Because the DiVA Project has already produced systems to support some of these activities, these solutions will be used as a starting point for the project. Considering the lack of practical implementations of solutions supporting long-term preservation and access within the library community, we believe the results of this project will be broadly useful.

Title: From DTD generation to XML conversion: Structured ETDs at the Document and Publication Server of the Humboldt University
Authors: Uwe Müller (Humboldt University)
Abstract: Since 1997 PhD students at Humboldt University Berlin can fulfil their publishing duty concerning their dissertations by using a digital publication. While we were one of the first universities having extended the bunch of accepted publishing possibilities to this method nowadays institutional document and publication servers are regarded as a standard service provided by almost every German university. In contrast to the vast majority of domestic university libraries Humboldt's electronic publishing group has pursued a structured document approach from the very beginning of its activities in this area. The originally developed DiML (Dissertation Markup Language) derived and adapted from an SGML-DTD evolved at Virginia Tech for Electronic Theses and Dissertations has now been transformed to an XML-DTD (xDiML).

In this context a DTD generation system has been developed allowing for the compilation of individually assembled DTDs. For this purpose the elements of the DTD have been grouped into modular units. These modules which are XML files themselves are stored in the DTDBase. Using the DTDSys – a transformation system on the basis of XSLT and Java – the modules can be combined to an individual DTD – e.g. the xDiML DTD used for Theses and Dissertations. Due to its modularity the system can easily be used to supply new publication series with appropriate DTDs, which can contain special elements and which are as slim as possible and thus more easily applicable than a universal "Mega"DTD. The DTDSys also facilitates the integration of externally managed (standard) DTDs such as SVG, SMIL, MathML, or MusicML and thereby allows the generation of DTDs with multimedia extensions. The use of a controlled and centrally managed set of modules provides the advantages of shared semantics beyond the borderline of different DTDs – a feature which is used e.g. for qualified fulltext retrieval.

The XML based publishing approach is currently applied for dissertations and master theses, university serial publications, as well as a few electronic journals and conference proceedings. Different approaches have been developed for the conversion process from text processing systems to XML. They were adapted for the various requirements of the particular author groups or editors. The developed methods include styles and

add-ons for OpenOffice/StarOffice, FrameMaker and MS Office. They especially exploit the XML support most software vendors of office systems have newly integrated as a genuine standard interface into their products. The generated XML files form the source for different presentation formats, the basis for longterm preservation activities, and a prerequisite for value added retrieval techniques.

The paper will reveal both, the DTD generation system and the XML conversion process.

Title: Avoiding the Digital Dark Age with PDF & XML: How your institution can ensure access to digital theses and dissertations and avoid technological obsolescence
Authors: Chuck Myers (Adobe Systems)
Abstract: If you suddenly couldn't access the building where your valuable research data or study was kept, what would you do? In a paper-based world, archiving graduate research meant storage of paper or microfilm. But what happens as more and more theses and dissertations are created electronically? How do you preserve valuable research in a consistent format? How do you keep the exact look and feel of a document today, 30 or 300 years from today?

The lack of a recognized and accepted electronic standard for archiving theses and dissertations -- particularly as new generations of hardware and software make previous digital technology obsolete – could lead to the loss of significant amounts of valuable information. For example, over the past several decades, military files from the Vietnam War, records from the Viking Mars Mission, Census Bureau data and land use records have been lost due to the inability to read data formats and the deterioration of magnetic tapes used to store that data.

Archiving standards such as PDF/A and PDF/XML can ensure secure access to digital data across the enterprise and Internet. The speaker will discuss how graduate schools could use PDF/A and PDF/XML to archive and preserve their digital works enabling the fidelity of digital documents to be preserved for generations to come. He will examine how XML as a neutral format for creating Web, print, and wireless content can be formatted and presented as a PDF document. The session describes the new place of XML and Web standards for printable representation of dynamic text and graphics as well as static content. He will discuss how these standards are being developed under the publicly available and highly prevalent PDF standard, offering protection from technological obsolescence over the ages.

Attendees will learn about:

- The dramatic growth of information due to the digital age and the challenges involved in preserving digital data.