STANDARDISATION OF MUSEUM DATA – A THREAT TO THE HERITAGE?

Hans Rengman
Sweden
hans@meta.se

By looking into classification systems, subject heading lists and social tagging initiatives etc we can find new knowledge about how museums, curators, information scientist and researchers use the terminology and standards provided or under development in our neighbourhood. I will here discuss some examples and their effects and analyse the situation. The bearing theme will be the importance of curatorial awareness in standardisation process – both in in-house work, system development and museological research.

We often think of technical issues – such as machine platform, operating systems and the program environment of the databases - as the main obstacles in collaboration and standardisation. This is not the whole truth, if even a problem at all. We are today facing many more problems areas, which will in fact not be obvious to curators nor to technicians and systems architects which groups both tend to look at "their" favourite issues, of course.

The increasing need for cooperative management of heritage knowledge – or at least possibilities to search and find information without being aware of the sources and their special structures – makes it necessary to add a curator's deeper look into the field of meta-knowledge between the "real" knowledge and the machines and systems.

Heritage institutions haven't been so focused at this in earlier days – general public knowledge were deeper, while the sources were local and the experts too. The same goes for many of the end-users. A large amount of pre-understanding was built into the professionals routines and was transferred from person to person in order to be able to read old ledgers, cards and registers.

Standardising the databases force museum staff not only to be experts on 18th century paintings, wooden furniture, old maps or a certain area of industry history but also to be able

to handle the language in a standardised way, to be aware of syntax and structure, spelling and multilingual issues in a much higher extent than earlier.

CLASSIFICATION AND SUBJECT HEADINGS – SOME ARCHIVE, LIBRARY AND MUSEUM DIFFERENCES

Being a bit square minded, for the sake of argumentation, one can say that libraries mainly use the meta information as a tool, a part of the retrieval process. The main purpose in library workflow is to provide the user with the book (or another source) best suited for answering a question or an information need – as efficient as possible and not really considering the purpose or making any analysis of the content on behalf of the user. The better the metadata is, the more relevant the amount of knowledge the user carries home. Libraries have a tradition going back to somewhere in the mid 19th century in subject classification and in archives the need for subject classification on item level or within documents is a quite new phenomenon.

Archives are known to have a good and strict structure in organising the documents, series etc. It is well functioning and sufficient for a researcher familiar with the scheme and trained to dig into the material. One can say that the aim is to deliver a common and understandable structure to all documentation of society bodies and social life. Metadata is presently used to facilitate the access points; persons, geography, time etc and by that telling the user: ”Here is a list of possible sources, they are of a type that are known to house information of the kind you asked for. Please feel free to dig further.”

Museums, can be supposed to combine both these aspects, whether it is about finding a portrait of Alfred Nobel or browsing through all furniture looking for a special type of carving or colour pattern.

While libraries have their Dewey and UDC with international and academic acceptance and nationally adopted structures as the Swedish SAB which is common to almost all citizens and used in public and school libraries, museums tend to have a local system from one of two families if they keep general collections. Cultural Heritage museums often stick to their traditional system, developed over years and trimmed to the physical shape of the local

collection content - a great tool as long as it isn't any need for cooperation outside the museum. In other cases, the museum can chose to adopt any of the intellectually developed systems or models that covers the entire world or a well defined aspect of it. AAT, Iconclass, ICOM Costume Class, OCM and a wide range of other systems. Some of these tools are specialised and might be regarded as being too narrow for a general cultural heritage collection as we see them in many regional and local museums.

Libraries also use the system of subject headings like the LCSH. Subject headings are an emerging trend within the entire ALM sector and sometimes regarded as the "new" and maybe "last" solution to the retrieval problems.

My point is that museums, on top of search facilities – or maybe at the real base – also have a tradition of arranging and categorising objects and items, cultures, persons, occupations, techniques etc just for the sake of it, just to be able to understand our heritage. So, museums have something to learn from libraries in getting the stuff out to the end users – and yet we might see a wider gap between specialists and developers and the curator's daily use and work.

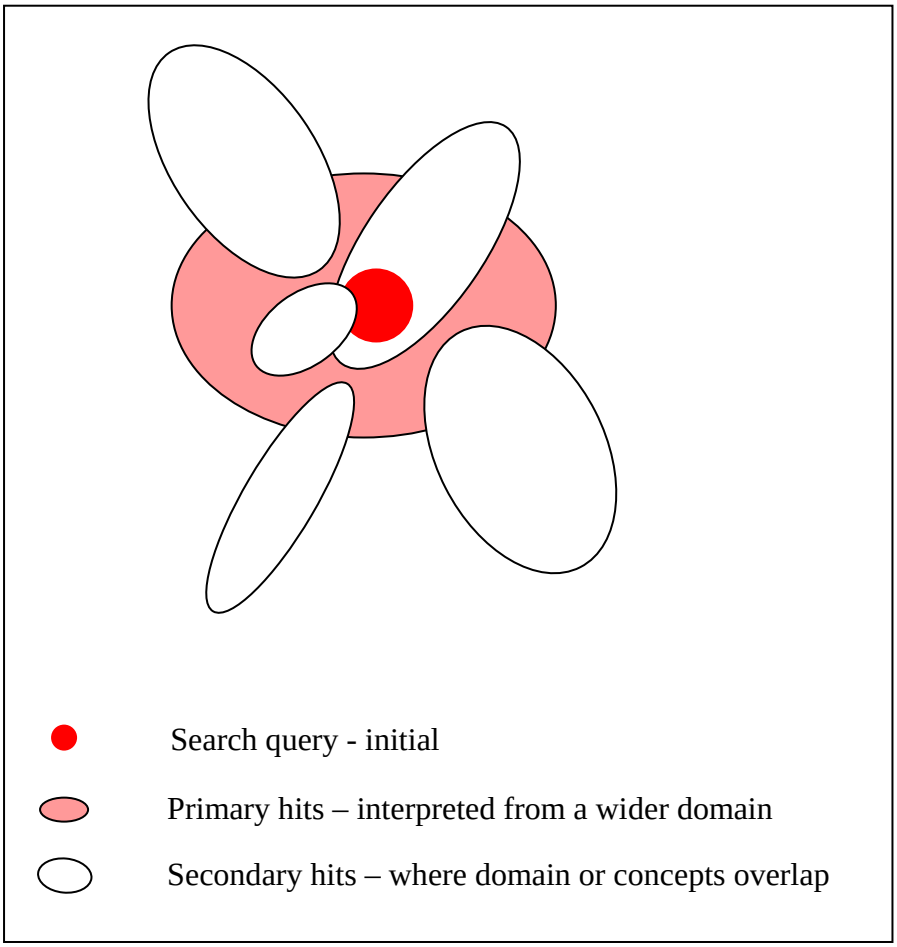
It is not a collections management issue to divide oil paintings from those made with acrylic base – it is a way to **understand the development** of the artist.
It is not a practical matter to see whether an axe is made of bronze or stone – it is a way to **draw conclusions about the socioeconomics** of Europe 2000 BC.

This means that museums have to deal with metadata for at least two reasons; organising, saving knowledge and understanding for the sake of its own, thus carrying the underlying content into future as well as a general need for search and retrieval routines – and the metadata best suited for one task might not be best trimmed for the other.

It is true that good modelled and well known standard terminology and systems that support the fantasy helps us to translate the user's questions into really interesting and useable answers. Standards also tend to dumbing down the content quality as it supports the tendencies to simplify, to use a general and broad term instead of a less known and context specific term which is narrower. As a result a wider question gives a wider answer – and the

range of possible hits is even wider as those target concepts have a wider range of interpretation.

A search question for an image, even if built on standards like classification, subject headings or social tagging will all lead to “one”, “far to many” or “zero” hits. However – this also hides all other similar pictures – not tagged or not classified - or if all of them really are well tagged and classified – we might get thousands of hits.



Whichever method we chose, we also have to deal with some linguistic issues. Either we can chose to use English language (British or American), or we can try to translate – finding that many of the elements are so bound to conceptual content with a cultural context and therefore not translatable. Some experiments in multilingual systems are known but to my knowledge no well-running ones. That leaves us with the idea of translating and modifying for local need. One example is the Outline of Cultural Materials, OCM, which were translated to Swedish in 1970ies and used for museum collections in general in Sweden, translated and adopted into

other languages in Scandinavia and suddenly was found not comparable between Sweden, Norway and Iceland. The underlying small cultural and economic differences in fishing or cattle-breeding traditions just to mention two of many examples were pushing the systems in different directions.

TERMINOLOGY COMPARISON

During some years I worked with a group of experts from the four museums within National Museums of World Culture. The museums were then quite recently merged into one organisation. Starting in investigating a possible solution for an authority-wide system for collection management it ended up in content quality analysis and cleaning of databases. Collections numberings cover some 400.000 objects and the museum profile differs – two museums having ethnographical collections, while the other two can be categorised more as having a fine art profile, collecting items from the Far Eastern and from the Mediterranean areas.

The discussion about the terminology lists is interesting as it raise the issue of end-user and target groups. The museums are by tradition focused on different user groups and also on persons with different language skills. One of the museums has almost all contacts in the international world of researchers and within a limited terminology context and thereby all documentation in English. One of them needs to handle Chinese signs and the level of knowledge is far above the normal levels of translation and transcription. The other museums have mainly Swedish documentation. The users are in the range from international experts to school children. Curators, conservators etc also treat info in different ways. This is not a specific problem in itself – but connected to the other aspects of quality, it makes the total image quite complex when combining the four museums context fields.

We found out that for the element “material” we could go on one of three tracks.

- Collect all actual terms in the museums and arrange them in a strict structure – a thesaurus.
- Leave it all for single museums to decide.
- Try to make a list of terms that is close to everyday language and a tool for common end users.

The decision was to agree on a very basic list of 15 words in natural language that gives the names for large or important groups of material often asked for. The list is not hierarchic and there can be an ambiguity in some cases. Different museums had different areas where they wanted to enlighten specific terminology. The museum with much objects of "mother of pearl" can have difficulties in accepting that this was just a part of "animalic" and so on.

Animaliskt material	Other animalic material
Ben, horn	Bone
Läder, skinn, päls	Leather
Mänskliga kvarlevor	Human remains
Papper	Paper
Textil	Textile
Trä	Wood
Växtmaterial	Plant material
Lera/keramik	Clay/ceramics
Metall	Metal
Sten/Mineral	Stone/Mineral
Glas	Glas
Plast	Plastic
Syntetiskt material/konstfiber	Syntetic material - other
Ej bestämt	Not classified

This was agreed as a tool for first entry for searching from general public. Looking at the content in the dozen of databases within the NMWC we can see how both tradition, competence, awareness, and differing ways of meeting new technical demands, colours the way the content is described. Also how end user's needs set rules that seems to be more pragmatic than strictly structured.

Another case shows the linguistic range of word use. Different function and different personal language skills colour the ontology not strictly defined. It is obvious that this is not a way of saying that these museums have low quality in collections or documentation. Yesterday, this was not a problem – today computers help us to be more fundamentalistic and strict. Similar examples are to be found in any group of selected heritage institutions of heterogeneous origin.

Målad	Painted	(*adjective - common gender*)
Målat	Painted	(*adjective - neuter*)
Bemålat	Painted	("old" form – more valuable?)
Färgad	Coloured	(dyed)
Painted	!	(done by guest researcher)
Röd-färgad	Red-painted	(a house)
Blå-målad	Blue-painted	(a boat)
Målning	Painting	(*noun* – the technique - can also be used as "a painting")

The first two formats are equal and reflect the grammar gender in Swedish language. A common way to make retrieving of data easier is to have writing rules saying "use the t-form in all situations, to be consistent".

The third is an older form, a synonym, that is today loaded with a value – this object is sometimes supposed to be of higher quality than a "målat" – but it might also mean that there is flowers painted over the covering paint - but trying to stretch this in a discussion, we often ends up in laughter and red flushing faces as the arguments decomposes. It is overrepresented in museums documentation! It is obvious that this archaic language is not well functioning as retrieval tools when interacting with new Swedes or in the international community

"Rödfärgad", "Red-painted" is a composed word that we do not want to have in a strict and standardised terminology. (At least not when discussing in this fundamentalist perspective.) Still it's there – used by many and understood by more. Some times even a part of even more complex terminology where "rödfärgad" is not just "painted in red" – it is painted with a special type of pigment and colour base used in Sweden – so "rödfärgad stuga" is a strong culture biased concept not equal to any red house…

"Färgad" seems to be another puzzling word, in many persons mind restricted to the meaning of dyed – the way you colour textiles by dipping them into the pigment. As seen from above the word "färgad" can have a wider meaning even on its own.

All these examples are subject to contextual prerequisites. Even if it, at the beginning, looks like there has been no or not sufficient proof reading, the second impression must be that

- there is a lot more behind the single words than we can see at the first sight
- personal bias makes a definition better or worse - both writing and interpreting
- research traditions differ and tend to invent language and terminology of their own
- in house traditions strengthen different ways of writing
- the context can be fuzzy and giving wrong signals

NAME CONVENTIONS AND GLOBALISATION

While one museum uses a German based name convention in documentation – due to the research history in late 19th century while other museums use an English transliteration based on their traditions. This points us to two problems today.

- The need for having a geographic master pointing to the "conceptual place" and a link for the German and English words, instead of just changing one of them and lose context and quality.
- Whichever language selected, the quality of matching a name to Google Earth for example, is not at all good. Today Google Earth uses an American English name form and a modern naming convention while museums use the old one from context.

One might ask if this is a mere language issue, a cultural one or a more political. Whichever - today a "non-answer" on the web in general and in search tools like Google is often considered equivalent to "non-existing" and an unstandardised nomenclature puts you aside. Heritage institutions do not deal with the culture politic aspects of these fine-tuned problems in their everyday work.

THE TIME DIMENSION – AN AGING QUALITY

The time dimension of quality is often forgotten. Many items have early documentation content which is outdated today – this is often the case in ethnographic or anthropologic

collections. Many objects – considered strange by the early researcher are classified as "ceremonial" or "ritual" and coloured by our ethnocentric view. Even if not accepted today for humanistic reasons the terminology mirrors a viewpoint of its time and were once accepted as intellectual and research based knowledge of high degree.

By definition this is valuable for its own sake in museum documentation, showing differing viewpoints of different times. Maybe not useable in modern IR but indeed a special effect of museum documentation compared to the rational information providing in image bureaus or research libraries.

We have a task here to both be correct and exact as of our knowledge today – meeting the needs from users of today – and at the same time show the history and development of research. Examples of this phenomenon can also be taken from geography names where country names changes with the political situation and just by using them you take a position in the war.

Conclusions from the example fields mentioned above, is that we should behave critical to the results of standardisation that is brought to us by systems that tend to set global standards – both large and small scale! Relevant is also to consider the different end user groups, their skills and their bias. It is important to save, rescue and maintain the cultural history aspects of linguistic differences in cultures and language as well as improving the standardisation tools and efforts from an IR – and supplying perspective.

MASTERING THE AUTHORITIES

Initiatives to translate classification systems, to merge them or make a megasystem have sometimes been tried or discussed. Benefits and outcome has not been satisfying and we have seen no real good results of this. The reason is obvious – all systems have to be used faithfully to their main idea – and they differ! A general rule, very often forgotten in daily routines, tells us that one has to be aware of the prerequisites and the context for each of the systems to have them making sense.

Classification systems carry their context and viewpoints and are not connectable or possible to overlay. A numismatic scheme can not be the natural extension to a general scheme like OCM or SHIC, they have to be regarded as two different, complementing, tools. Two costume class schemes can not be used to climb unlimited just because there is an agreement on one single concept as a crossroads.

In KMM there is a challenge to test some of these issues in a more offensive way. Two main reasons are discussed besides the practical goods of having the systems running in an environment where many museums help to develop the content quality.

- To have a general technical system developed, called the Masters – suitable to handle as many authority systems as you wish and to provide them to the users.
    - A general resource for categorisation and documentation with authority lists, thesaurus etc. The conceptualised name master can easily handle the transliteration issue between languages.
    - A general tool for building such authorities
    - A way to force – or help - the user to the high level of standardisation in input phase, and to lead the user around any free format traps that can lead to later difficulties.
- To have a test bed for looking at the possibilities of decision support and expert systems within the field of museum classification- Is it possible to make automatic classifications – and to what extent?
    - A study object for the field of automated services and expert systems use in museums material and workflow.

Finally – as a result of these experiments – we will seek an answer if it is possible to build such decisions and support into a workflow system within the heritage sector?

SUBJECT HEADINGS

Comparing the subject heading from different museums gives us interesting information. In this case we can see how three museums use a subject heading list with same origin – customised for local use. The example is chosen from the field of transport and shows mainly

concepts of "what" is transported and "how" it's done and some other facets. The examples are in Swedish but the pattern should be general. It is obvious how some words are ambiguous. "Car transport" is – in a museum context – clear. The same goes for "Timber transport". The term "Water transport" can in Swedish wording indeed be both the transportation of drinking water in a tank as well as transport made by boat at sea. There is a significant difference in the use and adoption of the lists. It is also interesting to think about what will be the best strategy for future collaboration.

Quite a lot of concepts seem to be unique as well as there is just a few common concepts. This is true already in comparing two lists. Adding a third one increases the differences, and the difficulties for collaboration.

These lists show, of course, a difference in collection content or profile, and a local adoption of the terms best suited for describing it. It also shows how curators discuss and recommend for future use a group of words or concepts that suits a certain aspect or view. The latter is not so easy to detect in a small scale example but is indicated by how words are chosen, the use of nouns or verb etc.

Two ways to go can be seen – both with at least one major drawback. Merging all lists together gives us a "complete" ontology where we can find all concepts that we are used to, no matter what museum context we are into. The bad thing is that we have to deal with up to 100% more concepts which are not useful in our world. Starting at the other end, to join by the concepts that are common gives us a slim list, surely accepted to be true – but leaving a lot of concepts out, where the agreement isn't easy to find in the broad community.

The first of these models will lead to a fuzzy terminology and a risk for unwanted synonyms etc. The second might point to a simplified scheme, where users are forced into using listed terms instead of adding new ones.

Biltransport	**Biltransport**
	Busstransport
Godstransport	**Godstransport**

Grustransport		
Hästtransport	**Hästtransport**	
	Hästtransport	
	Hötransport	
Järnvägstransport	**Järnvägstransport**	
Koltransport	**Koltransport**	
Lastbilstransport	**Lastbilstransport**	
Malmtransport		
Mjölktransport	**Mjöktransport**	
Mopedtransport		
Oxtransport	**Oxtransport**	
Posttransport		
Sjötransport	**Sjötransport**	
Sligtransport		
Spannmålstransport		
Timmertransport	**Timmertransport**	
	Transport	
Transportband		
	Transportföretag	
Transportvagn		
	Transportör	
Vattentransport	**Vattentransport**	
	Virkestransport	
8 unique values	**7 unique values**	**11 common values**
tot 19 values	tot 18 values	of total 26 values

Biltransport	**Biltransport**	**Biltransport**
	Busstransport	
		Djurtransport
		Fångtransport

Godstransport	Godstransport	
Grustransport		
Hästtransport	**Hästtransport**	**Hästtransport**
	Hästtransport	
	Hötransport	
Järnvägstransport	**Järnvägstransport**	**Järnvägstransport**
Koltransport	Koltransport	
Lastbilstransport	**Lastbilstransport**	**Lastbilstransport**
		Livsmedelstransport
Malmtransport		
		Militärtransport
Mjölktransport	**Mjöktransport**	**Mjölktransport**
Mopedtransport		
Oxtransport	Oxtransport	
Posttransport		Posttransport
		Resa och transport
		Sjuktransportbil
Sjötransport	**Sjötransport**	**Sjötransport**
Sligtransport		
		Soptransportvagn
Spannmålstransport		
		Stentransport
		Sädestransport
Timmertransport	**Timmertransport**	**Timmertransport**
		Torvtransport
	Transport	Transport
		Transportbana
Transportband		Transportvagn
	Transportföretag	
Transportvagn		
	Transportör	Transportör
		Trupptransportbil

Vattentransport	Vattentransport Virkestransport		
7 unique values	**5 unique values**	**12 unique values**	**7 common values**
tot 19 values	tot 18 values	tot 23 values	of total 38 values

	unique			common	
	um	ylm	vgm	um-ylm	um-ylm-vgm
unique values	7	5	12		
within a museum	19	18	23		
% unique values	37	28	**52**		
within three museums	38	38	38		
% unique values	18	13	32		
common values				11	7
within museums				26	38
% common values				42	**18**

Half of the values in one museum are unique
1/3 to 1/2 unique values in all museums

Consensus for two museums is as low as 42%
Consensus for three museums is as low as 18%

Obstacles to overwin

The KMM project and the experiments will show a structured way of handling more and more of the actual data and lead curators through the process towards a higher degree of quality in the sources, but we may also be aware that this can have the opposite effect. We might have as much to gain by study the effects of these tools and routines, as we win in practical support in the first stage. Here are some examples.

When we use a standard terminology we provide the curator with a set of words to choose from. If there aren't any suitable alternatives, you are stuck in the same situation as the child trying to put small pieces of different shape into a box with holes - round, square, triangular – it is all manageable until it comes a star. You end up in fetching a saw to trim the star or trim the box. But what is the best way? - If you get just one star, it's easier to trim the star to a round – if you will get many stars, it will be needed to trim the box to swallow all of them.

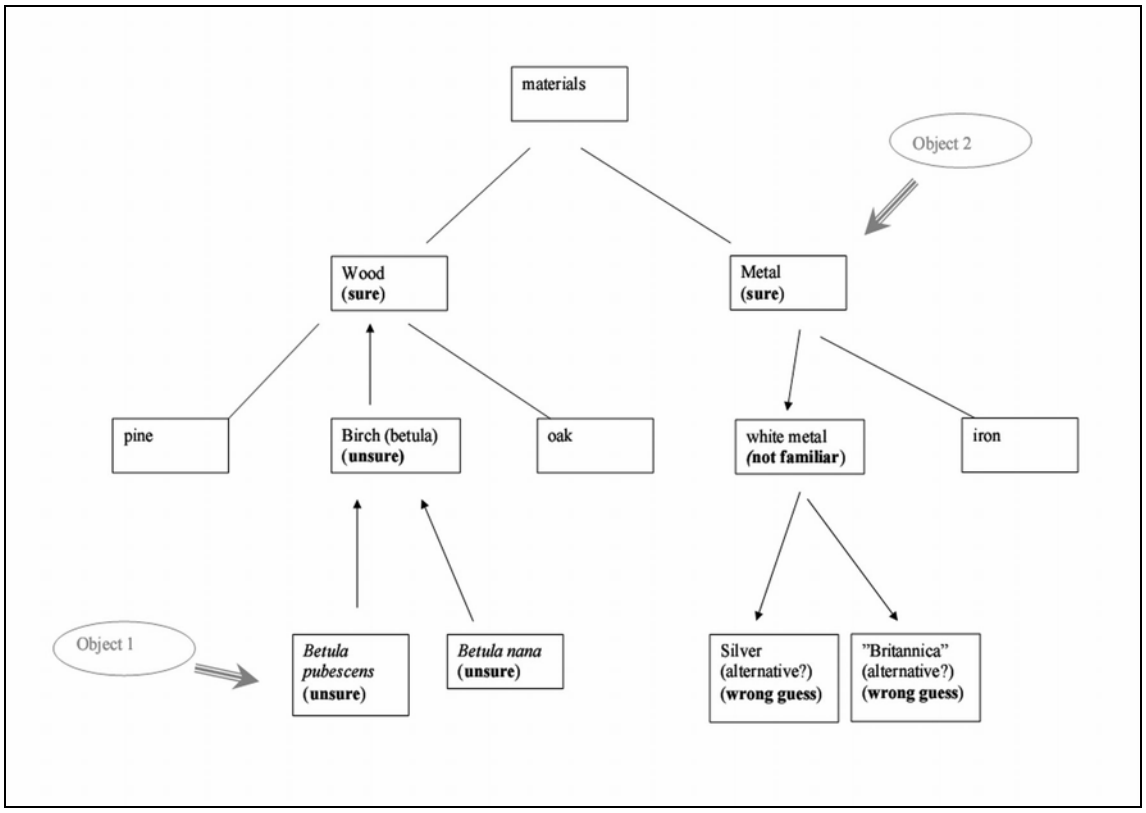
- You risk to be forced to define an object as belonging to another category – to be something else than it is, due to simplification.

In a case where you have quite good information and detailed knowledge and the list of terms to use is built just on words or concepts at a basic level, you will risk to be forced to declare an object belonging to a more general group than your expertise can find out. If there is a "Windsor chair" and you do not have it in your list, you look at the level above if there are "chairs", If not you climb another step and decide that the object is a "furniture" – which is indeed all right, but not precise enough.

- You have lost a lot of knowledge due to generalisation.

From the other side of view, the end users, we can see a similar risk. The question put to the database might be so broad or out of focus that the answer is either a mix of relevant and irrelevant objects, or a far too big, and wide, group of objects to be interesting. We can also see the general risks over time of having forwarded a heritage nomenclature or a language domain, simplified to that extent that is not usable for more complex purposes. This fear is relevant for both the end user and the professional staff that is leaning too much towards the tools of standardisation.

We risk – putting it in another way – either to loose precision in levels or precision in directions.



DISCUSSION

A concern worth discussing for a moment is whether old data is bad, (wrong, inconsequent, irrelevant), unstructured (low standard in form and content) or just too complex [personal customised, context bound, composite) to be handled by automated processes and in "simplified" dialogue with end users – whoever they might be?

Imagine an old catalogue card from the card box catalogue. Printed structure, typed info, layers of handwritten comments, drawing, photographs, exhibit history and references to literature and attachments. A bit of a lifetime picture of its own.

For the trained reader, this tells much more than the database record of the same object – even if all text is transferred correctly and stuffed with modern retrieval terminology.

So lets scan and make a PDF of the card – Fair enough – it gives a real good opportunity to interpret the story. But, in the standardisation process we also risk to judge parts of the old message as uninteresting, misleading or faulty, overrun by time or politically incorrect etc.

We need to look upon the metadata from two different viewpoints – in parallel and interacting:

- Information about an object, its context, the history, the user's relatives etc can never be regarded as bad – it is the story we are told and can retell
- Strategic and structured info that helps us to organise, retrieve, combine etc should be considered as an extra layer of info, a tool, a set of handles to be able to turn the object around in our minds and observe from another viewpoint.

On a concrete level the information elements used for retrieval or for knowledge understanding, can be identical but on an abstract level we are talking about two different roles for the metadata.

An often present discussion focus on whether "object name"/"simple name" can or should be standardised, put into and used from authority lists, thesaurus etc – or be regarded as a name primitive and left alone to live freely. A range of different solutions, based on varying ideologies (more or less well defined) is presented.

- simple names as third level of classification (furniture/chairs/Windsor chair)
- simple names used "in the format the donor used it"
- simple names in an over structured syntax due to old computer history (Lamp, kerosene-, part of)

While a lot of information (like material, technique) benefits from standardisation, for retrieval reasons, (and local or other special context info might be handled as comments), the simple name in itself carries an image of the past.

In technical /reason/ it might just be easy solved by multiple fields for simple names – with an option to flag the context

- simple name – from donor
- simple name – from authority
- simple name – dialect

For ideological and museologic reasons we must be aware of the possibilities to enrich the story by using one type of metadata for the retrieval and another (more vivid) for documentation and retelling and to give space for a set of parallel words.

The terminology comparison and the analysis of the content in a database or in a group of databases can give us direct knowledge to syntactic and structural handling of the language. It can also give us a higher degree of awareness of methodological approaches to the used language and help us making better rules for future work. There can be openings for looking at museum documentation terminology in a museology research perspective at any of our universities.

Working in a project like KMM gives the opportunities to use a large amount of real data from different environments for testing. The group of museums in the project provides both a wide range of objects (=terminology) and a varying set of user history, in the sense of different habits and writing rules. The merging of these data, to clean it all up and to make it generally usable opens both technical and museological discussions.

MULTIPLICATION EFFECT – A MEGAHIT PROBLEM

We have to deal with both general concepts – generic in definitions and wide in its definition quality, and more specific concepts with an exact and well defined meaning together with an interpretation or rule for use that adopts the concept to a context.

In a wide area retrieval situation, a question based on a generic concepts returns “millions of hits” and the search based on a more specialised concept, which seems to give high precision in a limited situation, happens to give a result that is excluding many relevant hits. This is the case when documentation stops on a varying and general level. When there is not a class, a

concept or a term used – the question might return no answer, and the same happens when there is a faulty value. This might be obvious and can be worked out by adding new values and correcting old values in the documentation. Some of this struggle can be solved with synonyms or a mainly contextual inherited structure like CRM as long as we see it as a system or structure issue. Still it leaves us with a need for the museological analysis and that is why we have to stick to the theoretical perspective.

Left to study are all the objects and metadata in between – the objects that are not possible to put on a more specific value than "tool" or "furniture" – be it that we don't know for sure or that the conceptual structure leaves without a more detailed alternative. This might be the case with old tools – used in a much specialised context but where tradition never put a detailed term on it, or being it the case that our expertise shows a lack of knowledge in this area.

Sticking to a terminology that only partially catches the view of the world - a not fully covered ontology – we might get problems to put our index finger on the right spot. But – we don't want to invent a synthetic terminology and we don't want to ruin the linguistic aspects of the heritage.

There will always be gaps between concepts and words thus making concepts and material not reachable. (We know that all "pliers" can't be placed on level five or six of detailed terminology even if some of them can.) There will always be "pliers" or "tools" and it leaves us with at least a minor feeling of uncertainty – is there a more detailed word or concept that fits in here. A database search for "Pliers" gives a certain amount of hits. What we can't find is those which only have a more detailed name and not also categorised as "pliers". Hit rate in a traditional museum will be something like 10-200 for the word pliers. "Drawing pliers" (a special form of "side cutter") will give a single hit in some specialised museum (specialised collections OR general collections with – far too? – specialised knowledge on personal basis)

Regarding a large scale platform, where the single collection might be multiplied with all museums in a country, or with all European countries, or both. The Swedish KMM, Europeana, etc is a general concept which will generate enormous amount of hits on a general level of abstraction while a specialised term used for searching might in worse case leaves us with a single hit and a curious feeling that we might have missed something important.

By connecting the large volumes to authority nomenclatures and standardised structures of concepts we will be able to translate and exchange – but there is a risk that we talk about fruits instead of apples, pears and bananas.

The secondary result, the real effects of this might be that some objects or images – with a good documentation and a well suited set of the right terminology – by accident returns on the first row of the list of hits. The same photo will appear in all books and a certain object will be well known and wanted for several exhibitions. Some objects, photos and paintings will continuously sleep forever while other will be lifted up to fame and glory and worn out – both physically and as ideas or content.

There seem not to be a nuanced scale between these outposts. Simplifying gives MegaHits and individualisation gives NoHits.

SOCIAL TAGGING

Is there a help to find in the social tagging initiatives? Looking at the present content of Steve.museum we find that also here you will get the answers you ask for. The BIAS of people interested in taking part in the experiment has to be taken into consideration as well as other parameters. If there is a real "social" approach there will be a lot of terms used from natural language (and natural way of looking at art – as images). A study of a random series of artworks in steve.museum gives a selection of terms used. Placing them in groups by characteristics shows a pattern where the pre-iconographic – and physic terms; apple, balcony, street, hat etc are the most used. Analytic iconographic or iconological terms or concepts are almost totally unrepresented. "Medusa" is one of the rare ones – but is seems to be just a girl with curly hair in fact.

steve.museum – sample tags from 10 random images							
Object type	**Style**	**Material**	**Form**	**Technique**	**Pre-iconographic / abstr**	**Pre-iconographic / concrete**	**Iconographic**
Category					**Motive / content**		

art	brushwork	bronze	cube	carving	advertising	Man	mandala
						Men	
						mujeres -	
figure	pointillism	clay		incised	apostles	apprentice	Medusa
genre						nose	
painting	postimpressionist	gold		painted	assistant	Poster	tanka
Portrait	social realism	stone			blacksmith,	artists	
						ribbons	
portraits					blacksmiths	big nose	
						squares	
						swirls	
pottery					Chest-up	billboards	
statue					City	black hat	
statuette					cityscape	blue	
						woman	
textiles					ciudad - stad	women	

curly hair	Boy
fashion	buildings
geometric	chimneys
linear	circles
	edificios -
market	väktare
pattern	fire escapes
relief	garden
seated figures	geometric
street market	hair
streetscape	head
trabajo -	incense
arbete	burner
urban	
landscape	line

So the question raised is: How do we benefit from a search in art museum databases resulting in a hit list of thousands of painting showing (among other things) an apple?

To reach further and to keep quality in museums knowledge management there is a need for curatorial and museological analysis. Otherwise we end up in statistics. Considering standardisation as a tool of importance, and cleaning and maintaining of databases as necessary work we should not forget the human aspect of the content in our documentation.

Besides being labels and handles for getting in touch with the content in a structured way we also have to face the fact that the information is a message about the context in itself, the language has a cultural level that has to be documented in itself and it all shows for the future a snapshot of the time of origin – both ancient cataloguing and today's electronic documentation.

There is need for two levels – or two groups of terminology. We must keep the difference between the structured and formalised terminology that is needed for comparing and exchange and the free forms of language that bears the time, the person and the flavour of history into the future. In both cases professional and layman approaches can join in increasing the quality.

In a large scale the amount of objects in a group or class will be unmanageable. It might point to a revision of the classes and subclasses. If so done – there is a risk that we construct patterns and syntax not yet seen. If not – we will have to stick to very large groups of items and data base records that carry a similarity, thus hiding the originality of the object.

By studies and research, by discussing methods and comparing results we can share experiences and raise the level of meta-knowledge in this field. Here I can see many interesting fields of comparison not only within museums, but also in the entire heritage sector where we all might gain from earlier results and experiences

CONCLUSION

What's so good about standardisation? Well – As I see it, Standardisation will be the tool for us to make contacts outside the institutional box possible. To be able to connect to other museums, to meet between archives, libraries and museums and to support the end users in having easy access to the data. On the other hand, a believe in the magic of standardisation makes us prisoners in the system, peeling away uncomfortable odd data and driving us to use more square boxes than necessary.

The NMWC studies have put a finger on the concrete examples and their background. Systems for catching multi-institutional data must be constructed in a way that does not cut away one aspect on behalf of the others.

The KMM project gives us a possible tool to large scale manipulation of information and a platform to discover ways of rational data handling. The risks in driving the standardising too far will be noticed.

Subject headings points to new curatorial traditions not yet analysed and social tagging is yet not thoroughly studied to its effects on knowledge. The results are to be taken carefully and analysed before use in wider context.

Large scale international projects push an information structure that benefits from, and enhance standards. As a result there is a risk that we both are dumbing down the heritage and risk a lower quality in results than we intend. Specific knowledge are Googlified. Amateurism on the net gives a polarised positioning between popular/simple/meaningless image elements and an expert culture (and cult?) based on iconographic analysis, quite difficult to use by a layman.

Studies of different museum catalogues and use of terminology and standards is needed for further analysis. The awareness of conceptualisation, context, the linguistic issues in heritage etc is instrumental to keep the level of quality and museology or information science research must be stimulated and initiated. This is essential to balance an image where the more technical approach to "standardisation of museum information" today separates the developers and the curators. The goal should be to keep the natural language as untouched as possible, to adapt standard terminology as tools and keys "outside" the documentation in itself

and to try to hang on as many subject access points, tags or information viewpoints as possible – a challenge both to the processes and the professional role. We need discussion and we need research within these fields and we can benefit from looking at each others systems with new eyes within the entire ALM and heritage sector.

SO – THE CORE MESSAGE OF THIS PAPER IS:

Take some minutes to think over and analyse the content and the way we handle it. Data put in a new order might show new knowledge. This is true both for the content itself and for the mirroring of how we are able to handle and manage the heritage knowledge. It leads to reflection – museological and professional as well as about the museums tasks in a wider political perspective. Curatorial awareness together with increasing quality in the development of standards and systems leads to better documentation and mediation of the heritage.