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

Fecher, Benedikt; Sokolovska, Nataliia; Völker, Teresa; Nebe, Philip; Kahn, Rebecca

Arbeitspapier / working paper

Making a Research Infrastructure:

Conditions and Strategies to Transform a Service into an Infrastructure

Benedikt Fecher[1] (fecher@hiig.de; corresponding author)

Nataliia Sokolovska[1] (sokolovska@hiig.de);

Teresa Völker[1] (teresa.voelker@hiig.de);

Philip Nebe[1] (philip.nebe@hiig.de);

Rebecca Kahn[1] (kahn@hiig.de);

[1] Alexander von Humboldt Institute for Internet and Society, Department "Knowledge and Society", Französische Str. 9, 10117 Berlin, Germany.

Abstract

In this article, we investigate the making of research infrastructures. Our aim is to identify the relevant context factors, actor constellations, organizational settings, and strategies for a mere service to become an actual infrastructure. To this end, we conducted 36 case studies of non-commercial and commercial research services. Our research sheds light on the motivations and logics behind infrastructure development and the reasons why not every service succeeds in becoming one. We believe that the results of this study are therefore of practical relevance, especially for persons and organizations that want to create and sustain research infrastructured.

Introduction

Over the last few decades, "infrastructure studies" have emerged in the humanities and the social sciences, as scholars have begun to question and investigate the systems which connect, control, support, constrain and transform our work, leisure, learning and interactions. Infrastructures influence which subjects, topics or individuals are positioned on the margins or in the centre of social configurations, and are the systems against which social, political and technical identities are defined (Liu 2016). Research services only become what most infrastructure scholars consider an actual infrastructure if they succeed in penetrating scholarly practice and structured social relations. In these instances we might witness the process in which emerging research infrastructure generates effects which loop back upon the social organization of science. In this light, a plethora of new digital services emerge that aim to reconfigure research and are a testament to the firm belief in scientific progress due to technology. They challenge established infrastructures and try to become infrastructures themselves.

In this article, we investigate the making of research infrastructures. Our aim is to identify the relevant context factors, actor constellations, organizational settings, and strategies for a mere service to become an actual infrastructure. To this end, we conducted 33 case studies of non-commercial and commercial research services (see table 5). Apart from desk research and document analysis, we conducted semi-structured interviews with representatives of these technologies (mostly founders and CEOs / project leads) between March and December 2018. This study builds on and further refines the rich work of the humanities in dealing with information systems and sets them in relation to current developments. Our research sheds light on the motivations and logics behind infrastructure development and the reasons why

not every service succeeds in becoming one. We believe that the results of this study are therefore of practical relevance, especially for persons and organizations that want to create and sustain research infrastructure.

Background

Examining the scholarly discourse on research infrastructures, there is a shared consensus that infrastructures are deeply relational, i.e. they represent or support an organised practice. According to Star & Ruhleder (1996), infrastructure is a relational property, not a thing stripped of use. The authors shift the emphasis from looking at changes in technological components as a source of infrastructure development, to looking at changes in infrastructural relations, thereby linking the technical and the social. An important aspect of this shift is the emergence of the notion of "infrastructural inversion" (Bowker, 1994, Bowker and Star, 2000). In this formulation, the moment of inversion takes place when those examining an infrastructure see the system for itself, as a whole, without allowing it to disappear into the background. Typically, according to Bowker and Star (p.34), this happens when an infrastructure is experiencing a moment of breakdown and its technical and social components become visible. Inversion allows the observer to see the depth of the dependencies and interdependencies that make up the infrastructure, as well as making the influence of external forces, such as politics and knowledge production, visible (ibid). It is this moment that we have tried to pinpoint, and interrogate, in this study, particularly as, for many existing research infrastructures, the move into the digital realm is a significant, and irrevocable moment of inversion. This moment of digital inversion dovetails neatly with Kaltenbrunner's (2014) argument that rather than thinking of infrastructures as physical things, such as wires, tubes, and pipes we should consider them as relational states, which emerge when actors (some of which may be non-human) work simultaneously in a shared setting, in order to produce a shared reality, or coordinated activities. Kaltenbrunner goes on to argue that it is important to resist the temptation to think about the moment of infrastructural inversion in terms of a dichotomy, but that it should rather be "[...] conceptualised as a specific form of articulation work". This approach, he argues, makes it possible to accept that inversion does not reveal an entire infrastructure as it really is, but rather provides a situated moment in which it is possible to reconstruct the infrastructure.

Where the perspectives on infrastructure often differ is whether it is the material nature of the infrastructure that shapes the social practice or vice versa, and the extent to which they continue to relate to one another (see table 1 in the attachment). Wouters (2014), for instance, sees infrastructure as a set of social processes which provide the taken-for-granted context that enable everyday life and work and

which cannot be constructed top-down. Larkin (2013) positions infrastructure as materiality in the sense that it is a set of material artifacts that shape social practice and thereby generate the ambient environment of everyday life. Edwards at al. (2013) describes infrastructure as adaptive systems which emerge and evolve rather than appear fully formed. In this line, Bowker and Star (2000) describe infrastructures as deeply embedded in both practice and technology.

In order to arrive at our working definition, we have drawn elements from all three of these, and have seen all three reflected, to varying degrees, in the services we examined (see table 6). In line with Bowker and Star (2000), we understand research infrastructures as deeply relational and adaptive systems, i.e. the technical (e.g., the technological configuration) and social aspects (e.g., user and stakeholder needs) of a research infrastructure are in permanent interplay. In addition, we accentuate the environmental factor of research infrastructures (similar to Wouters 2014). They do not emerge in an experimental setting far away from interfering factors from the environment. Research infrastructures are part of the social practice of scientific work and thus responsive to the environment (e.g., legal, political and cultural developments). This influences the internal functioning of infrastructures, i.e. how infrastructures are set up in order to remain adaptable. This approach highlights the episodic appearance of certain indicators of the work of an infrastructure — although these do not always equate to finished products or completed tasks (Star and Strauss, 1999). This approach tends to see the productive practice as generative, and asserts that it is iterative practice that shapes technology.

Method

Case selection

We focus on services that emerge along the research life-cycle. For the case selection, we used the well-established circular model of the research life-cycle and identified services along each step (see table 2). In order to bolster the level of validity of our findings, we included different types of services in our case selection, such as large and small scale, as well as for-profit and non-commercial services, as we suspected that the size and way of financing affects organizational workflows typical change processes (e.g. agile design in for-profit vs. formal change requests in publicly funded projects). It has to be noted that size and source of funding are not clear-cut categories and we needed to make informed decisions in the sampling. For instance, we approximated the size of a service by numbers of employees indicated in the interviews and other available information such as sales volume, profit and number of users. The distinction between for-profit and non-commercial on the other hand comes from the source of funding.

Some services have mixed business models (e.g., funding from a research funder as well as individual payments by customers). Furthermore, some services are publicly funded but do not exclude becoming commercial entities (as is intended by many funding policies, e.g, in the European Commission's research and innovation framework[1]). We generally considered a service non-commercial if the main source of funding is public or from a foundation. Through this systematic approach we aimed to achieve the necessary case variance in order to draw generalizable conclusions.

In total, we analyzed 36 different cases of emerging services along the research cycle between June and October 2018. We targeted mostly CEOs and founders (in some cases also managers). A list of the services together with a short description can be found in the Appendix; table 2. Our sample contains 16 for-profit digital research services and 17 mainly publicly funded projects (many have mixed sources of funding). Some of the services might—according to our working definition—already count as infrastructures (i.e. they are entangled with the social practice of research) as a result of their crucial importance in certain communities. An example for this might be the repository arXiv, a repository for preprints, that was founded in 1991 by Paul Ginsparg and is widely used in many natural science fields (14,675 article submissions in October 2018). For the sake of clarity, we consistently speak of services.

Data analysis

Following our working definition for research infrastructure and the desk research we conducted in advance, we designed the semi-structured questionnaire based on the following core themes and general research interests (see table 1).

Themes	**Research interests**
Environment	What are relevant legal frameworks in which the service operates (e.g., changes in data protection law such as the GDPR)? What are relevant political discourses for the services (e.g., political initiatives for data sharing such as FAIR data)? What are relevant social developments for the services (e.g., social movements such as the open science movement)?
Practice	What problem does the service aim to solve? Which strategies does the service apply to penetrate research practice?

[1] https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020 (last opened January 24 2020)

	How does the service implement changes?
Organization	What are the typical features of the organizational design (i.e. workflows, team location)? What are typical features of the workforce (i.e. background of the team members)? What business model does the service use?

Tab.1. Themes and research interests

We used a word-exact transcription of the interviews for our qualitative content analysis (see Mayring 2000). To this extent, we proceeded from a rough, descriptive analysis framework informed by the aforementioned themes and research interests and built and refined the category system through multiple rounds of thematic coding and coder discussions. We used MAXQDA for the coding of the texts. Once we arrived at the final category system, each interview was coded by at least two coders separately. The code assignment in each category was again discussed in the group of coders until agreement was reached. In order to facilitate the comprehensibility of our results, most of the materials (interview guideline, transcribed interviews) can be accessed online. Not all interviewees agreed to publish the full transcripts; some wanted to stay anonymous. In these instances we used the generic term "service". Generally, in the analysis we will not refer to the interviewed persons by name, but to the services they represent.

Results

Following our working definition for research infrastructures and the core themes (environment, practice, organization), we will present here the main findings of our case studies.

Environment

Research infrastructures are responsive to legal, political and social developments. Accordingly, we subsumed these aspects in the main category *environment*. The following part will report on the different manifestations of these environmental factors and point to interesting areas of tension.

Legal

When asked about which legal provisions are of relevance for running their service, the respondents largely referred to copyright, privacy, and standard licences. The majority of codes refer to privacy

regulations (40 codes), followed by copyright compliance (23 codes) and references to standard licences (7 codes).

At the time of the interviews, the General Data Protection Regulation had just been introduced in the European Union. This is a far-reaching regulation in EU law on data protection and privacy. Many of the respondents (including those services that are outside of the European Union) referred to this provision – presumably because it affects their business models and workflows. Noncompliance with EU privacy and security standards can lead to penalties reaching into the tens of millions of euros (GDPR Website 2019). 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. When GDPR came out, we also did a lot of the head work. 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*."

Service 6

When it comes to copyright, many services mention the need to be compliant with existing copyright regulations, while at the same lowering the threshold for users to enable the exchange of creative outputs (e.g., by using standard licenses). This becomes obvious in the interview with DOAJ, a platform that registers open access journals:

"*Copyright issues are important, because we are working very much with publishers to help them apply relevant licenses […]. Open access is about liberal licenses as much as possible, a lot of our work is actually to convince publishers to support that part. Copyright and licenses are an important legal aspect of our work.*"

DOAJ

The three legal categories identified here are central legal concerns for any web-based service. This reveals the nature of the observed services, all of which are digitally enabled. The core operational challenge for the services 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 try to avoid individual licensing solutions instead, by using standard licenses (e.g. Creative Commons).

Policy

Here we understand political references to be those that refer to concrete research policy actors or developments. When it comes to the geographic scope of these references, respondents refer mostly to national policies passed by governmental institutions or national funders (12 codes). Supra-national policies, such as those passed by the European Union (10 codes) and institutional mandates at the level of the library, university or company (3 codes).

Across the geographical categories, open science is the central theme that the respondents refer to when asked about relevant political developments for their services. This largely refers to national or international policies that advocate making articles available, such as transformative agreements for open access. Multiple respondents, for example, refer to the German DEAL negotiations between major scientific publishers and consortia of scientific institutions. After all, many of the services support practices that can be assigned to a general open science theme (as delineated in the background section). For instance, the data repository figshare enables data sharing and the Directory of Open Access Journals (DOAJ) supports open access to articles by listing trustworthy open access journals. Some of the respondents even lobby for open science, which can be seen as creating favourable boundary 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."

DOAJ

Many of the respondents say 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. Here the interviewee refers to multiple impact assessments on national levels:

"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."

Altmetric

It is self-evident that services would monitor policies that affect their business model. This is particularly pronounced in the case studies of the services that emerge along the research cycle. Of particular interest here is the relevance of open science to both for-profit and non-commercial services. They perceive this as a dominant development to which they have to react (for some, for example DOAJ, open science is the reason for their existence) and which they aim to steer into their direction. It also appears that services endeavour to influence the growing field of research assessment . This indicates a certain anticipation and appropriation of scientific and social discourses.

Discourses

We define discourses as public discussions which the service takes into consideration and is (in)directly affected by. Here, too, the importance of open science is evident. This confirms the picture that we have outlined above with reference to policy developments. The respondents referred to discourses around open science 24 times, 8 times to commercialization and 4 times to other political debates. Services which relate to open science often align themselves to a movement and intend to change scientific practice through their actions and service. The open science discourse seems to influence the evolution of non-commercial research services in two ways. In the first place, it appears to be the point of departure for many of the business models. Secondly, they support narratives of open science discourse which reinforces their goal to change scientific practice through their service.

Interestingly, different understandings of open science stand out, especially with regard to commercialization. For commercial services, open science is a lucrative business model 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 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?"

Dryad

This shows 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 see open science as a liberation from commercial interests and articulate a desire to create alternative publishing infrastructure. 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.

Practice

As stated in the background section, for services to become infrastructures, they must be entangled with the social practice of research. In this context, this means that they are responsive to user and stakeholder needs. Accordingly, the results of this category are dedicated to the identification of user and stakeholder groups, the strategies used to involve them, and the resulting logic of technical implementation. Since the services can be considered infrastructures in the making, we further identified which problems they aim to solve in the research practice and thus how they explain their raison d'être.

Strategy

Our aim here was to identify how exactly services intend to become research infrastructures, i.e. which strategies they employ in order to connect to potential and actual users and stakeholders and how they implement changes.

Users

In our understanding, users are actors that use the services, in contrast to actors who do not use the service, but are still related to it (here stakeholders). Based on the responses of the interviewees, we identify eight user groups: These are public officials (8 codes), students (8 codes), research managers (5 codes), research funders (2 codes), librarians (9 codes), lay persons (6 codes) and finally authors (15 codes) and researchers (45 codes). The last two categories certainly overlap, to the extent that authors are typically researchers. However, this was not always explicitly stated by the interviewees and in rare cases, a service is used by non-academic authors as well. Nevertheless, it becomes clear that researchers are by far the most important user group.

The services offer functions that support researchers in every phase of their research, from idea generation (e.g., Paper Hive, Lazy Scholar), data management (e.g., figshare), preprint archiving (e.g., EarthArxiv) to peer review (e.g. Service 10) and dissemination/recognition (e.g. Altmetric, Browzine). This is not

[2] https://openaccess.mpg.de/Berlin-Declaration (last opened January 24 2020)
[3] https://www.budapestopenaccessinitiative.org/ (last opened January 24 2020)

surprising, especially since the selection of cases was based on the complete research life cycle. It is also noteworthy that many services support substantial parts of the whole research cycle, not just individual elements of it. For example, one of the services publishes articles, organizes peer review, and disseminates outputs. Services that are not built exclusively for researchers, on the other hand, generally serve monitoring, evaluation and orientation purposes. For example, Altmetric, a service that creates performance metrics mainly based on social media engagement with research outputs, considers research managers as users, since they typically monitor scientific performance.

We furthermore identify eight different strategies to adapt to user needs. These are teaching and training formats (12 codes), feedback systems (32 codes), advisory boards (3 codes), support teams (7 codes), lead user engagement (9 codes), data analytics (14 codes), prototyping (9 codes) and user surveys (18 codes). The effort to respond to user and stakeholder needs is reflected in a quote from a representative of Service 1 , a service that 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."

Service 1

We can generally differentiate between pull formats (when a service reaches out to users or monitors their behaviour), push formats (when users reach out to the service) and dialogue formats (when user and service engage in a dialogue) - see table 2.

Type of format	**Formats**	**# codes**
Pull	Data analytics, prototyping, user surveys	41
Push	Feedback systems, support team	39
Dialogue	Teaching & training, advisory boards, lead users	24

Table 2. Formats of user engagement

It is noticeable that services typically use different strategies at the same time to ensure usability. In addition, it can be stated that both, more digital-savvy (e.g., data analytics) and analog methods (e.g. advisory boards) are applied together. In particular, we find teaching formats of interest because they

illustrate how services anticipate that users may need to adopt newer practices and that services may thereby lower the entry threshold.

Stakeholders

We define stakeholders broadly as actors that are related to a service, but are not users nor do they work for the service. We identify four different stakeholder groups: These are publishers (12 codes), customers (22 codes), other services (17 codes), research institutions (11 codes), hosts (12) and data providers (21 codes). These form the stakeholder cosmos of the observed services.

The broad category customers refers to actors who purchase a service. These are generally publishers, libraries and research institutions as well as research funders. The category 'other services' typically refers to services that are of relevance for the business model of a service, for example because it supplies data. For example, QOAM—a marketplace for scientific and scholarly journals which publishes articles open access—uses journaltocs, a service that offers access to a database that indicates which journals are open access. They are in place to ensure the functioning of a service, typically as data sources, extended service features or online space (see 'hosts'). This shows that interfaces to other services are elementary and that services connect to other services to create and integrate into a decentralized research environment.

We identify four different strategies to anticipate stakeholder needs. These are: Customer outreach (8 codes), i.e. reaching out to potential customers and buyers; monitoring work (16 codes), i.e. strategically monitoring a discourse that is relevant to a service; awareness work (14 codes), i.e. create awareness for a problem that a service intends to solve, and mediation work (18 codes), i.e. mediating between different stakeholder groups.

Significantly, we find the stakeholder strategies particularly intriguing because they demonstrate what a service—in addition to usability work—is doing in order to become interwoven with the research environment. The strategy *customer outreach* refers to the relationship with existing or potential customers. *Monitoring work* and *awareness work* are related to the extent that they are intended to observe and subsequently influence the public discourse in favour of the problem that a service intends to solve. Both for-profit and non-commercial services attempt to influence a discourse in their favor, as the statements of a representative of Service 30 (for-profit) and DOAJ (non-commercial) demonstrate:

"We try to influence the trend. In order to influence a trend, we need to try and shift the NIH for example, the European Union and that we managed to do. […] This is how you change a trend."

Service 3

"At a national level we have been 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."

Directory of Open Access Journals

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., demonstrating to a potential customer, for instance a librarian, that a service is needed by users, for instance a researcher), between a service and other services (e.g., to be technically connectable), and finally between internal coders 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. But then, especially visually, the UI/ UX had to be very simple and plain. 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. That was a big challenge for us."

Knowledge Unlatched

Generally, we observe that non-commercial services express themselves more politically, mainly by ascribing to an open science movement and criticising the commercialization of research. It is especially non-commercial services that articulate problems in engaging stakeholders due to a lack of resources. This resonates with the results in the organization section, where we find that for-profit services generally spend more on sales and marketing. Noteworthy is the amount of mediation work that services engage in and which illustrates the conscious effort that is undertaken to connect a service with a research

environment and integrate a service into the social practice of research—in short: making it a research infrastructure.

Implementation

This category refers to the modes of technical implementation, i.e. how a service adapts technological configurations to user and stakeholder needs. Here we coded references that illustrate how changes are implemented. We are able to differentiate two modes of implementation:

Phased implementation (6 codes), i.e. screening user needs first and then build the service accordingly; iterative implementation (15 codes), i.e. constantly screening user needs and adapting the service continuously.

Generally, we observe that mainly non-commercial services refer to phased implementation whereas for-profit services exclusively refer to iterative implementation. Below there are two example quotes, the first referring to iterative implementation, the second to phased implementation (see table 3):

Iterative implementation	**Phased implementation**
*"[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."* *Scholarcy*	*"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**."* *Moving*

Table 3. Iterative implementation vs. Phased implementation

This result is significant, since it potentially reflects the funding logic of many non-commercial services, who typically describe implementation in consecutive work phases, whereas for-profit services appear to have to search for exposure earlier and permanently. It should also be noted that all the infrastructure projects we observe are software projects. For software production, iterative (or agile) implementation has become somewhat of a production standard, and it remains to be discussed to what extent phased implementation is suitable for creating software products that meet current user needs.

Motivation

We define the category motivation as the reasons why the services exist, i.e. the problems that they aim to solve and the added value that they promise. Even though we coded problems and added value separately, the sub-categories almost perfectly mirror each other. Generally, we find eight different types of added value (see table 4).

Added Value (#codes)	**Explanation**	**Example**
Access (35 codes)	Providing or improving access to research outputs	Supporting open access to research articles through repositories (e.g. EarthArXiv, DOAJ).
Dissemination (31 codes)	Disseminating research outputs to different publics.	Supporting new formats for research communication (e.g. Browzine).
Transparency (18 codes)	Increasing the comprehensibility of the research process.	Facilitating data storing and management (e.g., figshare).
Orientation (30 codes)	Filtering and providing an overview of research topics.	Curating open access journals (e.g., DOAJ)
Compliance (12 codes)	Supporting the compliance to rules and regulations.	Providing structured guidelines (e.g., Service 6).
Recognition (17 codes)	Providing recognition for practices and outputs.	Using alternative metrics for practices and outputs (e.g., Altmetrics, Publons)
Collaboration (38 codes)	Facilitating collaboration among different actors.	Providing tools for sharing and communicating (e.g., Paper Hive)

Efficiency (33 codes)	Increasing the efficiency of the research process.	Mining content from large amounts of data (e.g, moving)

Table 4. Subcodes for "Motivation"

Accordingly, the problems that the services aim to solve are missing access to outputs (4 codes) and lack of hosting services (4 codes), missing recognition for new and overseen practices and outputs (9 codes), missing transparency of the research process (4 codes), lack of orientation on current research topics (7 codes), missing compliance to rules and regulations (8 codes) and inefficient workflows (15 codes). The only problem identified which was not explicitly mirrored in the added value category was the commercialization of research (10 codes). Here the respondents, both from for-profit and non-commercial services, articulated discomfort about the commercialization of research by large publishers.

Here again, it is noticeable that many of the services align with an overall open science discourse. Almost all respondents refer to improvements through digital technologies in their problem analyses and value propositions. In many instances they contrast the added value of their services with the deficits of 'old' practices. 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."

Service 1

Among the respondents who refer to the potential for improvements through digital technologies, old infrastructures (e.g., journal-based scholarly communication) are described as outdated and criticised for being slow, inefficient and insufficient. This shows that, in part, services are competing against established players and seek to replace them. It also brings to light the aforementioned moment of infrastructure inversion so that they articulate a breakdown of the established infrastructural order for scholarly communication.

Organization

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

Design

In this category, we subsume features of the organizational design. A common feature is decentrality - for example, seven respondents indicate that they engage volunteers or employ freelancers. These external team members are described as indispensable for the functioning of the services (e.g., Pre-prints or DOAB). This makes it possible to react to demands situatively and still remain operational.

Another feature of the organizational design are strategic partnerships. Typical cooperation partners for the services are research organizations and libraries, to which members of a service are affiliated with (e.g. Service 8). In some cases, services merged (e.g. Sharelatex, Dryad and the Dash platform) or were partly acquired by a larger service (e.g., figshare by Digital Science).

Team background

The services that we observed typically employ staff members with the following profiles: librarians (9 codes) and information scientists (3 codes), researchers with field-specific knowledge (15 codes), developers (23 codes), data scientists (8 codes; which may refer to either former researchers with coding skills or specifically trained technical staff), sales and marketing staff (12 codes). For-profit services emphasize the importance of having a strong technical team. Several for-profit research services claim that at least one third or even half of the staff has technical skills and a technical background, including Symplectic, a scholarly information management software:

"Obviously, the main bulk of the people that work for us are developers, not all of whom are graduates in computer studies. We have many people who've joined us with physics degrees, maths degrees. Quite a few of our developers have PhDs and that's true of certain other parts of the organization."

Symplectic

Many services acknowledge the importance of finding skilled staff, in particular 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."

DCC

Especially for-profit entities have designated sales teams that make up a significant part of the staff. These teams make sure that a service fulfills user and customer needs and thereby helps adapting a service by working closely with the developer team (e.g., Scholarcy). This resonates with the observation that for-profit services typically implement changes iteratively.

The observations in this category are of particular pertinence. They show that for-profit services put more effort into adapting their product by investing more resources in development and sales compared to non-commercial services. There are also indications that non-commercial services have problems finding technical expertise. This may be due to the fact that the salaries for coders in non-commercial services (mostly at scientific institutions) are lower than in the private sector and that there are limited reputative gains for infrastructure work in academia (compared to research work). 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 this or only for a short period at the beginning of the project.

Financing models

In this category, we describe different funding mechanisms that are applied among research services. We identified the following types among non-profit organizations (66 codes): institutional funding (18 codes), public funding (17 codes), fees (5 codes), donations (8 codes) and out of their own pocket (4 codes), meaning that the service is funded by the founder/s. Profit-oriented services (49 codes) mentioned the following types: subscription model and licensing (26 codes), individual payments (5 codes) and private investments (13 codes). In most cases, we cannot ascribe only one funding model to a service. Several services have mixed funding models, or at least emphasize the intention to seek other/additional sources of funding, as the interviewee from the Director of Open Access Books (DOAB) indicates.

"Our goal was that we would be sustainable based on services that we would provide to our stakeholders and we haven't quite managed that until now. We started out with subsidies from some Dutch institutions and gradually we have increased our income from services. About 70% of our income is now service based. We get paid by publishers and research founders and hopefully in the near future, also libraries

for what we do. We're not quite sustainable so we are still looking at the developing and modifying our business plan".

DOAB

Several non-profit research services are funded by public funders, for example by the German Research Foundation DFG. DRYAD Digital Repository was initially funded by a start-up grant by the US National Science Foundation. Non commercial services often point out the necessity to stick to the foreseen implementation phases and mention that after their funding runs out, they will have to look for a sustainable funding model in order to secure further existence of the service. Noticeably, some of these services point out a lack of funding opportunities after the funding ends. Some state that planning for a sustainable business model was not part of the application process for receiving such 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"

DRYAD

A common way to sustain financially is similar for non-profit and profit-oriented organizations. Both tend to introduce one of the following options:

- subscription fees that can be paid either by individual researchers or institutions (e.g. DRYAD)
- membership fees
- publishing fees such as data publishing charges DPCs and article publishing charges APCs (e.g. Dryad, QUOAM)

Most respondents explained that these payments are made by institutions, only in rare cases individual researchers bear the cost. Both nonprofits and commercial entities used seed funding at the beginning. Other than non-profits that get initial funding from public funders, profit-oriented services often rely on investments from companies like Digital Science at the start. Several companies started with seed investment (e.g. Tetrascience), 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 funders.

Discussion

In conclusion, and with regard to the research interest, namely understanding the making of research infrastructures, some common features can be identified. All observed services are highly sensitive to political, legal and social developments and try to anticipate them and use them to their advantage. All the observed services deal with privacy and copyright regulations. Most adhere to and try to influence policies that are relevant to their business models (e.g., open science policies on a national and supranational level). Most of the problems that the observed services try to solve stem from shortcomings of the established system for scholarly communication (e.g., lack of access to articles, lack of recognition for alternative scholarly outputs). Open science is here regarded as an umbrella term for solutions to these problems. All services implement strategies to penetrate research practice, engage with and convince stakeholders and set favorable boundary conditions (e.g., lobbying). Many of them apply decentral organizational designs and draw their staff from a research and library background. Developers are especially important for the functioning of the services. The business models applied are typically mixed, with slight differences between commercial and private entities.

Of particular interest are the differences between the services examined, especially between commercial and non-commercial services. At various instances in the case studies it becomes clear (partly implicitly) that the source of funding determines the practice of making research infrastructures. What we are dealing here is a recursive effect on infrastructure making that does not arise directly from the relationship with users, but with relevant stakeholders (here funders). In particular, for non-commercial services, the logic of public funders can be challenging, as several observations suggest: (1) Publicly financed infrastructure projects typically refer to phased implementation (i.e. assessing user needs first, then implement) whereas commercial infrastructure projects typically showcase an agile implementation logic (i.e. constantly assessing user needs and implement changes). The phased implementation logic could be a result of the funding logic for non-commercial infrastructure projects. Funders in this realm typically expect a project plan with clearly defined and chronologically ordered work packages. (2) Non-commercial services articulate problems with finding qualified technical staff. This could be the case because resources for highly qualified non-scientific staff (e.g. coders) are typically limited in public research infrastructure funding. (3) Representatives from non-commercial infrastructure projects articulate problems in finding follow-up financing. This does not come as a surprise: only a few funding agencies ask for a business model when an application is submitted, or offer opportunities for follow-up funding for promising projects. In addition, it can also be assumed that staff in academia (who typically apply for research

infrastructure projects) rarely have sufficient marketing and business skills to keep a service alive, outside the realm of public funding. This assertion is supported by the observation that non-commercial projects rarely have sales personnel at their disposal, although staff with these skills could occupy an important interface position and articulate the needs of their users internally.

We believe that this can be to the long-term detriment of non-commercial research infrastructure projects. While agile implementation seems to be the standard for many novel software products, due to rapidly changing user needs, this logic seems to be at odds with the logic of public research funders. Moreover, programming skills seem to be at the core of successful research infrastructure projects. However, highly qualified programmers orient themselves towards the free market, especially since they do not receive adequate recognition for their work on infrastructure projects (which in turn motivates highly qualified researchers not to work in the free market). Furthermore, it appears difficult for non-commercial infrastructure projects to be competitive in the long run if follow-up funding is scarce, considering that a change of business model is difficult even for services that were already on the free market before.

Conclusion

To a certain extent, a reorganisation of scholarly communication is taking place in such a way that previously closed or implicit processes (such as the management of data, the review of publications) can be displayed online. Previously, the scientific process culminated in the publication of articles and books; now there are many neuralgic points where alternative scientific outputs (e.g. data, software) can be exchanged and practices can be reorganized (e.g. peer review). In this respect, infrastructure can be understood as a lens through which the social reorganisation of science can be observed. It can be observed that the traditional players in the market for scholarly communication (publishers and their journals) remain and partly reinvent themselves (e.g. through the purchase of new services) and new players are entering the market. These try to provide added value and to become infrastructures themselves. The observed services try to anticipate and accelerate this change. With regard to the changes in scholarly communication patterns, there is a valid market potential for these new players and their services. The shift of scholarly communication to digital is an example of infrastructure inversion. The story of current research infrastructure projects is thus a story of the old and the new. The old refers to the journal-based communication system and the world of large scholarly publishers who (still) manage large parts of the scientific value chain. The new, on the other hand, is the world of online services that surface one after the other and which address every step of the research cycle (and are partly also owned by large publishers). They position themselves against the old world, which is in their view inefficient, opaque and

inaccessible. The vehicle to do this is open science (Fecher & Friesike 2014). Along the development lines of open science—access, transparency, collaboration—they offer services that are conducive to an opening of science. However, it remains unclear what exactly constitutes open science and to what degree it relates to the infrastructure for or just the outputs of research. In this respect, it is worthwhile to take a look at the transformative deals on open access (e.g. project DEAL in Germany), in which offsetting agreements are concluded with established publishers, under which, against payment of a large sum of money, all publications from these publishers are to be made open access. In this case, the sustainable development of non-commercial infrastructures is more or less abandoned in favour of a licensing solutions that only apply to articles but not to the underlying infrastructure. In this way the often criticised dependence on large publishers is reproduced to a certain extent to the digital world. Which poses the question how transformative an agreement truly is, whose result is the reproduction of the status quo in the digital age. It shows that digitization might be a grand moment of inversion but does not necessarily imply that scholarly communication is fundamentally transformed.

Practical policy conclusions can also be drawn from this research, particularly with regard to the development of non-commercial infrastructures. At present, non-commercial services have a competitive disadvantage over commercial services in that they follow a largely anachronistic linear production logic, have to get by with few programmers and rarely receive sustainable follow-up funding (and are thus basically built under the wrong premises). We propose that the system of funding for infrastructure be fundamentally reformed so that agile production logics are enforced (e.g. where changes in the work plan are made informally) and competitive follow-up funding is provided. In addition, there should be either a higher monetary remuneration or more recognition for work on infrastructure for research, for example. We propose that, given its importance for research, infrastructure work should be considered the fourth mission alongside research, teaching and transfer. In general, programming needs to play a much greater role in the training of librarians; after all, it is often librarians who are responsible for setting up and maintaining non-commercial research infrastructures. If this does not happen, public-private partnerships would be one possibility and the partial renunciation of funding for public research infrastructures the other. The latter, of course, contradicts the original idea of open access, which refers not only to outputs but also to infrastructure. In the end, we conclude that looking at the making of infrastructures allows understanding changes in the social organization of science as well as the external and internal forces that influence them.

References

Anderson, S., & Blanke, T. (2015). Infrastructure as intermeditation – from archives to research infrastructures. *Journal of Documentation*, *71*(6), 1183–1202. https://doi.org/10.1108/JD-07-2014-0095

Barjak, F., Eccles, K., Meyer, E. T., Robinson, S., & Schroeder, R. (2013). The emerging governance of e-infrastructure. *Journal of computer-mediated communication*, *18*(2), 113–136.

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

Blanke, T., & Hedges, M. (2013). Scholarly primitives: Building institutional infrastructure for humanities e-Science. *Future Generation Computer Systems*, *29*(2), 654–661. https://doi.org/10.1016/j.future.2011.06.006

Bowker, G. C. (1994). *Science on the run: Information management and industrial geophysics at Schlumberger, 1920-1940*. MIT press.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences* (First paperback edition). The MIT Press.

Bowker, G. C., Timmermans, S., & Star, S. L. (1996). Infrastructure and organizational transformation: Classifying nurses' work. In *Information technology and changes in organizational work* (S. 344–370). Springer.

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

Carse, A. (2012). Nature as infrastructure: Making and managing the Panama Canal watershed. *Social Studies of Science*, *42*(4), 539–563.

David, P. A. (2005). *Towards a cyberinfrastructure for enhanced scientific*. University Library of Munich, Germany.

De Roure, D., Jennings, N. R., & 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., Bowker, G. C., Borgman, C. L., & Burton, M. (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Deep Blue.

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In S. Bartling & S. Friesike (Hrsg.), *Opening Science* (S. 17–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_2

Goldthau, A. (2014). Rethinking the governance of energy infrastructure: Scale, decentralization and polycentrism. *Energy Research & Social Science*, *1*, 134–140. https://doi.org/10.1016/j.erss.2014.02.009

Hanseth, O., & Monteiro, E. (1997). Inscribing behaviour in information infrastructure standards. *Accounting, management and information technologies*, *7*(4), 183–211.

Hanseth, O., Monteiro, E., & Hatling, M. (1996). Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology, & Human Values*, *21*(4), 407–426.

Harvey, P., & Knox, H. (2012a). The Enchantments of Infrastructure. *Mobilities*, *7*(4), 521–536. https://doi.org/10.1080/17450101.2012.718935

Harvey, P., & Knox, H. (2012b). The Enchantments of Infrastructure. *Mobilities*, *7*(4), 521–536. https://doi.org/10.1080/17450101.2012.718935

Haug, C. (2013). Organizing spaces: Meeting arenas as a social movement infrastructure between organization, network, and institution. *Organization Studies*, *34*(5–6), 705–732.

Hepsø, V., Monteiro, E., & Rolland, K. H. (2009). Ecologies of e-Infrastructures. *Journal of the Association for Information Systems*, *10*(5), 2.

Hewitt, C. (1988). Offices are open systems. In *Readings in distributed artificial intelligence* (S. 321–329). Elsevier.

Hey, T. (2005). Cyberinfrastructure for e-Science. *Science*, *308*(5723), 817–821. https://doi.org/10.1126/science.1110410

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

Isaksson, K., Richardson, T., & 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. https://doi.org/10.1016/j.eiar.2009.01.007

Jensen, C. B. (2007). Infrastructural Fractals: Revisiting the Micro—Macro Distinction in Social Theory. *Environment and Planning D: Society and Space*, *25*(5), 832–850. https://doi.org/10.1068/d420t

Jewett, T., & 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), 246–275.

Kaltenbrunner, W. (2015). *Reflexive inertia: Reinventing scholarship through digital practices*. Faculty of Social and Behavioural Sciences, Leiden University.

Larkin, B. (2013). The Politics and Poetics of Infrastructure. *Annual Review of Anthropology*, *42*(1), 327–343. https://doi.org/10.1146/annurev-anthro-092412-155522

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Liu, A. (2018). Where Is Cultural Criticism in the Digital Humanities? In M. K. Gold (Hrsg.), *Debates in the digital humanities*. Univ Of Minnesota Press. https://dhdebates.gc.cuny.edu/

Mayring, P. (2004). Qualitative content analysis. *A companion to qualitative research*, *1*, 159–176.

Morita, A. (2017). Multispecies infrastructure: Infrastructural inversion and involutionary entanglements in the Chao Phraya Delta, Thailand. *Ethnos*, *82*(4), 738–757.

Mueller-Langer, F., Fecher, B., Harhoff, D., & Wagner, G. G. (2019). Replication studies in economics—How many and which papers are chosen for replication, and why? *Research Policy*, *48*(1), 62–83. https://doi.org/10.1016/j.respol.2018.07.019

Pipek, V., & 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., & 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), 521–556.

Ribes, D., & Finholt, T. A. (2009). The long now of infrastructure: Articulating tensions in development. *Journal for the Association of Information Systems (JAIS):*, *10*(5), Article 5.

Ruhleder, K., & King, J. L. (1991). Computer support for work across space, time, and social worlds. *Journal of Organizational Computing and Electronic Commerce*, *1*(4), 341–355.

Schmidt, K., & Bannon, L. (1992). Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW)*, *1*(1–2), 7–40.

Star, S. L. (1999). The ethnography of infrastructure. *American behavioral scientist*, *43*(3), 377–391.

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Star, S. L., & Strauss, A. (1999). Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, *8*(1), 9–30. https://doi.org/10.1023/A:1008651105359

Wouters, P. (2014). The citation: From culture to infrastructure. In B. Cronin & C. R. Sugimoto (Hrsg.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*. The MIT Press.

Yates, J. (1993). *Control through communication: The rise of system in American management* (Bd. 6). JHU Press.

Appendix

School of thought	Central Assumption	Authors	Example quote
practice	Infrastructures are a routinized and relational set of human interactions.	Bowker (1994), Star & Ruhleder (1996), Star (1999), Star & Strauss (1999), Hughes (1983), Kaltenbrunner (2015), Lave & Wenger (1991), Ruhleder & King (1991), Jensen (2007), Jewett & Kling (1991), Wouters (2014), Schmidt & Bannon (1992); Bowker & Star (2000), Isaksson, Richardson & Olsson (2009)	"(...) [Infrastructures] are the taken-for-granted context that enables our life and work. Infrastructures are multi layered and complex and cannot be constructed top-down. Instead they evolve bottom-up." (Wouters 2014 ,p.62)
materiality	Infrastructures are a set of technical artifacts that shape social practice.	Bowker, Timmermans & Star (1996), Harvey & Knox (2012), Star (1999), Hanseth, Monteiro, Hatling (1996), Hanseth & Monteiro (1997), Carse, A. (2012), Larkin (2013), Morito (2017), Pipek & Wulf (2009), Pollock & Williams (2010), Yates (1989), Ruhleder & King (1991)	"Infrastructures are material forms that allow for the possibility of exchange over space. They are the physical networks through which goods, ideas, waste, power, people, and finance are trafficked." (Larkin 2013,p.327)

environment	Infrastructures are adaptive networks of services forming a loosely-coupled production line.	Edwards, P. N. et al. (2013), Hewitt (1988), Hey (2005), Blanke & Hedges (2013), Anderson & Blanke (2015), De Roure, Jennings & Shadbolt (2001), David (2005), Hepsø, Monteiro & Rolland (2009), Ribes, Finholt (2009), Bilder et al. (2015), Anderson, S., & Blanke, T. (2015), Camarinha-Matos (2002), Haug, C. (2013), Barjak et al. (2013), Goldthau, A. (2014)	"[I]nfrastructures are ecologies or complex adaptive systems; they consist of numerous systems, each with unique origins and goals, which are made to interoperate by means of standards, socket layers, social practices, norms, and individual behaviors that smooth out the connections among them. This adaptive process is continuous, as individual elements change and new ones are introduced — and it is not necessarily always successful (Edwards 2013,p.3).

Table 5. Infrastructure schools

Cases	Description of services	Research cycle	Funding	
			Profit	Non-profit
Altmetric	Data science company that tracks where published research is mentioned online.	Share/Impact	■	
Service 8	Open access repository for preprints.	Discover		■
Bepress	Offers products and services to support scholarly communication (e.g. institutional repository and publishing software).	Share/Impact	■	
Browzine	Helps finding, reading and monitoring scholarly journals.	Discover	■	
Digital Curation Center (DCC)	Supports research organisations in building capabilities and skills for research data management.	Gather, analyze		■
Directory of Open Access Books (DOAB)	Discovery service that points to open access books that are deposited elsewhere.	Write and publish		■
Directory of open access journals (DOAJ)	An online directory that indexes and provides access to open access and peer-reviewed journals.	Write and publish		■
Dryad	A curated general purpose repository that makes data discoverable, freely reusable and citable.	Discover, gather, analyze		■
EarthArXiv	A preprint server and volunteer community devoted to open scholarly communication.	Write and publish		■
Service 3	A publisher of services for life scientists and clinical researchers. It includes e.g. an open access, open peer-review publishing platform.	Write and publish	■	
Figshare	An online open access repository where researchers can preserve and share their research outputs, including figures, datasets, images, and videos.	Share/Impact	■	
Knowledge Unlatched	An online marketplace that provides libraries and institutions worldwide with a central place to support open access collections.	Discover/Write and publish	■	
Service 2	A cloud-based platform, through which researchers can accelerate and broaden the positive impact of their research outputs.	Share & Impact	■	
Lazy Scholar	A browser extension that provides metrics, quick citation and sharing links for a selected scholarly output.	Discover, gather, analyze		■
Moving	Moving is a platform for research, collaboration and training for a wider public to improve information and data literacy. It includes e.g. a database, visualisation and analysis tools.	Share/Impact		■

Service 1	An open-source software for the management of peer-reviewed journals.	Discover		
Open Access Button	A browser application that helps to access closed access academic articles by requesting them from authors.	Write and publish		
Service 6	Provides a persistent digital identifier (an iD) that you own and control, and that distinguishes you from every other researcher.	Share & Impact		
Service 5	An online LaTeX editor for collaborative work on texts with a collection of templates.	Write and publish		
Service 9	A web platform that offers a framework for collaborative reading and commenting on academic publications.	Write & Publish, Share & Impact		
Preprints	A multidisciplinary preprint platform.	Write and publish		
Publons	A website that provides a service for academics to track, verify and showcase their peer review and other contributions for academic journals.	Write and publish		
Service 10	A website that allows users to discuss and review scientific research.	Write and publish		
Quality Open Access Market (QOAM)	A market place for scientific journals which publish articles in open access and provides information about the quality of these journals.	Write and publish		
RDMO	A service that enables institutions and researchers to plan and carry out their management of research data.	Gather & analyze		
Service 4	A collaborative data science platform.	Gather & analyze		
Researcher	A mobile app to discover research content, enables users to browse through more than 200 journals.	Discover		
Scholarcy	An online summarizing tool that converts long articles into summary flashcards.	Discover & gather		
Symplectic	A software that helps researchers, librarians, and their institutions collecting, managing, analysing and showcasing their research.	Share & Impact		
Tetrascience	A service that offers data integration, lab monitoring and lab scheduling for R&D labs.	Gather and analyze		
Viziometrics	A service that studies visual information (more than 8 million figures from PubMed) of academic articles.	Discover		
Xinder	A mobile app that helps discovering academic publications in a mobile environment.	Discover & Share		
Zenodo	An open access repository that allows researchers to deposit data sets, research software, reports and other digital artifacts.	Write & publish		

Table 6. Research infrastructures