**Draft Notes from "The Future of Art Bibliography Supplemental Meeting"**
**June 21-23, 2010**
**The Getty Research Institute, Los Angeles, California**
**Submitted by Kathleen Salomon, Getty Research Institute**

**Background:**

This meeting was conceived as a supplement to the initial Kress Foundation-funded "Future of Art Bibliography in the 21st Century" meetings held in New York on April 21-22, 2010 and organized by the Getty Research Institute in collaboration with the Kunsthistorisches Institut in Florence and the Zentralinstitut für Kunstgeschichte, Munich. Acknowledging recent events in the art historical community, including discussions of art library closures, scant funding resources for ongoing support of art libraries and projects internationally, and the cessation of the Getty's support for the continuation of BHA, the grant provided support to convene an international group of art librarians, art historians, and information specialists, including representatives from ARLIS/NA, IFLA, and CAA. The goal was to review current practices, take stock of changes in the field, and consider developing more sustainable and collaborative ways of supporting the bibliography of art history in the future. A summary report of the New York meetings is available online at: http://www.getty.edu/research/institute/fab/index.html Because the international representation at the New York meetings was hindered by the volcanic ash from Iceland, it was concluded at the end of those first meetings that supplemental meetings would be necessary to fulfill the goals of the grant.

**Attendees:**
Attendees included those unable to attend the New York meetings and a representative sampling of those who attended in New York. In addition, new invitations were extended to Jim Michalko (OCLC) and Jeremy Frumkin (University of Arizona). Thomas Gaehtgens (Director, GRI), and Marcia Reed (GRI) also participated. (See List of Participants, Appendix A.)

**Goals for GRI meetings:**

**Initial agenda:** The agenda for the second set of meetings, hosted and funded by the Getty Research Institute, was conceived before the Getty Trust entered into an agreement with ProQuest to continue the International Bibliography of Art. (See Agenda, Appendix B.) The original goals were:

1) To update and include those who could not be at the April meetings so that we would be on the same page as we discussed further action.
2) To consider the feasibility of some of the proposals regarding artlibraries.net and any other proposals going forth.

**Monday, June 21**

**Summary Discussion**

**IBA and ProQuest**

Thomas Gaehtgens, Director of the Getty Research Institute, announced the Getty Trust's new agreement with ProQuest to continue the International Bibliography of Art (IBA), at the same time underscoring the continuing need for a new initiative to bring together art historical information.

The group's questions and concerns regarding the new IBA included: the need for an advisory committee, use of authority files, the scope of ProQuest's commitment to the various and esoteric kinds of materials indexed by the IBA when it was under the Getty, and the problem of data from the former BHA being hosted in various places without federated search capabilities, due to legal issues with INIST.

**Does the agreement with ProQuest have an impact of the "Future of Art Bibliography" discussion and focus?**

It was agreed, as it had been discussed in New York, that even with the current coverage concerns being alleviated by ProQuest, this group still needs to consider future models for providing access to art information from around the globe, or, "the future of art bibliography." The news about ProQuest might allow us to consider certain problems more carefully while setting aside others that might now be taken care of by ProQuest.

**Further recurring comments over the 3 days of meetings:**

- We should complement the commercial products, (where there is already redundancy and overlap) with what is truly our niche: access to specialized, primary, or "medium rare" material, as well as to born digital resources, web-based content, etc., paying attention to the many gaps in the literature that escape other coverage by reason of medium or nationality.
- As noted in New York, the old traditional bibliography model is not the model for the future. However, whatever is envisioned should combine the efficiency of automated capture of bibliographic data enhanced with expert-contributed analytics.
- It is now the time to perform the stakeholder study and environmental scan agreed upon in New York. We have time to be reflective and analytical since the immediate concern about the gap in coverage is being addressed by ProQuest.
- We must involve other players: publishers and authors.
- What we are envisioning is open access and available to everyone. It is easy to use, visible, and discoverable. The user should be able to control data and enter data.
- "As a scholar, I only want to look in a couple of places, not in several – we need a way to link all these resources."
- We have proprietary information, third-party aggregators, and libraries. Would combining the efforts of those areas result in a complete bibliography?
- An exemplar solution already exists in artlibraries.net, but there is redundancy in the catalogs.
- We must remember the interdisciplinarity of art historical research.

**Data aggregation discussion:** Exploring the pro and cons of mining/harvesting existing resources from existing systems by members or third-party aggregators (**discussion facilitated by Carole Ann Fabian, Doug Dodds, and Jim Michalko)**

- Jim Michalko discussed how OCLC's business is aggregation of information about assets that libraries steward and the syndication of it. OCLC is narrowing its focus to aggregation and syndication and at the same time is expanding its international focus, having added recently the Hungarian National Library, National Diet Library of Japan, and French Sudoc records.
- Most referrals to WorldCat.org come from Google, i.e. OCLC records are highly discoverable on open web.
- OCLC has arrangements with ProQuest and other vendors to be able to expose their data to licensed libraries through WorldCat Local.
- OCLC can present a custom view of the big union catalog with WorldCat local.

**What can this group do vs. what can OCLC do?**

- OCLC can expose large-scale data sets in a common integrated discovery environment.
- There is a possibility of a WorldCat for art history, a customized "group" view with the functionality to expand an end-user search outwards against the broader OCLC database to discover relevant content in other disciplines and from global sources.
- An OCLC art history group view would dynamically update as libraries (in the course of their normal workflows) upload revised and new data to the OCLC database, thus providing timely updates on the bibliography – as opposed to periodic 'published' updates in a more static product.
- We might want to change our local behavior and workflow and come up with a standard way of creating records and exposing information so that aggregators can expose it more effectively.
- Some of the group would like to aggregate all subject indexing together in one record.

**What is the goal?**

- A portal for information in the arts?
- Creation of a sort of huge metacatalog for art libraries?
- We are putting the how before the what.
  - We need to find a common way to expose institutional data and then list potential consumer services needed. Some could be done by institutions and some by hosted services.
  - We should review the workflow and see where gaps are.
- We should develop, endorse, and promote behaviors for our colleagues through an art history case study. We need to view the world in interrelated ways and consider the potential and role of different resources including ProQuest, OCLC, artlibraries.net, etc.

***************

**Tuesday, June 22**

**Summary Discussion**

**Data creation:** Collaborative efforts to generate records across institutions (**discussion facilitated by Jan Simane and Rüdiger Hoyer)**

**What do we want? What are we trying to achieve?**

- The model of bringing bibliographic information together leveraging the OCLC framework is a promising vision. Such a partnership would introduce scale and marketplace innovation.
- Goal is to become a bibliographic instrument, not just a metasearch instrument.
- We are discussing both a bibliography and a search engine. Will there be a benefit on the workflow level?
- We want a unified discovery experience for art historians and art researchers, with the ability to enhance these records and push those records out. This would have an impact on the local work processes, which could be difficult.
- Are we building a repository or an index?
- What is the user experience and what do users want it to do? We need a stakeholder study (not just for users, but for publishers, librarians, indexers, etc.).
- We would like to address the problem of multilingualism of art history, but that is a larger problem than just for art libraries. However we should consider the Virtual International Authority File (VIAF) and the Getty vocabularies.
- We want to connect libraries but also organize the sharing of work. And to organize a centralized group of partners.
- What are we trying to achieve here? We need a roadmap. We want to improve user experience and to improve workflow.
- What level of integration are we talking about? Integration implies a level of data manipulation that is quite difficult.
- What is the information that people want to find? Is there a gap in the data?
- We are trying to provide more unified and seamless discovery to users.

**Artllibraries.net vs. OCLC:**

- A common strategy is needed for:
  - Content creation
  - Aggregation and data integration
  - Discovery mechanisms
  - Parsing search results in a useful display
  These criteria could be used to evaluate features of OCLC demonstration project and artlibraries.net
- Artlibraries.net is a service providing bibliographic information about the holdings of major representative art libraries.
- OCLC database is about aggregating library data.
- How can artlibraries.net grow with its current model? There are significant display and response time problems.

- What is needed, a bibliographic tool (like artlibraries.net) or a documentation of holdings (like OCLC)?
- Artlibraries.net has high participation and low cost, with a great bibliographic value.
- Artlibraries.net provides the "bibliographic value" of material from representative art libraries rather than trying to be a comprehensive access tool.
- Artlibraries.net is very slow and delivers redundant results, but with richer metadata.
- OCLC group catalog has higher cost and lower participation at this time in terms of art libraries; it does not have the coverage of artlibraries.net.
- OCLC provides access to and knowledge of holdings.
- OCLC aggregates holdings in an easy way.  However, there is a loss of the richness of bibliographic information due to OCLC's "first in" policy, where records put in first override those put in later, regardless of the quality of the metadata.

**OCLC proposal:**

Jim Michalko proposed that OCLC mount an example of a group catalog (similar to that of the Transportation Libraries ( http://ntl.bts.gov/cgi-bin/fs.scr) for a subset of art libraries that have their data represented and identifiable in WorldCat (some of whom are also represented in artlibraries.net). This would be a demonstration project to be prepared for the Lisbon artlibraries.net meetings on October 28-29, 2010.  This proposal takes into account both the bibliographic value of artlibraries.net and the aggregation, search, display and holdings features of OCLC.  This would be an experiment—it is not just to see what art libraries in OCLC own, but to see what other information and uses could be identified for an OCLC group catalog approach.  For example, while the larger OCLC community does not use social tools available via OCLC, the transportation libraries use them heavily in their group view. There is also the ability to resume a search out to the whole OCLC database from a group view. In terms of process, there would be a group evaluation and discussion once the model has been experimented with, and OCLC would consider potential changes recommended in order providing a compelling service for art libraries.  OCLC wants to understand the needs of researchers and research institutions and that is why they would be interested in pursuing this demonstration project and bearing the costs of it.

Important concerns were raised regarding the ability to put the breadth of descriptive materials into OCLC that would be needed (traditionally it has been only monographs and serials.)  It may be that making some material discoverable may not be possible, but OCLC can try.  OCLC can include in a local view all the services subscribed to by that institution, e.g. ProQuest.

**************

**Data Discovery**: Single source discovery; shared data files in redundant systems and discovery interfaces; 'discoverable' records distributed across multiple systems worldwide.  **(Facilitated by Joe Shubitowski and Jeremy Frumkin)**

**What do we want to achieve?**

- Do we see this as a destination-- a searching platform?
- Whatever we build, it should be exposed to search engines.
- How about less discovery and more access and delivery?
- Content concerns: What is not being picked up by ProQuest and how can we get access to it?
- OCLC demonstration project can help to explore gaps such as: art exhibition catalogs, festschriften, excavation reports, etc.

- What are the service layers we envision to overlay the aggregate index of library holdings? Discovery is one service element of the scholarly workflow, but support for evaluation and selection, access and delivery processes are also needed.

**Hybrid proposal based upon Solr/Lucene technology:**

- Joe and Jeremy suggested a hybrid proposal for a second demonstration project for Lisbon, based on the Solr/Lucene platform (like the Smithsonian and Oregon State University). For example in the Smithsonian Collections Search Center, they have extracted data from all 25 repositories and mapped the information to facets. One can combine browse facets and search terms; there is the ability to have different interfaces looking at the same core data.
- In a demonstration project, content would be limited, and we would then take a look at results to see what is missing and why.
- An advantage of the hybrid proposal is that it would not require libraries to be OCLC members.
- However either the OCLC group catalog or a hybrid would eventually be open access.
- Possibility of full-text and tables-of-contents.
- Social tools are provided in these types of interfaces.
- Possible elements of the feature set:
  - Must be open access
  - Crawlable by public search engines (e.g. Google, Bing, etc.)
  - Offer full-text or match against full-text providers (e.g. Hathi Trust)
  - Full text indexing for purpose of ranking search results
  - T.O.C. indexing
  - Filter by content type
  - Spell check prior to query
  - Social services (e.g. bookbag, RSS feed, tagging/annotation)
  - Ability to pass search string (query) onto broader research base
  - Automatic/dynamic data refresh
    - Z39.50 capable (on-the-fly per search)
    - OAI-PMH (scheduled data extract/harvest)
  - De-duplication of search results
  - Semantic search / concept searching

**Questions and concerns:**

- What are we building? Who is building? Is building something the key? What are our needs?
- We've got to build something.
- What do we want out of the comparison between OCLC group catalog and a hybrid example?
- The hybrid idea could set us apart.
- If we want to go beyond OCLC we need to figure out what we need and how it will be funded – someone could come up with server room. Talk to the Smithsonian. Oregon State University has software (LibraryFind.) We need to figure out where we want to be at the end of the project.
- Basic technical needs: server space, software, system architect/developer time and effort, QC/analysis/rebuild cycles, collaboration with local technologists to prepare and ship data.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Wednesday, June 23:**

**Summary discussion:**

Over the past days we have been groping toward a list of requirements and a strategic direction.

It was concluded by many that the group made some progress toward something new and important for global art history research; perhaps it could be termed a "virtual special library on an international scale."

**3 models were discussed:**

1) Existing artlibraries.net
2) OCLC group catalog demonstration project
3) Hybrid demonstration project based on Solr/Lucene architecture

In each of these models, unlike commercial ones, we are looking at post-selection by the viewer instead of pre-selection by a third-party provider (e.g. abstracting and indexing service, vendor database, etc.).

It was agreed that the group would move forward with two demonstration projects:

- OCLC via Jim Michalko for the group catalog demonstration
- Solr/Lucene model with Joe Shubitowski taking the lead.

We will contrast and compare the two demonstration projects to learn about the total experience. We will have the beginnings of a best-of-breed analysis with the purpose of gathering technical and user requirements.

Points of discussion regarding the demonstration projects:

**OCLC:**

- Problem of the large academic and national libraries not being able to partition out an art collection from the larger library holdings.
- OCLC must be working on faceting results and this would be a good group to experiment with.
- This would be a great way to be able to work with OCLC and get them involved, and to advocate for our needs.
- OCLC will bear the costs for the demonstration project.
- Kathleen and Carole Ann will be in contact with Jim Michalko regarding this.

**Solr/Lucene:**

- Offers more independence and flexibility than the OCLC model
- This is potentially a very large aggregation infrastructure project, which could cost millions initially. For example, the Hathi Trust, which also takes data from other institutions, has an enormous operating budget.
- Requires staffing, even for demonstration project.
- Could investigate already existing software such as LibraryFind or Getty's GAIA software which was in fact built for community collaboration

- Perhaps we should just use what already exists at the Getty (which has been working with Solr/Lucene for a major internal project) and perhaps add Avery to it, and add a front end. This could be enough for a practical review at Lisbon.
- Joe Shubitowski will take lead on this, consulting with Doug Dodds.

**Administrative matters:**

- Name of project:  Future of Art Bibliography (FAB)
- Communication: Initial communication can be done via a GRI webpage that might eventually move to its own domain.   It is not a Getty-lead initiative, but it is currently supported by a grant to the GRI from the Kress Foundation, and the Getty will continue to be an active and supportive participant.
- Organization: A small task force was appointed by consensus to work between now and the Lisbon meetings in October:  Douglas Dodds, Carole Ann Fabian, Rüdiger Hoyer, Kathleen Salomon, Jan Simane, and Joe Shubitowski.
- It was noted that strong management of this project is key, and while this small task force can function in the short term, something more needs to be in place for the future.  Funding from grant or other external or collective sources is essential.

**Action items to be addressed by the small task force in the short term:**

- Move ahead with Jim Michalko for the OCLC project demonstration for Lisbon. (Kathleen and Carole Ann)
- Solr/Lucene based project demonstration for Lisbon (Joe and Doug)
- Develop agenda for Art Libraries Section at IFLA in Gothenberg in mid-August, and plan to start stakeholder surveys/discussions there. (Jan and task force)
- Develop agenda for Lisbon  artlibraries.net meeting
- Set up web page to communicate regarding project (Kathleen and Joe)

**Additional action items to be discussed at IFLA and in Lisbon:**

- Define the project and requirements by:
  - Developing a grant proposal to provide project funding and a strong project manager(s) who will address the necessary research components of this project (first recommended at the New York meetings) including  :
    - an environmental scan
    - a stakeholder study.
  - Community involvement and outreach:
    - Inform community of project development
    - Work with groups such as the artlibraries.net target libraries, CAA, IFLA, ARLIS, ARIAH, and CIHA and other stakeholders to develop use cases and user needs.

## PARTICIPANTS

Mary Clare Altenhofen
Associate Librarian of the Fine Arts Library for Public & Research Services
Harvard Fine Arts Library
Boston, Massachusetts

Stephen Bury
Andrew W. Mellon Chief Librarian
Frick Art Reference Library, The Frick Collection
New York, New York

Douglas Dodds
Head of Central Services, Senior Curator, Computer Art,  Word and Image Dept.
National Art Library, Victoria & Albert Museum
London, United Kingdom

Carole Ann Fabian
Director
Avery Architectural & Fine Arts Library, Columbia University
New York, New York

Terence Ford
Head, Research Databases
Getty Research Institute
Los Angeles, California

Jeremy Frumkin
Assistant Dean / Chief Technology Strategist
The University Arizona Libraries
Tucson, Arizona

Thomas Gaehtgens
Director, Getty Research Institute
Los Angeles, California

Rüdiger Hoyer
Chief Librarian
Zentralinstitut für Kunstgeschichte
Munich, Germany

Roman Koot
Head of Public Services
Chief Curator Library and Foreign Art
RKD/Netherlands Institute for Art History
The Hague,  Netherlands

Jim Michalko
Vice President
OCLC Research
San Mateo, California

Martine Poulain
Chief Librarian
Institut National d'Histoire de l'Art (INHA)
Paris, France

Marcia Reed
Chief Curator and Head, Collection Development and Exhibitions
Getty Research Institute
Los Angeles, California

Kathleen Salomon
Head, Library Services & Bibliography
Getty Research Institute
Los Angeles, California

Joseph Shubitowski
Head, Library Information Systems
Getty Research Institute
Los Angeles, California

Jan Simane
Chair, Art Libraries Section of International Federation of Library Associations (IFLA)
Chief Librarian, Kunsthistorisches Institut in Florenz
Florence, Italy

Kenneth Soehner
Chief Librarian, Thomas J. Watson Library
Metropolitan Museum of Art
New York, New York

Cameron Trowbridge
Manager, Research Services
Getty Conservation Institute
Los Angeles, California

Lois White
Head of Reference Services
Getty Research Institute
Los Angeles, California

AGENDA


Monday, June 21 – Getty Research Institute, L3 Seminar Room

- 9:30 - 10:00    Arrival and breakfast

- 10:00 – 12:00   Welcome  (Thomas Gaehtgens)
  Introductions and Logistics (Kathleen Salomon)
  Review of meetings in April and discussion of conclusions;
  (Kathleen Salomon and Martine Poulain)

- 12:00 – 1:00    Lunch (Eucalyptus Garden)

- 1:00 - 3:00     Discussion Session 1:  Data aggregation—explore the pros
  and cons of mining/harvesting existing resources from existing systems by
  members or third-party aggregators (Carole Ann Fabian and Doug Dodds)

Tuesday, June 22 – Getty Villa Ranch House, VR212

- 9:45 - 12:00    Welcome and plans for the day (Kathleen)
  Discussion Session 2: Data creation—collaborative efforts to generate
  records across institutions (Jan Simane and Rüdiger Hoyer)

- 12:00 – 1:00    Lunch – Ranch House Terrace

- 1:00 – 3:00     Discussion Session 3: Data Discovery—Single source discovery; shared data
  files in redundant systems and discovery interfaces; 'discoverable' records
  distributed across multiple systems worldwide (Joe Shubitowski and Jeremy
  Frumkin)


Wednesday, June 23– Getty Research Institute, L3 Seminar Room


- 9:30 - 12:30    Review of past discussions; conclusions and next steps  (All)