustainability planning (Gretchen Wagner/ARTstor)

A preliminary legal agreement (Framework for Governance, Sustainability, Intellectual Property and Similar Matters) has been drafted to define and articulate the relationship between the three founding partners—ARTstor, Avery and Getty Research Institute. Major points within the agreement include the following:

- Establishes a Standing Committee composed of one representative member from each of the three Founding Partners. Each member would have one vote.
- Specifies that responsibilities of each Founding Partner during and beyond the IMLS grant project.
- Establishes that data contributions are accurate and do not infringe any third party rights.
- Addresses the issue of sustainability in the event that any one Founding Partner terminates their relationship with the BWR for any reason.
- Establishes an Institutional License Agreement to facilitate institutional contributions of content to BWR.
- Establishes a Terms of Use Agreement to facilitate individual contributions to the BWR.

Discussion:

- There was agreement as to importance of including images (thumbnails) in BWR records.
- BWR will function as a data resource; records will be contributed to both Shared Shelf and Getty's Cultural Objects Naming Authority (CONA).

- Relationship between Shared Shelf and BWR is very important. Shared Shelf will provide the ongoing technological infrastructure for BWR and both projects will share works records.

Shared Shelf platform & BWR (James Shulman/ARTstor)

A history of the development and evolution of Shared Shelf (SS) was provided. Particular note was made of a similar relationship between SS and SAHARA. SS now has the capability to integrate the Getty Vocabularies and has established a linked relationship which allows for automatic updates to records. Goal is to create a symbiotic relationship between SS and the BWR.

Discussion:

- There was some discussion about Linked Open Data (LOD) models and its relevance to the BWR. It was noted that at the time when the IMLS grant was written LOD models had not been fully developed. While LOD strategy was not included in the BWR IMLS proposal and grant, BWR is aware of its importance and will chart a linked data strategy for future implementation.
- It was noted that there is a distinction between “Linked Data” and “Linked Open Data” both of which exist in the Open Data Cloud. It was noted that in Europe 60% of the data available in the Open Data Cloud is not open—it is linked but not open.

BWR: Metadata schema overview (Dustin Wees/ARTstor)

Discussion focused on the BWR schema and how many fields should be required for a minimal or core record. The current BWR schema uses four elements--name, BWR ID, location and geocode—to establish this core record. These minimal data requirements are intended to uniquely identify a built work *enough* to differentiate one work from another.

Discussion:

- Add ‘Creator’ as required field.
- It was agreed that longitude and latitude are probably the most important identifiers of a built work—these elements most effectively and accurately distinguish one work from another. It was also noted that, although very important, latitude and longitude will not always be present in legacy data and will sometimes be challenging to determine.
- SAH Archipedia is currently working with UVA Press and Scholars Lab on a project to develop geocoding system. Suggested that the BWR might become involved with this project.
- There was some discussion about buildings whose use has changed over time. Also, questions were raised about how to include parts of buildings. Perhaps it would be helpful to view BWR records as a “slice of time.”
- The BWR should first concentrate on “simple” built works and avoid those which might be more complicated, i.e. parts of buildings or buildings which have changed use over time.
- It was suggested that it might be a good idea to begin modeling some “complex” built works.
- Many built works have already been described by cultural heritage organizations. These organizations already possess valuable data/information about built works, including location. It will be important to identify these organizations and to partner with them.

BWR: Tools (Dustin Wees & William Ying/ARTstor)

Several potential use case scenarios were discussed each with different models of data ingestion and promotion. The five seed data sets contain many different kinds of information and we now have multiple records for the same building. First step is to isolate one data set and eliminate duplications within that set, and then to iteratively 'match' and 'merge' the other seed data sets to that list.

Two scenarios for growing the BWR were discussed—one through institutional contributions and the other through individual/public contributions. In the first scenario (Phase II), institutions would be responsible for "cleaning up" their contributed records. In second scenario (Phase III), individual/public contributions would be evaluated by an editorial process. Other models of use case scenarios were also discussed.

Discussion:

- The sheer volume of ingested data into the BWR might make it difficult for an editorial board to maintain proper oversight. Machine processing might be able to do some of the initial vetting. If machine processing is not able to vet a record then it would be moved to an editorial process. Decision-support logic and editorial policies are needed for both machine processing and editorial processing to resolve such issues as: how many fields need to be matched before a record is added.
- SAHARA's editorial process was described and discussed. Contributions are first evaluated by an editorial board which then directs it to an identified expert/scholar. Final verification is provided by a Visual Resources (VR) librarian who then promotes the record to public. This model has worked well thus far and might be used for the BWR, although there is the question of volume which is relatively small when compared to the BWR.
- Questions were raised as to whether it would be more beneficial to spend effort building programs which could be used to match information or investing in an editorial board.
- It was noted that the BWR is intended to function as a registry and not an authority. As a registry, the BWR will provide the means for confirming the identity of a building. The BWR is intended to function as a data resource which will feed other data resources and/or authorities.
 - Authorities:
 - vetted
 - warranted
 - tool for term control in external systems
 - Registry:
 - data resource
 - feed to other data environments
 - normalized
 - uniquely identified
 - variable depth of records
- Tension between controlled data and web-enabled social processes
- Question of what constitutes an acceptable 'machine match' for contributed data about the same work
- OpenStreetMap was suggested as a possible source for geocoding data.
- Images were suggested as a means of identifying buildings and possibly linking records.
- Suggestion to include addition typology of works; clarified that BWR will include actual built structures – not designs for un-built works.

BWR: Data strategies (Dustin Wees/ARTstor & Murtha Baca/GRI)

Examples of Getty's CONA records were presented to illustrate variations which can be found in both name and in work type. A brief history of the development and evolution of AAT was provided. AAT now includes 150,000 concepts or 200,000 terms and was built primarily from contributions—very few records were actually created. The bulk of the labor was focused on clustering, enhancement and building of hierarchy/records.

An overview of various disambiguation options was presented:

- Machine matching (unambiguous = pre-matched)
- Machine matching (suggested matches; records need to be evaluated)
- Manual/human (reports needed; must have ability to edit if match is not correct)

Discussion:

- There was some discussion as to whether the "Creator" field in BWR record should be treated as a text or controlled field in BWR. Many BWR records will be associated with a culture instead of a creator.

BWR: Policies for collaboration

- **VIAF & the International Terminology Working Group (Murtha Baca/GRI)**
- **SAHARA editorial model (Pauline Saliga/SAH)**

An overview of the International Terminology Working Group (ITWG) was provided. As noted Getty Vocabularies are becoming increasingly multilingual. Significant efforts are underway to incorporate international databases. Institutions need to understand that it is to their advantage to contribute records—it will not only benefit respective institutions but the general public.

Identification of contributors to the BWR is important. There was general consensus that all contributions to BWR should be credited.

An overview of the Virtual International Authority File (VIAF) was also presented. VIAF was initiated by Barbara Tillet at the Library of Congress and functions as a union catalogue of authority files. The technical infrastructure was provided by OCLC. All of the matching/clustering was done by machine which was somewhat messy, but overall the results were fairly good. Currently includes more than 20 million names, linked data via URIs, Unicode throughout.

An overview of various collaborative projects of the Society of Architectural Historians were presented and discussed:

- JSAH online: provides a publishing platform for scholarly journals which includes dynamic media. This was a University of California Press effort, scaled up to the JSTOR Current Scholarship program.
- SAHARA: SAH online image archive developed in collaboration with ARTstor. SAHARA content (data and image) are contributed by SAH members, vetted through an editorial process by scholar/librarian teams. Images which had been approved are noted as "Editor's Choice" and promoted to ARTstor Digital Library. All SAHARA contributions are available to SAH members. The SAHARA project will migrate to the Shared Shelf platform.
- Archipedia: National Endowment for the Humanities (NEH)-grant funding development of the project over the next two years. Initial focus is on American content from the SAH Buildings of the United States (BUS) series NEH & University of Virginia Press-effort will

map buildings and include references. Intent is to develop this into a global resource and to integrate databases.

SAHARA's editorial workflow was outlined. An individual can upload an image along with metadata. Images are sent into "editorial buckets" whose contents are reviewed by a Visual Reference (VR) librarian who verifies all information.

Discussion:

- Individuals have various incentives for making contribution to the SAHARA's collections. Some feel strongly about a particular type of architecture or may simply want their visual collections to be digitized.
- Statistics indicate that the SAHARA collections are being used. Would like to encourage more individuals to make contributions to collection and need is to find new models for making contributions.
- RIBA offers to both digitize and to create metadata for architects who contribute their collections. Willingness to deal with metadata has significantly increased contributions.
- Suggested that librarians might provide a good source for future editorial work for the BWR.
- It will be important to develop contribution policies and protocols for both individuals and institutions.

Closing discussion (CA Fabian/Avery & James Shulman/ARTstor)

This upcoming year's work will focus on tool development and merging the five seed data sets. The BWR has the advantage of using other similar projects as a model of how this can be done.

Another goal for this coming year is for the BWR team and the Advisory Council to raise awareness of the BWR project in our professional community, with special effort to identify those individuals or groups who would be interested in becoming more involved. It will be important for the BWR to intersect with various communities and engage with them. Suggested affiliations include:

- Society of Architectural Historians (SAH)
- International Confederation of Architectural Museums (ICAM)
- International Federation of Library Associations – Art Libraries Section (IFLA)
- Art Libraries Society of North America (ARLIS/NA) & other ARLIS organizations worldwide
- Association of Architecture School Librarians (AASL)
- Coalition for Networked Information (CNI)
- Forum UNESCO

Discussion:

- Need to develop online presence that demonstrates the project. Perhaps a video could be developed which describes the intent of the project.
- "State of the Map" conference was suggested as possible venue for further promotion of the BWR.
- More communication with the Advisory Council would be helpful, either through the website or via email communications.

Meeting adjourned