Unlocking Thesis Data

Phase 1 survey of UK Higher Education Institutions

Sara Gould, Stephen Grace, Rachael Kotarski and Michael Whitton

DOI: 10.15123/PUB.4274

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

Contents

Introduction 3

Aims of the survey 4

Key findings 5

Analysis of the results 6

Conclusions and indicators for further activities in Phase 1 and beyond 16

Appendices 18

Publication details

|  |  |
| --- | --- |
| Title | Unlocking Thesis Data: Phase 1 survey of UK Higher Education Institutions |
| Authors | Sara Gould, The British Library (http://orcid.org/0000-0003-2763-9755)  Stephen Grace, University of East London (http://0000-0001-8874-2671)  Rachael Kotarski, The British Library (orcid.org/0000-0001-6843-7960)  Michael Whitton, University of Southampton (http://orcid.org/0000-0001-6838-1100) |
| Date | 23 June 2015 |
| URL | <http://dx.doi.org/10.15123/PUB.4274> |
| Version | 1.0 |

Introduction

Unlocking Thesis Data (UTD) is a community-driven project to promote the use of persistent identifiers for theses, their underlying data and their authors.

By their very nature, PhD theses break new ground and advance scholarly knowledge. Most make use of newly-created data but these data can be trapped in an appendix or DVD – either unavailable or not suited for reuse. UTD will make data more discoverable and citeable, thereby offering incentives to students to share their data in more appropriate formats, in the context of a sustainable national thesis framework.

UTD is led by the Universities of East London and Southampton and EThOS, the UK national thesis service at the British Library.

The project is envisaged in three sections: Phase 1 will explore current thesis practice through an online survey and individual case studies which will also look at how institutions might apply Digital Object Identifiers (DOIs) and ORCiD identifiers. If successful, Phases 2 and 3 are expected to enhance metadata and software for applying DOI and ORCiD identifiers, test them in live settings, and offer comprehensive advice for institutions to adopt them. By summer 2016 we expect to have a sustainable infrastructure covering the whole UK, driving the wider availability of research data emanating from PhD research, and introducing doctoral students to new norms of scholarly communication.

Data availability

**The data from the survey is available for reuse and analysis at** <http://dx.doi.org/10.15123/DATA.12> under a Creative Commons CC0 licence**.**

Aims of the survey

The first task for the project was to understand the current situation as it relates to thesis identifiers, thesis authors, uptake of DOI identifiers in HE institutions for other research outputs, and the management of supplementary data files and raw data produced in the course of the PhD.

The survey aimed to:

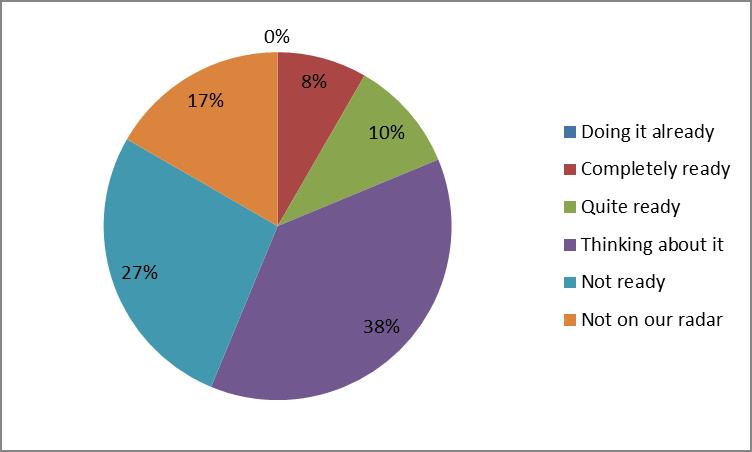
* Gather quantitative data on the level of DOI and ORCiD usage so far for outputs other than theses
* Understand typical workflows for the handling of theses in institutions, from student deposit to repository access and citation
* Begin to note ‘non-typical’ scenarios for which solutions may be needed
* Gauge the level of interest in assignment of unique identifiers to institutions’ theses and the UTD project
* Begin to identify how assignment of DOIs might fit into current workflows, and spot potential pinch points
* Gather volunteer institutions for the next UTD Phase 1 task, institutional case studies.

The survey was sent to a named contact at 138 HE institutions which offer research degrees. This list was drawn from the contact list held by the British Library for the EThOS E-Theses Service.

The survey was launched on 27 April 2015 and ran for three weeks.

Key findings

1. 49 institutions (35.5%) responded to the survey, indicating that *Unlocking Thesis Data* is of interest to a significant proportion of HE institutions. A list of respondents is provided in Appendix A.
2. At the time of the survey, no institution assigned DOI identifiers for their theses, although DataCite DOIs were used by 33% of institutions for their research data.
3. Around 59% of institutions require students to submit both print and e-copies of their final thesis, and this often results in double-handling, for example in creating separate records for the catalogue and repository. This may have implications for UTD.
4. The most ‘typical’ scenario is an institution which uses an EPrints repository for its e-theses and supporting files, students must submit both print and electronic versions of their thesis, the thesis is uploaded and the metadata created by the repository staff (though students are a close second), and the institution assigns DOIs for its datasets but not theses. This suggests such a scenario might form the first case study or the core focus for UTD.
5. In response to the question *How ready are you to begin assigning DOI identifiers to your theses?* institutions varied from ‘*Completely ready’* to *‘DOIs are not on our radar at all’*:



*Q16: How ready are you to begin assigning DOI identifiers to your theses?*

The intention is to ask this question again at the end of the full UTD project; this will form a key indicator as to the success of the project.

1. Twenty-four respondents (49%) volunteered their institution to be a case study for UTD. The aim is to deliver just six case studies under Phase 1, and we hope those institutions not selected will be willing to host UTD clinics, become early adopters or have other opportunities to be closely involved.

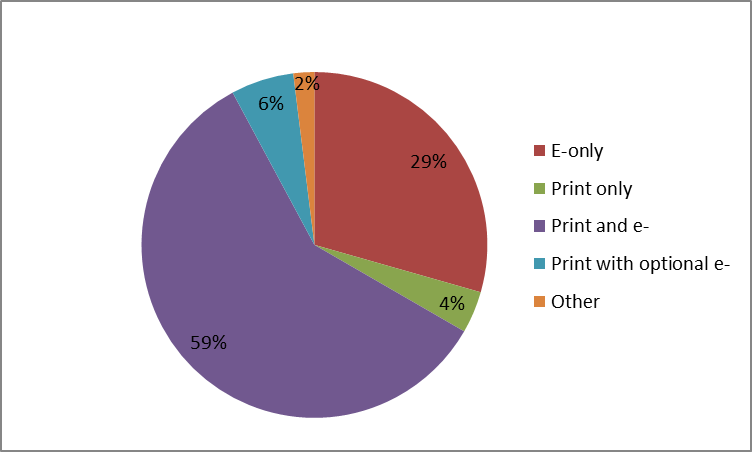
Analysis of the results

1: Student submission

The majority of institutions (59%) require their students to submit both print and electronic copies of the final thesis. This often results in doubling handling and the creation of two sets of metadata, presenting implications for UTD since either copy has the potential to be cited and accessed. Although DOIs can in theory be assigned to digital records of physical items, UTD expects to restrict its attention to digital theses only. In any case, many (but by no means all) institutions which still handle print appear to be in transition:

*“Print deposit is under review and we hope to move to e-only.”*

*“Both must be deposited with library for archiving but not mandatory to make available.”*



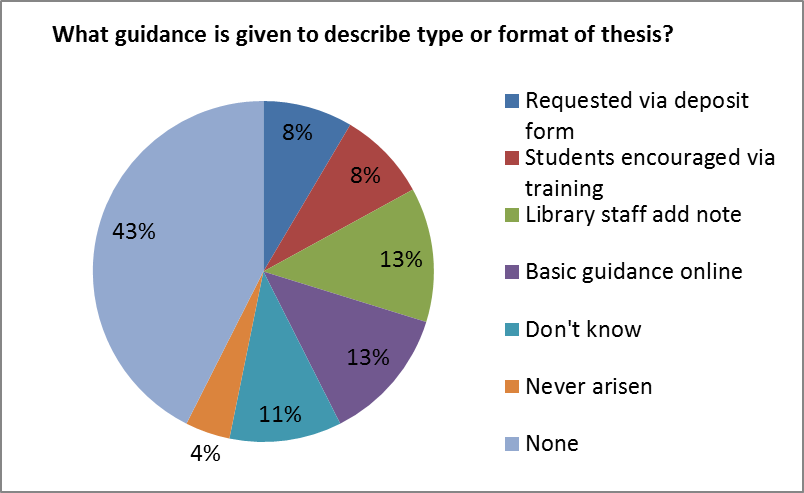
*Q2: What are your institution’s current requirements for student deposit of doctoral theses?*

The survey also asked whether there are requirements in place for students to describe the type or format of the thesis and the existence of any supplementary material. For example, are students required to state that the thesis consists of a “main work plus supporting DVD”, or a “published novel plus supporting critical analysis”? This would seem desirable so that users (and indeed the library of the awarding institution) would know the extent of the full thesis and would not miss the existence of any supplementary data files, creative elements or appendices. A note in the abstract would also help the user understand the relationship between the files attached to the record where there is more than the single main thesis file.

The question invited free text responses and 47 people responded, many giving links to online guidance and/or useful notes about future plans:

*“Guidance will be in our website, but currently is given to students attending training sessions.”*

Responses can be grouped to see the overall picture – and almost half of responding institutions have no guidance in place at all. Online guidance is often scant and in any case requires the student to read and follow the guidance with no checks in place; similarly, some institutions rely on the student training sessions to ask them to note the existence of supplementary data.



*Q12: What guidance or requirement is there for students to describe the type or format of the thesis and the existence of any supplementary material?*

*“No requirement but there should be!”*

Just two institutions mentioned a descriptor field where such information would be recorded:

*“We do the description within our team and can ask the student for more information. We have a dc.description field, though we would like something richer.”*

A possible exemplar for uploading and describing additional files is provided by University of Glasgow, <http://theses.gla.ac.uk/format.html>**.** Guidance includes file naming for additional files and depositing associated multimedia files. However even these extensive guidelines do not ask the student to list the constituent parts of the thesis or additional files beyond appropriate naming of each file. Similarly the University of Southampton responded that *“Guidance for Completion of Research Degree states that the number of volumes and a list of accompanying material should be provided.”*

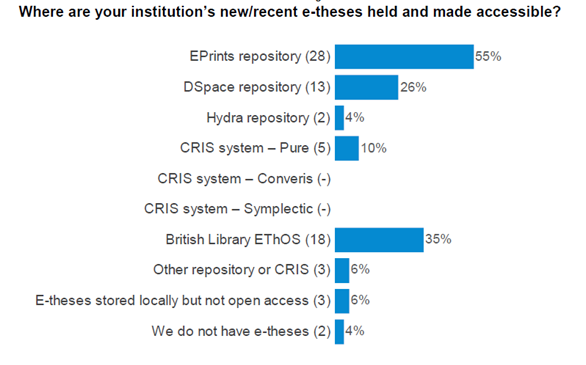
2: Workflows for theses

Institutions use a range of repositories for their e-theses but EPrints is by far the most used system, with 55% of respondents using EPrints, 26% using DSpace, 10% a Pure CRIS (Current Research Information System), and 4% using Hydra. Eighteen respondents also use EThOS for making their theses available in addition to their local system. As with student deposits, many institutions are in the throes of moving to new systems and/or moving their theses solution to align with their other systems. There is a lot of activity around CRIS's. Theses are not always handled in the same system or workflow as other research outputs, and some institutions have quite complicated situations:

*“We are looking to incorporate e-theses into our repository …”*

*“The service is currently under development and is expected to be formally approved shortly.”*

*“We have a Pure CRIS for all University publications. This currently excludes theses. University publications with open access full text are also in our DSpace repository so this is a subset of the total publications in the CRIS.”*



*Q3: Where are your new/recent e-theses held and made accessible?*

Thinking about supplementary data files relating to the thesis, the majority of institutions which manage such items store the files in the same repository or CRIS as the thesis itself (67%). Many institutions indicated that students may deposit or send any type of supplementary file alongside the main thesis, which is almost always a PDF. These would be ingested into the repository and attached to the main thesis record, unless the file is too big and needs to be stored in an alternative local store. Several institutions reported that students are invited to contact the library to discuss the options for their additional files, again suggesting that the sharing of supplementary files is not yet the rule, workflows are not yet established and their value not yet recognised by the students.

*“[supplementary files] Uploaded as separate files attached to main record.”*

*“We offer two approaches – files added to the theses record in the repository and/or a separate dataset record. We prefer all examined material to be in the same record.”*

On the other hand, 22% indicated that the situation had not yet arisen; considering the many comments along the lines of “*We don’t know*” and “*This is very much a work in progress*”, it is clearly early days for institutions in supporting or encouraging their students to deposit material other than the core thesis document itself. It will be useful to repeat this question in a new survey at the end of UTD to assess progress in making supplementary material more widely available.

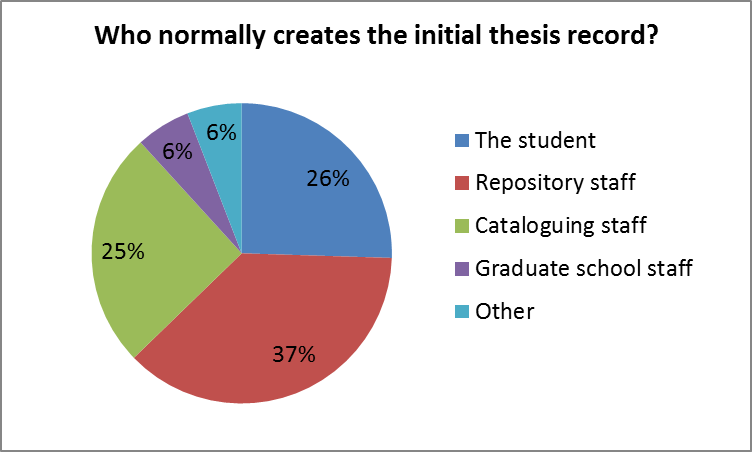
*“We don’t currently have a process for depositing supplementary data files but we expect it to be the Research Data Repository eventually.”*

3: Creating the metadata

Institutions were asked where the initial metadata record for the thesis is created and by whom. The DOI for the thesis needs to be created and included as early in the workflow as possible so that it is included whenever the record is cited, linked or harvested by aggregators. One institution suggests the DOI needs to be assigned even before the record is created while the student is still writing up the thesis so they can include the DOI in the thesis itself. This means it would be picked up by the first full-text indexing by Google and others. For now though, it is useful to note that 73% of institutions create the first record directly in the repository or CRIS. Although 24% create a catalogue record, this invariably relates only to print copies of the theses.

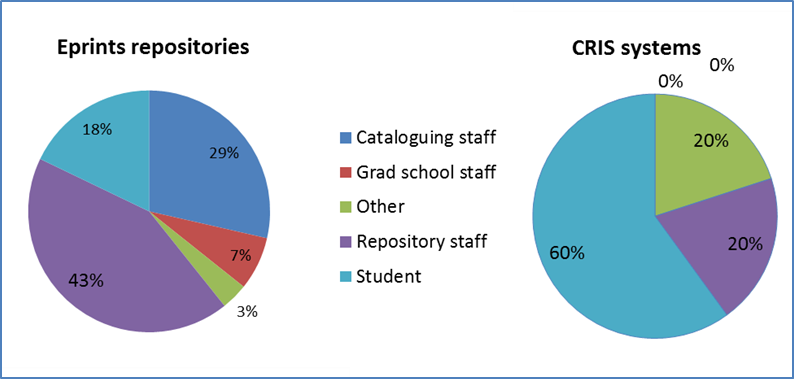
Less clear-cut is the question of *who* creates the first record. This elicited a much more equal balance of responses, with repository staff, cataloguing staff and the student almost all equally likely to be responsible for the task.

This may become a key factor in understanding DOI workflows further down the line. If the record is not created until almost the end of the process and is managed entirely by library or graduate school staff, then there may be little argument for assigning the DOI at an earlier stage in the process. This is because neither the student nor their academic school is likely to have engaged in thinking about the future storage location or citation of the thesis at that stage. On the other hand where the student creates the original record for their thesis, this seems to be done at an earlier point in the process - *“On submission of softbound version pre viva. This is updated on receipt of final version after award*” – and it could make a lot of sense to reserve or ‘pre-assign’ a DOI for the student to input at the same time as the rest of the metadata such as title, supervisor, abstract and keywords, thereby raising the student’s awareness that the thesis will become openly accessible and more easily citeable thanks to a DOI.



*Q7: Who normally creates this initial record for the thesis?*

Digging a little further, it seems that the repository system used does not determine who has the responsibility for creating the metadata record. The exception being that responsibility appears more likely to fall on the student if the theses are stored within a CRIS:



*Q7: Who normally creates this initial record for the thesis?; responses from EPrints users and CRIS users*

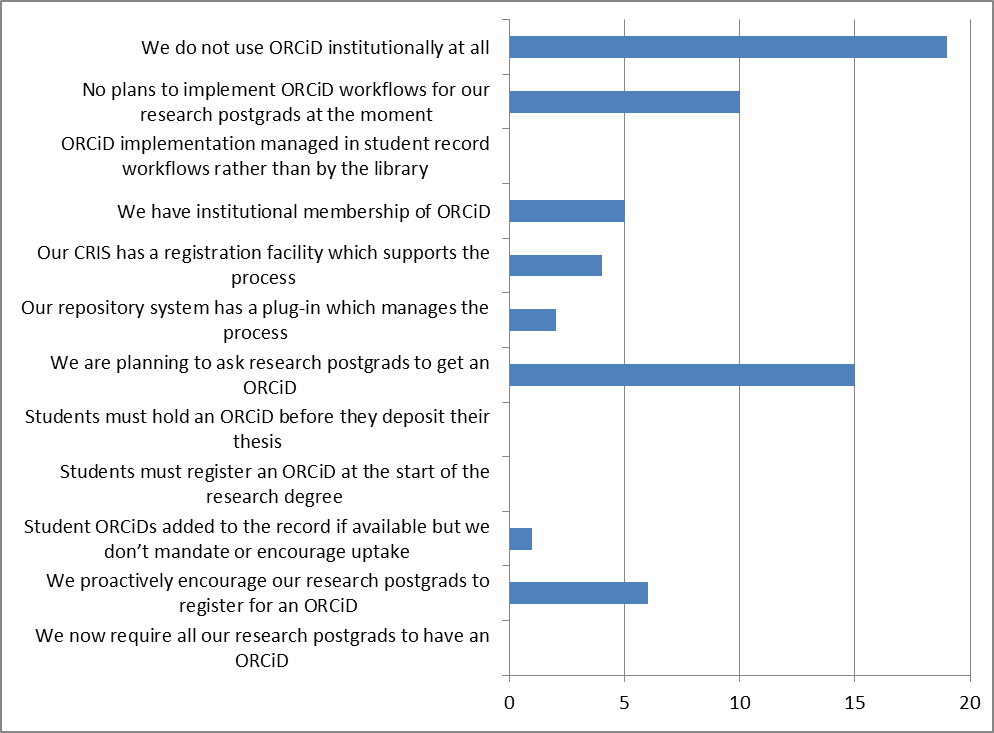
This is a trend borne out by several of the comments as well: *“[Cataloguing staff] Although we are considering moving to e-deposit by the student.”*

*“Currently repository staff but this may well change as currently the upload of the thesis is the only part of the PhD life cycle that is not done by the student/supervisor or Grad school.”*

4: Use of ORCiD identifiers

Institutions were asked about their current level of activity relating to research postgraduates and ORCiD author IDs. Respondents could tick more than one response. The most common position of institutions is to be planning ORCiD uptake for postgraduates while not yet having anything in place. On the other hand six institutions (12%) already encourage their research students to have an ORCiD while far more (40%) have no plans to implement ORCiD workflows for research postgraduates.

We expect to see a good deal of activity around ORCiD uptake in institutions following the successful Jisc/ARMA ORCiD Pilot project and the proposal by Jisc Collections to offer a national HE subscription to ORCiD for all interested institutions. While many institutions may start with assigning ORCiDs for their academic staff, we might quickly see this extended to include postgraduate researchers as well. With doctoral research often the first step in an academic career and the thesis one of the first research outputs, this will be an important step in getting good research and scholarly communication habits embedded in researcher behaviours and institutional workflows early on.



*Q13: Which of these describe your current level of activity relating to your research postgraduates and ORCiD author IDs? [Tick all that apply]*

Additional comments also indicated that ORCiD implementation will continue to grow:

*“This will be mandatory in the near future.”*

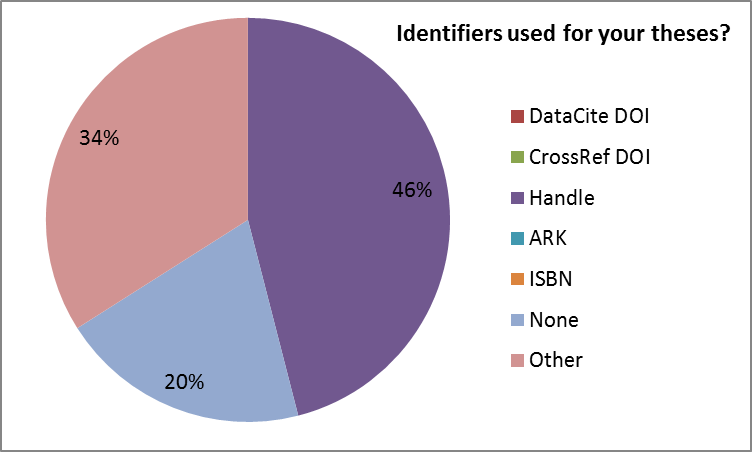
*“We should get a DSpace plugin* [to deal with ORCiDs] *with our next upgrade.”*

A second question on ORCiD asked institutions to provide an example if they could of a thesis whose author has an ORCiD. Perhaps surprisingly no examples were reported at all. This is not to say that no author of a completed thesis has yet to hold an ORCiD profile; in fact we know from other evidence that some academics have already added their EThOS record to their ORCiD profile[[1]](#footnote-1). It would seem there are no workflows yet in place for the author ORCiD to be recorded in the thesis metadata and made available in the repository record – although this is beginning to be taken up by RIOXX[[2]](#footnote-2).

5: Use of DOI and other output identifier

The only identifiers currently in use for doctoral theses in those institutions responding to the survey are Handles. Handle is the identifier system that DOIs are built on. DOIs are based on the Handle identifier system, however there is an additional layer of governance and infrastructure on top of DOIs that helps to further guarantee the persistence of their resolution. Handles are used by 23 responding institutions (46%), including all DSpace repositories (13), eight EPrints repositories and two CRIS. Use in DSpace repositories is unsurprising as Handles come as standard with DSpace.

The response to this question provides an excellent starting point for the UTD project: we will be able to see evidence of the project’s impact as institutions begin to think about how object identifiers could be assigned to their theses and the benefits of doing so.



*Q14: What identifiers do you use now for your theses?*

Ten respondents indicated they use “Other” identifiers, and these included a record ID (named by 4 institutions), persistent URLs (6 EPrints, 1 Fedora, 1 CRIS), EThOS ID (1), URL in the catalogue record (1) and HESA (1).

Several institutions noted their progress towards DOI implementation:

*“We will begin to issue DataCite DOIs for electronic theses in July 2015.”*

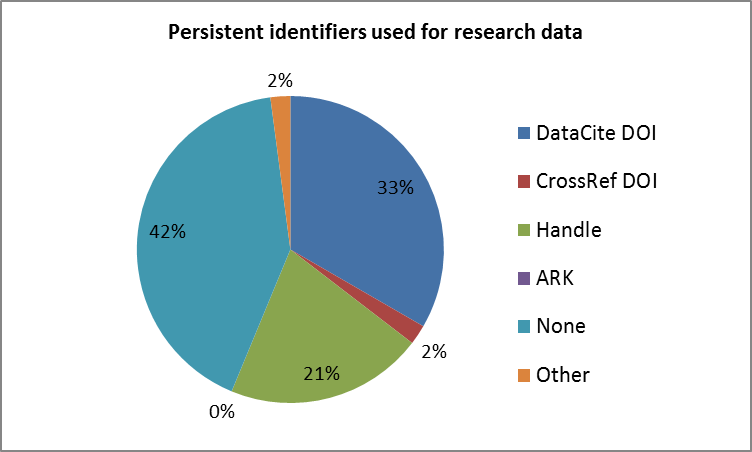
*“Repository URLs are the current identifiers but we want to start using DOI.”*

The contrast between the lack of use of DOIs for theses and their use for research data is striking. When asked what persistent identifiers they use for research data, 17 institutions indicated DOIs (35%). Of these 17 respondents, 16 are DataCite users and one is a CrossRef user. Slightly more institutions reported they do not use any identifier for their data (42%) than those who reported that they do not use any identifier for their theses (34%). This can be explained by the fact that many of them do not currently have responsibility for any research data, especially from postgraduates (although the question did not specify):

*“We do not regularly collect research data from postgraduates.”*

*“At present we do not hold research data or have a workflow for it.”*

*“The data is usually attached to the main record for the thesis, so shares the URL. No other identifier is currently used.”*

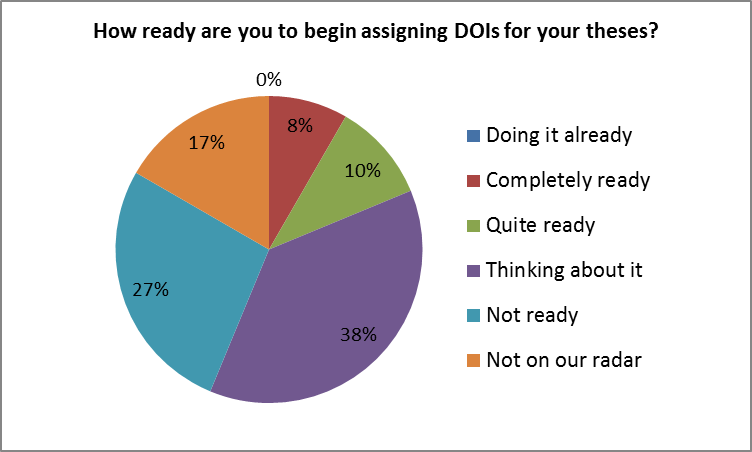


*Q15: What identifiers do you use for your research data?*

6: Readiness for thesis DOIs

We wanted to assess the level of readiness amongst institutions to start assigning thesis DOIs. This information will be useful to help identify exemplar institutions at each stage of the process – potential early adopters for example - but would also provide a baseline against which to measure the success and impact of UTD.

As also indicated by responses to previous survey questions, no UK institution is currently assigning DOIs for their theses, although one (Cambridge) has indicated their intention to start to do so in July 2015. Others (4) responded that they are completely ready and could start at any point. These might emerge as proof of concept institutions under Phase 2 of UTD if further project approval is given by the Jisc Data Spring programme. Thinking about drawing up case studies as part of the Phase 1 mapping work, however, a range of institutions is probably required, taking account of DOI activity for research data, repository system in use, size and nature of institution, as well as the level of preparation already underway. Rather than seeking to map only fully thought-out workflows, we will aim for a range of scenarios from the well-prepared to those institutions who are only just beginning to think about the benefits and challenges of persistent identifiers for theses.



*Q16: How ready are you to begin assigning DOIs for your theses?*

We also asked how institutions are progressing in thinking about assigning DOIs for more complex theses with multiple volumes, supplementary DVDs or data which is deposited with the thesis. In contrast to the 56% who are completely ready, quite ready or thinking about thesis DOIs, very few respondents had begun to consider more complex theses at all. *“Not thought about it.”, “I think we will need more guidance on this.”* and *“Would take advice/see how others are handling this”* are typical of the comments received.

Of those institutions which have started to think about more complex theses, initial thoughts can be divided into two camps, summarised by these two responses:

1. *“Separate record in data repository for each significant data collection, linked to thesis in publications IR”*

A few institutions responded along these lines. The main thesis would sit in the publications repository with its own DOI while supplementary files could sit in a separate repository (or the same one), each file having its own DOI. The records would be linked using the DOI.

1. *“We will probably use a single record with single DOI”*

And some institutions responded like this. Their initial thought would be to have a single record with a DOI, and hold all files relating to the thesis within that record. The single DOI would ensure users always end up at the main record for the thesis rather than at a record for a dataset whose relationship to the original thesis may be unclear.

In reality, the solution will probably lie in a combination of these approaches depending possibly on the ability of the supplementary file to stand alone as a citeable resource in its own right, and the workflows used to ensure the data file is linked back to the main thesis record. It remains to be seen whether it will be possible to develop detailed UTD guidelines to support decision making in assigning single or multiple DOIs to a thesis and its supporting materials. An alternative approach offered by one institution might also be valid:

*“ … users will be free to choose granularity of DOI minting.”*

7: Clearing the hurdles

The survey’s final question asked respondents to say what immediate task would need to be completed to start getting DOIs assigned to their theses. The aim of this free-text question was to reflect on the wider challenges beyond the immediate operational tasks such as repository mapping or DataCite subscription. The responses can be summarised as shown below:

1. Support from library management

Funding for project implementation as well as resource commitment to ensure additional workload can be accommodated. One institution indicated that an analysis of benefits over cost would be needed before a decision would be made to adopt DOIs for theses.

1. Engagement and support from Graduate School and academic committees.

This includes ensuring that guidance to students is updated to include information about DOIs, as well as approval of policy changes by the university. Communicating the benefits of developing good scholarly communication behaviours and the value of easily citeable theses and data to students will be key.

1. Workflow planning

Many respondents indicated that outcomes and guidance from UTD will be valuable in pushing ahead with their own planning. A significant number indicated that theses workflows are not as well established as those for other research outputs so work may be needed to streamline existing workflows before adding DOIs to the pipeline.

1. Repository changes at institution, including agreement to employ the DataCite plugin.
2. System enhancements

In particular, respondents referred to the need for the PURE Student Thesis template to be extended to include a DOI field.

Conclusions and indicators for further activities in Phase 1 and beyond

The survey has allowed the UTD project to develop a better understanding of the level of DOI and ORCiD usage at institutions for research students, map some typical thesis workflows, and gauge the level of interest in *Unlocking Thesis Data*. This baseline data will be valuable when we come to assess the success of the project in supporting institutions to roll out DOIs to their theses and underlying data or supplementary files.

The survey has also thrown up a number of indicators which UTD may not have considered at the launch of the project but will need to be taken into account during the Case Study interviews and beyond. These are informally noted here to provide a checklist for future reference:

* There was no evidence to indicate a difference in the view of DOIs for theses between institutions of different research intensity. On the other hand the level of external funding for postgraduate research training at an institution may well be one driver for encouraging student deposit of thesis-related datasets and other supplementary files since funder open access mandates for data may come into play.
* The planned case studies need to take account of as many nuances of approach as possible. The case studies are already planned to cover a range of research, repository in place, existing use of DOIs & ORCiDs and other factors, but smaller differences in workflows may turn out to be significant, for example whether the student or the library inputs the initial data for the thesis record and at what stage in the workflow.
* Institutions which allow students to upload the thesis need to have systems which ensure the thesis is the “correct” version and that all the necessary checks (copyright, confidentiality etc.) have been completed.
* Graduate schools or their equivalents are a key player in the thesis workflow, sometimes uploading the thesis itself and creating the initial record. UTD should ensure it takes the views and expertise of graduate schools as well as library colleagues into account during the case studies and beyond.
* Evidence has started to emerge that UTD will need to think about scenarios when the data should be part of the thesis record and when it should be separate. We may need to treat differently supplementary files considered to form part of the thesis even if they were not submitted for examination, and wider datasets collected as part of the research for the thesis.
* UTD should not assume that workflows and storage solutions are already in place for data and that the challenge lies only in thesis/DOI processes. Many institutions are only now working through Research Data Management plans. Early indications suggest that it will be relatively straightforward for some institutions to start minting DOIs for simple text-based theses stored as PDF. The devil as always is in the detail – and much of that detail comes down to data.
* Finally, anecdotal evidence is emerging that students and other researchers are rapidly becoming aware of the value of DOIs and beginning to ask for guidance from institutions on storing their datasets. Libraries and PhD processes will need to act fast to stay one set ahead of the researchers if we are to unlock thesis data quickly and effectively.

Appendices

A: Survey design

**Unlocking the UK’s Thesis Data: Survey of current workflows**

This survey on current thesis workflows in your institution will provide baseline information to help the UTD project understand how persistent identifiers might be assigned to UK theses. It will also ask if your own institution has started to think about identifiers, and map the level of interest in this national initiative.

The survey may take up to 20 minutes to complete.

Thank you very much for supporting the UTD Project.

Please read the following carefully before agreeing to take part in this survey. I understand that:

* All views and results from this survey will be treated in the strictest confidence in accordance with the Market Research Society's Code of Conduct
* Individual results or comments will not be shared with any third party unless this is required by law. The project may wish to use some quotations for illustrative purposes, but my specific permission would be sought before any responses are attributed to my institution
* I am free to withdraw from this survey at any time without penalty
* I am free to decline to answer particular questions

[Mandatory tick box]

**Q1. Please select your HE institution** (tick one only)

[Dropdown list]

If ‘Other’, please tell us your institution here [free text]

**Q2. What are your institution’s current requirements for student deposit of doctoral theses?** (Tick one)

Mandatory electronic deposit only

Mandatory print deposit only

Mandatory deposit of both print and e-

Mandatory print deposit with voluntary e-deposit

Student may choose any option

Other

Would you like to expand or comment on your response?

**Q3. Where are your institution’s new/recent *e-theses* held and made accessible?** (Multiple tick)

EPrints repository

DSpace repository

Hydra repository

CRIS system – Pure

CRIS system – Converis

CRIS system – Symplectic

British Library EThOS

Other repository or CRIS

E-theses stored locally but not open access

We do not have e-theses

Would you like to comment or expand on your response? [free text box]

**Q4. If supplementary data files relating to the thesis are also deposited, where are these held?** (tick multiple)

In the same repository or CRIS as the thesis

In a separate data repository and linked to the thesis

In a separate data repository, but not linked to the thesis in any way

Locally in a separate store and linked to the thesis

Locally in a separate store, but not linked to the thesis in any way

Situation has never arisen

Don’t know

Please comment or expand on your response [free text box]

**Q5. If your institution has more than one repository, please briefly describe your arrangements**

*e.g. A main repository plus a separate one for theses. A ‘publications’ repository and a ‘data’ repository. A single repository, but theses held locally elsewhere.* [Free text box]

**Q6. *Where* is the initial metadata or a bibliographic record for the thesis first created?** (Tick multiple)

In the repository

In the CRIS

In the library cataloguing system

In the student record system

On a standalone system

Other

Would you like to comment or expand on your response? [free text box]

**Q7. *Who* normally creates this initial record for the thesis?** (multiple tick)

The student

Repository staff

Cataloguing staff

Graduate school staff

Other

Would you like to comment or expand on your response? [free text box]500

**Q8. At what point in the entire PhD process is the metadata or bibliographic record for the thesis first created?** E.g. *Immediately after award*

**Q9. What file formats are accepted for PhD theses in your institution?** Please give as much detail as possible. [Free text]

**Q10. How are supplementary files such as DVDs handled, and how are they connected to the main thesis in your repository or other system?** [Free text]

**Q11. What guidance or requirement is there for students to describe the *type or* *format* of the thesis and the existence of any supplementary material?** *e.g. This thesis consists of the main work and two DVDs; This creative studies PhD consists of a published novel plus supporting critical analysis.* [free text]

**Q12. What guidance is given to students in writing their thesis *abstract*? If you have online guidance, the URL would be very useful**. [free text]

**Q13. Which of these describe your current level of activity relating to your research postgraduates and** [ORCID](http://orcid.org/) **author IDs?** Please tick all that apply (Multiple ticks)

We now require all our research postgrads to have an ORCID

We proactively encourage our research postgrads to register for an ORCID

Student ORCIDs are added to their publications record if available but we don’t mandate or encourage their uptake

Students must register for an ORCID at the start of the research degree

Students must hold an ORCID before they deposit their thesis

We are planning to ask research postgrads to get an ORCID

Our repository system has a plug-in which manages the process

Our CRIS system has a registration facility which supports the process

We have institutional membership of ORCID

ORCID implementation is managed as part of student record workflows rather than by the library

We have no plans to implement ORCID workflows for our research postgrads at the moment

We do not use ORCID institutionally at all.

If you can provide an example of a PhD thesis whose author has an ORCID, please provide the URL here [free text]

**Q14. What persistent identifiers do you use now for your theses?** (Tick multiple)

DataCite DOIs

CrossRef DOIs

Handles

ARKs

ISBNs

Other [please state]

None

Would you like to comment or expand on your response? [free text]

**Q15. What persistent identifiers do you use for your research data?**

DataCite DOIs

CrossRef DOIs

Handles

ARKs

Other [please state]

None

Would you like to comment or expand on your response? [free text]

**Q16. How ready are you to begin assigning DOIs for theses?**

We are doing it already

Completely ready, we could start next week

Quite ready, just waiting for guidance from the ‘Unlocking Thesis Data’ project

At the “thinking about it” stage

Not ready; it will happen at some point in the next five years

DOIs are not on our radar at all

*If you have plans for or are thinking about DOIs for theses, please answer Q17 – Q22 if you can. If you have not started to think about DOIs yet, please go to Q21.*

**Q17. Which departments in the institution have been involved in any planning/discussions about assigning DOIs to theses?** e.g. library, repository, academic departments, grad school, DTC … [free text]

**Q18. Who / Which department will lead on DOI assignment for theses?** [free text]

**Q19. What about more complex theses with multiple volumes, supplementary DVDs, or data which is deposited with the thesis?** How do you think you might assign DOIs in such cases? [free text]

**Q20. What’s the first thing that needs to happen, in your opinion, to start getting DOIs assigned to your theses?** [free text]

**Q21. If you would like to be included in a mailing list to be kept informed of our progress relating to DOIs and theses, please provide your name and email address here:**

**Q22. Finally we plan to do some case studies of institutions describing the detailed processes relevant to DOI assignment. If you would like to be a case study please indicate here**.

No thanks

Yes, we’d like to be a case study institution and I have provided our contact details above.

*We won’t be able to speak to everyone who volunteers, but will aim for a small range of different repository systems, levels of DOI development and size of institution.*

**Included in the thank you pop-up after submission –**

Thank you very much for supporting the Unlocking Thesis Data project.

We would now like to gather examples of existing material to help develop templates and a toolkit as part of the project. If you have any flow diagrams or guidance documents describing thesis deposit; thesis management; thesis record creation; or plans for assigning DOIs to theses, please would you share them? Please send any such documents you think would be useful to [Sara.Gould@bl.uk](mailto:Sara.Gould@bl.uk), or simply send the URLs if they are available online.

Thank you.

*Unlocking* Thesis *Data project team*

B: List of responding institutions

Aberystwyth University

Anglia Ruskin University

Bangor University

Birkbeck, University of London

Cardiff University

City University London

De Montfort University

Durham University

Edge Hill University

Goldsmiths, University of London

Keele University

King's College London

Kingston University

London School of Economics and Political Science

London School of Hygiene and Tropical Medicine

Northumbria University

Queen Margaret University

Royal Holloway, University of London

Sheffield Hallam University

Swansea University

University of Aberdeen

University of Birmingham

University of Bristol

University of Cambridge

University of Chester

University of Derby

University of Dundee

University of East London

University of Edinburgh

University of Glasgow

University of Hertfordshire

University of Huddersfield

University of Hull

University of Leicester

University of Lincoln

University of Liverpool

University of Newcastle Upon Tyne

University of Northampton

University of Oxford

University of Plymouth

University of Reading

University of Southampton

University of St Andrews

University of Stirling

University of Surrey

University of Sussex

University of the Arts London

University of the West of England, Bristol

University of Westminster

Data availability

**The data from the survey is available for reuse and analysis at** <http://dx.doi.org/10.15123/DATA.12> under a Creative Commons CC0 licence**.**

About the project

Unlocking Thesis Data (UTD) is a community-driven project to promote the use of persistent identifiers for theses, their underlying data and their authors. By their very nature, PhD theses break new ground and advance scholarly knowledge. Most make use of newly-created data but these data can be trapped in an appendix or DVD – either unavailable or not suited for reuse. UTD will make data more discoverable and citeable, thereby offering incentives to students to share their data in more appropriate formats, in the context of a sustainable national thesis framework.

Funded by Jisc, UTD is led by the Universities of East London and Southampton and EThOS (the UK’s national thesis service at the British Library). Phase one will explore current thesis practice through an online survey to EThOS member institutions, and individual case studies looking at the issues in more detail – including how institutions might apply DOI and ORCID identifiers. The survey and case study findings will be combined into a report with recommendations for further phases of the project. These are expected to enhance metadata and software for applying DOI and ORCID identifiers, to test them in live settings, and to offer comprehensive advice for institutions to adopt them. By summer 2016 we expect to have a sustainable infrastructure covering the whole UK, driving the wider availability of research data and introducing doctoral students to new norms of scholarly communication.

1. In 2014 the [ODIN](http://odin-project.eu/) project developed a prototype EThOS-ORCiD importer tool; still in beta, it is available at <http://ethos-orcid.appspot.com/> [↑](#footnote-ref-1)
2. Guidelines and a metadata application profile designed to help institutional repositories comply with the RCUK policy on open access. <http://rioxx.net/> [↑](#footnote-ref-2)