Coming to terms with concepts
From index cards to conceptual thesauri or the struggle to keep up

Erik Buelinckx
Royal Institute for Cultural Heritage (KIK-IRPA)
Department of Documentation: Art Historical Research & Inventory
Brussels, Belgium
erik.buelinckx@kikirpa.be

Abstract—When in 1989 the Royal Institute for Cultural Heritage (KIK-IRPA) started with the encoding of textual information about mainly Belgian art works and buildings, ranging from cathedrals over the fine and decorative arts to archaeological findings, based on over 1,000,000 original negatives, which are conserved in the institution and since many years in an ever changing process of digitisation, the fundaments for this enormous task consisted however of written, dactylographic and printed sets of index cards and long lists of names and themes. From the start the department of documentation had to manage database software to create, for a general and scientific public, external and internal (the institute has an interdisciplinary team of art historians, restorers, chemists and physicists), access to a bilingual environment, without realizing what the misconception of real 100% equivalence between words or terms in different languages could result in. The introduction of conceptualised thesauri seemed thus a great solution to take the next step to real multilingualism. Being from the beginning a partner in the translation of the Getty AAT into Dutch, their recent conceptualisation has put us before a choice. Mixing terms with concepts, difficulties in explaining this with words and terms, and trying to keep up with onterminologies while people are still translating word-couples back home, is certainly fascinating and even necessary, but in this paper I would like to use this real world situation to figure out how, or maybe if, we could ever come to terms with concepts.

Index Terms—terminology, thesaurus, ontology, concepts, cultural heritage, photographic archive.

INTRODUCTION

As a federal public institution the Royal Institute of Cultural Heritage (KIK-IRPA), located in Brussels, Belgium, has its official mission focused on the study, restoration and conservation of the Belgian cultural, mainly artistic, heritage. Three departments – Documentation, Laboratories and Restoration/Conservation – work together to achieve this goal. The Royal Institute for Cultural Heritage takes root in 1900 as the central unit. In these older times of paper and ink, the photographs to be glued on. Next to, above or under

1934 when the chief curator of the KMKG-RMAH, Jean Capat, appointed Paul Coremans as head of this documentation service to establish also a Physics and Chemistry Research Laboratory. With the start of World War II, and rather visionary, Stan Leurs, professor at the University of Ghent and general adviser for the conservation of monuments to the Restoration of the Country General Commission and Jozeƒ Muls, general director of the Fine Arts, instigated the photographic inventory of the Belgian cultural heritage, because of fear of wartime destruction. This resulted in 160,000 negatives, which were especially helpful to reconstruct damaged works, and they constitute in some cases the only visual traces left of destroyed works of art. After the war, in 1948, a new and independent institution was created out of the merging of the Central Iconographic Archives of National Art and the Central Laboratory of Belgian Museums (ACL). In 1957, by royal decree, it became the Royal Institute for Cultural Heritage, which in 1962 moved into its new building, specifically constructed in view of the multidisciplinary work to be carried out by its three departments.

As the focus of this article is on terminology used in describing and enhancing the different kinds of documentation generated in the three departments of the institute, it should be mentioned that the works of art that were studied, documented, analysed and/or restored at KIK-IRPA count into the ten-thousands, and that the documentation created out of this workload is open not only for scientific research but to a broad general public as well. This means, these days, a necessarily online platform, including the files on masterpieces by Van Eyck, Breughel, Rubens, and many more. Some decisions of the past, most of the current practices and the roadmap for the future have the openness of all this information in common, where and whenever possible.

PAPER AND INK, INDEX CARDS, LISTS AND CARTONS

Long before computers, the human need to classify, label and (thus) control things, was already put into practice in libraries, museums and other cultural heritage institutions. In KIK-IRPA, being responsible for creating a photographic inventory, of course the photograph was the central unit. In these older times of paper and ink KIK-IRPA used more or less standard size cartons for all the photographs to be glued on. Next to, above or under
and on the verso side, textual information was written. Since for many artworks more than one photograph existed, the same information was copied over and over again, with some slight differences as inventory number of the negative, year photo taken, photographer and some other more technical details (Ill. 1 to Ill. 7). A multitude of index cards was created to give the user the possibility to search on artist, location, iconographic theme, and more, and lists and repertoires were kept in written and dactylographic ways. One of the more amusing anecdotes or horrific stories regarding these times of ink and paper was the practice of scratching away textual information on the cartons to replace with new information – each artist historian got a special scratch tool to accomplish this work – in complete disregard of an important aspect of information management. Similar problems were also encountered during the different phases in the digitisation of this information and based on the set-up of database systems where, as a standard practice, information tended to be too easily overwritten, without trace of what was put into fields in the first place.

AUTOMATISATION

At the end of the 1980s an important decision was made. According to the then available information, the collection consisted of 800,000 photographs of 250,000 objects, more or less 3,500,000 referring index cards (Ill. 8 to Ill. 11) and 6,000 restoration and conservation files. Automatisation – as it was then called – of all this information would be done in such way to offer the most efficient services to research in art history and archaeology. Contact was made with the Mardoc foundation in the Netherlands, who had already implemented similar systems. Thanks to the first initiatives of CIDOC from 1979 on, structuring of data about museum objects was already being practiced, so that was something that could be built on.

This resulted in a system consisting of two main databases, OBJECT and PHOTO, linked to each other through the KIK-IRPA object number. OBJECT consisted of several subsets, some regarding the geographical location of an object (Flanders, Brussels, Wallonia, outside Belgium, unknown or undisclosed) and one to describe the more than 4,500 museums and churches where the objects were located. At last there was a specific dataset for photographs of objects not linked to an institution, such as landscapes, person and events. The second database PHOTO was used for the description of the photographs. Around and in-between these two databases a set of bilingual Dutch-French thesauri and authority-lists was created. Already at that time also rules for chronological input were given, necessary because of the wide range of objects, from prehistorical artefacts to contemporary art. A system of a begin date and an end date was chosen.

A BILINGUAL ISLAND IN A MONO-/MULTILINGUAL OCEAN

KIK-IRPA is, according to the in Belgian administrative and political landscape an officially French-Dutch bilingual institution. Care has to be taken to keep a perfect equilibrium between these languages, as well regarding staff as for the information collected, archived, published and presented. What can be seen as an advantage, this need to cater from the early beginning for two languages long before most of the other institutions had to, created also some problems, but some of them only visible much later. In the database system implemented in the institute, bilingualism was resolved by the use of equivalence fields. Each term should have its equivalent in the other language. There was no possibility to indicate different degrees of equivalence, or equivalence from one to more terms. For the sake of clarity this may be just fine, but as far as being useful for a more correct interpretation of meaning and language one can see that it created a much too rigid framework. A few – some could say weird – decisions in the past had a long-lasting influence. For instance the dropping of the indication of the language of a thesaurus record, during a migration from one platform (UNIX) to another (Windows) of the database software made the recent migration to a more performing system rather difficult without a lot of preparatory work. Or the fact that while terms can be forced (which is normal since many of the encoders are not specialists) not enough attention was given to the cleaning, controlling, updating of these terms in the thesaurus on a regular base. This leaves the KIK-IRPA thesaurus as a very rich, but also incomplete, and sometimes illogical terminology. And a last decision that was made, one very difficult to accept but luckily only for one language, had to do with the lack of understanding of how a controlled hierarchical vocabulary system could be used in the now emerging web environment.

An example:

<table>
<thead>
<tr>
<th>(fr)</th>
<th>(nl)</th>
<th>(nl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>église abbatiale</td>
<td>kerk[abdij-]</td>
<td>abdijkerk</td>
</tr>
<tr>
<td>église basilicale</td>
<td>kerk[basilica-]</td>
<td>basilicakerk</td>
</tr>
<tr>
<td>église de pèlerinage</td>
<td>kerk[bedevaart-]</td>
<td>bedevaartkerk</td>
</tr>
<tr>
<td>église de béguaing</td>
<td>kerk[begijnhof-]</td>
<td>begijnhofkerk</td>
</tr>
</tbody>
</table>

The idea was to offer the public a view on the thesaurus so one could choose a good term to perform a search. This view was a flat alphabetical list. To recreate the logical alphabetical view existing in French, due to specifics of that language, an artificial mash up of the Dutch terms was carried out to get a same flat alphabetical view of terms. The reason is that in Dutch many words are made of combinations of two or more words, and often the second or last word is the denominator. Since a couple of years our search environment offers a hierarchical walkthrough possibility for picking a thesaurus based search term. But we’re still in a system which works with equivalence based bilingualism. The process of cleaning this up is almost done, and this now offers us the possibil-
ity to get more involved into open collaborations with the outside world.

**TOO EARLY, TOO MUCH?**

Over the last years, in a two-step operation, we finally migrated to a new version of the content management system, Adlib Xplus, in which museums, library and archives are combined, but we had to adapt and stretch the standard software to make it fit with the needs of our photographic archive, library and restoration/conservation files management. It was a complicated process of many years because the different evolutions on how working with terminologies in the library and in the photographic archive resulted in almost two separate worldviews. This was reflected into differences in the authority files for creators and depicted persons, but even more in what in the library is called the ‘subject’ thesaurus, compared to the multifaceted thesaurus in use in the photo archive. So just by a simple putting these together one would create a really messed up thesaurus and authority files system. A great deal of the cleaning up until now, and for quite a time in the next months (or years) consists in making sure that hierarchies for terms which were up until now in different datasets, but would be merged together, follow the same rules. The disadvantage of early adopters is that when new and more interesting possibilities arise already too much work was done the “old way”.

As one can understand the biggest problem here is not only the way to do the work, but also finding efficient and professional ways to do the work when so many data and metadata exists. The “burden” of the past is reflected in current situation realizing that sometimes – often understandably – decisions were not taken because the sheer amount of data involved, or decision were taken without thorough understanding of the consequences. Certain resilience in not thinking too quickly that suddenly we know everything now has to guide our next steps, combined with a balanced combination of drive and restraint. But the world around us is moving much faster and technological developments promise interesting ways to offer new ways of using the open data of public institutions.

**FUZZY EXACTNESS**

Gladly the world is not a black and white place, and gradations are the norm, or at least should be. Realizing this makes one nevertheless wonder why we still have to cope too often with black-and-white thinking in our day-to-day work in cultural heritage. Our “bilingual world” thesaurus records are rigid pairs in which a translation was given as pure equivalence. So every record had its equivalent in a second one (the other language), and adding more languages was impossible. The new system has the possibility to get multiple languages in one record, but still as a kind of pure translation. The field refers still to a term, with its translations, but not to a concept. We are struggling to fit the multitude into this strict harness, realizing that considering this package of terms and their translations as concepts and there labels will not do the job, but also that an alternative is not at hand, for the time being. This of course will not help, neither even facilitate, the work we do in international multilingual projects. To name just one example we can look at the recently ended European project about art nouveau, Partage Plus. Parts of the Getty AAT were used as little central hubs and all partners were asked to give translations in their own language for the terms considered important to describe art nouveau objects. To my own understanding, the simplification of presenting concepts as terms that not only could be simply translated, but where the plural and singular representations were also accepted, laid down a framework where conceptual thinking could sneak into the “real” world of art historians and other specialists. A great part of the work was still manual-intellectual. It was considered a great improvement, because the lack of decent multilingualism in Europeana-projects, but it showed also that to advance in a quick and productive way conceptual thinking of thesauri had to hide itself a bit to implement a more or less simplified translation. This is not a critique on the advancement in science, but too often the real world situation doesn’t give us enough time and means to restart in a more efficient way, and forces us to use a patchwork method.

**KIK-IRPA is involved in a series of collaborations regarding terminologies:**

- AAT translation Dutch (since the 1990s, together with the RKD and Flemish museums and cultural institutions)
- Optimisation of French parts of the KIK-IRPA thesaurus (with the Walloon Region, since 2011)
- MULTITA (a Belgian Science Policy funded project between KIK-IRPA and KMKG-MRAH, with other players in Belgium, France and Holland to create a subset of a common multilingual thesaurus and to study methodologies to do so, 2012-2014)
- European projects Partage Plus (2012-2014) and AthenaPlus (2013-2015). For this last project we hope to finally see a working terminology management platform.

New opportunities will be grasped if possible, and a first one is the picking up of the French translation of the AAT, either as a collaborative project with Belgian Science Policy funding, or as an individual initiative of KIK-IRPA (starting end of 2014). The main problem, as always is means and personnel, but also the way to fit in all these initiatives into a more efficient multilingual “end product” to be used in KIK-IRPA and to put and maintain it into the open. Our experience in what at first was a one-on-one translation in the early AAT Dutch project evolved into working with concepts and terms.
THE GHENT ALTARPICE(S) AS AN EXAMPLE

Because of the intimate relation between KIK-IRPA as an institution and the Ghent Altarpiece by the Van Eyck brothers as a masterpiece of Flemish art, this work will be used for a few examples (Ill. 12). The advantage is that through its fame many institutions worldwide have information about this masterpiece, and it also the subject of a yearlong collaboration between KIK-IRPA and the Getty Institute. And of course it gives the possibility to show some magnificent images (Ill. 22 to Ill. 23).

What exactly is The Ghent Altarpiece, which in Dutch is simply called Het Lam Gods, or The Lamb of God, and in French L’Agneau mystique or The Mystic Lamb? In a recent discussion it was mentioned the work should in fact be called The Ghent Altarpieces, in plural, because it is not one piece but a combination of objects. Also this is exemplary for the discussion about terminology, where on one side we could easily create a seemingly logical descriptive system, but on the other side all human emotional and intellectual paradigms ask for their right spot.

In KIK-IRPA this object was “labelled” altaarveelluik (nl) and polyptique d’autel (fr) (Ill. 13). In our database not many other objects are identified with the same term. Most of the similar objects are identified as polyptichs, retables (altar appendages), and so on. This is a typical case where the use of a concept/term and how to put them in a conceptual hierarchical relation leads to endless discussions. The Ghent Altarpiece is combination of objects: a painting, an altarpiece, a (wooden) construction consisting of multiple parts. All of these have their labelling terms and all of these can be thought up as a concept. Although there can be a great hesitation in making up a description in which a set of combinations is used while a simpler labelling seems much more effective, one can’t throw away the corpus of scientific (and other) literature in which this combined term was used.

To understand the terms used we could check their place in the different hierarchies. When we look at the term altaarveelluik in our object names thesaurus we notice two broader terms [BT].

altaarveelluik
[BT] altaarschilderij
[BT] veelluik
[NT] altaardrieluik
[NT] altaarveelluik
[ET] polyptique d’autel (fr)

The broader term veelluik, which is part of the polyptych concept, can be traced down in the AAT. This means that this is the level where to make a connection between the thesaurus of KIK-IRPA and the LOD vocabulary of Getty AAT (Ill. 19 to Ill. 21).

The scope note states that “[t]hey are typically altarpieces, but may be objects with other functions as well”, so the practice in Dutch, where words can easily be glued together, gives an opportunity to refine, but also the complication that two different domains (a set of panels and a religious function) come together in one word.

Apart from our BALaT-website information on this work of art is also available through other online databases. At the RKD website, the chosen object category is veelluik, while at Marburg Bildindex multiple terms are used: Retabel, Tafelmanerie, Flügelaltar (Ill. 15 to Ill. 18).

As a more efficient service to the public we need to effectivity bring it all together – it could even be considered part of our official mission now that the world is a global village – keeping in mind that this is not an easy task, due to the number of terms in the KIK-IRPA thesauri and in the AAT, especially because so many terms exist in the one but not in the other and vice versa. Since the 1990s research has been done to find solutions for such kind of questions. Another problem, not mentioned yet, is that in KIK-IRPA, due to our financial difficult situation, the IT department has no means to aid fundamentally. While several projects helped us transforming parts of our thesauri in SKOS, at least in an experimental way, it is not clear how to get it all back in. This is the same problem existing in for instance Europeana projects and other collaborations where there is a flow of metadata to the project, eventually being enriched there, but none of it returning to home base. We also depend on the willingness of capabilities of software vendors to adapt their software for such developments.

ONTOTERMINOLOGY

As the need is still there to find other ways to tackle the problems we face, a look into Ontotermology, “i.e. a terminology whose conceptual system is a formal ontology, emphasizes the difference between the linguistic and conceptual dimensions of terminology while unifying them”, is another possibility. Is it workable to adapt an existing thesaurus of object names used in the field of cultural heritage into an ontotermology? “The variations in terminology and its use in different organizations mean that terminological ontologies are not useful for cultural heritage data integration.” For the moment this seems inconclusive, but nevertheless interesting enough to delve in.

CONCLUSION

In this article, which tried to address some important aspects of the documentary work of an art historian in a small cultural heritage institution with lots of data and metadata about over a million objects, on photographs, the idea was not to come up with a solution. What is needed is a kind of simplified framework to do several things at the same time:

- Preparing for the outside world
- Knowing the outside world
- Linking to the outside world
- Letting the outside world in

This simplified framework should be integrated in the daily workflow and it is clear that this controlled intrusion of the outside world will have a deep influence on the work inside the institution. Therefor it is also im-
important that such proposed schemes, theories or workflows regarding new paradigms in handling terminologies don’t lose sight of the fuzzy real world situations where they have to be implemented.

Closing Remark

A last remark concerns my worries about the recent evolutions regarding open data and is as such relevant to this article. It has to be clear that it is certainly not the openness of the data that is worrying, but the way openness is used as an excuse to diminish funding to cultural heritage and scientific institutions. When at one hand such public institutions are more and more obliged to find external funding just to survive and be able to do their work laid down in their official mission, they have on the other hand to compete with new commercial players who can just use the free and open cultural data right now, without the cost of collecting and maintaining it. My great fear is not the opening up of our data now, but the misconception that by opening up cultural industry will flourish, and thus more funding will flow back to cultural institutions; this a very naive and even dangerous way of thinking because it risks to jeopardise the creation of new cultural and scientific information. What we share now with the world didn’t come into existence just like that, but is the result of decades (centuries) of work funded by public and private bodies, and often by anonymous unselfish individuals. Will we be able to keep on doing this in the near and distant future, when cultural heritage information becomes just a commodity?
ILLUSTRATIONS

Ill. 1. Ghent Altarpiece photograph on carton (verso)

Ill. 2. Ghent Altarpiece photograph on carton (recto)

Ill. 3. Ghent Altarpiece photograph on carton

Ill. 4. Ghent Altarpiece photograph on carton

Ill. 5. Ghent Altarpiece photograph cartons
Erik Buelinckx (KIK-IRPA) Coming to terms with concepts CIDOC2014 (Dresden) 8/10 (20957)

Ill. 12. Adlib Xplus: object module

Ill. 13. Adlib Xplus: thesaurus module

Ill. 14. BALaT website (http://balat.kikirpa.be)

Ill. 15. RKD website

Ill. 16. RKD website

Ill. 17. Bildindex der Kunst und Architektur (Bildarchiv Foto Marburg)

Ill. 18. Bildindex der Kunst und Architektur (Bildarchiv Foto Marburg)
Ill. 19. AAT-Ned (Dutch translation project): polyptieken

Ill. 20. Getty AAT Thesaurus Online: altarpieces

Ill. 21. Getty AAT Thesaurus Online: polyptychs

Ill. 22. Closer to Van Eyck (http://closertovaneyck.kikirpa.be)

Ill. 23. Adam: macrophotography & X-radiography

Ill. 24. Adam: infrared reflectography & infrared macrophotography


2 This paper is written from the field, looking at some of the scientific theories and developments regarding terminologies and the like. The probably flawed interpretations of these developments and theories are not on purpose but reflect problems often underestimated in our day-to-day work.


4 Current count: over 1.000.000 photographs, over 450.000 object descriptions, over 16.000 restoration/conservation files.

5 The Dutch MARDOC Foundation (Maritime Documentation) was founded in 1981, to study what existed outside The Netherlands as standardized description of museum objects. Museums in Great Britain were already more advanced. [http://www.erfgoedwjs.nl/index.cfm?art_id=40, 2014-08-19]

7 Compare to the now withdrawn ISO 5964:1985 Guidelines for the establishment and development of multilingual thesauri.

8 (nl) and (fr) will be used to refer to Dutch (nl) or French (fr) words/terms/concepts.

9 The vocabularies and authority files are available on the project website [http://www.partage-plus.eu/en/contents/12_Deliverables+and+documents, 2014-08-19]

10 The Getty Vocabularies as Linked Open Data (LOD) is a most promising example for an institution as KIK-IRPA [http://www.getty.edu/research/tools/vocabularies/lo/, 2014-08-19]


12 Multiple broader terms are a blessing. They show our ability to think in multiple or flexible ways. But multiple broader terms are also a pain, because they show us that in the real world consistent or strict categorization is impossible. And as such it is of course also a technical problem.

13 BALaT: Belgian Art Links and Tools (http://balat.kikirpa.be)


17 It touches the subject of this research, but needs a separate study.