The use of Metadata in Denmark

Leif Andresen
Library Advisory Officer
Danish National Library Authority
lea@bs.dk

Abstract

A wide range of metadata projects in Denmark over the last 4-5 years are described.

A new Danish legal deposit act in 1997 facilitated co-operation in the creation of a common application form for Danish Dublin Core including the basic fifteen elements and four sub-elements. The form is used for creation of metadata in government publications and as application form for legal deposit and inclusion in the national bibliography.

Then several public libraries initiated projects developed with different views on how to make the Internet accessible for the common user. Not all the projects use Dublin Core from the start but that will change. Actually some research library based subject gateways have been launched and all of these projects use Dublin Core based on a common solution. In 2001 a very ambitious governmental information project was started, where the ambition is to cover all documents on governmental and municipal servers on the net. Finally are described the efforts to gather all registrations from the projects to be included in the National Union Catalogue.

Keywords: *Denmark, National Union Catalogue*

1. Introduction

In 1997 the Danish National Library Authority decided as its general policy that Internet documents should contain metadata as part of the document itself - and not only as cataloguing of parts of the Internet documents in the library catalogues. This policy is being implemented in a number of national initiatives as well as in positive efforts in the areas of international standardisation.

This happened in response to the increasing volume of electronic publications, which was already at that time becoming apparent. In various contexts a need emerged for the cataloguing and description of documents published on the Internet. Several projects had been launched in Denmark - which originally were not really co-ordinated.

Before January 1998 three institutions had implemented application forms for registration. Two of them also offer the possibility of generating metadata to include in HTML-files.

The institutions were:

- The Royal Library (for legal deposit)
- Danish Library Centre (for national bibliography)
- The Danish State Information Service (for government publications).

2. Legal deposit

All Danish static documents on the Internet are since 1st January 1998 subject to legal deposit. Static documents are publications, which the publisher does not intend to alter. Producers of these documents must register them on an electronic application form on The Royal Library's web server and provide them with the necessary access codes and other relevant information needed in order to access the works. This form is now based on Danish Dublin Core [1].

One question presents itself: Why only static documents? The reason for not yet including dynamic documents is the need for further clarification both concerning the legal aspects and the technical problems associated with collection, preservation and use.

The legal deposit act is actually under discussion. The main focus is the need for a broader preservation of documents on the net.

3. The National Bibliography: InDoReg

In 1996 The Danish Library Centre launched a project, InDoReg (Internet Document Registration) [2] to clarify some fundamental questions concerning bibliographic control with Danish Internet documents. The project focuses on the following areas in particular: Inclusion criteria, cataloguing procedure, metadata, PURL, and publication of registrations.

Proposed principles for national bibliographic inclusion criteria were drawn up. They operate with the concepts of static and dynamic publications - with homepages as an independent category under

dynamic publications. These principles reflect the criteria that exist for publications in fixed physical form, since there are formal requirements with regard to both size and (to a certain extent) content. For instance, it is proposed that publications of a commercial, internal, highly local or private nature should not be included.

A follow-up was published in 1999: *InDoReg 2* [3]. Cataloguing of static and dynamic publications and electronic periodical articles, which describes the experiences gained from the work with Internet publications during recent years. In addition the report contains a completely separate part, dealing with the description of electronic articles.

The report operates with three types of publications:
- Static publications: corresponding to traditionally published books, pamphlets, reports, etc.
- Dynamic publications: are ongoing like databases, homepages, etc.
- Electronic periodical articles.

The work described has accomplished that an increasing quantity of Internet publications relevant to libraries is included in the Danish National Bibliography.

The new act on legal deposit and the InDoReg report formed the basis for an extension of the national bibliography in 1998, which meant the inclusion of static electronic documents. The agreement on the extension of the national bibliography also contained further development work on dynamic publications.

4. Government information policy

The Danish Ministry of Research and Information Technology and The Danish State Information Service has worked out a standard describing how government publications on the Internet should be encoded. The standard was published *in Net Publications* [4] in 1997. All new printed publications issued by Danish ministries, government offices and agencies must also be published on the Internet.

The publication includes guidelines on how to produce web publications. One recommendation is always to have a HTML page with metadata - human and machine-readable - even if the resource itself is PDF. Metadata embedded in the resource itself is mandatory. The metadata format was not Dublin Core, but was a specific format inspired by Dublin Core. The Danish Dublin Core has replaced this format.

5. A common application form

As part of the preparations for the implementation of the new legal deposit act an 'Application form committee' was formed. The goal for the committee was co-operation in connection with the implementation of legal deposit of static electronic publications. The main object was co-operation on cataloguing - especially on government publications - and to promote the use of metadata embedded in the resource itself.

We worked out a common definition of an application form based on Dublin Core. A basic idea in the design was that when you fill in the application form the programme copies metadata (if any) from the resource into the form. People should only have to type the information once!

There is an application form at The Royal Library (for legal deposit of static documents) [1]. Notice that this project is based on the idea of metadata embedded in the resource itself - not on databases with metadata.

There is an application form at the Danish Library Centre (for dynamic documents) [5] and it works as metadata generator (for both dynamic and static documents). These two forms have references to each other.

The application form of The Danish State Information Service is replaced by reference to The Royal Library.

The interesting point is that it has become possible to minimise the volume of information that a number of different institutions need.

The original format defined by the Danish State Information Service had a lot of specific tags. But it became apparent that even people who worked in government institutions had difficulties in working out how to fill in a specific form. It was necessary to lower one's level of ambition in order to encourage a more widespread use of metadata. This is particularly true when official policy dictates that government institutions must publish their publications electronically and enhance them with metadata.

6. Danish Dublin Core

We decided to use the fifteen basic Dublin Core elements. It is important to realise that this is not a Danish version of Dublin Core. It is a recommended subset of Dublin Core.

In order to consider some special needs in the recording of these documents it became necessary to use more than the 15 Dublin Core elements. This happens by bringing in four sub-elements.

There is specified one sub-element *DC.Title.Release* to edition/version/release as it is in the resource: '1 ed', 'Release 2.1.3' or 'Revised Edition'. This qualifier is never adopted in spite of several attempts to find an internationally acceptable model.

Three different applications of DC.Relation have been specified:

- *IsVersionOf*: Specifying parallel printed edition. A special need of The Danish State Information Service
- *HasPart*: List of files in the resource (a file list). This is a special need of The Royal Library. They need to know the files they have to copy to their server
- *Requires*: List of program(s) necessary for using the resource in question. Especially The Royal Library needs this information to know what software to use to read.

7. Guidelines

To support use of Danish Dublin Core the Danish National Library Authority has published *Metadata in Danish net publications. Technical description* [6]. The target group is Web designers and major publishing houses with a view to supporting the generation of metadata as an integrated part of the creation of a document. The idea is that only if the creation of metadata is integrated with the creation of the document will it become the common thing to do.

This publication was followed by *Metadata in Danish net publications. Guidelines* [7]. The target group is anyone who has to create metadata as an integrated part of the creation of documents.

Both publications contain a description of Danish recommended use of Dublin Core.

We have been working on a comprehensive Danish solution to the problem with a unique identifier for electronic documents. The draft solution would mean that already known unique identifiers like ISBN and DOI will be reused in a URN and publications without unique identifiers will be allocated a special URN based on the national bibliographic number.

As a result of these two additions to the Danish model for the handling of metadata in net publications it has been decided that the two publications on metadata in Danish net publications will be published in a new edition later this year. A campaign to encourage more authors and publishers of net publications to add metadata to their publications is going to be launched at the time of publication.

8. E-government use of metadata

Early this year a new and ambitious project in governmental information in Denmark emerged. The project is called OIO - Open public Information Online. The main object is easy access to information published by the Danish public sector. Other goals are easy management, interoperability of information and linking of coherent information.

The project covers not only the governmental level of public information, but also regional and local information. The National Association of Local Authorities in Denmark and the Association of County Councils in Denmark are involved in the project, too. It is funded by the Danish Ministry of Information Technology and Research and is implemented by The Danish State Information Service.

The first part of the project has been a study of state of art for metadata in Denmark including a study of relevant international standards and de facto standards. Based on this a proposal has been worked out for more comprehensive Danish Metadata Core including encoding schemes. This is based on Dublin Core supplemented by local extensions. The extensions are particularly made for administrative use.

Another part of the project is a link-service application. The idea is to develop a 'more relevant information' button to provide deep linking between public web pages based on metadata. The basis for this function is a classification scheme with about 500 headings on three hierarchical levels.

The project includes a unique combination of metadata embedded in the resources and a metadata-based search engine. A metadata harvester and the classification scheme secure the coherence.

All the tools are to be completed during summer 2001 and in the early autumn work starts on implementation including comprehensive work on contacting all the public information providers.

One of the key persons in the project is Palle Aagard from The Danish State Information Service. He is also co-chair of DCMI Government Working Group.

9. Projects as part of DEF - Denmark's Electronic Research Library

The DEF-project covers a wide range of initiatives of the digital library for research libraries in Denmark. One very visible to the public is the DEF Directory [8], which covers all subjects at the general level. To cater for subject specialists requiring in-depth access within narrow subject fields, some subject gateways [9] are currently under development. These gateways are being developed cooperatively by research libraries working together in consortia, which complement each other in their respective subject fields. Already in operation are:

- BiziGate: access to Danish and foreign business economics web sites, etc. within management, marketing, companies' trades and statistics

- Virtual Music Library: access to Danish and foreign web sites, Danish special collections and a few sound tracks and printed music

- Clinical Information: specialist information targeted at Danish health professionals.

Four other gateways will be launched during the second half of 2001:

- Food: a guide to nutrition and food science and technology

- Energy: access to key information about energy and energy research

- Traffic: access to traffic and transport research web sites

- ARAKADE: a guide to art, architecture and design.

Both the DEF Directory and the subject gateways use the same metadata tool and same metadata scheme. The scheme is an implementation of Dublin Core including some qualifiers complemented with a special DEF Name Space. This name space includes several administrative metadata such as date for next verification, metadata creator and comments. The Virtual Music Library uses a few extensions to Dublin Core: qualifiers for composer, lyricist and conductor.

10. Projects of public libraries

A number of Danish public libraries have individually - or together - developed a number of net services. Some of these services, which are of a national character, are presented with links on the front page of bibliotek.dk. All of them are listed at a next level page [10]. The net services include:

- BibHit: a minor number of quality-checked web sites as supplement to the other materials in the public libraries

- DotBot: childrens library on the net – including 'Kloge Åge' ('Clever Dick') with links for children

- FINFO - Information for ethnic minorities in Denmark: general information on which rights, obligations and options ethnic minorities have in Denmark as well as references to other countries.

- Folkebibliotekernes Netguide [Netguide of the public libraries]: subject-related access to 3500 selected web-sites within the subjects: sport, diseases, the world, job searching etc.

Another project is *NetKat*, where several libraries include Internet documents as part of their online catalogue. Dansh Bibliographic Center does the cataloguing and the libraries make the selection. The cataloguing records are part of the Danish National Union Catalogue.

The projects of public libraries have chosen different solutions for registering web pages. The BibHit project was from the beginning using Dublin Core with some extensions, e.g. a more detailed list of values of DC.Type. The FINFO project is after a recent reconstruction also based on Dublin Core.

The Netguide and the DotBot projects use their own internal format based on an application form. But it can easily be translated to Dublin Core and they plan to do it in the near future. The records of NetKat are registered in the Danish national MARC-format danMARC2.

11. Integration in National Union Catalogue of metadata from projects

Since March 2001 several net services developed by Danish libraries have been integrated in bibliotek.dk [11] - the Danish National Union Catalogue accessible to everyone free of charge. There are links to these services from the front page of *bibliotek.dk*. The next step in integration is import of registrations of net resources into bibliotek.dk.

The idea is that bibliotek.dk will covers a wide range of Internet resources, quality- checked by Danish libraries. First the plan is to import data from the projects listed above. This is a challenging job. The projects of the public libraries used different formats

and the level of registration is different. For instance the level in BibHit is very detailed and maybe these regsitrations will have to be compared with regsitrations of articles.

A plan and a time schedule for implementation of this integration was decided in September 2001.

12. International standardisation

The decision to play an active part in securing international standardisation of Dublin Core has at European level meant the participation in two workshops in the European standardisation organisation CEN. First in CEN/ISSS Workshop on Metadata for Multimedia Information (MMI) and now in CEN/ISSS Workshop on Metadata for Multimedia Information - Dublin Core (MMI-DC) [12], where the Danish National Library Authority acts as secretariat. The Dublin Core version 1.1 was endorsed in January 2000 as a CEN Workshop Agreement.

CEN Workshop Agreements (CWAs) are consensus-based specifications, drawn up in an open workshop environment. They are developed along straightforward lines, with a minimum of bureaucratic rules. A CWA reflects the consensus of identified companies and organisations responsible for its contents. The CWA therefore does not represent as high a level of consensus as a European standard and is not designed to support legislative requirements. Instead, its purpose is to offer market players a flexible and timely tool for achieving a technical agreement.

Parallel to these activities the American standardisation organisation NISO has now finished the final balloting on a similar formal standardisation of Dublin Core (Z39.85). The process is expected to end early this autumn. The next step will be standardisation at international level in ISO - International Organisation for Standardization, Technical Committee 46 Information and Documentation. In all likelihood this will happen by turning the future NISO Z39.85 standard with Dublin Core into an ISO standard. It is my assessment that it will be possible by the so-called fast-track procedure in ISO TC46.

13. Conclusion

When the Danish National Library Authority defined a metadata policy as mentioned in the introduction, focus was centered on embedded metadata. The development over the past four years in Denmark as well as in the rest of the world has shown that Dublin Core, as a tool for metadata registration projects is widely accepted and more powerful than embedded metadata. But exactly the widespread use of Dublin Core metadata in registration projects makes it more relevant to control the registration of your web site by using embedded metadata in standard format.

All projects mentioned in this paper use different solutions for administrative metadata for the metadata records. This emphasizes the need for the work started at DC-8 for a detailed scheme for administrative metadata. A proposal was early September 2001 sent for discussion in DCMI Administrative Metadata Working Group [13].

The mentioned Danish projects emphasised the need for a mechanism for simultaneously general metadata as Dublin Core, local/national/project metadata and administrative metadata. The mechanism must be the function of the metadata application profile, which was one of the major steps forward in 2000 in the ongoing development of metadata.

The Danish National Library Authority maintains a web page in English with updated information and links about metadata in Denmark [14].

References

[1] Reference to the Legal Deposit web site (Danish text)
http://www.pligtaflevering.dk/

[2] InDoReg: INternet DOcument REGistration. Project report, 1999
http://purl.dk/rapport/html.uk/

[3] INDOREG 2: Registration of static and dynamic publications and electronic periodical articles. Project report, 1999
http://www.dbc.dk/english/indoreg/

[4] Netpublikationer. Statens standard for elektronisk publicering, 1999
http://www.si.dk/vejl/publ/netpubl2/

[5] INROREG application form, see:
http://purl.dk/metadata/meta_lang.htm
and a shorter version:
http://purl.dk/metadata/meta_kort.htm

[6] Metadata i danske netpublikationer. Teknisk beskrivelse (Danish text), 1999
http://www.bs.dk/publikationer/vejledninger/2/index.htm

[7] Metadata i danske netpublikationer. Brugervejledning,
(Danish text), 1999
http://www.bs.dk/publikationer/vejledninger/3/

[8] DEF Directory
http://www.deff.dk/vejviser/?lang=eng

[9] DEF Portal >> Subject Portals
http://www.deff.dk/vejviser/?p=subportals&lang=eng

[10] www.library.dk. See also:
http://www.library.dk/links.htm

[11] bibliotek.dk has an English user interface at:
http://www.library.dk

[12] MMI-DC. Meta-data for Multimedia - Dublin core
http://www.cenorm.be/isss/Workshop/MMI-DC/

[13] DCMI Administrative Metadata Working Group
http://dublincore.org/groups/admin/

[14] Dublin Core in Denmark
http://www.bs.dk/metadata/english.htm