



Five Ways to Add Value to PhD Theses...

...and Two Ways to Do It Better

Joachim Schöpfel, Lille (France)

<http://www.citeulike.org/user/Schopfel/tag/gl14theses>

joachim.schopfel@univ-lille3.fr

ANRT <http://www.diffusiontheses.fr/>



Methodology

- State of the art
 - NDLTD/ETD 2007-2012
 - USETDA 2010 & 2011
 - GL (in OpenGrey)
- Studies on open access and grey literature
- Case studies



Five ways to add value to PhD theses

- Quality
- Metadata
- Format
- Interoperability
- Services



Quality

- *“The content (...) is the most important factor that has been cited by researchers to show the success of a repository”* (Macha & Jager 2011)
 - Policy, standards
 - Only validated version
 - Review system
 - Requirement of minor or major revisions
 - Minimal requirements for acceptance (OA, fulltext, research theses, metadata...)
 - Born-digital and older repurposed digital materials
 - Retrospective digitization of selected sources
 - Unrestricted access
 - Copyright cleared



Metadata

- *“Considerable variations in practice exist (...) Many of these variations involve local qualifiers and/or the use of metadata elements to record information that does not correspond to the prescribed meaning of the elements”* (Park & Richard 2011)
 - Standards, good practice
 - OAI DC simple (if networking, cf. DART Europe)
 - NDLTD standards for DC/MARC (ETD-MS)
 - ProQuest schema (UMI XML DTD)
 - TEF (France)
 - EVSKP-MS (Czech Republic)
 - Especially for small and medium sized institutions
 - Bibliographic description for searchability
 - Value added metadata for scientometrics
 - Metadata for print on demand (book publishing)
 - Metadata must change with new policies (Heyse 2007)
 - Easy conversion to other formats



Format

- *“In the sciences theses should be regarded not simply as objects destined for preservation and archival status, but rather as unique resources containing potentially valuable data that must be made extractable and re-usable”* (Morgan et al. 2008)
 - Fulltext and not just metadata
 - PDF
 - LaTeX → XML
 - Postscript → RDF
 - MS Office Word DOC
 - MS Office Open XML DOCX
 - Format processing
 - OCR
 - Semantic querying



Interoperability

- *“Demonstrate research efforts of the state... Central finding tool for research”* (Dowling & Steans 2011)
 - OAI-PMH
 - Policy, statement, commitment, collective will
 - National network
 - Statewide architecture
 - International network
 - Single hub for research
 - Discoverability of research



Services

- *“User-Centered Process of Service Development Prioritization”* (Halbert 2007)
 - Flexibility and capacity for rapid adaptation
 - Interactivity, discussion forum, mailing list, comments
 - Discovery tools, federated search
 - User-friendly and reliable software
 - Statistics and metrics, citations
 - Video with presentation of thesis
 - XML export
 - Print on demand
 - Information literacy education
 - © or CC licensing
 - Preservation (multiple copies)
 - Extended (distant) learning environment



Two ways to do better

- CRIS
- Data



CRIS

- *“A well managed repository can simplify the task for the researchers and their respective departments, their university administrators and (...) their assessors”* (Hey et al 2006)
 - ETDs as important part of IR integrated in national infrastructure
 - Assignment of unique identifier
 - OrgID and PersID
 - Analysis of research trends



Data

- *“It may be possible to enhance both the research options and the richness of access. “Deep access” to the research could include data, visual and aural representations, links to supporting material such as notebooks or collaborative activity.” (White 2007)*
 - Supplemental content
 - Video and multimedia (avi)
 - Images
 - Spreadsheets
 - Music (mp3)
 - Attn with copyright and formats



Examples

- Emory ETD Program (Halbert 2007)
- Peruvian Network of Digital Theses (Huarato 2008)
- Humboldt University edoc server (Schirmbacher 2009)
- Butler University (Tamminga 2010)
- EthOS (Gould & Rosie 2012)
- DART E-theses @ UCL (Moyle 2009)
- ProQuest metadata (Marsh & McLean 2008)
- Format (Morgan et al 2008)
- Quality (Lewandowska 2010)
- CRIS Soton (Hey et al 2006)



Bibliography and contact

[http://www.citeulike.org/user/Schopfel/tag/
gl14theses](http://www.citeulike.org/user/Schopfel/tag/gl14theses)

Email joachim.schopfel@univ-lille3.fr

ANRT <http://www.diffusiontheses.fr/>