A Portal to Dutch Academic Heritage: www.academischecollecties.nl

HENRIETTE REERINK*

ABSTRACT

The UNICUM development project, commissioned by the Dutch Academic Heritage Foundation Stichting Academisch Erfgoed (SAE), has been carried out by the five classic Dutch universities in 2010–2012. UNICUM, short for 'University Collections and University Museums', has received a national government grant to create a digital portal to Dutch academic heritage. The portal, which can be reached via www.academischecollecties.nl, presents both academic archives and museum and library collections. Images, collection metadata and items can be found on one site. The UNICUM idea is inspired by the Online Archive of California.

The project is important because it crosses the traditional sector boundaries between museums, libraries and archives, it creates awareness of the opportunities which this cross-sectoral approach offers, and it retains the context of – and the relation between – objects within collections as a whole. Moreover, the joint effort brings to light the importance of creating metadata according to international standards to stimulate re-use and exchange of content. In addition, UNICUM intended to be a technical project in which multi-level descriptions are presented and can be browsed in a structured way (collections linked to objects, and archives linked to separate documents). Now the project is finished, the focus will shift to generating content.

Keywords: academic heritage; university history; university museum; library; archives

Introduction

Recently the UNICUM portal website www.academischecollecties.nl was launched, a joint effort of the five classic Dutch universities. The portal presents the academic heritage of the Dutch universities. Academic heritage comprises the pre-1850 collections that have historically grown, or actively been collected to meet the educational and research purposes

Fig. 1: www. academischecollecties.nl.

of the universities. Examples are, for instance, historic microscopes, anatomical models or photographs. In addition, faculty archives, and paintings belonging to universities, as well as rare book collections are part of the academic collections. Because of this diversity, the portal presents both academic archives and museum and library collections.

One of the sources of inspiration for the academic heritage portal was the Online Archive of California, which was conceived in 2002 and has expanded ever since. In the Online Archive of California more than 200 Californian cultural heritage institutions present their material at both collection and item level. In the UNICUM project, we started off with five institutions. Our project was commissioned by the SAE, the 'Stichting Academisch Erfgoed' or Dutch Academic Heritage Foundation, which at the start of the project represented the five classic Dutch universities: Utrecht and Groningen, being represented by their university museums, and Leiden, Delft and Amsterdam by their university libraries. Last year four other universities joined the SAE – Maastricht, Eindhoven, Wageningen and the Free University of Amsterdam. These four institutions are on the verge of uploading their academic heritage into the recently built portal. The Digital Production Centre (DPC) of the University Library of Amsterdam built the portal and is responsible for the technical infrastructure, the tools developed in the project and the hosting of the content. DPC uses open, international standards and open source software.

The portal was designed to serve the interested general public, the researcher, as well as the collection manager himself. During the project we concentrated on developing the project as a whole. Delivering content or creating a website with the newest features, was not our main priority but they are among some of our future challenges. The project has demonstrated that museums and libraries have different types of expertise to bring to the process: libraries tend to have more hands-on experience with information technology, with for instance applying international standards and using controlled vocabulary,

whereas museums are experts in presenting and preserving their material. Both sectors have benefitted from the interaction.

Object metadata

UNICUM can be considered a metadata project, and especially the object metadata of all the partner institutions formed a major challenge. During the project their metadata was converted to CDWA Lite, the data structure standard of the Getty Museum, which DPC chose as the portal's format.[1] Now the project has been completed the partners are expected to work according to standard mappings, which have been created for this purpose.[2] They will offer their metadata to the UNICUM portal in such a way that these mappings can be used. The partner institution is responsible for its own metadata within the portal, not the DPC.

Standardization facilitates mapping to the aggregation but it was clear that many partners did not describe their objects according to international standards. This hampers the (international) exchange of data. There were cases where a single institution did not use one particular method for describing their various collections, but where different collections were described in markedly different ways. The pitfalls of describing diverse materials are well known, and academic collections are usually varied. It must be emphasized that *consistency* is a key factor here. Even when items have been catalogued incorrectly, as long as this has been done consistently, the errors can sometimes be easily corrected. Two examples of mistakes we came across that resulted in data loss:

- At some point in time, an institution transferred its metadata to another database system. This was not done as carefully as necessary and all distinguished elements were placed in one or two fields in the new database. At the time, no one was aware of the consequences, and no back-up copies were kept. As a result, years of work were lost.
- When trying to obtain a dump of metadata from a university museum collection, we came across a file published on the Internet that contained the required metadata. The html file differed considerably from the file of metadata extracted from the database; it was much richer in data. It turned out that records were updated in the static html file, instead of in the source database.

To prevent such mistakes in the future, as well as to guide a light through the tricky field of cataloguing, the UNICUM partners were recommended to make use of a metadata content standard.[3] The project did not *impose* standards, but tried to convince the partners

---

1 A data structure standard is a formal guideline specifying the elements into which information is to be organized. By establishing a set of elements to be included, a data structure standard also excludes other types of information. EAD, CDWA and MARC formats are examples of data structure standards.

2 In the project standard mappings have been created from the data structures used at the partner institutions (MARC, AdLib) to the CDWA Lite data structure of the portal.

3 A data content standard is a set of formal rules that specify the content, order, and syntax of information to promote consistency. A content standard goes beyond identifying the general type of information and indicates how to select between different, equivalent representations of the information and the manner the information is to be structured. For example, a content standard for a field called 'creator' might indicate whether an individual's common or full name should be used and whether the name should be inverted. For example, Lewis Carroll might be entered as 'Dodgson, Charles Lutwidge'. *Anglo-American Cataloguing Rules* (AACR), *Cataloguing Cultural Objects* (CCO) and *Archives, Personal Papers, and Manuscripts* (APPM) are examples of content standards.

by illustrating the benefits. We advised CCO (Cataloguing Cultural Objects), the content standard for the cultural heritage community. The CCO guide to describing cultural works and their images is available online.[4] These guidelines are illustrated by clarifying examples and answer many questions about how to fill the defined fields in database records of cultural heritage institutions. The CCO content standard is based on a subset of categories of the CDWA data structure standard, the native DPC format for the UNICUM portal. A language field was added to this format by us, since language is not used as a distinguishing criterion in the museum world, whereas it is an essential prerequisite in the library and archive domains.

Thesauri

The project showed in a tangible way the advantage of consistent metadata described according to a content standard. The use of controlled vocabulary also proved beneficial when it was time to publish the metadata on the portal website, as the site can be searched by both word and attributed keywords. Whereas the two partner university libraries (Leiden and Amsterdam) make use of controlled vocabularies and (inter)national thesauri, most partner museums work with lists of keywords of their own design. Obviously, such lists are neither conductive to international data exchange, nor show up well in the future semantic web. Ideally these lists should be linked to (inter-)national thesauri. The facet searching option within the portal on the development server clearly illustrated the pitfalls of not using controlled vocabulary or of not using it correctly (figure 2).

The partner institutions, while not striving for perfection, considered it necessary that DPC corrected the keywords automatically via specified conversion scripts in order to present the data in a good enough way. Most of those conversion scripts cannot be used for new metadata batches because of metadata inconsistency within the institutions themselves. As such new metadata uploads will be quite labour intensive for the partner institutions.

In UNICUM we recommend to make use of controlled vocabulary to solve this problem. We advised the partner institutions to make use of at least the AAT (Art and Architecture Thesaurus) and the NBC (Nederlandse Basis Classificatie, or Dutch Basic Classification), which classifies according to academic discipline. Both thesauri contain Dutch keywords that are related to their English equivalents.

One of the conclusions of the project was that a common metadata cleaning and enrichment project would be of interest to all partner institutions.[5] The Dutch Academic Heritage Foundation SAE was advised to take up the initiative and to find money for such a project. In preparation to such a project, the legacy records at the involved heritage institutions should be carefully screened as to draw up the specifications. Various open source software is available on the internet to handle the process of metadata cleaning and enrichment.[6]

---

4 Murtha Baca, Patricia Harpring, Elisa Lanzi, Linda McRae, Ann Baird Whiteside (eds.), *Cataloging Cultural Objects: A Guide to Describing Cultural Works and Their Images* (Chicago 2006). Downloadable through http://cco.vrafoundation.org.

5 With metadata enrichment one can think of adding persistent identifiers, controlled vocabulary to subsets of metadata and information about the rights.

6 S. van Hooland, R. Verborgh, M. De Wilde, J. Hercher, E. Mannens, R. Van de Walle, 'Evaluating the success of vocabulary reconciliation for cultural heritage collections', *Journal of the American Society for Information Science* (forthcoming).

Object work type
- Tekening: Portret (19)
- Tekening: Plattegrond (3)
- Tekening: Architectuur (22)
- Tekening: en aquarel: Architectuur (1)
- Sculptuur: Portret (2)
- Schilderij: Portret (18)
- Prent: Topografie (1)
- Prent: Tekening en architectuur (1)
- Prent: Tekening (1)
- Prent: Portret (4)
- Prent: Plattegrond (8)
- Prent: interieur (2)
- Prent: Groepsportret en architectuur (2)
- Prent: Groepsportret (2)
- Prent: Bouwtekening (3)
- Prent: Architectuur, plattegrond (2)
- Prent: Architectuur en spotprent (1)
- Prent: Architectuur en naamlijst (1)
- Prent: Architectuur en groepsportret (6)
- Prent: Architectuur en genrestuk (1)
- Prent: Architectuur (335)

Fig. 2: Self-designed keywords, instead of controlled vocabulary, as delivered to the portal's repository.

Collections, items and stories

The portal has been built around different related components: collection descriptions (museum and library collections) with inventories (archives), item descriptions (museum, library and archival objects), stories and images.

1. Collections

Analogous to the Online Archive of California, we chose to use the archival EAD format (Encoded Archival Description) to describe both museum and library collections. Best Practice Guidelines for collection registration were formulated and all partners now describe their collections accordingly. The archival EAD standard is known for its multi-levelled complexity and xml-encoding. DPC has made a specially designed input module which simplifies the process and which can transform the delivered content into EAD xml right away. Institutions are supplied with a login to use the input form through the Internet. University libraries already working with EAD can supply the DPC with their own generated xml and do not have to make use of the input module.

With the input module and the guidelines at their disposal, the museums in the project are content with the EAD format, and they have already described some of their collections in this way. They have actually requested a similar format and Best Practice Guidelines for describing their items. That was beyond the scope of the project, but the content standard CCO, which we chose during the project, has met their needs in many ways already.

2. Inventories

Multi-layered archives can be very well described by means of the inventory levels offered by EAD. It is possible to link various archival series to a basic, upper-level description. An online input module for inventories has not yet been realized, and this will remain a challenge to be tackled in the near future. EAD xml for inventories can be delivered directly to DPC, which implies that libraries with EAD experience will be able to upload their
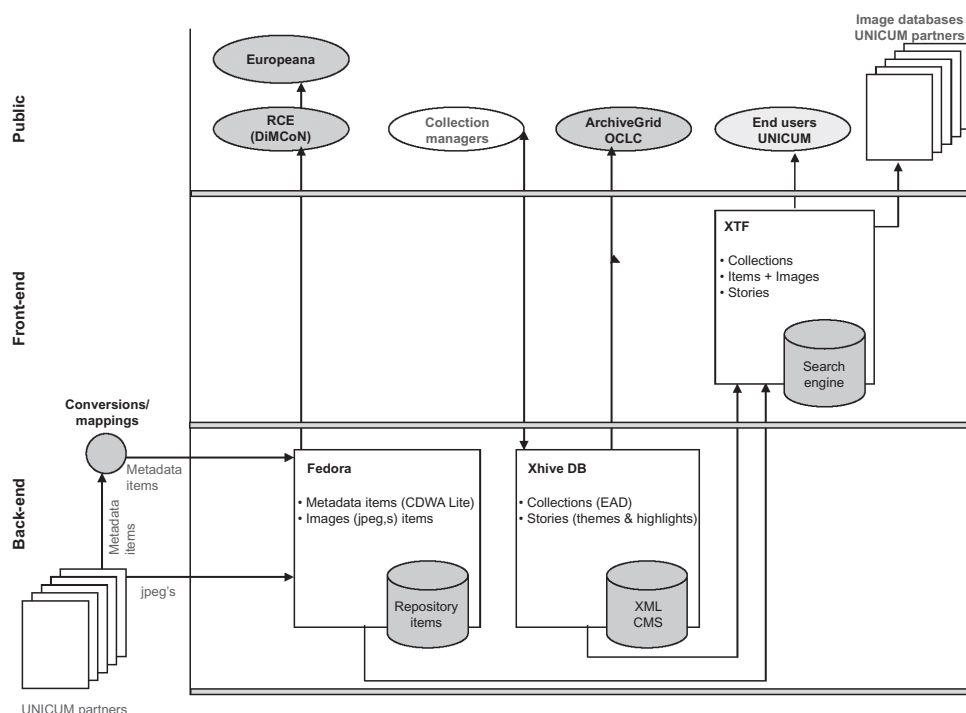
Fig. 3: UNICUM flow chart.

inventories to the portal. The design of the inventory part in the web portal and the routing from the DPC infrastructure should be adapted in a later stage, when the remaining issues around inventories have been resolved.

3. Items

As shown in figure 3, the partner institutions will recurrently supply DPC with metadata exports of their items to be uploaded into the UNICUM aggregation. The accompanying conversion and mapping procedure have been discussed in the previous section on object metadata. The idea of harvesting the partner's metadata by DPC still remains wishful thinking.[7] At the moment, it is not yet possible for either the museums to be harvested, or for DPC to harvest data from the partner institutions. We hope to tackle this issue in a future project.

4. Stories

To enliven the portal's website, the partners may publish stories about special themes or objects. These stories can be uploaded by means of an online input form designed by DPC, analogous to the EAD input module for collection descriptions. In this way universities can work together to create thematic profiles of their academic heritage.

---

7 Harvesting is an automated, regular process of collecting metadata descriptions from different sources to create useful aggregations of metadata and related services.

Fig. 4: URL referring to the image database of the owning institution.

5. Images

As for the images, large thumbnails of the items are presented to the user in the portal.[8] The heritage institution which possesses the material is responsible for the copyrights, not DPC. The thumbnails are linked to the image databases of the universities, which own the items. To see the complete picture instead of the thumbnail, the user is directed to the particular website of the owning institution. This also applies to composite objects, such as books, which will only be presented as a single thumbnail within the portal. Smaller museums often do not use web-based image databases and the portal www.academischecollecties.nl offers them the possibility to increase the visibility of their holdings in a relatively easy and instant way.

International exchange: ArchiveGrid and Europeana

The collection descriptions in the EAD format will be periodically sent to ArchiveGrid, the OCLC database of archival and collection descriptions.[9] For this purpose the abstracts of these descriptions have been translated into English, and the keywords also are submitted in English. Europeana will harvest the object metadata and thumbnails of the aggregated items in the portal. Europeana is an initiative by the European Commission to provide a

---

8 Partners have to supply their images to DPC in a 750 pixel wide jpg format.

9 In 1967, the boards of several academic institutions in the state of Ohio in the U.S. have founded the Ohio College Library Center (OCLC). Their aim was to develop an automated system that allows the academic libraries to share sources and reduce costs. At the moment OCLC serves more than 72,000 academic and other libraries in 170 countries worldwide.

single point of access to the digital content of Europe's cultural heritage institutions such as (audio-visual) archives, museums and libraries. Presently, Europeana is not yet able to process metadata at the collection level. The on-going European ApeNet project aims at contributing multi-level archival descriptions to Europeana.

Europeana does not do business with individual institutions or new portals, and only works with national aggregators. In the Netherlands, the Ministry of Education, Culture and Science distributed the cultural heritage sector roughly into four Europeana aggregators according to material types:

1. The 'Rijksdienst voor het Cultureel Erfgoed' (RCE), the Dutch National Cultural Heritage Service for museum material.
2. The 'Koninklijke Bibliotheek' (KB), the National Library of the Netherlands for text material.
3. The 'Nederlands Instituut voor Beeld en Geluid', the Netherlands Institute for Sound and Vision, for audio-visual material.
4. The 'Nationaal Archief', the National Archives of the Netherlands for archives.

The newly created portal www.academischecollecties.nl will deal with the RCE, since the majority of the content can be marked as museum material and, not in the least, because the RCE is already applying a well-functioning tool to convert the UNICUM metadata to the Europeana format. Delving, the software company that wrote the Europeana software, developed this tool. This, of course, is a major advantage, since Delving knows all the technical ins and outs of Europeana.

The 'Delving SIP-Creator', as the tool is called, is an open source conversion tool, which can be used and adjusted by anyone interested according, to his or her needs. If the input is meeting international standards, as it is in our case (the data structure of the portal being CDWA Lite), a sustainable mapping to the Europeana format can easily be created. And if it is not, the tool creates practical out-of-the-box conversions to Europeana.

The portal's content is harvested by the RCE and has been technically incorporated into DiMCoN (Digital Museum Collection Netherlands). The Dutch Academic Heritage Foundation (SAE) is on the verge of signing the contract with the RCE and thus with Europeana. Last year Europeana adopted a new contractual agreement in which the Creative Commons Zero (CC0) license was accepted for the metadata in Europeana (the images still fall under the CCby license). The UNICUM project partners discussed the conditions of the contract before they consented to the harvesting of their content from the UNICUM portal by the RCE (Europeana).

A few issues still remain unsolved, for instance the risk of uploading the same content of the same organisation to Europeana more than once. Europeana has established a working group to deal with this issue. Each individual institution also has to decide by which aggregator it wants to be harvested in the end, because, obviously, institutions will only want to invest for this one time.

Added value

What is the added value of the portal? One may wonder whether it would not be more practical if the five original UNICUM partners would deal directly with the RCE or the National Library to exchange their data. That might be the case if the only goal of the portal's partners would have been becoming part of Europeana. However, the merits of using

standards with regards to metadata were immediately clear within the UNICUM project. Every single project partner has profited in its own way from the knowledge gained in the project of building a portal. Eventually the project might lead to cooperative initiatives to strengthen the back-end infrastructure, especially where it concerns harvesting and metadata cleaning and enrichment.

By commissioning this portal, the Dutch Academic Heritage Foundation (SAE) can stimulate the presentation of academic heritage and use UNICUM to create a distinct profile for itself. Last year four more Dutch universities joined the Academic Heritage Foundation, resulting in the SAE now housing almost all Dutch universities. The SAE has become a stronghold, also in applying for grants or other subsidies.

The portal may also work for Dutch collection managers to fine-tune their collections. But, more importantly, the university museums that have joined UNICUM will not have to initiate the project by themselves, as we can do it collectively. And that is what we wanted from the start: co-operation to meet the challenges and opportunities of globalization.
Websites cited:

- UNICUM – www.academischecollecties.nl
- Online Archive of California – www.oac.cdlib.org
- Stichting Academisch Erfgoed – www.academischerfgoed.nl
- Leiden University Library – www.library.leiden.edu
- Universiteitsmuseum Utrecht – www.uu.nl/NL/universiteitsmuseum
- Groningen University Museum – www.rug.nl/museum
- Amsterdam University Library – http://cf.uba.uva.nl
- Delft Library – www.library.tudelft.nl
- Digital Production Centre (DPC) – http://uba.uva.nl/diensten/overige/dpc
- Open source software used by DPC (XTF) – http://xtf.cdlib.org
- EAD – www.loc.gov/ead
- CCO – http://cco.vrafoundation.org
- CDWA Lite – www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.html
- AAT – www.aat-ned.nl
- ArchiveGrid – http://archivegrid.org/web/index.jsp
- Europeana – www.europeana.eu/portal
- ApeNet – www.apenet.eu
- Dutch Europeana aggregators – http://digitalecollectie.nl
- RCE – www.cultureelerfgoed.nl
- The National Library of the Netherlands – www.kb.nl/index-en.html
- The Netherlands Institute for Sound and Vision – http://instituut.beeldengeluid.nl
- Delving – www.delving.eu
- SIP-Creator – http://vimeo.com/19291418
- DiMCoN – www.digitalecollectienederland.nl
- Freeyourmetadata – http://freeyourmetadata.org