# The research librarian of the future: data scientist and co-investigator

**blogs.lse.ac.uk**/impactofsocialsciences/2016/12/14/the-research-librarian-of-the-future-data-scientist-and-co-investigator/

*There remains something of a disconnect between how research librarians themselves see their role and its responsibilities and how these are viewed by their faculty colleagues. **Jeannette Ekstrøm**, **Mikael Elbaek**, **Chris Erdmann** and **Ivo Grigorov** imagine how the research librarian of the future might work, utilising new data science and digital skills to drive more collaborative and open scholarship. Arguably this future is already upon us but institutions must implement a structured approach to developing librarians' skills and services to fully realise the benefits.*

Imagine research librarians as equal partners in the research process, helping a researcher in any discipline to map existing knowledge gaps, identify emerging disciplinary crossovers before they even happen, and assist in the formulation and refinement of frontier research questions.

Imagine a librarian armed with the digital tools to automate literature reviews for any discipline, by reducing thousands of articles' ideas into memes and then applying network analysis to visualise trends in emerging lines of research.

What if your research librarian could then dig deeper and use an ami-2word plug-in to map in which sections of articles your key research terms appear? Imagine the results confirmed that your favourite research term almost never appears in the results sections, but cluster only around introductions and perspectives.



**Image credit: Imagine… by Gabi L. This work is licensed under aCC BY 2.0 license.**

And what if the librarian did not stop there, but zoomed into the cloud of data with savvy statistics, applying the latest text and data mining techniques to satisfy even the most scrutinising scientific mind, before formulating an innovative

research question?

Imagine a librarian who understands, in pragmatic terms, the benefits of Open Science to the discovery process. Imagine a librarian who also has practical advice on how to make those ideas part of your daily workflow. Would you like that librarian to help you kick-start your academic career?

It may sound too good to be true, but in a way it is already happening.

## Core duties versus 'stretch' services

The research librarian community is not in consensus as to what exactly are the emerging roles of future librarians in a rapidly evolving digital scholarship environment (see #libraryfutures). Added to the polarised views within that community, a recent survey shows there is also a clear gap in perception and expectations between librarians and faculty staff. While librarians surveyed agreed that "information literacy" and "aiding students one-on-one in conducting research" are primary and essential roles, they viewed "supporting faculty research" as less important than their faculty colleagues. So does this present an opportunity in the digital age?

## Librarian as co-investigator, not an overhead

In the digital age, many of the skills and competencies librarians develop to perform 'core' services can actually directly serve the research lifecycle and workflow. Competencies such as mapping the knowledge landscape, digesting volumes of heterogeneous data or presenting in understandable formats are not things every researcher is armed with but which every hypothesis can benefit from.

By using their data science and digital skills, research librarians have the opportunity to make an impactful contribution to the workflow of their faculty colleagues. Librarians' data science skills can help navigate through the deluge of information, and can truly change how they are perceived: from an overhead service to research co-investigators.

Despite the opportunities, it would be easy to frustrate the community of librarians by calling for a skills upgrade without contributing in small steps to filling any skills gap. If data science skills are not part of an institution's strategy, it is difficult to find time and resource to upgrade an individual's skillset while fulfilling existing contractual obligations. This is where existing structured training philosophies can help with both skills and convincing institutional managers of the strategic benefits of boosting data science capacities. Some examples of data science training include Data Carpentry, Data and Visualization Institute for Librarians, Library Carpentry, Data Science Training 4 Librarians and there are likely many others too.

## Data science as a driver of Open Science by default

As the need for more open and transparent scholarship permeates through funder mandates, research librarians become an indispensable partner in optimally disclosing the diverse outputs of the research process; from advice on choice of appropriate licenses for re-use, to best long-term curation and persistent identifier (PID) assignment in synergy with existing intellectual property rights practices.

But to keep in step with the trend, "librarians must provide research data supporting services in the digital age", and institutions need a structured approach to "enhance research librarians' data skills, RDM & data services". The data science skillset of librarians is also considered by some graduate schools (e.g. those engaged with FOSTER Project 2014-2016) to be a deciding factor if Open Science is to form part of the standard skillset taught to postgraduates.

To meet Open Science implementation needs, calls for the boosting of institutional data skills extend across disciplines (NSF-NITRD Federal Big Data Research & Development Strategic Plan, May 2016) further downstream to "new skills in data science, data analysis and visualisation" and "text and data mining of content".

Making the future librarian an indispensable research partner to faculty would not only close the gap in how the role is perceived, but also create a self-sustaining conduit for including best practices in collaborative and open scholarship, and implementing Open Science by default. Ultimately, everyone would get more impact.

*Note: This article gives the views of the authors, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our comments policy if you have any concerns on posting a comment below.*

**About the authors**

*Jeannette Ekstrøm holds a Master of Library and Information Science from the Royal School of Library and Information Science, Copenhagen and has been working with information literacy, user education and openness at DTU Library. Jeannette coordinates Data Science Training 4 Librarians in the EU in 2015 and 2016 (@DST4L) and can be followed on Twitter at @JEkstroem.*



*Mikael Elbaek  holds a Master of Library and Information Science from the Royal School of Library and Information Science, Copenhagen and is the coordinator of the Research Data Management Team at Technical University of Denmark. He has been a strong advocate for openness in science and been involved in policy development and infrastructures since 2005. Mikael can be followed on Twitter at @melbaek.*



*Chris Erdmann holds a Master of Library and Information Science from the University of Washington iSchool and a BA from the University of California, Davis. He is the Chief Strategist for Research Collaboration at the NCSU Libraries, and he formulated the Data Science Training 4 Librarians curriculum in 2014. Chris can be followed on Twitter at @libcce.*



*Ivo Grigorov is an Open Science enthusiast and research coordinator at Technical University of Denmark. As a member of the FOSTER Consortium, he advocates for the strategic and career benefits of Open Science to both individuals and research institutions. Ivo can be followed on Twitter at @OAforClimate.*