The vocabulary work of RKD and the National Strategy Digital Heritage in the Netherlands

Reem Weda
Information Retrieval, RKD-Netherlands Institute of Art History, M.A.

The RKD-Netherlands institute for Art History is involved in maintaining three vocabularies that are be used by the heritage field. Our controlled vocabularies are: RKDartists, Iconclass and the Dutch translation of the Art & Architecture Thesaurus® As a set these vocabularies we aim primarily for usage with art collections, where they function as authority files and standard thesauri for indexing and for end-user information retrieval. The AAT is being developed by the Getty Research Institute, but the RKD has been participating in the translation and development since 1996.

These systems have been developed over decades, sometimes from before the digital age, like Iconclass. The systems adapted to digital demands and are now entering the domain of Linked Open Data (LOD). Although the vocabularies are available as linked open data in various variants, there is still further development necessary from the side of the data publisher(s) but also from the side of the systems that could utilize the LOD.

In the full paper I will explain what our vocabulary work entails, how it's organized and how the datasets are being used at the moment. Furthermore I will present the plans of the Dutch National Strategy for a distributed network for cultural heritage data and what role the RKD and our vocabularies could(and should) play in that network.

About the RKD

The RKD is one of the world's leading documentation and research institutes. As a knowledge institute we work in collaboration with museums, heritage institutes, universities and private institutes. Our staff are involved in research, publications and organizing events. We manage unique archives, documentation and photographic material and the largest art historical library on Western art from the Late Middle Ages to present, with the focus on Netherlandish art. Our collections contain not only paintings, drawings and sculptures, but also cover monumental art, modern media and design. The collections are present in both digital and analogue form (the latter in our study rooms). Our mission is to provide knowledge and information on Netherlandish art in an international context to scholars and

members of the general public worldwide.

The RKD strives to actively involve the general public in its developments and ambitions by providing a valuable (digital) service, to reach a wide audience and to achieve digital sustainability. The RKD is developing and maintaining the following vocabularies to be used by the broader heritage field:

RKDartists&, a name authority file

This is our oldest database with biographical data for Dutch and non-Dutch artists from the Middle ages till present. Other persons that are of importance to art, like scholars, collectors, and traders are also added to the dataset. First reason for the RKD to develop this controlled list is that it functions as an index to the collection of art historical documentation (pictures, library, archives).

Another important reason was that here we could make persons relate to one another and to images and documents through a relational database. Since then a lot of useful information about persons related to the arts in the Netherlands where and is being recorded in RKDartists.

At present, the database contains records of approximately 250,000 biographical records. New records are added daily by RKD staff, but we are also exploring the possibilities of having trusted partners from outside the organization work on the data.

Open Search Artists is an RKD web service for software developers and others who want to interact with the data directly. So it can be used for collection management systems, apps or web services among other. Open Search Artists is a derivative set of biographical data from the extensive dataset RKDartists&. The displayed elements are made available under the Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF) standard. Each object (artists record) can be accessed through a URI (Uniform Research Identifier). Also there is a deep link in the data to the full record in RKDexplore.

This data set is made available under ODbL-BY (Open Database Licence – Attribution). Please note that the license concerns the descriptive data. Images can be retrieved in low resolution with watermark, but these are not yet available as open content.

Iconclass, a art historical subject classification system

Iconclass is a multilingual classification system for cultural content. It is a hierarchically ordered collection of definitions of objects, people, events and abstract ideas that serve as the subject of an image. Art historians, researchers and curators use it to describe, classify and examine the subject of images represented in various media such as paintings, drawings and photographs.

The Iconclass structure was largely developed by Henri van de Waal (1910-1972), a professor of Art History at the University of Leiden (photo). In 1950, the RKD already began its collaboration with Van de Waal on the development of Iconclass, with the publication of mounted and annotated photographs of Dutch works of art, known as DIAL (Decimal Index of the Art of the Low Countries). Published 1973 and 1985 in book form by the Royal Netherlands Academy of Arts and Sciences (KNAW), of which Van de Waal was a member. The publication was followed by the development of several computerized editions of Iconclass by the University of Utrecht in the years 1990 to 2001. Published online in 2004 and updated in 2009 in partnership with RKD. RKD took over the role as maintaining institute from the Academy of Arts and Sciences.

To sum up a the key elements of Iconclass: It's a classification system for Iconographical topics with 28,000 hierarchically ordered definitions. It consists of Alfa-numerical codes with text correlates in several languages. It's hierarchically ordered so it works like a thesaurus. It has 14.000 keywords linked to the subjects, that can help guide the user to the subject he/she is looking for. It's multi-lingual, and now available in five languages. English, German, French, Italian and Finnish. Other projects for Portugese, Spanish, Polish and Chinese being worked on. To have in Dutch is still a wish. These translations have mostly been conducted by contributors from other countries and have been not funded by a Dutch Institute. This shows the intrinsic value of the system.

Iconclass has been published as LOD in RDF and Json format and the concepts can be found online though URI's like http://iconclass.org/rkd/52D1/. You can query using the API or download the full set of raw data.

The Dutch translation of the Art & Architecture thesaurus, a generic heritage thesaurus

Without doubt, the best thesaurus available for the cultural heritage sector is the Getty Art & Architecture Thesaurus (AAT). Work started in the 1970s and it was first published in 1990. The scope is 'terminology' needed to catalog and retrieve information about the visual

arts and architecture. The content goes wider than the name indicates, it covers the broad domain of cultural heritage. It is compliant with the ISO standards for thesauri, and it is often used as an example when thesaurus technology is discussed, also outside of the cultural heritage sector.

The AAT-Ned or Dutch AAT is a translation of the Getty AAT and in an way extended version of it. Because AAT-Ned also hosts several thousand Dutch candidate concepts that are used in the Dutch heritage field. An editorial board consisting of Flemish and Dutch professionals is working to validate and translate these concepts to have them accepted by the Getty.

RKD handles translation of Getty AAT. We have translated 42.0000 of concepts in the Getty AAT. This is regarded as a full translation. We do our work in a different thesaurus database than Getty. The software used is Xtree, and is shared with the German translation project of AAT. We are working on a structural and regular way of exchanging data with Getty Vocabularies.

The Getty Research Institute is publishing their vocabularies as LOD and is part of the Getty's ongoing effort to make their knowledge resources freely available to all. The AAT, TGN, and ULAN are now available as LOD. They are published under the Open Data Commons Attribution License (ODC-By) 1.0. The Dutch translation of the AAT is fully integrated within this publication.

A distributed network of heritage information

In 2015 the Dutch Cultural Heritage sector started a joint effort to improve the usability of its cultural heritage collection data. The challenge is to develop a digital heritage infrastructure that overcomes the necessity of aggregating and post-processing data. Instead they aim to realize a true distributed network of digital heritage information. With this new infrastructure they expect to improve the usability of the collection data maintained by the heritage institutions. Implementing Linked Data principles in the collection registration systems is one of the central building blocks of this approach. Part of the strategy is to urge the maintainers of the collections to align their data with formal Linked Data resources, like thesauri (persons, places, time periods, concepts) and to publish data as Linked Open Data.

The Dutch Digital Heritage Network (NDE) is a national program aimed at increasing the social value of the collections maintained by the libraries, archives and museums in the

Netherlands. The partners in the NDE network are the Ministry of Culture, the National Library, the National Archives, the Institute for Sound and Vision, the Cultural Heritage Agency and a number of Research Institutes for Dutch Culture and History, among which the RKD

These partners will formalize their co-operation through the installment of a new organization that will be responsible for realizing a joint strategy program for the Dutch cultural heritage network. The goal is a distributed network build by the institutes and their stakeholders (including commercial parties), each contributing from their own perspective. The program consists of three layers with a functional division between the management of data collections ('sustainability'), facilities for connecting that data ('usability'), and applications for presentation and use of the data ('visibility').

The work at the usability layer is focused on the development of a lightweight cross-domain infrastructure that is build on a distributed architecture. The core functionality consists of a network of terms that references all common definitions for persons, places, time periods and concepts. These terms are made accessible through a SKOS API that collection registration systems can implement in order to search for relevant terms when annotating their cultural heritage objects. As a result the URIs of the terms will be added to the object descriptions. The NDE program works on getting all relevant terminology sources available as Linked Data and provides facilities for term alignment and even support for building new thesauri. Several tools for this work ([CultuurLink, PoolParty](#)) are being provided by the NDE network.

Having cultural heritage institutions publish their data as Linked Open Data with references to established definitions for persons, places, time periods and concepts is one part of the challenge. The other part is to provide means for browsing in a cross-domain, user-centric fashion. Based on possible relevant URIs identified in the user query it should be possible to browse the available Linked Data in the cultural heritage network.

The downsides of existing practices have led to the decision to build a (preferably distributed) registry that records the back links for all, from discovery perspective relevant, terms used in the Digital Heritage Network. With this new approach they hope to move away from a traditional repository-centric approach to a more web-centric approach. Optimizing the usability of resources in their source environment is the main starting point. A Proof-of-

Concept is being developed for the distributed network of digital heritage information and they hope to demonstrate the first results before the end of 2017.

Conclusion

This is broadly the state of affairs that the RKD operating in. Although we have a solid base to build on, the developments are going very fast. Besides these technological developments, the content of the vocabularies also need to be consistently worked on. At the moment this is being done decentralized by people in and outside the organization and without much coordination. This is a situation that needs to be addressed by RKD. Especially when more users in the Netherlands and beyond will make use of the linked open data versions of these vocabularies.