MINUTES

Meeting of the artlibraries.net committee
with colleagues from the FAB project and Uwe Dierolf

Munich, Zentralinstitut für Kunstgeschichte January 26th /27th, 2012

Participants

artlibraries.net committee
Véronique Goncerut Estèbe, Musées d’art et d’histoire de la ville de Genève
Rüdiger Hoyer, Zentralinstitut für Kunstgeschichte, Munich
[Deborah Kempe, The Frick Collection, New York (temporarily connected on both days via Skype)]
Geert-Jan Koot, Rijksmuseum Amsterdam
Paulo Leitão, Gulbenkian Foundation, Lisbon
Michael Rocke, Villa I Tatti, Florence
Jan Simane, Kunsthistorisches Institut in Florenz

Guests:
Uwe Dierolf, Karlsruhe Institute of Technology (KIT)
For the ‘Future of Art Bibliography’ project:
Kathleeen Salomon, Getty Research Institute, Los Angeles
Joe Shubitowski, Getty Research Institute, Angeles
Christian Gries, Janusmedia, Munich (on January 27, from 2 to 3 p.m.)

Apologized:
Irena Murray, Royal Institute of British Architects, London

This joint meeting of the artlibraries.net committee with Uwe Dierolf (KIT Karlsruhe), Kathleen Salomon and Joe Shubitowski (Getty Research Institute) was motivated by the intention to elaborate a proposal for the short- and mid-term technical und structural development of artlibraries.net against the background of the FAB initiative, with the perspective to discuss this proposal at the next bi-annual meeting of the artlibraries.net partners in autumn 2012.

Current organizational matters were discussed in the afternoon of January 27, without the guests from Karlsruhe and Los Angeles
Discussion of proposals for an improved technical bases for artnet

Continuing the discussions we had at the last general meeting (Lisbon, October 2010) and at the last committee meeting (Zurich, June 2011), the participants re-analyzed the proposals and prototypes for a renewal of the artnet search infrastructure and tried to come to a definite conclusion.

(See also Art Libraries Journal 36.2011,3, Special issue : the future of art bibliography.)

1. Federated Search Proposal by Deep Web Technologies, Santa Fe, NM¹, based on its ‘Explorit Research Accelerator Federated Search Solution’²

The Deep Web Federated Search prototype, developed for artnet on the initiative of Deborah Kempe and first presented to the committee in October 2010, has already been discussed at the last committee meeting in Zurich in June 2011 (the minutes have been sent to all ALN partners). The free demo site³ is still available (as of April 5, 2012) but shall be closed soon. It works with a selection of 10 targets. (The ‘real’ artnet has 43 targets.)

The committee discussed the following topics:

- functional characteristics of the Explorit system
- ranking criteria
- possibility of deduplication
- re-examining of Deep Web’s cost-estimate

**Functional principles of the Explorit system**

Users can choose between the standard “full text” and “advanced search” option. Selecting single targets is not possible with the “full text” search. In the “advanced search” option, the set of search fields is very basic compared to the existing artnet (no ISBN/ISSN, publisher or subject heading fields).

¹ [www.deepwebtech.com](http://www.deepwebtech.com)
However, the main difference lies in the fact that the Deep Web solution is producing just one unified hit list whereas we currently show individual result lists for each target. The results are divided into screens with 10 items maximum. Compared to the KVK technique, the Deep Web prototype has thus a search engine ‘look and feel’ and gives the impression to react faster. The search engine feeling is underlined by faceting and clustering elements. Compared to the existing artlibraries.net, Explorit is apparently not concentrating on representing individual holdings but rather on producing quickly some results from the fastest targets.
As Deep Web explained in an e-mail from January 23, 2012, to Deborah Kempe: “Our federated search software is designed to respond to the needs of busy users who are used to almost-instant results from other search engines. As such, it immediately displays the fastest results that are returned. This allows the user to review some results while waiting for the rest of them to be brought back. Then, when the search has finished, the user sees a window pop up that says ”Additional Results Available” (see attached screenshot).”

Furthermore there are some additional features which do not exist in the current artlibraries.net interface like:

- thumbnail images (where applicable),
- making a selection in the hit list
- printing and emailing results
- “did you mean?” suggestion
- automatic updates

The system is indicating the total number of results found and producing an additional list of result numbers arranged by target. Chronological clustering allows to restrict the display according to publication date. It is also possible to sort the results according to different criteria and to limit the results to those from a single target.

**Ranking criteria**

The ranking criteria have been summarized by Deep Web as follows (transmitted by Deborah Kempe in an e-mail from September 13, 2011):
"Keyword density and keyword position are both used to rank. Words appearing at the beginning have a higher weight than those at the end, which in turn has a higher weight than words in the middle of an occurrence. Longer records, like those with Table of Contents notes, have a de facto weight given their higher word content. We could assign a higher weight to more recent publications (within the last 2 or 3 years for instance). The "star" ranking display may be turned off (others have chosen to remove them, including Stanford), but removing ranking completely is not advised. Display by Rank is the display default, but it could be changed to Alphabetical if preferred." The committee raised the question whether Deep Web’s ranking would interact with rankings implemented by the targets themselves.

**Possibility of deduplication**

Explorit’s search engine look and feel may cause the impression that results are being deduplicated. However, we quickly had to realize that the first impression was wrong and that even with Deep Web deduplication is not realizable:

"Based on earlier conversations, Brian [i.e. Brian DeSpain from Deep Web Technologies] had turned on a "title and date" de-duping feature. Because verification is not 100% effective and because duplicates disappear, rather than cluster, he will turn that feature off, and essentially there will be no de-duping or clustering. Clustering of results (i.e. one master record with a list of holdings) will not be possible. Users can refine/limit the initial search using faceting on the sidebar or dropdown menus." (transmitted by Deborah Kempe in an e-mail from September 13, 2011).

In the same way as the KVK, Explorit is a federated search tool, not an indexed based search engine which generates entire hit lists in a fast manner. As only few (basic) title elements are taken into account in order to produce a limited hit list, deduplication is actually not possible.

**Financial proposal**

Deep Web Technologies has presented a calculation of the costs for a basic implementation:

- one-time charge of $250 per target (= $10,750 for all current targets)
- current maintenance (annual fee $600 per target)
- setup for new targets ($1500).

The setup costs for new targets are lower than the current fees. However, maintenance will become ca. 50% more expensive. Furthermore, the basic implementation would of course require an additional financial effort. For $250 per target, only a minimal amount of customization is assured. This does not include a multilingual interface. According to an e-mail from
January 26, 2012, each additional language would mean a one-time charge of $3000 (= $12,000 for three languages).

Conclusions of the committee

The Deep Web solution is centred on producing as fast as possible a limited list of hits. As explained by Deep Web in an e-mail from January 26, 2012 to Deborah Kempe: "When our federated search software searches sources, it does not always bring back all the available results at all the sources. This is especially true for a very broad search term, such as Full Text: Picasso. This is for a couple reasons, including timing and quality of results. It takes longer to bring back more results. If the software tries to bring back thousands of results, it's very likely that sources will time out and stop communicating with us, which means we might bring back 0 results instead. Also, users are only willing to wait so long. In addition, our application aims to bring back the best, highest-ranked results for the user. That means that some lower-ranked results are not brought back and displayed.

For each connector/source, we set a threshold of how many results can be brought back. This is done based on the speed of the source and ease with which results can be retrieved. Generally, the upper limit is 100 results per source. In some cases, this can be raised to 200 results per source. We do not recommend going beyond that threshold, as most sources will time out and be effectively returning 0 results."

This means that the ‘seductive’ rapidity of producing a unique hit list is finally possible only due to implementing strict limits concerning the time available and the information returned, with the aim of displaying a kind of first choice. And even if the origin of the results is no ranking criterion, it is evident that the first choice of ranked results always prefers the fastest targets.

"Sources can vary greatly in their speed. Because we return incremental results, this may look like some sources are getting ranked highly and others are not. This is not what is happening. Very likely, the source that looks unranked is still being searched.” (e-mail to Deborah Kempe from January 23, 2012).

To the committee it seemed that Deep Web’s attempt to confer a ‘search engine look and feel’ to what in fact remains a federated search solution - just like the KVK technique - is impressive rather in itself and may be useful for many domains. In the artlibraries.net context, however, the product seems to counteract some of the project’s generally accepted core principles: Based on a selection of – to a certain degree – complementary targets, artlibraries.net has been conceived to produce results in a bibliographically reliable way. And although the current KVK model displays in the first moment also only a selection of results, the total number of hits is visible as well as all responding targets. Moreover, the possibility to compare the displayed hit lists of all responding targets allows to get an
overall impression of the quantity of literature involved and to identify at
first glance the probably most interesting collections. Of course, the tech-
nique is suffering from the relatively slow reaction of some of the targets
and of a certain general clumsiness (as described in R.Hoyer: Meta cata-
logues and search engines : artilibraries.net and the state of the art. – In:

Compared with the existing artilibraries.net model, the variety and rich-
ness of the participating institutions’ holdings do not seem to be repre-
sented adequately by the Deep Web pilot. The test queries done by the
committee seem to confirm that results from certain targets will be dis-
played most of the time on the first search screens whereas other targets
will hardly ever be visible at this stage. This fact has been judged by the
committee as a serious problem and a contradiction to the artilibraries.net
policy rooted in the plurality and multiplicity of the partners, since the full
visibility of all the participating targets’ contributions is one of the principle
goals of the project.

The committee finally decided not to recommend to pursue this proposal,
given that closer analysis has shown that implementing the Deep Web
solution instead of the current KVK technique would not be a way to ob-
tain a really fundamental improvement of the artilibraries.net project and
that the considerable financial and organizational effort requested could
after all not be justified. Nevertheless, Deep Web’s pilot application due to
Deborah Kempe’s initiative was very appreciated because it helped enorm-
ously to develop a better awareness of the problems involved. There-
fore, the committee members assembled in Munich asked Deborah Kempe
to express our sincere thanks to Deep Web Technologies for their com-
mitment and for all what we have learned through discussing with them.

2. Search Engine Based Solutions

Background: For the general meeting in Lisbon in October 2010, Joe Shu-
bitowski from the Getty Research Institute has prepared a search engine
prototype, based on Solr/Lucene and developed as a result of the discus-
sions within the FAB initiative. At the same occasion, Bruce Washburn
from OCLC has presented the prototype of a WorldCat-API for art librar-
ies.4

As the artilibraries.net project is based on a rather informal co-operation,
so far not involving a real institutional and financial basis, and as it has
not been so far possible to secure third-party financing for such a search
engine, pursuing this solution seemed not practicable for the near future,
especially because of the enormous organizational efforts and, on the
other hand, without being able to guarantee that all present artilibrar-
ies.net partner institutions could participate in such a solution.

4 See Joseph Shubitowski / Bruce Wasburn: Imaging the future of art bibli-o-graphy:
using prototypes to evaluate technical approaches. – In: Art Li-braries Journal 36.2011,
3, pp. 40-45.
Hybrid solution for integrating a WorldCat API into artlibraries.net

Given the predictable difficulties with both the Deep Web pilot and an index-based search engine, the committee turned towards a more pragmatic approach: When committee member Geert-Jan Koot reminded us of his libraries’ efficient cooperation with OCLC and the fact that his library’s users seem to be much more attracted by WorldCat than by artlibraries.net, we finally came back to the proposal made by Bruce Washburn in context of the FAB initiative: As a result of the discussions on the future of art bibliography, started on the initiative of the Getty Research Institute in early 2010, and continued at the General Meeting of the ALN partners in Lisbon, October 2010, the OCLC representatives (in particular Bruce Washburn), involved in both the Getty meetings (New York City (April 2010) and Los Angeles (June 2010)) and the Lisbon meeting as well, agreed to install an experimental bibliographical discovery tool, based on the data pool of the OCLC/WorldCat and a special pilot API taking into account the data of 30 art libraries participating in WorldCat.6 ‘Art Libraries Discovery Experiment’ is based on the WorldCat Search API, a web service providing access to specific library holdings.7

WorldCat art libraries API home page

Geert-Jan Koot pointed out that several of the current artlibraries.net partners are already OCLC-libraries or that they send their data regularly to the WorldCat pool via national or discipline specific consortia. The experimental WorldCat art libraries API not only covers a part of the existing artlibraries.net targets but adds ‘automatically’ outstanding North American art libraries (for instance the Art Institute of Chicago or the National Gallery of Art Library) which for some reason are not part of the artlibraries.net network but partners of OCLC/WorldCat.

Uwe Dierolf then explained that a combination of the existing KVK-based artlibraries.net federated search with the OCLC/WorldCat API would probably be the most pragmatic approach to renew our project in an organic

---

5 cf. http://www.getty.edu/research/scholars/research_projects/fab/
6 This pilot API is to be found at: http://experimental.worldcat.org/aldp/.
7 http://www.worldcat.org/affiliate/tools?atype=wcapi
manner by rendering the search more efficient without restricting the number of participants and without having to abandon the existing working structure!

*Characteristics of the WorldCat art libraries API*

The home page lists all WorldCat-art libraries identified by OCLC as pertinent for this pilot interface. Some of them are already artlibraries.net partners, others not. If this list is exhaustive, the question should be raised under which conditions pertinent holdings could be added to WorldCat. As for the technical basis, we have to stress the fact that we are dealing here with a real database, i.e. the WorldCat database viewed through special filters, not with a federated search tool and not with a search engine. This explains the very fast availability and completeness of the results, but also the very interesting possibility to extend the search query to the world outside art libraries, i.e. to the entire WorldCat and to special materials provided by OCLC/WorldCat.

*Screenshot: WorldCat art libraries API results screen*

Because we are dealing with a real database the results are more complete: no partial short title lists, no fragmentary title information, but complete data as we expect them from a normal online catalogue. Books, articles, sales catalogues and digitized texts are presented in separate hit lists on the same screen.

What is most interesting is the fact that the records produced by the art libraries API are intertwined with the whole WorldCat world: Thus it is possible to search for holdings anywhere and to link seamlessly the art libraries planet to the rest of the library and information universe. Fur-
thermore, ‘Art Libraries Discovery Experience’ benefits from the additional data pools available through WorldCat (online resources, article databases, special databases like SCIPIO, digital collections etc.). Not every resource may be accessible for every user everywhere but at least the basic bibliographic information is made available. Thanks to the fact that the bibliographic records shown are part of the entire functional and semantic WorldCat universe, it is possible to continue a hyperlink search in the wider WorldCat context or to provide relevant subject terms which are automatically drawn from literature catalogued elsewhere in WorldCat.
After re-examining this pilot API, it seemed to the committee members that it has decisive advantages compared to any federated search or search engine approach. But since the inclusion of the catalogue data of all artlibraries.net partners is an inevitable condition for the efficiency of this model it is obvious that it cannot be regarded a short-term solution for artlibraries.net.

Therefore, Uwe Dierolf made a suggestion for a short- or middle-term solution, which is the following hybrid approach:

- Integrating the WorldCat art libraries.net API as one target into artlibraries.net
- Eliminating all targets already covered by the WorldCat API
- Maintaining as ‘individual targets’ only those which cannot be accessed via WorldCat

Organizationally speaking, this would mean to reduce significantly the current number of targets. Less targets means, technically speaking, improved reaction time. Financially speaking, the solution could strongly reduce the general maintenance costs, for the WorldCat API would be counted by the KIT Karlsruhe as one target, no matter how many library holdings it may cover. Relying as much as possible on the powerful options offered by the WorldCat API would probably confer much more international significance and acceptance to artlibraries.net, especially in North
America. artlibraries.net could thus perhaps even replace the more ambitious search engine project conceived within the FAB initiative or, at least, become FAB’s bibliographic core element. artlibraries.net and the FAB initiative would finally come together.

In order to get a clearer and a more precise idea how such a vision could be put into practice, Uwe Dierolf is ready to implement a WorldCat target into artlibraries.net to provide a basis for discussions at the next general artlibraries.net meeting.

However, before we can be sure that this ingenious proposal will be realizable, a number of questions have to be answered:

- Will OCLC be willing to cooperate with the artlibraries.net and above all to allow Uwe Dierolf to experiment with their API without the KIT Karlsruhe being an OCLC member? [Note: In the meantime OCLC provided to Uwe the appropriate and unlimited access to the API.]
- How would it be possible to integrate the holdings of other libraries, being already artlibraries.net targets but not OCLC members, into WorldCat in order to get the maximum benefit from the WorldCat API? How is the OCLC policy, what are the conditions and costs for such an integration?
- Is the described ‘hybrid’ approach realizable without bigger investments which are impossible with the present informal structure of artlibraries.net? Or would a cooperation with OCLC require a formal institutionalization of the artlibraries.net cooperation?
- So far and also due to the technical restrictions imposed by the search engine technique, artlibraries.net targets have been carefully selected, with the aim of producing a maximum of bibliographically significant results with a minimum number of targets. The WorldCat API does not call for such restrictions. How would this influence the general artlibraries.net policy?

The committee agreed to clarify these questions quickly and to contact OCLC in order to help Uwe Dierolf to realize a pilot target as soon as possible (see above). Only such as pilot target can prove whether the combined approach - maintaining the artlibraries.net site on one hand and using it as a gateway to the WorldCat API on the other hand - would be satisfying from a functional point of view. Given the excellent contacts between OCLC and the Getty Research Institute, the colleagues from Los Angeles declared that they were ready to help establishing the necessary dialogue with OCLC representatives to start discussion on a possible initiative. Apart from this, committee members who will attend the IFLA conference in Helsinki in August (Kathleen Salomon, Rüdiger Hoyer, Deborah Kempe, Jan Simane) will try to organize there a meeting with OCLC representatives.

3. Short-term improvements to artlibraries.net
Besides the general modifications discussed above, there is a certain number of improvements that could be implemented into the existing artlibraries.net interface, probably without additional costs. Uwe explained that although no important new developments are for the moment being prepared for the KVK, some new features have been developed within the KVK context, for example an option to filter the results in the hit lists according to criteria otherwise available only on the start page (as ‘electronic resources’ or ‘year of publication’), thus allowing to refine search results in a facet-like manner. The committee appreciated this proposal very much and encouraged Uwe to prepare it also for artlibraries.net, if possible.

Christian Gries from Janusmedia, Munich (agency in charge of the artlibraries.net user interface), joined the meeting for an hour on Friday afternoon. He reported that slight modifications and additional features could be implemented into the current interface without additional costs, in particular link buttons which connect to the most frequented social media networks and the option to save personal preferences (realized in the meantime). The committee agreed with this proposals. Another important information from Janusmedia: As the additional artlibraries.net interface, existing so far within the portal arthistoricum.net, has been replaced by a direct access in context of the recent relaunch of arthistoricum.net, the monthly maintenance costs can be reduced to EUR 150 from January 2012 on.

4. Portal of Digitized Art-Historical Texts (DAHT)

Kathleen Salomon and Joe Shubitowski presented the current Getty project which aims to create a portal to digitized art-historical source texts available on the internet. This project is based on a cooperation between the GRI, the Avery Architectural and Fine Arts Library, the Institut national d’histoire de l’art, and the Universitätsbibliothek Heidelberg. The main scope is to build a trusted portal for descriptive and technical metadata for digitized scholarly sources in the field of art and architectural history. This discovery environment, aimed at researchers who wish to access full-text versions of canonical art-historical texts, rare books, and related literature, will make these resources easier to find and will help to avoid unintentional redundancy of digital texts. Joe gave a demonstration of the technical prototype which in basic is a Solr/Lucene search engine. The starting point for content selection are the standard bibliographies of art-historical sources like Cicognara’s ‘Catalogo ragionato’ and Schlossers ‘Kunstliteratur’. On the first stage of expansion the discovery tool will offer rather basic features. In the long run a number of additional functionalities and search options shall be added. The project is explicitly open for
cooperation with all libraries and archives which can contribute to the register. Against the background of the discussion on a metamorphosis of artlibraries.net from the federated search to a index based discovery system the upcoming DAHT is a very promising source which certainly will be integrated in a future moment.

5. Forthcoming general artlibraries.net meeting

According to the traditional bi-annual rhythm, the next general meeting of the artlibraries.net partners should be scheduled for autumn 2012. The initial idea to organize this meeting in Vienna, Austria, had to be abandoned because the colleagues from Vienna will already host the next annual conference of the German art libraries association (AKMB) and therefore see no possibility to host an additional meeting. Paolo Leitão extended an invitation from the Gulbenkian Foundation to come again to Lisbon in November, at the time when the Portuguese library association is organizing its next conference there. This generous proposal was appreciated very much. However, most of the participants thought it should be avoided to go to the same place twice in a row and that it would be practical to have a venue which can be reached most easily by most of the partners. In the meantime, we have received an invitation to Paris, where there are several artlibraries.net partners, for September 27/28/29, 2012. The meeting will take place or at the Deutsches Forum für Kunstgeschichte / Centre allemand d’histoire de l’art or and at the Musée des arts décoratifs. Precise information will follow as soon as possible.

(Minutes: Rüdiger Hoyer, Jan Simane)