Unlocking Thesis Data

University of Southampton case study

Stephen Grace

<http://orcid.org/0000-0001-8874-2761>

Sara Gould

<http://orcid.org/0000-0003-2763-9755>

July 2015

DOI:[10.15123/PUB.4302](http://dx.doi.org/10.15123/PUB.4302)

 © The authors 2015. Licenced under Creative Commons Attribution 3.0 Unported License.



Overview of Southampton

1: Interviews

The following people were interviewed as part of the case study:

1. Aline Giordano, PGR Support Manager, Student & Academic Administration
2. Isobel Stark, Academic Liaison Librarian, Library
3. Kate Walker, Academic Liaison Librarian, Library
4. Dorothy Byatt, Academic Liaison Librarian, Library
5. Michael Whitton, Academic Liaison Librarian, Library
6. Simon de Montfalcon, Senior Editor, Library

2: Summary data

|  |  |
| --- | --- |
| **Summary Table** |  |
| Institution | University of Southampton |
| Higher Degrees Awarded in 2012/13 | 600 ([[1]](#footnote-2)) |
| Publications repository | EPrints (ePrints Soton) |
| Publication repository IDs | EPrints IDs / URIs |
| Data repository | EPrints (ePrints Soton) |
| Data repository IDs | DataCite DOIs |
| Theses held in… | Hard-copy and electronic |

Southampton case study

Southampton has an in-house IT system for managing PGR students (PGR Tracker), and most interactions are recorded on this system.

Administration of students is shared between the Faculty Graduate School and supervisory teams. Southampton aims to support the personal and professional skills development of research students using an academic needs analysis (in line with the Vitae Researcher Development Framework ([[2]](#footnote-3))) and subsequently as part of annual monitoring reports.

1: The student arrives

A student either arrives at Southampton with an outline research idea, having applied and been allocated a supervisor, or the university has a clear proposal and recruits a student to undertake a particular project (typically with funding from a Research Council, Charity or University studentship). Students are currently registered for MPhil/PhD and upgrade to the latter following an official process, but from 2016 entry there will be direct PhD registration with a confirmation process.

Full-time students have three years to research their project and another year to write up. They then submit their thesis for examination within their period of candidature.

2: The thesis is submitted

When ready, the student applies for examination using an Intention to Submit form which includes the provisional title (this may be subject to change following examination). Examiners are appointed and the thesis is submitted via the Faculty Graduate School, with soft copies for each of the examiners. A viva is held, and the examiners recommend one of five actions to the research degrees committee:

* Award of the thesis as submitted
* Award with minor amendments: resubmission of an amended thesis within one month, with a document addressing issues raised by the examiners
* Award with modest amendments: resubmission should normally be no later than six months
* Student is required to attend a further oral examination
* Student is permitted to resubmit their thesis – normally no later than 12 months
* Student is permitted to apply for the award of MPhil
* No award or resubmission is permitted

Students submit their thesis to their Faculty Graduate School (print and electronic) with the necessary forms for deposit, access and any embargoes. The e-thesis may be presented on CD, DVD, USB stick or via email. Administrative staff enter information (Name, Title, Supervisor, Abstract, Department and Publication Date) on ePrints Soton. Faculties have different practices and there can be significant delays in uploading theses or transferring their print equivalent to the library. There is no feedback to the student, who does not know what has happened to his or her thesis or whether it has been made available.

Library staff offer monthly training for students on preparing for submission covering formats and templates, copyright, when and how to submit, research data, ORCiD identifiers and reference management. First and third year students typically attend, with the occasional second year depending on discipline. Southampton requires both print (bound) and e-theses from students, and award of a PhD (or MPhil) degree is conditional on the final version being submitted in both forms.

Regulations on the format of the thesis are detailed in the Guidance for Completion of Research Degree, available on the website [http://tinyurl.com/qespmzs, also linked via <http://library.soton.ac.uk/thesis> ]. The Faculty of Physical Sciences and Engineering has more detailed guidance on their web pages concerning layout, data, etc.

Ownership of data is a complex area, governed by regulations on IPR and Research Data Management, and not always well understood by academics. If no (other) intellectual property rights exist in the Research Data or subset thereof, the University is the owner of all legal rights. Clarifying how data will be managed etc. can be important to agree at an early stage in a student’s studies.

There is guidance on appendices and on ‘unusual’ objects which may be included in a submission. Any appendix can have its own embargo period, different to that of the main thesis. In practice, supplementary files of data objects in their ‘native’ formats are rarely submitted – students still tend to submit data as appendices in a single PDF text file.

3: The final thesis is processed

A team of Senior Library Assistants in the Hartley Library process the two versions of theses. They check the deposit form and embargo information accompanying the print thesis, contacting the Graduate School at the end of any embargo period. Students may request an embargo of up to three years, countersigned by their supervisor, and they may also renew an existing embargo using a standard procedure managed by the Library.

Meanwhile, the e-thesis is briefly inspected to check it opens as a valid PDF (with similar checks on other formats for supplementary files), has adequate metadata and conforms to the student’s declaration form. The quality of metadata submitted by faculties varies, and is best where there is an established administrator familiar with using ePrints Soton for theses.

There is a more detailed inspection of data deposits by the Research Data team in the Library for file integrity and documentation. Because students can choose to add data as supplementary files for their thesis or separately as data collections, it would help if a statement or similar was required to clarify the relationship between the various objects that make up a thesis “collection” – or that a data collection is related to a thesis. (Note there is a requirement to include a ‘List of accompanying material, if any’ in the Guidance for Completion of Research Degree.)

Currently details of funders and ORCiDs for student or supervisor are not collected routinely. Details of associated funding can be added to ePrints Soton, and students can register their ORCiD with the University. Southampton does not apply Creative Commons licences to theses, but does offer a standard licence allowing access while respecting copyright and moral rights.

4: Cultures of data sharing

Southampton has a well-established culture of research data management support, with a cross-functional support service co-ordinated by the library. It has made provision for handling datasets in ePrints Soton.

Thesis data is mentioned in the “Introduction to RDM” workshop run by the library, and students can choose to deposit data with the thesis or separately (there are not many of the latter yet).

Separate deposit of data is permitted. Students require approval from their supervisor to deposit data as a separate output. Where data are deposited separately they could precede or follow thesis deposit, and it is not always clear there is a relationship between the two; it would help if the deposit form prompts users for this information. Deposit requires a current University of Southampton account, so there is a point in time after which students cannot deposit their data themselves (but staff in the Library can do this on their behalf on request).

Embargoes can be applied to datasets where they underlie a publication, and separate embargo periods can apply to publications (including theses) and datasets.

Data deposits can have a DOI for discovery and citation; where data relates to a thesis, a DOI will not be created until after the viva has been held. The process for creating DOIs is manual, requiring both an XML editor to create the necessary metadata and manual use of the DataCite Metadata Store. DOIs for data are registered by the Research Data team on request from the student (at an appropriate stage). Southampton is in the process of implementing Pure (while retaining ePrints as the public interface). DOI registration for Datasets is expected to happen in Pure (which includes this functionality).

There are cultures of sharing in specific disciplines, but others face potential barriers: music, for instance, has copyright and performance rights issues while physics has a requirement to share data to enable the progress of science. Social science (and some humanities) research involving human participants necessitates more detailed checking of the content of theses. Students, in the main, were thought to lack confidence in exposing their data, while some academics resisted data sharing with their peers.

5. Reflections on the process

Southampton has clear procedures for handling doctoral theses, which reflect the more devolved nature of the institution. Faculty and central Graduate School administrators and Library staff have developed working practices to ensure the high quality of records on ePrints Soton and availability (subject to any embargoes) of e-theses. Due to the high volume of awards each year there can be delays in making the final versions of theses available. Southampton still requires bound (print) and electronic versions of a student’s thesis, with metadata being created for each.

More supporting information for students which welcomes data deposits, with examples of the types of data that could be submitted, would help students to consider including such data with their thesis. It might also help students to discuss data (and software) availability with their supervisors.

Students can self-archive their data and other publications within the scope of ePrints Soton, but not their thesis. A number of paper forms are submitted which could be replaced by more efficient web forms or by using an appropriate system with the right permissions and approvals. This is not a trivial matter for the university, however.

Applying Persistent Identifiers

It would be relatively easy for Southampton to adopt Persistent Identifiers (PIDs) for students and theses.

ORCID

The Southampton ORCiD system (<https://orcid.soton.ac.uk>) allows students and staff to create an ORCiD account if needed, and link it to their Southampton identity. This was created as part of the Jisc ARMA ORCiD project at Southampton (<http://blog.soton.ac.uk/orcid>).

This service could be further promoted, at a point soon after the arrival of new students. Also it may be possible to integrate this service with appropriate forms and systems.

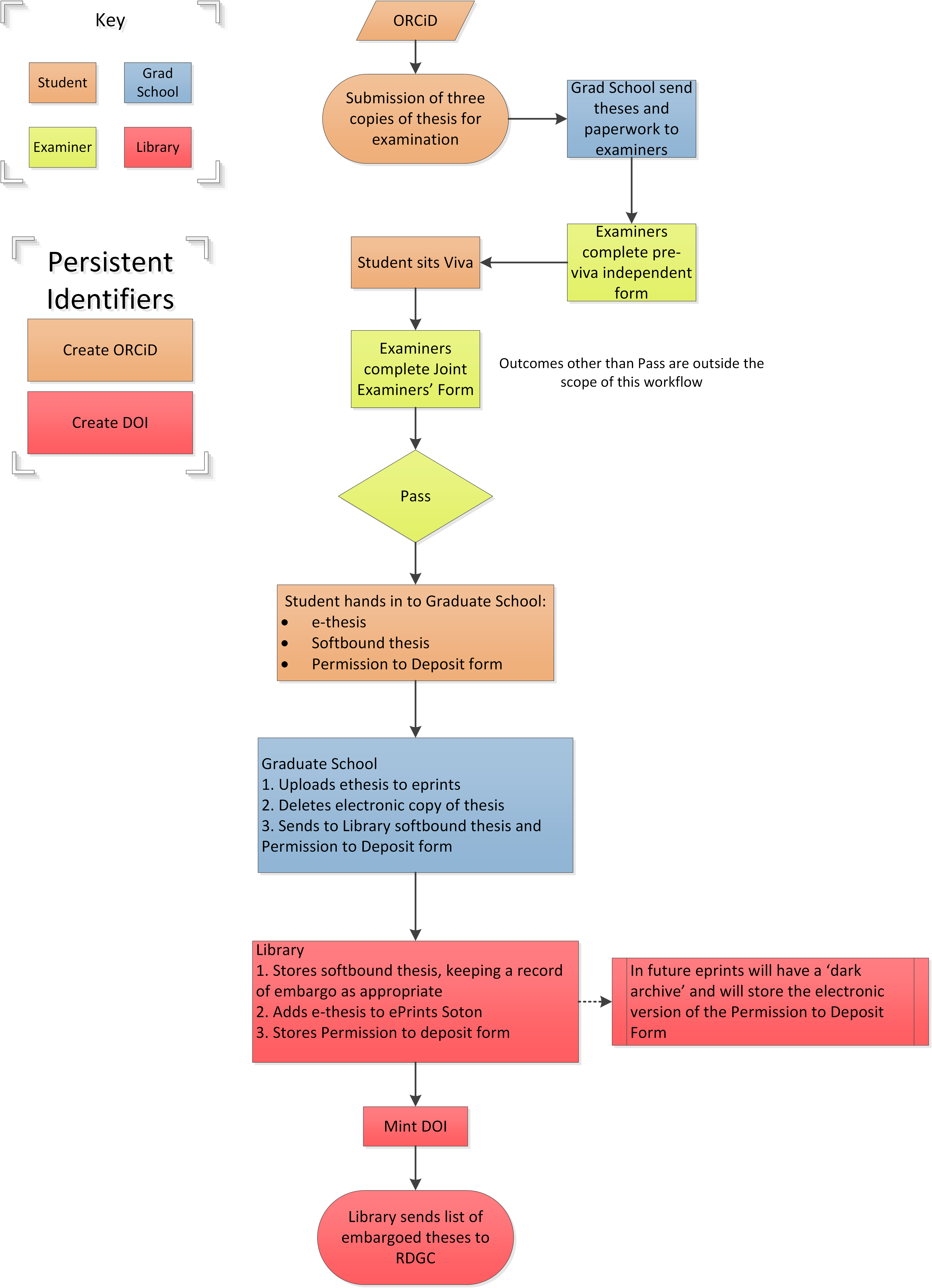
DOI

DataCite DOIs could be assigned on publication of the thesis, making use of the existing DataCite membership. A more automated service, such as using the EPrints DataCite plugin would speed up processes. Southampton has a preference for assigning a single DOI to each record in the repository (i.e. to the thesis, and to the datasets where they exist) .Assigning DOIs to each file (where there is more than one), is not currently possible and is not a priority.

Recommendations

1. Adopt ORCID and DOI identifiers as outlined above.
2. Consider replacement of paper forms where data could be entered onto systems or extracted from them to improve efficiency and reduce delays.
3. Provide further guidance welcoming submission of data/software files with theses, together with examples to stimulate consideration by students and their supervisors.
4. Consider the continuing value of requiring both print and electronic versions of the same thesis, and the costs incurred by students and the university as a result.
5. When the thesis is made live in ePrints Soton, send an email to the student informing them of its availability; include the DOI in the email, and encourage the student to use it whenever citing or linking.

Appendix: Suggested workflow with PID overlay



1. From HESA Table 18a, Higher Degree (Research) Qualifiers by Institution and Subject of Study [↑](#footnote-ref-2)
2. Vitae (2010), Introducing the Researcher Development Framework. Cambridge: Vitae. Available at <https://www.vitae.ac.uk/vitae-publications/rdf-related/introducing-the-vitae-researcher-development-framework-rdf-to-employers-2011.pdf> [↑](#footnote-ref-3)