

A portal to Dutch academic heritage: www.academischecollecties.nl

HENRIETTE REERINK

Abstract

The UNICUM development project, commissioned by the Dutch Academic Heritage Foundation (SAE), has been carried out by the five classic Dutch universities in 2010-12. 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 that 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 was 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 that the project is finished, the focus will shift to generating content.

UNICUM aimed to create structured and integrated access to academic heritage by:

- *using international standards (CCO and CDWA Lite) to stimulate exchange of metadata; examining the potential of EAD as an exchange standard for (non-archival) collections and for structuring related items;*
- *determining a common method and creating an online input module for registration at collection level using 'Best Practice Guidelines';*
- *developing an integrated format and online input module for thematic and highlight descriptions;*
- *choosing available and established thesauri which cover all aspects of the future content; focusing on the interrelation between items on one hand and collections as a whole on the other, and - vice versa; an interrelation that tells the story of, and gives meaning to cultural heritage; harvesting of the aggregation's content by Europeana.*

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 those pre-1850 collections which have grown historically, or actively been collected to meet the educational and research purposes of the universities. Examples are, for instance, historic microscopes, anatomical models and 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 California¹, 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

¹ Online Archive of California www.oac.cdlib.org/ (accessed July 31, 2012).

collection and item level. In the UNICUM project, we started off with five institutions. Our project was commissioned by the SAE, the Dutch Academic Heritage Foundation², 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 libraries³. Last year four more universities joined the SAE – Maastricht, Eindhoven, Wageningen and the Free University of Amsterdam. These four are on the verge of uploading their academic heritage into the recently built portal.



Fig. 1 - www.academischecollecties.nl

The Digital Production Centre (DPC)⁴ 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 software⁵.

The portal was designed to serve the interested general public, the researcher, and 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 were not our main priorities 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 (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.

² SAE (Stichting Academisch Erfgoed), Dutch Academic Heritage Foundation www.academischerfgoed.nl/ (accessed July 31, 2012).

³ Leiden University Library www.library.leiden.edu/ (accessed July 31, 2012).

Universiteitsmuseum Utrecht www.uu.nl/NL/universiteitsmuseum/Pages/default.aspx (accessed July 31, 2012).

Groningen University Museum www.rug.nl/museum/index (accessed July 31, 2012).

Amsterdam University Library cf.uba.uva.nl/nl/ (accessed July 31, 2012).

Delft Library www.library.tudelft.nl/ (accessed July 31, 2012).

⁴ Digital Production Centre (DPC) www.uba.uva.nl/digital_production_centre/home.cfm (accessed July 31, 2012).

⁵ Open source software used by DPC (XTF) xtf.cdlib.org/ (accessed July 31, 2012).

Object metadata

UNICUM can be considered a metadata project, and the object metadata of all the partner institutions was especially challenging. During the project, their metadata was converted to CDWA Lite⁶, the data structure standard⁷ of the Getty Museum, which DPC chose as the portal's format. Now that the project has been completed the partners are expected to work according to standard mappings⁸ which were created for this purpose. 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 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 easily be 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 institutions through the tricky field of cataloguing, the UNICUM partners were advised to use a metadata content standard⁹. The project did not impose standards, but tried to convince the partners by illustrating its benefits. We recommended CCO (Cataloguing Cultural Objects)¹⁰, the content standard for the cultural heritage community. The CCO guide to describing cultural works and their images is available online¹¹. These guidelines are illustrated by examples and answer many questions about how to fill in 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. We added a

⁶ CDWA Lite www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.html (accessed July 31, 2012).

⁷ 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.

⁸ 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.

⁹ 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.

¹⁰ CCO cco.vrafoundation.org/ (accessed July 31, 2012).

¹¹ cco.vrafoundation.org/index.php/about/news_events_entry/258 (accessed July 31, 2012).

language field to this format, 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 illustrated 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 in the portal website, as the site can be searched by both word and attributed keywords. Whereas the two partner university libraries (Leiden and Amsterdam) use controlled vocabularies and (inter)national thesauri, most partner museums work with lists of keywords of their own design. Obviously, such lists are not conducive to international data exchange in these days of globalization.

objectWorkType

- 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)
- prent: allegorie (3)
- prent (3382)
- portret: tekening (1)
- portret (632)
- pentekening en architectuur (1)
- penning met judicium (1)
- penning met bijbeh. cassette (1)
- penning (8)
- groepsportret (2)
- gebeurtenis (38)
- fotomateriaal: fotoafdruk papier zwart wit (662)
- fotomateriaal: fotoafdruk papier kleur (1)
- fotomateriaal (439)
- fotoafdruk (439)
- dubbelportret (Ravetz. Prof. dr. J. R. (1965) (1)
- dubbelportret (4)

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

For example, at the moment 'portrait' (2507 hits) is a keyword in the portal, as well as 'portraits' (1146 hits). If all partner institutions had made use of controlled vocabulary, these two keywords would have been combined. This example is not directly fundamental while searching the portal. But it does mean that the facet searching option within the portal loses its function, as fig. 2 illustrates.

In UNICUM we recommend using controlled vocabulary to fix this problem. We advised the partner institutions to use at least the AAT (Art and Architecture Thesaurus)¹² and the NBC (Dutch Basic Classification)¹³ 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 data cleaning and enrichment project for the metadata would be of interest to all partner institutions. The Dutch Academic Heritage Foundation (SAE) was advised to take the initiative and find money for such a project. Before such a project is launched, the legacy records at the involved heritage institutions should be carefully screened in order to draw up the specifications for the data cleaning and enrichment. Various open source software is available on the Internet to handle the process of metadata cleaning and enrichment.

Collections, items and stories

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

¹² AAT www.aat-ned.nl/ (accessed July 31, 2012).

¹³ NBC www.kb.nl/vak/basis/bc04.pdf (accessed July 31, 2012).

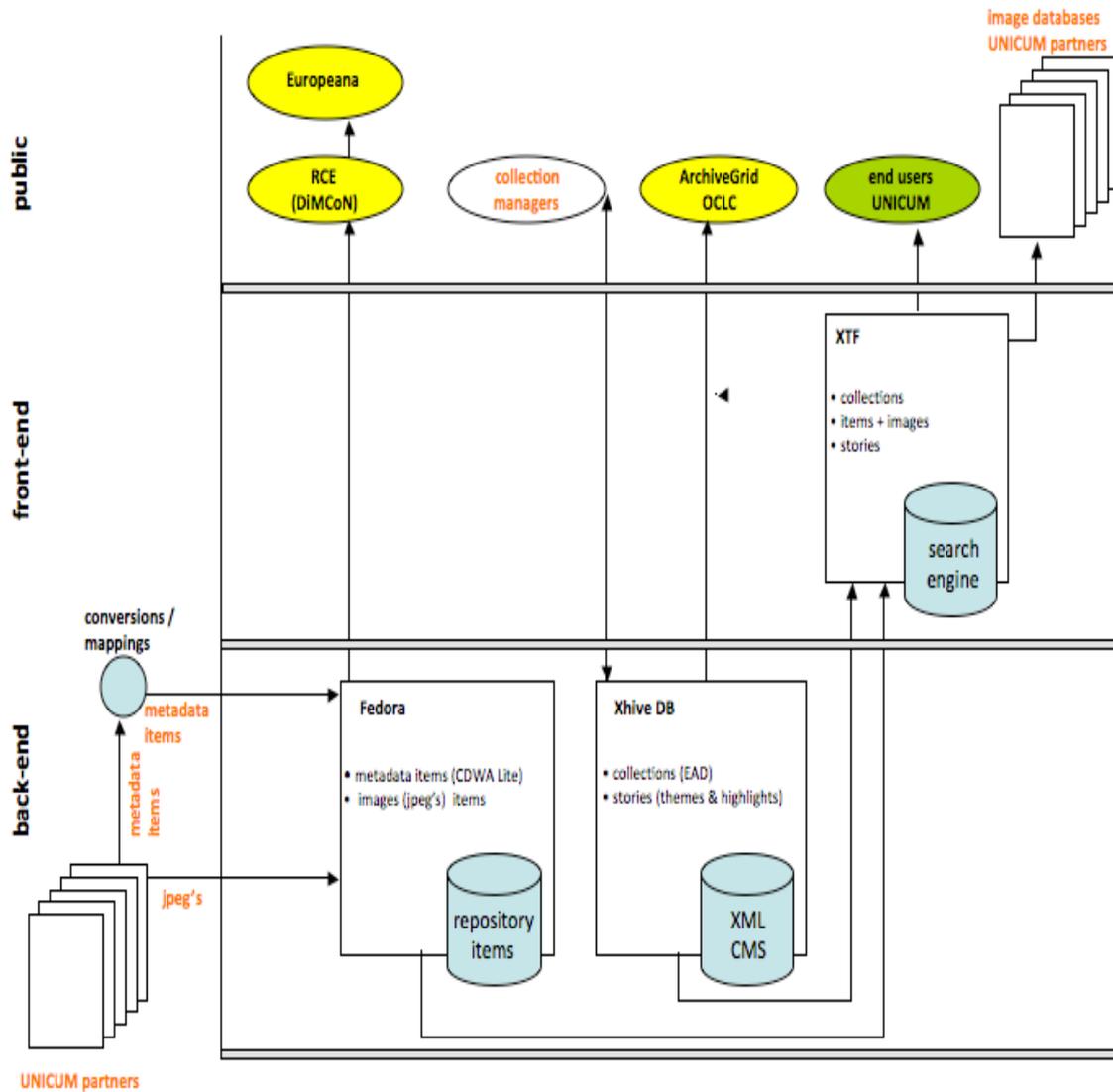


Fig. 3 - UNICUM flow chart

Collections

Like the Online Archive of California, we chose to use the archival EAD¹⁴ format 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 module¹⁵ 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 use the input module.

The museums in the project are content with the EAD format. They have already described some of their collections in the EAD input module following the guidelines. They have actually requested a similar format and Best Practice Guidelines for describing their items. That was beyond the scope of

¹⁴ EAD www.loc.gov/ead/ (accessed July 31, 2012).

¹⁵ with Windows X-Forms.

the project, but the content standard CCO, which we chose during the project, has met their needs in many ways already.

Inventories

Multi-layered archives can be described very well 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 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.

Items

As shown in Fig. 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 harvesting¹⁶ the partner's metadata by DPC still remains wishful thinking. 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.

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.

Images

Previews of the items are presented to the user in the portal¹⁷. 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 be presented only 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 sent periodically to ArchiveGrid¹⁸, the OCLC database of archival / collection descriptions. For this purpose the abstracts of these descriptions have been translated into English, and the keywords are also submitted in English.

The object metadata and thumbnails of the aggregated items in the portal will be harvested by Europeana¹⁹. Europeana is an initiative by the European Commission to provide a 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

¹⁶ Harvesting is an automated, regular process of collecting metadata descriptions from different sources to create useful aggregations of metadata and related services.

¹⁷ Partners have to supply their images to DPC in a 750 pixel wide jpg format.

¹⁸ ArchiveGrid archivegrid.org/web/index.jsp (accessed July 31, 2012).

¹⁹ Europeana www.europeana.eu/portal/ (accessed July 31, 2012).

level. The European ApeNet²⁰ project aims to contribute multi-level archival descriptions to Europeana.

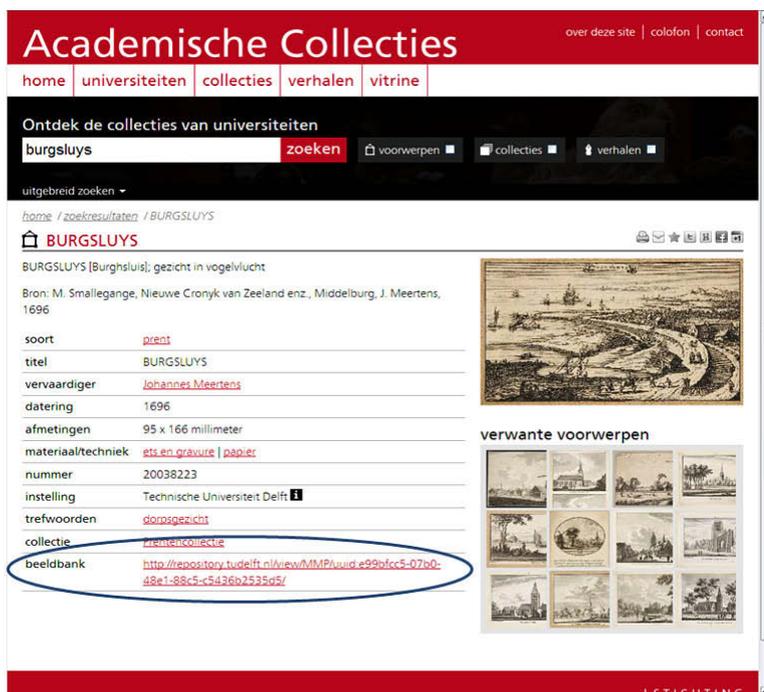


Fig. 4 - www.academischecollecties.nl

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 types²¹:

1. RCE²², the Dutch National Cultural Heritage Service for museum material
2. The National Library of the Netherlands²³ for text material
3. The Netherlands Institute for Sound and Vision²⁴ for audio visual material
4. The National Archives of the Netherlands²⁵ 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, more importantly, because the RCE is already applying a well-functioning tool to convert the UNICUM metadata to the Europeana format. This tool was developed by Delving²⁶, the software company which wrote the Europeana software. This, of course, is a major advantage, since Delving knows all the technical ins and outs of Europeana.

The Delving SIP-Creator²⁷, as the tool is called, is an open source conversion tool which can be used and adjusted by anyone according to his or her needs. If the input follows international standards, as it is in our case (the data structure of the portal being CDWA Lite), a sustainable mapping to the

²⁰ ApeNet www.apenet.eu/ (accessed July 31, 2012).

²¹ Dutch Europeana aggregators digitalecollectie.nl/ (accessed July 31, 2012).

²² RCE www.cultureelerfgoed.nl/ (accessed July 31, 2012).

²³ The National Library of the Netherlands www.kb.nl/index-en.html (accessed July 31, 2012).

²⁴ The Netherlands Institute for Sound and Vision instituut.beeldengeluid.nl/index.aspx?ChapterID=8532 (accessed July 31, 2012).

²⁵ National Archives of the Netherlands en.nationaalarchief.nl/ (accessed July 31, 2012).

²⁶ Delving www.delving.eu/ (accessed July 31, 2012).

²⁷ SIP-Creator vimeo.com/19291418 (accessed July 31, 2012).

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 Netherlands²⁸. The Dutch Academic Heritage Foundation (SAE) is on the verge of signing the contract with the RCE and therefore 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, there is a risk that the same content of the same institution will be uploaded 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 have been more practical for the five original UNICUM partners to deal directly with the RCE or the National Library to exchange their data. That might have been the case if the only goal of the portal's partners was to become 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.

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, so that the SAE now houses almost all Dutch universities. The SAE has become a stronghold, also in applying for grants or other subsidies.

The portal may also help for Dutch collection managers to fine-tune their collections. 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.

Acknowledgements

With thanks to Klara Ashley Bonder for editing the text.

Contact

Drs Henriette A. Reerink

Projectmanager Electronic Services

Address: University Library Amsterdam, Netherlands, P.O. Box 19185, 1000 GD Amsterdam,
The Netherlands

E-mail: [henriette.reerink\(at\)uva.nl](mailto:henriette.reerink@uva.nl)

²⁸ DiMCoN www.digitalecollectienederland.nl/ (accessed July 31, 2012).