

XMetaDiss - Metadata for ETDs in Germany

Authors: Frodl, C., Korb, N., Wollschlaeger, T.

Die Deutsche Bibliothek, Frankfurt am Main, Germany 60322

Tel. +49-69-1525-1404

Fax. +49-69-1525-1010

Mail : dissonline@dbf.ddb.de

Abstract: In the project a new metadata set "XMetaDiss" for online university theses and dissertations (ETDs) is being developed. This metadata set is supposed to relieve the set "MetaDiss". XMetaDiss is based on DCMI of 2004 and will be used in Germany for interchange of metadata and corresponding full text of online university theses and dissertations (ETDs). The potential of XMetaDiss consists in the use for an automatic compiling-process of metadata by OAI protocol, the targeted compatibility with the NDLTD set ETD-MS, the use of hierarchical patterns and the simple transformation potential by means of XSLT into other metadata formats. The embedding within an international framework points to a high acceptance of a new metadata format for ETDs, corresponding to the state of the art.

Keywords: digital library; metadata applications; ETD; online theses; OAI; transfer format; XMetaDiss; Germany

1 Introduction

Since 1998 a metadata set MetaDiss, based on DCMS of 1998, is being used in Germany for interchange of metadata and corresponding full text of online university theses and dissertations (ETDs).

The great majority of German university libraries (73 of 100 possible ones) participate in this process.

In the project a metadata set XMetaDiss for ETDs is being developed and will be defined by means of XML DTD as well as XML Schema. This metadata set is supposed to relieve the set MetaDiss which is up to now embedded into HTML 4. The embedding in HTML 4 does not represent the state of the art for the data exchange anymore.

The potential of XMetaDiss consists in

- the use for an automatic compiling-process of metadata of ETDs by OAI protocol,
- the targeted compatibility with the NDLTD set ETD-MS
- the use of hierarchical patterns and

the avoidance of allocation errors (up to now the allocation occurred only through the sequence of the elements)
- and the simple transformation potential by means of XSLT into other metadata formats, as ETD-MS and DC Simple.

2 XMetaDiss Framework

XMetaDiss is based on MetaDiss and is compatible to it. It contains all content descriptive and administrative data elements of the MetaDiss format. XMetaDiss is downward compatible to the "Electronic Theses and Dissertations" metadata sets (ETD-MS) and DC Simple (OAI_DC).

XMetaDiss can be embedded by means of XSLT transformation into a Resource Description Framework (RDF) structure.

The use of XML allows hierarchically structured data that facilitate a correct automatic evaluation, transformation and if necessary conversion. With metadata according to ETD-MS German online university publications can be integrated via international metadata search engines (NDLTD). The inclusion in an international framework, depending on sufficient agreement with libraries and library service centres in Germany, points to a high acceptance of a new metadata format for online university theses and dissertations, corresponding to the state of the art.

The definition of XMetaDiss is being made both by DTD and Schema.

XML DTDs have their own clear syntax that differs from the XML syntax.

Other than DTDs, XML Schemas (defined by XML syntax) have the advantage of being developed considerably further but more complex, and therefore offer more chances to define very exactly the content and the structure of document classes. The application of Schema also offers the chance to execute more complex automatic plausibility checks as they'd be manageable with DTDs.

At first, metadata producers will receive both types of document definitions. The long-term priority lies within Schema.

3 Metadata Set Description

3.1 Usage

The basis for the data elements described in the XMetaDiss format is the Dublin Core Metadata Element Set. However, the Dublin Core Set does not meet all requirements considered necessary to exploit dissertations and post-doctoral theses. Therefore, within the Dublin Core range the metadata set has been expanded for several positions: for the metadata set of the Networked Digital Library of Theses and Dissertations ETD-MS (Electronic Theses and Dissertations Metadata Set), for metadata for long-time preservation on the basis of the New Zealand National Library Preservation Metadata, and for a set to register personal related data

(METAPERS).

The metadata set can be created by the online publication's author or by the appropriate university's office or the university library. The data set can be used for administrative purposes within the individual university. Moreover, it can be stored on the university library's server (or other local servers, respectively) as an evidence for the online publication. The data set is being used for the data transfer to Die Deutsche Bibliothek (the German National Library) as well. There it will be used during the workflow and will be provided at the frontpage of the online publication for enquiries via the internet. Furthermore, the metadata set can be used for data exchanges in the international context (e.g., via OAI). The data on authors will be added to the German Personal Names Authority File (PND).

3.2 Structure

XMetaDiss consists of a bibliographical part that contains all fifteen Dublin Core elements including refinements and qualifiers (DC Qualified). Additional elements were taken from ETD-MS, as for example "degree". Below dc:creator, metadata on persons were inserted, for example the distinction of fore- and surname (mandatory) and the date of birth (optional). Below dc:publisher, metadata on the publishing institution(s) were inserted, for example the distinction between name and place of the institution. This is for

the metadata to meet the requirements of the German Personal Names Authority File (PND) and the German Corporate Names Authority File (GKD).

In dc:identifier, the Uniform Resource Name (URN) of the document is stated as it has been assigned by Die Deutsche Bibliothek according to the URN namespace "national bibliographic number" (urn:nbn).

A further part of XMetaDiss consists of technical and administrative information. For the administrative information, for example a contact identification number, a special namespace has been defined. The technical information is based on the New Zealand National Library Preservation Metadata and contains format specifications with release, file size and file directory information etc.

If the element name is introduced by "dc", the element derives from the Dublin Core Set. Personal related data are indicated with "pc". Data concerning corporate authorities have "cc" that stands for "corporate core". The names of elements used for the workflow and for the additional elements for long-term preservation are introduced by "ddb". Elements derived of ETD-MS belong to the namespace "thesis".

4. Perspective

XMetaDiss was supposed to be available in the course of the summer 2004 in a usable release. Accordingly,

on June 1, the reference description of the XMetaDiss version 1.0 has been released to inform all libraries, universities and other parties involved in the online publication workflow in Germany as well as in other interested countries. It is planned to use XMetaDiss for the data exchange of online university publications within Germany by OAI protocol.

Furthermore, an international data exchange with the dissertation database of NDLTD is possible via automatic conversion to ETD-MS. A conversion to Dublin Core can easily be realised, too. That would achieve the aim of an international applicability. Simultaneously, the use of technical and preservation metadata in XMetaDiss sets the prerequisite of the long-term preservation and hence the long-term availability of online university publications.

References

1. <http://www.dublincore.org>
2. <http://www.ndltd.org/standards/metadata/current.html>
3. <http://www.openarchives.org/OAI/openarchivesprotocol.html>
4. http://www.natlib.govt.nz/files/nlnz_data_model.pdf
5. <http://deposit.ddb.de/metadiss.htm>
6. http://www.ddb.de/standards/xmetadiss/ref_xmetadiss_v1_e.pdf
7. <http://www.ddb.de/professionell/pnd.htm>
8. <http://www.persistent-identifier.de/>
9. http://www.ddb.de/professionell/dissonline_e.htm

10. <http://www.langzeitarchivierung.de>