

# Why openness makes research infrastructure resilient

Helena Cousijn, DataCite, Amsterdam, Netherlands ([hcousijn@datacite.org](mailto:hcousijn@datacite.org))

ORCID – <https://orcid.org/0000-0001-6660-6214>

Ginny Hendricks, Crossref, Oxford, UK ([ghendricks@crossref.org](mailto:ghendricks@crossref.org))

ORCID: <https://orcid.org/0000-0002-0353-2702>

Alice Meadows, NISO, Baltimore, USA ([ameadows@niso.org](mailto:ameadows@niso.org))

ORCID: <https://orcid.org/0000-0003-2161-3781>

**Corresponding author**

## Key Points

- Open research infrastructure provides the building blocks of scientific progress, which must be available to everyone, with no barriers to access.
- Organizations enabling open research infrastructure must endorse these fundamental principles: equity, value, trust, interoperability, sustainability and community governance.
- Finding ways to invite co-creation and community participation engenders a strong sense of 'buy-in' and is therefore essential to developing successful research infrastructure.

## Introduction

First, let us define what we mean by open research infrastructure. In this article, we are defining research infrastructure as the core elements — persistent identifiers, metadata schemas, standards, and so on — that enable the creation of tools and services used by researchers. Such tools and services, for example, manuscript submission systems, content hosting platforms, profile systems, current research information systems/research information management systems, tracking and analysis tools, and so on, are layers built on top of the infrastructure. In this article, we are talking about the building blocks on which these services are constructed: foundational infrastructures. Specifically, we are talking about what Bilder, Lin, and Neylon (2016), refer to as SIMA:

- Storage: places to put stuff that is generated by research, including possibly physical stuff
- Identifiers: means of uniquely identifying each of the – sufficiently important – objects the research process created over time
- Metadata: information about each of these objects
- Assertions: relationships between the objects, structured as assertions that link identifiers

By *open* research infrastructure, we mean that these building blocks must be available to everyone, with no barriers to access. This includes, for example, providing open-source code, software, open data, openly available documentation, and tools and services that are free (both gratis and libre) to researchers. Typically, research infrastructure providers are not-for-profits; both local and global, community-specific and cross-disciplinary. Most of these organizations were established starting from the 1990s when the move to digital made open access—and open research—achievable as well as desirable, and ensuring that open research was supported by an open research infrastructure became essential (Martone, 2015)). Recent developments in science because of the COVID-19 pandemic have emphasized the need for open research and open infrastructure to maximize efficiency (see Homolak et al., 2020).

Next, what do we mean by *resilience*? Most dictionary definitions talk about this in terms of an organization or individual's ability to successfully adapt to change (for example the Merriam-Webster definition: <https://www.merriam-webster.com/dictionary/resilience>). While many open research infrastructure providers are still relative newcomers, given the amount of change we have seen in the scholarly communications landscape over the past 20-30 years, those that have survived and thrived for a decade or more of disruption—digital, economic, cultural, and more—can, arguably, be seen as resilient. COVID-19, and its impact on the whole research community, is the latest test of resilience for open research infrastructure organizations and many others. As Thaney says: whilst challenging, the pandemic is also an opportunity, “a moment to adapt, rethink the status quo, and build resilience by co-ordinating diverse efforts across different institutions and organizations as part of a long-term strategy” (Thaney, 2020)).

## Attributes of Open Research Infrastructures

We propose six attributes of open research infrastructure organizations, which we believe contribute to their resilience and which, ideally, should be enshrined in their governance. To be successful, we propose that research infrastructure organizations be equitable, provide (demonstrated) value, be trustworthy (and, therefore, trusted), as well as being interoperable, sustainable, and community-driven, and governed.

### Equitable

It's critical that everyone who participates in the research process has equal access to the research infrastructure—and not just the obvious elements like open and reusable metadata, persistent identifiers, and standards—but also some less obvious “public goods,” like common business practices and common technology solutions. This equity of access applies not just to researchers (across disciplines, geographies, career stages, etc.), but also to research managers and administrators, librarians, funders, publishers, and editors. In addition to equitable access, it is just as important to ensure that all stakeholders have an equal voice at the table, for example, a single but equally-weighted vote in governance, and broad representation when making decisions about fees. This means giving equal weight to the needs

of different types and sizes of organization and of different disciplines and geographies. Being equitable ensures that there is a level playing field for everyone using the research infrastructure, which is critical to the health of global society in general. Tools and services can be built on the open infrastructure for the benefit of all, and everyone is equally invested in the organization's long-term success.

## **Value**

For a research infrastructure organization, especially an open one, it can be difficult to demonstrate a clear value proposition. This is, in part, because many such organizations' services can be technically challenging to implement and require investment in terms of both direct costs and development time. This is exacerbated by the fact that the value proposition may not be straightforward, the return on investment may not be immediate, and the benefits for both the organization and the users may not be obvious. Another challenge is that the return on investment is not necessarily just for one's own organization, but sometimes for mutual benefit including for one's competitors. Infrastructure organizations tend not to have marketing departments and rely on organic word-of-mouth to build and share the value story. Articulating the value of both the research infrastructure overall and of any one infrastructure provider is essential to ensuring collective resilience.

## **Trustworthy and trusted**

Users and organizations alike expect the research infrastructure to run smoothly and they get frustrated when it does not, whether that is being confronted with a 404 error or finding that an incorrect DOI has been linked to their ORCID record. According to Bilder, Lin, and Neylon "Infrastructure at its best is invisible. We tend to only notice it when it fails. If successful, it is stable and sustainable. Above all, it is trusted and relied on by the broad community it serves. Trust must run strongly across each of the following areas: running the infrastructure (governance), funding it (sustainability), and preserving community ownership of it (insurance)" (Bilder, Lin, and Neylon, 2015). There have been initiatives working towards defining criteria for trust to enable users to use this as selection criteria (for example, CoreTrustSeal: L'Hours et al., 2019). Transparency goes a long way towards building trust, whether it relates to board elections, processes and procedures, system downtime, fee decisions, audit outcomes, or anything else. And trust, in turn, goes a long way towards building adoption and, therefore, resilience.

## **Interoperable**

Interoperability is one of the four FAIR principles (data should be Findable, Accessible, Interoperable, and Reusable) (Wilkinson et al., 2016). It is vital to ensure that the research infrastructure is connected, not fragmented, so that research objects, including content and data (and their respective metadata) can be readily indexed and cross-referenced, and information can move seamlessly between different systems. For example, a persistent identifier (PID) on its own has little use; it only becomes powerful when it includes standardized metadata and is

connected to other PIDs. Assertions are made between metadata records, identified by PIDs that, together, enable interoperability between different systems, saving time and reducing the risk of errors. Interoperability is a clear benefit of (and an ongoing opportunity for) open research infrastructure. Unlike proprietary systems, those built on open infrastructure can be connected to—and exchange information with—any other system, freely and without the need for bilateral agreements. Open interoperability frameworks can play an important role here (Burton et al., 2017).

## **Sustainable**

Sustainability is, of course, essential for resilience. Importantly, sustainability means not just breaking even financially, but also making a surplus that can be reinvested in the organization when times are good and used to support it when the going gets tough. However, as Bilder et al (2016) point out, in order to be trustworthy, that sustainability must be built on revenues that are consistent with the mission of the organization. For open research infrastructure organizations, this typically means allowing everyone to have access to the basic data or metadata, and charging a fee—membership or other—for access to services built on that data. Organizations that rely entirely on grants are, arguably, by definition, not self-sustaining. Indeed, some funders require a plan for achieving sustainability as part of the funding process. Skinner notes that (whether grant- or member-funded) scholarly communication infrastructure providers are chronically underfunded and understaffed. “If payment is not required, most users will not voluntarily contribute funding. And the programs and organizations building scholarly communication tools, platforms, and services in partnership with the academy end to ask clients/members for the bare minimum. When we have so little capital in hand, we are unable to invest properly in both maintenance and in R&D, which means that we struggle to offer compelling packages to potential client-members” (Skinner, 2019).

## **Community-driven and governed**

Some open research infrastructure organizations never reach full sustainability, which is one reason why community ownership is so critical to resilience. It helps ensure the long-term viability of an organization’s core mission so that the openly available elements like code and metadata, can be forked, which, as Bilder et al (2015) put it, means that, “the community could replicate the entire system if the organisation loses the support of stakeholders, despite all established checks and balances. Each crucial part then must be legally and technically capable of replication, including software, systems and data.” Being community-driven is an equally important component of a resilient organization. Research infrastructure is much more likely to be used if it is driven by community needs, whether national or global, disciplinary or cross-disciplinary. People and communities are often passionate about what they need and why, as are many of the staff at the infrastructure organizations that serve them. Finding ways to invite co-creation and community participation engenders a strong sense of ‘buy-in’ and is, therefore, essential to developing successful research infrastructure.

# A Model for an Open Research Infrastructure Organization

So, what would a model open research infrastructure organization actually look like? How would these six attributes work in practice? As noted, ideally, they should be enshrined in the organization's governance. Considerations include:

## **Equity**

Membership is open to all, with a fee or fee-assistance program that is appropriate for organizations of different types, sizes, disciplines, and geographies. Members who pay an additional fee may obtain additional services, but they do not have more rights than those that pay a lower fee. One member, one vote is an essential element of an equitable organization. In addition, the tools and services provided by the organization are freely available to all users.

## **Value**

The value proposition for all stakeholders is clearly defined and articulated, so that members—potential and actual—understand what they are obtaining, including any additional commitments required in terms of technology, implementation, and communication. Likewise, users understand the service or tool they are registering for and the specifics of how it helps them.

## **Trustworthy and trusted**

The infrastructure elements provided by the organization should be consistently reliable; outages or other technical issues should be communicated to the community (in advance, where possible). The organization's operations should be transparent, with easily findable and up-to-date information about governance, decision-making, policies, practices, financials, and strategies.

## **Interoperable**

To be fully interoperable, the research infrastructure must essentially be openly mineable. At the very least, key elements of it, such as the metadata, should be open and free to use. Robust and supported public APIs, with plugin or widget options, are good examples of an organization that is demonstrating a commitment to interoperation. Committing to interoperability means committing to openness.

## **Sustainable**

The organization should have a formal plan for its own sustainability and an exit strategy that ensures the future of the infrastructure element(s) it is responsible for in case that plan cannot

be fulfilled. Reliance on a single source of funding does not usually constitute sustainability, so the organization should ensure that its income is spread across multiple contributors. And, ultimately, it should have a contingency fund to enable its operations to continue for some period of time in case of emergency.

## **Community-driven and governed**

It is essential that the organization's governance ensures continued community ownership. This almost inevitably means that it will be a not-for-profit of some sort, rather than having a board of shareholders who regulate the business priorities. And it may well include bylaws that, for example, require a majority of not-for-profit organizations on the board, and/or a large proportion of members to approve any significant changes to the governance structure, pricing model, or ownership.

In the scholarly communications arena, very few organizations exist that meet all of the principles for open scholarly infrastructure put forward by Bilder, et al. And very few are truly global or cross-disciplinary. On local and subject-specific levels, there is increased activity in recent years, as governments invest in achieving open research goals. For example, OPERAS is a research infrastructure for the development of open scholarly communication in the social sciences and humanities across the European Research Area (ERA) (<https://www.operas.unito.it/>). Further west, Coalition Publica (<https://www.coalition-publi.ca>) was born in 2017 as a partnership between Canadian university-based initiatives the Public Knowledge Project and Érudit. Backed by the Social Sciences and Humanities Research Council of Canada and the Canada Foundation for Innovation, Coalition Publica aims to develop a non-commercial, open-source national infrastructure for digital scholarly publishing, dissemination, and research in the humanities and social sciences. Amelica (<http://amelica.org/>), which is a communication infrastructure for scholarly publishing and open science in Latin America and the Global South, is another example of a regional initiative aiming to solve specific problems for its community by building upon foundational open research infrastructure. Even while needing local governance, management, and funding, all of them are working towards or have already demonstrated a commitment to the principles, values, and attributes of open research infrastructures.

## **Conclusion**

Resilience is a journey. To arrive successfully at the destination—wherever and whatever that might be—research infrastructure organizations must embrace these proposed attributes, both individually and collectively, in order to give greater assurance of future resilience for all who depend on them. As Aspesi writes, “it is time to start thinking about collaboration and sharing, rather than competition, as a basis for rebuilding the academic community ... [so that] ... data infrastructure may become what it was always meant to be: a support for the research community” (Aspesi and SPARC 2020).

The COVID-19 pandemic is testing the resilience of many—if not most—organizations, including those that provide and support open research infrastructure. In fact, Thaney believes that “openness is going to be more radically accepted (even demanded) than ever before post-crisis” (Thaney, 2020). It’s not yet clear what the long-term (or even medium-term) impact of the pandemic will be but, assuming there is agreement about the critical importance of continued support for research and researchers, then there is an equally strong case to be made for the continued support of the infrastructure on which they depend. Many services and tools that are heavily used by researchers and their organizations are fully dependent on the existence and availability of the current infrastructure. The resilience of open infrastructures that are built on equity, value, trust, interoperability, sustainability, and community, can help ensure both their own longevity and that of the tools and services that rely on them.

There is no single existing example of an entirely good and successful open infrastructure, nor do we suggest that there is one size that fits all. The proposed attributes are intended to provide the basis for further ideas and next steps towards what an open research infrastructure should behave like and how the communities that sustain them could assess them.

As a community, we need to identify which attributes, values, and principles are important to us and to support the organizations that apply these. We hope that our suggestions will provide a good starting point for a more formal model for evaluating open infrastructure, to be developed by the broader community, including a checklist to assist in the assessment of these attributes. The recently codified Principles of Open Scholarly Infrastructure (<http://openscholarlyinfrastructure.org>) should also be a good tool. Open infrastructure organizations could use resources like these to identify areas of strength as well as areas of improvement, while communities could use it to identify the open infrastructure they want to use and invest in. This will help ensure the building blocks that enable the academic community to advance scientific knowledge continue to be robust and resilient.

## References

- Aspesi, C. & SPARC (Scholarly Publishing and Academic Resources Coalition). (2020, June 22). A Time for Radical Change. *2020 update*, <https://infrastructure.sparcopen.org/2020-update/time-for-radical-change>
- Bilder G, Lin J, Neylon C (2015) Principles for Open Scholarly Infrastructure-v1, retrieved June 27, 2020, <http://dx.doi.org/10.6084/m9.figshare.1314859>
- Bilder G, Lin J, Neylon C (2016) Where are the pipes? Building Foundational Infrastructures for Future Services, retrieved June 27, 2020, <http://cameronneylon.net/blog/where-are-the-pipes-building-foundational-infrastructures-for-future-services/>

- Burton, A et al. 2017 The Scholix Framework for Interoperability in Data-Literature Information Exchange. D-Lib Magazine, 23 (1/2). DOI: <https://doi.org/10.1045/january2017-burton>
- Homolak, J., Kodvanj, I. & Virag, D. Preliminary analysis of COVID-19 academic information patterns: a call for open science in the times of closed borders. Scientometrics 124, 2687–2701 (2020). <https://doi.org/10.1007/s11192-020-03587-2>
- Kaitlin Thaney (2020), <https://blogs.lse.ac.uk/impactofsocialsciences/2020/06/19/the-open-scholarship-ecosystem-faces-collapse-its-also-our-best-hope-for-a-more-resilient-future/>
- L'Hours, H., Kleemola, M., & de Leeuw, L. (2019). CoreTrustSeal: From academic collaboration to sustainable services. IASSIST Quarterly, 43(1), 1-17. <https://doi.org/10.29173/iq936>
- Maryann E. Martone, FORCE11: Building the Future for Research Communications and e-Scholarship, BioScience, Volume 65, Issue 7, 01 July 2015, Page 635, <https://doi.org/10.1093/biosci/biv095>
- Skinner, Katherine (23 July 2019) Why Are So Many Scholarly Communication Infrastructure Providers Running a Red Queen's Race? Educopia Institute blog. <https://educopia.org/red-queens-race/>
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>