

Workflow Analysis

In June 2010 UCL conducted a sector-wide survey into thesis deposit and open access in UK Universities and HEIs. A rich body of qualitative data was provided by respondents describing in their own words how they approach the issues surrounding open access deposit of electronic theses. The data was analysed and evidence is presented here of consensus and best practice. A survey summary report and full text responses are available on the project website¹.

This document outlines the routes via which e-theses are processed in UK HEIs. There does not appear to be any particular pattern of workflow associated with a particular type of HEI (e.g. Russell Group members, post-1992 or specialised research institution) but it is worth noting that 67% of the institutions examined here either accept e-theses as part of the normal submission process and then send them to the library for upload to the IR or require students to upload directly to the IR as a part of the submission process.

Figure 1 below, shows the departments which are involved in e-theses workflows at each of the respondent HEIs. Although departments at each university may have similar names but different functions, or different names but similar functions, it is possible to cluster patterns of workflow according to the nature of the department using generic classifications such as ‘central department’ (which may be the registry, graduate school, research office, student progression office or another designation) or ‘library’ (which may be termed the learning resource centre, the information service, book library, main library or similar).

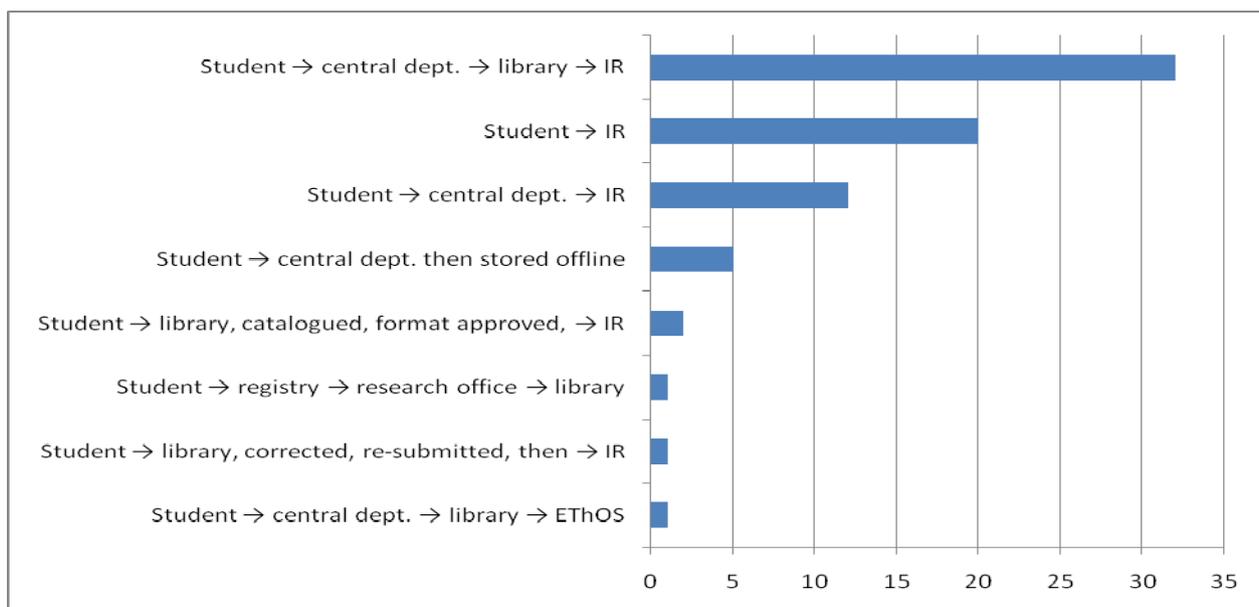


Figure 1: Identified workflows by number of HEIs.

¹ Project website <http://www.ucl.ac.uk/ls/etheses>

At 32 HEIs, students deposit their theses at a central department. This department then passes the thesis to the library, which archives the thesis in an institutional repository. The second most frequent pattern (20 institutions) is one in which the student deposits their thesis directly to the institutional repository. 12 HEIs receive e-theses at a central department which then uploads them directly to the IR.

This representation of the workflow is of necessity simplified. There are tasks, such as approving the award of the doctorate, checking copyright compliance or creating metadata records that are not included. A more detailed analysis of responses to this question reveals more diversity in the ways that HEIs manage these tasks across the departments.

Student → central dept → library → IR

There is a generally consistent approach to the distribution of tasks within this group of HEIs. In all cases, after assessment the thesis is submitted to a central department who check the thesis is the examined version, register receipt of the thesis and approve the award of the doctorate. They then pass the thesis on to the library, where staff members catalogue the thesis, and archive it in the IR, as in example 1:

- 1. Thesis (1 print, 1 pdf/electronic) received by Registry*
- 2. Print copy sent to the academic department of the author*
- 3. Electronic copy sent to Library (along with metadata details) for cataloguing*
- 4. Full text placed in Repository by Library cataloguing staff.*

In this example, copyright checks are carried out once the thesis reaches the library, and embargoes are managed using embargo functionality built into their repository software (EPrints). Example 2 highlights the risk of duplication of effort within the library when handling print and e-copies:

- 1. Final submission - one hard copy plus electronic copy on CD handed into College Office.*
- 2. College Office periodically sends batches to Library for cataloguing*
- 3. CD copy is archived in Institutional Repository*
- 4. Hard copy currently catalogued for inclusion in OPAC*
- 5. Work is ongoing to merge steps 3 and 4*

Example 3 also reveals some duplication of effort, but also emphasises that the IR is not the endpoint of e-theses, but a route into a wider infrastructure for the dissemination of theses.

- 1. Thesis received in print and electronic copies by the Registry*
- 2. Deposit form sent to library with print and electronic copy*
- 3. Print copy catalogued by library on catalogue, catalogue record sent to BL for Index to Theses*
- 4. Electronic copy catalogued and deposited in repository by library*
- 5. EThOS harvests electronic copy*

These more detailed summaries of workflow processes show that this model, adopted at 41% of the HEIs in this sample, benefits from relative simplicity, since each stage tends to be self-contained.

Student → IR

The highlighted examples of this workflow illustrate the level of variation which hides behind its apparent simplicity, especially in the area of communication, in which matters can quickly become complex.

1. School sends viva outcome & official documentation to Graduate School
2. Graduate School sends Completion Memo to Subject Librarian
3. Candidate submits to Research Administrator a printed copy for School (if required)
4. Candidate submits the following to Subject Librarian:
 - Printed copy of thesis;
 - Electronic copy of thesis;
 - Deposit of and Access to Thesis Form (including details of restriction if relevant)
5. Subject Librarian checks the integrity & completeness of submitted thesis and:
 - passes thesis to Senior Library Cataloguer (adds metadata to Library Catalogue, deposits thesis in repository, sends metadata to BL)
 - inform Graduate School (original signed copy of Deposit of and Access to Thesis Form)
- 6 Graduate School review all documentation and:
 - recommend conferment of award to Senate (via memo to Registrar);
 - inform Graduation Officer for Graduation arrangements;
 - send award letter to candidate.

It is possible however to simplify the lines of communication by using automated messages and triggers within the process:

1. Print copies received by Research Student Office (RSO), along with Thesis Deposit Agreement form
2. Student submits electronic copy via the online submission form (an automatic email notifies RSO - thesis is stored in the repository at this point, but not visible online)
3. Thesis approved by Senate and Faculty
4. RSO send a print copy to the Library to be catalogued and the electronic version is made visible online in the repository (if the status is "Open")

In this case, uploading to the repository is part of the submission process, but embargoes are managed by the school. This ensures that embargo requests are handled at the point of submission, further clearing the lines of communication and ensuring that responsibilities are clearly demarcated.

1. Student submits print copy and Access Form detailing any embargo requests (plus e-version on CD-ROM/memory stick if an embargo has been requested) to Faculty Graduate School.

2. *Student registers with the [thesis repository], submitting metadata for the thesis and uploads an e-version in PDF format (if no embargo issues apply).*
3. *Library confirms registration with both Faculty and student. Library requests the print copy of the thesis and Access Form (as well as the e-version if an embargo has been requested).*
4. *Once Access Form has been received Library makes the thesis record available publically via the [thesis repository] site.*
5. *If an embargo has been requested Library files the CD/memory stick and backs up the e-version on an external drive.*
6. *Once the print copy has been received Library catalogues both the print and e-versions linking the repository record with the catalogue record.*

These examples make it clear that the apparent simplicity of authors submitting their own theses to the repository is deceptive. Links between departments and processes can be structured flexibly in this model, but trigger points for essential tasks and communications must be carefully considered in advance to reduce the risk of broken or overly complex workflows.

Student → central dept. → IR

This process seems to work best in institutions only accepting e-theses (as opposed to print and electronic) as it obviates the need for library staff to process print copies separately, essentially removing one stage or point of communication from the procedure.

- 1 *Theses emailed with forms, Copyright forms and statements and permissions (if any) to Graduate research office by students*
- 2 *Graduate Research Office perform checks re eligibility*
- 3 *Graduate Research Office upload eThesis*
- 4 *IR staff perform checks re metadata and copyright*
- 5 *IR staff release the item*

Workflows are also naturally compartmentalised in this workflow, which serves to highlight the potential efficiency benefits from a move to electronic-only.

Conclusion

It is clear that there are a number of issues in managing e-theses workflows. While there are immediate and obvious efficiency gains to the HEI from the electronic archiving and dissemination of theses, these gains are not always realised in the internal processing of theses.

The apparent simplicity of direct upload to a repository is deceptive, as the award of a research degree involves numerous departments within the university and the arrangement of points of communication and handover of tasks require careful consideration. A structure in which the

completion of tasks by one department automatically triggers the next stage in the process (for example, upload triggers a message to the registry which approves the award which triggers a message and so on) combined with clear demarcation (for example, schools handle embargoes while the repository handles copyright checks) is the the most effective way of minimising the potential inefficiencies or complexities of adapting workflows to electronic thesis deposit.

Converting to an electronic-only environment in which a minimum number of departments handle the theses offers the greatest efficiency savings, and avoids duplication of effort; but those institutions operating a 'hybrid' or 'parallel' system and archiving both electronic and print theses can still realise significant benefits from the change by, for example, arranging systems to eliminate duplication, such as generating catalogue records by converting Dublin Core to MARC.