

February 2007

(Last edited: sd: 21/03/2007 10:10)

# Institutional Repositories Workshop Strand Report

## Strand title: e-theses

### Authors:

Susanne Dobratz, Humboldt-University Berlin, [dobratz@cms.hu-berlin.de](mailto:dobratz@cms.hu-berlin.de)

Hubert Krekels, Wageningen University, [Hubert.Krekels@wur.nl](mailto:Hubert.Krekels@wur.nl)

Maurice Vanderfeesten, SURF Utrecht, [Vanderfeesten@surf.nl](mailto:Vanderfeesten@surf.nl)

Gerrad van Westrienen, SURF Utrecht, [vanWestrienen@surf.nl](mailto:vanWestrienen@surf.nl)

## 1. Executive Summary

The workshop showed that e-theses are an important part of a university's research output and yet not sufficiently integrated into a European repositories infrastructure and searchable for the specific e-theses information. As preparation of the e-theses strand a small demonstrator, showing how an integrated search using the OAI-PHM could be used to offer a retrieval portal on a European scale and where the difficulties in terms of metadata are.

The participants of the discussion agreed upon the fact, that e-theses should be seen as an integral part of broader institutional repositories. As such they should not follow an own metadata set, but be integrated into a metadata profile for the whole repository. It was suggested to develop a European e-prints Application profile. Such a profile should work out the very specific metadata that only apply to doctoral e-theses.

During the workshop the work was started to identify key issues in handling doctoral e-theses and to prioritize those issues. This table can be used for planning further European and country specific activities to reach the goal: make e-theses in Europe better retrievable and therefore visible via institutional repositories.

The results of the demonstrator and discussions will be further elaborated and presented as "European e-Theses Demonstrator" at the ETD 2007 conference in Uppsala.

## 2. Summary of recommendations

The discussions at the workshop brought the following results:

1. E-theses have to be seen as part of an overall institutional repositories infrastructure and content. They should not be handled differently from other scientific and scholarly e-papers.
2. The demonstrator showed that it is generally possible to harvest European e-theses using the OAI-PMH. But it was shown in the study and discussions that Simple Dublin core is not enough. The participants agreed that a richer metadata set is necessary to offer a retrieval portal with good quality.
3. Country specific best practise examples like the *XmetaDiss* approach in Germany, the *uketd metadata set* in the UK have been presented and compared to older e-theses specific metadata sets like the *NDLTD etd metadata set*, the latter was seen to be outdated.

- 
4. It was suggested to develop a European e-theses application profile for metadata at a European level, meaning within a European working group. It was suggested that the GUIDE group could be a good address for such further activities.
  5. Investigations have to be made to handle e-theses in terms of metadata as part of e-prints. Therefore e-theses specific informations has to be encoded into metadata. A first list of e-theses specific elements has been produced during the workshop.
  6. National authorities should fund national development of richer metadata schemes. This allows to achieve a demonstrator that shows what is possible in a short term. It is the easiest to build 4 or 5 filters at the demoservice level, but it is not scalable. The main goal is to get at a higher level based on national richer metadata, is a good way to put first steps in achieving a European e-prints application schema.
  7. The key issues in handling doctoral e-theses have been identified by the workshop strand participants and they have been prioritized as follows:
    - I. Richer metadata (11 points)
      - Use qualified dc
      - Compound docs
      - Complex objects
      - Technical and preservation metadata
    - II. Wider IR perspective (8 points)
      - How to integrate richer metadata
      - Is document type a good choice to base services on
      - How to get them
      - Compatibility
    - III. ETD specific issues (7 points)
      - Degree, level, definition
      - Various dates
      - Define minimal requirements
    - IV. Cultural aspects (5 points)
      - Language
      - Interpretation and definitions
      - Local versus national services
    - V. Audience (2 points)
      - Keep it simple
      - Focus on added value metadata
      - Think about target groups
    - VI. Subject classification ( 0 points)

### 3. Discussion

The following issues in the context of e-theses have been mentioned during the discussion:

<b>Pool of issues</b>	<b>Richer metadata</b>	<b>Theses specific</b>	<b>cultural</b>	<b>Audience &amp; services</b>	<b>Subject classification</b>	<b>Repository perspective</b>
complex or compound objects	x					
subject classification					x	
minimum metadata requirements	x					x
national differences			x			
avoid simple dc	x					
thesis specific metadata		x				
access rights						x
target groups		x				x
cultural aspects			x			
Crosswalks on an abstract level						
Graduation Degree + level		x				
Keep it simple	x					x
Identify most important fields for added value				x		
Technical information in metadata	x					
Abstract mandatory	x					
Richer metadata	x					
Part of a wider picture						x
Document type is not a good choice for managing reps						x
Roles of sponsors etc		x				
Etd specific data integrated in repositories						x
Interoperability vs. manageability						x
How to get content						x
Advisor info		x				
Rights author and user	x					x

---

## 4. Outcomes

As a result of the workshop several questions could be answered by the group:

### A) What is the dimension of the problem?

- Metadata at unqualified level are not rich enough to really add value in a service demonstrator
- Unqualified DC does not accommodate some essential ETD specific metadata
- E-theses have to be seen as e-publications as part of the wider e-publications picture
- The problem is not very big, it is solvable in our opinion

### B) What can be achieved in term of e-theses during the next 12-24 month?

Within the next six months:

- We want to enhance the demonstrator
- Use existing richer national metadata which are e-thesis specific
- Test, if possible, the e-prints application profile as an input format and output format for the demonstrator
- Figure out the thesis specific elements on a semantic level

Within the next two years

- Use e-prints application profile as a kick off towards developing a European Application profile which incorporates e-theses specific information
- Demonstrator for European E-prints (cooperate within KE, maybe EU- 7th framework programme)

### C) Which approaches are needed to achieve this goal?

- Identify e-thesis specific elements on a national level or in broader int. coop.
- Identify e-thesis specific encoding on a national level
- Identify cultural issues, language, graduation level
- Enhance the demonstrator
- Encourage countries to adapt or adopt existing profiles (for a longer term adopt European application profile)
- **Cooperation with other KE workshop strands .....**

### D) What role could the KE partner organisations play?

KE partner organization could initiate, coordinate and finally fund national initiative to reach national agreements on national e-theses specific metadata sets. They could revise current programs and projects from the perspectives explained in the e-thesis group. KE partners should take responsibility to organise a network on European E-thesis (and investigate if the GUIDE working group can be function as umbrella organisation).

### E) How much can be done within national programmes?

Within national programs and activities awareness could be risen and promoted, directed towards the key issues (see above), national players as universities, libraries and library organisations should be persuaded that it is worthwhile investigating the problem. The national partners have to give advise to local projects and initiatives, discuss the appropriate issues with them and finally fund the necessary developments. This is a matter for the KE-partners at a local level.

### F) Where is an international approach mandatory?

- Developing the European application profile

- 
- Make recommendations (KE) to data-providers targeted at balancing between data- and service providers. (in this project we are data and service provider)

**G) The over-arching findings and challenges of the discussions have been**  
the following ones:

- Changing the balance between data and service providers
- Moving from toddlers speech to adult conversation
- Richer metadata
- Wider IR perspective
- ETD specific issues
- Cultural aspects
- Subject classification
- Rights

**H) What will be the major breakthroughs if the advice leads to partner actions?**

- European Eprints Application profile incorporating ETD spec. Metadata Elements
- European Eprints portal prototype

## 5. Annexes

### Annex 1: List of Participants

<b>MODERATORS:</b>		
Gerard van Westrienen	SURF	<a href="mailto:vanwestrienen@surf.nl">vanwestrienen@surf.nl</a>
Hubert Krekels	Wageningen University and Research Centre	<a href="mailto:hubert.krekels@wur.nl">hubert.krekels@wur.nl</a>
Susanne Dobratz	Humboldt University Berlin	<a href="mailto:dobratz@cms.hu-berlin.de">dobratz@cms.hu-berlin.de</a>
<b>PARTICIPANTS:</b>		
<b>GERMANY:</b>		
Maren Brodersen	DNB (German National Library)	<a href="mailto:m.brodersen@d-nb.de">m.brodersen@d-nb.de</a>
Kim Braun	University Oldenburg Library	<a href="mailto:kim.braun@uni-oldenburg.de">kim.braun@uni-oldenburg.de</a>
<b>UK:</b>		
Paul Needham	Cranfield University / EThOS project	<a href="mailto:paul.needham11@btinternet.com">paul.needham11@btinternet.com</a>
<b>The Netherlands:</b>		
Maurice Vanderfeesten	SURF	<a href="mailto:vanderfeesten@surf.nl">vanderfeesten@surf.nl</a>
Arent Bosman	Technical University Delft	<a href="mailto:a.j.bosman@library.tudelft.nl">a.j.bosman@library.tudelft.nl</a>
<b>DENMARK:</b>		
Claus Vesterager Pedersen	Roskilde University Library	<a href="mailto:cvp@ruc.dk">cvp@ruc.dk</a>
Henrik Juul-Nyholm	University of Copenhagen	<a href="mailto:hjn@adm.ku.dk">hjn@adm.ku.dk</a>
<b>SWEDEN:</b>		
Ronnie Kolehmainen	Uppsala University Library, Electronic Publishing Centre	<a href="mailto:ronnie.kolehmainen@ub.uu.se">ronnie.kolehmainen@ub.uu.se</a>

### Annex 2: Liste of e-theses specific metadata to be discussed at a broader level

Element	Recom- mended	Man- datory	optional	notes
Title		X		
Creator		X		
Description	X	-		
Date (published)		X		To decide on what kind of date
Type		X		
Language		X		
Contributor	X			
Identifier		X		There has to be found a solution for compound objects
Relation			X	
Subject			X	
Publisher			X	
Rights			X	
Source			X	