ELEPHANT
IN THE LAB

SHORT ANALYSIS

Challenges in Building Innovative, Sustainable and Open Research Infrastructures

Title	Challenges in Building Innovative, Sustainable and Open Research Infrastructures
Authors	Ilire Hasani Mavriqi[2], Nataliia Sokolovska[1], Tony Ross-Hellauer[2], Benedikt Fecher[1]
Author affiliation	[1] Alexander von Humboldt Institute for Internet and Society, Berlin, Germany [2] Technical University Graz, Austria
Author bios	Ilire Hasani-Mavriqi is a post-doctoral researcher at the Institute of Interactive Systems and Data Science (Graz University of Technology) and deputy group leader of the Open and Reproducible Research Group (ORRG). She has a PhD. in Computer Science from Graz University of Technology. Her research interests include Social Computing, Complex Networks, Agent Based Modelling, Computational Social Science, Web Science and Research Data Management. Nataliia Sokolovska is a researcher at the programme 'Knowledge & Society' at the Alexander von Humboldt Institute for Internet and Society (HIIG). She has a background in international relations and worked for several years in journalism and PR. Tony Ross-Hellauer is leader of the Open and Reproducible Research Group at Graz University of Technology, Senior Researcher at Know-Center GmbH, and Editor-in-Chief of the MDPI open access journal 'Publications.' His research focuses on a range of issues related to open science evaluation, skills, policy, governance, monitoring and infrastructure. Benedikt Fecher is the programme director of the research programme Knowledge Dimension and heads the Open Science research group at the Alexander von Humboldt Institute for Internet and Society.
Author social links	Benedikt Fecher: LinkedIn - ORCiD - ResearchGate - Mail Ilire Hasani Mavriqi: LinkedIn - ResearchGate - Mail Nataliia Sokolovska: Mail Tony Ross-Hellauer: LinkedIn - ORCiD - Mail
Date published	18 February 2020
DOI	10.5281/zenodo.3671740

Cite as (APA)	Fecher, B., Hasani-Mavriqi, I., Sokolovska, N., Ross-Hellauer, T. (2020). Challenges in Building Innovative, Sustainable and Open Research Infrastructures. *Elephant in the Lab*. DOI: 10.5281/zenodo.3671740

Introduction

The way science is conducted is changing. The internet and other digital means alter the scale and pace of data-driven research and collaborative possibilities. Most modern scholarly research now relies to some extent on digital research infrastructures. Recognition of this has led to large-scale, public-funded initiatives to develop digital infrastructures for research at the institutional, national and international levels. Research funding focuses in particular on Open Science and thus on access to scientific products and the transparency of the research process. Despite the focus on infrastructure development, however, publicly funded infrastructure projects often find it difficult to prevail in the long term. Innovative, user-friendly and open research infrastructures remain urgently needed to make scientific knowledge available on a long-term and sustainable basis.

In this post, we present the results of a stakeholder-focused workshop at the Open Science Fair 2019 in Porto (September 16-18). The Open Science Fair is one of the leading conferences for public research infrastructure projects in Europe and is funded by the European Commission. The 2019 event was organized by the EC projects OpenAIRE, FIT4RRI, the EOSC Secretariat and FAIRsFAIR. Hence, many attendees are working directly or indirectly on the creation of research infrastructures for the 21st century and know the problems and opportunities first-hand. In the interactive workshop, we garnered the views of more than 50 expert participants on the current barriers to developing innovative and sustainable research infrastructures. Key areas for discussion were how resources are allocated and managed, how people are motivated and teams are organized. In synthesising these results, we here identify key problem blocks and preliminary solutions (with commentary on their feasibility) as proposed by participants.

Problem Areas

Our discussion was ordered around three thematic tables with different focuses:

- **Political-economic**: This table focused on topics such as science policy, funding and governance of research infrastructure
- **Technical**: This table focused on topics such as technical maintenance, usability and interoperability of services
- **Social**: This table focused on topics such as incentivisation and team coordination.

These problem areas were discussed in a world café scenario where the participants formed groups that moved from table to table, discussing each issue in turn during the course of the workshop. Each table had a moderator and a notekeeper. In the following we will report the main results of each area.

Political-economic Dimension

- **Fragmentation of services:** Participants identified the sometimes chaotic and competitive emergence and evolution of research infrastructure projects as key. A common theme was competition vs. collaboration: Participants felt that many publicly funded research infrastructure services are in competition with each other (for resources, users, etc). This was felt to stem from competing national and supra-national funding initiatives as well as different national Open Science priorities and mandates. According to the participants, this could lead to a fragmentation of services (i.e., lack of interoperability, use of competing standards) and the duplication of effort. While some said that competition could help to design competitive, user-focused services, many argued that national funding initiatives should be better coordinated, in particular to ensure the long-term interoperability of services.
- **Funding and Sustainability**: Participants also identified an issue with the funding logic for (mainly software) services along the whole funding cycle, including a) funding decisions, b) adjustments during the implementation process and c) the long-term sustainability of services. Participants expressed concerns that while seed funds for initiatives, there is a lack of funding opportunities for the long-term maintenance of services. In this context, some expressed doubts about the way funding decisions are made; they expressed concerns that user involvement and long-term sustainability are often neglected in funding schemes and decisions. Some referred to the (administrative) problem of the inflexibility of project work-plans - where working agile (e.g., in order to respond to changing technological possibilities or user needs). An explanation for this problem area is the bureaucratic logic of public funds that is many respects at odds with the way software is built. The main proposed solution was to remove barriers of bureaucracy and embrace a more agile funding logic, as well as to specifically allocate (non-project-based) funding opportunities for long-term maintenance according to schemas more suited for service development.

Technical Dimension

- **Interoperability:** The technical interoperability of research infrastructures was identified as another crucial challenge. Semantic interoperability, in particular, was identified as a seemingly intractable issue not solvable, even within industry. Existing research infrastructures are often tailored to institutional needs. There is a lot of variation between

research institutions in terms of the amount and complexity of data, data collection and analysis, departmental structures and disciplinary context. Thus, the interoperability of existing research infrastructures is very limited. Metadata and standards are often very discipline-specific and difficult to transfer across fields, necessitating cumbersome discussions on interoperability. To address these challenges, participants suggested many possible solutions. Firstly, to learn from best practices, international guidelines and technology standards (TCP/IP) that worked. Secondly, to focus more on the balance between general, cross-discipline and discipline specific solutions and standards. The European Open Science Cloud (EOSC) was mentioned as an exemplar of the latter. Thirdly, to follow a coordination strategy, operating on EU & national level, i.e., discipline specific clusters that operate on EU level, but are implemented on national level. Finally, to provide training on ontologies for non-technical staff.

- **Lack of User Engagement:** The governance of publicly funded infrastructure services, especially the relationship between users (mostly researchers) and the builders and funders of these services (often libraries through public project funds, e.g., from the European Commission as part of the EOSC) was also identified as a problem area. Many participants expressed concerns that services are too-often designed in a top-down, rather than bottom-up, fashion. Resulting infrastructures can be isolated solutions based on very specific requirements that are ill-suited for broader use. Users, for whom these services are nominally built, are excluded from governance while projects and funders are steered more by political currents than community needs. Some participants articulated frustration with a lack of collaboration between different stakeholders and cooperation across disciplines. They expressed concerns that this leads to services that lack vision and a clear understanding of community needs. Many participants expressed the wish that infrastructures be closer to the research cycle, i.e. enhance the efficacy of data production, data use and data curation. Suggested solutions involved including users in the creation of infrastructure projects (e.g., through constant beta-testing, co-creation, prototyping) and making sure that users are part of the governance structure (e.g., in advisory boards, through surveys) once the service is running.

Social Dimension

- **Lack of Rewards for Infrastructure Work**: Infrastructures for scholarly communication rely heavily on good software developers. However, acquiring and retaining developer talent is difficult for public-sector players, as the payment they get in public research infrastructures is far less than in the private-sector, where their skills are also in demand. Furthermore, the effort for infrastructure work is not credited in the traditional academic value system. For example, code is not considered a valued research output and does not feed into the reputation record. Participants suggested the need to provide competitive salaries in order to make infrastructure work more attractive in academia; this refers

especially to technicians (e.g. coders) whose expertise is highly demanded in the commercial sector. At the same time, excellent infrastructure work should should bring reputational benefits in academia.

- **Cultural Diversity:** All participants agreed: research infrastructures require a very mixed team with expertise in diverse areas ranging from scholars who know the respective research field, and librarians with competencies in information management, to software developers who create and maintain technicalities. Furthermore, it was mentioned that international infrastructures mostly operate in a decentralized manner with lots of volunteers and staff members contributing from different parts of the world. Two problems appear in this context: there is no clear-cut understanding of which experts are needed to set up and run an infrastructure. This makes it complicated to launch an infrastructure for those who have an innovative idea. Then, there is no clear description of the responsibilities and functions the team members should have. This makes it difficult to manage expectations, day to day work and recruit new staff in decentralized and international teams. A consent on a core team constellation and a set of responsibilities could make the management of diverse and international teams easier, more efficient and more understandable for newcomers.

Way Forward

Services become research infrastructures when they integrate into research practice; when they do not stick out anymore and recede into the background of everyday activities. This is also why break-downs of established infrastructures are so annoying – because many people depend on them to do something (think, for example, of train cancellations). Most research infrastructures that are built today are cyberinfrastructures; they consist of information technology that provides particularly powerful and advanced capabilities. It is the underlying infrastructure that enables novel research practices (e.g., data sharing) and at the same time responds to it. In that sense, services that emerge along the research lifecycle aim to become research infrastructures and thereby penetrate research practice. Open Science, in this regard, can be understood as the *best possible use* of technological infrastructure for the scientific value creation. A key to support truly open science is therefore the development of open, innovative research infrastructures that are able to compete against commercial services. In this workshop, we were able to capture the view of people working in and around publicly funded research infrastructures, mainly working on projects funded by the European Commission.

Our workshop revealed a couple of important problem areas that call to rethink the current practice of funding and governing research infrastructures. This concerns the rigid funding logic for research infrastructures, missing adoption of common standards, the lack of recognition for infrastructure work and thereby the lack of talent for infrastructure development. The results of the workshop provide insights into a logic for public infrastructure development that is in many respects at odds with user needs and standards for software development and urge funders and

policy makers to rethink the funding of research infrastructures. Many of the issues here can be related to a current lack of recognition of this central fact, which diminishes the importance of infrastructure work and denies us the utopia of truly innovative, non-commercial research infrastructures.