kubikat in a transatlantic context?

Paper presented at the

Artlibraries.net and the Future of Art Bibliography conference

Copenhagen, October 9-11, 2014

The „kubikat" union catalogue (www.kubikat.org):
- Kunsthistorisches Institut in Florence, Max-Planck-Institut (Florence, Italy)
- Zentralinstitut für Kunstgeschichte in Munich
- Deutsches Forum für Kunstgeschichte / Centre allemand d'histoire de l'art, Paris
- Bibliotheca Hertziana, Max Planck-Institut für Kunstgeschichte, Rome

is one of the most important specialized data sets joining WorldCat on the occasion of the new 'Art Discovery Group Catalogue' project.

The complete kubikat data are accessible on www.kubikat.org which is the Aleph-based online catalogue for the records of all four institutions.

Since 2011, the kubikat libraries are cataloguing in the "Bibliotheksverbund Bayern (BVB)", the renowned Bavarian network, with immediate download to kubikat.

kubikat currently (10/2014) comprises approximately 1,690,000 records. More than 50 % (ca. 850,000) of these are records for articles from periodicals and collected writings. Due also to several important conversion campaigns of the old subject catalogues, a high percentage of these records comes with subject headings created by the participating research institutes. kubikat is one of the most important bibliographical tools for the history of art, besides the historic or current bibliographies as RAA, RILA, BHA, IBA. That is to say that these data are of crucial importance for the character of the Art Discovery Group Catalogue project. The upload of the kubikat data to WorldCat and the way they are finally indexed and displayed can certainly serve as a case study which should be important for similar uploads during the current project phase.

At the occasion of the launch of the Art Discovery Group Catalogue at the ARLIS/NA conference in Washington in May of this year, I had already the opportunity to talk about what we had planned concerning the upload of these data to WorldCat. At the time, just ca. 25,000 sample records had been uploaded.

The kubikat data are being uploaded to WorldCat via the Bavarian network BVB. The participating kubikat libraries receive individual OCLC symbols. The kubikat is not visible as such in WorldCat (nor is it in the BVB database). The necessary mapping from the German MAB2 format to MARC21 is being defined together with the specialists from the Bavarian State Library in cooperation with OCLC.

Depending on the origin of the data, especially when dealing with converted subject headings from the old 'Sachkatalog' of the Zentralinstitut with its nearly 1 mio. subject entries, the kubikat records may contain valuable subject heading strings whose structure and contents is not in line with the current German cataloguing rules for subject indexing, the RSWK. Furthermore, the old German MAB2 format, still valuable in German networks, does only allow 10 repetitions for orthodox RSWK fields. As kubikat is participating in the BVB network, this leads to inconsistent and heterogeneous solutions already on the level of the Bavarian union catalogue. In short: If there are too much subject heading strings accumulating e.g. from kubikat conversion campaigns, part of these will exceed the repetition possibilities of the MAB2 format. In such cases, we can still recur to an alternative MAB2 field (field 710 = subject headings following other rules than RSWK), albeit with no authority data linking. But still, in many cases, the complete information cannot be preserved on the Bavarian or German level. For this reason, we have defined special, additional fields (K00 to K99) in our proper kubikat database which allowed to preserve the original field content when we joined the Bavarian network some years ago and which allow us also to preserve the entire information coming from conversion campaigns.

When we joined the 'Art Discovery Group Catalogue' project, our aim was to free ourselves from the limitations dictated by the MAB2 format and by the German RSWK rules and to provide the community worldwide with the complete information available in our kubikat database. The colleagues from the BVB supported this aim in a very generous manner, although the BVB is normally not uploading subject heading strings to WorldCat but just single subject headings characterized as 'swd' or 'gnd'. They were ready to upload to WorldCat the complete kubikat strings. Therefore they helped us to define an export format within MARC21, together with colleagues from OCLC. Our project simply was to preserve the subject heading strings as they are in kubikat and to add them to the data already existing in WorldCat. The MARC21 format allowed to prepare the data in a seemingly convenient form, although abandoning the links to GND authority records, using fields 600 to 655 with second indicator 4 (Source not specified) and subfield $e ("Relator term") for

additional free text explanations ('footnotes') sometimes added to the strings. OCLC did not object.

This functioned apparently for the sample of 25,000 records uploaded for the ADGC launch in Washington, although the 'footnotes' in subfield $e were not displayed.

At the time of the launch, we were proudly looking forward to have our kubikat string systematically added to the records and, as the case may be, in addition to international subject headings as LoC subject headings or RAMEAU subject headings.

The data were handed over to OCLC in August 2014, before the IFLA conference in Lyon, and rapidly uploaded to the WorldCat database. The upload was finished on August 24, 2014.

1,289,912 bibliographical records from kubikat out of 1,689,141 were added to WorldCat as completely new records.

But in most cases we analysed, we were looking in vain for our subject headings.

We saw records with multiple international holdings where our subject headings from Bavaria had not been added, although there were no German subject headings so far. We saw cases of records with holdings only from kubikat libraries where nevertheless the subject headings furnished are not displayed. In other cases, the BVB fields for subject heading strings following other rules than RSWK (MAB2 field 710) were accepted as such, the even more complete proper kubikat fields (K00) having been rejected.

It took a while before we received relatively conclusive, but frustrating explanations from OCLC (Mail from Larry Wolkan, 28.8.2014):

"OCLC transfers headings on the basis of recognized schemes. Second indicator 4 is not a recognized scheme, and adding one unique heading string after another with the same second indicator does not improve the record. In cases like this example, we worked with the library to get a MARC code to identify their particular set of subject headings […] and had them output their records with 6xxx second indicator 7 and subfield $2 with new code.

This is most likely what should happen here."

OCLC also wrote us that "all the subject headings have been included when the record is 'added as a new record. It looks like 76+% of the records sent for this

round were added as new records." (ibid.) We have not been able to verify this statement so far.

Finally, OCLC informed us that

"In order to get something like a Kubikat code approved, it requires LC staff to establish new codes. As noted on the LC web site: 'Requests for the assignment of a new source code should include a full bibliographic citation for the source documentation or web site for the code or term list, code or term database, scheme (e.g., for classification schedules), or rules (e.g., for description conventions). For printed sources, include a copy of the title page. For online sources, include the URL.'"

In addition to this, we discover another problem concerning article records. Apparently the kubikat article records are recognized as such and indexed accordingly for the article facet. However, we were not able to reproduce the linkings from the article records to the record for the containing periodical oder collective writing. Apparently, the linking is not done with OCLC identifiers but with ISSN/ISBN, which arenot contained in these records. So far, the user has to go to the kubikat in order to get this information. Moreover, there is no de-duplication of our article records with thoses coming from JSTOR, ArticleFirst etc, where an apparently ISSN-based link leads to the record for the containing work in our local catalogue. So we are in a situation where kubikat is contributing to WorldCat for example hundreds of additional articles for outstanding art historians like Willibald Sauerländer without providing the user without a practical way to localize the item outside the kubikat libraries.

Several months after the project has been launched, it now appears that WorldCat and Art Discovery Group Catalogue are not so sophistically administered and not so open to new specialized information from specialized sources as we might have imagined. We are facing problems which never occurred with artlibraries.net where each participating institution is free to present its data as they appear in the target system. The idea of having the specific historic kubikat subject indexations evaluated and eventually rejected by LoC staff proves that after all the Art Discovery Group Catalogue project has but limited freedom and that the statement that the Art Discovery Group Catalogue is not a project by OCLC is simply misleading. The intended "ability to discover aggregated bibliographic data from all the participating art libraries" (information text on the website) may be in fact be restricted. In any case, the hope to provide a well-functioning additional important data layer within our transatlantic project is very much put into perspective by the final facts. You

will have to continue to go directly to the kubikat and not to the Art Discovery
Group Catalogue in order to get the real thing.

Rüdiger Hoyer