Institutional Repository Project

Discussion Paper:
Towards Electronic Theses and Dissertations

Author: Dr PR Butler

Date: 21st January 2008

Filename: IRProject_DiscussionPaper_TowardsETDs_v1-0

Version: Draft 1.1

Contents

1 Purpose

This document considers issues towards the implementation of electronic theses and dissertations (ETDs) at the University of Manchester. Topics considered include,

a. what are the key recommendations (section 2)
b. how we have arrived at this paper and who has been consulted so far (section 3)
c. what other universities have adopted ETDs (sections 3)
d. what are the benefits and costs (section 4)
e. what needs to be discussed and decided (section 5)
f. what are the known obstacles and risks (section 5 and 6)

For further details contact,

Dr Phil Butler
Institutional Repository Manager
Tel: 0161 275 1514, Internal: x51514
Email: p.butler@manchester.ac.uk
Website: http://www.manchester.ac.uk/institutionalrepositoryproject

2 Summary and recommendations

The overall goal of this paper is to present issues for discussion on the adoption of electronic theses and dissertations by the University of Manchester.

The paper seeks the views of academic, administrative support and library staff from across the University. The intention is to identify an electronic submission and preservation system that is as acceptable to students and staff as print is currently.

The paper has the following recommendations,

Recommendation 1 The University initiates a Project with the aim to adopt electronic submission of theses and dissertations, for PhD and Masters (research and taught) degrees. Priority should be given to research-based work.

Recommendation 2 A Project Sponsor is identified and a Working Group is convened to direct and oversee the implementation of the necessary works.

Recommendation 3 Project management and implementation efforts are supported by the John Rylands University Library within the established framework of the University's Institutional Repository Project.

Recommendation 4 The Project aims to enable students to self-submit an electronic version of their thesis/dissertation and that electronic submission is made mandatory in some form.

3 Background

The University of Manchester currently supports around 9,300 postgraduate students, of which around 4,800 graduate annually. Towards Manchester 2015, the number of completing postgraduate students is set to double. As a consequence the workload required too manage thesis and dissertations in print-form, as well as the shelf-space required to store and preserve these works, is set to increase.

The information held within theses and dissertations in a non-digital format is notoriously difficult to access. As a consequence these scholarly works often remain invisible to the research community. This paper recognises that increasingly researchers expect the full-text of their work, plus the work of their students and colleagues, to be easily accessible to themselves and others via the World Wide Web. Simple, direct and instant access to theses and dissertations enables researchers to promote themselves as well as more easily retrieve valuable knowledge that may not necessarily appear in journal articles and other academic publications.

An individual's and university's research reputation is in part dependent upon the visibility and impact of their research outputs. Many research-led universities have already established institutional repositories of digital scholarly works (including ETDs) in efforts to raise the exposure of their research using the internet. Ultimately, such developments will favour the reputations of individuals and organisations by

enabling the research community to more easily access and cite these research outputs. In the absence of a Manchester repository, the developments of others could threaten Manchester's reputation and its goal to be a top 25 research-led institution in the world by 2015.

Noteworthy repositories of theses and dissertations are,

a. MIT Theses in DSpace is a collection of over 15,000 selected theses from all MIT departments. [http://dspace.mit.edu/handle/1721.1/7582]

b. Nottingham University has established separate services for the management and storage of eTheses and eDissertations. [http://etheses.nottingham.ac.uk/] [http://edissertations.nottingham.ac.uk/]

c. Oxford University Research Archive initiated a project in Jan 2007 that is ongoing and aims to manage and store theses submitted as part of research degree awards. [http://www.ouls.ox.ac.uk/ora/oxford_etheses]

d. University of Edinburgh Research Archive includes access to full-text digital Theses and Dissertations. [http://www.era.lib.ed.ac.uk/]

e. University of Florida Electronic Theses and Dissertations (ETD's) service provides access to over 4,500 full-text works. [http://www.uflib.ufl.edu/etd.html]

f. University of Glasgow DSpace service includes access to selected eTheses. [https://dspace.gla.ac.uk/]

g. University of Warwick's Research Archive Project (WRAP) aims to provide infrastructure that will accommodate 350 eTheses per annum. [http://www2.warwick.ac.uk/services/library/main/research/instrep/erepositories/]

h. Virginia Tech's ETD service hosts over 10,000 records. [http://etd.vt.edu/]

In addition to established repositories, a number of significant projects exist that aim to promote the adoption of ETDs nationally and internationally. These include,

i. In the UK, the EthOsNet project has been set up to establish the Electronic Theses Online Service (EThOs). A live service is due to launch in summer 2008. [http://www.ethos.ac.uk/]

j. In Europe, DART-Europe is an academic consortium of university partners, who are undertaking work on e-theses. [http://www.dart-europe.eu/]

k. The Australian Research Repositories Online to the World (ARROW) project aims to identify and test software or solutions to support best practice institutional digital repositories including electronic theses. [http://www.arrow.edu.au/]

l. The Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations. [http://www.ndltd.org/]

These projects and other university repositories represent a significant body of knowledge and experience that we may apply to a University of Manchester ETD repository.

Some Schools already receive and store electronic versions of theses and dissertations. There are different ways of collecting this information. Overall, the take-up of these routes by students is low. Examples are,

- the School of Computer Sciences has an electronic repository of mainly abstracts and some full-text PhD thesis [http://intranet.cs.man.ac.uk/Intranet_subweb/library/]
- the School of Mathematics is using ePrints software and encourages self-submission of PhD theses by students [http://eprints.ma.man.ac.uk/]
- specific areas within Schools are accepting and storing electronic versions of theses/dissertations on CD

John Rylands University Library has recently agreed to participate in the EThOs Project [http://www.ethos.ac.uk/]. This will enable the digitisation and preservation of around 180 theses annually over the next three years.

In response to a range of national and international developments, and in support of the University of Manchester's 2015 agenda, the Vice President's Research Group (co-chaired by Professors Nancy

Rothwell and Simon Gaskell) requested John Rylands University Library initiate a project with the aim to establish electronic repository services for the University of Manchester. As a consequence, the University's Institutional Repository Project officially started on 1st October 2007. Professor Carole Goble is the Project Sponsor and a Steering Group with representatives from across the university has been convened. A fulltime dedicated Project Manager and Implementation Team has been appointed and is working within John Rylands University Library. The Project runs for 2 years.

The Institutional Repository Project will focus on providing services that offer members of the university a means of storing, managing and disseminating their scholarly works. This includes academic publications, grey literature and audio visual items. The Project identifies electronic theses and dissertations as one type of scholarly work for consideration. See the Project website for further information [http://www.manchester.ac.uk/institutionalrepositoryproject]

This paper summarises views expressed by the following individuals during meetings held in December 2007/January 2008,

- Clare Atherton (Acting Head of Graduate Education and Skills Development Co-ordinator, Research Office)
- Janet Barratt (Theses Librarian, John Rylands University Library)
- Phil Butler (Institutional Repository Manager, John Rylands University Library)
- Keith Elliott (Head of the Learning Development)
- Neil Ferguson (Head of Faculty Academic Services, Faculty of Humanities)
- Nicola Ellis (Senior Graduate Administrator, Faculty of Humanties)
- Anna Lawless (Postgraduate Administrator, Faculty of Life Sciences)
- Helen Myers (Graduate Education Manager, Faculty of Medical and Human Sciences)
- Lee Wilkinson (Postgraduate Research Administrator, Faculty of Engineering and Physical Sciences)

4 Outline business case and benefits

The implementation of ETDs represents a long-term commitment for any institution. Benefits tend to accrue overtime and are difficult to quantify. As a consequence, this paper considers the benefits of ETDs in terms of a strategic development and the cost savings against current practices.

The most prominent reason for adopting ETDs is their increased visibility and impact. Evidence suggests that usage of theses and dissertations by the research community can increase dramatically when made easily accessible. This was shown, in particular, by the spectacular success of the US-based Virginia Tech service - the number of requests for PDF files (mostly full-text) of Virginia Tech ETDs increased 33 times in five years, from 221,679 in the year 1997/98 to 7,320,818 in the year 2002/03. In contrast, annual requests for paper-copies of theses from the British Library Thesis Service are around 12,000. Although difficult to quantify, one would expect the reputations of students, researchers and institutions to benefit from this increased visibility.

ETDs enable educational innovations. Examples are,

a. they enable remote submission by students enrolled on distance learning Masters courses, or by international students who have returned home prior to the final examination,
b. they enable easier showcasing of the best theses and dissertations for learning, teaching, marketing and communication purposes,
c. they can incorporate innovative content including multimedia as part of a submitted works (for awarding winning examples, see http://www.ndltd.org/awards/)
d. they enable examiners who prefer to work with an electronic version to do so
e. they present opportunities for better plagiarism detection via services such as "turnitin" [http://turnitin.com/]

The costs of submitting a thesis or dissertation in paper-form are invariably absorbed by the student and/or supervisor. At 6p per printed page and £25 per hard-bound copy, three copies of a 200 page PhD thesis would cost a student/supervisor around £110. With 770 PhD students graduating in 2006/2007, this would have represented a cost of around £85,000 to the University community. Submission of an ETD in Manchester, using existing computer network infrastructure, effectively has zero cost. However, we note that these monies may be equally viewed as a source of revenue for University functions such as binding, photocopying and printing services.

To preserve a bound thesis or dissertation in paper-form requires physical space to store the work in an environment that protects it from damage for posterity. Paper copy theses and dissertations are stored on

office shelves, Faculty/School libraries and/or in John Rylands University Library. Currently, John Rylands University Library stores around 55,000 hard-bound theses. Given the expected rise in graduating PhD students, this is set to increase by 2-5% per annum by 2015. In comparison, the resources required to store and preserve an ETD are small.

The introduction of ETDs would not reduce the storage space used by the existing collection. However, it would offset future storage requirements and any resource issues that may arise.

To store and manage ETDs the purchase of the necessary hardware and software are the main capital costs. Both commercial and open-source software solutions exist for implementing electronic repositories. Open-source software has no purchase costs but does require more technical support staff. Overall, staff time is the major ongoing cost. Staff are required for advocacy efforts (initially to promote the concept of ETDs and later, if submission is not optional, to encourage students to provide content), to provide training for students and staff, and oversee the integrity of the repository records.

Consideration of ETDs is one of the objectives of the Institutional Repository Project. The Project has existing resource to examine issues, establish working pilots, implement appropriate technical solutions and assess user experiences. This resource constitutes the major staffing costs for setting up ETDs. As a result, this Project represents a significant opportunity for the University. Further finances are required to scale-up pilot to production hardware and the Project is currently investigating how best to obtain these.

5 Known issues

5.1 What will be stored?

Discussion and decisions are required on what can be submitted and stored in an ETD repository.

This paper recommends the University of Manchester ETD repository accepts ALL PhD Theses and Masters Dissertations (research and taught) generated by students registered with the University from a defined date. Where storage of all theses and dissertations is deemed undesirable, impractical or (financially) unjustified, this paper recommends PhD Theses and Research Masters Dissertations are prioritised. This is because the quality and usefulness of the knowledge in these works is deemed to be greater than that in taught Masters Dissertations. Hence these works are more suited to long-term preservation and dissemination.

Undergraduate dissertations and other project-based student works are excluded here.

Electronic submission and storage of theses and dissertations would commence for all postgraduate students from some given future date (see Issue 5.4). The storage of past theses and dissertations is primarily limited by the cost of digitisation. At around £50 per text, the expense of digitising all past University theses and dissertations is prohibitive. However, digitisation and storage of selected past works chosen on a case-by-case basis or using some strategic criteria (e.g. most accessed) is possible if finances were made available.

The repository would store meta-data describing the work and the full-text of the work (with images and other media as necessary). Decisions are required on what meta-data is compulsory. For example, this might be the title, author(s), supervisor(s), date of submission, abstract and table of contents. The repository would support submission of one or more files making up the full-text.

To accommodate the range of electronic formats used by students, the repository would support various file formats. However, to help with the preservation of the files and to prevent problems with file-conversion that may be required as client software evolves, it is expected that only standard formats would be accepted. Decisions are required on what standard formats are supported (see Issue 5.9). These might include,

- Text Formats: ASCII (.txt); PDF (.pdf);
- Image Formats: CGM (.cgm); GIF (.gif); JPEG (.jpg); PDF (.pdf); TIFF (.tif)
- Video Formats: MPEG (.mpg); QuickTime – Apple (.qt and .mov); Encapsulated Postscript (.eps)
- Audio Formats: AIF (.aif); CD-DA, CD-ROM/XA (A or B or C); MIDI (.midi); MPEG-2; SND (.snd); WAV (.wav)
- XML/SGML according to the document type: "etd.dtd" (.etd) ETD-ML
- Special Formats: AutoCAD (.dxf); Excel (.xls)

Importantly, an electronic version of a thesis/dissertation should be an accurate representation of the examined work. It is worth noting that conversion of media between file formats can degrade the quality of the information stored (especially images, audio and movie files). As a consequence, students will

need to be informed of any conditions that exist for electronic submission and how they impact on the preparation of their thesis/dissertation. Training, guidance materials and support will need to be made available where appropriate. It is envisaged that this knowledge would be incorporated into existing graduate education programmes at a Faculty/School level and as part of training on transferable skills and scholarly publishing.

5.2   Who will submit content?

It is anticipated that students will submit their thesis/dissertation themselves directly to the ETD repository. A significant concern of graduate administrators is the additional workload that electronic submission might create. This is especially the case following the recent restructuring and rationalisation of the University's administrative support workforce.

The submission process would be, simple to use, available 24/7, web-based and possible from any internet-enabled computer. Submission would require the student to identify themselves via a logon using their University network username and password.

Only on occasion, do we expect a student would need help with submitting their thesis/dissertation electronically. This may be the case, for example, where a physical disability prevents the student from using a web browser. We expect such situations would be accommodated by existing University structures e.g. Disability Support Office and/or appropriate Faculty/School/Library support functions.

We assume that if not all, a very high fraction of postgraduate students already prepare their thesis/dissertation in an electronic form. Furthermore, we assume that the modern postgraduate student is fully conversant with the use of the World-Wide-Web, the use of web browsers and submission of content via online forms. As a consequence, we believe student training requirements in this area are negligible. It may be necessary to survey the postgraduate community to ensure this is the case.

5.3   What will be mandatory and what will be optional?

Decisions are required on what content is mandatory and what is optional. This paper recommends that once decided, the University's rules and regulations on submission of theses and dissertations are amended to include submission of an electronic version.

Any submission policy should be sustainable. Evidence from other universities suggests that when submission is optional only a small fraction of theses and dissertations are submitted electronically by students. Furthermore, if an optional submission policy was adopted significant advocacy efforts by administrative staff would be required to encourage students to submit their works electronically.

The electronic submission of the final version of the thesis/dissertation is most important for the purposes of preserving and disseminating research knowledge. For PhD theses and Research Masters Dissertations, the electronically submitted version would be equivalent to the final fully-corrected hard-bound print copy as submitted after the viva voce. In contrast, for a taught Masters Dissertation the work is often awarded a mark by an examiner (e.g. fail, pass, credit, distinction) and the student has no opportunity to make corrections. In this case it is more sensible that the version submitted electronically is that that is first submitted.

For both research and taught theses/dissertations, electronic submission of the work prior to examination would present examiners the option to mark the works electronically and help with the detection of plagiarism (using digital services like 'turnitin'). Discussions with Graduate Education administrators suggest that different disciplines will have different requirements and expectations in this area. As a result, further discussion is required.

Six of the seven UK Research Councils (EPSRC is the exception) and a number of significant charities (e.g. Wellcome Trust, Arthritis Research Campaign, Cancer Research UK) have now mandated that research outputs from there funded grants should be made available for free (where copyright allows) by depositing the work in an electronic repository. At the time of writing these mandates do not explicitly refer to research theses and dissertations. However, it is important to note there is a clear trend amongst research councils and HEFCE to encourage institutions and individuals to make every effort to reduce access barriers to research knowledge and affirm that it is the intrinsic merit of the work, and not the means by which it is published, that should be considered in making value-judgements.

In conclusion, this paper recommends that submission of an electronic version of theses and/or dissertations is made mandatory in some form (see section 2 recommendation 4). Further discussions are required to refine this form.

5.4   When will electronic submission start?

If mandatory electronic submission is adopted, there may be concern about potential additional demands on students who are part way through their studies. It may be best to propose that mandatory regulations take effect with regard to the next new intake of doctoral and/or masters students or at the next registration point. Voluntary submission would be encouraged for existing and past students, if an electronic copy of the material was to hand.

5.5   How will electronic submission be verified?

A thesis or dissertation submitted electronically will need to be verified that it is a true and accurate representation of the examined material. This might involve checking that the submitted meta-data is accurate and that the full-text includes any necessary corrections. Verification of the hard-bound print copy is currently required. If submission of the electronic copy was mandatory and verified in preference to the hard-bound print copy, additional effort to complete this task should not be required. In fact, by implementing appropriate workflow and validation techniques, the verification of the electronic copy could at least semi-automated.

Further discussion is required on what aspects of verification are required and how these might be implemented in practice.

5.6   What impact will electronic submission have on print submission?

Mandatory electronic submission theses and dissertations do not mean that submission of a print version stops. It is clear that most examiners prefer access to at least a soft-bound print version. In most cases a print version may easily be generated from an electronic version (audio or moving visual content are exceptions), however, it is unreasonable to expect examiners (internal and external) to support the printing costs. Hence, this paper recommends that submission of soft-bound copies for examination purposes remains mandatory as is the current practice in the University.

A range of views have been expressed regards retaining mandatory submission of hard-bound versions of theses and dissertations. Some academic supervisors are likely to prefer having access to a hard-bound copy, whether it is located on an office shelf, in a faculty/school library or in the University library. Similarly, students may prefer to have a hard-bound copy of there work as something they can be proud of and show to their peers and relatives. However, in terms of preserving the work for posterity, the ease by which it may be accessed and the opportunities for the work to be disseminated and cited, the consensus is, electronic is by far more flexible than print. One model might be, make submission of the hard-bound print version less of a University requirement and more determined by faculty, school or discipline requirements. Furthermore, this could be made a personal choice of the student and/or supervisor. Further discussion on this issue is required. Consideration of the financial impact any changes may have on University/Library bindery, photocopy and print services is also required.

5.7   What about copyright, intellectual property, levels of access and embargo periods?

To realise the benefits to the author, University, funder or collaborator, of any intellectual property contained within a thesis or dissertation, the University submission policy already allows for setting of restricted access to the work. Electronic submission and storage would need to allow for this situation equally. Work is required to develop and integrate procedures to address all aspects of intellectual property, royalties, levels of access and embargo periods related to theses and dissertations, including where appropriate digital rights management.

5.8   What about dissemination, take-down and preservation?

It will be necessary to put in place polices and guidelines to cover matters of dissemination, take-down, and preservation of content. These will need to take account of relevant national legislation (e.g. *'Freedom of Information Act'*, *'Data Protection Act'*).

A dissemination policy would consider levels of access (e.g. freely available, university-only, and restricted), what embargo options are enabled (e.g. 6 months, 12 months, with or without notification, review and extension options) and what dissemination mechanisms are possible (e.g. direct to websites, request via email submission, and request with moderation).

Structures and processes need to be established for dealing with suggestions and complaints from people accessing ETDs. It is wise, from the outset, to have clear guidelines about the circumstances in which it would be advisable or acceptable to remove content, for example, in cases where some work is subsequently found to break copyright or bring individuals or the University into disrepute. A decision should be made regarding the extent to which the details of the 'take down' policy are publicised.

A preservation policy would consider what period the University guarantees access to the electronic content. This might include conversion of the media to future file formats necessary to accommodate changes in software used by people.

5.9 What are the cultural obstacles?

Section 4 outlines the benefits of adopting electronic submission and storage of theses and dissertations. Initial discussions suggest that staff and students would see the introduction of ETDs as a positive development. However, some concerns have been raised. These tend to be minor issues and often result from misunderstandings. Other sections of this document attempt to address some of these concerns. Nonetheless, further communication and advocacy efforts combined with discussion, planning and development are necessary to ensure that such issues are addressed.

The following list represents common concerns,

- electronic submission should not be an additional burden to completing a thesis/dissertation
- the submission process needs to be simple and flexible
- the creation of the necessary digital files should not require special skills and software
- electronic submission should not be the first step to replacing paper submission and burdening examiners with printing costs
- efforts should be made to offset increases in plagiarism by ensuring detection mechanisms are effective
- the copyright and IPR of individuals and the University should be protected
- a hard-bound paper copy is a measure of the effort required to complete a thesis/dissertation, electronic submission should not dilute this
- a central repository should be flexible and support the requirements of individuals and organisations, not just central University functions,
- operation of the repository should result in an overall cost-saving
- electronic submission should not create more work for administrative staff
- electronic submissions will require student and staff training and long-term commitment
- the repository will need to share data with existing information systems e.g. Student Records System (Campus Solutions), Library catalogue
- the indefinite secure preservation of electronic works will need to be guaranteed
- electronic submission should not negatively impact on existing services (e.g. bindery, printing, photocopying) without due consideration and planning
- the necessary infrastructure to store and preserve content, needs to be sized, costed and planned for

This paper recommends that a project is initiated to address all issues and concerns (see section 2 recommendation 1). This would involve appointing a Project Sponsor and convening a Working Group (see section 2 recommendation 2). The Working Group would discuss issues and ensure the necessary works are undertaken. The Institutional Repository Project represents an ideal opportunity to manage and coordinate these efforts (see (see section 2 recommendation 3).

5.10 What are the technical obstacles?

Technologies already exist that enable the submission, preservation and dissemination of ETDs (see section 3). Work is required to assess and pilot these technologies in the University of Manchester. In particular it will be necessary to ensure technologies interface with existing University systems and external services, including,

- Student Portal Service
- Student Records System
- Research Information Management System
- Library Catalogue
- University web farm and websites
- EThOs service

6 Known risks

6.1 Inertial resistance

Adoption of ETDs is likely to encounter some reluctance from, in particular the University's academic community. This may in part be driven by a perception that electronic submission is burdensome and of little value. We believe the likelihood of any project encountering this type or resistance will vary across the University but in general this risk is MEDIUM.

6.2 Uninformed resistance

We recognise that a significant amount of misunderstanding and lack of awareness may exist in the University community about repositories and electronic theses and dissertations. As a consequence, individuals can become misinformed and ultimately their expectations may not be met. These misconceptions can only be addressed through communication and advocacy efforts. This risk is MEDIUM.

6.3 Informed resistance

Cultural, political and philosophical issues remain subjects of significant debate in the institutional repository community. The impact of access to electronic content on plagiarism is believed to be of particular concern. It is possible that such issues may distract from the benefits of ETDs. These issues are rare and invariably confused with 'uninformed resistance'. As a result this risk is LOW.

6.4 Financial constraints

These may prevent up-scaling of infrastructure from pilot to production service or limit the functionality that a project could implement. This risk is HIGH.

6.5 Lack of cooperation

We recognise that consequences of the 2004 merger continue to impact on the workload of individuals and that a degree of 'change fatigue' may exist within certain areas of the University. As a result, we anticipate that this Project may be perceived as yet another centrally-driven activity that is aimed primarily at benefiting administrative functions and reducing costs. As a consequence the Project may experience a lack of cooperation. This risk is MEDIUM.

6.6 Technical obstacles

It is inevitable the Project will encounter technical problems with the hardware and software. We expect by using well established solutions with an active support community and agreed university frameworks that we will minimise these risks. This risk is LOW