Preserving Electronic Theses at the university of st andrews libraries and museums

|  |  |  |
| --- | --- | --- |
| **Sean Rippington** |  | **Janet Aucock** |
| *University of St Andrews Libraries and Museums**UK**sbr2@st-andrews.ac.uk* |  | *University of St Andrews Libraries and Museums**UK**ja@st-andrews.ac.uk https://orcid.org/**0000-0001-9616-0612* |
|  |  |  |

**Abstract – The University of St Andrews Libraries and Museums have been preserving and making our PhD theses accessible since they were introduced in 1917. Since 2007 theses have been required to be submitted in both print and electronic form, the electronic files being used to make full text theses digitally available on our institutional repository https://research-repository.st-andrews.ac.uk/. There have been aspirations for progressing to ‘electronic only’ thesis deposit since 20081, but this was held back by a lack of digital preservation support and infrastructure. However, during the COVID pandemic many students were only able to submit an electronic copy of their thesis and print submission was temporarily suspended – forcing a temporary move to e-only thesis deposit. This move to ‘e-only’ thesis deposit became business as usual in March 2022, pushed by various stakeholders and supported by changes to governance, training, and technical solutions.**

**This poster explores the emerging good practice, lessons learned, and future steps for preserving electronic theses using the University of St Andrews as a case study in ensuring that these vital scholarly outputs have integrity, are accessible, and can be used by future generations.**

**Keywords – Theses, Open Access, ETD, PDF**

**Conference Topics – Exchange, Resilience**

# Introduction

The University of St Andrews is moving to electronic-only thesis deposit for PhD students in March 2022 in response to emerging stakeholder needs and expectations, including the streamlining of student experience, reduction in printing and binding costs for students, and the need to save on expensive physical library storage space. While we have considerable experience in receiving electronic (mostly PDF) copies of thesis for publication online since 2007, these have been seen as dissemination copies of paper theses, rather than authoritative ‘master’ or ‘archive’ versions in their own right. Although we do make efforts to preserve these electronic dissemination versions, including surveying deposited file formats to identify and monitor preservation risks, most of our past preservation effort and resource has been directed to paper over electronic theses.

The move to electronic only thesis deposit requires a shift in mindsets (that the electronic thesis will now be the ‘master’ or ‘archive’ copy) and requires a shift in resourcing and effort towards robust digital preservation of the electronic copy. Theses are unique scholarly outputs, and it is vital to ensure that the electronic files submitted have integrity, are accessible, and can be preserved.

# Emerging good practice

Prior to the move to electronic only theses deposit, we researched:

* Which institutions already preserve e-only theses, and how
* Who does this well
* What concerns have been noted

* 1. Which Institutions Already Preserve E-only Theses, And How

We could not find a definitive list of institutions already undertaking electronic-only thesis deposit, though some research via Google, professional mail lists and projects such as the Educopia ETD+ Toolkit revealed some institutions and individuals with experience in this area. Broadly speaking they took one of four approaches:

* Accept a pdf version of the thesis only
* Accept a pdf version of the thesis only, specifying a pdf/a variant
* Accept the ‘original’ file(s) and a pdf version of the thesis
* Accept the ‘original’ file(s) and a pdf version of the thesis, specifying a pdf/a variant.

The more detailed governance, workflows, systems, training and advice were usually quite specific to the individual institution and were surprisingly heterogenous.

* 1. Who Does This Well

It was difficult to establish who manages preservation of e-only deposited theses well for several reasons; there are no specific external benchmarks for measuring success in this process; relatively little experience in the process is shared publicly in detail; it can be difficult to assess the success of institutions in this process without access to their systems and content; relatively few institutions (only three found at the time of writing) had undertaken and shared the results of detailed reviews of their electronic theses corpus for preservation and access issues; we could not find any published data (at the time of writing) on preferred access formats for users of electronic theses.

* 1. What concerns have been noted

Common concerns that arose during our research into good practice included:

* What to ask students to deposit
* How to deal with increasingly common ‘non-traditional’ theses formats including video, web-based and ebook formats such as EPUB.
* The value of PDF/A as an archive format
* How to ensure and audit the integrity of formats derived from ‘original’ files
* Who should undertake any file format changes, and what support they need
* How to encourage the creation of ‘accessible’ theses
* How to audit the success of e-thesis preservation processes
* How to assess and handle any inability of students to meet technical deposit criteria (e.g. in format specification, including validity and integrity of deposited files)

# conclusions and Lessons learned

Following our research and internal discussions, we concluded that a successful transition to e-only thesis deposit requires significant non-technical changes, including to governance, training, culture, and potentially staff resource, with input from a range of stakeholders including relevant committees, senior academic staff, PhD supervisors, library staff, and accessibility experts.

Our systems and processes required review and development to ensure the integrity of electronic files throughout their journey from deposit, to storage, access and reuse. This review has been supported by the publication of our digital preservation policy and submission of an outline business case for additional business analysis and funding for technical solutions for digital preservation.

In terms of what we asked of depositing students,

our own process settled on:

1) Always ask for the source or ‘original’ files to act as the ‘master’ or ‘archive’ versions, as requesting or automating conversion of a thesis to a new format introduces risks around the integrity and validity of the resulting file that may be difficult to mitigate and document.

2) Ask for embedded files (video, etc.) in a separate folder so that their preservation can be managed separately if needed.

3) Ask for a PDF, which will be used as the ‘access’ version to be published online.

3) When converting to PDF for access provide Adobe Pro, or Nitro Pro, to produce valid PDFs.

4) Provide guidance for making minor adjustments to make the pdfs more accessible in the long term e.g. Create PDF 1.6:

• turn off additional compression of images

• Change the colour management to Gray Gamma 2.2 and Adobe RGB (1998)

• Choose ‘save original JPEG images in PDF if possible”

• Also choose “embed all fonts”

• Embed hyperlinks

• Stabilise hyperlinks using a web archiving service.

Note that many of the above settings will make the PDFs bigger.

5) Provide guidance for students to look for and correct common pdf migration issues, including migration errors introduced by use of non-western characters, unusual fonts, raster images, tables, and embedded multimedia.

6) Ask students submitting their e-thesis in more unusual file formats (e.g. video, ebook formats including EPUB) to discuss their submission with library staff to identify and mitigate against preservation and access issues.

7) Allow students to submit a valid PDF-A if they wish, and provide training and guidance to support this.

8) Request that any security settings on deposited files be disabled

9) As our particular workflow will involve staff inserting a coversheet into the PDF, ensure that this process does not affect the integrity or validity of the file.

Going forward, it will be vital to secure resourcing to review and act upon changes in good practice across the sector, and to respond to our own changing capacity and user needs and expectations. This cannot be a one-off project, and should seen as part of an evolution in a service that has existed since 1917.

We are grateful for the support and advice we have received from various related communities of practice, including the EThOS community, DPC members, SCURL Repository Shared Services Group, and other universities in Scotland. Greater communication between institutions undertaking electronic-only theses deposit, and perhaps some funding to undertake an evidence-based review of their processes, would also be welcome. Any lessons learned across the sector are likely to be useful in other digital preservation contexts, especially those involving deposit of content by third-parties and accessibility of scholarly communications.

# REFERENCES

1. Aucock, J. (2008). Electronic theses at the University of St Andrews: institutional infrastructure, policy and support to establish an electronic theses service. Research-Repository.st-Andrews.ac.uk; University of St Andrews. http://hdl.handle.net/10023/513