Bringing two LOD vocabularies together
Linking Iconclass keywords to AAT concepts

Abstract
In this paper I’m exploring the following question: would Iconclass as a system benefit from enrichment of it’s keywords by linking them to AAT concepts?

Both Iconclass and AAT are widely used multilingual systems to catalogue, describe and annotate works of art. The content and scope of AAT and Iconclass differ noticeably, but there is some overlap. Especially within the Iconclass keywords. I have done mapping of these keywords to AAT and present the results - positive and negative - in this paper.

Content and structure of Iconclass
Iconclass was developed by Henri van de Waal (1910-1972), Professor of Art History at the University of Leiden (photo). His ideas for a systematic overview of subjects, themes and motifs in Western art, which later became the Iconclass system, took shape in the early 1950s. The complete Iconclass system was finished in the years after 1972 by a large group of scholars and was published between 1973 and 1985 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.

The main divisions of the Iconclass system are represented by digits 0 to 9. Of these ten 'main divisions', the numbers 1 to 5 are 'general' topics, designed to comprise all the principal aspects of what can be represented. Divisions 6 through 9 accommodate 'special' topics, coherent subject matter of a narrative nature, with an emphasis on the Bible (7) and Classical Mythology (9). A tenth division, represented by the number 0, was added in 1996 at the request of Iconclass users, to accommodate abstract art.

0 Abstract, Non-representational Art
1 Religion and Magic
2 Nature
3 Human being, Man in general
4 Society, Civilization, Culture
5 Abstract Ideas and Concepts
6 History
7 Bible
8 Literature
9 Classical Mythology and Ancient History
Within each division of Iconclass, definitions are organized according to a logic of increasing specificity. Within each division of Iconclass, definitions are organized according to a logic of increasing specificity. A main division is divided further into a maximum of nine subdivisions by adding a second digit to the right of the first one. The third level of specificity is attained by adding a letter in upper case. From the letter(s) onward, all subsequent descents in the hierarchy take place by extending the notation to the right with more digits.

**Hierarchical path**

The following example from division 7 Bible shows the hierarchical principle: all subsequent descents in the hierarchy take place by extending the notation to the right with more digits:

7 Bible  
71 Old Testament  
71H story of David  
71H7 David and Bathsheba (2 Samuel 11-12)  
71H71 David, from the roof (or balcony) of his palace, sees Bathsheba bathing  
71H713 Bathsheba receives a letter from David  
71H7131 Bathsheba (alone) with David's letter

Bathsheba with letter of King David, by Rembrandt in 1654, Paris, Musée du Louvre

6 Results in RKD database on search for iconclass code ‘71H7131 Bathsheba (alone) with David’s letter’. When searching on ‘71H7*’ to look for all subjects relating to David and Bathsheba it retrieves 104 results from the RKD database making use of the hierarchical principle.
Keywords
A standard entry in the Iconclass system consists of a notation and its textual correlate. The Iconclass system offers the user additional features to increase the accuracy of meaning of a notation in a more or less systematic way. These features are: bracketed text, keys, doubling of the letter and structural digits.

I will not elaborate further on the other elements here, because they have no correlation to the question I have. Instead I want to focus on the Iconclass ‘keywords’.

Keywords are added to the notations as separate elements and they aid the information retrieval from the iconclass system. They are inspired on the textual correlate, but they mostly reflect the hierarchical relations that exist between the notations, they are hereditary. By this we mean that a keyword that has been assigned to a certain notation, is a valid keyword for that notation and all notations on a lower hierarchical level.

For instance: 92B32 love-affairs of Apollo ~ females

Has the keywords: Apollo · ancient history · classical antiquity · gods · heaven · history · love-affair · mythology

The keyword ‘mythology’ for example, which is present from the beginning of notation 9, is valid for the entire division 9. The keyword ‘Apollo’, placed at level 92B3 (story of) Apollo (Phoebus), is valid for all notations below that notation. Like with the notation for his ‘love-affairs’.

Keywords show are considerably larger degree of standardization than the words in textual correlates, they also follow certain conventions:

- with a few exception, the singular form of nouns is preferred.
- There are keywords that are verbal nouns, denoting activities and processes (e.g. burning, walking, singing, laughing).
- Modified letters (diacritics) in the textual correlates – é, ö, ü, ï, etcetera have been replaced by their unmodified equivalents when they are part of a keyword. ¹

All 14,000 keywords have been alphabetically indexed and can be used in the online Iconclass browser to locate the notation and its textual correlate to describe and/or index an image.²

- The Iconclass keywords are a powerful means of retrieving information from the Iconclass system, but they are not (yet) fully linked semantic elements within the LOD publication. Mapping and structuring these keywords would benefit the publication.

---

¹ This was done in the early 90-ties to enhance searching. Today this would not be necessary anymore. ² http://www.iconclass.org/
Content and structure of the AAT

The Art & Architecture Thesaurus (AAT) is part of the Getty Vocabularies, developed by the Getty Research Institute, Los Angeles. The current managing editor of the Getty vocabularies is Patricia Harpring. Work on the AAT began in the late 1970s in response to a need expressed by art libraries and art journal indexing services that were beginning to automate their cataloging and indexing procedures. Soon catalogers of museum objects and visual resource collections also expressed a need for similar controlled vocabulary to encourage consistency in cataloging and more efficient retrieval of information.

Visualization of the thesaural relationships of ‘rhytha’ in the AAT

The Getty vocabularies are intended to provide terminology and other information about the objects, artists, concepts, and places important to various disciplines that specialize in art, architecture, and material culture. The AAT contains generic terms; it contains no iconographic subjects and no proper names. That is, each concept is a case of many (a generic thing), not a case of one (a specific thing). For example, the generic term cathedral is in the AAT, but the specific proper name Chartres Cathedral is out of scope for the AAT (it would be included in one of the other vocabularies instead).

The multilingual AAT contains more than 42,000 generic concepts, organized in 8 facets, about material heritage, like art and architecture, but also techniques, materials and living creatures. Since 2014 AAT is published as LOD.

The AAT is a compiled resource; it is not comprehensive. The AAT grows through contributions. Among others the RKD-Netherlands Institute for art history, coordinates a full Dutch translation and proposes new concepts coming from the Dutch heritage field.

Scope and Structure

The AAT is a structured vocabulary containing terms and other information about concepts. Terms in AAT may be used to describe art, architecture, decorative arts, material culture, and archival materials. The target audience includes museums, libraries, visual resource collections, archives,
conservation projects, cataloging projects, and bibliographic projects. Terms for any concept may include the plural form of the term, singular form, natural order, inverted order, spelling variants, scientific and common forms, various forms of speech, and synonyms that have various etymological roots. Among these terms, one is flagged as the preferred term, or descriptor. There may be multiple descriptors reflecting usage in multiple languages.

The AAT is a thesaurus in compliance with ISO and NISO standards. The focus of each AAT record is a concept. In the database, each concept's record (also called a subject) is identified by a unique numeric ID. Linked to each concept record are terms, related concepts, a parent (that is, a position in the hierarchy), sources for the data, and notes. The temporal coverage of the AAT ranges from Antiquity to the present and the scope is global.

The AAT is a hierarchical database; its trees branch from a root called Top of the AAT hierarchies (Subject_ID: 300000000). There may be multiple broader contexts, making AAT polyhierarchical. In addition to the hierarchical relationships, the AAT has equivalence and associative relationships.

Part of the Styles and Periods facet from the AAT

The conceptual framework of facets and hierarchies in the AAT is designed to allow a general classification scheme for art and architecture. The framework is not subject-specific; for example, there is no defined portion of the AAT that is specific only for Renaissance painting. Terms to describe Renaissance paintings will be found in many locations in the AAT hierarchies.

A lot more can be said about the AAT content and structure, but for now I would refer the reader to the Getty Vocabularies documentation for more in depth information.

- AAT concepts can contain multiple equivalent terms that have equal authority. The terms connected to the concept all have warrant from authoritative sources. Users do not have to use the ‘preferred term’ the AAT chose, if needed they can use a equivalent.

3 http://www.getty.edu/research/tools/vocabularies/aat/about.html
**Similarities and differences**

Both systems are structured as thesauri. Both are available as linked open data, although in different formats. Both Iconclass and AAT are published as SKOS and SKOS-XL. Downloadable in several dataformats like RDF/SKOS, JSON. Getty vocabularies offers more formats. Also the Getty vocabularies data model is a lot more elaborate, containing much more elements than Iconclass.

The content and structure of the two vocabularies differ substantially. It goes too far to point out all these differences here, but in general you could say that the AAT is a more modern thesaurus based on ISO and NISO standards, and Iconclass was largely developed outside of that framework. This is not to say that the Iconclass structure is not sound, it’s just organized in a way that followed the iconographic practice of those day’s.

For AAT the concept is found through the unique ID number that is part of a URI. Examples of URI’s for AAT: *rhyta*

- Human-readable full record: http://vocab.getty.edu/page/aat/300198841
- Human-readable hierarchy view: http://vocab.getty.edu/hier/aat/300198841
- Semantic RDF concept (several dataformats): http://vocab.getty.edu/aat/300198841

In Iconclass the notation has meaning in itself, but also it doubles as a unique ID for the textual correlate that you could call a concept. For 52D2 Clarity, Precision the URI’s looks like this:

- Human readable record: http://iconclass.org/rkd/52D2/
- RDF concept in RDF or JSON: http://iconclass.org/52D2.rdf

It is also possible to link concepts in Iconclass and AAT. For instance: *47H6 textile fabric, cloth* with *cloth* (AAT: 300162391), but this Iconclass notation also maps with *textile materials* (AAT: 300231565), which is a broader term of ‘cloth’ in the AAT. Also the hierarchal position of the concept differs a lot.

- Conclusion: a concept in Iconclass is a somewhat different thing than in AAT. Iconclass annotations do not always describe one single topic. Linking on this level will only work for a relatively small number of instances because of described differences. However: connecting the iconclass keywords with AAT concepts will deliver more results.

**Mapping exercises**

At the moment iconclass keywords are not ‘concept based’. So they are not represented as unique elements with an ID, description or place in a hierarchy. Linked data demands that these element have a unique URI (Uniform Reference Identifier). The keywords for 47H6 textile fabric, cloth are: civilization · cloth · craft · culture · industry · occupations · society · textile industry

---

4 SKOS: Simple Knowledge Organization System. SKOS-XL is a adaptation with more elements.
5 http://www.getty.edu/research/tools/vocabularies/loa/at_semantic_representation.pdf Accessed on 2017-08-29
Here the mapping rate IC to AAT is high and looks like this.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Totals in use</th>
<th>unique</th>
<th>AAT matches</th>
<th>%Match per facet, rounded</th>
</tr>
</thead>
<tbody>
<tr>
<td>facet 0_1</td>
<td>27826</td>
<td>3417</td>
<td>1282</td>
<td>38%</td>
</tr>
<tr>
<td>facet 2_3</td>
<td>5844</td>
<td>2812</td>
<td>1213</td>
<td>43%</td>
</tr>
<tr>
<td>facet 4</td>
<td>9459</td>
<td>5299</td>
<td>2258</td>
<td>45%</td>
</tr>
<tr>
<td>facet 5_6_7_8</td>
<td>18120</td>
<td>5058</td>
<td>1208</td>
<td>24%</td>
</tr>
<tr>
<td>facet 9</td>
<td>20364</td>
<td>3223</td>
<td>877</td>
<td>27%</td>
</tr>
<tr>
<td><strong>average</strong></td>
<td><strong>81613</strong></td>
<td><strong>11843</strong></td>
<td><strong>3774</strong></td>
<td><strong>35,3%</strong></td>
</tr>
</tbody>
</table>

The results per facet would give a more clear view on mapping results, but that was not possible at this time. The result gives the impression that keywords used in facets 1 till 4 map better than the other facets. Keywords from facet 5 to 9 do not map very well, mostly because they are made up of many named entities that are out of scope for AAT.

---

NB: Textile industry is a so-called ‘compound term’, not present in the AAT. It’s possible to map to 2 concepts representing this (industry AND textile), or to a related one; for instance ‘textile centers’ (300387290).

**Reconciling iconclass keywords through Refine**

For this test I collected all the keywords from the Iconclass data and put them in a spreadsheet. The keywords that are used in Iconclass are grouped into 5 directories, combining two to four facets in one directory. For each group I first deleted all the duplicates. Then I uploaded the remaining list into Google Refine to reconcile with the AAT. By reconciling I mean linking/or mapping the data to structured databases online, in this case the AAT.

The results of the you see below.

6 A tool developed by Google, now open source, for working with messy data: cleaning it; transforming it from one formation into another; and extending it with web services and external data. [http://openrefine.org/index.html](http://openrefine.org/index.html)
The average percentage of the facets is 35.3%. This differs a bit from the percentage that mapped when I reconciled all the unique keywords (11843) at once. Here the result was around 32%.

**Linking with CultuurLINK**
CultuurLINK is developed in the Netherlands for cultural heritage institutions to help them link their vocabularies, such as thesauri and term lists, with the Dutch cultural heritage Hub. With CultuurLINK you upload your vocabulary, select a target from the Hub and build your unique link strategy.²

CultuurLINK loads thesauri expressed in SKOS. Simple Knowledge Organization System (SKOS). The strategy I built yielded approximately the same results as with Refine. Around 3700 links for 11843 keywords. The difference with this method is that you see how and to what you are linking. This way it became clear that about 600 matches can be made with a ‘snowball’ method. This means that the string of a keyword matched with a string in the AAT keyword. But they don’t necessarily match 100%, so these need to be manually checked. Another feature of this tool is that you can save and reload your strategies for future use.

**What kind of concepts match?**
The kind of concepts that match are the ‘generic’ ones. See for what generic means the description of the AAT content in the paragraph above.

For instance the term ‘art’. This links to AAT concept ‘art (fine art)’. What information do we get by linking to the AAT concept? For starters there are the terms for art in several languages: English, French, Chinese (3 variants), Dutch, German, Italian, Spanish and Swedish. Then there is a ‘scope note’ which describes the concept in several languages (not all), and disambiguates it from related terms, like ‘the arts’ and ‘works of art’.

² [http://cultuurlink.beeldengeluid.nl/app/#/](http://cultuurlink.beeldengeluid.nl/app/#/)
It is possible that the Iconclass keyword does overlap other concepts in meaning. This needs to be investigated further. In theory the AAT hierarchical structure could also be used to structure the Keywords in Iconclass as a thesaurus.

Another match is on the term ‘battlefields’. This concept has several ‘alternate descriptors’, which means that these terms are synonyms that refer to the same concept. That is also the reason that the reconciliation was successful: the word ‘battle-field’ -look at the different spelling - is an English Use for in the AAT. Please note that although AAT always has a ‘preferred term’, this does not mean that this term is the only term you may use. The other terms should be full equivalents to each other.
Looking at the keywords it seems that with a more sophisticated mapping, or a manual interference, the percentage mapped keywords to AAT could be higher.

**Keywords that don’t match**

There is a considerable amount of keywords that don’t match with AAT. Sometimes that’s because there is a variant in spelling, but for a lot of instances it’s because the term is out of scope for the AAT. Names for people, mythical figures, gods or other entities together with named events and places are not in scope for AAT.

Names like: **Cleopatra, Arthur, Ophelia, Odin, Mars, Paris, Marseille.**

Also specific iconographical (biblical) subjects like ‘Man of Sorrows’, ‘hand of God’ and ‘Jacob’s ladder’ are out of scope for AAT.

And then there are the unusual activities, events or states of being, in Iconclass that are often compound terms, like: ‘**apparent death**’, ‘**love unrequited**’ and ‘**not too poor, not too rich**’ These are keywords that are probably not linkable to any vocabulary.

I discovered that the keywords for named entities and iconographical subjects often have a counterpart in a Wikidata-element. Wikidata is a free and open knowledge base that can be read and
edited by both humans and machines. For instance linking: 11H(LAURENCE) the martyr and deacon Laurence of Rome, with keyword ‘Laurence (St.)’ to Wikidata ‘Laurentius of Rome’ https://www.wikidata.org/wiki/Q17590 (note the name variant). Through this link there is access to additional name variants in 57 languages. As well as links to the article, sources and other authority files. Although very promising, further mapping of Iconclass to this knowledge base is out of scope for this paper.

Other issues to consider
One user of the Iconclass LOD pointed out that the keywords represented in the JSON data, where not always structured symmetrically between the language variants. For instance the data for the Roman godess "Victoria": www.iconclass.org/96A5(VICTORIA).json

The English keywords start with “idea”, “Roman god”, “Victoria”etc. But the French starts with “Victoria”, then “antiquité classique”, "concept", etc. They do not line up

This way, it seems difficult to use the Iconclass keywords in any consistent multilingual environment via its Linked Open Data interface, where one would need for any language interface the consistent contents.

Mapping the keywords to an external thesaurus like AAT, could perhaps help solving this problem. Because then the URI for each language variant of a concept would be the same. Unfortunately, the AAT does not have a full translation in all the languages that are available in Iconclass. If that was the case then we would in theory only need one AAT URI providing all language equivalents for one concept. for instance: the AAT concept ‘violins’ only has English, Dutch and Spanish translations, we therefore still miss the German, Italian and Finnish for the Iconclass keywords.

Conclusion
Connecting knowledge systems enhances the usability and value of these systems used within the cultural domain. But we need more developed user interfaces to let people benefit. The connection between Iconclass and AAT provides Iconclass users with ready to use AAT-concepts that can be added to object description besides the subject notation. Thus providing access to the full semantic content of a AAT record, with multilingual labels, equivalent, alternative terms and description of the concept. Doing this could provide a useful bridge between collections indexed with Iconclass to other collections using AAT concepts. If the mapping of keywords proves correct it could become a integral part of the iconclass LOD. As far as I know, this would be the first time integration between two distinct vocabularies was done in this way.

Reem Weda
The Hague, September 1, 2017

---

8 Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others. https://www.wikidata.org/wiki/Wikidata:Main_Page accessed on 2018-08-31