

Moving Samoreg into LIDO

Experiences from a mapping exercise

Hans Rengman

META – kunskap om kunskap
Uddevalla
Sweden

Abstract:

Is it possible to successfully map a 30 year old national de-facto standard (SAMOREG) for museum object documentation into a modern event based scheme (LIDO)?

SAMOREG set a standard for museum terminology in mid 1980-ies. Divided in a set of six sector specific forms, information was collected in a coordinated and well defined structure, covering cultural history as well as fine arts, archaeology and nature history. This model has, explicit or implicit, been the platform for almost all CMS development in Sweden since then. Most terminology today can be referred to the SAMOREG system in some extent. Two Swedish projects have over the last years gained experiences from delivering Museumdat-formatted content to Europa/Athena. The mapping exercise referred here is part of the Athena work.

The work is in progress during spring 2010 and this paper will give a report of reached experiences and hopefully contribute to the LIDO development as a globally accepted standard.

Introduction

SAMOREG set a standard for Swedish museum terminology in mid 1980-ies. Divided in a set of six sector specific forms, information was collected in a coordinated and well defined structure, covering cultural history as well as fine arts, archaeology and nature history. This model has explicit or implicit, been the platform for almost all CMS development in Sweden since then.

Most terminology today can be referred to the SAMOREG system in some extent. Two Swedish projects have over the last years gained experiences from delivering Museumdat-formatted content to Europeana/Athena.

The mapping exercise referred here is part of the Athena work. Its goal is to develop a general mapping overview for Samoreg based information towards LIDO.

Problem

Mapping of a well structured set of information like Samoreg open for a bit of discussion and tweaking to fit into a schema like LIDO. We can see structural problems in mapping, but the experiment also puts a finger on several content issues within the Samoreg system itself. LIDO can be used as a tool to increase the content quality in the local systems.

Spectrum gives the logical home for the information content in the first step. LIDO organizes the naming of the element in a well-structured syntax.

This project is a work in progress. The presentation will give latest experiences and results from the mapping process.

SAMOREG

History

The system was based on early computer experience carried out at Nordic museum and The Skokloster Castle. One of the lead persons in this development – and Chair of the working group behind SAMOREG – was the late Göran Bergengren at the Nordic Museum in Stockholm, former Chair of the CIDOC executive board. Another CIDOC person in this work was Anne Murray at the Ethnographic Museum in Stockholm.

Description

Samoreg structure is based on a set of element grouped into documentation blocks. Common information is to be found in the same block for all sectors. The following blocks are used – some of them are context dependant and not used for all sectors. A couple of elements are to be found in a non-typical block due to sector specific context.

100 – Basic object and collection administrative info
200 – Find and collection context
300 – Production context
400 – Description (including technical, some production context and some find context)
500 – Use context
600 – Comments
700 – Museum/acquisition context

1. SAMOREG information blocks

Within – and also in comparison between the blocks – the element id tells us which element we are looking at. Same number is applied to same sort of information, i. e. 81 is a person name in all blocks. Thus 381 is the maker of an item while 581 is the user.

Block 600 is a free text field used for notations and remarks in a very wide sense. By tradition there has been a loose context based syntax developed, structuring the content starting with further technical descriptions, comments on use of the object, curatorial remarks including the conservator's comments and finally info about exhibitions and publications.

Most geographical info is coded – or at least meant to be coded. The underlying authority systems were not well established at the time of Samoreg development. In more accurate terms, there were different systems at hand, and it seems like the Samoreg group wasn't in position to fully handle this issue even if there is recommendations for one alternative. The fields are not mandatory and in practice there has been several ways of dealing with this field of information. Free text content is partially mixed with authority file based values.

Person, Time and Geo information are the main groups where information is structured in high extent. Other elements are in higher degree open to customization. Some elements are tied to authority file use (in theory) while others are open for free text information.

In short, this system takes Swedish museum documentation from the old – locally flavoured systems of card boxes to a modern approach in documentation. Structured for computer based information handling the SAMOREG brought a general viewpoint to collection documentation and focused on finding common elements in all sectors of the museum field. This systematic approach has influenced most theory and all system development for the last 25 years.

Looking at the Samoreg elements one can find that most of them are targeted to documentation of the natural and cultural history documentation of the object. Swedish museum tradition puts more relevance to the descriptive information about the object and its user context than to the administrative needs of accountability and collections Management. This is mirrored in the terminology of the system.

In archaeology the site and the context aspects of documentation wasn't fully carried out. In nature history the system was considered to general to be useful. Fine art, ethnography, cultural history seems to form the base for use. A photographic object form has taken shape as a secondary development.

LIDO

History

LIDO is a harvesting format for providing core data from museum holdings, technically specified in a XML Schema. It is the joint effort of the CDWA Lite, Museumdat, and SPECTRUM communities. LIDO is also a preferred format in the Athena project – harvesting data from most central museums in Europe, for use in Europeana and by that the most widespread model for exchange of heritage documentation based on museum standards.

The history and current state of LIDO is presented in other papers of this conference and will not be repeated here.

- Object Classifications - Object / Work Type (mandatory) Classification	- Events - Event Set
- Object Identifications - Title / Name (mandatory) Inscriptions Repository / Location State / Edition Object Description Measurements	- Relations - Subject Set Related Works - Administrative Metadata - Rights Record (mandatory) Resource

2. Descriptive and administrative elements of a LIDO record

Mapping

The need for mapping in Sweden is raised by the development of two large projects inside Sweden, and the contacts made within the EU-financed Europeana project.

Knowledge Management in Museums, now finished, was an early adopter of the Museumdat format and provided an extensive data quantity to the early test beds in Europeana. All material reachable on line was presented in a Museumdat profile beside the local and traditional web page format. The K-Samsök project also made experimental export routine for the Museumdat but stayed to the internal format for harvesting. K-Samsök is today able to produce a LIDO mapping of information structured in their internal format.

The museums in the K-Samsök project are the same as those in the target group of Athena, within the Europeana framework.

From this perspective it is interesting to map and test a total mapping of the terminology used in Sweden to the LIDO format and the structure of Spectrum.

Elements missing in Samoreg

In the Samoreg system view there is a number of information elements missing in comparison with the structure of elements foreseen in the LIDO schema and the Spectrum model.

Some of them are neglected, some were probably not found central in 1985. Most of the information elements not covered by the structured element fields in the form, however were supposed to be described in the free text format field S601 – Comments.

This was due to what we today can call a non mature approach to computerized information handling and a clearly outspoken need to limit the human readable version to the A4 paper format – which also set limits for the layout of the on screen versions.

Associated elements

The complex of associations (associated person, associated place etc) is another way of turning the PTG elements (Person, Time, Geography) around. Instead of the “old” model with explicit fields for each role, the event/association – pair of bindings opens for a more flexible way of handling – and connecting – information elements.

We can also see a better way to adapt the information at the right place in the model. Earlier models – like Samoreg – often gather secondary information like remarks about the family of the donator and such things on the catalogue card belonging to a chair or a table. This gives a referencing problem and a quality problem. We are now quite aware of the problems of redundancy etc this brings to the museum knowledge base.

Multiple ways of mapping

The Samoreg elements collect information in the syntax of that system. Information that has to be decomposed and put into the right context in LIDO.

The man who built the chair – is to be seen as a person associated to the chair in LIDO, by the event in which the chair is made. In Samoreg the element S581 – Brukare/Ägare (user/owner) by definition shows the name of the person who has used the tool or owned the book. (And in some cases extra information.) In LIDO we have to associate the person with the object and give the association event a label of the use context event.

Elements missing in LIDO

This far no elements in Samoreg are found that not will fit into the LIDO model. In some cases the need for qualifying attributes is a condition for not losing granularity. These attributes will be specified in later reports as they might be of general use.

Heterogeneity

One area of problems is the heterogeneity of the S601 element, which collects all information that does not fit into the structured elements. There are at least five types of info sharing the S601 field.

Trailing info – Info that is technically “too large” for the form field. This is a left over from early computerization where fixed field lengths were necessary and not a big volume problem. Still there exist a lot of info divided between the structured fields and the free text field.

Specifications – Info that goes deeper in knowledge than the format of the fields accepts. (ie the distribution of different materials in a composed object)

Comments – Info that goes a bit beside the scope of the format. (ie history of the user family)

Metadata – Explanation for how to interpret structured info (i.e. – measurements)

References – Direct or indirect references to other objects, to exhibitions and to publications.

There is sometimes a context based structure in the way of writing, growing from experience and developed within a museum. Even if a general outline of this is quite easy to note, it would be impossible to find any automated help for deducting the limits of different elements.

Field structures

In some cases the same element ID is used a little bit unclear, as the different sectors shows differing needs. So, for example S354 is used for the nationality of the artist in the fine art sector while in the ethnography the same field is used for a classification of the culture.

When discussing the Samoreg system it has been a suggestion to divide the S601 into several comment fields – fore each block or even for each element. This is however not taken into any running application of CMS systems.

From a mapping viewpoint this might not seem like a problem – the field in itself contains comments of varying quality. Some info will look a bit misplaced anyhow.

The underlying problem is a problem of content quality **within** the source system. The large element with a mix of information is very unsatisfying as it has to be dealt with manually – to help the miscellaneous info parts find their way to their right structural element.

Information split

The result is though, that information content need to be split up in parts and mapped to a series of elements to be able to retain knowledge.

This point us to a major issue that has to be taken care of – the quality of the sources. We need to put effort in manually interpretation of all documentation records before exporting them to Europeana.

Horizontal/vertical structure and attributes

The difference between horizontal and vertical orientation of the models are a bit tricky to handle. The set of explicit time elements in the Samoreg system should be mapped into the single – but more complex - structure of “associated time” etc. Here is need for a equivalent set of event types as attributes to label the info in a meaningful way and avoid ambiguity.

These attributes are of two kinds. Attributes that are general and should be built into the LIDO system by default, and a possibility to locally add attributes for special reasons and to repeat elements and attributes.

Person issues

One issue is to deal with the information about persons. In Samoreg – as well as in other older systems there is a field box containing title, first name, last name, age/year of birth, and in some cases even geo info. This possibility to use the field as a concatenated field for a set of elements opens for a localized, or even a personally customized, use of the field. In practice, even other biographical data will be found here, instead of in the comments field.

In addition the fact that data' about person's lives in five blocks – tell that the amount of combinations of data is enormous. Any attempt to find an automated way of mapping these elements will not succeed. Parts of a field will map to relevant LIDO fields, and attributes are needed to distinguish different aspects from each other.

No real test of this has been done so far, but it will surely be interesting to examine the possibilities.

IPR issues

Information about Rights management is not at all covered in the structure of the Samoreg system. All such information has to be placed in the free text element, or maybe more common, in a combination of attached text documents and archival documents. If IPR information exists – there is a need for cut-and-paste of content from the comments fields.

Analysis

The models

Samoreg gather information in a more structured way than earlier local systems – needing knowledge of how to interpret the context of the card – and by that opened for a recognizable approach to documentation information in a lot of museums. It also put focus on content quality which was very important in Sweden at that time. LIDO, in a way, breaks this logical framework down to a more fragmented presentation and opens for a far more flexible way of connecting information aspects. The XML-styled output is merely not readable in practice and not to be taken for a catalogue record. The event based approach and the associations of persons, time geography and other objects will make it possible to build chains of information in a new way.

Though LIDO is a harvesting format – not meant for building full documentation – it will certainly have impact on ways of saving and presenting heritage knowledge within a local museum documentation context in future.

One has to discuss whether this impact is all good or bad. We know that standardization of structures and models on the one hand are of great help, not allowing the user to forget any perspective. It opens for distinctions between fine granular differences in meaning, which are not easy to note in other case. We can also see that systems demand force user to produce information “to fill the form”. From Samoreg perspective we can see a tendency to fill the object record with secondary information, trying to catch as full documentation as possible. LIDO in itself will NOT force this effect – but improper use in combination with the concatenated information in fields like S601 etc. might lead to negative effects if particular pieces of information are duplicated, or not properly mapped.

Those local attributes must be suggested with a bit of open-minded fantasy, still with a chance for users to deduce the meaning. Extensive use of local attributes to qualify the content might lead to a plethora of possibilities which might give a redundancy as well as will be the case with a too open set of general fields. A central service for comparison and updating of the common set of attributes might be a valuable service.

In other cases elements in Samoreg contains information which directly belongs to a set of combined elements in LIDO. Simple mapping of a Samoreg element into a single field in the LIDO structure will end up in information that (partly) does not make sense.

It might also be elements which we can discuss whether they should be mapped in different ways as they might fit into two separate LIDO elements.

In a way of saying – there is a “one to many” – option as well as there is a “many to one” situation to take care of.

Some of these difficulties are due to systematic issues, some are direct or indirect a matter of content quality.

Mapping the schemas and – in next step – transforming the content, should lead us to a better content quality level in the local systems.

Experiences

To make a general summary so far in the experiment, the LIDO structure will take care of the SAMOREG data in a decent way as the model is developed so far. Some additional attributes will be needed to ensure the fine granularity or any local needs. These additional attributes can be of interest for all users and should positively be added to the LIDO model. Some information sets in Samoreg will have to be cut up in parcels and divided to different parts of information, and thus mapped into different places in the LIDO structure. The LIDO mapping exercise will be of great value – pointing out such knowledge areas in Samoreg.