



Platforms, programmability, and precarity: the platformization of research repositories in academic libraries

LSE Research Online URL for this paper: <http://eprints.lse.ac.uk/118824/>

Version: Published Version

Article:

Plantin, Jean-Christophe ORCID: 0000-0001-8041-6679 and Thomer, Andrea (2023) Platforms, programmability, and precarity: the platformization of research repositories in academic libraries. *New Media & Society*. ISSN 1461-4448

<https://doi.org/10.1177/14614448231176758>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>



Article

Platforms, programmability, and precarity: The platformization of research repositories in academic libraries

new media & society

1–21

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448231176758

journals.sagepub.com/home/nms



Jean-Christophe Plantin 

London School of Economics and Political Science, UK

Andrea Thomer

University of Michigan, USA

Abstract

We investigate in this article how repository platforms change the sharing and preservation of digital objects in academic libraries. We use evidence drawn from semi-structured interviews with 31 data repository managers working at 21 universities using the product *Figshare for institutions*. We first show that repository managers use this platform to bring together actors, technologies, and processes usually scattered across the library to assign to them the tasks that they value less—such as data preparation or IT maintenance—and spend more time engaging in activities they appreciate—such as raising awareness of data sharing. While this platformization of data management improves their job satisfaction, we reveal how it simultaneously accentuates the outsourcing of libraries' core mission to private actors. We eventually discuss how this platformization can deskill librarians and perpetuate precarity politics in university libraries.

Keywords

Data repository, labor, libraries, maintenance, metadata, platforms

Corresponding author:

Jean-Christophe Plantin, Department of Media and Communications, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

Email: J.Plantin1@lse.ac.uk

Introduction

In this article, we analyze how information professionals working at academic libraries use research repository platforms, what universities and libraries gain from them, but also how platforms potentially conflict with the core missions of libraries. Institutional repositories are a set of services offered by a university library to support the archiving and sharing of scholarship produced by a university's faculty and students (Lynch, 2003); data repositories are institutional repositories that specifically archive research data. We refer to them collectively as research repositories, and the work of maintaining them as repository management. Several researchers have already highlighted the larger politics of digitization in libraries, for example, by revealing the ideologies underpinning mass digitization across institutions (Thylstrup, 2019), the difficulties for long-term preservation of social media data (Zimmer, 2015), or the dangers of privatization of the public role of archives when tech giants fill this role (Jeanneney, 2008; Vaidhyanathan, 2011). We complement this scholarship by focusing here on the capacity of digital platforms (Plantin et al., 2018a; Helmond, 2015; van Dijck et al., 2018) to reorganize repository management. In this article, we ask: How do platforms shape repository management in university libraries? What are both the practical consequences for the work of repository managers, and the larger implications for the long-term mission of libraries when universities rely on platforms as archives?

To answer these questions, we focus on the cloud-based repository environment, *Figshare for institutions*. The company Figshare was created in 2011 by Dr Mark Hahnel, a PhD graduate from Imperial College London, and its main product is the website *figshare.com*, where researchers can publish a wide range of research outputs (such as datasets, graphics, or presentation slides). The company launched *Figshare for institutions* in 2013 to specifically target academic libraries. *Figshare for institutions* provides "software as a service" (SaaS) in which university libraries pay a subscription to get a customized cloud-based Figshare module that they can use as either an institutional or a data repository, or both.^{1,2} *Figshare for institutions* has seen considerable uptake by libraries, and count 87 universities and research institutes as clients at the time of this writing.

Literature review

Scholarship on Figshare

Figshare has already generated an extensive scholarship in library and information science (LIS). This literature is, however, mostly instrumental and quantifies, for example, the differences of upload and use of open data depending on whether a document is viewed or shared (Thelwall and Kousha, 2016) or the distribution of document and usage data on Figshare (Kraker et al., 2015). Figshare also appears in review articles listing the pros and cons of research data management products (Amorim et al., 2017). Our own past research went beyond such an instrumental view and analyzed Figshare's platform architecture to show its interaction with existing knowledge infrastructures (Plantin et al., 2018b). We complement this research by investigating the specific point of view of repository managers using *Figshare for institutions* at university libraries, and how this platform changes their role and status within the institution.

LIS scholarship on institutional data repositories

Academic libraries and other knowledge institutions have over the last 30 years dedicated significant resources to meeting the technical needs of the information age and building capacity to support next-generation modes of scholarly communication in libraries. These efforts have been spurred by the increasingly computational nature of scholarship (Hey et al., 2009), calling for increased reproducibility and transparency in scholarship (Baker, 2016; Lynch, 2008; Sayre and Riegelman, 2018), and numerous mandates from governments, publishers, and funders to make science more open (Bill & Melinda Gates Foundation, n.d.; Nature, 2023; Nelson, 2013) and FAIR (i.e. Findable, Accessible, Interoperable, and Reusable, cf. Wilkinson et al., 2016). A major focus of this effort has been the development of institutional repositories, defined by Cliff Lynch as:

a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution. (Lynch, 2003: 3)

Our work here is concerned with organizational impacts of the technical components of research repositories: namely, the platform used to manage digital materials and the standards used to describe materials. Early institutional repositories were largely built on open-source, community-developed software, such as Dspace, and many of these initiatives additionally developed their own in-house IT teams, and digital curation and preservation best practices (Lynch and Lippincott, 2005); later, open-source solutions such as Dataverse and CKAN were eventually developed to more specifically support research data management (e.g. Crosas, 2011; Winn, 2013). This community-developed repository software typically uses metadata standards developed by the LIS community (by both practitioners and researcher) to share and archive materials in a standardized way. Metadata is colloquially defined as “data about data,” which are used to find and retrieve resources in digital and physical collections (Gilliland, 2016). By using a common metadata standard, libraries, archives, museums (LAMs), and other memory institutions make their holdings easier to search and aggregate (e.g. via initiatives like the Digital Public Library of American [DPLA] or Europeana, both of which harvest metadata records from many institutions to make them searchable in one place). Metadata standards are critical for interoperability between collections and the long-term preservation of data (McDonough, 2009; Sandy and Freeland, 2016; Zeng and Chan, 2006); however, even when a common standard is used, care must be taken to ensure that it is implemented in a consistent way (Maron and Feinberg, 2018; Millerand and Bowker, 2009).

As we describe below, Figshare is emblematic of a continuing shift away from community-developed infrastructure within libraries. First, it is a commercial product (though by no means the only commercial software adopted by libraries). Second, it is a cloud- and subscription-based service, which is licensed (not owned) by an institution. Finally, it prioritizes the use of industry standards over library-developed standards to share metadata—specifically a standard called schema.org, which was developed by tech

giants like Google and Yandex. *Figshare for institutions* is not the only commercial or SaaS platform for data repository management (others include Digital Commons or the Open Science Framework for Institutions, respectively); we focus on it as a case study because it has been broadly adopted by academic libraries and our findings can be applied to similar platforms.

Platform studies

The success of *Figshare for institutions* comes from its capacity to leverage the properties associated with platforms (Plantin et al., 2018b): it is programmable (Helmond, 2015) as it provides a platform that repository managers can customize to their needs. It is an intermediary between the different users of the service (Gillespie, 2010). It is modular (Baldwin and Woodard, 2008; McKelvey, 2011) with existing university IT infrastructure and other scholarly communication modules via Application Programming Interfaces (APIs).

Fundamentally, the appeal of *Figshare for institutions* for repository managers is that it creates a platform ecosystem (van Dijck et al., 2018) that brings together previously separated components of scholarly communication. As we show below, the platform brings together researchers or students, platform tech support, discoverability standards, and data mandates into a stable ecosystem. Once in place—and echoing a movement of centralization between core and complementary components typical of platforms (Blanke and Pybus, 2020; Gerlitz and Helmond, 2013; Helmond, 2015)—repository managers use the platform to redistribute many parts of their work to the components linked together.

Sample and methodology

We conducted semi-structured interviews with university library staff members responsible for or working with Figshare at their institution. Using the list of 81 research institutions using *Figshare for institutions* on the Figshare website at the time,³ we searched online for the names and email addresses of the Figshare administrators at each university. We found and contacted 68 of them, and 31 contacts working at 21 universities responded positively to our interview request. While we use in this article the generic term of “repository managers,” our respondents identified with several different titles, mostly librarians, but not exclusively, cf. Table 1. As noted above, Figshare can be used as either an institutional or data repository and our respondents used Figshare in a range of ways; however, we did not find that the use of Figshare as an institutional versus data repository impacted our results so we pool them together below. The participants in our study work at universities located in Europe, North America, and in British Commonwealth countries in Australasia, and Africa which is representative of the global customer base for this UK-based product. We conducted all the interviews via Zoom between April and June 2021 after piloting the question guide. Interviews were with one to three respondents at a time and lasted between 26:32 and 66:44 minutes. Respondents were not compensated. Prior to the interview, we sent an information sheet and consent form via email, and we collected their consent. Audio files were then transcribed via a paid service. We

Table 1. Universities' location and professions of our respondents.

Institution	Continent	Title of respondents
University 1	Australasia	Research Infrastructure Librarian
	Australasia	Research Infrastructure Librarian
University 2	Australasia	Manager, Research and Strategic Projects
	Australasia	Repositories and Access Librarian
University 3	Australasia	Manager of Digital Stewardship
University 4	Australasia	Manager, Research and Scholarly Information Services
University 5	Australasia	Senior Coordinator of the Scholarly Publications Team
University 6	Australasia	Senior Librarian
	Australasia	Manager for Library Information and Research Services
	Australasia	Senior Librarian, Research
University 7	Australasia	Digital Projects Officer
University 8	Africa	Electronic Resources Librarian
University 9	Africa	Principal Librarian: Digital Scholarship
University 10	Africa	Acting Assistant Director: Digital Scholarship Center
University 11	Africa	Institutional Repository Manager
University 12	Africa	Librarian: Research Data Management
University 13	Europe	Research Support Team
	Europe	Research Support Team
	Europe	Research Support Team
University 14	Europe	Research Data Analyst
University 15	Europe	Head of Research Services
	Europe	Research Service Consultant
University 16	Europe	Library Research Services Manager
University 17	Europe	Research Data Manager
University 18	Europe	Data and Systems Officer
	Europe	Technical Analyst
	Europe	Education and Research Services Manager
University 19	North America	Research Data Management Specialist
	North America	Assistant Director for Data Management & Curation Services
University 21	North America	Institutional Repository Specialist
	North America	Scholarly Communications & Research Curation Consultant
	North America	

then conducted a thematic analysis (Braun and Clarke, 2021) of the transcript by proceeding both deductively (by coding for themes that we had already identified in the literature and our past research) and inductively (by coding for new themes that emerged from the transcripts and were engaging with our research questions or with past literature). This thematic analysis took the form of a codebook that we collectively created in NVivo and that contained the deductive and inductive codes. After the separate coding of

three interviews, we obtained an inter-coder reliability score of .825 using the Holsti method; this score is commonly agreed to be reliable when paired with discussion between coders (Mao, 2017). We then used the codebook to code the whole corpus of interviews. In parallel to coding, we wrote extensive memos that we shared during meetings to conduct the analysis.

Despite the global reach of our respondent pool, we did not note significant differences due to geography beyond differences in the specifics of national data sharing policies. Consequently, we did not separate our findings by geography. Our study is additionally limited in its exclusive focus on repository managers; however, while administrative and user perspectives would certainly be valuable, a focus on practitioners makes it possible to understand how library roles shift in response to platform adoption.

Summary of results and outline

Our analysis brings two sets of results. We show in the section *Creating an ecosystem for data sharing with Figshare* how platforms like *Figshare for institutions* reshape repository management in university libraries. We reveal that repository managers use this platform to bring together actors, technologies, and processes usually scattered across the library: researchers or students with datasets to share, support for installing and maintaining the platform, discoverability standards to increase visibility of data sets, and mandates for data management. Once part of the same platform ecosystem, repository managers can then redistribute to the actors now on the platform some tasks that they value less, such as data preparation or IT maintenance. In section three: *The Platform Trade-off: Renouncing Control for More Ease-of-Use*, we investigate the practical consequences *Figshare for institutions* has on the work of them. This platformization of data management generates a trade-off for repository managers: on the one hand, by reducing the technical and curatorial part of their work, it increases the part of their job they value more such as user engagement. On the other hand, this delegation of curatorial and technical work to commercial platforms threatens to undermine the LIS community's governance over their tools, standards, and practices, which eventually weakens their capacity to shape data management and policies. The last discussion section invokes the larger implications of this platformization of data management for the long-term mission of libraries. By outsourcing libraries' core mission to private actors, platforms potentially deskill librarians and perpetuate precarity politics in university libraries. We conclude with suggestions to combine community standards with commercial platforms.

Creating an ecosystem for data sharing with Figshare

Research repository as a platform ecosystem

In describing their reasons for adopting *Figshare for institutions*, participants in our study refer to the positive changes it brought to their curatorial workflows. The work of repository managers is quite diverse (Bishop et al., 2022) and has generated various lifecycle models (e.g. Faundeen et al., 2013; Higgins, 2008), but we identified a common core

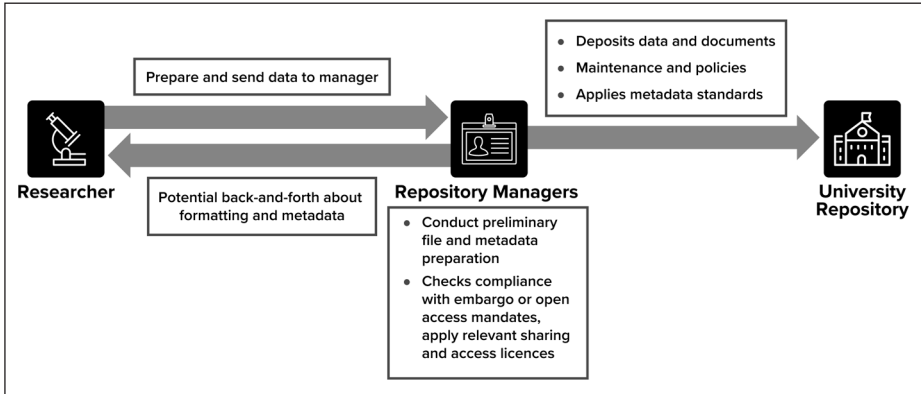


Figure 1. Typical data management workflow.

pattern for our respondents (Figure 1). In a typical pre-Figshare workflow, a researcher sends a dataset to the repository manager who will sometimes conduct preliminary file and metadata preparation (the amount of preparation varies by institution), check compliance with any embargo requests or open access mandates, apply relevant sharing and access licenses, and then deposits the dataset in the repository. Often, there is back-and-forth between the repository manager and the researcher regarding file formatting, application of licenses, embargo timelines, and so on. The repository manager and others in the library are additionally in charge of maintaining the repository itself, setting sharing and deposit policies overall, and applies metadata standards to make the content discoverable beyond the library. This workflow is distributed across multiple digital and physical spaces.

Figshare for institutions alters this workflow by providing a platform that largely brings together the various actors, technologies, and processes involved during data management into one digital space. Once in place, repository managers can redistribute via Figshare several tasks of data management, as detailed in Figure 2 and in this section.

Repository managers use *Figshare for institutions* to create a stable ecosystem (van Dijck et al., 2018) which operates a movement of decentralization and recentralization typical of platforms (Blanke and Pybus, 2020; Gerlitz and Helmond, 2013; Helmond, 2015)—but in reverse order. First, the Figshare platform and the data repository managers (at the center of Figure 2) bring onto the same platform all the different actors involved (centralization): researchers or students (A), Figshare employees (B), web standards for discoverability (C), and data sharing and data localization management features (D). Once in place, repository managers use the platform to redistribute parts of their work to these components now linked together.

Second, *Figshare for institutions* allows repository managers to redistribute tasks to these actors (decentralization). Repository managers more effectively distribute the curatorial work to researchers themselves (A) via simpler data preparation and deposit. While there is still some back-and-forth between researcher and repository manager, Figshare

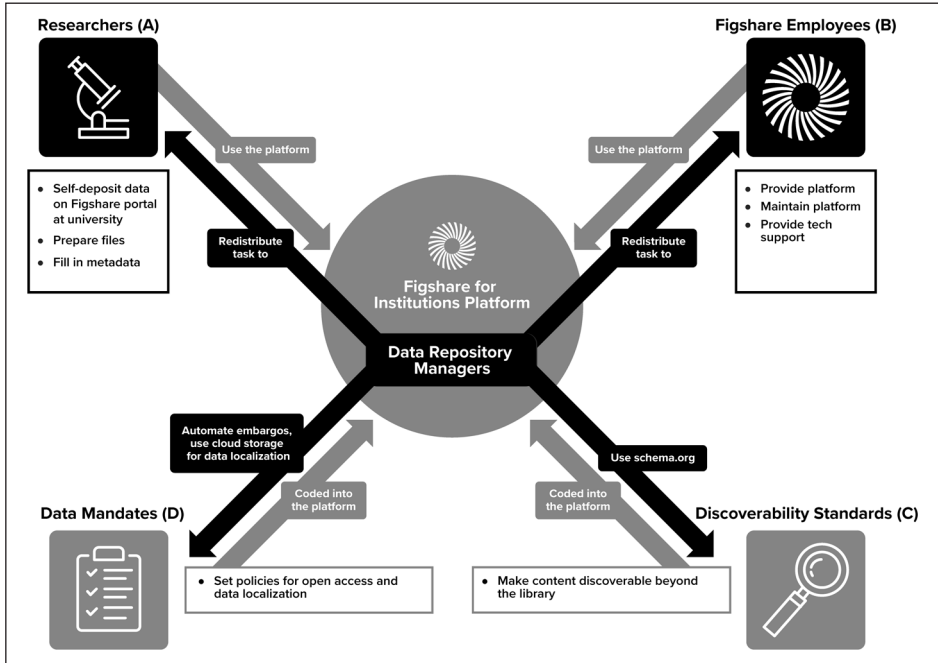


Figure 2. Data management via Figshare for institutions.

has an extremely user-friendly interface that repository managers credited for making uploads easier for researchers and therefore reducing the need for revision. The platform also distributes the technical works to Figshare employees (B) hereby reducing the work for repository managers to install and maintain the repository. *Figshare for institutions* then distributes the tasks of increasing the discoverability of researcher's outputs on web search engines or online libraries (C) by integrating web standards like schema.org by default to its platform. Finally, it streamlines the application of open access and data localization mandates (D) mostly by automating them as much as possible. We describe these four redistributions in this section.

Reducing curatorial work before submission

Participants in our study described Figshare as a time saver for repository managers as it redistributes curatorial work back to the academic depositing data or papers. Per one respondent, with Figshare:

More work [is done] by the academic who [has] actually [written] the paper, then they can decide which is the ones that they want to make open access in our system. And they can do it themselves and we [as repository managers] can just support them, rather than us actually getting the PDF off them and asking about the missing data that we need, and us uploading the work ourselves.

This reallocation of responsibilities, as the same person describes it, from “conducting” the actual submission to “supporting” researchers to do it themselves quickly raises the question of researchers’ participation: if we rely on researchers to submit their outputs, how to make sure they are actually going to do it? While researchers are generally positive toward data sharing, many obstacles prevent them from sharing data, including lack of time, trust, or training (Borgman, 2015; Curty et al., 2017; Tenopir et al., 2020).

Repository managers praise many characteristics of Figshare as potentially solving these challenges and leading to an increased researchers’ submission rate. First, they emphasize its simplicity. They compare Figshare to social media platforms: with Figshare, “all you really need to use it [are] the skills you need to use Facebook”—for example, uploading a media file, entering minimal description, all within an easy-to-use web interface. As a result, repository managers state that it is much easier for researchers to submit datasets by themselves: “very few people struggle” with this stage. Second, they appreciate that the Figshare website possesses a contemporary design, as opposed to traditional institutional repositories—which have a “library look and feel” and a “nineties sort of website feeling.” Taken together, the simplicity and modern look of Figshare are supposed to increase submission rates. These two properties are positioned in strong contrast with past or competing repositories that they used or considered adopting (such as Dspace or ContentDM), which require much more refined metadata—a preliminary step described as “pulling teeth before you can even get to a point of submitting your dataset.” Repository managers know that these characteristics of Figshare do not in themselves guarantee a high submission rate, but they feel that at least “the technology [isn’t] in the way of uploading the data set.”

Reducing technological work on the platform

Figshare reduces to the minimum the technical burden of managing a repository, either at the installation or subsequent stages. First, since Figshare is an “out of the box solution” as a repository manager put it, it reduces considerably the amount of technical know-how and skills required by repository managers to set up Figshare on top of their existing IT infrastructure. In addition, Figshare staff assist repository managers when they install the module on top of existing library infrastructures (especially regarding collection migration, metadata design, or customization of their Figshare portal). They also offer a test bed for repository managers to try the features of Figshare before committing to it, and they are generally praised for their availability and responsiveness.

After the implementation stage, the upgrade and maintenance tasks of the platform are similarly conducted by Figshare staff in a seamless way for repository managers. Consequently, this delegation of platform maintenance to Figshare saves the repository managers the constant worry of fixing the repository. As one confesses,

I can rely on the product, I don’t go to sleep thinking: “what happens if Figshare breaks tomorrow, what happens if I’m presenting to a faculty and the software doesn’t work?” That used to happen all the time with the old software, it doesn’t happen with Figshare.

Repository managers also strikingly praise Figshare for its *limited* set of features, not for its abundance. They have learnt the hard way that the more customized a repository is, the harder the update and maintenance will be later on:

[with] the previous repository, the customizations ha[d] to be reapplied every time you did a patch, and so you had a long list of changes which you had to reapply and then check that they didn't cause other problems. So, it was a little bit too customized.

Another decided during preliminary setup to keep the platform “as vanilla as possible,” changing only “some data types that go in,” for the same reason that “it’s easier for upgrades, it’s easier for everything down the line”—as opposed to what he calls the “nightmare” that upgrades of their past customized system constantly generated.

Increasing discoverability via web standards for metadata

Figshare provides a *de facto* integration to web standards for metadata, making the collection more visible on major search engines. It stores basic bibliographic metadata about each object in its repositories (e.g. title, authors, keywords, abstract) and assigns each object a DOI (Direct Object Identifier, a unique identifier commonly used for journal articles and other scholarly products); this metadata is then published via an API and harvestable by search engines via the use of *schema.org* metadata. Schema.org metadata is an industry-driven (rather than library-driven) standard developed by companies like Yahoo, Google, and Yandex. Its original purpose was to make information on websites easier to harvest in a structured way (e.g. movie times, or descriptions of products on e-commerce sites), but the standard has since been expanded to describe datasets at a high level as well.

This online visibility was seen as a major benefit by many repository managers over their prior repository systems:

[Figshare] pushing and advocating for their systemic position in that global scene, we benefit from that without having to resource that ourselves. With DSpace for example, we have to worry about whether or not Google is harvesting us, we have to worry about whether or not our materials are searchable, or whether or not our handles [a type of digital object identifier] are working.

With this industry standard, Figshare’s holdings are more reliably available in Google Search or Google Scholar. However, additional work is needed to crosswalk and share metadata via established LIS standards (we discuss this further below).

Streamlining compliance with data sharing mandates

Participants in our study described how *Figshare for institutions* eases the application of the various data mandates that repository managers must abide by. One repository manager said that the platform made it easier to manage embargo periods, that is, the period before a submitted dataset can be made public. Figshare back-office automatically manages such embargoes. Previously, this repository manager had to do this manually, with a

spreadsheet that they would revisit once a month to see which items could be taken off embargo. They appreciate that Figshare automate this part of their job. Another example concerns data sharing mandates, the requirement that researchers share any data resulting from public or foundation funding. This has required university libraries to be able to accommodate such requirements and led some to contract with Figshare:

It was the requirement from funders that we needed a system, and needed to be seen to have a system that could do all of the things that they wanted, things like the EPSRC [UK Engineering and Physical Sciences Research Council] Mandate.

Other mandates include various data localization constraints, which require domestic storage of research data. *Figshare for institutions* offers this capacity to choose the location for data storage:

That actually was another factor when we were doing the purchasing [of Figshare]. No one else really was able to guarantee us that our data would be stored in Australia. They were like, well, it could be in Australia or it could be somewhere in Europe or it could be in Singapore

Figshare does not host the data themselves; they simply mediate access to data storage via Amazon Web Services, which allows users to choose which region the data will be hosted in.

The platform trade-off: renouncing control for more ease-of-use

With its self-depositing SaaS model, adopting Figshare means that repository managers have less curatorial and technical work to do. While in some ways this is a positive outcome, we argue that reliance on third-party platforms for repository management raises concerns regarding the ability of information professionals to keep control over LIS repository solutions and discoverability standards.

Fostering more engagement with library users

Echoing scholarship on the invisibility and low status of professions caring, repairing, and maintaining technology and society (Jackson, 2014; Puig de la Bellacasa, 2012), LIS researchers have described the considerable work needed to maintain software projects (Ensmenger, 2014; Geiger et al., 2021) or to clean, curate, repair, and otherwise prepare data before it can be shared (Plantin, 2019; Borgman, 2015; Pink et al., 2018; Scroggins and Pasquetto, 2020). These authors also emphasize the low desirability of these jobs. Similarly, the repository managers we spoke with do not enjoy the technical tasks inherent to managing a repository, and they praise Figshare for delegating them to others or removing them altogether, as we described in the previous section.

Consequently, the platform frees up time for repository managers to engage in “tasks that are a bit more valuable to [them].” They use this spare time to present the larger value of sharing and reusing data for knowledge:

[With Figshare] the energy shifted from the IT side of things to very much where the library needed to be, which was promoting and hopefully trying to change the culture of the research staff toward research data management, towards storage and publication and preservation of their data.

In the words of another repository manager, Figshare makes their job less about data preparation or IT management and “more engagement focused.” Relatedly, *Figshare for institutions* allows repository managers to engage in more concrete conversations about data sharing, especially regarding FAIR data initiatives on campus:⁴

We always try to include conversations about how to upload and what metadata, and then we try and talk about metadata for impact and for discovery and reuse. And our conversations are much more around FAIR data and those sorts of things, rather than just give me your file, we'll upload it for you.

Since Figshare provides the infrastructure for data sharing at their university, it substantiates repository managers' effort to support data sharing. With Figshare, data sharing “is not a hollow conversation. It's not: ‘Oh, we'd love you to be more FAIR.’ [It is instead:] ‘Actually we have a possibility to help you.’” However, while this is described as a positive outcome for repository managers, along this shift to hosted solutions comes the risk of losing of control over the solutions and standards they use, as we describe below.

Promoting (paid) commercial versus (free) open-source solutions

We found during our interviews that the choice between LIS-driven solutions or Figshare was not between open-source or proprietary software. In fact, repository managers value the freedom that open-source solutions bring: “I do like the idea of open-source that you can just work on and customize and program to make it your own.” However, the choice of solutions simply came down to a question of budget:

Can I be honest with you? At this stage in time, I must say we are concerned about cutting costs, but we were trying to find a tool that could help us get the work done to our satisfaction. Because we believe that the important things to offer to ensure that we achieve the goals, we support the mandate of the university in as much way possible.

While open-source solutions might be free, they generate a significant increase of labor (and labor cost) for libraries. They require substantial technical skill to set up and maintain, such as in-house staff to set up servers, to install idiosyncratic software, and to sometimes develop graphical user interfaces. In many cases, they require extensive customization via software development to deploy within a particular institution. Participants in our study know from experience that open-source solutions like Fedora “take forever to install and then forever to troubleshoot,” in addition to being particularly “buggy.”

For example, one repository manager mentions how their past repository based on a Fedora solution “was falling over on itself, essentially” due to accumulated problems—mostly:

[It took an] awful lot of time trying to keep that repository up and going, like just keeping it running, essentially, so it could be accessed [. . .]. It was becoming quite old and we'd customized it a lot and that had become a bit of a problem.

None of these characteristics are surprising to users of open-source software, which are known to have high maintenance costs (Geiger et al., 2021). However, the problem emerges when institutions, such as the libraries studied here, do not have the resources or staff to do this maintenance work:

If we had chosen an open-source repository, there'd be a lot of work that would have had to be done to that repository to customize it for our own purposes, and there's not really that resource of people with those skills available in the library to do that, unfortunately.

Other teams may have the necessary skills but are simply stretched too thin to be able to allocate someone to open-source software maintenance. A repository manager illustrates this: "research data management is one full-time position" and this person cannot take yet another demanding task of platform maintenance.

Open-source solutions come with a higher operation cost and also rely on the repository managers' skills to solve a problem when it emerges, usually after lengthy research on community forums. Figshare, on the opposite, comes with a support service included in the subscription price. This capacity to have someone to contact when needed is for many repository managers the main reason to choose Figshare over an open-source solution:

We could have gone for something like DSpace for our institutional repository which is open-source. We definitely didn't want to go that route and so paying for our institutional repository, it's worth triple [. . .]. And the reasons behind that is you have all this support structure that goes with it. And Figshare has delivered on that support structure.

The choice between proprietary and open-source solutions unfolds as a paradox which reproduces a financial divide between university libraries. While well-resourced university libraries can hire or train dedicated staff to maintain (free) open-source solutions to power their repositories, modest ones subscribe to one or several (paid) solutions like Figshare to lower the operational cost of running a repository.

Privileging industry standards for discoverability

As we reviewed above, Figshare publishes data to the web using the metadata standard schema.org, which was initially built to support generic web search. Internally, Figshare's metadata fields are customizable to include, for example, a deposit agreement and description of files and folders. Figshare also makes it possible to export data in library standard formats such as Dublin Core.⁵ However, several of our respondents were unaware of this functionality, or said they wished that integration with LIS standards was easier:

We're used to having standards like Dublin Core . . . But in Figshare, you'd have to say, I'm going to make a custom field for volume, or I'm going to make a custom field for the journal title. So . . . [laughing], we did have to look at what we expected our publications to have, and

also try to start to see into the future about what changes Figshare might make in that area and match our data with them. I wish that they would just qualify Dublin Core and make everyone's life easy and you wouldn't have to decide any of this stuff.

By making a repository's holding findable via web standards first, but library pathways second, Figshare subtly nudges repository holdings into Google's ecosystem. This might make digital objects easier to find, but at the cost of some control over the mechanisms for discoverability. This capacity of search engines to make their infrastructure unavoidable represents a crucial aspect of the power of tech giants over data curation.

Discussion: impacts of the platformization of institutional repositories in university libraries

Prior work on the platformization of knowledge infrastructures has speculated that the introduction of private platforms into public scholarship may risk "turning over critical scientific functions to private firms" and may "contribute to the further 'splintering' of knowledge infrastructures" (Plantin et al., 2018b). Our work here shows that transition in action. Our respondents choose (and praise) Figshare because it fills a gap in their workforces and budgets and allows users to focus on the person-to-person interactions that many enjoy. Yet, though Figshare has made great efforts toward backwards compatibility with existing library metadata and sharing mechanisms, the use of schema.org paves the way for other platforms—in this case, Google's platform for scholarly search and discovery—to further splinter scholarly discovery away from scholarly communities.

Like all standards, schema.org "represent[s] a form of control over technology" (Abbate, 2000: 147) that reflects the worldview of its makers (Svenonius, 2000) and fundamentally shapes how we encounter information on the web (Iliadis et al., 2023). In the case of Figshare, the use of schema.org makes it much easier to encounter scholarly output via generic search engines—but specifically, Google, which has put the most effort into scholarly search (via Google Scholar and the still nascent Google Dataset search). The dominance of Google Scholar already has shifted citation patterns and discovery norms in academia, and in ways that the scholarly community has little ability to influence (Portenoy et al., 2022; West, 2019). By sharing repository holdings via schema.org terms, Figshare provides Google with the critical mass of data it needs to monopolize (and monetize) scholarly search. Academic libraries have already had to reckon with the consequences of the privatization and platformization of their publishing infrastructure, in that many are now beholden to increasingly expensive and increasingly restrictive "Big Deal" contracts with commercial publishing companies (Aiwuyor, 2020); the platformization of repository infrastructure risks repeating this recent history.

Deskilling university librarians

In addition to moving scholarly infrastructure closer to tech giants like Google, the adoption of hosted solutions like Figshare brings a potential deskilling of repository managers and librarians—or, at the very least, a problematic shifting of their skills. One hope underlying the development of technical solutions and standards is that by shoring up

technical capacity in libraries (and technical skills in LIS graduates), the field of librarianship would further defend its professional status against the reduction of library funding and the replacement of library staff with entry-level workers (Marshall et al., 2009; Witt, 2008). The adoption of hosted platforms like Figshare has the potential to reverse some of this work, that is, to make library jobs *less* technical and librarians potentially lose key skills, for example, “key information and preservation techniques and methods” or “expertise on effective product valuation” (ALA Outsourcing Task Force, 1999: 9).

Many of our respondents were relieved to not have to learn new technical skills and used Figshare to get back to the social aspects of their work that they preferred. But while working with patrons or supporting information access are cornerstones of librarianship, they are not valued as much as technical skills in the labor market. Librarianship has long been considered a “pink-collared” job (Nicholson, 2019): librarians are first in line to deal with users’ expectations, emotions, and frustrations (Kaun and Forsman, 2022). Just like scientific discoveries rely on “invisible technicians” (Shapin, 1989), libraries rely on the largely unacknowledged labor (Thomer et al., 2022) of an often racialized, gendered, and precarious workforce (Chalmers, 2019) that provides the necessary digital care work and trouble-shooting for these projects to succeed (Kaun and Forsman, 2022; Ringel and Ribak, 2021). LIS workers have at times defended their professional status by going against this invisibility, for example, by pointing to the skilled technical labor that goes into information organization and description (Sloniowski, 2016). Digital curation and preservation work are just the most recent versions of this. By removing librarians’ opportunities to engage in this technical work, platforms may also remove their ability to defend their status as a technical, professional—and more highly paid—field.

Perpetuating austerity at university libraries

By providing an affordable way to run a data repository, *Figshare for institutions* can either be seen as a solution to the defunding of libraries or as a justification to cut budgets further. Mark Hahnel positions his product in a context where “cash-strapped libraries” need to look for new business models for scholarly publishing—such as Figshare (Hahnel, 2012). He emphasizes that “Figshare primarily sells its platform to universities ‘for less than the price of a full-time employee’” (reported by Brinded, 2015).

Hosted solutions appear in a context of systemic precarity for university libraries. Focusing on US numbers, this has been a long-term trend: library resources peaked in 1974, with 3.83% of university budgets, down to 1.95% in 2009, that year marking the 14th straight year of budget decrease (Kolowich, 2012). This trend was recently exacerbated by university reorganization following the Covid-19 pandemic (McKenzie, 2020). Library faculty lines are also being eliminated, the working population is aging out and either not being replaced, or with non-permanent workers (American Library Association, 2019; Henninger et al., 2019).

While looking for alternative ways to conduct some missions at a lower cost seems logical in such a context, the American Library Association has been raising serious concerns about outsourcing in libraries as early as 1999 (ALA Outsourcing Task Force, 1999). Specifically, it warns against outsourcing core library services as it transforms them into a “simple commodity that can be quantified [. . .] and contracted to the lowest bidder” hereby obliterating

that “much of the important work of librarianship is abstract and non-quantifiable” (p. 8). Long before Figshare, libraries have already moved toward SaaS modules, for instance, via cloud-based catalogs, curation platforms, e-readers, and numerous systems for discovery (Mavodza, 2013). Figshare constitutes another step in this direction.

Conclusion

Figshare for institutions leverages the properties of platforms to create an ecosystem linking together separated components of the library and data repository, such as researchers, technical staff, web standards, and data mandates. This model reduces the curatorial and technical tasks for repository managers by redistributing or automating them, and since it is a part of their job they value less, they welcome Figshare for improving their work conditions.

However, while this is in many ways a positive outcome, this platformization of data management brings several risks. It can undermine efforts of the LIS community to develop their own solutions, technical frameworks, and skillsets. The prioritization of web standards for discoverability via schema.org increases the short-term visibility of research outputs, but as opposed to LIS-based standards, represents a loss of control over visibility of research outputs. This shift takes part in a general trend of deskilling library staff, leaving them only with “softer” skills. Finally, by making data management “cheap” in terms of funding and staff required, platforms like Figshare put this library mission—and repository managers positions—at the bottom of budget priorities. Making the scarcity of resources more manageable transforms a situation of precarity into a normal state of functioning for knowledge institutions.

Libraries, repository managers—or even Figshare—are not to blame for the shifts in power dynamics and work arrangements described above. The dynamics described above are reflective of larger societal and organizational forces rather than individual library or repository manager’s choices. Figshare is not by any means the only commercial or SaaS research data management platform. For instance, Esploro by ExLibris is part of Clarivate, a multi-billion dollar, publicly traded company that has been criticized by open access advocate as exercising an “effective monopoly” on library systems (SPARC, 2021). It is important to think about ways that libraries could reap the benefits of platforms without worsening precarity politics or incentivizing the deskilling of the field. One avenue could be for libraries to develop standards to guide the adoption of platforms—whether for data management, curation, or other purposes. Participants in this study were of frameworks for repository design and assessment. However, they felt that they did not suit their institution or simply did not apply to the architecture of Figshare. New guidelines designed for a platformized library could help repository managers communicate their needs and priorities to companies like Figshare, while also asserting some of their design-making control and technical power when working with for-profit organizations.

Authors’ Note

Andrea Thomer is now affiliated to University of Arizona, US.

Acknowledgment

The authors would like to thank the editor of *New Media & Society* and the reviewers for their productive feedback on earlier versions of the article. They are also very grateful to all the respondents who time to be part of this research. Additional thanks go to Alexandria Rayburn and Joseph Starks for their work during data collection, as well as Chris Kette for his work on the figures. Any errors or omissions remain our responsibility.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Jean-Christophe Plantin's research and authorship of this article was supported with funding by the Department of Media and Communications at LSE. Andrea Thomer's work was funded in part by IMLS grant # RE-07-18-0118-18.

ORCID iD

Jean-Christophe Plantin  <https://orcid.org/0000-0001-8041-6679>

Notes

1. Examples of features include a customized Figshare portal at a custom URL; 100TB of Amazon Web Services S3 cloud storage; custom branded DOI minting capacities; dashboards to report on the module's activity; administrative features for data management, such as custom metadata creation or moderation of submission; and access to the Figshare API.
2. *Figshare for institutions* is meant to function as a data repository, which stores research data, an institutional repository, which stores publications, or both. A similar product is sold to academic publishers as *Figshare for publishers*.
3. Data collected in January 2021. "Who we work with," Figshare website: <https://knowledge.figshare.com/type-of-client/institutions>.
4. "FAIR" are guidelines for data management based on four principles: Findability, Accessibility, Interoperability, and Reusability.
5. See https://docs.figshare.com/#oai_pmh; Dublin Core is a metadata standard developed by a collective of librarians, digital archivists, and digital curators and is meant to describe both born-digital and physical collections in libraries, archives, and museums. We note that since the time of our interviews, Figshare has updated their website to include more FAQs about the integration of library and information science (LIS) standards like Dublin Core or OAI-PMH (a protocol for metadata harvesting developed by early digital library developers); see <https://web.archive.org/web/20210510170122/https://help.figshare.com/admin-faq>.

References

- Abbate J (2000) *Inventing the Internet*. 58839th ed. Cambridge, MA: The MIT Press.
- Aiwuyor J (2020) *More Research Libraries Decline "Big Deal" Subscription Contracts with Publishers*. Association of Research Libraries. Available at: <https://www.arl.org/news/more-research-libraries-decline-big-deal-subscription-contracts-with-publishers/> (accessed 27 January 2023).
- ALA Outsourcing Task Force (1999) Outsourcing and privatization in American libraries. Available at: <https://www.ala.org/tools/outsourcing/background> (accessed 17 August 2022).
- American Library Association (2019) The State of America's Libraries 2019. *American Libraries*, April. Available at: <https://www.ala.org/news/sites/ala.org.news/files/content/2019-soal-report-final-accessible.pdf>

- Amorim RC, Castro JA, Rocha da Silva J, et al. (2017) A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society* 16(4): 851–862.
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533(7604): 452–454.
- Baldwin CY and Woodard CJ (2008) *The Architecture of Platforms: A Unified View*. Harvard Business School. Available at: <http://hbswk.hbs.edu/item/6025.html> (accessed 12 June 2015).
- Bill & Melinda Gates Foundation (n.d.) *Bill & Melinda Gates Foundation Open Access Policy*. Bill & Melinda Gates Foundation. Available at: <https://www.gatesfoundation.org/about/policies-and-resources/open-access-policy>
- Bishop B, Orehek A, Eaker C, et al. (2022) Data services librarians' responsibilities and perspectives on research data management. *Journal of EScience Librarianship* 11(1): 4.
- Blanke T and Pybus J (2020) The material conditions of platforms: monopolization through decentralization. *Social Media + Society* 6(4): 2056305120971632.
- Borgman CL (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: The MIT Press.
- Braun V and Clarke V (2021) *Thematic Analysis: A Practical Guide*. 1st ed. London; Thousand Oaks, CA: Sage.
- Brinded L (2015) Figshare CEO Mark Hahnel: drug discovery, open data, academia, coding. Available at: <https://www.businessinsider.com/figshare-ceo-mark-hahnel-drug-discovery-open-data-academia-coding-2015-9?r=UK> (accessed 1 February 2019).
- Chalmers M (2019) *On a mission to scan: visibility, value(s), and labor in large-scale digitization*. PhD Dissertation, University of Michigan. Available at: <https://deepblue.lib.umich.edu/handle/2027.42/153473> (accessed 27 August 2020).
- Crosas M (2011) The Dataverse Network®: an open-source application for sharing, discovering and preserving data. *D-Lib Magazine* 17(1/2). Available at: <https://doi.org/10.1045/january2011-crosas>
- Curdy RG, Crowston K, Specht A, et al. (2017) Attitudes and norms affecting scientists' data reuse. *PLOS ONE* 12(12): e0189288.
- Ensmenger N (2014) When good software goes bad. The surprising durability of an ephemeral technology. In: *MICE (Mistakes, Ignorance, Contingency, and Error) conference*, Munich. Available at: <http://homes.soic.indiana.edu/nensmeng/files/ensmenger-mice.pdf>
- Faundeen JL, Burley TE, Carlino JA, et al. (2013) *The United States Geological Survey science data lifecycle model*. U.S. Geological Survey Open-File Report 2013-1265. Reston, VA: U.S. Geological Survey.
- Geiger SR, Howard D and Irani L (2021) The labor of maintaining and scaling free and open-source software projects. In: *Proceedings of the ACM on human-computer interaction*, April, pp. 1–28. Available at: <https://dl-acm-org.gate3.library.lse.ac.uk/doi/10.1145/3449249> (accessed 2 March 2022).
- Gerlitz C and Helmond A (2013) The like economy: social buttons and the data-intensive web. *New Media & Society* 15: 1348–1365.
- Gillespie T (2010) The politics of “platforms.” *New Media & Society* 12(3): 347–364.
- Gilliland A (2016) Setting the stage. In: *Introduction to Metadata*. Getty Institute, pp. 1–19. Available at: <https://www.getty.edu/publications/intrometadata/setting-the-stage/>
- Hahnel M (2012) Exclusive: figshare a new open data project that wants to change the future of scholarly publishing. In: *LSE Impact of Social Sciences*. Available at: <https://blogs.lse.ac.uk/impactofsocialsciences/2012/01/18/can-we-do-better-with-scholarly-publishing/> (accessed 17 August 2022).
- Helmond A (2015) The platformization of the web: making web data platform ready. *Social Media + Society* 1(2): 1–11.

- Henninger E, Brons A, Riley C, et al. (2019) Perceptions and experiences of precarious employment in Canadian libraries: an exploratory study. *Partnership: The Canadian Journal of Library and Information Practice and Research* 14(2): 2.
- Hey T, Tolle K and Tansley S (2009) *The Fourth Paradigm Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research.
- Higgins S (2008) The DCC curation lifecycle model. In: *Proceedings of the 8th ACM/IEEE-CS joint conference on digital libraries—JCDL '08*, Pittsburgh, PA, 16–20 June.
- Iliadis A, Acker A, Stevens W, et al. (2023) One schema to rule them all: how Schema.org models the world of search. *Journal of the Association for Information Science and Technology*. Epub ahead of print 24 February. DOI: 10.1002/asi.24744.
- Jackson S (2014) Rethinking repair. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. 1st ed. Cambridge, MA: The MIT Press, pp. 221–240.
- Jeanneney J-N (2008) *Google and the Myth of Universal Knowledge: A View from Europe*. Chicago, IL: University of Chicago Press.
- Kaun A and Forsman M (2022) Digital care work at public libraries: making digital first possible. *New Media & Society*. Epub ahead of print 23 June. DOI: 10.1177/14614448221104234.
- Kolowich S (2012) Smaller servings for libraries. Available at: <https://www.insidehighered.com/news/2012/02/21/library-budgets-continue-shrink-relative-university-spending> (accessed 28 January 2022).
- Kraker P, Lex E, Gorraiz J, et al. (2015) Research data explored II: the anatomy and reception of figshare. arXiv:1503.01298 [cs]. Available at: <http://arxiv.org/abs/1503.01298> (accessed 1 March 2022).
- Lynch C (2003) *Institutional Repositories, Infrastructure for Scholarship*. CNI: Coalition for Networked Information. Available at: <https://www.cni.org/publications/cliffs-pubs/institutional-repositories-infrastructure-for-scholarship> (accessed 27 January 2023).
- Lynch C (2008) How do your data grow? *Nature* 455(7209): 28–29.
- Lynch CA and Lippincott JK (2005) Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine* 11(9). Available at: <https://doi.org/10.1045/september2005-lynch>
- McDonough J (2009) XML, interoperability and the social construction of markup languages: the library example. *Digital Humanities Quarterly* 3(3). Available at: <http://www.digitalhumanities.org/dhq/vol/3/3/000064/000064.html>
- McKelvey F (2011) A programmable platform? Drupal, modularity, and the future of the web. *Fibreculture Journal* 18. Available at: <http://eighteen.fibreculturejournal.org/2011/10/09/fcj-128-programmable-platform-drupal-modularity-and-the-future-of-the-web/> (accessed 6 April 2015).
- McKenzie L (2020) Library leaders brace for budget cuts. Available at: <https://www.insidehighered.com/news/2020/12/09/university-library-leaders-prepare-uncertain-financial-future-amid-pandemic> (accessed 28 January 2022).
- Mao Y (2017) Intercoder reliability techniques: Holsti method. In: Allen M (eds) *The Sage Encyclopedia of Communication Research Methods*. Sage, pp. 741–743.
- Maron D and Feinberg M (2018) What does it mean to adopt a metadata standard? A case study of Omeka and the Dublin Core. *Journal of Documentation* 74(4): 674–691.
- Marshall JG, Marshall VW, Morgan JC, et al. (2009) Where are they now? Results of a career survey of library and information science graduates. *Library Trends* 58(2): 141–154.
- Mavodza J (2013) The impact of cloud computing on the future of academic library practices and services. *New Library World* 114(3/4): 132–141.

- Millerand F and Bowker G (2009) Trajectories and enactment in the life of an ontology. In: Lampland M and Star SL (eds) *Standards and Their Stories*. Ithaca, NY: Cornell University Press, pp. 149–165.
- Nature (2023) Reporting standards and availability of data, materials, code and protocols. *Nature Editorial Policies*. Available at: <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>
- Nelson A (2013) *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research [Memorandum]*. Office of Science and Technology Policy. Available at: <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>
- Nicholson KP (2019) “Being in time”: new public management, academic librarians, and the temporal labor of pink-collar public service work. *Library Trends* 68(2): 130–152.
- Pink S, Ruckenstein M, Willim R, et al. (2018) Broken data: conceptualising data in an emerging world. *Big Data & Society* 5(1): 2053951717753228.
- Plantin J-C (2019) Data cleaners for pristine datasets: Visibility and invisibility of data processors in social science. *Science, Technology, & Human Values* 44(1): 52–73.
- Plantin J-C, Lagoze C, Edwards PN, et al. (2018a) Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society* 20(1): 293–310.
- Plantin J-C, Lagoze C and Edwards PN (2018b) Re-integrating scholarly infrastructure: the ambiguous role of data-sharing platforms. *Big Data & Society* 5(1): 1–14.
- Portenoy J, Radensky M, West JD, et al. (2022) Bursting scientific filter bubbles: boosting innovation via novel author discovery. In: *CHI conference on human factors in computing systems*, New Orleans, LA, 29 April 2022, pp. 1–13. New York: ACM.
- Puig de la Bellacasa M (2012) “Nothing comes without its world”: thinking with care. *The Sociological Review* 60(2): 197–216.
- Ringel S and Ribak R (2021) “Place a book and walk away”: archival digitization as a socio-technical practice. *Information, Communication & Society* 24: 2293–2306.
- Sandy HM and Freeland C (2016) The importance of interoperability: lessons from the Digital Public Library of America. *International Information & Library Review* 48(1): 45–50.
- Sayre F and Riegelman A (2018) The reproducibility crisis and academic libraries. *Sayre*. Available at: <https://doi.org/10.5860/crl.79.1.2>
- Scroggins MJ and Pasquetto IV (2020) Labor out of place: on the varieties and valences of (in)visible labor in data-intensive science. *Engaging Science, Technology, and Society* 6: 111–132.
- Shapin S (1989) The invisible technician. *American Scientist* 77(6): 554–563.
- Sloniowski L (2016) Affective labor, resistance, and the academic librarian. *Library Trends* 64(4): 645–666.
- SPARC (2021) *SPARC Statement on Completion of Clarivate-ProQuest Merger*. SPARC, 2 December. Available at: <https://sparcopen.org/news/2021/sparc-statement-on-completion-of-clarivate-proquest-merger/>
- Svenonius E (2000) *The Intellectual Foundation of Information Organization*. Cambridge, MA: The MIT Press.
- Tenopir C, Rice NM, Allard S, et al. (2020) Data sharing, management, use, and reuse: practices and perceptions of scientists worldwide. *PLOS ONE* 15(3): e0229003.
- Thelwall M and Kousha K (2016) Figshare: a universal repository for academic resource sharing? *Online Information Review* 40(3): 333–346.
- Thomer AK, Akmon D, York JJ, et al. (2022) The craft and coordination of data curation: Complicating workflow views of data science. *Proceedings of the ACM on Human-Computer Interaction* 6(CSCW2): 1–29.
- Thystrup NB (2019) *The Politics of Mass Digitization*. Cambridge, MA: The MIT Press.

- Vaidhyanathan S (2011) *The Googlization of Everything*. 1st ed. Berkeley, CA: University of California Press.
- van Dijck J, Poell T and de Waal M (2018) *The Platform Society*. New York: Oxford University Press.
- West J (2019) Echo chambers in science? Keynote speech, Stanford, CA. Available at: <https://www.metascience2019.org/presentations/jevin-west/> (accessed 16 March 2023).
- Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1): 160018.
- Winn J (2013) Open data and the academy: an evaluation of CKAN for research data management. In: *IASSIST2013*, Cologne. Available at: <http://eprints.lincoln.ac.uk/id/eprint/9778/>
- Witt M (2008) Institutional repositories and research data curation in a distributed environment. *Library Trends* 57(2): 191–201.
- Zeng ML and Chan LM (2006) Metadata interoperability and standardization—a study of methodology part II: achieving interoperability at the record and repository levels. *D-Lib Magazine* 12(6). Available at: <https://www.dlib.org/dlib/june06/zeng/06zeng.html>
- Zimmer M (2015) The Twitter archive at the Library of Congress: challenges for information practice and information policy. *First Monday*. Available at: <https://firstmonday.org/ojs/index.php/fm/article/view/5619/4653>

Author biographies

Jean-Christophe Plantin is associate professor at the department of Media & Communications, LSE. In his research, he investigates the hidden work of data processors (or “data cleaners”) and their contribution to knowledge production; the social consequences of the platformization of several sectors (data sharing, cybersecurity, cloud gaming); and the increasing power that tech giants gain over global communications networks.

Andrea Thomer is an assistant professor at the University of Arizona School of Information. She conducts interdisciplinary research in the areas of scientific data practices and data(base) curation; the maintenance and evolution of knowledge infrastructures: metadata and the conceptual foundations of information organization.