



Impact of Digitization of Theses at London School of Economics

The London School of Economics [LSE] thesis digitization project was one of two pilot projects funded by ProQuest to ensure that a large body of theses was available from a UK institution. (The other project took place at the University of Surrey.) ProQuest's hypothesis was that critical mass was important for attracting visitors to an online thesis collection. The project commenced in 2014. By May 2015, 2,000 digitized theses had been uploaded to LSE Theses Online [LSETO].

The LSE had a free hand in which theses to digitize, and decided to choose theses from 2010 -11 'backwards' to the early 1990s. Authors were contacted and told they could opt out if they wished; only 14 chose to do so. After digitization, five take-down requests were received and complied with immediately.

Study arising from the project

It was quickly realised that valuable insights to support the hypothesis could be gained by conducting a complementary study when the digitization and uploading process was complete. The aim was to gain a greater understanding of how digital theses, clearly an important academic resource, are used and how they fit into the scholarly resources landscape. The research focused particularly on the LSE collection of theses, because they had been digitized most recently and were from a smaller, more interrelated group of academic disciplines than the Surrey collection.

“The aim was to gain a greater understanding of how digital theses, clearly an important academic resource, are used and how they fit into the scholarly resources landscape.”

Methodology

Usage statistics – measuring full-text PDF downloads both from the LSE’s own institutional repository and the ProQuest PQDT [ProQuest Dissertations and Theses] database – provided the main quantitative basis for the study. Where they were available, citation statistics were also used. Qualitative information to complement the statistics was obtained by carrying out three focus groups at the LSE, with undergraduates, postgraduates and librarians respectively, and by means of four semi-structured telephone calls with LSE academics working in different disciplines.

Results

There was a rapidly escalating trend in downloads as the collection increased in size. (See Figure 1.) The indisputable conclusion to be drawn from this was that having a high volume of theses available attracts much more traffic. The resulting pressure on the LSE’s repository then had a temporarily adverse impact on downloads per item, which dropped briefly to an average of 10 per item per month, but rapidly returned to the 2014 average of 15 downloads per item per month and at the time of the study was continuing to rise. Users were coming into the site from across the globe, preponderantly from Western countries and those from further afield with large economies.

Methods of entry and objectives of visitors

Figure 2 illustrates the key methods of entry to LSETO from January 2011 to February 2016 (on the left) and from January 2015 to February 2016 (on the right). The figure shows that the expansion of the repository did not substantially change access methods. The entry point for around 80% of users remained constant, with Google dominant in directing traffic. The decline in traffic share from the LSE’s own website suggests that the LSE could have done more to promote the theses collection on its website. Many LSE referrals came from the past PhD students’ page (e.g. for Statistics) or the research students’ profiles (e.g. for Sociology). Others came from clicks on collection profiles on the library pages.

The top searches via Google landed on the LSETO home page, which showed that many users were carrying out general searches for LSE theses, perhaps just to view an example of how a thesis had been laid out in their own area of research. However, other searches also led to specific theses, which demonstrated both Google’s strong indexing capabilities and that LSE researchers had contributed sought-after scholarly material to the theses collection.

The key question that the study wanted to answer was how much impact the theses collection was having on scholarly activity. Figure 3 shows download figures and Google Scholar (GS) citations for the LSE’s top ten downloaded theses. There was no demonstrable correlation between the numbers of downloads and the citations. Even some of the older (more than five years old) theses, which had had a reasonable timeframe in which to make an impact, had only achieved one or two citations.

It is interesting to speculate why theses are viewed for reasons other than their direct academic impact. It may be that they address a topical subject, have a broader societal impact, or are useful for reading lists. Delving further into access of the top LSE downloaded thesis yielded a little further information: this thesis appears on an international relations theory website and is cited in many foreign Masters theses, which probably accounts for its popularity.

Figure 4 digs deeper into the download/citation relationship, by examining the LSE theses with a strong (at least ten) citation count. The numbers of downloads for these varied considerably: some were impressively high, some quite low. The cross-hatched bars show works for which there is both an original thesis and a subsequent publication with a similar title, and for which the records have been merged in Google Scholar. The cross-hatched

bars stand out because they account for larger numbers of citations. This demonstrates that to achieve maximum academic impact, a researcher would be well advised to publish a journal article/book chapter/book deriving from the original work carried out for the thesis. However, since not all the theses achieved a large number of downloads, clearly readers only sometimes go back to the original research output (the thesis) to investigate the author's work in greater depth.

Table 1 gives further detail about the relationship between citations and downloads. It shows the top ten downloaded theses from the ProQuest digitization project. (They had therefore been available for approximately ten months online when the study was undertaken). The theses that had been most accessed did not necessarily have the highest citation figures. The top downloaded thesis is highly topical: it is on the Greek economy. But it was written in 1999, before Greece's current financial problems. Perhaps it was downloaded by researchers to get an idea of whether economists at the time were aware of nascent problems in the Greek economy. The most important finding for this project was that the number of times it had been downloaded made a powerful case for digitizing older theses, thus giving alumni the opportunity to reintroduce their research to contribute to relevant contemporary issues. Furthermore, if LSE theses are still considered important and worth consulting ten, 15 or 20 years after they were submitted, this can only enhance the reputation of the institution itself.

Routes to Access

Facebook is the largest social media application to direct traffic to LSETO, different from the largest application to the main institutional repository, [LSE Research Online](#) (Twitter). A possible explanation for this is that new PhD graduates still hesitate to engage with a wider audience to promote their research, preferring to choose a platform that enables them to share their work with a smaller, closer circle.

Figure 5 shows two access 'spikes' that occurred. On 19 June 2015, 473 people landed on the 'browse by year' page, which suggests that they were looking for a thesis but could not find it in LSETO. It was discovered that the majority of searchers came from Taiwan and may therefore have been looking for the thesis of the presidential candidate, who had recently been the subject of a Time magazine article which mentioned that she had completed her doctorate at the LSE. The Library at

the LSE had at the same time received e-mails enquiring about the availability of the thesis. (This thesis had not, in fact, been digitized.)

In the second spike, on 12 February 2016, 847 people landed on the Finance Minister of Finland's thesis after he promoted it through Facebook and Twitter. He was clearly proud of his work and wished to bring it into the 'Brexit' debate, 15 years after the thesis was originally submitted.

Awareness and perceptions of digital theses

Qualitative information to complement the statistics was obtained by carrying out three focus groups at the LSE, with undergraduates, postgraduates and librarians respectively. Of the seven undergraduates who took part (none of whom was British), only four knew of the digitization project. None had been recommended to consult theses as a scholarly resource, though one had been advised to look at a digital thesis for the layout. They were enthusiastic about the opportunities that digital theses offer for accessing cutting-edge research, pointing out that they become available much more quickly than monographs or journal articles. Some also said that, since each piece of research builds on the body of work that precedes it, it is useful to have a collection of theses going back some distance in time to provide a kind of audit trail. Nevertheless, most had reservations about making their own work available as a digital thesis. Some already had ambitions to publish their research with a traditional (they used the word 'proper') publisher and were worried that, if it appeared as a digital thesis, their chances of this would be damaged. This was a concern shared also by the postgraduates and academics.

"They were enthusiastic about the opportunities that digital theses offer for accessing cutting-edge research, pointing out that they become available much more quickly than monographs or journal articles."

The six postgraduate focus group participants (again, none was from the UK) were extremely motivated and quite high-powered: for example, one was working with the Bank of England, one was conducting research on far-right propaganda in France and one on food distribution and its effect on poverty in Asia. Only one knew of the digitization project and none had been told of it by their supervisors.

Like the undergraduates, they believed that complying with what is now an LSE requirement to provide an electronic copy of their thesis would undermine opportunities for commercial publication. They had some other concerns, too: one thought that publishing what she described as 'academic juvenilia' might impair her reputation later on, 'when I am famous', as she put it. This group also raised concerns about copyright and permissions issues. They were unanimous in their view that the number of citations made a digital thesis of potential interest while the number of downloads did not.

The study found that the perception that publishing a thesis digitally affects future publication opportunities has almost become a *bête noire* in academic circles. The concept is so ubiquitous that it is one of the areas relating to digital theses in which a considerable amount of research has already taken place. Marisa L Ramirez, with several colleagues on each occasion, has carried out two fairly large-scale pieces of research entitled 'Do Open Access electronic theses and dissertations diminish publishing opportunities in the Sciences?'¹ and 'Do Open Access electronic dissertations and theses diminish publishing opportunities in the Social Sciences and Humanities?'². Ramirez and her colleagues found that in the Sciences, 'a slim majority of science journals (51.4%) reported that manuscripts derived from openly accessible electronic theses and dissertations ETDs are always welcome for submission, and an additional 19.4% of science journals would accept revised ETDs on a case-by-case basis', and in the social sciences and humanities (HSS), 45% of respondents considered that 'manuscripts that are revisions of openly accessible ETDs are always welcome for submission' and 27% of respondents would consider such manuscripts on a case-by-case basis. Only 12.5% of editors in the sciences and 4.5% in SSH indicated that they would under no circumstances consider such material for further publication. There is some slight risk, therefore, in making a thesis available via electronic Open Access, but it is nowhere near as great as most researchers imagine.

The librarian focus group participants agreed that having the digital collection of theses was good for the reputation of both the University and the Library itself – 'Academics expect us to be there to sort out this kind of thing' – but that better metrics were needed in order to be able to assess impact more effectively. They also pointed out that there is a lot more to building and maintaining a

digital thesis collection than straightforward digitization. They, in conjunction with colleagues, had had to do a lot of time-consuming training and hand-holding, especially in the form of providing support on such issues as copyright and permissions.

Managing expectations was also a problem: 'We have 3,500 digitized theses. We have that number again and more that haven't been digitized. This means that we are bound to disappoint people who think that all our theses have been digitized.'

Of the four academics who took part in the semi-structured telephone calls, two knew of the digitization project. Two were in favour of digital theses, while the other two had reservations, again connected with copyright and publication issues. Three had consulted digital theses and all could see where their value lay: helping promote cutting-edge research and putting new research in a historical context were particularly mentioned. Like the postgraduates, they were interested only in citations, not in downloads.

"Digitization of a large proportion of its thesis collection has brought many more users to LSETO."

Conclusions

Digitization of a large proportion of its thesis collection has brought many more users to LSETO. It is difficult to assess with accuracy the impact this has had on scholarly research: there is no demonstrable direct relationship between downloads and citations. The study does seem to indicate that if researchers wish to make an impact on the body of scholarly resources, subsequent publication in a book or journal as well as publishing the thesis digitally seems to achieve the best results; and published research consulted to support the project suggests that making the thesis available digitally has only a minor effect on curtailing publication opportunities. Furthermore, impact at the micro level should also be recognised as important. The LSE's alumni deserve the opportunity for dissemination that having their work included in a digital thesis collection offers, especially

when it may have languished in unjustified obscurity because of more limited opportunities to promote it when it was originally submitted.

The project demonstrated that more proactive work by librarians and academic departments needs to take place to enable the better promotion of digital theses. The LSE has now used the evidence that very few authors opted out of the ProQuest digitization project and that few take-down requests were received to push for a change in its policy with ETHOS. It has introduced a policy adopted by many other UK universities, including Surrey, no longer to chase author permissions, but instead to view digitization as merely a format change which therefore requires no permission. A take-down policy is still in operation for the few authors who object. This approach allows inclusion of more older theses in the collection and helps to generate repository growth.

Endnotes

1 See Ramirez M L, McMillan G, Dalton J T, Halon A, Smith H S and Kern C, Do Open Access electronic theses and dissertations diminish publishing opportunities in the Sciences?, College and Research Libraries, Early view: <http://crl.acrl.org/content/early/2013/10/29/crl13-524.short>.

2 See Ramirez M L, Dalton JT, McMillan G, Read M and Seamans N, Do Open Access Electronic Dissertations and Theses Diminish Publishing Opportunities in the Social Sciences and Humanities? Findings from a 2011 Survey of Academic Publishers, College and Research Libraries, crl 356 – 380; DOI: <http://dx.doi.org/10.5860/crl-356> (accessed 9 June 2016).

Appendix: Figures and Table

Figure 1.

Relationship between number of theses in LSETO and downloads each month

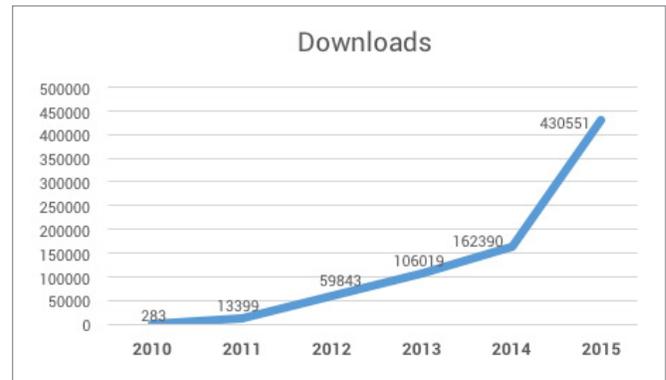


Figure 2.

Largest sources of traffic to LSE Theses Online

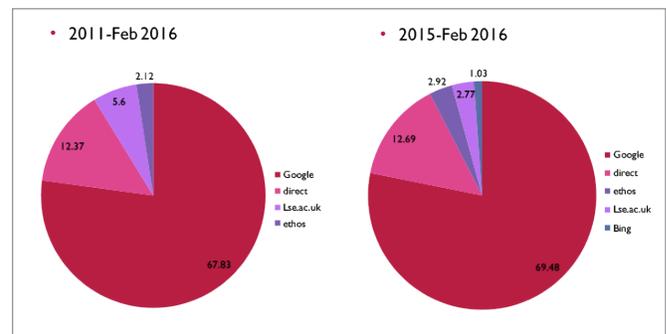
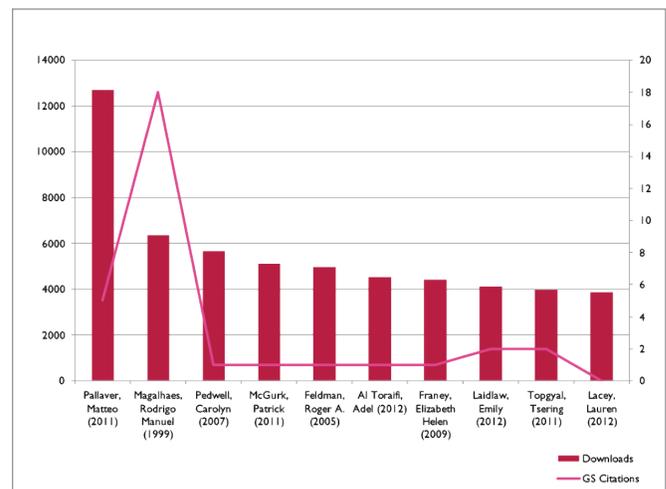


Figure 3.

Top ten downloaded theses in LSE Theses Online



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

Figure 4.

Sample of theses with at least ten citations according to Google Scholar (GS).

The cross-hatched bars show subsequent publications which have a similar title to the thesis. * denotes download statistics are from EThOS as the full text is not in LSETO. NB: The Dolan thesis has a related book publication, but the records are not merged in GS as the titles are substantially different.

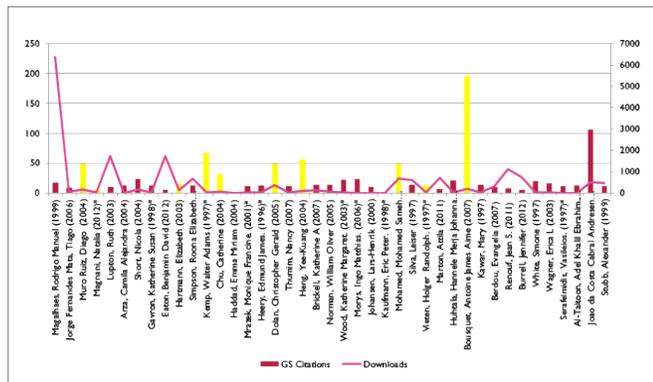


Table 1.

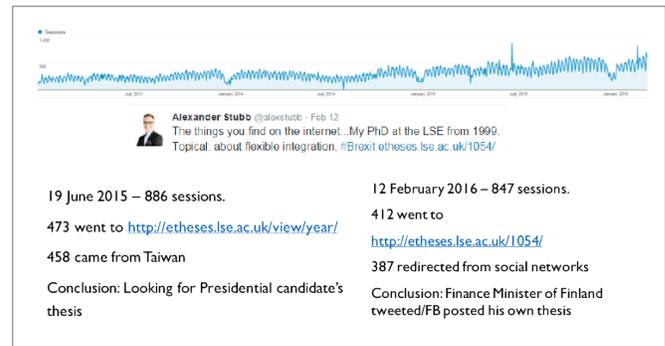
Top ten downloaded theses from the ProQuest digitization project

AUTHOR & TITLE	DOWNLOADS	GOOGLE SCHOLAR CITATIONS	PUBLISHED
Konsolas, Ioannis (1999) The competitive advantage of nations: The case for Greece	1705	0	
Yaffe, Helen (2007) Ernest 'Che' Guevara: Socialist political economy and	927	1	
Ahuja, Monika Sangeeta (1996) Public interest litigation in India: A socio-legal study	900	1	
Holt, Andrew Derek (2005) The role of management accounting within the development of environmental management systems	895	1	
Mohamed, Mohamed Sameh Ahmed (1997) The role of the International Court of Justice as the principal judicial organ of the United Nations	686	49	Book 2003
Baxell, Richard (2002) The British Battalion of the International Brigades in the Spanish Civil War	677	1	
Michelutti, Lucia (2002) Sons of Krishna: The politics of Yadav community formation in a north Indian town	520	4	
Joao da Costa Cabral Andresen Guimaraes, Fernando (1992) The origins of the Angolan civil war. International politics and domestic political conflict 1961-1976	514	105	Book 1995
Stubb, Alexander Cai-Goran (1999) Flexible integration of the Amsterdam Treaty: Negotiating differentiation in the 1996-97 IGC	491	12	
Fuller, Harcourt (2010) Building a nation: Symbolic nationalism during the Kwame Nkrumah era in the Gold Coast/Ghana	483	1	

© Copyright to this case study is jointly owned by ProQuest, Dimity Flanagan, Research Librarian at the LSE, and Linda Bennett.

Figure 5.

Spikes in traffic to LSE Theses Online



Alexander Stubb @alexstubb · Feb 12
The things you find on the internet... My PhD at the LSE from 1999.
Topical: about flexible integration. #Brexit etheses.lse.ac.uk/1054/

19 June 2015 – 886 sessions.
473 went to <http://etheses.lse.ac.uk/view/year/>
458 came from Taiwan
Conclusion: Looking for Presidential candidate's thesis

12 February 2016 – 847 sessions.
412 went to <http://etheses.lse.ac.uk/1054/>
387 redirected from social networks
Conclusion: Finance Minister of Finland tweeted/FB posted his own thesis