

Legal Deposit Web Archives and the Digital Humanities: A Universe of Lost Opportunity?

Dr. Paul Gooding (University of East Anglia); Prof. Melissa Terras (University of Edinburgh); Linda Berube (University of East Anglia).
www.elegaldeposit.org @elegaldeposit



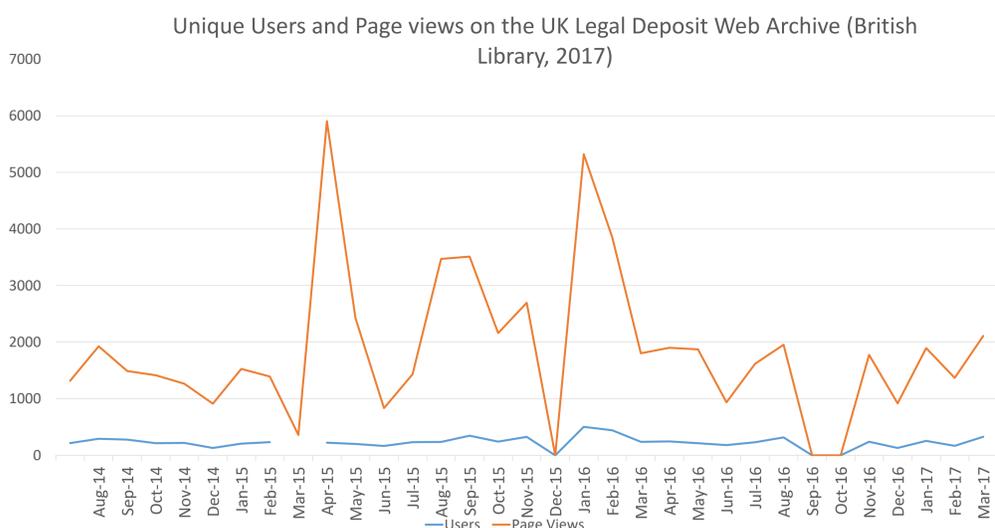
What is the *Digital Library Futures* Project?

Digital Library Futures is a two-year research project to investigate the impact of e-legal deposit collections on two key stakeholders in the United Kingdom: academic deposit libraries, and existing and potential users of e-legal deposit collections. Legal Deposit is the legal requirement that a person or group submit copies of their publications to a trusted repository, and ensures the preservation of a nation's published output. Since the UK's introduction of electronic legal deposit regulations in 2013, digital publications are now systematically collected for posterity. We are exploring what it means for the UK's national libraries, and for Digital Humanities Researchers.

Research Context: The UK Legal Deposit Web Archive (UKLDWA)

The British Library began harvesting the UK web domain under legal deposit in 2013, a step towards preserving the UK's "digital universe" (National Library of Scotland, 2013). Although the boundaries of this universe are unclear, its scale as an archival collection is huge: the first web crawl in 2013 produced 30.84 terabytes of data; by 2017, the total collection had grown to 500Tb (British Library, 2017). A web archive of this scale promises significant opportunities for DH scholarship. However, this potential is affected by the UK Legal Deposit regulations, which are founded on the centuries-old model applied to print resources within the walls of legal deposit libraries. This means that materials that were freely available via the web become inaccessible outside library reading rooms as soon as they are deposited. This raises the following vital research questions:

- How does the current framing of e-legal deposit in the UK affect usage, both in terms of user numbers and digital innovation?
- Is legal deposit, a concept defined and refined for print materials, the most suitable vehicle for collecting web archives that allow reuse by DH researchers?



Overall Usage

National library collections received under legal deposit in the UK are traditionally reference-only. In other words, readers are usually not able to access legal deposit works outside the reading rooms of the six UK legal deposit libraries. Unlike printed works, electronic resources can be accessed remotely; indeed, user expectations have shifted towards ubiquitous remote access. Despite this, the UK regulations still treat legal deposit electronic materials in the same way as print: in the case of the UK Legal Deposit Web Archive, discovery and retrieval is only available within the reading rooms of the six UK legal deposit libraries.

The graph to the left shows overall usage of the UKLDWA, which receives an average of 226 users per month, accessing a combined 1859 pages per month. This chimes with the impressions of the staff we spoke to, one of whom commented that "very few people are sitting down and... doing a piece of academic research with it." Given the posterity-driven objectives of legal deposit, though, what looks like success in terms of usage?

Text and Data Mining

Winters argues that "it is the portability of data, its separability from an easy-to-use but necessarily limiting interface, which underpins much of the exciting work in the Digital Humanities" (2017, p.246). Restrictions in using derivative data from the UKLDWA mean that researchers must often look to not-for-profit organisations like *Common Crawl* (<http://commoncrawl.org/>) and the *Internet Archive* (<https://archive.org>) for portable data. These organisations can explore the boundaries of copyright legislation to support their mission.

This contrast between deposit libraries, described by our interviewees as risk averse due to regulations and publisher relationships, and risk-enabled independent organisations, challenges library/DH collaboration models. What does this changing information ecology mean for libraries and researchers in the future?



Next Steps

The current framing of legal deposit in the United Kingdom has a clear impact on reuse of web archival materials. Limitations on access ensure that very few people use the UK Legal Deposit Web Archive, while barriers for text and data mining challenge existing models for library/DH collaboration. What is less clear is how these patterns of use are consistent with the strategic aims of legal deposit: in other words, do we need to reconsider the conceptual frameworks for legal deposit in light of the differing affordances of the born-digital universe? The next steps in answering the questions raised here are to:

- Provide empirical case studies of the impact of e-legal deposit collections on our project partners, the Bodleian Libraries and Cambridge University Libraries, and their users;
- Develop analysis of e-legal deposit usage metrics in the United Kingdom, combined with user surveys and interviews with library staff, practitioners, and key stakeholders.