

XMetaDiss meets ETD-MS

Christine Frodl

Co-ordinator of Metadata Standards, Office for Library Standards,
Die Deutsche Bibliothek, Frankfurt am Main

Nikola Korb (Ph.D.)

Head, Co-ordination Agency DissOnline,
Die Deutsche Bibliothek, Frankfurt am Main

Keywords: XMetaDiss, ETD-MS, online theses, metadata, Germany

ABSTRACT

The XML-based metadata set "XMetaDiss" will be used in Germany for interchange of metadata and corresponding full text of online university theses and dissertations. XMetaDiss is downward compatible to the "Electronic Theses and Dissertations" metadata set ETD-MS and to Dublin Core Simple (OAI_DC). With metadata constructed according to ETD-MS German online university publications can be integrated in international metadata search engines like the ETD Union Catalog. For applying the international metadata standard of the Networked Digital Library of Theses and Dissertations ETD-MS, a mapping from XMetaDiss to ETD-MS is being developed. For the transformation of XMetaDiss to ETD-MS with a XSLT processor, a XSLT file is provided.

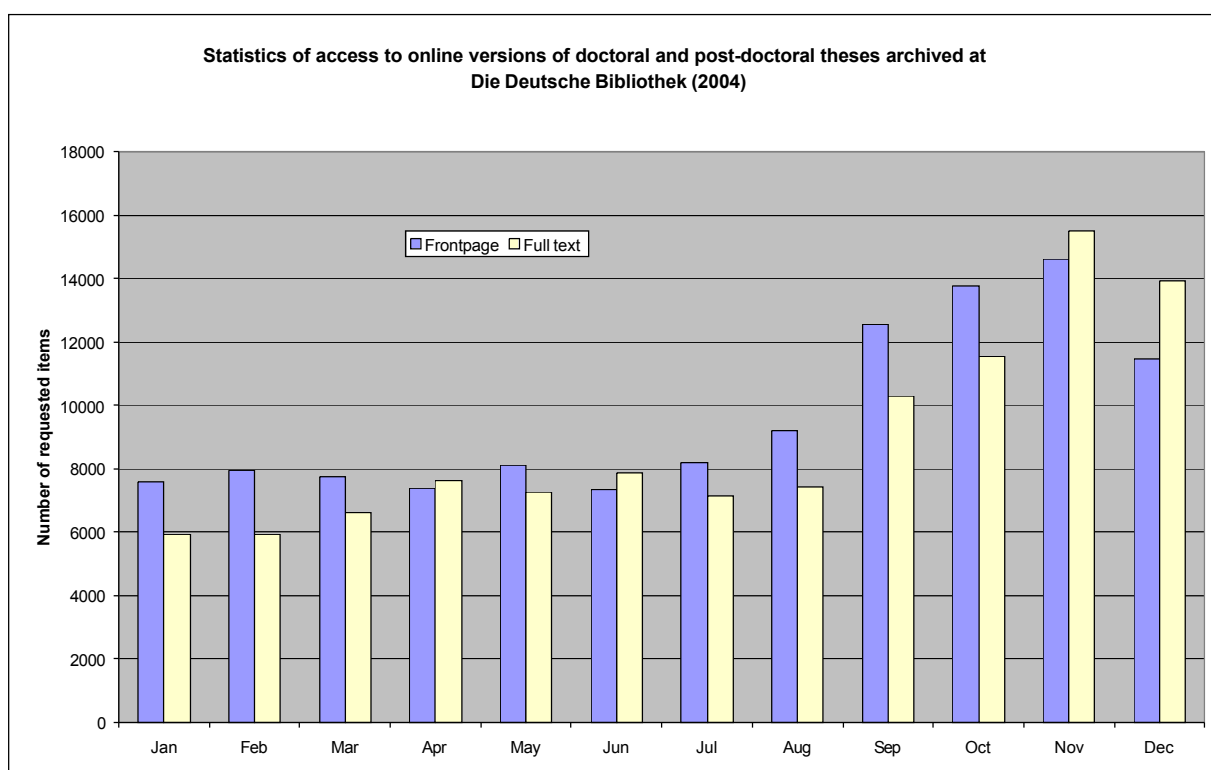
1. INTRODUCTION

Since July 1998 Die Deutsche Bibliothek (German National Library)¹ has been collecting online dissertations and postdoctoral theses while applying the accompanying metadata to indexing and making them available for further utilisation. Die Deutsche Bibliothek is responsible for the metadata formats MetaDiss² and XMetaDiss³ and the transfer interface for online dissertations. It is also responsible for the long-term access⁴ to electronic publications. Die Deutsche Bibliothek applies its experience to the development of metadata standards in the "Arbeitsstelle für Standardisierung" (Office for Library Standards)⁵. In this context, Die Deutsche Bibliothek has participated in the project "Dissertation Online"⁶ financed by the Deutsche Forschungsgemeinschaft (German Research Foundation)⁷ from 1998 to 2000. The results from the seven project parts can be found via http://www.dissonline.de/index_e.htm. Following the recommendations of the project "Dissertation Online", the "Koordinerungsstelle DissOnline" (Co-ordination Agency DissOnline)⁸ at Die Deutsche Bibliothek was founded in February 2001. The Co-ordination Agency DissOnline is responsible for co-ordination and counselling of institutions participating in the DissOnline procedure. To keep the outcome of the project "Dissertation Online" state-of-the-art, to initiate necessary further developments, and to build up a network of competence, the German Research Foundation supported the project "Building-up a Co-ordination Agency for online dissertations and post-doctoral theses" from January 2003 to December 2004.

Since 1998 the metadata format "MetaDiss", which had been developed within the project "Dissertation Online" is being used in Germany for interchange of metadata and the corresponding full text of online university theses and dissertations. This format is based on DCMES⁹ of 1998 and on HTML 4. The embedding in HTML 4 does not represent the state of the art for the data exchange anymore. Therefore it was necessary to develop a new and more modern format. In cooperation with university libraries and scientific associations in

Germany, the Co-ordination Agency DissOnline and the Office for Library Standards at Die Deutsche Bibliothek have developed a new metadata set for online university theses and dissertations, named "XMetaDiss". XMetaDiss is an XML-based format and refers to DCMES of 2004. The potential of XMetaDiss consists, among other things, in its use for an automatic compiling-process of metadata by OAI protocol, the targeted compatibility with the NDLTD set ETD-MS¹⁰ and technical metadata for long-term preservation. XMetaDiss is downward compatible to the "Electronic Theses and Dissertations" metadata sets (ETD-MS) and Dublin Core Simple (OAI_DC).

To point out the relevance of this effort, note that the annual increase of new electronic theses and dissertations amounts to 10.000 items. The figure below shows the number of times online theses and dissertations were accessed within the year 2004. By January 2005 Die Deutsche Bibliothek had archived more than 30.500 ETDs.¹¹



2. METHODS

Metadata are needed for the interchange of metadata and corresponding full text of online university theses and dissertations. Because it was an aim, to use international standards, the basis for the data elements described in MetaDiss, and therefore in XMetaDiss, is the Dublin Core Metadata Element Set. Because the Dublin Core Set does not meet all requirements considered necessary to exploit dissertations and post-doctoral theses, XMetaDiss falls back on other international standards like the metadata set of the Networked Digital Library of Theses and Dissertations ETD-MS, the metadata set for long-term preservation on the basis of the New Zealand's National Library Preservation Metadata¹², and a set to register metadata on personal names called "XMetaPers"¹³. XMetaDiss also takes advantage of persistent identifiers¹⁴ like URN, DOI and others. The persistent identifier URN can be assigned at Die Deutsche Bibliothek.

3. RESULTS

With metadata constructed according to ETD-MS German online university publications can be integrated in international metadata search engines like the ETD Union Catalog. To apply the international metadata standard of the Networked Digital Library of Theses and Dissertations ETD-MS, a mapping¹⁵ from XMetaDiss to ETD-MS was being developed.

Table 1 provides a mapping between XMetaDiss version 1.2 and ETD-MS version 1.00, revision 2.

Table 1. Mapping XMetaDiss version 1.2 – ETD-MS Version 1.00, revision 2

XMETADISS	ETD-MS	Notes
<dc:title >	<title>	repeatable, Mapping of attribute lang (ISO-639-2) in XmetaDiss to xml:lang (ISO-639-1) in ETD-MS
<dcterms:alternative>	<alternative Title>	repeatable, optional, Mapping of attribute lang (ISO-639-2) in XmetaDiss to xml:lang (ISO-639-1) in ETD-MS
<dc:creator> <pc:person><pc:name><pc:surName>, <pc:foreName>	<creator>	repeatable surname, forename(s)
<dc:subject>	<subject>	repeatable, scheme in ETD-MS results from xsi:type in XMetaDiss
<dcterms:abstract>	<description>	repeatable, optional, Mapping of attribute lang (ISO-639-2) in XmetaDiss to xml:lang (ISO-639-1) in ETD-MS
<ddb:note>	<description role="note">	repeatable, optional
<dc:publisher><cc:universityOrInstitution><cc:name><cc:place>	<publisher>	repeatable, in ETD-MS optional
<dc:contributor thesis:role="advisor", thesis:role="referee", thesis:role="committeeMember", thesis:role="chair", thesis:role="co-chair"> <pc:person><pc:name><pc:surName>, <pc:foreName>	<contributor role="advisor", role="referee", role="committee_member", role="chair", role="co-chair">	repeatable, optional
<dcterms:issued> oder <dcterms:dateAccepted>	<date>	dcterms:issued date of first publishing dcterms:dateAccepted date of doctoral graduation overriding "issued", only if not available, use dateAccepted.
<dc:type>ElectronicThesisandDissertation </dc:type>	<type>Electronic Thesis or Dissertation</type>	repeatable, more types of dcmi-type standard vocabulary possible
<dcterms:medium>	<format>	repeatable, in ETD-MS optional
<dc:identifier> and <ddb:identifier>	<identifier>	repeatable
<dc:language lang= triple-digit language code in according to ISO 639-2>	<language>	repeatable, in ETD-MS optional
<dc:coverage> and <dcterms:spatial> und <dcterms:temporal>	<coverage>	repeatable, optional; scheme in ETD-MS results from ddb:Scheme in XMetaDiss
<dc:rights> and <dcterms:accessRights>	<rights>	repeatable, optional
<thesis:degree><thesis:name>	<degree><name>	repeatable, optional
<thesis:degree><thesis:level>	<degree><level>	repeatable, optional in ETD-MS
<thesis:degree><thesis:discipline>	<degree><discipline>	repeatable, optional
<thesis:degree><thesis:grantor> <cc:universityOrInstitution><cc:name> <cc:place>, <cc:department>	<degree><grantor>	repeatable, optional in ETD-MS
	<description role="release">	The German peer review system allows only one end version. Therefore errata like in <description role="release"> can not be specified.

The ostentatious listing of details about a referee, advisor, chair of the committee reviewing the thesis etc. has its reason in the fact, that a lot of scientists want to know more about the genealogy of the scientific family of their subject. The Mathematics Genealogy Project¹⁶ allows the discovery of scientific ancestors. The intent of this project is to compile information about all mathematicians of the world who have received a doctorate in mathematics. It is planned to show the following, beside others, for each individual: the complete name of the degree recipient, the name of the university which awarded the degree, the complete name(s) of the advisor(s).

Other elements are also being applied within XMetaDiss to fulfill the requirements of libraries and scientists: administrative elements, elements for long-term preservation and elements which provide additional information about the doctorate.

The metadata element `<type>Electronic Thesis or Dissertation</type>` is a topic which has to be discussed for future developments and applications. It is likely to be true, that the dcmi-type standard vocabulary will not be able to be used but a list of types specific to thesis material may be required, right down to individual university level. Exploring how to deal with this on an international as well as a national scale could be significant in the future. For the transformation of XMetaDiss version 1.2 to ETD-MS Version 1.00, revision 2 with a XSLT processor, a XSLT file is provided at <http://www.ddb.de/standards/xmetadiss/etdms.xsl>

4. DISCUSSION

XMetaDiss provides the possibility to state relations between the electronic dissertation and other publications like earlier or later ones. It has to be discussed if these relations are seen as important to describe the resource.

It has also to be discussed if MARC¹⁷ or MODS¹⁸ relators and roles could be used, especially for the specific graduation process in Germany. A conversion from XMetaDiss to MODS and MADS¹⁹ had kindly been done by R. Guenther, Library of Congress, Washington D.C.

MARC (**MA**chine-**R**eadable **C**ataloguing) is a format standard for the storage and exchange of bibliographic records and related information in machine-readable form. As an XML schema, the "Metadata Object Description Schema" (MODS) is intended to be able to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records. Metadata Authority Description Schema (MADS), is an XML schema for an authority element set that may be used to provide metadata about agents (people, organizations), events, and terms (topics, geographics, genres, etc.). MADS was created to serve as a companion to the Metadata Object Description Schema (MODS). As such, MADS has a relationship to the MARC 21 Authority format, as MODS has to MARC 21 Bibliographic.

The validation of the exemplary files, has been performed with the XML editor "XMLSPY 2004 Edition" ('Enterprise Edition' and 'Home Edition'). Validations with different parsers may result in failure notices, due to the application of the Dublin Core Schema (dcterms.xsd; see the special explanations at <http://dublincore.org/schemas/xmls/qdc/2003/04/02/notes/> in DC, Appendix A : Parser Behaviour). The validation of the Schema 'xmetadiss.xsd' has been performed with the "XMLSPY 2004 Enterprise Edition". The dcterms.xsd uses a construct that was controvertible for some time: The W3C Schema Specification did *not* permit it until the release of the "Proposed Edited Recommendation" (18 March 2004). Therefore some parsers (including widely-used versions of Xerces-J) reject the schema.

5. CONCLUSION

What are the benefits of these results for managers and librarians?

MetaDiss is a widespread and accepted format for online university theses and dissertations in the German-speaking countries. XMetaDiss is based on this successful appliance. Furthermore the potential of XMetaDiss consists in modern and fast automation processes which enable library and IT staff to deal with ETDs in a more economical and simultaneously in a richer way. Because the distributors of ETDs have the chance to deliver personal data on authors, this data can be transformed in authority records. Scientists and scientific managers obtain high-level information on electronic theses and dissertations. If an institution has to ensure long-term preservation and accessibility of online-publications, it could refer to the technical metadata which is implied in the XMetaDiss format. The compatibility with the NDLTD set ETD-MS allows retrieval and access to German and German-speaking online theses and dissertations via international metadata search engines like the ETD Union Catalog.

What are the next steps?

The main aims are to provide a qualitatively high description level of ETDs, to insure the long-term preservation and access to these resources, and to support the ETD Union Catalog. The main aims also include considerations about international cooperation processes on the basis of ETD-MS. It is recommended, that if ETD-MS will be enhanced, regional, respectively national developments should be comprised. That means, that ETD-MS should include useful metadata elements which are being applied by international partners to promote electronic theses and dissertations worldwide.

It is also recommended to put ETD-MS in a RDF²⁰ structure and to think about applying OWL²¹. OWL, the Web Ontology Language, a specification of the W3C (World Wide Web Consortium) is useful, to create, to publish and to disseminate ontologies with a formal description language. Terms of a specific domain and their relationships can be described in a formal way so that software (agents) can process their semantics. Therefore OWL is a fundamental part of the idea of a semantic web. OWL is technically based on the RDF syntax and provides better retrieval possibilities.

XMetaDiss now is available in the usable release 1.2. An OAI harvester (OAI-PMH version 2.0)²² is implemented and running. Some partners have switched to XMetaDiss and deliver metadata either by providing XMetaDiss records at their OAI interface or by submission of XMetaDiss records as e-mail attachments. As soon as most of the metadata providers will have changed to submission of the new format, XMetaDiss records will be imported automatically into the bibliographic database. New services (e.g. a portal for search and retrieval) exploiting the greater opportunities of XMetaDiss will be implemented in a new project "Building-up a Co-ordination Agency for online dissertations and post-doctoral theses" that is sponsored by the German Research Foundation and is about to be launched. The inclusion in an international framework, depending on sufficient agreement with libraries, scientists and library service centres in Germany points to a high acceptance of XMetaDiss.

6. CONTINUATIVE URLs¹

1. <http://www.ddb.de>
2. <http://deposit.ddb.de/metadiss.htm>
3. http://www.ddb.de/standards/xmetadiss/index_e.htm
<http://deposit.ddb.de/xmetadiss.htm>
4. http://www.natlib.govt.nz/files/nlnz_data_model.pdf
5. http://www.ddb.de/professionell/afs_e.htm
6. http://www.dissonline.de/index_e.htm
7. <http://www.dfg.de/en/index.html>
8. http://www.ddb.de/professionell/dissonline_e.htm
9. <http://www.dublincore.org>
10. <http://www.ndltd.org/standards/metadata/current.html>
11. http://deposit.ddb.de/netzpub/web_online-hochschulschriften_stat.htm
12. http://www.natlib.govt.nz/files/nlnz_data_model.pdf
<http://www.langzeitarchivierung.de>
13. <http://www.ddb.de/professionell/pnd.htm>
14. <http://www.persistent-identifier.de/?lang=en>
15. http://www.ddb.de/standards/xmetadiss/konkordanz_xmetadiss_etd_ms_e.pdf
<http://www.ddb.de/standards/xmetadiss/etdms.xsl>
16. <http://genealogy.math.ndsu.nodak.edu/>
17. <http://www.loc.gov/marc/>
18. <http://www.loc.gov/standards/mods/>
19. <http://www.loc.gov/standards/mads/>
20. <http://www.w3.org/RDF/>
21. <http://www.w3.org/TR/owl-features/>
22. <http://www.openarchives.org/OAI/openarchivesprotocol.htm>

¹ Links accessed August 2005