



elegaldeposit.org

Towards Critical Addressability in Digital Library User Studies

Paul Gooding
(University of
Glasgow)



“Digital Library Futures”

- Two year AHRC-funded project to investigate the impact of e-Legal Deposit on UK Academic Deposit libraries:
 - Case study partners: Bodleian Libraries & Cambridge University Library.
- Focus on academic deposit libraries:
 - Shift focus away from national libraries and towards the specific problems faced within academic libraries.
- First ever public user-centric study of the impact of e-legal deposit.
 - Normally focused on technical, preservation, and long-term aspects of these collections, not contemporary usage.
- Aim to address several challenges created by tension between user and publisher rights.

What is Legal Deposit?

- Legal Deposit – “the legal requirement that a person or group submit copies of their publications to a trusted repository or repositories.”
- **Electronic Legal Deposit:** broad term to denote legal deposit regulations that apply to digital materials.
- **Non-Print Legal Deposit:** the specific term for the UK’s e-legal deposit regulations.
- **The Legal Deposit Libraries (Non-Print Works) Regulations 2013** refer to work in writing – “(a) transmitted by electronic means; (b) received in legible form; and (c) capable of being used for subsequent reference” (2013).

USEFUL

Use, *n.*

- “The act of putting something to work, or employing or applying a thing, for any (esp. a beneficial or productive) purpose; the fact, state, or condition of being put to work, employed or applied in this way; utilization or appropriation, esp. in order to achieve an end or pursue one’s purpose. Frequently with *of*” (Oxford English Dictionary).

What do I mean by “Critical Addressability”?

- Thomas Padilla refers to “Critical Addressability” as the notion that one should be able to evaluate the technical and social forces that shape data, through data documentation and transparent workflows:

"A researcher should be able to understand why certain data were included and excluded, why certain transformations were made, who made those transformations, and at the same time a researcher should have access to the code and tools that were used to effect those transformations" (Padilla, 2016).

The “Black Box” problem in Library User Studies

- e.g. Google Analytics widely adopted in studies of digital library usage:
 - These largely adopt a “black-box” approach – some customisation but very little (almost no) critique of what’s going on inside that box.
 - Lots written about the benefits and drawbacks of web analytics, but (almost) nobody in LIS discusses *how* these tools collect, process and analyse data.
- Two key points:
 - Raw data is hidden from GA users for reasons including data privacy, resulting in a “lack of transparent, reproducible data” (Gooding, 2016).
 - “There is no reason to believe that library patron data and its analysis will be free from bias. On the contrary... we should expect sexism, racism, homophobia, and bias against additionally societally marginalized groups” (Asher et al., 2018, p.4).

The datasets

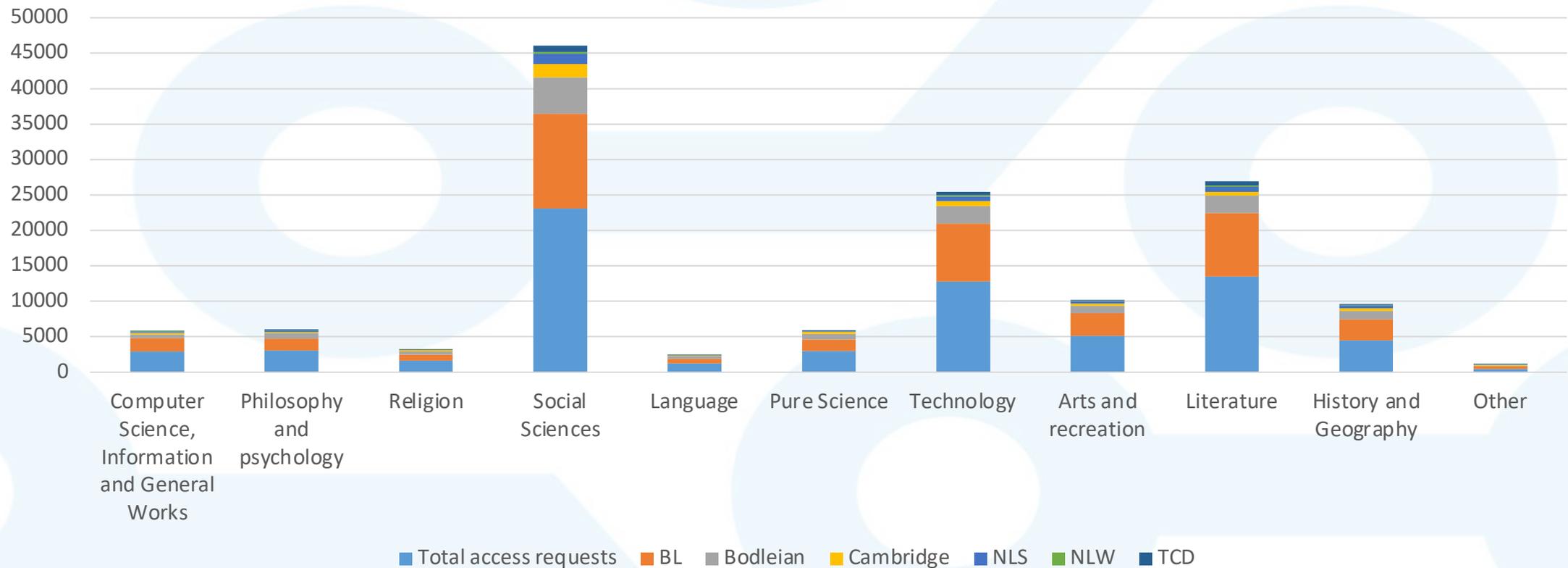
- Two datasets: both contained all NPLD requests within UK Legal Deposit Library reading rooms – 31st July 2015 to 31st March 2017:
- Bibliographic metadata relating to titles requested from fixed terminals:
 - Metadata for all **eBook** title requests – total 91,809 requests (title-level).
- Metadata provided: date and time of access request; originating legal deposit library; title of book or article; journal title (where applicable); publisher; and ISBN or ISSN.
- Provided as CSV file and cleaned in OpenRefine.

The *Subjectify* Tool

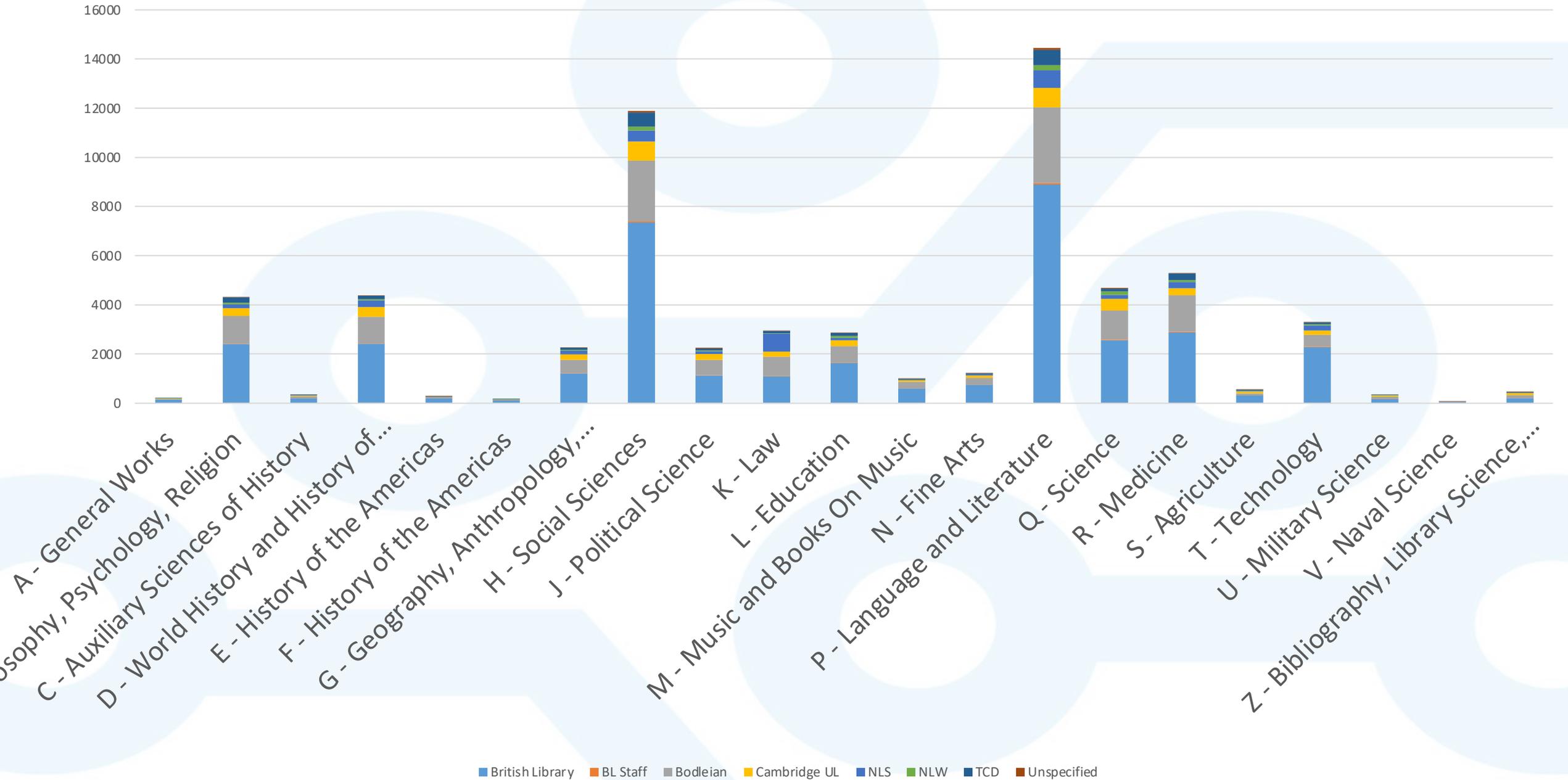
- [Subjectify](#) is a Python-based tool:
 - Required to preserve anonymity: no identifiable information about users, but possible to infer information about users from the works they consulted.
 - Ethically necessary to consider microanalytic approaches – and to identify meaningful usage patterns.
- Uses the [OCLC Classify2](#) API Service to automatically obtain Dewey Decimal (DDC) and Library of Congress (LCC) classmarks:
 - Author, title, ISBN, and ISSN data taken from a provided CSV file;
 - Tool designed to work on varied data sources, with different options for how to locate relevant fields;
- Discarded unclassified records and used the remaining records to identify subject-based patterns of usage of NPLD materials.

NPLD Reflects Long-Established Disciplinary Usage (eBooks, Dewey Decimal Classmarks)

Comparison of NPLD Book Access Requests by Subject (May 2015 to April 2017)



The same data, but using Library of Congress Classmarks (LCC)



The big(gest) problem with Library Classification

- Library classification is a subjective process undertaken by humans that reflects existing biases (Mai, 2010).
- DDC provides distinct categories for English, American, and classical European schools of literature, while lumping the rest of the world under “other literatures”:
 - Bias emerges from the 19th Century North American perspective of DDC (Kua, 2008).
- Automatic matching of this kind embeds existing bias into our data, and problematic perspectives:
 - This bias works well for UK-centric library collections (NPLD is a record of UK publications);
 - But what about the wider applicability of this method? How do we become “ethical stewards” (Weingart, 2014) of library usage data?

Possible next steps?

- Compare these findings to subject usage of non-NPLD materials.
- Investigate other ways to derive subject data from metadata records.
- How might a fruitful conversation between Information Science and other fields develop more ethically and critically informed approaches to “representing” (Unsworth, 2000) library user data?
- Ethical and critical approaches to the creation and interpretation of library user data - recognise user analysis as an act of interpretation, not mere “reporting” of statistics.

Project Partners and Funding

- This research was supported by a grant from the Arts and Humanities Research Council under grant number AH/P005845 between June 2017 and May 2019.
- The project white paper is available to download here: <http://elegaldeposit.org/dlf-white-paper>





elegaldeposit.org