OXFORD

Making a Research Infrastructure: Conditions and Strategies to Transform a Service into an Infrastructure

Benedikt Fecher[1,2]**, Rebecca Kahn** ⬤ [1,]*****, Nataliia Sokolovska**[1]**,
Teresa Völker**[1]**, and Philip Nebe**[1]

[1]Knowledge and Society, Alexander von Humboldt Institute for Internet and Society, Französische Str. 9, Berlin 10117, Germany and [2]Research Infrastructure, German Institute for Economic Research (Berlin), Mohrenstraße 58, Berlin 10117, Germany

*Corresponding author. E-mail: kahn@hiig.de

Abstract

In this article, we examine the making of research infrastructures for digital research. In line with many scholars in this field, we understand research infrastructures as deeply relational and adaptive systems that are embedded in research practice. Our aim was to identify the relevant context factors, actor constellations, organizational settings, and strategies which contribute to the evolution of a basic service into an actual infrastructure. To this end, we conducted thirty-three case studies of non-commercial and commercial research services along the research life cycle. By examining how these services emerge, we hope to gain a better understanding of the conditions and strategies to transform a service into an infrastructure. We are able to identify competitive disadvantages for publicly financed infrastructure projects with regard to the mode of implementation and the resources invested in development and marketing. We suggest that the results of this study are of practical relevance, especially for individuals, communities, and organizations wanting to create research infrastructures, as well as for funders and policy makers wanting to support innovative and sustainable infrastructures.

Key words: research infrastructure; open science; infrastructure development; research life cycle; funding models; digital science

1. Introduction

Digital communication technologies have proved instrumental in changing practices across all sectors of society, including academia. The hope of many researchers and science policy makers alike is that the Internet will help foster scientific progress and ultimately to make science more open, that is more inclusive, accessible, and transparent (cf. Fecher and Friesike 2014; Heck 2021). However, realizing efforts such as this require concrete policy initiatives behind them, if they are to endure and become part of everyday research practice. To date, many policies tend to focus on getting the technical aspect of research infrastructures off the ground, such as the development of major scientific equipment, sets of archival or scientific data, or communication and computing networks (European Commission, 2016). As a result, we have seen a plethora of services emerge in recent years, which stand as a testament to the firm belief in scientific progress due to technology. While these are a valuable step in trying to meet new user and stakeholder needs and thereby integrate into the research life cycle (and, in some cases,

attempt to reconfigure it), we argue in this article that there is more to research infrastructures than technical black boxes.

Infrastructure studies offer a fruitful perspective from which to study how technical innovations might generate effects which loop back upon the social organization of science. Scholars in this field largely agree that only when a technical service is embedded in practice, when it becomes 'invisible' (Star and Ruhleder 1996; Bowker and Star 1999), can it be considered part of an infrastructure. In this understanding, infrastructures are much more than the technical assemblage of things; only when these are part of practice, can they be considered part of the infrastructure. Bowker and Star (1999) refer to the depths of interdependence between the technical networks and the real work of knowledge production as 'infrastructural inversion' and suggest that infrastructures become examinable, when they break down. In this light, the transformative potential of the Internet on scholarly practice can be seen as an ongoing irritation for routinized academic work, which offers us an opportunity to study changes in scholarly practice through the infrastructural lense (Kaltenbrunner 2015).

In this article, we present the results of an empirical study on the emergence of research infrastructures for digital science that we conducted as part of a research project funded by the German Federal Ministry of Education and Research (BMBF). In particular, we are interested in the relevant environmental (i.e. legal, political, and social) factors for research services (RQ1), the strategies services apply to engage users and stakeholders (RQ2), and the typical organizational characteristics (i.e. team constellation, workflows, and financing) that services feature (RQ3). To approach these questions, we conducted thirty-three case studies of emerging services along the research life cycle between March and December 2018. We used desk research and semi-structured interviews with representatives of these services (mostly founders, CEOs, and project leads). Our results shed light on the motivations and logics behind infrastructure development and the interdependencies between new technical services and academic knowledge production. We are able to identify competitive disadvantages for publicly financed infrastructure projects with regard to the modes of implementation and the resources invested in development and marketing. The results of this study are of practical relevance, especially for persons and organizations which want to create and sustain research infrastructures and for funders and policy makers who aim to create the conditions for research in the twenty-first century.

2. Conceptual background

2.1 Defining research infrastructures

For the purposes of this article, it is necessary to review the scholarly discourse on infrastructures and to derive a robust definition for an empirical investigation. To this extent, we conducted an extensive literature review drawing from infrastructure and information studies (see Online Appendix Table 1).

We find that there is a consensus in the scholarly discourse that infrastructures go beyond the pure material framework and also take into account social and environmental factors. Bowker and Star (1999) understand an infrastructure as a practical match among routines of work practice, technology, and wider-scale organizational resources. In their understanding, infrastructures are sunk into other structures of social arrangements and technologies and support communities of practice (cf. Bowker and Star 1998). In that line, Wouters (2014) defines infrastructures as a routinized and relational set of human interactions that are multilayered and cannot be constructed top-down. This echoes the work of Pollock and Williams (2010) who argue that infrastructures should be viewed iteratively over time, as entities with their own biographies and which only exist in social contexts. The bottom-up nature of infrastructures is further explored by Blanke and Hedges (2013) who argue that such an understanding is essential if an infrastructure is to adequately meet the needs of its users. Edwards (2013) describes infrastructures as ecologies or complex adaptive systems that incorporate technological standards, social practices, and norms. Similarly, Hanseth et al. (1996) propose that infrastructures rely on a degree of standardization and compatibility if they are to function effectively (see also Larkin 2013). Drawing on Strauss (1985, 1988), Kaltenbrunner (2015) describes infrastructures as a result of articulation work, that is the activity of meshing distributed elements of labor in cooperative settings. He differentiates the production task (e.g. a research report) from the articulation work (i.e. everything that is necessary to write the report). These settings, as previously described by Schmidt

and Bannon (1992), are increasingly distributed, thus requiring the kinds of cooperative, digitized support infrastructures that form the basis of this study.

We suggest that these general conceptions of infrastructures can be transferred to research infrastructures. Drawing from this, we proceed from an understanding of research infrastructures as deeply relational and adaptive systems where the material and social aspects are in permanent interplay. They are embedded in the social practice of research and influenced by environmental factors. This allows us to consider the examined services as infrastructures in the making, that is they are not (yet) part of research practice but try to become part of it, and informs our central research interest: by examining how these services emerge, we hope to gain a better understanding of the conditions and strategies to transform a service into an infrastructure.

2.2 Conceptual framework

Three conceptual dimensions appear particularly relevant in the context of this study and for answering our three research questions:

- Environmental perspective, that is the ecology in which services operate.

This conceptual dimension relates to the first research question and thus which and to what extent environmental factors play a role in the development of an infrastructures for digital science. As adaptive systems, it can be assumed that research infrastructures do not emerge without context and are indeed influenced by environmental factors. Here, we distinguish between legal norms (e.g. with regard to data protection) as well as societal and political discourses (e.g. science policy developments) with regard to the influence of digitalization on science.

- Social perspective, that is the practice that services try to penetrate.

This conceptual dimension relates to the second research question, that is the strategies services apply to engage users and stakeholders. Services must be embedded into the social practice of research in order to be part of the research infrastructure. In this context, two large (and occasionally overlapping) groups of social actors appear crucial to us. These are the actual users (i.e. people who use a service) and relevant stakeholders (i.e. people who do not use a service but are directly relevant to its provision). For example, repositories are used by researchers (i.e. they are the users), but they are funded by research funders and hosted by libraries (i.e. they are stakeholders). We assume that both groups are relevant for a service to become part of practice. Empirically, we are interested in what practical problems a service wants to solve (i.e. motivation), which users and stakeholders they address and what strategies they employ to engage them, i.e. to become part of the practice.

- Organizational perspective, that is the resources that services have to adapt.

This conceptual dimension relates to the third research question, that is the organizational characteristics that services feature. Taking the perspective of technical services, we are interested in the organizational capacities that a service has with regard to the team constellation, modes of implementation of changes, as well as the financial resources. Thereby, we assume that the interplay between

the material and the social does not only relate to the relationship between the service and its (external) users and stakeholders but also to the internal, social, and material, capacities.

3. Method

This study is part of the BMBF-funded research project DREAM (Digital Research Mining), which deals with infrastructures for digital science (i.e. scholarly practices that rely on digital resources).[1] The aim of this study was to better understand the conditions and strategies to transform a service into an infrastructure. We assume that the transformative potential of the Internet makes it possible to study infrastructures for scholarly practice insofar as new services challenge existing infrastructures and seek to become part of the infrastructure themselves.

To this end, we conducted thirty-three case studies of non-commercial and commercial research services along the research life cycle between March and December 2018. We used a purposeful, theoretical sampling, guided by three criteria: size, source of funding, and functionality. Regarding functionality, we chose cases that can be assigned to different phases of the well-established research cycle (cf. Wilkinson 2000; Humphrey 2006). This is to ensure that sufficient cases are included in our analysis for all practices and phases in a typical research project. Accordingly, we differentiated five broad phases (think and plan; discover; gather and analyze; write and publish; share and impact). Many services in our sample cover more than one phase. For instance, the service Knowledge Unlatched offers features for discovering and publishing. We approximated the size of a service by the numbers of employees indicated in the interviews and other available information such as profit and number of users. It was important to include both large and small services in order to better assess the impact of organizational resources on infrastructure development. Similarly, it was important to include both commercial and publicly financed services, as the two are subject to fundamentally different operational conditions (e.g. accountability to a research funder versus accountability to shareholders). It has to be said that many services have mixed business models. For instance, it is quite typical that services that receive public funds also receive individual payments by customers. A table of the cases in our sample can be found in the Online Appendix Table 2.

We conducted semi-structured interviews with representatives of the services (mostly CEOs, founders, or project managers). For the instrument, we converted the aforementioned conceptual categories into questions. This resulted in three topics:

- Environment (i.e. relevant political and societal discourses, and legal frameworks),
- Social practice (i.e. motivations, user, and stakeholder strategies), and
- Organization (i.e. team constellation, business model, and technical implementation).

The personal interviews have resulted in rich, textual data for the comparative analysis. We used a word-exact transcription of the interviews for our qualitative content analysis (cf. Mayring 2004). To this extent, we proceeded from a rough, deductive framework informed by the aforementioned categories and research interests and refined the category system through multiple rounds of thematic coding and coder discussions. In order to establish inter-coder reliability, all interviews were analyzed by two coders, using MAXQDA. Not all interviewees agreed to allow us to use their institutions' names or to publish the full transcripts. In these cases, we speak generally of 'service + number' and avoid identifiers in quotations. In general, the results will not refer to the interviewed persons by name, but to the services they represent.

4. Results

Here, we present the main findings of our research, relating to (1) environment (i.e. relevant political and societal discourses, legal frameworks), (2) social practice (i.e. motivations, user, and stakeholder strategies), and (3) organization (i.e. team constellation, modes of implementation, and business model).

4.1 Environment
We defined the external context in which the services operate as their environment, which consists of the legal frameworks within which it may operate, as well as relevant political and societal discourses. How the service anticipates these influences its ability to become embedded in research practice.

4.1.1 Legal framework
When asked about which legal provisions are of relevance for running their service, the respondents largely referred to copyright, privacy, and standard licenses. The majority of codes refer to privacy regulations (forty codes), followed by copyright compliance (twenty-three codes), and references to standard licenses (seven codes). The core operational challenge here is presented by different national legal regimes, to which the services—most of which operate internationally—must respond. In addition, when it comes to copyright, services aim to keep the threshold for sharing material low and often try to avoid individual licensing solutions by using standard licenses (e.g. Creative Commons). In order to comply with this set of legal obligations, research services need to invest in monitoring, compliance, and implementation work, as the interview with Service 6, a service that offers a unique identifier for individual researchers, demonstrates:

> We do a huge amount of work around privacy. Privacy regulations in every country are different. [...] We've gone through an external privacy audit since 2013 to ensure that we're meeting international standards. [...] We are fully compliant with GDPR, we also have to look outside of Europe, what are the other privacy regulations that we need to comply with.

4.1.1.1 Service 6. It is noteworthy that the three legal categories identified are central legal concerns for any web-based service (also in non-academic contexts). This reveals the digitally enabled nature of the observed services. As with other web services, a key challenge is anticipating different legal regimes.

4.1.2 Political and societal discourses
Open science is the dominant theme that the respondents refer to when asked about the relevance of political developments to their services. At the time of the interviews, this largely referred to policies that advocate for open access and open data. Multiple respondents, for example, refer to transformative open access agreements (e.g. the German DEAL negotiations between major scientific publishers and consortia of scientific institutions) and data policies (e.g. FAIR). When it comes to the geographic scope, respondents refer mostly to national policies passed by governmental institutions or national

funders (twelve codes), supra-national policies, such as those passed by the European Union (ten codes) and institutional mandates at the level of the library, university, or company (three codes). Many of the respondents state that they are monitoring policy developments closely, as these affect their business models. Here for example, a representative from Altmetric, a service that provides attention metrics for scholarly outputs, refers to developments in the realm of research evaluation.

> We pay attention in the UK and Australia and Hong Kong, the Research Excellence Framework type of thing. So in Australia it is ERA, in the UK it is REF. So the guidelines on how to assess research. Obviously, we want to be the people you go to as a research admin at the university, to get the evidence to write this case and so you can get the money you deserve.

4.1.2.1 Altmetric. Most services align themselves to open science and the aforementioned dimensions (i.e. transparency, accessibility, and inclusivity). Some of the respondents even lobby for open science, which can be seen as creating favorable environmental conditions for the service and are thereby beneficial for becoming an infrastructure. This becomes obvious in the interview with the Directory for Open Access Journals (DOAJ), an online directory that indexes and provides access to open access, peer-reviewed journals:

> We have been very much involved in pushing for open access policies, open access mandates in the European Union, for instance. At a national level we have been active behind the scenes lobbying for open access policies. We, together with many other organizations, have been quite successful in the last decade to motivate decision makers to go in the direction of open access and open science.

4.1.2.2 DOAJ. Interestingly, different understandings of open science stand out, especially when it comes to commercialization. Commercial services describe open science (implicitly and explicitly) as a business opportunity, whereas some non-commercial services articulate reservations about the commercialization of open science and even try to counter it strategically. This becomes obvious in the following quote from a representative of Dryad, a non-commercial repository for research data:

> Universities and university libraries are concerned about commercial publishers and commercial entities sort of taking over the research infrastructure space. That's part of what we are trying to combat with this new partnership with [name of a non-profit service] is how do we make nonprofit infrastructure that is more aligned with values of academia?

4.1.2.3 Dryad. On the one hand, the results show how closely digital science is associated with Open Science by the interviewees. On the other hand, the results show a divergence in what is perceived as open science. In particular, non-commercial services are dedicated to the early activist understanding of open science as articulated in the Berlin Declaration in 2003[2] or the Budapest open access Initiative in 2002.[3] They often see open science as liberation from commercial interests. Commercial services, on the other hand, relate to open science as a practice (e.g. sharing data, making articles openly accessible) and not necessarily to the underlying ideologies.

4.2 Social practice

For services to become infrastructures, they must be embedded within the social practice of research. Accordingly, our aim here was to identify how exactly services intend to become part of research infrastructure, that is which motivations they have and what strategies they employ in order to engage users and stakeholders.

4.2.1 Motivations

We found that interviewees referred to eight different types of motivations. It is noteworthy that many of the motivations relate to the aforementioned open science dimensions, that is accessibility (e.g. access), inclusivity (e.g. dissemination and collaboration), and transparency (e.g. transparency). Beyond that, the motivations mirror efficiency (e.g. orientation) and research governance considerations (e.g. compliance, recognition, and efficiency). These motivations are further delineated in Table 1.

When asked about their motivations, almost all respondents refer to potential improvements in the scientific workflow through the adoption of digital technologies. For example, they refer to better access to articles through platforms, increased transparency through archiving data, and better ways to disseminate results through social media formats. In many instances, they contrast the added value of their services with the deficits of established

Table 1. Subcodes for 'Motivation'.

Motivations (#codes)	Explanation	Example
Access (thirty-five codes)	Providing or improving access to research outputs	Supporting open access to research articles through repositories (e.g. EarthArXiv, DOAJ)
Dissemination (thirty-one codes)	Disseminating research outputs to different publics	Supporting new formats for research communication (e.g. Browzine)
Transparency (eighteen codes)	Increasing the comprehensibility of the research process	Facilitating data storing and management (e.g. figshare)
Orientation (thirty codes)	Filtering and providing an overview of research topics	Curating open access journals (e.g. DOAJ)
Compliance (twelve codes)	Supporting the compliance to rules and regulations	Providing structured guidelines for data sharing (e.g. Service 6)
Recognition (seventeen codes)	Providing recognition for alternative outputs	Using alternative metrics for practices and outputs (e.g. Altmetrics, Publons)
Collaboration (thirty-eight codes)	Facilitating collaboration among different actors	Providing tools for sharing and communicating (e.g. Paper Hive)
Efficiency (thirty-three codes)	Increasing the efficiency of the research process	Mining content from large amounts of data (e.g. moving)

infrastructures. This shows that, in part, services are competing against established players and seek to replace them. This becomes obvious in a statement from the representative of Service 1, a journal management and publishing system that has been developed to support open access publishing.

> You hand over the finished articles to publishers, including all rights. The publisher prints and distributes, so the rights are gone. The state basically paid twice, for paying the people who do the editing and for the libraries that buy the articles back. On the Internet, researchers have the opportunity to do this themselves.

4.2.1.1 Service 1. The motivations are of importance here because they show where the services see problems in current practice and thus how they justify their raison d'être. In many cases, services position themselves against other, already established services and in some cases even articulate a need to replace them.

4.2.2 Users and stakeholder strategies

Discovering how these motivations are translated into a strategy required identifying users and stakeholders and the activities designed to engage with them and meet their needs. It is important to distinguish between users and stakeholders when analyzing strategies, because user strategies tend to refer to technical adaptation needs (i.e. making a service useful), whereas stakeholder strategies tend to refer to outreach activities and customer relations (i.e. making a service accepted). Based on the responses, we identified eight user and six stakeholder groups (see Fig. 1). It became clear that researchers are by far the most important user group, bearing in mind that there are potential overlaps between the researchers and authors categories. The most important stakeholder groups are customers and data providers. The latter has potential overlaps with the other service category and shows how important other technical services and their APIs are for a service (e.g. Altmetric uses the Facebook and Twitter APIs to build an impact metric).

To a certain extent, the illustration of users and stakeholders provides a map of the relevant actors for digital research infrastructures. It shows that, in addition to the actors already expected, the platform and cloud services play a significant role in the making of research infrastructures and that services relate to other services outside of the academic sphere.

We identified eight strategies implemented by the services to adapt to user needs. We differentiated these between pull (i.e. when a service reaches out to users or monitors their behavior), push (i.e. when users reach out to the service), and dialog strategies (i.e. when user and service engage in a dialog)—see Table 2.

It is noticeable that services use different strategies at the same time to ensure usability. In addition, it can be stated that both, more digitally enabled (e.g. data analytics to monitor user engagement)
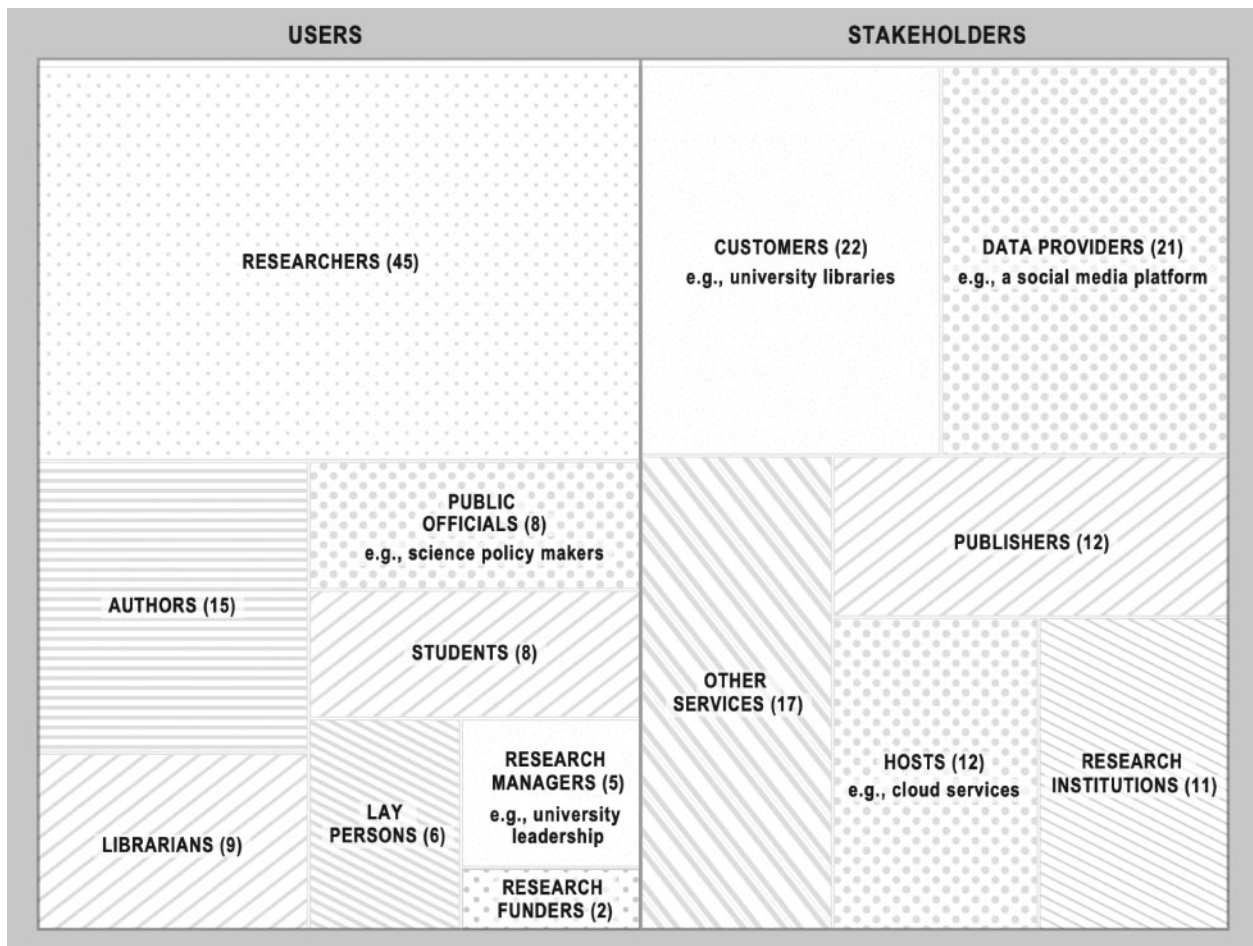

Figure 1. Users and stakeholders.

Table 2. Strategies to anticipate user needs.

Type of strategy	Strategies (# codes)	# codes
Pull	Data analytics (14), prototyping (9), user surveys (18)	41
Push	Feedback systems (32), support team (7)	39
Dialog	Teaching and training (12), advisory boards (3), lead users (9)	24

Table 3. Strategies to anticipate stakeholder needs.

Strategy (# codes)	Explanation
Customer outreach (8)	Building a relationship with existing or potential customers
Monitoring work (16)	Observing a political, legal, or societal discourse that is relevant to the service
Awareness work (14)	Influencing a discourse by raising awareness of the problem that the service was created to solve
Mediation work (18)	Mediating between different stakeholder groups (e.g. libraries and policy makers)

and analog methods (e.g. advisory boards), are used together. It is noteworthy that many of the strategies make use of typical software development practices (e.g. data analytics, prototyping, and feedback systems). In addition, we find the teaching formats of interest because they illustrate how services anticipate that users may need to adopt newer practices. The effort required to respond to user and stakeholder needs is reflected in a quote from a representative Service 3 which supports users in text and data mining activities:

> If the customers are still interested, there will be another very intensive discussion, in which we really discuss all features and go into the contractual details, so that everything is really transparent and clear. The customers can then do a training session. We currently offer a basic training course, which ideally takes place before commissioning. As soon as the installation has gone online, after a while we offer intensive training in which individual questions can be answered.

4.2.2.1 Service 3. We find the stakeholder strategies particularly intriguing because they demonstrate what a service is doing in order to become interwoven with the research environment. We identified four different strategies to engage stakeholders (see Table 3).

Non-commercial services articulate problems in engaging stakeholders due to a lack of resources. Both for-profit and non-commercial services attempt to influence discourses in their favor (i.e. awareness work). The largest category, *mediation work*, shows that services go to great lengths in order to connect and translate between different stakeholder groups which are considered relevant to the service. These are generally users and customers (e.g. researchers and librarians at an institution), between a service and other services (e.g. to be technically connectable), and finally between the programmers and users (e.g. in order to match technical possibilities with user requirements). The latter illustrates the negotiation of the technically possible with the socially desired as indicated in the working definition for infrastructure. This becomes obvious in an excerpt from an interview with a representative from Knowledge Unlatched, a platform that supports open access to books:

> I was with a team of very young developers, they all knew about the latest technologies and of course, they wanted to use these technologies, because that is most interesting for them [...]. That was a challenge, because these designers and front-end developers; they all wanted to have some fancy moving buttons. When we asked librarians to login and to use it, they were like, what is this? They have no idea, give me an Excel sheet, and I'll do it.

4.2.2.2 Knowledge unlatched. It becomes apparent that, in addition to the research communities as the biggest user group, other actors are of great relevance for the services—for example, because they guarantee the technical operation (e.g. data providers) or grant favorable institutional conditions (e.g. research institutions and research libraries). Furthermore, remarkable differences between commercial and non-commercial services can be seen, in that non-commercial or publicly funded services in particular articulate a lack of resources for outreach and implementation.

4.3 Organization

Here, we focus on the internal aspects of research infrastructures, in particular the roles that organizational design, team background, financing models, and technical adaptation play for the emergence of an infrastructure.

4.3.1 Team constellation

The services that we observed typically employ staff members with specific skills, either in information management, such as librarians (nine codes) and information scientists (three codes), researchers with field-specific knowledge (fifteen codes), developers (twenty-three codes), data scientists (eight codes), and sales and marketing staff (twelve codes). Furthermore, a common feature of the team constellation is decentralization—that is the involvement of volunteers and the employment of freelancers. These external team members are described by many as indispensable for the functioning of the services (e.g. Pre-prints or DOAB), and make it possible for the service to react to demands situatively and still remain operational. For-profit services emphasized the importance of having a strong technical team, and several claimed that at least one-third or even half of their staff has technical skills and a technical background. Both non-profit and commercial services acknowledged the importance and difficulty of finding skilled staff, particularly developers:

> One of the problems we have had is that it is always hard to have sufficient developers. People have a lot of demands on a service naturally. They start using it, they like things, they have ideas for how they would like to innovate and it is hard to always have sufficient developers and to be able to offer people everything they would like.

4.3.1.1 DCC. In contrast to non-commercial services, for-profit entities described sales teams as an important part of their staff. These teams help the service to adapt by ensuring they are able to fulfill user and customer needs, thereby deepening their ability to embed themselves into the research practice. There are also indications that non-commercial services struggle to recruit staff who have technical expertise. This may be due to the fact that the salaries in non-commercial services (which are mostly based within scientific institutions) are typically lower than those in the private sector and that there are limited reputative gains for infrastructure work in academia.

4.3.2 Business models

Regarding the business models, we broadly distinguish between rather non-profit and profit-oriented services. Among non-profit services (sixty-six codes), we differentiated between those who received institutional funding (eighteen codes), public funding (seventeen codes), charged fees (five codes), accepted donations (eight codes), and services that were exclusively financed by the founder/s (four codes). Profit-oriented services (forty-nine codes) included subscription models and licensing (twenty-six codes), individual payments (five codes), and private investments (thirteen codes) as their funding sources. Most services have mixed funding models, or at least emphasized the intention to seek other/additional sources of funding.

The fact that there are typically no sales employees for non-commercial services could have something to do with the funding logic of many public research funders, who do not provide funds for hiring staff with this type of skills for the duration of a project. The lack of funding opportunities and short-term nature of public funding was mentioned by several of the services that depend on public funding, as the quote from DRYAD shows:

> [...] currently there's just sort of the grant model, temporary funding that is designed to do some special project and then it ends and you're left with no means for continuing the work

4.3.2.1 DRYAD. Access to initial seed funding was common to both types of entities, but while non-profits often received initial funding from public funders, profit-oriented services often relied on investments from external companies in their startup stages. Several services started with seed investment (e.g. Tetrascience) and angel investment or were part of a startup incubator. The issue of sustainability for services that receive public funding is notable. There appears to be a need for follow-up funding that has not been satisfactorily addressed by research funders. Strategic partnerships are another feature of the organizational design. In some cases, strategic partnerships led to services becoming merged (e.g. Sharelatex, Dryad, and the Dash platform) or were partly acquired by a larger service (e.g. figshare by Digital Science).

4.3.3 Technical implementation

We are able to differentiate two modes of technical implementation: phased and iterative implementation. Phased implementation (six codes) describes an approach that begins with the users, that is screening their needs and then building the service accordingly. Iterative implementation (fifteen codes) is a process whereby user needs are constantly screened and adaptations are continuously

made. Generally, we observe that it was mainly non-commercial services which used the phased implementation approach, whereas for-profit services exclusively referred to iterative implementation. Below, in Table 4, there are two example quotes, the first referring to iterative implementation, and the second to phased implementation:

We consider this to be an important result, since it seems to reflect the funding logic of many non-commercial services, who typically expect implementation in consecutive work packages, whereas for-profit services appear to have to search for exposure earlier and permanently. This, we suggest, may further limit the adaptability and thereby competitiveness of non-commercial services.

5. Discussion

In our observations, it became clear that open science is the dominant discourse to which new online services for research refer. They use open science as an umbrella term to describe possible solutions to what they perceive as the shortcomings of the established system and infrastructures of the scholarly research life cycle, such as a lack of access to articles and the lack of recognition for alternative scholarly outputs (cf. Fecher and Frieske, 2014). What differs, however, are the services' responses to this discourse: although open science was initiated as a movement against the commercialization of research, it has been anticipated as a business model by many of the commercial services we observed. Meanwhile, non-profit services see open science as a set of principles, which framed an activist approach to research support. This finding echoes critical voices that have pointed to the appropriation of open science by commercial players (cf. Mirowski 2011).

The differences between commercial services and non-profit services permeated almost every aspect of their responses to their environment' (e.g. which public debates they participate in), how they engage with users and stakeholders, and how they implement changes. For instance, it is noteworthy that commercial services devote more resources to marketing and sales. Non-commercial services, on the other hand, articulate a lack of resources for marketing and sales. The distinctions between commercial and non-commercial services were also clear in the observations related to organization: Both types of services followed a fairly straightforward version of a decentralized digital service and both place similar importance on the need to hire staff with strong technical backgrounds. However, non-commercial services report that they do not have the resources to hire highly qualified programmers on a long-term basis. Further, non-commercial services often adopt phased implementation, possibly due to the funding logic of many public research

Table 4. Iterative implementation versus phased implementation.

Iterative implementation (commercial services)	Phased implementation (non-commercial services)
'[The] alpha version of the extension was available in the middle of February, so six weeks. And we've been iterating since then. So it's kind of a continuous process, but it took another three months before the Web Library was ready for example. So I suppose, yeah, so it's been in continuous development since January this year. We've just pushed an update today in fact to the Chrome Store. So there's an updated Chrome extension with a few new features, and the API is continually being developed and updated. We have a continuous release cycle, so pretty much every day a new release goes up'.	'We have had a very extensive empirical phase in which we have conducted interviews with our stakeholders, or representatives, as it were. We then modeled the use cases from these stakeholders. We had an abstract idea what it should be about, which of course was also described in the project planning and then in this first phase we actually conducted interviews with teachers, students and auditors. These were practically qualitatively evaluated and then the use cases were modeled'.
Scholarcy	Moving

funders. Commercial services generally adopt an agile implementation logic, possibly to be responsive to changing market needs.

Herein, we see a severe competitive disadvantage for non-commercial services. We suggest that there are three reasons for this: the first might have something to do with the phased implementation logic of public research funders, which restricts the capacity of a service to adapt to user needs. The second is a general lack of resources for hiring highly skilled staff, which puts non-commercial services at a disadvantage in a competitive market, and the third is a short-funding runway, which makes it difficult for non-commercial services to plan for future continuation. The implications of these three factors might be that in a competitive landscape, it is the commercial services, and their market-driven approach to open science, who have a better chance of embedding themselves in the research life cycle, and thereby co-shaping the scientific practices of the future.

6. Conclusion

In this research paper, we examined the making of research infrastructures for digital science, that is the relevant environmental factors, the strategies deployed to penetrate practice, and the organizational conditions necessary for a service to become part of a research infrastructure. We defined infrastructures as deeply relational and adaptive systems where the material and social aspects are in permanent interplay and which are influenced by environmental factors. The ways in which the services respond to these environmental factors and anticipate user and stakeholder needs create effects that might loop back into the overall social organization of science. It can be seen in our study that the services position themselves against shortcomings of the established infrastructures with regard to the access and transparency of research or the dissemination and curation of results online. In this regard, the study of emerging infrastructures might provide us with a glimpse into the future of an increasingly 'open' academic value creation.

At the same time, however, many services hold ties to established infrastructures, including mergers and acquisitions by the established publishers. In addition, the non-agile funding logic of public infrastructures and the limited financial possibilities of public institutions for highly trained staff could mean competitive disadvantages for publicly funded services. It therefore remains to be assumed that although the range of available services will change, the dominant players for research infrastructures may remain unchanged with digitization. This might explain why some scholars see open science as a neoliberal project in which market logics define the shape of research and non-lucrative services (e.g. for niche communities) are neglected (Mirowski 2018). In this respect, the dependence on commercial research infrastructures seems to be reproduced for digital science. If there is an interplay between research policy developments and research infrastructures, and if public funding for infrastructure works do not take community needs sufficiently into account, then certain communities who coalesce around non-commercial services risk being left out of research policy debates. The risk of funding logics contradicting infrastructure logics, especially for digital services, increases as the relative dominance of commercial services grows (cf. Morris and Rip 2006; Fry et al. 2009; Lilja 2020). Although our study is limited in terms of the cases studied and the depth of survey, it gives reason to critically reflect on public research infrastructure investments, for instance by revising funding policies and increasing incentives for highly skilled non-research staff. It appears sensible to us to revive infrastructure

research as a meta-scientific field of research especially now, in a time of transition to an increasingly digital ecosystem for scholarly work. This could help to ensure that public funds are used sustainably and moreover help to understand how possible futures of academic work might look like. Future, and in our eyes highly relevant, research questions could, for instance, concern the increasing interconnectedness and dependence on platforms, the long-term success of public infrastructure funding, and new governance models for critical infrastructures.

Supplementary data

Supplementary data are available at *Science and Public Policy Journal* online.

Conflict of interest statement. None declared.

Notes

1. https://www.hiig.de/en/project/dream-digital-research-mining/
2. https://openaccess.mpg.de/Berlin-Declaration (last opened 24 Jan 2020)
3. https://www.budapestopenaccessinitiative.org/ (last opened 24 Jan 2020)

References

Anderson, S. and Blanke, T. (2015) 'Infrastructure as Intermediation—From Archives to Research Infrastructures', *Journal of Documentation*, 71/6: 1183–202.

Barjak, F., Eccles, K., Meyer, E. T. et al. (2013) 'The Emerging Governance of E-Infrastructure', *Journal of Computer-Mediated Communication*, 18/2: 113–36.

Bilder, G., Lin, J., and Neylon, C. (2015) *Principles for Open Scholarly Infrastructures-v1* [Data set]. Figshare <https://doi.org/10.6084/M9.FIGSHARE.1314859> accessed March 2021

Blanke, T. and Hedges, M. (2013) 'Scholarly Primitives: Building Institutional Infrastructure for Humanities E-Science', *Future Generation Computer Systems*, 29/2: 654–61.

Bowker, G. C. (1994) *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920–1940*. MIT Press.

—— and Star, S. L. (1999) *Sorting Things out: Classification and Its Consequences (First paperback edition)*. MIT Press.

——, Timmermans, S., and Star, S. L. (1996) 'Infrastructure and Organizational Transformation: Classifying Nurses' Work' In: Orlikowski, W.J., Walsham, G., Jones, M.R., DeGross, J. (eds) *Information Technology and Changes in Organizational Work*, pp. 344–70. Berlin: Springer.

—— and Star, S. L. (1998) 'Building Information Infrastructures for Social Worlds—The Role of Classifications and Standards'. In: Ishida, Toru (Ed.) *Community Computing and Support Systems, Social Interaction in Networked Communities* , pp. 231–48. Berlin, Heidelberg: Springer-Verlag.

Camarinha-Matos, L. M. (2002) *Erratum to: Collaborative Business Ecosystems and Virtual Enterprises*, pp. E1–E1. Berlin: Springer.

Carse, A. (2012) 'Nature as Infrastructure: Making and Managing the Panama Canal Watershed', *Social Studies of Science*, 42/4: 539–63.

David, P. A. (2005) *Towards a Cyberinfrastructure for Enhanced Scientific*. Germany: University Library of Munich.

De Roure, D., Jennings, N. R., and Shadbolt, N. R. (2001) *Research Agenda for the Semantic Grid: A Future E-Science Infrastructure* (Report commissioned for EPSRC/DTI Core e-Science Programme).

Edwards, P. N., Jackson, S. J., Chalmers, M. K. et al. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Deep Blue.

European Commission. (2016) 'Open Innovation, Open Science, Open to the World: A Vision for Europe' *Publications Office of the European Union*.

<https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe> Accessed, March 2021

Fecher, B. and Friesike, S. (2014) 'Open Science: One Term, Five Schools of Thought'. In: S., Bartling and S., Friesike (eds) *Opening Science*, pp. 17–47. Springer International Publishing. <https://doi.org/10.1007/978-3-319-00026-8_2> Accessed, March 2021

Fry, J., Schroeder, R., and den Besten, M. (2009) 'Open Science in e-Science: Contingency or Policy?', *Journal of Documentation*, 65/1: 6–32.

Goldthau, A. (2014) 'Rethinking the Governance of Energy Infrastructure: Scale, Decentralization and Polycentrism', *Energy Research & Social Science*, 1: 134–40.

Hanseth, O. and Monteiro, E. (1997) 'Inscribing Behaviour in Information Infrastructure Standards', *Accounting, Management and Information Technologies*, 7/4: 183–211.

——, ——, and Hatling, M. (1996) 'Developing Information Infrastructure: The Tension between Standardization and Flexibility', *Science, Technology, & Human Values*, 21/4: 407–26.

Harvey, P. and Knox, H. (2012) 'The Enchantments of Infrastructure', *Mobilities*, 7/4: 521–36.

Haug, C. (2013) 'Organizing Spaces: Meeting Arenas as a Social Movement Infrastructure between Organization, Network, and Institution', *Organization Studies*, 34/5–6: 705–32.

Heck, T. (2021) 'Open Science and the Future of Metrics'. In: R., Ball (ed.) *Handbook Bibliometrics*, pp. 507–16. De Gruyter.

Hepsø, V., Monteiro, E., and Rolland, K. H. (2009) 'Ecologies of E-Infrastructures', *Journal of the Association for Information Systems*, 10/5.

Hewitt, C. (1988) 'Offices Are Open Systems.' In: A. H., Bond and G., Les (eds) *Readings in Distributed Artificial Intelligence*, pp. 321–9. Elsevier.

Hey, T., and Trefethen, A. E. (2005) 'Cyberinfrastructure for e-Science', Science (New York, N.Y.), 308/5723: 817–21. 10.1126/science.1110410 15879209

Hughes, T. P. (1983) *Networks of Power: Electric Supply Systems in the US, England and Germany, 1880–1930*. Baltimore: Johns Hopkins University.

Humphrey, C. (2006) 'e-Science and the Life Cycle of Research: Report to the Association of Research Libraries', *ERA* <https://doi.org/10.7939/R3NR4V> Accessed March 2021

Isaksson, K., Richardson, T., and Olsson, K. (2009) 'From Consultation to Deliberation? Tracing Deliberative Norms in EIA Frameworks in Swedish Roads Planning', *Environmental Impact Assessment Review*, 29/5: 295–304.

Jensen, C. B. (2007) 'Infrastructural Fractals: Revisiting the Micro–Macro Distinction in Social Theory', *Environment and Planning D: Society and Space*, 25/5: 832–50.

Jewett, T. and Kling, R. (1991) 'The Dynamics of Computerization in a Social Science Research Team: A Case Study of Infrastructure, Strategies, and Skills', *Social Science Computer Review*, 9/2: 246–75.

Kaltenbrunner, W. (2015) 'Infrastructural Inversion as a Generative Resource in Digital Scholarship', *Science as Culture*, 24/1: 1–23.

Larkin, B. (2013) 'The Politics and Poetics of Infrastructure', *Annual Review of Anthropology*, 42/1: 327–43.

Lave, J. and Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.

Lilja, E. (2020) 'Threat of Policy Alienation: Exploring the Implementation of Open Science Policy in Research Practice', *Science and Public Policy*. <https://doi.org/10.1093/scipol/scaa044> Accessed March 2021

Mayring, P. (2004) 'Qualitative Content Analysis.' In: U., Flick, E., von Kardoff, and I., Steinke (eds) *A Companion to Qualitative Research*, pp. 159–76. London: Sage.

Mirowski, P. (2011) *Privatizing American Science*. Cambridge, MA: Harvard University Press.

—— (2018) 'The Future(s) of Open Science', *Social Studies of Science*, 48/2: 171–203.

Morita, A. (2017) 'Multispecies Infrastructure: Infrastructural Inversion and Involutionary Entanglements in the Chao Phraya Delta', *Ethnos*, 82/4: 738–57.

Morris, N. and Rip, A. (2006) 'Scientists' Coping Strategies in an Evolving Research System: The Case of Life Scientists in the UK', *Science and Public Policy*, 33/4: 253–63.

Mueller-Langer, F., Fecher, B., Harhoff, D. et al. (2019) 'Replication Studies in Economics—How Many and Which Papers Are Chosen for Replication, and Why?' *Research Policy*, 48/1: 62–83.

Pipek, V. and Wulf, V. (2009) 'Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology', *Journal of the Association for Information Systems*, 10/5: 1.

Pollock, N. and Williams, R. (2010) 'E-Infrastructures: How Do We Know and Understand Them? Strategic Ethnography and the Biography of Artefacts', *Computer Supported Cooperative Work (CSCW)*, 19/6: 521–56.

Ribes, D. and Finholt, T. A. (2009) 'The Long Now of Infrastructure: Articulating Tensions in Development', *Journal for the Association of Information Systems (JAIS)*, 10/5.

Ruhleder, K. and King, J. L. (1991) 'Computer Support for Work across Space, Time, and Social Worlds', *Journal of Organizational Computing and Computing*, 1/4: 341–55.

Schmidt, K. and Bannon, L. (1992) 'Taking CSCW Seriously', *Computer Supported Cooperative Work*, 1/1–2: 7–40.

Star, S. L. (1999) 'The Ethnography of Infrastructure', *American Behavioral Scientist*, 43/3: 377–91.

—— and Ruhleder, K. (1996) 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces', *Information Systems Research*, 7/1: 111–34.

—— and Strauss, A. (1999) 'Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work', *Computer Supported Cooperative Work*, 8/1: 9–30.

Strauss, A. (1985) 'Work and the Division of Labor', *The Sociological Quarterly*, 26/1: 1–19.

—— (1988) 'The Articulation of Project Work: An Organizational Process', *The Sociological Quarterly*, 29/2: 163–78.

Wilkinson, D. (2000) *The Researcher's Toolkit: The Complete Guide to Practitioner Research*. London: Routledge.

Wouters, P. (2014) 'The Citation: From Culture to Infrastructure'. In: B., Cronin and C. R., Sugimoto (eds) *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*, pp. 47–66. Cambridge: The MIT Press.