

The Impact of the Electronic Digitisation of Theses

A case study from the University of Surrey

The University of Surrey has been 'digitising on demand' for a number of years, beginning with the pilot of the British Library's EThOS project, to which it was a contributor. A significant number of Surrey theses was requested to seed this service. For the first three years of the EThOS project, Surrey agreed to pay for the digitisation of any of its requested theses when permission to digitise was given. £38,000 was quickly spent and the cost of servicing requests began to outstrip the annual budget allocated to digitisation; in the final year the allocation was used-up within

the first four months of the academic year. Consequently, Surrey had to revert to the 'requester pays' model, and requests for digitisation of its theses decreased dramatically. All digitised theses were added to Surrey's local research repository (<http://epubs.surrey.ac.uk/view/type/thesis.html>).

During the academic year 2014 – 2015, the University of Surrey decided to move to an e-only library deposit model for all new theses, to become effective from 1st May 2015. Students would no longer be required to pay for a hard-bound copy destined to take its place in a closed local access-only store. Today, all theses are digitised and self-deposited in the repository. The British Library harvests the metadata and the file containing each thesis to keep EThOS up to date.

In February 2015 there were 1,342 digitised Surrey theses, all available via Open Access. There was a much larger number of theses held in hard copy format only, and no funds available to digitise them. ProQuest approached the University and generously offered to pay to digitise a

"The lessons we learnt will make life much easier in the future and help feed into any business case for moving to a complete digital environment."

further 2,000 theses. Fiona Greig, Head of Strategy and E-Resources at the University of Surrey Library, says that "this was a wonderful and an exciting prospect, but as we looked into selecting the titles we realised that we needed to create some kind of selection process."

It was decided that the theses for ProQuest to digitise would be chosen on the basis of past usage by Surrey's own staff and students (those theses that had been requested by non-Surrey researchers had already been digitised by the British Library). This reduced the number of theses Surrey wanted ProQuest to digitise to a little over 2,000. However, about one third of these had to be excluded because they contained third party copyright material for which the Library was uncertain it would be able to clear the rights. Surrey's guidelines for postgraduate students instruct them to remove any copyrighted material in the version of the thesis deposited in the Library, but apparently no checks were made to ensure they had fulfilled this requirement. Fiona Greig says "it appeared this was not something that the academic community was too concerned about. They reasoned that

the item would not be used frequently and would always be consulted in the library context. But the Library knew that the risk of allowing the University knowingly to breach copyright law was too great."

The Library therefore had to reject any theses with uncertain rights permissions. It then started to think more creatively about which of Surrey's other theses might be of interest to both its own and the wider academic communities. In 2011 a box of very old theses had been found, written before Surrey had acquired university status: they demonstrated that long before the University received its charter, some of its brightest students were breaking new technological ground. The theses were added to the group earmarked for digitisation. Lastly, theses written by Alumni who are still active in research, teaching, learning or advancing their subject areas were sought out and added. Thus the final figure of 2,000 was reached.

Fiona Greig says: "While the whole exercise was rather painful for our staff—they were checking these documents page by page—we recognised the value of knowing what our future digital activities were likely to be. The lessons we learnt will make life much easier in the future and help feed into any business case for moving to a complete digital environment."

Challenges

As well as the main challenge of selecting the theses they wanted to digitise, librarians at Surrey had to deal with a number of other challenges and obstacles before the project could get under way.

Are your school's dissertations and master's theses digitally archived and preserved alongside the works of other premier institutions worldwide?
Promote your school's global profile and reputation for scholarship with ProQuest's Digital Archiving and Access Program. **Find out more at www.proquest.com/go/DAAP.**

Securing author permissions was a subject of some debate. Fiona Greig says, “For some reason, libraries who have undertaken digitisation projects have included as part of the process going back to the author to seek permission. We had an internal debate about this; our final conclusion was simple. The ‘contract’ of study signed

“...we are seeing genuine evidence of the increased use and impact on scholarly thinking of those theses that we have been able to make available.”

by the student states that a copy of the thesis will be deposited in the library for public consultation. At the time of deposit, the student has the option to apply an embargo on access. This means that the thesis is either available immediately or at the end of the embargo. We have never sought the author’s permission for people to access a hard-copy thesis; nor did we seek permissions when the British Library requested a thesis to be turned into a microfilm version because they had received an Inter-Library Loan request—so why would we seek permissions for digitisation? I understand that people think the format (or medium) should enforce policy, but our view is that policy concerns content and not the delivery mechanism. What we’re talking about is simply a format shift. Rather than build a cottage industry on to-ing and fro-ing to get permissions, we need to ensure our policies, agreements with students and academics, and indeed our profession, are future-proofed.”

Another issue with potential for triggering high emotions was whether to destroy the hard copy after scanning. ‘Destructive’ scanning, whereby an item is cut from its binding and pages are batch-fed through a scanner, is much cheaper than non-destructive scanning, which involves the use of a ‘cradle’ or similar apparatus that slows down the process and is very labour-intensive. For this project, Surrey had three options: complete destruction, meaning that the items would be destructively scanned and then the physical residue destroyed in the appropriate manner; destructive scanning that involved returning the sheaves of loose paper to the Library (each volume thereafter to be kept in its own elastic band); or non-destructive scanning. Some of the Library staff took the pragmatic view that a digital copy could be guaranteed to exist ‘in perpetuity’ and advocated full destructive scanning. Others were worried that, as author permissions were not being sought, this could leave the University in the

vulnerable position of being unable to fulfil a request from a past student (or member of their family) to access a hard copy of their thesis. It was felt that a print-on-demand solution would not provide sufficient insurance against this possibility. Therefore, Surrey elected to choose the ‘middle ground’ option of retaining the separated sheets of paper alongside a number of hard disks containing the scans in two formats and the metadata. The librarians are aware that this was a risk averse decision that they may wish to revisit and question in the future.

Usage of the Surrey collection of digitised theses

Some ‘snapshots’ of usage statistics of the Surrey e-theses collection are included at the end of this case study as **Appendix A**. They provide a powerful vindication of the making of theses available digitally.

More than 2,370,000 requests for Surrey theses have been received via Google at the time of writing (February 2016). The next highest source of requests has been internally, from students and academics working at Surrey itself. Almost 376,000 requests have been received in this way. (See **Appendix B**.)

The Library at Surrey has not yet carried out citation analysis from these downloads as this would be difficult to achieve using the Library’s own resources. However, it is working with ProQuest to gain some insights about citations from theses that have been on the ProQuest platform for some time.

Comparison of usage of digital theses with print ones

From both the Library’s and the University’s point of view, one of the most important reasons for holding a digitised thesis collection is to increase usage. Surrey has collected some indicative information about this. Its most downloaded e-thesis is “A Study of the Factors of Consumer Behaviour Related to Overseas Holidays from the UK”, which has been downloaded from the repository 21,079 times in five years. Prior to its digitisation, the paper version was used fewer than eighty times since it was first made available in 1992. Between the beginning of January and early February 2016, the e-thesis was downloaded 751 times. During the same period, a total of forty-eight separate theses were downloaded a total of 23,843 times, while there were no requests for access to items held in the hard-copy collection of theses.

Promoting the digital thesis collection

Perhaps owing at least in part to their earlier limited accessibility in hard copy format, digital theses are often not the first resource that comes to mind when academics are recommending resources, either to undergraduates or postgraduates. The Library at Surrey recognises that promoting awareness of online theses is essential to securing greater usage of the collection. Its main method of doing this is via its discovery tool, which both highlights and promotes the digital theses collection.

The Library also offers training sessions for new PhD students, Masters students and postgraduate tutors as the students begin their research projects. The main reasons for drawing attention to the collection are that reading previous theses is often useful for finding a research subject to build on; and that existing theses serve as examples of the type of work that is expected of the new researcher.

In addition, the Library has employed its general communications and publicity tools to increase awareness of the collection. These include blogs, Twitter and other social networking devices. It has tended to focus on promoting the older digitised titles (rather than those just deposited) as it is keen to find information that supports Surrey's links to important research both within and beyond the University. For example, the recent news about Major Tim Peake heading to the International Space Station gave the Library an opportunity to highlight some of the work carried out at the Surrey Space centre.

In conclusion

Overall, Fiona Greig says that the main problems encountered have been connected with authors' disregard of third party copyright. She adds: "This is a historic and embedded issue in the culture we are working with. While we can 'train' our students, their supervisors are requiring they include the copyrighted material in the work academics comment on or are marking. The change to the UK law that says 'for examination' is no longer a protected use has not yet penetrated the wider academic community; this is a real issue for us moving forward.

"The time we took to identify copyright breaches in our historical theses was significant; to have to spend staff time retrospectively gaining permissions or redacting information is not practical, nor does it represent value for money. This is a real pity, as we are seeing genuine

evidence of the increased use and impact on scholarly thinking of those theses that we have been able to make available."

Conversely, the Library has observed significant concerns among students that they will be able to protect their own Intellectual Property (IP) and copyright: since the University has mandated online access, the requests for embargos have increased. The Library had anticipated this and has worked with the University Research Degrees Committee to come up with University-agreed criteria on when an embargo would be accepted and for how long. Nevertheless, supervisors are sharing with students their fears that material published in theses in e-format is unlikely to be accepted subsequently as suitable for traditionally-published monographs. Consequently, Surrey is approaching all the major publishers in the UK to try and bust this myth.

Fiona says that she thinks that e-theses will constitute a major contribution to the Open Access movement, because if a postgraduate's first solo publication is via Open Access, demands from research councils and others that their future research outputs are delivered via OA will not challenge their perceptions in the way that they do at present. However, she also says that she thinks the data issue is one that librarians will be fighting for a while, and with the whole of the research community. While research councils and universities can insist that publicly-funded PhD research projects make their findings and data accessible to everyone, many students are funded either by employers or sponsoring companies; often their contracts say the data belongs to the 'home' company. Similarly, in the case of studentships funded privately or by other nations, the IP of any outputs resides with the student.

ProQuest offers a full suite of services that can help to increase the reach and impact of your institution's graduate works through all available channels, including your institutional repository. **Learn more at www.proquest.com/go/dissertations**

Appendix A

Some snapshot illustrations of how the success of the Surrey e-theses collection has built up.

Comparison per year



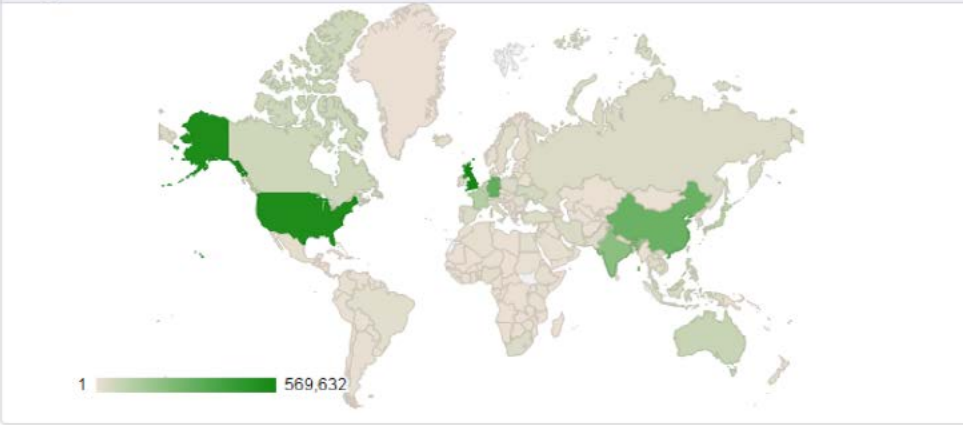
Appendix B

Sources of requests for Surrey e-theses

All items

Filter Items Dates Available Reports

Origin of downloads



Top Referrers

1. Google	2,370,416
2. Internal	395,720
3. Internal (Abstract page)	137,188
4. www.surrey.ac.uk	50,904
5. Internal (Browse view)	45,711
6. Internal (Search)	22,523
7. MSN/Bing	21,018
8. Yahoo	16,729
9. yandex.ru	7,749
10. dcms.dlibrary.net	5,946

10 25 50 all

Browsers

1. Apple Safari	1,986,167
2. Firefox	1,432,551
3. Microsoft Internet Explorer	1,070,742
4. Opera	79,834
5. Android	9,453
6. BlackBerry	501
7. Netscape	369
8. AOL	99
9. Google Chrome	78
10. Opera Mobile	4

10 25 50 all

© The data are the copyright of the University of Surrey. Copyright of the case study is held jointly by the University of Surrey, ProQuest and Linda Bennett.