Unlocking the PhD: From Behind Closed Doors to Immediate International Availability – a UK University Perspective

Fiona GREIG

University of Surrey, United Kingdom
f.greig@surrey.ac.uk

Introduction

This paper aims to explore the ways that we have supported the PhD Thesis as a Library service, the opportunities and benefits that the digital environment has brought, how we can use historic digitisation to improve access and looks at the challenges we have faced to date. I also ask if we are in the best possible place to support how this unique type of artefact is evolving.

Setting the scene

Definitions

This paper has used a shorthand of PhD to cover all the various qualifications that are offered at doctorial level. I am very aware that not all disciplines or countries use the same qualifications. During this paper the use of PhD should be used as a catch-all phrase for all of these qualifications. The loose definition used to outline what I mean by PhD is from the ERC Policy on PhD and equivalent doctoral degrees[1] and states:-

"The research doctorate is the highest earned academic degree. It is always awarded for independent research at a professional level in either academic disciplines or professional fields. Regardless of the entry point, doctoral studies involve several stages of academic work."

There is also an international element around the use of the word Thesis (or Theses). In the UK the "doctorate degree" is only complete with a written component that we call a thesis. Dissertations are only produced by Undergraduate or Masters level students and are not made available outside the institution. I am aware that in the US and many other countries have these reversed. This is a UK perspective so I will be using our native nomenclature.

A little about the University of Surrey

 We are currently celebrating our 50th birthday as the University of Surrey based in Guildford. Surrey is south of London very close to both Heathrow and Gatwick airports. You can get into the centre of London's "theatre land" in under 45 mins by train from our station. Yet we are also surrounded by the Surrey hills and the main road into the city is a designated European Area of Outstanding Natural Beauty".

---

[1]

https://erc.europa.eu/sites/default/files/document/file/ERC_policy_on_PhD_and_equivalent_doctoral_degrees_2016.pdf

While we have only been on our current site since 1967 our institution traces its history back to 1891 and the Battersea Polytechnic Institute which specialised in "technical" education for the poorer students in a part of west London. Since those early days we have continued to have long-standing strengths and recognition in the sciences, engineering, technology and health sciences. However we have been pioneers in academic education and where the first University to aware a degree in Hospitality and Tourism anywhere, and the first academic qualification for Dance in the UK.

As a University we value both our research reputation and student experience and are in the top 10 UK Universities. We also excelled at the last Government review of the quality of the research outputs from the University with over 94% of our research being regarded as world leading or world renown. Today we have over 15,500 students covering Physical Sciences, Health and Medical Sciences and Arts and Social Sciences all on one unique landscaped campus.

So what is it about Theses?

Many people may wonder why there is a need for a paper at a conference of this standing to deal with theses – after all information professionals have been looking after these artefacts for a long time and it seems to be going well. And I have to agree, form a security of storage stance a library is exactly the place you would trust with this unique material – but is that really what we are primarily here to do? The doctoral thesis is a window into the most cutting edge thinking in any discipline. Certainly the expectation of PhD courses in the UK is for independent research into a topic. A failed hypothesis or experiment is just as likely to be important and have a degree awarded as one where the student "proved" their contention. But that hypothesis must be new and build on or challenge existing thinking or theories. With the increase in those seeking a PhD actually finding new topic areas is problematical. By having easy access to what has gone before we help improve the journey of current and future candidates, and help advance science and thinking in general.

Let's just look at what the PhD student's life has looked like:-
- They have to find a subject for research and put forward a compelling proposal for approval to study at the institution
- They spend 4-7 years undertaking the research, hopefully with great support from a Supervisor and researcher development team helping them make the most of the development opportunities going on.
- They might be undertaking a funded course and so have to work within a larger research project where they have to follow the demands and interests of the Principal Investigator and work around any of their own research needs as they can.
- They then have to find a way to write-up their thesis – there is no standard guidance but in general theses are expected to be 500,000 words or so. That is a lot of writing to do and it has to focus on their learning approaches and methodologies as well as the results of their research. Over the years it is fascinating to see the methods the students have used to write their material, from wonderfully artistic handwriting, through lithographs, typewriters (with correction fluid) and onto the computer age of dot-matrix and laser printers. There may be a thesis in charting the art of written communication in just a cross-section of old theses
- They then have to present copies of this work to the examination panel.
- The dreaded Viva comes after that where they have to defend the thesis (again not just the research project, but how it has been documented and commented upon)
- No student gets off unscathed and so there are always corrections to be made – and in theory at this stage they are supposed to remove any 3rd party Copyright material.
- Finally they are able to have at least two copies hard-bound (at an average cost of £35 per volume) and pass these to the Department and University Library.

- Only once the Library has its copy are they granted their Doctorate.

Once the eager student brought their hard-earned thesis we contacted the Faculty to ensure we should have it and to inform them that the deposit had happened – and finally let the student graduate!

Our catalogue team then looked at the thesis and did their best to catalogue the items. As we said the thesis is becoming more and more niche and it is becoming harder for non-specialists to understand; so for the last few years we have not really "catalogued" it is more like using a template and copying what was written on the title pages and, if you are lucky, the abstract.

Of course we also receive requests for the thesis to be restricted. This is something that on the whole we avoid, there have to be very strong reasons why the item cannot be made public, these revolve around security and commercial considerations. If restricted the catalogue record is "hidden" from the public view.

Then it is put away securely – in Surrey the hard-copy theses live in our Archives area where there are three locked doors and an alarm that needs to be turned off anytime we want to retrieve a copy.

Finally the bibliographic information was sent to be added it to the *Index of Theses* – more on that in a moment.

Is that fit for the 21st century?

What is interesting is that the main activities of research for the student has not really changed. As I said we might be able to make finding a topic easier, but writing a proposal and actually doing the research is still very much the same. But from the "writing-up" element is beginning to change, and change much more profoundly than libraries or even Universities are prepared for.

Students are now using multimedia to document, record and present their research – and some of this material is not suited at all to a long narrative document, so the student's thesis outputs are changing. There is also a feeling that asking students to pay to bind copies of a document for the institution to keep is unethical, let alone unfair, and that this final economic hurdle is being removed.

But the main contention and change is around discoverability and access of these wonderful works. We live in a connected and interconnected world and the traditional expectation that to see a thesis I would have to know it existed and then approach the institution for a copy is not how academics work anymore – and that is a good thing!

In the UK we have the British Library (BL) which is our national deposit library. Of course as theses are not "published" there is no legal requirement to send a copy to the British Library; but they did collate an Index of Theses which outlined the bibliographic information of any thesis that institutions made the British Library aware of. The BL also co-ordinates our national inter-library loan service and so they were in a position to offer a lending service for theses. If the British Library received a request for a University of Surrey thesis we would be approached and we would send them the Library copy (under heavy guard of course – or at least via registered courier). The BL would then microform it – mainly as a microfilm but they did also offer microdots and microfiche over the years. These would be sent to the requesting library and the researcher keen to read it who then had to master the "state of the art" technology of a film reader/printer.

Moving to the digital

In 2009 the BL decided that things had to change and they launched EThOS[2] a service that provides access to scan or born digital theses free at the point of use to a vast range of UK PhD level theses. Some universities do require the first requester to pay for the digitisation of a thesis if it is not already available, but once scanned it is there for all.

The BL "seeded" EThOS by asking UK universities to provide the most requested theses again. These were then scanned using high quality digital scanners (using OCR where needed) to capture full digital copies. We were asked to provide 268 titles from Surrey. Our research tends to be very applied and so much more likely to be recalled by researchers than some of the more esoteric theses that other institutions are known for.

Once the title was scanned the physical copies were returned to us as was a digital file that we were able to upload into our own institutional repository. What we noted immediately, and Sara Gould (the EThOS Manager at the BL) has written on, is how the move to digital copies immediately available on demand has driven up use of these works exponentially. [3] After go-live of EThOS any request for a thesis via the BL was provided as a digital copy. Here at Surrey we paid for the digitisation of the titles irrespective of where the requester was from, we also paid for theses from other libraries to be digitised if their business model was to have the requester pay. Some might question why we paid to allow access to our old theses but for us it was logical; we had unique information that was required by a researcher, it is our duty to make that available. It also did not hurt that people were remembering – or discovering – just how rich our research is and how inspiring our academics and past students are.

We started uploading in 2010 and by the end of that year we had loaded 631 theses. From 2011 we automatically uploaded a digital copy of any new thesis that came in where the student had included it with their physical copy (we requested this but it was not mandatory) and in February 2015 we moved to electronic deposit only.

Now, as the last step of their student journey, we ask students to upload their final, corrected and copyright cleared thesis to our Surrey Research Insights (SRI) repository https://epubs.surrey.ac.uk/. And we are anticipating by the end of 2017 we will have had 1,000,000 downloads of our theses.

Why stop there?

So, Surrey had already been supporting retrospective digitisation (via on demand from the British Library), we had gone to only accepting new thesis electronically, so what do we do with our other theses – those that had not yet been used by someone outside our institution? And that is where ProQuest comes into the picture. We were approached by Cathy Boylan about possibly being involved in the ProQuest historical digitisation project. For decades ProQuest has been the legal dissemination route for US Dissertations (still PhD level remember) and were interested in ensuring that their Dissertations and Theses[4] database was truly global and to do that they were seeking more content. They wanted to harvest from our repository and also offered to scan up to 2,000 of our theses which we had not got around to yet. The scanning would be undertaken by a firm in Ireland and then full quality assurance and cataloguing would be handled by the theses team in Ann Arbor in Michigan, USA.

---

[2] http://ethos.bl.uk/

[3] Gould, S. 2016, "UK theses and the British Library EThOS service: from supply on demand to repository linking", *Interlending & Document Supply,* vol. 44, no. 1, pp. 7-13.

[4] http://www.proquest.com/libraries/academic/dissertations-theses/pqdtglobal.html

Of course we said yes.

But as with everything in libraries it was not that simple. There were a number of things we had to consider:-

What should be sent? We have a lot more than 2,000 theses in store, how did we select the ones we wanted to go. We started by looking at our very oldest titles, those where the quality of paper – and storage over the years – had left them fragile. This would give those titles a new lease of life while giving us some assurance over longevity. Then we looked at the theses that had been popular with Surrey staff and students but never requested by anyone outside our institution. And to fill the final few places, theses from some of our more prominent Alumni or in fields where Surrey made a breakthrough (of which there were many).

Author consent, do we need permission? This is, I know, something that many institutions have decided to do. However at Surrey we are more pragmatic. Our "contract" with the students makes it very clear that the Library copy is to be made available to the general public. My view, and that endorsed by the University Research Committee is that there is no need to seek further permissions. We had never sought permission to send a thesis to the British Library in the past, so why is "online" digitisation different to a microform digitisation? This is just a format shift and any attempts to trace academics to seek permission would be wasted staff resource that should be used supporting our services. We do of course have a "take-down" policy if an author were to request there thesis be removed. However this request would go through the same process as current theses undergo if there is a request to restrict access. There is no guarantee that we would remove the title from the repository.

What to do with those titles where there may be 3rd party copyright? This is trickier. The regulations have always been very clear that before a student sends the Library a copy of their work any material that they do not own, or have rights to use, should be redacted or removed. However this is not something that has been consistently done. Let's be honest after 7 years, thousands of words, a gruelling examination by panel and making the required corrections the very last thing you want to do is see your thesis again. What we have found is that around 1/3 of the theses that we have in the Library actually contain material that we are unsure of the source. As we could easily find 2,000 "clean" theses to send this was not a problem. It has heled us reinforce training to current students but we were able to skip-over the historic ones that might cause problems. We also had to decide what type of licence we would apply to digital theses – again we decided that historical digitisations are no different to our new submissions and so we use the same Creative Commons licence.

Should we allow a commercial company to profit? One discussion point we had internally was that ProQuest would be able to sell our theses to people who discovered them on their database. Again the pragmatism of Surrey came to the fore. ProQuest does not receive any thesis that has not already been made available for free on our repository. What subscribers to the ProQuest database (or one-off sales) are paying for is the ease of discoverability over a range of materials and for any aftersales support. There is nothing to stop anyone from going to OAIster, Google Scholar or any other search engine. If people are daft enough to pay for something they could find for free that is fine. And in fact how many of us have paid for something because it actually save that rarest of commodities – time. This is about choice and not forcing an economic burden on anyone.

Destructive or non-destructive scanning? The tricky one, can we go e-only for these historic documents or do we need to retain the physical artefact? This one lead to a range of conversations,

but in the end we decided to compromise. We would allow destructive scanning (slicing off the spine and "de-boarding" the thesis) but asked for these to be returned to us so that they could be placed back in a store "just in case".

How will we measure the benefit? At the end of the day, how we will know that we have achieved something? Are there any measures that we need to be putting in place? Is there a cost-benefit or efficiency saving that can be linked to historic digitisation? Again this one goes to the heart of what we as libraries, and in fact the wider university is about. How do you measure sharing knowledge? Yes I would love to think that we will see citations of our theses increase, but I am not sure how realistic this is. We know there are a huge number of students referencing theses as part of their studies, and even academics use thesis in their thinking but it is extremely rare for these to actually be cited in any papers or even other theses. There is a culture where a journal article or monograph that has been written from the research is more valued than the underpinning thesis and until that changes is it right to set a measure that I know will fail? But I can point to the staff and systems efficiencies of already having theses in digital format available, and we can look at downloads and possibly increased interest in potential PhD candidates applying to come and research at Surrey. But, to the chagrin of our Chief Finance Officer this is one of the "for the public good" activities that libraries are fantastic at championing.

The benefits

And while I may not be able to point to a cold-hard cash "saving" there are a number of benefits the historical digitisation approach.

I have already spoken about the increased discoverability of the works, this additional exposure to the work is critical. Any kind of access to this resource is something to be valued, after all they are costing us to be kept in a locked room; let's try and maximise the impact of our researcher in any possible way.

At Surrey we have a real commitment to Open Research, it is not just something that we do because funders or governments require us to it is a belief that runs deep and we have always invested in making Surrey research open and free to all – this project has just helped us revisit a set of content that we ourselves have at times overlooked.

As I said in academic year 16/17 we are celebrating our 50th Anniversary and being able to make more open the fantastic work of our Alumni is a great way to doff our caps to those who have gone before. As an institution we are where we are due to their passion, wisdom, creativity and downright perseverance; the least we can do is acknowledge their work and celebrate the wonderful things that have happened at Surrey.

It also helps underpin our reputation as a research intensive institution.

And now the prosaic benefit – by partnering with ProQuest we are able to achieve all of those benefits on a grander scale, but what is unique in the ProQuest "offer" is the resilience and disaster proofing that they offer. Remember I spoke about old and fragile documents, well without a systematic digitisation process these items are at risk. Of course I could have done my own digital copy but that pales into what we get from ProQuest – you see as well as having a second electronic version on their platform, they produce a microfiche archive copy, and if things look really bleak that microfiche will end-up in the original Iron Mountain store; many hundreds of metres under a mountain "somewhere" in the USA. My understanding is that ProQuest has the vault between Microsoft and the Walt Disney Company – if it is good enough for them than the University of Surrey unique content is safe – much safer than a physical copy in one place in Guildford could ever be.

Looking forward

Unlike most of my papers this one has been looking back, but I can't do a presentation that does not ask, possibly uncomfortable, questions about where we are going. I asked if we were doing enough for theses in the 21st century, and I admit we have improved, but are we ready for the next evolution of the "PhD Thesis"?

We are already seeing changes to what is accepted as a thesis, I know it has been common in Scandinavia for years, but the idea of a thesis by publication is new in the UK. This is where a student just has to publish a set number of articles in peer-reviewed journals and use that as their thesis. I can see how that is very attractive to the student, but where is the "original thinking" that explores all the details of the research that would never be published. The unique elements that are in the traditional document but that is not required in this form of award. And how can a Library make it public if it only comprises of Copyrighted work (the publishers will of course still insist on the Copyright Transfer Agreement)? Would the open access versions – without the "proof" of publication be good enough? Or are we back to just listing the bibliographic details and passing on links to pay walled content?

New forms of output. More and more PhD projects are not things that can be easily captured in a 500,000 word report. The "traditional" approach will continue to work for the social scientists for the longest period, but even the hard science types might transition to a "thesis" that results in a dataset or a piece of code. That type of result is not something we will easily manage in a pre-print type repository. And it gets worse as you move across the scale from hard science, via medicine and biochemistry (an expressed gene is a possible research output) and when you get to the arts, especially performing arts, fine art, design, textiles etc things become very strange. How do libraries allow long-term open access to what was actually a multisensory installation? Is it somehow less worthy of discoverability, access and long-term curation because it is not in a printed form?

And my favourite challenge is the, how can I put this kindly, hypocrisy that we are seeing in students (and Supervisors). Since we went to e-deposit the number of requests to make theses restricted has significantly increased, and it is normally prompted by the academics. The students are very keen that their work is not easy to copy – yet they are happy to use 3rd party copyright in the work and see no comparison. We need to work on education of the students and academics. Related to this request for restrictions however is the fear that having a thesis available on the Internet will stop them from being published. We have undertaken research with publishers and this is just not the case, no reputable publisher will just accept a thesis as a manuscript. The thesis might form the base of a monograph but it will need editing with all the "academic required" material removed.

A legal conundrum … I plan on continuing the historic digitisation and will be sending another 2,000 theses off to be scanned (ProQuest offer the use of their scanning facility at a very reasonable rate). What do we do about that 1/3 of the titles that we think might breach Copyright? We did work with our Copyright Librarian and determined the Academic purpose and Fair Use clauses that we have in UK law and, taking that alongside the likelihood of commercial harm to publishers, thought that there was arguments to be made in sending some items that we have previously rejected. Other options considered:-

- Should we redact the sensitive content in some way?
- Should we seek permissions (and where will the finance for that come from) for what was included?

- Just send anything published before a certain date – very little likelihood of a court upholding an argument about loss of commercial activities if the Copyrighted material is not significant.

I am currently looking at how much risk do we want to expose ourselves to? This is going to have to be discussed for the third-batch that we send.

Where is the Library in this?

I am proud of the heritage that information professionals have, our past is built on the tradition of pushing the technological boundaries as some of the very earliest users of computers and the Internet, long before the WWW, we have seen the opportunities information technology offers to extend the reach of our core values – knowledge should be available to all "free at point of use". But can we honestly say we are living up to those standards any longer? This is an exceptionally exciting time and we should be helping our academic colleagues develop these new scholarship skills. But are we?

If I were to look at the "information literacy" or "digital literacy" schemes in most university libraries would I find data carpentry sessions? Are we providing students (and academic staff) the skills they need to understand the discoverability, usage and long-term aspects of their multimedia projects? Do we offer sessions on the ethics and scope of virtual worlds? Are we pushing legislators on issues around copyright, licences, protection of academic works? Or are we still taking an hour to demonstrate how to search the catalogue, do advanced searching in specific databases and pointing out how to reference books and journal articles?

Are our library schools developing challenging curricular that will attract the kind of people we need in the profession, or are they following the money and aiming courses still at people who "love books" and see themselves spending hours cataloguing rare items? Just where do we want the profession to be and what are the core competencies of the services we are delivering now and evolving for the future?

It is not just libraries, how will information aggregators deal with this brave new world? If we, as universities, are not able to have strategies in place to deal with the emerging works than how can our traditional partners, the systems developers and content aggregators help us? We rely on them to bring together disparate information from around the world, and that has worked well for printed materials, but gets more and more complex (and expensive) the more that the work morph into something else, but as we are seeing already they must change if they want to remain in this business.

In my opinion we still have a critical place in the centre of the scholarly discovery and communications practices of our institutions but we need to start being confident in supporting, advising and yes providing expertise in the emerging techniques, tools and approaches. Our traditional searching skills change to helping understanding contextual searching, take linked data to the next level. People will no longer want, or need, to read the whole article or book but be looking for a segment of it and want to search across different types of artefacts. How can we ensure that the underpinning information architecture supports that and that we can explain it to our clients? And a critical one, can "librarians" become futurists? We have this pool of content and context across disciplines and international boundaries; why are we not using that content, and the data that we can generate as a sector, to help spot trends, opportunities and collaborations.

We have successfully unlocked our historical theses but are we already building more walls around the next-generation of doctorial outputs?

Fiona Greig

Head of e-Strategy & Resources, University of Surrey May 2017

References

Gould, S. 2016, "UK theses and the British Library EThOS service: from supply on demand to repository linking", *Interlending & Document Supply,* vol. 44, no. 1, pp. 7-13.

Julie Gould 2016, "FUTURE OF THE THESIS", *Nature,* vol. 535, no. 7610, pp. 26.

Pittayachawan, S., Macauley, P. & Evans, T. 2016, "Revealing future research capacity from an analysis of a national database of discipline-coded Australian PhD thesis records", *Journal of Higher Education Policy and Management,* vol. 38, no. 5, pp. 562-575.

ProQuest 2015, *Enhancing the Visibility and Impact of Electronic Theses and Dissertations* .

Rodriguez-Marquez, M.d.M., Daoutis, C.A. & Gravely, C. July 2014, "The training needs of PhD students: A case study from the University of Surrey", *ETD 2014*University of Leciester, , 24-25 July 2014.