Dublin Core as a tool for interoperability
Common presentation of data from archives, libraries and museums

Leif Andresen
Danish National Library Authority
Tel: +45 3373 3373
Fax+45 3373 3372
lea@bs.dk

Abstract
Over the last years there has been an increasing focus on co-operation between archives, libraries and museums – particularly in local and regional contexts. This was the reason for setting up the Danish ALM standard committee appointed by the three national authorities for the ALM sectors. In February 2006 a report was completed by the committee. Dublin Core is recommended as the common format for content information to which selected parts of individual databases can be mapped. The selection is based on relevance of data according to search and presentation. The 15 basis DC elements are complemented with DC refinements, Administrative Components and supplements from the developed dkdcplus schema. XML is recommended as exchange format and a XML schema called dkabm collects XML schemas belonging to the built-in content formats.
With the dkabm as starting point, Dublin Core as basis for technical interoperability between different domains is analysed. The analysis covers both the content that is processed and the handling of relation databases when mapping to record-based formats like Dublin Core. Furthermore the chosen extensions to Dublin Core are analysed according to possible disadvantages of Dublin Core as exchange format. The conclusion is that Dublin Core is useful. There are some minor disadvantages, but no alternatives.

Keywords:
Dublin Core, XML Schema, interoperability, library, museum, archive, common presentation.

1. Introduction
In 2003 the Danish Ministry of Culture entrusted the three national authorities for archives, libraries and museums to develop recommendations for data content, data formats and data transport. The goal was to facilitate presentation of joint information from archive, library and museum sectors (in the following ALM-sectors) for the public on the Internet. The initiative was in prolongation of ongoing local/regional projects (1).

2. The ABM standard working group
The steering group consisted of the heads of the three national authorities and the work has been done by a working group with six members, two from each sector. The first report (2) was completed in 2003, and contained essential recommendations for further work.

2.1 The first report

According to the content of data this first report includes some drafts for mapping between the sectors' specific formats to Dublin Core. It was acknowledged that a mapping from a database universe to fifteen predefined elements will cause loss of data.
Only part of the data in the original databases can be represented in Dublin Core. The problem was for each original data base to define how much and to which level data should be converted to Dublin Core.

The working group limited the requirements to:

- only data relevant to the public for search and presentation should be converted to the common format
- the requirement for representing details is not the same in a common database as in the sector specific ALM databases. But it is important for the user of the common data base to get a pointer (normally a link) to the original database, e.g. to retrieve more information, to ask for a copy or to send a loan request.

The working group didn't find alternatives to Dublin Core. But the group realized that mapping from more complex structures as used in museum and archive databases was not simple. So, the first report recommended making further investigations into this question and called for detailed examination of the mappings.

2.2 Exchange of data

The working group recommended XML as exchange format and chose the DCMI XML schema. This is also in accordance with recommendations from the Danish *National IT and Telecom Agency* attached to *Ministry of Science, Research and Innovation*. Concerning data transport, the working group found it important not to freeze to a particular model. The recommendations for content and exchange are not linked to a specific model.

Data transport can be handled in two different ways:

- search in the specific databases and establishing common presentation on the fly, including conversion on the fly
- search in one common data base, which implies data exchange beforehand

The conversion on the fly implies several problems. The defined conversions need to be converted including change of data (e.g. from code to text) and selecting only parts of databases. These kinds of requirements to be implemented on running systems are much more complex than regular export of data. The working group recommended a common database, but the metadata schema and XML schema work for a distributed model too.

2.3 Final report

Based on the first report a more thorough analysis was initiated which aimed at completing the mapping from the ALM-sectors' own formats to Dublin Core with the needed extensions, and at compiling a XML schema including the needed extensions. The working group drew upon consultants from The Royal School of Library and Information Science and from the company Index Data.

The starting point was some basic requirements:

mapping from the ALM-formats should point to one of the fifteen basic Dublin Core elements if possible
refinements of Dublin Core should only be used if absolutely necessary
other metadata formats should only be used if absolutely necessary
a new additional Danish metadata format should be the last option.

The consultants from the library school delivered input about mapping, and Index Data made the XML schema. Based on these contributions the working group made some further development and finished their report (3) in February 2006. This final report is out for hearing in the ALM-sectors with a deadline for comments mid-September 2006. Parallel with this process, the ALM working group is currently considering how to ensure continuous maintenance of the format and the XML schemas.

3. Common content format

The basic concept of mapping from domain format to Dublin Core is widely used. The Danish example for the ALM domains is subsequently presented and another example is for the geospatial domain (4).

The basis for mapping to Dublin Core is the ALM-sectors' specific formats for registration of collections. These formats are developed by the individual sectors partly based on - or inspired by - international standards. A consequence of this is different traditions for the selection of registration units as well as variances in registration levels, which can cause problems when converting to a common format.

For archives and museums the unit of registration is often done on a collection level, which means that one registration contains several units. The coherence of a collection can be due to many different circumstances: same excavation, same donor or just collected in the past by somebody. For libraries the collections are normally registered on a document level, describing identical units. An exception is multilevel published books.

The challenge was to map this hierarchical structure to the flat Dublin Core structure with a workable result for search and presentation.

One of the original goals for Dublin Core as metadata format was to support 'Resource discovery'. In the development of mapping, support of the sector-specific functionality has not been in focus. The focus has merely been on identifying the existence of an object. The consequence is a relative simplicity in the mapped format. To exchange data with a common system will be easy notwithstanding inequalities between the original registration formats. Another consequence can be that the users of the common system can encounter problems with interpretation of data because the original context is missing.

Mapping schemes to Dublin Core has been developed for four domain formats:

Daisy for governmental archives
Arkibas 4 for local archives
danMARC2 for libraries
Regin for museums

To ensure a fundamental level of retrieval in the system, it is important that a basic part of the original semantic retains are preserved. This has been an important parameter for evaluation of what kind a data have needed to use refinements to the fifteen basic Dublin Core elements.

One way to handle this problem is to add information about the source of the information. When the common system handles a record, the source of the record can be used to improve the presentation of data. To meet this requirement together with the requirement of linking to the original registration it is necessary to go outside Dublin Core. These two data elements are not descriptions of the resource, but information about registration of the resource – and then outside the scope of Dublin Core.

The AC - Administrative Components (5) was employed to meet these kinds of requirements. In part 7. the use of AC is further described.

Illustration of the dissimilarity based on source of the registration data:

> Using DC.Creator can be for different kind of originator. For museums creator can be the one responsible for composition of a museum's file. For a library the creator can be the author of a book, the composer of music or the band playing on a cd. For an archive the creator will often be the institution or part of the institution who established the archive. To make it possible to present the Creator-information in the right context it is necessary to use some refinements of Creator to preserve what kind of responsibility the described creators have. These refinements are described in part 5.
>
> Also using DC.Title will reflect different kind of titles. For libraries a title will normally be mandatory for all documents, but for archives and museums a title will often be constructed only for the common database – normally the source system does not contain a title in the same sense as a bibliographic title.

When trying to convert data for DC.Type and DC.Format some inconsistency must be anticipated. The different vocabularies seem not to cause problems because they reflect heterogeneous appearance forms.

Some of the basic fifteen elements are only used for libraries. DC.Publisher and DC.Language are not relevant as target for conversion because the Danish original databases for archives and museums do not contain these kinds of data. The lack of use for these elements is not assessed to give any problems for the use of a common database because they only exist for libraries.

In the model for the mapping of data from the different ALM-systems two levels of description are defined: a level for registration of collections and a level for registration of individual units. This way of splitting up into levels was chosen because archives and museums use these levels in their original registration. The different level is reflected in DC.Type using Collection for the first level and for museums normally Physical Object.

4. Mapping from ALM domains

4.1 Mapping from museums

The Regin data model consists of *primary entities* describing objects or concepts in the museums' domain and *secondary entities* describing properties and aspects of the primary entities. Furthermore the model consists of relation tables describing various connections between entities.

In the mapping most of the primary entities have been mapping to Dublin Core records:

- Archive file
- Magnetic mediums
- Case
- Photo
- Ships
- Objects
- Literature
- Reports
- Big formats.

The primary entities Case and Archive file are mapped to collection level and all the others to item level. The secondary entities are mapped as attributes to the above primary entities and together with two primary entities (Artist and Player).

4.2 Mapping from governmental archives

The Daisy data model used for governmental archives is a relational model and consists of three main parts:

- Agents
- Heuristic units
- Archive store units.

The occurrence of a heuristic unit causes a Dublin Core with the connected agents as DC.Creator's. The Daisy system doesn't contain keywords. To ensure reasonable possibilities for searching the names of the heuristic units are also mapped to DC.Subject and to DC.Title. Because the name of an agent often reflects a geographic area Agent is also mapped to Spatial as refinement of DC.Coverage.

4.3 Mapping from local archives

The format for the local archives is record-based and the mapping is from an Arkibas record to a Dublin Core record. The originator of an archive file is mapped to a DC.Creator refinement.

4.4 Mapping from libraries

Bibliographic records in danMARC2 are mapped to Dublin Core records. MARC records are supposed to be well known. Note that a top level record in a multilevel record is mapped to collection level.

5. dkdcplus
To handle the additions to DC and AC namespaces defined a namespace is and a XML schema called *dkdcplus*.

The specific Danish supplement contains three refinements for Creator, one refinement for Description and a list of values for Subject and a Danish-language version of DCMI type.

For Creator *preferredName* is used for the agreed version of a name. Several institutions have over time many names and variations of names and for archives it is suitable to collect documents from the same institution – also in presentation together with libraries and museums. As a logical consequence *alternativeName* is used for other variations and actPeriod to delimit the 'on duty' period.

The dumb-down test (6) for preferredName and alternativeName is easy. It is evident that different versions of names still at meaningful values of Creator. For actPeriod the values of start and end year of function for a creator together with the name of the originator/author/organizer/creator will be meaningful dumb-down together in the simple Dublin Core element Creator.

For Version/Edition the element *version* is defined as a refinement of Description. It is important for description of library books to distinguish between different editions, and it is no less important in presentation together with archives and museums.

To dumb-down content of version to Description is valid. Description includes all kinds of contents descriptions, and a version, edition etc will qualify a description of a resource.

The list Subject Type is tied up to DC.Subject. This list includes ontologies, classifications system and thesauri from the three sectors.

A reason for making a schema dkdcplus is to establish a Danish reference point for extensions to existing metadata schemas to ensure the interoperability between Danish metadata systems. This requirement has already showed validity because library projects in Denmark demand this facility.

6. The element 'Version'
Among the extensions to Dublin Core (fifteen elements and refinements) only version is a general refinement of Dublin Core.

Sometimes hasVersion is used for the edition, version, release etc. of the described resources. This refinement of DC.Relation doesn't meet the dumb-down test. DC.Relation is a link to another resource – not to the resource itself. The definition of DC.Relation is unambiguous: "A reference to a related resource".

Several application profiles based on Dublin Core have as one of the supplementary elements hasVersion, edition or other synonymous term. But none of these profiles are according to the working group's judgment relevant to use as international and

unencumbered reference source. The element version in MODS (7) is for example linked to bibliographic content.

To evaluate the general use of the element version the report of the IFLA Cataloguing Section on metadata (8) is used. This report lists ten core elements for metadata records and nine of these elements are analogous to Dublin Core elements.

Comparison IFLA metadata and Dublin Core	
IFLA metadata	*Dublin Core*
Subject	Subject
Date	Date
Conditions of use	Rights
Publisher	Publisher
Name assigned to the resource	Title
Language/mode of expression	Language
Resource Identifier	Identifier
Resource Type	Type
Author/creator	Creator
Version	[no equivalent]

By a mistake in appendix II in this IFLA report Dublin Core has a mark opposite Version indicating that this element is covered by Dublin Core. The reason for this mistake was that an earlier version of the DC Libraries Application Profile (9) included an element for Version. Originally version was a refinement of DC.Description, but in the actual version of this Application Profile a MODS element is used.

7. Administrative components

To handle connection between the original data in the ALM databases and the resulting Dublin Core records it has be vital to register this information about not the described resources but about the metadata itself. This kind of data is out of scope for Dublin Core. To solve this problem the metadata schema for administrative information about metadata Administrative Components (5) was selected.

Two of the AC elements are used:

AC.Identifier	Identification in original system. Can be used for linking to all information
AC.Source	Identify the delivering organisation/institution

8. The combined schema

Used elements for exchange of ALM data for shared presentation:

Element - refinement	Namespace
DC.Title	dc
- alternative	dcterms
DC.Creator	dc
- preferredName	dkdcplus
- alternativeName	dkdcplus
- actPeriod	dkdcplus
DC.Subject	dc
DC.Description	dc
- version	dkdcplus
DC.Publisher	dc
DC.Contributor	dc
DC.Date	dc
DC.Type	dc
DC.Format	dc
- extent	dcterms
- medium	dcterms
DC.Identifier	dc
DC.Source	dc
DC.Language	dc
DC.Relation	dc
- isPartOf	dcterms
- hasPart	dcterms
DC.Coverage	dc
- spatial	dcerms
- temporal	dcterms
DC.Rights	dc
AC.Identifier	ac
AC.Source	ac

The scheme Period from dcterms is used for ActPeriod and Temporal. A list of values SubjectType is used for DC.Subject. This list include the values DK5 (library classification), DBCS, DBCF and DBCM (library keywords) and SRKM (museums classification) in dkdcplus. For DC.Type is used the general list of values in DCMI-type and a Danish translation in dkdcplus.

9. XML Schemas

As a tool for interoperability a XML schema is composed collecting the XML schemas of the mentioned content formats. The name of this schema is dkabm – dk for Denmark, a for archives, b for libraries (biblioteker) and m for museums.

9.1 Schema survey

Names, URL's and description of the used XML schemas:

dkabm.xsd

(http://www.bs.dk/standards/schemas/dkabm_2006-01-13.xsd)
collecting schemas mentioned below.
dc.xsd
(http://purl.org/dc/elements/1.1/dc.xsd)
defines the fifteen basic Dublin Core elements.

ac.xsd
(http://www.bs.dk/standards/schemas/ac_2005-09-01.xsd)
defines Dublin Core Administrative Components.

dkdcplus.xsd
(http://www.bs.dk/standards/schemas/dkdcplus_2006-01-13.xsd
defines Danish elements and subject lists.

dcterms_ext.xsd
(http://www.bs.dk/standards/schemas/dcterms_ext_2006-01-13.xsd)
import Danish elements from dkdcplus.xsd together with dcterms and dc

dcterms.xsd
(http://purl.org/dc/terms/dcterms.xsd)
defines Dublin Core refinements.

dcmitype.xsd
(http://purl.org/dc/dcmitype/dcmitype.xsd)
defines Dublin Core resource types.

dktype.xsd
(http://www.bs.dk/standards/schemas/dktype_2003-05-01.xsd)
defines Danish language of DCMI types.

ISO639-2.xsd
(http://www.ddb.ds/standards/xmetadiss/ISO639-2.xsd)
defines valid language codes.

9.2 The XML dkabm schema

The completely *dkabm* XML schema:

```
<?xml version="1.0" encoding="UTF-8" ?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
 xmlns:dkabm="http://www.bs.dk/standards/#dkabm" xmlns:ISO639-
 2="http://lcweb.loc.gov/standards/iso639-2/" xmlns:ac="http://www.bs.dk/standards/#ac"
 xmlns:dkdcplus="http://www.bs.dk/standards/#dkdcplus"
 xmlns:dcterms="http://purl.org/dc/terms/"
 targetNamespace="http://www.bs.dk/standards/#dkabm" elementFormDefault="qualified"
 attributeFormDefault="unqualified">
<xs:annotation>
<xs:documentation>XML Schema for http://www.bs.dk/standards/#dkabm namespace This XML
 schema shall support the Archives-Libraries-Museum standard working group Dublin Core based
 format for common presentation of records from archives, libraries and museums. Created in
 accordance with the naming convention used in the Dublin Core Metadata Initiative (DCMI)
 Created 2005-09-09 by Per Moerkegaard Hansen (perhans@indexdata.dk) with assistance from
 Soeren Pedersen (sorenpedersen@makeitright.dk)</xs:documentation>
</xs:annotation>
```

```
<xs:import namespace="http://purl.org/dc/terms/"
 schemaLocation="http://www.bs.dk/standards/schemas/dcterms_ext_2006-01-13.xsd" />
<xs:import namespace="http://www.bs.dk/standards/#ac"
 schemaLocation="http://www.bs.dk/standards/schemas/ac_2005-09-01.xsd" />
<xs:import namespace="http://www.bs.dk/standards/#dkdcplus"
 schemaLocation="http://www.bs.dk/standards/schemas/dkdcplus_2006-01-13.xsd" />
<xs:import namespace="http://lcweb.loc.gov/standards/iso639-2/"
 schemaLocation="http://www.ddb.de/standards/xmetadiss/iso639-2.xsd" />
<xs:element name="metadata" type="dcterms:elementOrRefinementContainer" />
</xs:schema>
```

9.3 dkdcplus XML schema

The XML schema for *dkdcplus* is designed by copying methods from the DCMI XML schemas and among these the specification of SubjectType is modelled after DCMI Type. The substitutionGroup for the refinements to DC.Creator is Creator for Version Description.

Extract of dkdcplus XML schema:

```
<xs:element name="preferredName" substitutionGroup="dc:creator" />
<xs:element name="alternativeName" substitutionGroup="dc:creator" />
<xs:element name="actPeriod" substitutionGroup="dc:creator" />
<xs:element name="version" substitutionGroup="dc:description" />
```

10. Conclusion

The analysis of data in the Danish ALM databases revealed that the basic fifteen DC elements solve many problems when converting data to a top level. Several of the Dublin Core refinements are useful, but most of the data relevant for search and presentation for the interested public is covered by the fifteen elements. But some information cannot be handled satisfactory without special refinements. The Danish ALM-project introduced four new refinements. Three of them are domain metadata with limited relevance outside ALM and Denmark. Only the version refinement had a broader relevance and is missing as part of Dublin Core refinements.

Dublin Core has three different roles in practical work:

the original: a simple schema to register Internet documents
the extended: as basis for project and domain specific registration schemas
interoperability: as basis for interchange of information between domains

This paper shows a way to use Dublin Core for interoperability between three related sectors. Arguably the sectors are related but with major differences. This indicates that Dublin Core has a more general role as carrier for interoperability.

11. References

1. NOKS – Nordjyllands kulturhistoriske søgebase (The North Jutland history of civilization search data base)
   http://www.noks.dk/
2. Afrapportering fra arbejdsgruppen om ABM-standarder til styregruppen, august 2003 (Danish)

http://www.kulturarv.dk/tjenester/publikationer/emneopdelt/kulturarv-it/abm2003/index.jsp
3. Fælles præsentation af data fra arkiver, biblioteker og museer på internettet. Rapport fra arbejdsgruppen om ABM-standarder, februar 2006 (Danish)
http://www.kulturarv.dk/tjenester/publikationer/emneopdelt/kulturarv-it/abm2006/index.jsp
4. CWA 14857 - Mapping between Dublin Core and ISO 19115, "Geographic Information – Metadata", 2003
http://www.cenorm.be/cenorm/businessdomains/businessdomains/isss/cwa/cwa14857.asp
5. AC - Administrative Components. Dublin Core DCMI Administrative Metadata
http://www.bs.dk/standards/AdministrativeComponents.htm
6. DCMI Grammatical Principles / DCMI Usage Board, 2003.
http://dublincore.org/usage/documents/principles/
(reference to part 2.3. Dumb-down Principle).
7. MODS. Metadata Object Description Schema
http://www.loc.gov/standards/mods/
(sub element edition to element originInfo)
8. Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections. Report of the IFLA Cataloguing Section. Working Group on the Use of Metadata Schemas
http://www.ifla.org/VII/s13/guide/metaguide03.pdf
9. Library Application Profile
http://dublincore.org/documents/2004/09/10/library-application-profile/