Unlocking Thesis Data

University of Bristol case study

Rachael Kotarski

[http://orcid.org/0000-0001-6843-7960](http://orcid.org/0000-0001-8874-2761)

Michael Whitton

[http://orcid.org/0000-0001-6838-1100](http://orcid.org/0000-0003-2763-9755)

July 2015

DOI: [10.15123/PUB.4305](http://dx.doi.org/10.15123/PUB.4305)

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

Contents

Overview of Bristol 3

Workflows at Bristol 4

Data and supplementary files 7

Use of identifiers 8

Moving forward 9

Recommendations 10

Overview of Bristol

1: Summary data

|  |  |
| --- | --- |
| **Summary Table** |  |
| Institution | University of Bristol |
| Higher Degrees Awarded in 2012/13 ([[1]](#footnote-1)) | 590 |
| Publications repository | Pure |
| Publication repository IDs  | Handles |
| Data repository | CKAN (data.bris) |
| Data repository IDs | DataCite DOIs |
| Theses held in… | Hard-copy only |

2: Summary profile

The University of Bristol currently does not have a process for electronic (e-) theses. Their current practice and workflow is primarily for the deposit of hard-copy theses. This process is coordinated by Bristol's Exams Office.

Note that an electronic copy of the initial thesis submission is requested, but is currently only used for a pre-examination plagiarism check via TurnItIn ([[2]](#footnote-2)). The electronic copy is discarded after the plagiarism check. TurnItIn can be set to save the thesis after checking (for marking etc.).

Workflows at Bristol

The current process at the University of Bristol for thesis submission, examination and deposit is shown in Figure 1. This shows that the bulk of the information created prior to deposit of the thesis at the library is carried out within the examinations office and the student's school.

As this is currently a process for a physical thesis, there is little automatic information transfer through the chain between the departments involved.

The examinations office is first aware of a thesis when they receive an Appointment of Research Degree Examiners (ARDE) form ([[3]](#footnote-3)). This form will include the title of the thesis and its author (the student). There is central advice on thesis submission and more detailed advice within each school. There is a code of practice that covers submission, and each school's detailed guidance sits within the general framework.

Hard copy theses are then submitted alongside an electronic copy which is solely used for plagiarism checking using TurnItIn. The electronic copy is to TurnItIn via Blackboard. The process of plagiarism checking is out of scope of this case study, but details of the process can be found in Bristol's Regulations and Code of Practice for Research Degree Programmes 2014/15, Annex 8 (University of Bristol, 2014).

The TurnItIn report is handled only within the School, who will advise the Exams Office of the outcome of the plagiarism check. If the plagiarism check is satisfactory, the Exams Office will send the soft-bound copies of the thesis to the examiners. At this point, the ARDE form has already been used by the School to confirm that the examiners are suitable and arrange examination.

After the viva has been completed, the examiners reports are sent onto the Exams Office. The ultimate decision to award the doctorate lies with the Research Degree Examinations Board. From this point a number of situations may arise, depending on the decision of the Research Degree Examinations Board (e.g. PhD awarded, minor corrections, major corrections, etc.). In the majority of cases, a finalised, hard-bound thesis is submitted to the Exams Office.

Successful students are asked to complete a Thesis Deposit form to accompany the hard-bound thesis. Where agreement is given, the Thesis Deposit Form enables the Library to make the hard copy available for consultation; the author may also agree to make their thesis available for digitisation at a later date and accessible online via EThOS or the institutional repository. The Thesis Deposit form is held inside the physical thesis at the library. It should be noted that up until this point, all the forms in the process (ARDE, examiners reports and the thesis deposit form) require details of the title of the thesis and student ID.

Students may in some cases request deferred access to the thesis. This is made via a Request to Defer Access to a Research Degree Thesis form ([[4]](#footnote-4)). This request by the student must be recommended by the student's Supervisor and the Faculty Graduate Education Director and must then be approved by the Academic Director of Graduate Studies. Its intention is to protect the students' ability to take advantage of the research, but also in cases where the materials within the thesis are sensitive. Where the work is embargoed, it is stored separately in the Library until the embargo ends.



**Figure 1. The process of thesis submission, examination and deposit at the University of Bristol.**

The University of Bristol Library receives a box of hard-bound theses from the Exams Office. Each thesis is then catalogued in the library management system, Aleph (using the MARC21 metadata format for bibliographic data and using RDA cataloguing standards). This is a manual process and at busy times can create a backlog of work. These records are then harvested by the library's discovery service, which is an instance of Primo ([[5]](#footnote-5)) (see Figure 2).



**Figure 2. Example of a thesis record in Bristol's instance of Primo.**

Requests for online access to University of Bristol theses are directed to EThOS in the first instance and then passed on to the Library; where agreement has been given by the author the hard copy will be sent to EThOS for digitisation and then made available online via EThOS.

Data and supplementary files

The University of Bristol has developed a research data repository based on the CKAN platform ([[6]](#footnote-6)). Bristol's data policy is relatively mature, and they have been receiving and making data sets public on data.bris.ac.uk since 2013.

Researchers at Bristol have 5TB of dedicated file storage space, which integrates with CKAN for data publishing. They each have 400GB of publication space. However, postgraduate researchers do not have their own dedicated storage at present – they have to use their supervisor's storage.

When data is deposited, the metadata etc. are reviewed by the data.bris team prior to being published in the repository. The description of the data in the metadata record is checked for accuracy, relevance and readability. In addition all zipped files are extracted, folder structures explored and individual files opened (where possible). Once published, datasets in data.bris are assigned Digital Object Identifiers (DOIs), as shown in Figure 3.



**Figure 3. Example of a research data set record on data.Bris.**

Use of identifiers

The University's instance of Pure currently has a Handle server assigning Handles as the persistent identifier for the full-text publications held there. This functionality does not currently extend to theses, although a request has been made to the Pure software suppliers to add it.

As there are few records of electronic theses, there are currently no links between theses and datasets at the University of Bristol. Integration between data.bris and Pure is planned for the future, at which point links between theses and the dataset could be made.

The University has been an ORCiD institutional member since March 2015. The University’s instance of Pure is integrated with the ORCiD API, making it possible for researchers to register for an ORCiD from within Pure ([[7]](#footnote-7)). The University sees potential in being able to register or capture ORCiDs as part of student admissions and staff recruitment processes, and will be exploring this over the coming months.

Moving forward

The University of Bristol is just beginning an internal change process to move to full electronic submission and handling of theses and the theses examination process. A project has been proposed and a business case is being developed so that resources can be allocated, with an aim of having electronic submission in place for the 2016/17 academic year.

Our interview with representatives of the Examinations office and the Library highlighted that they are aiming for an end-to-end, streamlined process. This should help to remove the administrative burden, for instance currently the thesis title and student ID are entered into at least three different forms by a number of different roles (the student, their supervisor, the examiners) before the thesis itself is even catalogued.

They are aware the move to e-theses will require more consistency and advocacy for processes, compared with the current hard-copy submission. Some specific issues include:

* A noticeable apprehension about the copyright issues, and where the responsibility lies for checking copyright
* Potential greater need for embargos e.g. where the thesis (or parts of it) will be published or where the student has an industrial partner
* Potential opportunities to standardise processes across schools and faculties
* Dealing with supplementary files and datasets
* Changes in how data is exchanged with EThOS – sending metadata about electronic theses from Pure to EthOS.

A recommendation on the appropriate points for the University of Bristol to apply identifiers in the eventual electronic thesis workflow is difficult to provide at this point, and will be influenced by other decisions such as the identifier used, the submission routes of the thesis to the Library and how and where metadata is created.

Recommendations

1. Registration with ORCIDs can be promoted early in the PhD process after registration. Pure can enable researchers to register for an ORCiD or link their existing one to Bristol.
2. The institution needs to decide whether to use DOIs for Theses – or whether just to use handles. This may depend on the difficulties of using two persistent identifier systems within Pure and the outcomes of the Unlocking Thesis Data work.
3. Elsevier need to be lobbied via the Pure UK User Group to add any required functionality to allow DOIs to be registered for Theses.
4. Datasets stored in data.bris will be given DOIs as per current practice.

Appendix: Suggested workflow with PID overlay



Works Cited

Higher Education Statistics Agency. (2015). *Table 18a, Higher Degree (Research) Qualifiers by Institution and Subject of Study, 2012/13.* Higher Education Statistics Agency.

University of Bristol. (2014). *Regulations and Code of Practice for Research Degree Programmes 2014/15.* Retrieved June 06, 2015, from University of Bristol, Academic Registry / Examinations Office: http://www.bristol.ac.uk/media-library/sites/academic-quality/migrated/documents/pgrcop14-15.pdf

1. (Higher Education Statistics Agency, 2015) [↑](#footnote-ref-1)
2. <http://turnitin.com/> [↑](#footnote-ref-2)
3. <http://www.bristol.ac.uk/media-library/sites/exams/migrated/documents/arde-form.doc> [↑](#footnote-ref-3)
4. <http://www.bristol.ac.uk/media-library/sites/exams/migrated/documents/deferred-access.doc> [↑](#footnote-ref-4)
5. Primo is a discovery tool developed by Ex Libris, allowing aggregation of items from multiple metadata sources. <http://www.exlibrisgroup.com/category/PrimoOverview> [↑](#footnote-ref-5)
6. An open-source data portal platform developed by the Open Knowledge Foundation <http://ckan.org/> [↑](#footnote-ref-6)
7. <http://www.bristol.ac.uk/red/research-policy/orcid/pure-and-orcid/> [↑](#footnote-ref-7)