Doctoral e-Theses
Experiences in Harvesting on a National and European Level

Gerard van Westrienen
20 April 2007
1. Introduction

2. National developments: ‘Promise of Science’
   - Aims
   - Results
   - Lessons learned

3. European developments
   - Various meetings / workshops
   - New initiative: GUIDE
   - Project: European e-Theses Demonstrator
An e-Theses Service upon the (Institutional) Repositories Data Grid

Services
- Research
  - Subject repositories, refereed portals, e-Theses portals, (Open Access) journals, ...
- Education
  - Virtual Learning Environments, Course Ware, Readers, ...
- Society
  - Institutional windows, expertise, professional journals, personal Web sites, national windows, ...

Harvesters

Institutional Repository
- Rotterdam
- Delft
- CNRS
- MIT

Data Grid
Doctoral dissertations:
- High quality and standard; original research
- Show case for researcher: nice printed monograph
- Show case or visiting card of a university
- hardly visible or accessible

The Netherlands:
- Yearly output: between 2600 – 2900 (5% of formal scholarly output)
- No ‘student’ work; normally temporary employment contract
- formally 4 years; average >5 years
Promise of Science
the young and promising scholars;

Aims:
- creating a national e-Theses gateway with more than 10,000 e-theses
- To realise that by the end of 2006 more than 90% of all new theses will be digitally and Open Access available
Growth of completely Open Access available e-Theses

- October 2005: 6,400 e-theses
- June 2006: 9,800 e-theses
- November 2006: 12,100 e-theses
- January 2007: 13,580 e-theses
- April 2007: 14,470 e-theses

Growth of over 8,000 e-theses in 18 months
(almost 3 times the yearly output)
Number of e-Theses available per year of publication (1597-2006)
Percentage of Total Theses Output per year in ‘Promise of Science’
1995 - 2006

Theses
% in PoS
Some lessons learned

Organisational:
- Decentralised approach and integrated part of the Institutional Repositories infrastructure is great success (business model; workflow and expertise);
- Goal of 10,000 e-theses was easily reached
- Goal of attracting >90% of all new theses is more challenging; not been met by the end of 2006. Needed is a change of culture, procedures, regulations etc.
- Interoperability is not something that goes without saying...; appointments on standards need to be agreed upon explicitly and clearly
Some lessons learned

Legal issues:
- E-Theses are a good starting point for the creation of a university copyright policy
- In the Netherlands 6 out of 13 universities have a mandatory policy for the archiving the e-theses in and making it accessible via the institutional repository
- A mandatory policy however doesn't directly mean a 100% coverage of e-theses
- Separate archiving from making theses accessible via the IR
- Embargo periods can be useful to apply; especially in certain disciplines where a thesis is a merger of various articles, already published in (commercial) journals
Some lessons learned

Communication and awareness:

- Many **Cold feet**
- **Mandatory** policy is not sufficient
- Presentation of **intermediate results** of universities stimulates to take action
- **Launch** of National e-Theses site was important stimulus
Promise of Science

What is Promise of Science?
Promise of Science provides a simple way to search for doctoral e-theses in the Netherlands. You can find doctoral theses from all Dutch universities via the website. Once you find a relevant thesis, it is often available full text, and free of charge.

Features of Promise of Science

- **Search** doctoral e-theses from all Dutch universities from one convenient place.
- **Find** the full text.
- **Learn** about the most recent evolution of ideas of (young) researchers in their fields.

The importance of Promise of Science
Every year about 2500 doctoral theses are published, which is about five percent of the country’s entire formal academic output. Doctoral theses usually represent the most recent evolution of ideas by (young) researchers in their fields. With this, DAREnet has added ‘Promise of Science’.
19-20 Jan 2006: e-Theses workshop Amsterdam; 11 countries; overview

21-22 Aug 2006 e-Theses Workshop SURF, JISC, DIVA

16-17 January 2007: KE workshop Inst.Repositories; 6 strands, incl e-Theses

NDLTD; int network

GUIDE, Guiding Universities In Doctoral E-theses

European e-Theses Demonstrator Project

KE Workshop

ETD’07 Conference, June 2007 Uppsala
Eur. Workshop + Demonstrator

ETD ’08 Conference, UK

EU Commission; FP7...

Knowledge Exchange; SURF, JISC, DEFF, DFG

DRIVER

Eur.Commission; FP7...
**GUIDE: Guiding Universities In Doctoral E-theses**

**Aim**
To stimulate European doctoral ETD developments.

**Objectives**
- Share good practice; identify common problems and solutions
- Encourage compatibility where appropriate
- Identify where limited joint activity would be beneficial, and plan and pursue this joint activity
- Prepare the e-theses domain to participate fully in the emerging wider interoperable network of European digital repositories, within an international context
- Avoid unnecessary duplication of effort
GUIDE Steering Group:
- Stefan Andersson - Sweden
- Paul Ayris – UK
- José Borbinha - Portugal
- Susan Copeland - UK
- Neil Jacobs (Secretary) - UK
- Wilma Mossink – NL
- Eva Müller - Sweden
- Christopher Pressler (Chair until ETD’07) – UK
- Peter Schirmbacher - Germany
- Rita Voigt – Finland
- Gerard van Westrienen (Dep. Chair until ETD’07) - NL
Aim: facilitate ongoing international coordination

Activities:
- formally constitute GUIDE with links to the NDLTD;
- an advice and information point via a website;
- establish discussion forums;
- a launch at the ETD 2007 conference in Uppsala;
- elaborate on the European e-Theses Demonstrator;
- conduct synthesis and comparison activity;
- evaluate and make recommendations;
10th International Symposium on Electronic Theses and Dissertations

Uppsala, Sweden, June 13-16, 2007

News (2007-04-03)

Lists of accepted papers and accepted posters are on-line now.

Important dates:

November 11, 2006
First call for papers

January 25, 2007
Second call for papers

February 1, 2007
Registration open [Registration]

February 15
Final deadline for submission of abstracts of papers

http://epc.ub.uu.se/etd2007/
Pilot project (Oct 2006 – June 2007):
- Harvest repositories with e-theses on an international scale
- Set up a freely accessible European portal and test in practice the interoperability.

Aim:
- share current interoperability practices;
- get better insight in (critical) issues and potential solutions;
- set up a Demonstrator – an interoperable portal of European e-theses
Repositories involved from 5 countries:
- Cranfield QUEprints (UK)
- Delft University of Technology (the Netherlands)
- DiVA, the Academic Archive Online (Sweden)
- Humboldt University of Berlin (Germany)
- Roskilde University (Denmark)

Project funded by:
- DIVA (through BIBSAM, Sweden),
- JISC (UK)
- SURFfoundation (the Netherlands).
http://e-thesis.sharelab.cq2.org/
Lessons learned (up to now)

- Qualifications of doctoral theses differ amongst Europe (in spite of Bologna process)
- Language: metadata (incl. summary) in English
- Garbage in is garbage out
- Filtering *doctoral* theses: document types need standardisation
  
  dc:type = Doctoral thesis / Master thesis / Bachelor thesis
- Simple DC is not enough; richer and some ETD specific metadata are needed (and available)
- ‘Cross walks’ can be made between various metadata-formats, but scaling-up is a problem
- Standardisation needed at the level of *data providers* if we want to stimulate services
Thank you.

Gerard van Westrienen
vanwestrienen@surf.nl
www.surf.nl