

LSE

THE LONDON SCHOOL  
OF ECONOMICS AND  
POLITICAL SCIENCE ■

# LSE Research Online

## Jean-Christophe Plantin, Carl Lagoze and Paul N. Edwards Re-integrating scholarly infrastructure: the ambiguous role of data sharing platforms

**Article (Published version)  
(Refereed)**

**Original citation:** Plantin, Jean-Christophe and Lagoze, Carl and Edwards, Paul N. (2018) *Re-integrating scholarly infrastructure: the ambiguous role of data sharing platforms*. [Big Data and Society](#), 5 (1). p. 205395171875668. ISSN 2053-9517  
DOI: [10.1177/2053951718756683](https://doi.org/10.1177/2053951718756683)

Reuse of this item is permitted through licensing under the Creative Commons:

© 2018 The Authors  
CC-BY-NC-ND

This version available at: <http://eprints.lse.ac.uk/87289/>  
Available in LSE Research Online: March 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

# Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms

Big Data & Society  
January–June 2018: 1–14  
© The Author(s) 2018  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2053951718756683  
journals.sagepub.com/home/bds



Jean-Christophe Plantin<sup>1</sup>, Carl Lagoze<sup>2</sup> and Paul N Edwards<sup>3</sup>

## Abstract

Web-based platforms play an increasingly important role in managing and sharing research data of all types and sizes. This article presents a case study of the data storage, sharing, and management platform Figshare. We argue that such platforms are displacing and reconfiguring the infrastructure of norms, technologies, and institutions that underlies traditional scholarly communication. Using a theoretical framework that combines infrastructure studies with platform studies, we show that Figshare leverages the platform logic of core and complementary components to re-integrate a presently splintered scholarly infrastructure. By means of this logic, platforms may provide the path to bring data inside a scholarly communication system still optimized mainly for text publications. Yet the platform strategy also risks turning over critical scientific functions to private firms whose longevity, openness, and corporate goals remain uncertain. It may amplify the existing trend of splintering infrastructures, with attendant effects on equity of service.

## Keywords

Application programming interface, Big Data, data sharing, decentralization, infrastructure, integration, library, platform, scholarship

## Introduction

The advent of “Big Data” in scholarship has provoked debate about the paradigm changes—or lack thereof—resulting from increases in the size and number of datasets available for research (Anderson, 2008; Hey et al., 2009; Kitchin, 2014a). Another unexpected and perhaps more fruitful consequence has been to reveal the systemic under-theorisation of the definitions, roles, and status of data *per se*, as opposed to those of concepts (e.g. objectivity, cf. Galison and Daston, 2007), methods (e.g. statistics, cf. Hacking, 1990), or disciplines (Knorr-Cetina, 1999).

Recent works compensate this lack by providing social studies of the status of data in scholarship. The meaning of “data” and its role in the research process is not fixed and has shifted many times in history (Rosenberg, 2013). “Big Data” is better conceptualized not as large quantities of data but as interrelated “data assemblages” (Kitchin, 2014b), revealing the interrelated practices, organization and institutions, or systems of thought involved in the circulation of

scientific data. The “bigness” of data is itself dependent on disciplinary practices, and changes widely across types of science (Borgman, 2015). Data does not flow automatically, but requires careful “packaging” (Leonelli, 2016) before it can travel across sites of production and reuse (Edwards, 2010). We contribute to this body of work by analysing the entities responsible for the organization of scientific data sharing. We show how the rise of large datasets, in conjunction with a rising interest in data as scholarly output, contributes to the advent of data sharing platforms in a field traditionally organized by infrastructures. After detailing

<sup>1</sup>London School of Economics and Political Science, UK

<sup>2</sup>University of Michigan, USA

<sup>3</sup>Stanford University, USA; University of Michigan, USA

## Corresponding author:

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

Email: [j.plantin1@lse.ac.uk](mailto:j.plantin1@lse.ac.uk)



the relationship between these two configurations, we describe what this relation means for scholarship.

As an empirical case, we consider the data storage, sharing, and management platform Figshare, a company created in 2011 by Dr Mark Hahnel, a PhD graduate from Imperial College London. Figshare is both a website and a technology: on the one hand, it invites individual researchers to self-archive their outputs (including datasets, graphics, presentation slides, and almost anything else) on *figshare.com*. On the other, it is marketed to university libraries or scientific publishers as a middleware service (respectively *Figshare for institutions* and *Figshare for publishers*), providing a suite of features (web portals, data management tools, persistent identifiers) installed “on top of” existing infrastructural components.

This case study reveals how platforms, such as Figshare, insert themselves within infrastructures to organize data sharing. We first show that platforms emerge in the context of a *de-integration* of the existing infrastructures that traditionally organized scholarship. Unlike print publications, data as scholarly output has never been integrated into a clearly defined chain of norms and processes. Widespread demands for data deposit and citation, now much more feasible with web technology, have provoked new intermediaries to take over this function. In a second theme, we show how, in this context, platforms leverage these two trends to position themselves as assuming this integrating role. Figshare relies on the decentralized nature of platforms (Helmond, 2015) and their application programming interface (API)-based architecture to link scattered components of the scholarly infrastructure (researchers, libraries, journals, etc.), seeking to effectively integrate data in this chain of processes and actors. After detailing this ecosystem, we show in a third theme that such integration comes with the risk of platforms becoming a new centralizing entity in scholarship, but also of further splintering knowledge infrastructures.

To reach these results, we rely on a theoretical framework that combines *infrastructure studies* with *platform studies* (Plantin et al., 2018) and captures the complexity of this hybrid configuration for data sharing. On the one hand, scholarship in sociology of science and science, technology & society has shown how infrastructures organize the circulation of knowledge in society (Edwards, 2010; Borgman, 2007; Bowker et al., 2010). This first perspective describes scholarly infrastructures as a chain of interrelated actors, such as universities, academic publishers, data archives, and libraries, each of which serves a dedicated function. This lens is also helpful to describe the current dynamics of de-integration of scholarly infrastructure that facilitates the rise of digital platforms. Platforms are

defined by an architecture allowing programmability and reuse of content and data (van Dijck and Poell, 2013), typically through an API (Helmond, 2015), and organizing modularity between a stable core and variable complementary components (Baldwin and Woodward, 2008). By theorizing the migration of such natively digital objects (Rogers, 2009) in the social world of scientific research—as opposed to platforms endogenous to the scientific world (Keating and Cambrosio, 2006)—this second body of work allows us to show how platforms leverage these characteristics to re-integrate a splintering scholarly infrastructure, as well as to critically assess the risks for scholarship that emerge from this hybridity.

### Scholarly infrastructure in a “Big Data” age

Two recent transformations in the scholarly communication infrastructure have set the stage for the rapid development of data-sharing platforms. First, the increasing interest in data as scholarly output challenges the ability of scholarly infrastructures to fulfill their multiple functions (validity, certification, dissemination, etc.) in the context of new scholarly objects. Second, the online environment has spawned a series of web-based entities (e.g. e-print archives and institutional repositories) that duplicate and sometimes extend functions typical of existing actors in the scholarly infrastructure. After discussing these two transformations, we describe how they have facilitated the rise of new intermediaries, such as Figshare.

#### *The traditional organization of scholarly infrastructure*

Edwards et al. (2013) defined knowledge infrastructures as “robust networks of people, artefacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds.” The infrastructure behind the publication of scholarly outputs (such as journals and books) constitutes a quintessential example of the “share and maintain” aspects of this definition. Ever since the creation of the first scholarly journal in 1665,<sup>1</sup> scholarly communication has required four critical functions: *registration*, providing authors with the means to claim precedence; *certification*, establishing the legitimacy of a result (e.g. through peer review); *awareness*, enabling the community of scholars to access a result; and *preservation*, ensuring the longevity of the scholarly record (Roosendaal and Geurt, 1997).

Pre-Internet scholarly communication carried out these functions through a synergistic collaboration among three distinct categories of actors.

*Authors* created content (publications) and forwarded them to *publishers*, which performed the registration, certification, and awareness functions. *Libraries* subscribed to journals and collected books, preserving these materials and providing access. Over time, these actors formed an integrated infrastructure with well-established norms, routines, technological systems, shared practices, and differentiated social roles. The continuity and stability of these well-functioning systems ensured the confidence and trust of their participants (Borgman, 2007).

### *The rise of data as scholarly output*

For over 300 years, this infrastructure developed with print publications at its center. Across this period, complete research datasets were rarely formally published. Yet scholarship has changed dramatically in recent decades—especially since the 2000s—due to the increasing scale, role, and status of digital data. Both “pull” and “push” forces have transformed the role of data in scholarship in recent years.

**“Big Data” across disciplines.** Changes in the technical architecture for data manipulation and storage—including high-capacity storage systems, high-speed networks to easily move large datasets back and forth, and MapReduce algorithms for parallel computing—now permit a wider range of disciplines to engage with data-driven research. The focus on data-at-scale is well established in “big science,” as exemplified by high-energy physics (Knorr-Cetina, 1999), astronomy (Borgman, 2007), and biology (Leonelli, 2016). Similar trends are also emerging in the social sciences and humanities (e.g. computational social science, cf Lazer et al., 2009, or the many digital humanities projects and centres). In addition to the more central role data play in scholarship, the diversification of data sources is also a defining trait of this new context (Kitchin, 2014b). Traditional modes of data production (e.g. surveys in social science) have been complemented by new modes of data collection, such as citizen science (Conrad and Hilchey, 2010) and networked sensor grids (Wallis et al., 2007). Institutional actors such as government agencies and university libraries now routinely provide open access to many kinds of data. Meanwhile, social media such as Twitter, Facebook, and YouTube can themselves be constitute vast datasets which can be mined or treated as proxies for variables of interest (yet with limitations, as shown by boyd and Crawford, 2012; Lazer et al., 2009).

**New data sharing requirements.** In recent years, researchers have faced both pressures and incentives to share

the data they use or produce. These have been motivated by four reasons: to facilitate reproducibility of research, to make publicly-funded assets available to the public, to leverage investments in research data, and to advance research and innovation (Borgman, 2013). Funding agencies now often require data sharing and data management plans: the US National Institutes of Health (NIH) initiated these policies in 2003 for grants over \$500,000, while the US National Science Foundation (NSF) established data sharing requirements in 2010 (Borgman, 2012). Such requirements provide new opportunities to expand existing repositories and develop innovative archiving and storage systems.

**Materials newly considered as scholarly outputs.** Researchers, librarians, publishers, and archivists increasingly seek to register, certify, preserve, and access a range of research outputs beyond peer-reviewed articles and books. These include datasets, simulations, software, visualizations, and other non-traditional representations of knowledge and information (Van de Sompel et al., 2004). The expanding definition of “publication” poses new issues for the scholarly reward system, which is heavily biased toward traditional communication types (articles, archival conference papers, books); promotion and tenure decisions rarely consider data or software contributions (Howison et al., 2015). One notable development in support of broadened reward criteria has been the de facto standardization of infrastructural elements such as persistent naming for data (e.g. EZID) or for authors (e.g. ORCID), and mechanisms for data citation (e.g. DATACITE).

Combined, these three trends have vastly expanded both the size and the types of data available for scholarship, as well as the incentives to archive and to share them. Yet traditional questions around quality and trust remain—the very questions addressed by traditional scholarly knowledge infrastructures: how to guarantee the quality and authenticity of data? How to preserve them? How to foster data sharing and reuse? And more fundamentally: *Which entities should preserve and maintain data as scholarly resources?*

Until recent years, with certain important exceptions, data remained essentially external to formal scholarly communication. Under the print publication regime, these items were typically too expensive or impractical to disseminate on paper alongside the articles whose results relied on them. Where it did occur, data and software sharing took the form of informal researcher-to-researcher exchanges on paper, punch cards, tapes, or disk drives (Wallis et al., 2013). Data were implicit in publications as the basis for results and analysis, yet rarely were they formally registered, cited, published, or curated. In contrast to the

well-developed ecology of scholarly publishing, even in the early 2000s,

no comparable infrastructure exist[ed] for data. A few fields have mature mechanisms for publishing data in repositories. Some fields are in the stage of developing standards and practices to aggregate their data resources and make them more widely accessible. In most fields, especially outside the sciences, data practices remain local, idiosyncratic, and oriented to current usage rather than preservation, curation, and access. Most data collections—where they exist—are managed by individual agencies within disciplines, rather than by libraries or archives. Data managers usually are trained within the disciplines they serve. Only a few degree programs in information studies include courses on data management. (Borgman, 2007: 115–116)

As an example of a pre-digital shared data archive, one could cite the Inter-university Consortium for Political and Social Research (ICPSR), a specialized repository of census and survey data established in 1962. ICPSR is functionally similar to libraries in its levels of curation, control of access, and user (patron) support. More recently, large-scale cyberinfrastructures (Atkins, 2003)—heavily promoted by institutions such as the US NSF—have proliferated to provide data, tools, and software to various disciplinary communities (Edwards et al., 2013; Bowker et al., 2010).

### *The de-integration of scholarly infrastructure*

The advent of the commercial Internet and the World Wide Web in the early 1990s introduced both technologies and cultures that challenged traditional scholarly infrastructure. Previously, such infrastructure was by and large a “black box,” with highly integrated functionality bound up with technologies and institutions. Borrowing a term from Actor-Network Theory (ANT), it was “punctualized” (Callon, 1990) as a coherent, monolithic entity, i.e. as a single node in the network of interacting actors (Law, 1992). Depunctualization refers to “opening” that black box to expose the identity of the components, thereby opening up opportunities to replace, modify, or even eliminate those formerly hidden components. Depunctualization frequently occurs in the context of crisis or instability, when new or changed actors disrupt an established actor-network (Gehl, 2016; Latour, 1999).

The affordances of the web, with its openness for participation, essentially for free and without geographic constraints, produced just such an instability. It depunctualized the scholarly communication

infrastructure, exposing its many parts and their interconnections, leaving them fragile and contingent rather than robust and permanent. The result was the *de-integration* of the formerly tightly integrated chain of actors and institutions constituting the scholarly publishing infrastructure. We call this a “de-integration” in preference to the commonly used but often inaccurate term “disintermediation,” in which intermediary actors in a producer-to-consumer chain are bypassed (Gellman, 1996). While in some cases traditional intermediaries such as publishers were indeed simply bypassed, in many other cases *new* entities emerged to fulfil some tasks related to scholarly publishing. Often, instead of reducing the number of intermediaries involved (as “disintermediation” implies), de-integration actually increased it.

The e-print movement and the institutional repository effort of the early 2000s, both of which focused on textual scholarly artefacts, illustrate this rise of new intermediaries in a de-integrated context. The near-zero marginal costs of electronic publishing opened the door to a plethora of online genres. Many emulated traditional journals, using the then-emerging Portable Document Format (PDF) standard to provide electronic “pages” that looked exactly like their print counterparts, while maintaining traditional journal workflows. More significant (but still text-focused) were those institutions, learned societies, and even individual scholars who set up self-deposit, non-peer-reviewed repositories such as the well-known ArXiv for physics, formed in 1991 (Ginsparg, 2011); boutique digital libraries of scanned documents, such as Stanford University’s SiliconBase;<sup>2</sup> and numerous “institutional repositories,” usually managed by academic libraries (Johnson, 2002). By the turn of the millennium, web-based content management platforms such as DSpace and Fedora, as well as open protocols for federating their metadata and content (e.g. OAI-PMH), had emerged upon which enhanced services (e.g. metadata aggregation services and peer-reviewed “virtual journals”) could be built. In combination with a burgeoning “open access” movement, these new technologies disrupted traditional journal-based publishing.

The same new technologies that led to this de-integration also opened the possibility of accommodating new scholarly objects. In principle, at least, digital data and software could now join the same publishing, storage, and access framework as articles and books; this proved considerably more difficult than many imagined. By the mid-1990s, some journals were beginning to require the deposit of data and analysis scripts or codes along with articles. McCullough et al. (2006) analyzed 150 articles in one such journal published between 1996 and 2002. In the large majority of cases (135 out of

150 articles), the data and software deposited with those articles were insufficient to permit independent replication of results. Incomplete data, idiosyncratic formats, poor or non-existent metadata, and machine-dependent code were among the main reasons. This example illustrates the absence of well-defined registration and certification systems for data and software, which lacked analogues in traditional publishing; in the early 2000s, such systems remained to be invented and their form settled.

Significant movement toward shared norms and standards has led to competing efforts to re-integrate the traditional (and still-relevant) functions of the scholarly communication infrastructure. Examples of emerging data registration systems include Dryad, Dataverse, openICSPR, and Figshare (the focus of this article). Yet difficulties remain, and data systems that work for one community still do not translate easily to others (Borgman, 2015). Most of today's data repositories, portals, and curation systems initially developed in relative isolation, creating path dependencies or incompatibilities that rendered the eventual task of integration more difficult (David and Bunn, 1988; Ribes and Finholt, 2009). The combination of all these factors—vast data resources and incentives to share and archive them; scholarly needs and tools ready for experimentation; a de-integrated scholarly infrastructure—has opened the path for platforms to assume an important role in data sharing. In the next section, we develop an analysis of Figshare as a product of these converging trends.

Let's summarize the first part. The traditional infrastructure for scholarly communication relied on specialised intermediary institutions such as journals, publishers, and libraries. The pipeline carrying research from authors to readers was integrated through norms, routines, standards, and technical systems, primarily focused on text, graphics, and images. With isolated exceptions, no similar integration or standardisation was established for data *per se*. Since the early 1990s, new actors and institutions have arisen to manage scholarly data, but few have yet achieved broad acceptance. With respect to data and software, standards for such functions as identification, description, and citation remain immature. Absent the historical precedents established for textual artefacts, uncertainty reigns regarding which institutions should curate and manage data and under which procedures. Meanwhile, new possibilities opened up by the Internet and WWW have de-integrated the traditional communication system, resulting in numerous experiments, realignments, and efforts to re-integrate into an infrastructure that supports a range of products of scholarship including traditional text publications, data, and software.

## Re-integration through platforms: A case study of Figshare

Exiting the context of text-based scholarly materials, in this section we describe how Figshare leverages the properties of digital platforms to insert itself within the multiple components and actors organizing data sharing. Figshare adopts a data interoperability architecture typical of web-based platforms. It maintains an API and modular architecture that allows third parties to develop applications making use of its services. This architecture effectively allows Figshare to re-integrate scattered actors and institutions of the scholarly infrastructure.

### Modularity and programmability of platforms

As we have seen, data currently reside in an unsettled, shifting frontier where numerous incompatible standards, technologies, and institutions compete to serve the cacophonous practices of publishers, disciplines, and individual researchers. As a result, nothing close to an infrastructure for data currently exists. Figshare positions itself as a bridging entity that addresses this lack of integration among stakeholders and their interests. It does so by *reaching out* to a multiplicity of stakeholders (from researchers to universities to open science initiatives), offering to cover their diverse needs while simultaneously *linking them all together* through the modularity of the Figshare architecture. This integration of isolated components will (it hopes) eventually provide the glue to insert data into the standards and practices of scholarly infrastructure.

This strategy leverages two properties of platforms: the *modularity* of platform architectures, and the *programmability* of an API that serves as a gateway to the discovery and access of hosted content. Manovich (2001) called modularity a defining property of new media, but management-science literature had already described platform architecture as comprising three elements: a few core components with low variability, many complementary components with high variability, and interfaces for modularity between core and complementary components (Baldwin and Woodward, 2008). When interfaces to core components are openly described, third-party developers can readily build complementary components. Apple's iOS and Google's Android are excellent examples of core components, each using open APIs to attract tens of thousands of developers to build complementary apps. Enhanced by the inherent flexibility of programmable digital devices, platform architecture permits users and developers to go far beyond the original designers' project (Montfort and Bogost, 2009). In stark contrast

to the static text- and image-oriented world of traditional scholarly communication, standardized data access and interchange constitute the DNA of digital platforms (Helmond, 2015; van Dijck and Poell, 2013). In summary, Figshare organizes flows of scholarly data around a *modular architecture* that is *programmable* through APIs, and configurable via plugins to other components of the infrastructure with which it interacts.

### *Reaching out and linking together*

The neutral-sounding term “platform” connotes a discursive strategy whereby platform operators present themselves as transparent intermediaries serving (and profiting from) content providers, users, advertisers, and others (Gillespie, 2010). Figshare deploys this discursive strategy to position itself as intermediary between virtually all kinds of actors concerned with scholarly data—researchers, open science activists, applications developers, academic institutions (such as libraries), journals, publishers—some of which we describe in the following sections. We rely here on various Figshare documents, presentations, and case studies published on the company’s website, published interviews with company founder Mark Hahnel, and a semi-structured interview with Hahnel conducted by one of us at the Figshare headquarters in London in September 2016.

First, Figshare targets *individual researchers*. It presents itself as a “platform where researchers can store, share, and get credit for all of their research” (Wired, 2014). It invites individual researchers to self-archive their outputs (including datasets, graphics, presentation slides, and almost anything else) through personal profiles much like those of Academia.edu, Facebook, and other social media. Hahnel understands the disincentives for researchers to share their data, sometimes including outright “resistance towards data sharing.”<sup>3</sup> Several features of Figshare therefore aim to address these disincentives and adapt to researchers’ unique needs and practices. The Figshare interface is designed to make depositing data as easy and effortless as possible. It incorporates affordances (such as drag and drop) popular in other web-based platforms (in presentations, Hahnel frequently refers to Figshare as, e.g. “Dropbox and Youtube for science”<sup>4</sup>). Moreover, the interface simplifies data deposit by reducing the amount of metadata researchers must enter during that process. Similarly, it incorporates features that help researchers receive credit for depositing their data, for example by providing Digital Object Identifier (DOI) to make their datasets discoverable and citable.

Second, Figshare aims to engage the *open science community*, for which Hahnel has been a powerful

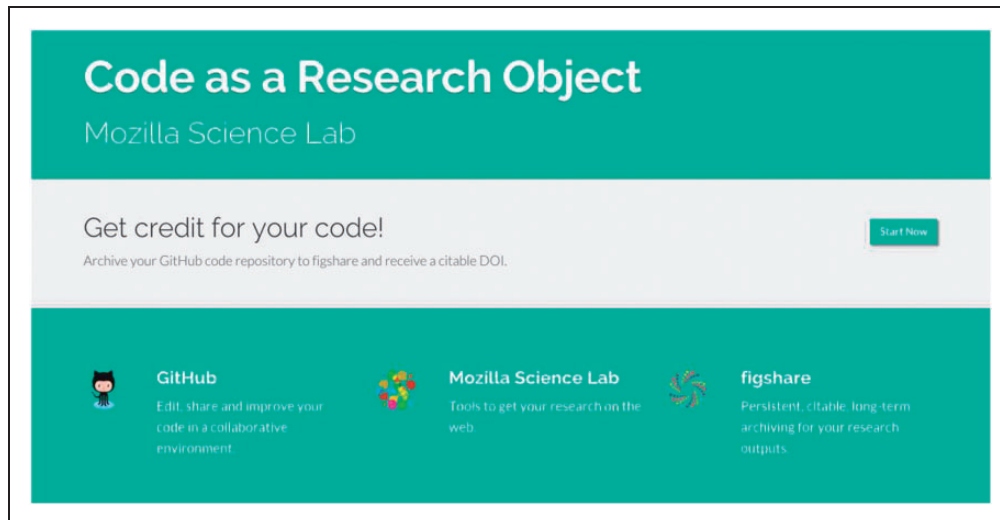
advocate. In 2013, he published a *Guardian* op-ed entitled “Open data: we need to share research results, even when they are wrong” (Hahnel, 2013). There he makes a strong case (without mentioning Figshare) for publishing negative results and providing easy access to primary raw data. Similarly, in another venue, Hahnel positioned Figshare in relation to recent scandals in the scientific world: “It has also been shown many times that there is vested interest, bribery, and corruption within the academic world of publishing, and this should stop” (quoted in Brinded, 2015). By invoking these elements of context, Hahnel clearly relates his platform to recent critiques of secretive data practices, unreplicable scientific findings, and the distorting effects of the highest-prestige journals on science. By emphasizing the benefits for science of increased transparency, Hahnel aligns Figshare with other government- and citizen-led open data initiatives seeking improved accountability through access to data (Ruppert, 2015). Such initiatives belong to a much larger Internet utopianism which hopes to “open” government, business, software development, and practically everything else—open meetings, open source, open peer review, open experiments, and so on—in the name of transparency, accountability, and replicability of results.

Third, Figshare aims to connect with *communities of developers* by allowing structured access to the metadata of deposited research output. To this end, Figshare participates in collaborative development projects. In 2014, Figshare, Mozilla Science Lab, and the code repository Github teamed up in the project “Code as a Research Object” (Figure 1). The collaboration aimed to design a Firefox browser extension that generates a DOI for data sets and code deposited on Github, to be released on Figshare. A well-functioning API is a sine qua non for such a project: “The partners have created a brokerage point between GitHub and Figshare, utilizing their APIs and developing systems so that any GitHub repository can be processed and received as a package” (Summers, 2014).

APIs and plugins play a central role in Figshare’s web-based platform, by allowing extensible and flexible application development. They are central to inserting Figshare in other platforms or code repositories such as Github. They also—in the terms of our argument—allow Figshare to attempt its grandest goal: re-integrating scholarly infrastructure by linking naming services, academic library web portals, institutional repositories, cloud storage, and electronic scholarly journals.

### *“Figshare for institutions”*

Shortly following its creation, Figshare broadened its offerings beyond simple self-archiving solution for



**Figure 1.** Project “Code as a Research Object”.

individual researchers to target institutions such as publishers and universities through two products: “Figshare for publishers” and “Figshare for institutions.” The offer to publishers started in 2012 with F1000 Research, followed by PLOS, Wiley, Taylor and Francis. In 2014, Figshare extended its offer to academic institutions, and has since contracted with Loughborough University, St Edwards University, and Monash University, among others. We focus here on the specifics of Figshare’s approach to academic libraries, as it illustrates most clearly the insertion of a platform within a traditional scholarly institution.

Figshare presents its service to institutions as “a simple and cost-effective software solution for academic and higher education establishments to both securely host and make publicly available its academic research outputs.”<sup>5</sup> The strategy adopted is not to replace existing institutions with a centralized web presence, but rather to insert an intermediary platform between localized features at both the backend (e.g. the institution’s own online storage) and the frontend (e.g. “branded” online portals, as shown in Figures 2 and 3).

For both publisher and institutional products, the Figshare software acts as a common intermediary, or middleware layer, mediating between the institution or publisher’s specific user interface and other services, such as search, storage, and long-term preservation. The Figshare API allows programmatic control of those features. The most common use of this ability is to provide localized user interfaces, or “skins,” over a basic function set. For example, users interact with an application that appears to be provided by their local institution (e.g. University of Michigan); in reality, the user-interface layer is simply translating user actions into API-defined commands addressed to the Figshare middleware. The institution’s programmers do not need

to modify the Figshare code, which may not even be available to them; they just “program the API.”

While APIs create “layer on top of” capability, plugins permit “layer beneath” functionality. This makes it possible for the middleware, in this case Figshare, to leverage a variety of subsystems to accomplish some function or service. A common example is storage, e.g. either on the local file system or in the cloud. To accomplish this, the middleware layer provides an abstract storage module that mediates between common storage system requests, such as write and read, and the APIs of specific storage systems. In such cases the middleware’s code may have to be modified in order to use the storage systems’ APIs, but this is usually accomplished by dynamically linking plugin-conforming code to the main middleware application (without requiring full disclosure of the middleware code base).

The component architecture specific to Figshare is illustrated in Figure 4. Note the separation of localized functions such as presentation (i.e. user interface), shown in the upper left quadrant of the figure, from core data access functions in the lower left, with communication between them mediated by the core API. Furthermore, note the provisions for multiple plugins (green rectangles in the upper and lower right quadrants of the Figure 4) that permit local customization of search, document storage, notifications and the like. Virtually all of Figshare’s functionality may therefore be “branded” and shaped to fit specific local needs, while the core Figshare logic remains as a switchboard mediating all infrastructure transactions.

The API-and-plugin architecture allows Figshare to connect with and use existing institutional infrastructures and standards such as institutional repositories, cloud storage (e.g. Amazon S3, DuraCloud), archival storage (e.g. CLOCKSS, Digital Preservation



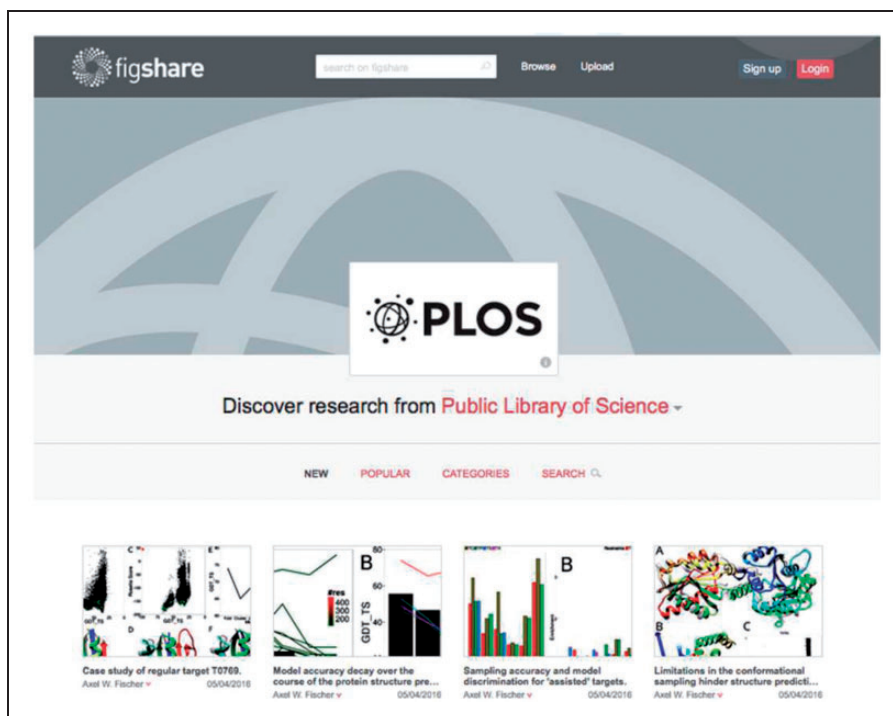


Figure 2. Figshare portal for PLOS.

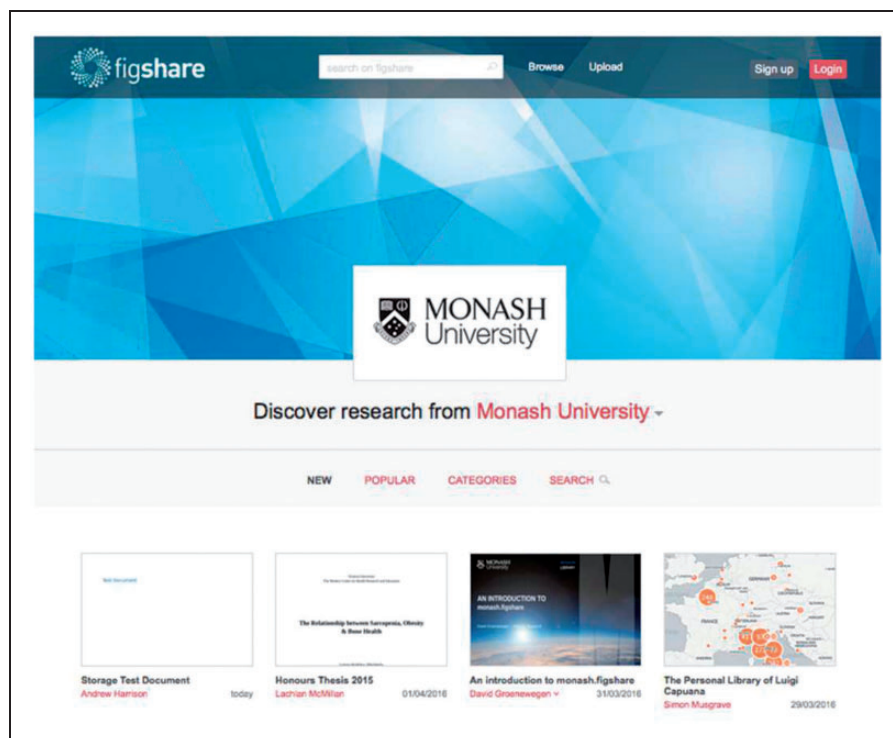
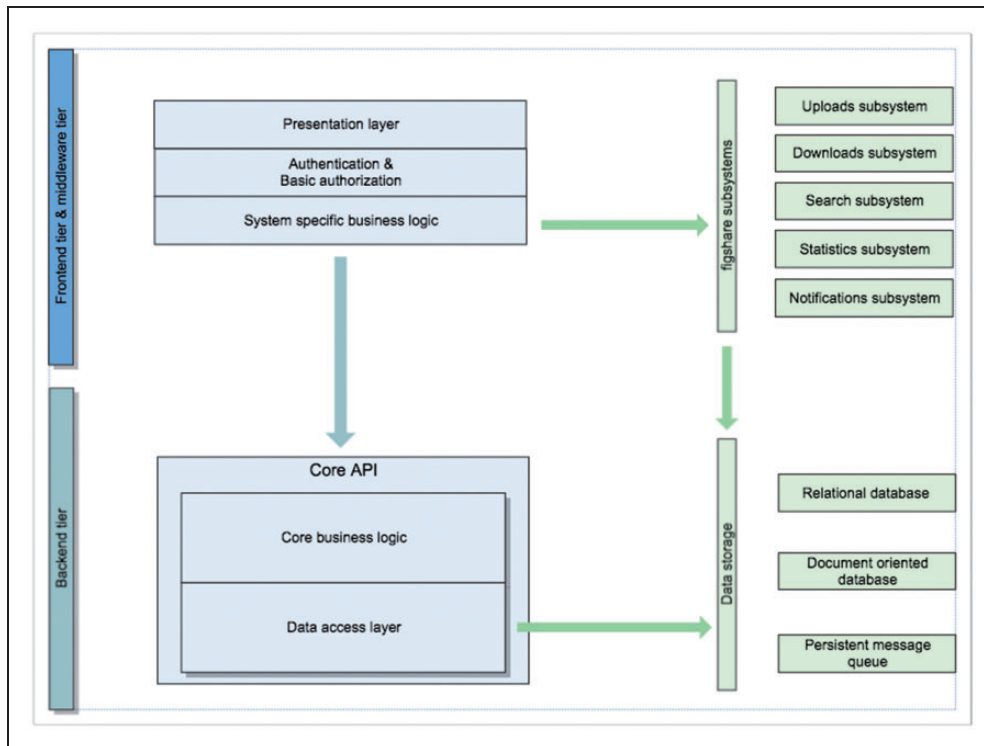


Figure 3. Figshare portal for Monash University.



**Figure 4.** Figshare system architecture.<sup>6</sup>

Network), research information management systems (RIMS), naming services (e.g. DOI, EZID), citation services (DataCite), and the like. This re-integration enables customized solutions for archiving and display research data that are simultaneously institutionally specific and globally interoperable (Hahnel, 2015).

Let's summarize the argument so far. The current ecosystem for data sharing lacks a tradition of standards and practices to treat data as a scholarly product, from production to sharing to archiving to reuse. Figshare seeks to become a new intermediary that can re-integrate many incompatible, non-standardized and/or localized components into a coherent, widely shared data infrastructure. It leverages both the discursive power and the technical architecture of platforms to connect multiple actors, such as researchers and libraries, in the chain of scholarship. In the next section, we examine how this re-integration by platform may affect scholarship, for better and for worse.

### Platforms within infrastructures: Risks and opportunities for scholarship

We described earlier how the Web and digital network technologies de-integrated the existing infrastructure for printed scholarship, while simultaneously creating a context in which platforms such as Figshare could flourish. Through their API architecture, such

platforms constitute a “link” that has been missing between the scattered components of this infrastructure, thus providing a valuable service to the scholarly community. While the library world has made and continues to make efforts to standardize APIs<sup>7</sup> for data access and deposit,<sup>8</sup> these efforts remain immature, which presents a barrier to a “plug and play” distributed system. Moreover, Figshare's simplicity alleviates some of the technical disincentives for data sharing—yet without addressing directly the systemic lack of positive incentives for data sharing (as shown by Borgman, 2012). Moreover, applying this new logic of platforms to data sharing and the scholarly world broadly also comes with multiple risks, which we discuss below.

#### Re-centralization through APIs

We described earlier the strategic role of the Figshare API in connecting a plurality of actors. The company presents its API as public and open (Hane, 2013; Hyndman, 2017), permitting the widest possible access to Figshare's functionality and content (data and metadata). This presentation is in accordance with its commitment to open science, which goes hand in hand with a principle of expanding data circulation. As Hahnel puts it: “the more raw data available, the greater the transparency and the easier it is to verify the results” (Wired, 2014).

Beyond such discourses, platform studies scholarship deconstructs the idea that APIs are merely neutral conduits. Instead, it unpacks the power relations hidden within them: APIs constitute specific constraints (defined by the platform that creates and controls them), which determine who can access data, in which forms, and under which conditions. They allow connection, but do so by operating like protocols (Bucher, 2013; Galloway, 2006), and are best conceived as a “management . . . technique for governing the relations [they] contain” (Bucher, 2013). Seen in this light, Figshare acts as a gatekeeper (McKelvey, 2011) in the emerging infrastructure for data sharing—even though, paradoxically, it does so by broadening the types of research output that can be deposited and shared through its technology.

Our view of recentralization through platforms complements prior work in STS that has highlighted similar trends. The increasing role of data in large-scale research (Edwards, 2010; Baker and Millerand, 2010; Star and Ruhleder, 1996) comes with its own ordering of scientific practices (Bowker and Star, 1999), and a “layering” of categories and standards that embrace some scientific practices, but exclude others (Bowker, 2005). Similarly, data sharing generates “metadata friction” (Edwards et al., 2011) as well as difficult conflicts between the local context of data production and the goal of global circulation and reuse (Leonelli, 2013). However, the rise of hybrid objects that combine properties of both infrastructures and platforms (Plantin et al., 2018) requires a closer look at the architecture and normative power of digital platforms.

Once we view open and public APIs as a technology of governance (following Bucher, 2013), it becomes apparent that Figshare intervenes in scholarly infrastructure at multiple levels. First, even as Figshare decentralizes data sharing by linking a large number and variety of actors, it also *re-centralizes* data flows around its platform, becoming what Callon (1986) famously called an “obligatory passage point.” Through this typical dynamic of decentralization and recentralization (Gerlitz and Helmond, 2013; Helmond, 2015), platforms take the role of *de facto* passage point in the circulation of data. On the one hand, platforms foster possibilities for sharing, programming, and “remixing” data, potentially generating a whole ecosystem of applications; on the other, they position themselves at the centre of this ecosystem, to become the single entity mediating all data circulation, with potential consequences for data openness and brokering.

### Financing and ownership

Following research that showed how the technical and economic dimensions of platforms are inseparable

(Langlois et al., 2009), an investigation of Figshare’s ownership and financing structure is useful to reveal a relation between the profit structure of Figshare and its features. In September 2011, Figshare became part of Digital Science, a for-profit “umbrella investor and startup incubator” owned by Macmillan Publishers. Digital Science’s portfolio includes a number of knowledge infrastructure components, such as Altmetric to measure the “impact” of online scholarly artifacts. The Figshare FAQ describes their relationship as follows:

- Figshare is an independent body that receives support from Digital Science. “Digital Science’s relationship with Figshare represents the first of its kind in the company’s history: a community based, open science project that will retain its autonomy while receiving support from the division.”
- Figshare operates independently of other Macmillan portfolio companies. However, we do work closely with our Digital Science sister companies such as Altmetric.com and Readcube [sic].<sup>9</sup>

However, the FAQ never makes clear the exact relationship between Figshare and Digital Science/Macmillan. Nor has Figshare publicized its corporate structure, board of directors, advisory bodies, or governing board (as noted by Murray-Rust, 2015). These murky corporate structures and relationships will not reassure critics of privatized scholarly infrastructure.

A comparison with article-sharing platforms can highlight potential concerns. The scientific publisher Elsevier acquired the paper-sharing platforms Mendeley and Social Sciences Research Network (SSRN) in 2013 and 2016 respectively. Open-access advocates feared that Elsevier would assert ownership of these repositories’ content—a worry apparently validated when SSRN promptly began pulling down papers for copyright infringement (Masnick, 2016). Like other scientific publishers, Elsevier has also shown strong interest in web-based data analytics for measuring scholarly “impact.” As Figshare provides both repositories and data analytics, it might be next in line for acquisition. Would Figshare resist an aggressive (or generous) acquisition offer from a leading publisher?

Additionally, some “open” paper-sharing platforms have begun brokering access to their databases of deposited outputs. For example, Academia.edu recently implemented a series of features, such as citations for papers in its repository, that are only accessible via paid premium accounts (McKenna, 2015). This move, too, created a controversy in the scholarly world, relayed by the Twitter hashtag #DeleteAcademiaEdu (Plantin, 2016).

Unlike stand-alone paper-sharing platforms such as Mendeley, Academia.edu, and ResearchGate, which deliberately bypass publishers, libraries, and institutional websites, Figshare currently seeks strong integration with the full scholarly infrastructure. For example, it attaches to the Digital Preservation Network (DPN) and CLOCKSS archives. Yet nothing in Figshare's architecture prevents future changes in its revenue model, which might alter access to data and features—changes which could be instantly imposed on the entire ecosystem via the API.

### *Splintering knowledge infrastructures*

In the age of the Web, how can scholarly infrastructures preserve reliable, minimally restricted access to registered, certified knowledge, including the data and software that may underlie it (Edwards et al., 2013)? Weinberger (2012) argues that libraries should reimagine themselves as “open platforms,” renewing their mission by providing tools (including APIs and other “social tools”) for the appropriation and reuse of scholarly resources. In opposition, Mattern (2014) defends a conception of “library as infrastructure,” with priority to combining and sustaining the multiple aspects—physical, technological, intellectual, and social—of its traditional mission.

Figshare enters squarely into the middle of these debates about libraries' role in a “Big Data” age. Hahnel emphasizes that Figshare complements, rather than replaces, existing knowledge institutions in their mission to provide a “trusted long-term repository for the scholarly record” (Hahnel, 2016). There is certainly a powerful logic to this goal. As Hahnel himself wrote, “when cash-strapped libraries threaten to end journal subscription deals over rising costs, alternative business models for scholarly publishing that take advantage of falling online storage costs and increasing global internet access must be considered” (Hahnel, 2012).

Yet Figshare may still further the existing tendency of many modern infrastructures to *splinter* (Graham and Marvin, 2001). From the mid-19th century through most of the 20th, a widespread “modern infrastructural ideal” valued public provision of socially important services, including knowledge services such as those of libraries and publicly funded science. Since the 1970s, that ideal has eroded in favor of private, for-profit provision, with greater access for those who can pay more and only residual, basic services for everyone else. This process entails hiving off profitable aspects of infrastructure to private concerns, which then entwine themselves in the infrastructural fabric, providing essential services but also changing the logics of provision. Just as UPS, FedEx, and other private delivery firms provided greater speed and reliability—but also eviscerated

traditional public postal services—so for-profit scientific publishers have taken over some roles once reserved for academic libraries and archives, as we observed earlier. With their large margins and oligopoly status, they can conjure resources and systems for data management and command obedience to standards they impose, and can do so across institutions and national borders.

Figshare positions itself as a solution to this splintering of traditional scholarly infrastructure, in two ways. First, as a platform, Figshare also offers integration across institutions and borders, but its “under-the-hood” approach permits distinctive “branding” and customized services at the level of the user interface, allowing libraries considerable latitude to experiment within their local communities. In this way it may open up a middle path between Weinberger's researcher-centered “open platform” concept and Mattern's community-centered “library as infrastructure.” If successful, this could permit academic libraries and archives to regain at least part of their traditional roles in propagating norms and practices for professional scholarly communication, while adding data management to their repertoire of services and offering the reach and power of Figshare's centralized holdings.

Second, it promises to do this at very low cost. Institutions would pay the organization “less than the price of a full-time employee” to provide data management capacities (quoted in Brinded, 2015). As proof of the concept that libraries and archives can do more with less, Figshare might thus play a role (even if unintentional) in justifying even further budget cuts and/or privatization. In the context of preservation and curation, such questions as what to preserve, for how long, and at what level of curation (extent and quality of metadata) are already difficult to answer. But while traditional libraries and archives were always affected by financial contingencies, historically their curation decisions were not driven by the return-on-investment criteria typical of for-profit organizations. This begs the question of whether the lower-cost alternative can provide the necessary level of service, especially for sensitive tasks such as anonymizing data, verifying quality (as determined by a variety of metrics), and registering/attributing data to the appropriate parties for eventual citation.

### **Conclusion**

We have argued that data, as a new genre of formalized scholarly object, currently reside in an uncertain position. The standards, norms, practices, and institutions developed over centuries for traditional text-based scholarly communication do not fit them well. Thus until recently, scholars shared data informally and

privately, much as they communicated via personal letters before the rise of scientific journals in the 17th century. These ad hoc communications remained outside the chain of actors and technical systems that publish, disseminate, and preserve text-based knowledge objects. Meanwhile, the traditional chain that characterized printed scholarly communication became “de-integrated” as the Internet and World Wide Web provided vehicles for new intermediaries and systems to arise.

Several entities currently present themselves as candidates to re-integrate the splintering infrastructures of scholarly knowledge. Initiatives such as the Research Data Alliance, the National Data Service, and some publishers’ data repositories attempt to build on existing structures. But in general, academic institutions and libraries are still trying to work out how data fits into their curatorial models. This uncertainty leaves room for platforms such as Figshare to develop new ways to archive and share scientific data. But where the logic of traditional infrastructure leads to a few large, centralizing entities, such as libraries and publishers, and converges on a relatively small number of standards and practices, Figshare leverages the logic of platforms to link a potentially very large range and number of functions and actors as complementary components, with APIs as the key gateway technology. Platforms rely on this lightweight structure to insert themselves as intermediaries within a chain of actors, providing the *apparent* integration of traditional infrastructure while permitting more, and more varied, entities to join.

Figshare thus provides one possible solution to the library community’s urgent need to accommodate an increasing heterogeneity and quantity of scholarly data. We have argued, however, that relying on Figshare (or any other privately owned platform) to perform this task entails two risks. The first is that the platform becomes an obligatory passage point, recentralizing data flows around itself, with unforeseeable consequences that might include an accumulation of brokering power over research data and the commercialization thereof. The second is that the success of a platform strategy may amplify the existing tendency of infrastructures to splinter, with their more profitable functions gradually pruned off and privatized. Should this trend continue, knowledge infrastructures—like their physical counterparts—might be starved of the funding and attention needed to preserve and maintain data, information, and knowledge over the long term.

#### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

#### Funding

This article results from research conducted during the two following projects: New directions in the Study of Infrastructures (2013-2014, Principal Investigators: Prof Paul Edwards, Dr Carl Lagoze, Prof Christian Sandvig, funded by the MCubed program, University of Michigan); The Science of Data Science (2014-2015, Principal Investigator: Dr Carl Lagoze, funded by the Alfred P. Sloan Foundation and the Gordon and Betty Moore Foundation).

#### Notes

1. Philosophical Transactions of the Royal Society.
2. Stanford University’s SiliconBase: <http://web.stanford.edu/group/itsp/>. Paul N. Edwards directed this effort from 1996 to 1998.
3. Mark Hahnel, interview, 1 September 2016.
4. Website AngelList: <https://angel.co/Figshare-2>
5. Digital Science website: <https://www.digital-science.com/press-releases/Figshare-for-institutions-solving-the-research-data-management-problem-for-educational-research-institutions/>
6. Figshare system architecture: <https://www.glify.com/go/publish/6041874>
7. <http://www.standardsforapis.org>
8. <http://swordapp.org/about/>
9. Figshare Support Forum: <https://support.figshare.com/support/solutions/articles/6000061086-what-is-figshare->

#### References

- Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 June. Available at: <https://www.wired.com/2008/06/pb-theory/> (accessed 29 January 2018).
- Atkins D (2003) Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Available at: <https://arizona.openrepository.com/arizona/handle/10150/106224> (accessed 29 January 2018).
- Baker K and Millerand F (2010) Infrastructure ecology: Challenges in achieving data sharing. In: Parker JN (ed.) *Collaboration in the new life sciences*. Surrey: Ashgate Publishers, pp. 111–138.
- Baldwin CY and Woodard CJ (2008) The Architecture of Platforms: A unified view. Available at: [hbswk.hbs.edu/item/6025.html](https://hbswk.hbs.edu/item/6025.html).
- Borgman CL (2007) *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Borgman CL (2012) The conundrum of sharing research data. *Journal of the Association for Information Science and Technology* 63(6): 1059–1078.
- Borgman CL (2013) Why you should care about open data. In: *Open Access Week at UCLA*. Available at: <https://works.bepress.com/borgman/274/> (accessed 29 January 2018).

- Borgman CL (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.
- Bowker GC, Baker K, Millerand F, et al. (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrup L and Allen M (eds) *International Handbook of Internet Research*. Netherlands: Springer, pp. 97–117.
- Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.
- Brinded L (2015) This startup thinks it can speed up a cure for cancer by bringing scientific research into the YouTube era. *Business Insider UK*, 18 September. Available at: [https://figshare.com/blog/The\\_importance\\_of\\_Institutional\\_APIs/155](https://figshare.com/blog/The_importance_of_Institutional_APIs/155) (accessed 29 January 2018).
- Bucher T (2013) Objects of intense feeling: The case of the Twitter API. *Computational Culture* 3. Available at: <http://computationalculture.net/objects-of-intense-feeling-the-case-of-the-twitter-api/> (accessed 29 January 2018).
- Callon M (1986) Elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay. In: Law J (ed.) *Power, Action and Belief: A New Sociology of Knowledge?*. London: Routledge, pp. 196–233.
- Callon M (1990) Techno-economic networks and irreversibility. *The Sociological Review* 38(S1): 132–161.
- Conrad CC and Hilchey KG (2010) A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment* 176(1–4): 273–291.
- David PA and Bunn JA (1988) The economics of gateway technologies and network evolution: Lessons from electricity supply history. *Information Economics and Policy* 3(2): 165–202.
- Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. The MIT Press.
- Edwards PN, Matthew S, Mayernik AL, et al. (2011) Science friction: data, metadata, and collaboration. *Social Studies of Science* 41(5): 667–690. <https://doi.org/10.1177/0306312711413314>.
- Edwards PN, Steven J, Melissa C, et al. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. University of Michigan School of Information. <http://deepblue.lib.umich.edu/handle/2027.42/97552>.
- Galison LD and Daston P (2007) *Objectivity*. Cambridge, MA: Zone Books.
- Galloway AR (2006) *Protocol: How Control Exists After Decentralization*. Cambridge, MA: MIT Press.
- Gehl RW (2016) The politics of punctualization and depunctualization in the digital advertising alliance. *The Communication Review* 19(1): 35–54.
- Gellman R (1996) Disintermediation and the Internet. *Government Information Quarterly* 13(1): 1–8.
- Gerlitz C and Helmond A (2013) The like economy: Social buttons and the data-intensive web. *New Media & Society* 15(8): 1348–1365.
- Gillespie T (2010) The politics of platforms. *New Media & Society* 12(3): 347–364.
- Ginsparg P (2011) ArXiv at 20. *Nature* 476(7359): 145–147.
- Graham S and Marvin S (2001) *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*. New York, NY: Routledge.
- Hacking I (1990) *The Taming of Chance*. New York, NY: Cambridge University Press.
- Hahnel M (2012) Exclusive: Figshare a new open data project that wants to change the future of scholarly publishing. In: The Impact blog, 18 January. Available at: <http://blogs.lse.ac.uk/impactofsocialsciences/2012/01/18/can-we-do-better-with-scholarly-publishing/> (accessed 29 January 2018).
- Hahnel M (2013) Open data: We need to share research results, even when they are wrong. *The Guardian*, 8 October. Available at: <https://www.theguardian.com/higher-education-network/blog/2013/oct/08/open-research-data-science-papers> (accessed 29 January 2018).
- Hahnel M (2015) The importance of Institutional APIs. In: Figshare blog, 16 June. Available at: [https://figshare.com/blog/The\\_importance\\_of\\_Institutional\\_APIs/155](https://figshare.com/blog/The_importance_of_Institutional_APIs/155).
- Hahnel M (2016) Whitelabelling the figshare software suite for full institutional control of RDM. In: Figshare blog, 12 May. Available at: [https://figshare.com/blog/Whitelabelling\\_%20the\\_figshare\\_software\\_suite\\_for\\_full\\_institutional\\_control\\_of\\_RDM/272](https://figshare.com/blog/Whitelabelling_%20the_figshare_software_suite_for_full_institutional_control_of_RDM/272).
- Hane PJ (2013) Sharing research data – New figshare for institutions *Against the Grain*, 16 September. Available at: <http://uk.businessinsider.com/figshare-ceo-mark-hahnel-drug-discovery-open-data-academia-coding-2015-9>.
- Helmond A (2015) The platformization of the web: Making web data platform ready. *Social Media + Society* 1(2): 1–11.
- Hey T, Tolle K and Tansley S (2009) *The Fourth Paradigm: Data-intensive Scientific Discovery*. Redmond: Microsoft Research.
- Howison J, Deelman E, McLennan MJ, et al. (2015) Understanding the scientific software ecosystem and its impact: Current and future measures. *Research Evaluation* 24(4): 454–470.
- Hyndman A (2017) Research space ELN—Figshare integrations. In: Figshare blog, 11 September. Available at: [https://figshare.com/blog/Research\\_Space\\_ELN\\_-\\_Figshare\\_Integrations/293](https://figshare.com/blog/Research_Space_ELN_-_Figshare_Integrations/293) (accessed 29 January 2018).
- Johnson RK (2002) Institutional repositories: Partnering with faculty to enhance scholarly communication. *DLib Magazine* 8(11). Available at: <http://www.dlib.org/dlib/november02/johnson/11johnson.html> (accessed 29 January 2018).
- Keating P and Cambrosio A (2006) *Biomedical Platforms: Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*. Cambridge, MA: MIT Press.
- Kitchin R (2014a) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1). DOI: <http://>

- journals.sagepub.com/doi/abs/10.1177/2053951714528481 (accessed 29 January 2018).
- Kitchin R (2014b) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Thousand Oaks, CA: SAGE.
- Knorr-Cetina K (1999) *Epistemic Cultures: How the Sciences make Knowledge*. Cambridge, MA: Harvard University Press.
- Langlois G, Elmer G, McKelvey F, et al. (2009) Networked publics: The double articulation of code and politics on Facebook. *Canadian Journal of Communication* 34(3): 415–434.
- Latour B (1999) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- Law J (1992) Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice* 5(4): 379–393.
- Lazer D, Pentland A, Adamic L, et al. (2009) Computational social science. *Science* 323(5915): 721–723.
- Leonelli S (2013) Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties* 8(4): 449–465.
- Leonelli S (2016) *Data-Centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.
- McCullough BD, McGeary KA and Harrison TD (2006) Lessons from the JMCB archive. *Journal of Money, Credit, and Banking* 38(4): 1093–1107.
- 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 29 January 2018).
- McKenna L (2015) The convoluted profits of academic publishing. *The Atlantic*, 17 December. Available at: [www.theatlantic.com/education/archive/2015/12/the-convoluted-profits-of-academic-publishing/421047/](http://www.theatlantic.com/education/archive/2015/12/the-convoluted-profits-of-academic-publishing/421047/).
- Manovich L (2001) *The Language of New Media*. Cambridge, MA: MIT Press.
- Masnack M (2016) Just as open competitor to Elsevier's SSRN launches, SSRN accused of copyright crackdown. In: Techdirt, 18 July. Available at: <https://www.theatlantic.com/education/archive/2015/12/the-convoluted-profits-of-academic-publishing/421047/>.
- Mattern S (2014) Library as infrastructure. *Places Journal*, June. Available at: <https://placesjournal.org/article/library-as-infrastructure/> (accessed 29 January 2018).
- Montfort N and Bogost I (2009) *Racing the Beam: The Atari Video Computer System*. Cambridge, MA: MIT Press.
- Murray-Rust P (2015) Is Figshare open? In: Personal blog, 23 April. Available at: <https://blogs.ch.cam.ac.uk/pmr/2015/04/23/is-figshare-open-it-is-not-just-about-open-or-closed-it-is-about-control/> (accessed 29 January 2018).
- Plantin J-C (2016) Algorithmic Accountability in Scholarship: What We Can Learn from #DeleteAcademiaEdu. Media Policy Project (blog), February 17, 2016. <http://blogs.lse.ac.uk/mediapolicyproject/2016/02/17/algorithmic-accountability-in-scholarship-what-we-can-learn-from-deleteacademiaedu/>.
- Plantin J-C, Carl L, Paul NE, et al. (2018) Infrastructure Studies Meet Platform Studies in the Age of Google and Facebook. *New Media & Society* 20(1): 293–310. <https://doi.org/10.1177/1461444816661553>.
- Ribes D and Finholt T (2009) The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems* 10(5): 375–398.
- Rogers R (2009) *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press.
- Roosendaal HE and Geurt PATM (1997) Forces and functions in scientific communication: An analysis of their interplay. In: *Conference on Cooperative Research in Information Systems in Physics*, University of Oldenburg, Germany, 1–3 September, pp. 1–32.
- Rosenberg D (2013) Data before the fact. In: Gitelman L (ed.) *Raw Data is an Oxymoron*. Cambridge, MA: MIT Press.
- Ruppert E (2015) Doing the transparent state: Open government data as performance indicators. In: Rottenburg R, Merry SE, Park S-J, et al. (eds) *A World of Indicators: The Making of Governmental Knowledge through Quantification*. Cambridge, MA: Cambridge University Press, pp. 127–150.
- Star SL and Ruhleder K (1996) Steps toward an ecology of infrastructure. *Information Systems Research* 7(1): 111–134.
- Summers N (2014) Mozilla Science Lab, GitHub and Figshare team up to fix the citation of code in academia. In: The Next Web, 17 March. Available at: <https://www.theatlantic.com/education/archive/2015/12/the-convoluted-profits-of-academic-publishing/421047/> (accessed 29 January 2018).
- Van de Sompel H, Payette S, Erickson J, et al. (2004) Rethinking Scholarly Communication: Building the System that Scholars Deserve. *D-Lib Magazine* 10(9).
- van Dijck J and Poell T (2013) Understanding social media logic. *Media and Communication* 1(1): 2–14.
- Wallis JC, Borgman CL, Mayernik MS, et al. (2007) Know thy sensor: Trust, data quality, and data integrity in scientific digital libraries. In: Kovács L, Fuhr N and Meghini C (eds) *Research and Advanced Technology for Digital Libraries*. Berlin: Springer, pp. 380–391.
- Wallis JC, Rolando E and Borgman CL (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE* 8(7): e67332.
- Weinberger D (2012) Library as platform. *Library Journal*. Available at: [lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/](http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/).
- Wired (2014) Startup of the week: Figshare, 25 July. Available at: [www.wired.co.uk/article/figshare](http://www.wired.co.uk/article/figshare).