DOI: https://doi.org/10.3897/arphapreprints.e163819

Engineering Open by Design into Research Infrastructures

Laurel Haak, Katherine Skinner, Kristen Ratan

ARPHA Preprints *Author-formatted, not peer-reviewed document posted on 02/07/2025*. DOI: https://doi.org/10.3897/arphapreprints.e163819

Engineering Open by Design into Research Infrastructures

Laurel Haak[‡,§], Katherine Skinner[|], Kristen Ratan[¶]

‡ Mighty Red Barn, Townsend, United States of America
§ University of Wisconsin, Madison, United States of America
| Invest in Open Infrastructure, Greensboro, United States of America
¶ Strategies for Open Science, SANTA CRUZ, United States of America

Corresponding author: Laurel Haak (laurelhaak@gmail.com)

Academic editor: Editorial Secretary

Reviewed v 1

Abstract

Research activities utilise and depend on interlocking infrastructures – tools, standards, protocols, and other systems and structures at local, national, and international scales that enable researchers to collaborate, analyze and share data and software, and discuss their research findings. Despite growing policy momentum towards open science, a significant challenge persists: a substantial portion of research infrastructure remains inherently closed or restrictive. This lack of openness undermines transparency, integrity, limits reproducibility, and constrains researchers' ability to fully engage with each other. In this paper, we examine how research infrastructures can be designed to embed open principles throughout their development and operation, borrowing elements from Manufacturing Principles, Systemic Service Design, and Open Science frameworks. We propose an open-by-design reference framework for infrastructure builders and guidance for enabling, making visible, and incentivizing specific elements of "openness" within research infrastructures that are prerequisites for a thriving research and knowledge ecosystem.

Keywords

open research infrastructure, design thinking principles, Manufacturing Principles, service design, human-centered design, systems service design, POSI, FAIR, CARE, FOREST, research policy, open-by-design reference framework, open science, research integrity, persistent identifiers, user surveys, telemetry, research evaluation, impact evaluation, sustainable development goals

Introduction

Over the past two decades, open science has gained significant momentum as a principled framework aimed at enhancing transparency, reproducibility, and innovation within research ecosystems (National Academies of Sciences, Engineering, and Medicine 2018b, Hicks et al. 2015). Driven by an evolving landscape of technology, policies, and practices, open science seeks to facilitate knowledge dissemination through accessible repositories, persistent identifiers, and robust data citation standards. The GO FAIR (Findable, Accessible, Interoperable, and Reusable) framework (Wilkinson et al. 2019, Hrynaszkiewicz et al. 2020, GO FAIR 2016,Schultes et al. 2020) is foundational to this work.

Balancing this effort to make research products machine readable, open science policies also are directed at processes of research (UNESCO 2023), in recognition that community engagement and relationship-building are fundamental to building trust (Leonelli 2023). A number open science frameworks and initiatives aim to foster normative socio-ethical behavior, centering on the people and communities engaging in and affected by research, including CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) (Carroll et al. 2021), DORA (Declaration on Research Assessment) (Gagliardi et al. 2023, Bladek 2014), and CRediT (Contributor Roles Taxonomy) (Allen et al. 2019).

Yet, despite these developments, a critical gap persists: research infrastructures themselves. The digital platforms, tools, and underlying standards, protocols, and technical systems that researchers depend upon to build communities of practice often are designed more to extract knowledge than to enrich communities. Research communities experience challenges arising from proprietary systems, unexpected infrastructure failure, lack of governance clarity, little to no interoperability, and the privatization or discontinuation of essential platforms. Such scenarios jeopardize sustainability, limit collaborative potential, and erode community trust in research and scholarship.

These challenges become even more pressing in the context of emerging technologies, such as Artificial Intelligence (AI), which amplify complexity through intricate dependencies among models, systems and data, each potentially governed by divergent openness standards (OECD 2025). For example, AI-driven research often integrates openly accessible algorithms with proprietary datasets or closed computational infrastructures, and lack of provenance standards complicates assessments of reproducibility and transparency. Without intentional openness embedded within their foundational designs, such emerging technological infrastructures risk enhancing existing issues around reproducibility, equity of access, and accountability (Miceli et al. 2022).

To address these critical concerns, this paper advocates for a strategic shift towards an "Open by Design" approach in (re-)designing and maintaining research infrastructures. We bring together conceptual frameworks of public goods (Ostram 2009) with manufacturing and systemic services to provide evidence and practical guidance for "Open by Design". We argue for proactively embedding community engagement and co-design practices into infrastructure development, including consideration of complexity and socio-economic contexts; resource availability; values and policies; and the users and beneficiaries from the earliest stages of infrastructure conception and throughout its operational lifecycle. Our goal is to provide infrastructure builders practical guidance to translate open science principles into infrastructure design components and evaluation frameworks.

Background and Literature Review

Historical foundations of "Open by Design" in Manufacturing

The origins of the "Open by Design" concept can be traced back to manufacturing methodologies developed during the post-World War II industrial expansion. Central among these methodologies is Design for Manufacturing (DFM), a framework introduced to proactively integrate manufacturing constraints, optimize production efficiencies, and systematically enhance product quality at the earliest design stages. DFM addresses manufacturing complexities by anticipating inefficiencies, reducing production costs, minimizing material waste, and ensuring superior reliability and product performance through early stage integrated planning processes (Boothroyd et al. 2011, Andreasen et al. 1987, Whitney 1988).

Further developments emerged in the form of Design for Assembly (DFA), a complementary approach specifically targeting simplified assembly operations. DFA methodologies prioritize reducing overall product complexity by minimizing the number of components, standardizing interfaces, utilizing modular assemblies, and facilitating straightforward, error-resistant assembly procedures. These design approaches have been instrumental in reducing production timelines, enhancing maintenance efficiency, and promoting the widespread adoption of enduring principles such as modularity, interoperability, and standardization. Baldwin and Clark 2014, through their analysis of modular and platform-based system architectures, further articulated these foundational concepts, emphasizing the benefits of modular designs in fostering technological adaptability and innovation.

Simultaneously, parallel advancements in manufacturing were realized through the adoption and implementation of Lean Manufacturing and Just-In-Time (JIT) production systems. Originating primarily within Japanese manufacturing practices, notably pioneered by Toyota, these methodologies institutionalized rigorous and disciplined production processes emphasizing workflow transparency, continuous incremental improvement (Kaizen), systematic elimination of waste (Muda), and enhanced inter-organizational collaboration. These principles cultivated an organizational culture characterized by openness, continuous iterative refinement, and collaborative transparency (Womack 2007), elements increasingly relevant in contemporary frameworks of open innovation and research.

The "Open by Design" concept extends these manufacturing practices by building upon Design Thinking principles. Design Thinking embodies a human-centred, iterative process that prioritizes empathy for users, rapid prototyping, and ongoing iterative testing. This methodological framework actively incorporates continuous feedback mechanisms, systematically refining products to ensure ongoing relevance, adaptability, and user satisfaction (Brown 2008). The iterative, participatory nature of Design Thinking aligns with open innovation frameworks (Chesbrough 2003), which emphasize openness, transparency, and collective stakeholder involvement as essential drivers of innovative capacity and sustained knowledge dissemination. Moreover, the broader influence of these manufacturing philosophies extends beyond mere technical improvements. They foster organizational behaviours conducive to cross-disciplinary collaboration, stakeholder engagement, strategy development and transparency in knowledge sharing (Platts and Gregory 1990).

Evolution of Design Principles in Digital Contexts

While industrial design emerged from mass manufacturing, service design has emerged from digital transformation of many former in-person transactions, from banking to publishing. Emerging from product design, interaction design and cognitive psychology (Rytilahti et al. 2015), at its best, service design considers all relevant stakeholders and is centered on user needs, not the needs of the business - but also integrates a business case and service sustainability model. Designers engage users in developing narratives and co-designing the service, then prototype and test the service with users, before deploying a minimum viable service, which is then iterated upon and improved based on user feedback (Reason et al. 2015; IxDF 2023). This approach ensures that services are based on evidence and genuine comprehension of the purpose of the service as well as capabilities to deliver value to users.

More recently, the field of service design has evolved significantly from UI/UX work to become a driver for organizational transformation (Souheimo et al. 2025). Service design can fail when the designers do not have a holistic understanding of the systems in place in and around the product they are developing. The systemic service design approach ensures that broader social, economic, and political environments are considered when developing a service so that the design process incorporates multiple stakeholder perspectives, emphasizes relations and interconnections, and considers the ethical implications of different pathways.

Systemic service design is a useful construct for considering design and management of public goods. (Miettinen and Sarantou 2017). Services transcend design objects and become a means for supporting the emergence of a more collaborative, sustainable and creative society and economy. At the infrastructure scale, these designs can become commons – community-scale interventions engage staff, organization, and publics and ideally build capacity rather than dependency and can empower knowledge exchange and re-distribution of power (Bollier and Helfrich 2015, Sangiorgi 2011).

Together, the focus on product from manufacturing design and the engagement of communities in developing that product from service design incorporate the goals of open science: engaging communities in the scientific process and ensuring access to knowledge. The challenge in managing community products, or public goods, has been well-described by Ostram (2009) as a framework of interacting subsystems. Applied to the research, design, and development space (Fig. 1), this framework helps to articulate the context and high-level components that must be considered when designing research infrastructure. These include the resources (people, facilities, data, funding) that drive scientific processes; the overarching resource system at varying levels of scale from local to international; as well as the values, rules, policies, and ethics that govern how research is carried out; and the communities that use research. Infrastructure designers must endeavor to understand how these subsystems interact with each other to be able to develop effective services.

Application of Design Principles to Open Research Infrastructure

The adaptation of design principles into research infrastructures provides a robust, structured framework for embedding openness systematically across conceptual and practical dimensions. Like manufactured products, research infrastructures necessitate comprehensive strategic planning that incorporates sustainability, standardization, interoperability, and lifecycle management from initial design through continuous operational phases. As digital products, research infrastructure requires explicit community partnership from ideation through prototype through launch and subsequent iterative improvement (Fig. 2). Open source code and software are examples of community goods with a rich history of iteration regarding maintenance and governance (Rossi et al. 2012, Lin and Maruping 2021). These open tools and infrastructures serve the research community, after all, not the other way around.

Consequently, the "Open by Design" approach explicitly integrates principles of transparency, accessibility, modularity, and interoperability into early-stage decisions. This approach mirrors the proactive methodologies central to DFM and DFA, systematically embedding manufacturability and ease of assembly into product designs from inception. Research infrastructures inherently share characteristics with complex manufactured products, particularly regarding modular architecture and standardized interfaces.

Such structural considerations facilitate seamless updates, replacements, and integration with complementary systems, substantially enhancing infrastructure adaptability and long-term resilience. This foresight in the initial phases mitigates risks associated with technological obsolescence, vendor lock-in, and operational disruptions, ensuring enhanced sustainability and operational effectiveness (Mons et al. 2017).

Illustrative examples of successful research infrastructures incorporating these openness principles include platforms such as the WorldWide Protein Data Bank and PubMed. These infrastructures demonstrate tangible advantages gained from embedding systematic openness within their governance frameworks, operational protocols, and technical architectures. Consistent stakeholder engagement, transparent governance practices, and rigorous adherence to open standards substantially contribute to their enduring operational effectiveness and sustained relevance within the research community.

Additionally, akin to Lean Manufacturing's commitment to continuous improvement and Systems Service Design's commitment to ongoing community engagement, open research infrastructures significantly benefit from iterative developmental processes. The importance of iteration is well recognized in design research. Wynn and Eckert (2017) reviewed the efficacy of a variety of iteration methodologies and suggest functional categories to be selected based on design needs, from exploring and refining, to correcting, and also coordinating efforts within and across groups. The goal of iterative methodologies is to integrate robust feedback mechanisms from user communities, leveraging user insights to refine operational processes continuously.

Transparency in operations, systematic use of persistent identifiers, and clear, comprehensive metadata standards substantially enhance reproducibility, promoting cross-disciplinary interoperability and comprehensive data stewardship (Wilkinson et al. 2019). Furthermore, research infrastructures adopting these principles achieve greater stakeholder confidence through transparency and collaborative engagement, facilitating improved user trust and enhanced community participation. This openness significantly fosters greater innovation potential and adaptive capacity, enabling these infrastructures to evolve dynamically in response to emerging research needs and technological advancements.

Thus, applying open design principles to research infrastructures provides an integrated and pragmatic framework for embedding openness systematically throughout the infrastructure lifecycle. Outputs generated by open infrastructures are more likely to be 'born with integrity' as the originating systems are auditable and the formats more likely to conform to data standards. Consequently, infrastructures developed under this paradigm could potentially demonstrate robustness, enhanced accessibility, superior interoperability, and sustained long-term operational sustainability, aligning closely with contemporary standards of open innovation and built upon relationship and trust building in scholarly communication.

Open Research Infrastructure Frameworks

The concept of "open" as a scholarly value or principle emerged by the mid-1990s and gained prominence in the early 2000s, when it was invoked widely and variously in open letters, statements, declarations, and manifestos in academic and research communities (see FORCE11 Scholarly Commons Working Group 2017), particularly in the contexts of open access and open science (Skinner and Wipperman 2020). Since that time, it has been used to signal a range of mechanisms and actions that make research available freely online without financial or technical barriers. Openness, in this sense, can be embedded in or enacted by scholars, research objects, and research infrastructures including standards, protocols, and tools.

Applications of the term "open" extend across open science, open access, open source, and open data; they also extend across many different types of functions, including technical (e.g., interoperability, flexibility) and community engagement (e.g., participation, contribution, governance) (see Leonelli 2023). This highly variable use arguably made claiming and/or evaluating "openness" too subjective, and by 2016, research stakeholders were actively working to build and refine clearer, more consistent standards by which characteristics like "openness" could be measured (see CoARA 2025).

The resulting promulgation of frameworks between and beyond 2016-2022 (see Bilder et al. 2020, GO FAIR 2016, GIDA 2023, COAR 2024, SPARC 2019, DPSC 2018, AmeliCA 2019, Group 2025, UCOLASC 2018, Badolato 2019, OANA 2016, Lippincott and Skinner 2022,Agate et al. 2022,Schultes et al. 2020) provide those who engage in or manage research outputs, tools, standards, processes and/or protocols with ways of both articulating and evidencing a broad range of values and principles. These frameworks focus on a relatively similar core set of values and principles - including openness, transparency, accessibility, ethics, equity, sustainability, discoverability, reusability, and representation. They have been largely created and issued by associations and working groups with very specific remits, and although there is some level of cross-referencing between them, each stands largely on its own, used by a subset of the landscape. For example, FAIR and CARE are geared towards data collection and sharing (data scientists and analysts); Exemplarity Criteria are geared towards funders and recipients of grant funding; and POSI, It Takes a Village, and FOREST are geared toward those who build and/or offer digital services (e.g., publishing, preservation).

Focusing on the infrastructure builder use case, among the myriad of open research frameworks we have found the Principles of Open Scholarly Infrastructure (POSI) (Bilder et al. 2020), FAIR (Jacobsen et al. 2020), CARE (Carroll et al. 2021), and FOREST (Lippincott and Skinner 2022) translate well into practice. We describe each briefly in **Table 1.**

While there is ongoing work to develop implementation advice for each of these frameworks (e.g., see Jacobsen et al. 2020 and Carroll et al. 2021), our experience as advisors has clearly demonstrated that the people building research infrastructures – entrepreneurs and researchers – are often not able and/or unwilling to invest the time to translate the advice into their design process. As discussed below in ***Use Cases***, a challenge with each of these frameworks is that their target audience and evaluation use case is post-hoc community evaluation. From a builder's perspective, the frameworks are at best directional. From an Open by Design perspective, there are many different and at times apparently contradictory requirements, making it difficult to prioritize and build strategically. Further complicating matters, funders of research infrastructure have not provided clear and specific guidance or incentives to the builders to encourage adoption of these frameworks.

Given this context, we have developed builder guidance for implementing Open By Design principles from the inception of an infrastructure. With the *technical* goals of modular architecture, standardized interfaces, sustainability, and operational effectiveness, and the *organizational behavior* goals of cross-disciplinary collaboration, stakeholder engagement, strategy development, and transparency in knowledge dissemination, we translate the open research infrastructure frameworks into design questions that make these Open by Design principles immediately applicable (Table 2).

Evaluating Research Infrastructure

Part of the Open by Design framework is an evaluative component. One of the benefits of the Open by Design is the prototyping-iteration approach coupled with ongoing community engagement and user feedback. This provides builders with feedback on user experiences for the complement of services offered through the infrastructure, ideas for new features, as well as information on the value of the offering to the user community. What can we learn from existing evaluation frameworks?

Typically, infrastructure is evaluated based on “classical” technical criteria such as performance and security. While these are crucial metrics, Edwards et al. (2003) argue that we need to distinguish between technical workability and value for end-users. Ideally, infrastructures aim to support ongoing iterated patterns of relationships between human beings (Bijl-Brouwer 2017). Bollier and Helfrich (2015) call for evaluation to be oriented toward the vitality of the system measured through sustainability, resilience, and support of living things, rather than efficiency or profit maximization. This ideal is embedded in the United National Sustainable Development Goals (SDGs) (United Nations 2015), a global partnership that has developed a shared framework for peace and prosperity for people and the planet. SDGs were created through an open consensus process, are openly available, are internationally recognized across many sectors, and are imminently suitable for open evaluation purposes. In particular, SDG 9 calls for building resilient infrastructure, promoting inclusive and sustainable industrialization, and fostering innovation. This is the overarching frame in which all research infrastructure can be assessed.

The OECD decision support framework for investments in research infrastructures (OECD 2019), while based on large facilities, is well-suited also for assessing digital platforms. Both facilities and platforms are infrastructures purpose-built for research, and are expected to serve multiple community stakeholders over a long time period (decades) through provision of tooling, educational resources, and/or economic impact. The community-developed OECD framework includes 25 Core and additional standard Impact Indicators. The Core indicators are categorized by Scientific impact; Technological impact; Economic impact; Training and education impact; Social and societal impact. Subsequent community work extended these infrastructure indicators to include global-scale interoperability (National Academies of Sciences, Engineering, and

Medicine 2018a). However, impact is difficult to assess because many outcomes are indirect and produced by users downstream rather than while using the infrastructure itself.

Starting in the 2020s, the Australian Research Data Commons (ARDC) has required infrastructures in their national research portfolio to integrate evaluation components, and encouraged the adoption of persistent identifiers as a strategy to “connect the dots” between researchers, data producers, data users, and the data itself and infrastructure(s) that house and deliver the data (ARDC 2020). Similar national-level policy efforts are underway in the UK, Canada, and Netherlands, and are captured in the Research Data Alliance (RDA) National PID Strategies Working Group (WG), which describes benefits of widespread and consistent PID adoption for open research infrastructure implementation frameworks (Brown et al. 2023).

As with the OECD frameworks, these national infrastructure evaluations have largely focused on downstream bibliometrics – publications and citations. While researchers use infrastructures and eventually write papers, the key value of infrastructure – research process support – as well as its openness is not captured using these metrics. Nor are infrastructures often designed using user experience principles to support generalist use (Osagie et al. 2017) or capture information that can be used for evaluative purposes (Seffah et al. 2006, Amorim et al. 2017, Haak et al. 2020).

Aligned with Open by Design principles, we propose an evaluation framework that is integrated into the design and build process. One that is based on the concept of research infrastructure as a public good as presented in **Figure 1**. Infrastructures typically operate outside the control of a single stakeholder, so it makes sense to relate the assessment of impact to its role in and fit in the environment it inhabits: Is the infrastructure adopted? Does it evolve? Does it scale? (Henfridsson and Bygstad 2013). Vora and Dolan (2022) suggest a framework that has both supply (design, implementation, and governance) as well as demand (scale, user experience, and outcomes/impacts) components.

The underlying question, then, is how does the infrastructure enrich the community it serves? **Fig. 3** illustrates an Open by Design evaluation framework, which encompasses adoption, business sustainability, governance, community engagement and support, and relationship and trust building. We will explore how this framework incorporates Open Science principles and review its implementation in Section 4.

Case studies: Real-world applications and lessons learned

We have been exploring the Open by Design concept from a variety of perspectives. One is policy: working with colleagues and community groups to explore open science frameworks for specific contexts (Haak et al. 2012). Another is practice: working with researchers to define open science evaluation strategies (IOI 2024). And finally, implementation: working with entrepreneurs to embed open science design elements and evaluation frameworks into their product offering at launch (Flanagan et al. 2021). Here we provide specific case studies that test the Open by Design concept in real-world infrastructure build environments.

ARDC Impact Reporting Framework

The ARDC has refined OECD impact indicators to create an Impact Reporting framework for Australian national-level data assets (ARDC 2020). The intent was to drive application of FAIR principles and persistent identifier (PID) strategies, and promote openness and transparency in data management. However, even with a funding supplement to promote implementation, the adoption barrier for those building the data assets was substantial. Infrastructure builders were focused on the specific use case they were solving for, and did not have the knowledge or capacity to learn about open principles and how to implement them.

Working in collaboration with the project teams in the Health Studies Australian National Data Asset portfolio and ten disciplinary National Data Asset projects, one of the authors (Haak) tested an Open by Design approach to translate the evaluation framework into practical implementation plans. These projects included field instrumentation and data processing and pipelines for ocean microbes, biomedical data discovery, geophysics, barley science, ocean wave dynamics, plant trait data, air pollution, species distribution data, electronic medical records, and a cancer registry for companion animals.

As a first step, teams reviewed webpages and initial funding proposals for each infrastructure project and extracted infrastructure services, beneficiaries, and impact goals for each project. The breadth of disciplines and use cases represented by the projects provided an opportunity to test for shared approaches and features while also identifying unique features and functions. Teams explored the question set described in **Table 2** in a series of online workshops with project leaders from April to June 2023. Using whiteboarding, teams mapped out the value proposition and core technical features of each infrastructure. Then, teams focused on the behavioral aspects: stakeholders and primary and secondary beneficiaries; options for user and community engagement; evaluation strategies.

From findings elicited in these conversations, project feature maps were updated and then and then options were explored for implementing open principles. After highlighting structural congruences between infrastructures, existing modular architecture components were discussed, as well as standard interfaces and workflows, and schemas and persistent identifiers that can engage the intended infrastructure users while also providing information about usage patterns and longer term impacts on the research field or area of practice.

Based on this work, teams revised their infrastructure outcome statements to incorporate data collection strategies, available telemetry and instrumentation toolsets, and data processing pipelines to support integrated evaluation, as well as explicit alignment with UN Sustainable Development Goals (United Nations 2015).

Through this iterative consultative process, we found that the Open by Design approach provided a general design process as well as practical and flexible design elements that allowed builders to engage with open infrastructure principle frameworks across research disciplines, target users, and infrastructure types.

Infra Finder

If asked, many stakeholders involved in building and managing research tools, standards, protocols, and services would readily say they believe in an open ethos. They may openly share (or be willing to share with anyone who asks) the content and software elements that they create and maintain. Some even use open licensing mechanisms (e.g., open source, open access) to indicate to others how their creations can be used. Still, the gap between professing openness and really practicing openness is wide.

Invest in Open Infrastructure (IOI), has been actively testing ways to improve investment in and adoption of “openness” (and other key scholarly values and principles) and to better understand what barriers stand in the way of its implementation for different stakeholders in the research and knowledge ecosystem. In 2022, IOI began inviting infrastructures to participate in Infra Finder (IOI 2024), an information tool that would showcase “open infrastructures” in research and scholarship, providing up-to-date and verified information about key attributes of these tools, protocols, and standards to help their prospective funders and adopters to quickly see what options are available and to gauge their fitness for purpose across a number of vectors. Infra Finder launched in April 2024 with an initial set of 57 infrastructures, and in its first year, it added more than 50 additional infrastructures through a rolling intake process.

The attributes Infra Finder focuses on and shares about these infrastructures are deeply informed by IOI’s work with values and principles frameworks, including POSI, FOREST, FAIR, CARE, and CHAOSS (CHAOSS 2025). Each infrastructure’s listing includes technical details (licensing, dependencies, road map) as well as information about its community engagement mechanisms, governance model, business form, and key policies.

To participate, the represented infrastructures submit information about their entity and its practices, including its launch date, maintenance status, technical documentation, licensing, programming languages, standards used (metadata, PIDs, etc), how and by whom it is hosted, integrations with other tools/standards, community norms and contribution pathways (e.g., code of conduct, contribution guidelines), governance, policies, organizational structure, and engagement with values frameworks (FAIR, POSI, etc). IOI vets the information for each entity at submission and again annually, ensuring clear evidence exists for each claim that is made. IOI also evaluates whether the infrastructure conforms to at least one of its key criteria. Along the way, there is usually some level of dialogue between IOI and the infrastructure to elaborate or clarify particular points about the infrastructure.

What IOI has learned in the first year of active hosting of this resource is just how much the process of engaging with IOI (and with the formal set of questions they answer in order to be included) has meant to its represented infrastructures. For these infrastructures, being represented publicly in this way has motivated and enabled them to clarify, both internally and externally, how they evidence their values.

Participating infrastructures have actively increased their values alignment, citing Infra Finder's intake process as "kind of an x-ray of our operations and way of working" that "certainly revealed some areas where we could improve" (Collister 2024). The intake process has prompted infrastructures to create or make more readily available their policies and documentation. It has also helped some to question and pursue changes in their governance models, and to prioritize work they may not have considered before, including creating a "living will" or sunset plans describing where open assets will (and won't) go in the case of a shut down or transition and budgeting an operational reserve to fund these activities if/when needed. Infrastructures that are not (yet) featured are also reporting their use of the framework to better understand and strategically build towards more of the mission-aligned characteristics that Infra Finder surfaces.

Much like the values and principles standards from which it has drawn (POSI, FOREST, etc), Infra Finder as a framework actively helps infrastructures to consider and evaluate their options for better values alignment; it also provides them with a crucial lens into the practices and behaviors modeled by their peers.

It also helps make others - the funders and adopters - more aware of differences in models deployed by various infrastructures and gives those funders and adopters ample details to help them structure a conversation with the infrastructures that interest them.

IOI created Infra Finder for a different, though related, purpose - to provide adopters and funders with additional, clear information about open tools, standards, and protocols that they could consider in hopes that it would increase their awareness and engagement with these infrastructures. In a 2025-2026 research study, "Measurement of Community Health Indicators," funded by the Digital Infrastructure Insights Fund, IOI has been exploring how community health frameworks (CHAOSS, POSI, FOREST) are experienced by three key groups: infrastructures, adopters (institutions that use the software or standard, for example), and funders (e.g., philanthropies providing grant funding). Preliminary findings based on discovery interviews with academic libraries, research institutions, funders, technology creators, vendors, and collectives, include:

- Many adopters/uses are primarily driven by practical needs like cost and fit-for-purpose. For this group, community health attributes are seen as value-added factors rather than decision drivers.
- Funders cite their strong desires to see underlying values like openness, transparency, sustainability, and community governance evidenced by the groups they fund, but also demonstrate very low awareness of the tools that could help.

- Adopters in well-resourced environments with capacity and desire to contribute (time or funding) to the open infrastructures with whom they work, saw community health as crucial for building and maintaining their trust (particularly elements like governance, community contribution, and financial transparency).
- Infrastructures themselves demonstrated low knowledge of community health frameworks, and their initial reaction to the idea was grounded not in “trust” but in “grading.” As one participant put it, “no one likes being judged.”

As reflected also in the ARDC example above, two key barriers to the use of open infrastructure principle frameworks are the time and attention that are required to understand, let alone explore use of them. Embedding the frameworks in a reference environment like Infra Finder incentivizes players to engage with the key values and principles in very practical ways, thinking explicitly about their value to the work they do rather than as a philosophical exercise or other time-consuming, nice-to-have but exhausting-to-undertake requirement.

In other words, part of what IOI is learning from running the Infra Finder service is that groups that resist spending time and energy on values and principles (either infrastructure providers evidencing them or funders and users finding evidence of them) may be highly motivated to provide clear signals of their open practices in an informational system that helps to spotlight these elements as design features (e.g., open licensing). The Infra Finder environment may help to encourage and incentivize “open by design” work without requiring separate evidence trails created for or provided by individual “principle” frameworks.

Next Generation Library Publishing

Library publishing, an early form of what is now referred to as diamond open access publishing, started in the early 2000s as a low cost way to put a range of content online, including journal and book content, as well as grey literature and digital humanities content. The California Digital Library (CDL), for example, launched the eScholarship Publishing program in 2002, offering free publishing solutions for the University of California (UC) faculty, students, and staff. eScholarship has become an exemplar for large scale diamond open access, providing publishing solutions for now over 90 scholar-led OA journals affiliated with the 10 campuses of the UC System. CDL supports publishing at this scale with a team of publishing professionals and technical staff, a level of dedicated resourcing that is not necessarily available to smaller libraries or other open access providers.

For smaller library publishing programs, externally hosted publishing platforms, including bepress have historically enabled them to serve their communities with diamond OA publication options. In 2017, when bepress, the main outsourced library publishing solution, was acquired by a commercial publisher, library publishers sought new options aligned with open, non profit, and community-led values. Many library publishers have also been looking for ways to expand their programs to provide a compelling alternative to commercial and society publishers in alignment with these values.

The Next Generation Library Publishing (NGLP) project launched in 2019, with the goal of knitting together open access components and filling in gaps in order to provide end-to-end infrastructures for submission, production, and distribution of journals at a level that would compete with proprietary commercial publishing options (NGLP 2022). With a grant from the Arcadia Fund, NGLP was formed by CDL, Strategies for Open Science (Stratos), and Educopia, and launched with a year of community engagement to gather requirements. The library publishing community indicated that, along with values-alignment, they wanted to build on existing infrastructures whenever possible and they would need hosting and production services since many could not manage hosting their own open source technologies in house.

Building on this needs assessment, then-NGLP Product Owner Sarah Lippincott and then co-Principal Investigator Katherine Skinner, created, vetted, and released the FOREST Framework (Lippincott and Skinner 2022), a new methodology and instrument for assessing alignment with community values. This resource codified and made explicit a set of open values that could inform the design process for the NGLP project. It provided guidance on how to embed and enact common values in open infrastructures, including Financial sustainability; Openness; Representative governance; Equity, accessibility, and anti-oppression; Sharing of knowledge, and Transparency (FOREST), at any stage of their development. For the NGLP team, FOREST served as a framework for both the project's development and business-planning work, from determining what OS components to use and how to build and host the code base, to planning for its future governance and management as an open infrastructure.

NGLP has been designed to unite systems run by open infrastructures such as Open Journal Systems (OJS), Janeway, and DSpace, with a newly built publishing front end called Meru, built by Cast Iron Coding. Leveraging a modular framework, Meru enables aggregated display of publications managed in multiple back-end systems, supporting the growth of library publishing programs without requiring migration off of the often disparate tools they use. The NGLP project ran several pilots in 2021-2022 and received a lot of positive feedback from partners. The University of North Carolina (UNC) Press and Longleaf Services were able to turn this pilot into a successful business offering for local libraries and other customers, leading the way for a new set of mission-aligned services for diamond journal publishing.

Then, in 2022, NGLP received a grant from IMLS to explore scaling the Meru effort with partner University of Iowa Libraries, a member of the Big Ten Academic Alliance (BTAA) of libraries. Recently, NGLP announced a partnership with BTAA to bring scale to the effort, working with multiple member libraries to explore a shared solution for the consortium to aggregate and display its combined publishing portfolio. Despite a funding shortage from IMLS, which required recalibrating goals for moving the University of Iowa to a production-ready system, the BTAA pilot is underway, and will be another proof point that the modular, open source solution is flexible and can meet the needs of different diamond open access publishers. With current funding challenges in institutions in the US, it is unclear whether diamond open access will receive enough support, but having choices in open source solutions will reduce costs and enable more cross-institutional collaboration and scale.

The NGLP example is demonstrating that being intentional at the start of a large-scale infrastructure initiative can lead to alignment with community values and that separate and even competing services can share open infrastructures and assemble novel solutions that meet the needs of their user community.

Toward an "Open by Design" practice

Working with and as entrepreneurs building open infrastructure, we have experienced first-hand the need for specific technical and behavioral design components that can be easily deployed in a modular architecture using standardized interfaces. We also realize that many entrepreneurs prefer to focus on technical requirements rather than socio-behavioral change, believing that it is easier to code than to converse. But, as design thinking teaches us, we must do both to position the infrastructure for adoption and impact.

Establishing the Design Framework

We have taken learnings from each use case and here propose a flexible evaluation framework that combines public goods and open infrastructure frameworks (Fig. 1 and Table 1) with Open by Design product and process principles (Fig. 2 and Table 2) and the Open by Design evaluation framework (Fig. 3) organized by SDG, that can be used by any research infrastructure. Ideally, this framework would be consulted during the design phase, but it can also inform existing infrastructures with goals to align with open science principles.

The Open by Design framework aligns accountabilities and metrics for research infrastructures designed as research resource technology (SDG 9, Table 3), as training tools (SDG 4, Table 4), and as foundations for research and policy studies (People & Planet SDGs, Table 5).

Each infrastructure can ensure it meets the facets of open by design while also customizing its framework depending on its functional specifications and features, and also can use the framework to ensure that the product offering includes components of FAIR, CARE, POSI, and FOREST frameworks, as is pertinent to the product offering.

Instrumenting Impact Evaluation

Once the design framework is established, the next step is the evaluation component. Given that each infrastructure has a specific constellation of services, activities, and beneficiaries, it is important that each customizes its community engagement and telemetry strategy. At the same time, there are similarities between infrastructures, and there is good reason to harmonise reporting when possible to enable effective internal and community oversight. To balance these needs, we have created a set of impact metrics for each Impact Area, aligned with the SDG pillars of Infrastructure, Education, and a grouping of SDGs around People-Planet. Within each, we ask a set of questions, and the infrastructure host can select which questions and measures are most relevant. The frameworks are shown below in Table 6, Table 7, and Table 8.

Telemetry Strategies

Measuring adoption and use of research infrastructures is complex and flexible, but do-able. As shown above, to capture impact we need to use a range of approaches, including user surveys, usage data, and outputs monitoring. A key component of any measurement strategy is a combination of pre-planning and telemetry (Gagnon and Swanson 2023), which can include the integration and use of persistent identifiers (PIDs) in data pipelines and workflows (Haendel et al. 2021, Haak et al. 2018, Brown et al. 2023) as well as in-workflow user surveys, monitoring of help desk inquiries, and community meetings.

User Surveys

Research infrastructures are built by and for research communities. They need to reflect and support the needs of those communities. To do so, infrastructure project teams need to build pathways for connecting with users on an ongoing basis. ORCID single-sign-on (SSO) is one way to both collect user information and also request permission to send survey materials. ORCID SSO can be integrated into the pathways used to access the infrastructure; to process and/or download data; as well as to access webinars, code, and educational materials. Completion of ORCID SSO and permission to contact can be used to trigger a quick turn-around (within 24 hours) 3-question Net Promoter Score survey.

Contact information collected in this way can also be used to carry out quarterly or annual surveys, and also compare users over time. An additional benefit of using ORCID SSO is the opportunity to collect user ORCID iDs to enable output tracking, as well as request permission to post information into a user's ORCID record, such as creation of a dataset, contribution to a codebase, or curation of a data asset. In addition to these transactional data, it is important to also understand whether the infrastructure is facilitating relationship and trust building. Building out local community groups, supporting working groups, participating in standards work, and co-developing strategy with the community are all examples

Usage Data

How people are interacting with the infrastructure should also be monitored. This can provide data to improve the user experience as well as improve understanding of which product features, such as workflows or datasets, are being accessed. Data collected can include number of page views, datafile or database accesses and/or downloads, and should encompass the data asset as well as related code and educational materials. The use of PIDs for the component parts ("items") of the data asset – at the very least, the entire asset (e.g., landing page), but also databases, data file subsets, codebases, and educational materials – will greatly facilitate this process. DataCite's Fabrica and Usage Tracker toolsets (DataCite 2025b, DataCite 2025c) can be used to create, find, connect, and track DOIs and associated metadata.

These tools align with COUNTER reporting practices. The COUNTER Project specifies total, unique, search, and session metrics for regular full-content databases (COUNTER 2023).

- Total_Item_Requests. Counts every click where a user downloads the full content of an item from the database.
- Total_Item_Investigations. Total_Item_Requests and clicks on other information and clicks on links to other details.
- Unique_Item_Requests. Counts first click to download the full content of an item, per item per session (see below).
- Unique_Item_Investigations. Counts first click to download and follow-on clicks, one click per item per session.
- Search_Regular. Counts the number of user clicks in a search interface; for a resource such as an infrastructure this could count the number of searches kicked off from the infrastructure landing page.
- Sessions. COUNTER uses IP addresses to measure the amount of time within one hour in a 24-hour block that a user is logged into a data resource. The metric logs the number of total sessions within each one-hour timeblock.

For some databases, standard Google Analytics reporting may be sufficient. While not open, GA is free for smaller-use websites; a subscription is needed for higher-volume websites. Google Analytics Engagement Reporting (Google 2025) can provide information on sessions, page accesses, user actions and "conversions'' – such as an action taken on a database or dataset, and usage by country. This type of reporting does not require use of DOIs, but does benefit from their use.

Code Repository Activity

Additional frameworks like Community Health Analytics in Open Source Software (CHAOSS 2025) have been developed to establish ways to gauge and measure aspects of an open source software based on specific topics, each of which includes explicit metrics. For example, CHAOSS can be used to understand the community health of a technical solution (software) developed and managed by an open source community by using metrics gathered through a github repository that relate to things like "Development

Responsiveness" (how well a project is responding to issues and comments), "Code Changes Lines" (measuring how many lines of code are touched via additions and removals over a period of time), or "Project Popularity" (measuring visibility and engagement through forks, stars, social media mentions, and downstream dependencies). CHAOSS includes a wide range of both topics and metrics, and users of this assessment model can pick those that are applicable to their industry, software focus, and/or community design in order to surface data that can help them to analyze what is working and what may need attention within an open source project.

Outputs Monitoring

We also care about what users are doing with the data they have obtained from or processed using an infrastructure. How is the infrastructure being used to support inquiry and problem-solving? Modular architecture, standard triggers and labels, and data mining techniques have been deployed with success in educational games research (Zoeller 2013, DataCite 2025a). In addition, PIDs can help to automate reporting. Infrastructures that integrate ORCID SSO and manage PID assignments for their asset and component items can construct a PID graph. DataCite Commons (Costa Kohwalter et al. 2017) can be used to track DOI citations, as well as connections with PIDs for people, places, and funding.

Discussion: Navigating openness and sustainability

As a unifying concept, "Open by Design" provides a useful construct for designing and assessing research infrastructure openness. Coupled with SDGs, Open by Design provides a community-validated standard that enables a structured conversation about openness and impact while also allowing each infrastructure team the flexibility to discuss and decide which framework components are relevant for their project.

From a design perspective, open research infrastructures should incorporate all of the Open by Design components. From an impact evaluation perspective, we posit that all open research infrastructures are aligned with SDG 9: Industry, Innovation,and Infrastructure, as they are all creating a new or novel research process resource in a governance framework with delivery tooling. While not always recognized by the infrastructure team, each also is aligned with SDG 4: Quality Education, as they are creating open data models, curation methods, and in some cases, open source code and training resources and workshops. Each of the infrastructures is associated with a constellation of other SDGs, which for the purposes of this project we have combined into "People and Planet".

The design framework developed in this project supports generative conversations about community benefit, and enables infrastructure teams to take a broad view of openness and impact. The design framework supports product design decisions, and together with the evaluation templates, informs design and implementation of usage telemetry in the present, near term, and farther into the future with a constellation of relevant impact metrics.

The Open by Design framework provides clarity for design of open research infrastructure product features. It also can serve as a tool for existing research infrastructures to conduct a gap analysis to determine what might be missing in their functional requirements to be considered a true open science infrastructure. In addition to continuing to test the Open by Design approach with entrepreneurs and established infrastructures, we seek to share this approach to enable broader adoption of open science principles by research infrastructures, clarity in impact reporting, and general community understanding and expectation of the fulfillment of open science principles by the research tools and infrastructures they use.

Conclusion and recommendations

Overall, design thinking is as applicable in the research community as it is in manufacturing and service design. The iterative and interactive processes used are effective for engaging with stakeholders to test the applicability of the Open by Design approach.

The Open by Design framework is relevant across a broad group of research e-infrastructure projects. There are salient examples of open infrastructures that have integrated these design components with great impact for their user and stakeholder communities, as recognized by the FASEB DataWorks Prize (Haendel et al. 2021, Abidi et al. 2020, Suraj et al. 2022).

Of course, as with any Open by Design construct, continuous improvement is critical. Ongoing iterative testing should be used to assess the Open by Design framework. The data and means of collection for each of the impact areas should be assessed regularly to capture new survey methods, web usage tools, telemetry modules and implementation of persistent identifier tracking systems such as COUNTER and SCHOLIX. And, it is critical that the community continue to strive for workflow transparency by sharing not just data and code, but also the modules that support organizational behavior improvements, including templates for governance, fundraising and business models, user surveys and telemetry modules.

Out of this work, we make the following recommendations for encouraging the adoption of Open by Design in the design, development, and evaluation of research infrastructure:

Infrastructure Builders

For individuals and organizations that design, engineer, build user communities, and maintain research infrastructures:

- Engage on international data standards

- Use modular architecture, standardized interfaces, and ongoing user engagement to support sustainability and operational effectiveness.
- Use multiple methods to engage with all direct beneficiaries specifically to gather feedback to assess if your infrastructure is meeting their needs.
- Expand your training offering, including open source models, sample analysis code, documentation, and workshops.
- Proactively determine your impact goals, build your evaluation strategy into your infrastructure, and regulrly report out.
- Perform regular self-assessments of infrastructure openness and continue to improve your offering.
- Share your work, including software code, governance documents, evaluation criteria, and fundraising and budgeting templates.

Infrastructure Funders

For organizations – including government sector Ministries and Agencies and philanthropic organizations — that develop policies, funding programs, and training related to research infrastructures:

- Work with infrastructure hosts and builders at project initiation and regularly throughout the project to signal the importance of openness and transparency and perform regular assessments.
- Work with infrastructure hosts and builders to map Impact Areas and determine reporting options, requirements, and workflows.
- Fund the development of standard interfaces and modules: plug-in “applets” for sign-on, permissioning, user surveys, web analytics, impact analytics.
- Continue to support international infrastructure standards work, through staff involvement and direct infrastructure participant support
- Build Impact reporting communities of practice through hands-on workshops with featured infrastructures
- Showcase infrastructure openness and impact in national and international venues, including workshops, articles, and demonstrations
- Include openness and transparency criteria in funding calls and funding agreements
- Support projects that work to enable ongoing infrastructure user engagement

Acknowledgements

Thank you to Leonidas Aristodemou for sharing his expertise on Design for Manufacturing frameworks; Catherine Mitchell, Sarah Lippincott, and Natasha Simons for comments on the paper; and to the ARDC National Data Asset project leads for their participation in workshops, thoughtful comments, and inputs into the evaluation frameworks. Also thank you to our reviewers Jonathan Starr, Zach Chandler, and Cameron Neylon for their insights and very useful suggestions for improving the paper. A portion of the work described in this paper was carried out under a contract to LLH with the Australian Research Data Commons, ARDC Contract NDA04.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Abidi M, Aboulafia D, Accordino M, Acoba J, Ahluwalia M, Ahmad S, Ajmera A, Alimohamed S, Altman J, Angevine A, Bakouny Z, Bar M, Bardia A, Barnholtz-Sloan J, Barrow McCollough B, Bashir B, Batist G, Bekaii-Saab T, Berg S, Bernicker E, Bhutani D, Bilen M, Bindal P, Bishnoi R, Blau S, Bohachek P, Boland G, Bonnen M, Bouchard G, Bouganim N, Bowles D, Busser F, Butt O, Cabal A, Cabalona W, Cabebe E, Caimi P, Campian J, Carducci T, Chen J, Cheng A, Chism D, Choueiri T, Clark M, Clement J, Connors J, Cook E, Curran C, Daher A, Dailey M, Davis E, Dawsey S, Deeken J, Del Prete S, Demetri G, Desai A, Doroshow D, Durbin E, Egan P, Elias R, Elkrief A, Elms D, Elshoury A, Faller B, Farmakiotis D, Fecher L, Feldman L, Ferrario C, Fiala M, Flora D, French B, Friese C, Fu J, Gadgeel S, Gainor J, Galsky M, Gantt G, Garcia J, Gartrell B, Gatti-Mays M, Gill D, Gillaspie E, Giordano A, Glace (G, Glover M, Goel S, Graber J, Griffiths E, Grivas P, Grover P, Gulati A, Gulati S, Gupta S, Gurley M, Hafez N, Halabi S, Halfdanarson T, Halmos B, Hausrath D, Hawley J, Hennessy C, Herbst R, Hershman D, Hoppenot C, Hoskins K, Hoyo-Ulloa I, Hsu E, Hsu C, Hwang C, Islam JY, Jabbour S, Jani C, Jha A, Jhawar S, Johnson D, Joshi M, Kasi A, Kelleher K, Kennecke H, Khaki AR, Khan H, Khan M, Kharofa J, Kloecker G, Knoble J, Kulkarni A, Kumar V, Lammers P, Leighton J, Lemmon C, Lewis M, Li A, Li X, Liu S, Lo KM, Loaiza-Bonilla A, Logan B, Loggers E, De Lima Lopes G, Loree J, LoRusso P, Low C, Lustberg M, Lyman G, Lynch R, Madhavan S, Mahadevan D, Mahmood S, Mansoor A, Marcum M, Markham M, Mashru S, Masters T, Mavromatis B, McKay R, McNair C, McWeeney S, Menendez A, Menon H, Mesa R, Mico V, Miller C, Mishra S, Monahan R, Morgans A, Mulcahy M, Mundt D, Mushtaq S, Nagaraj G, Nagle S, Nakasone E, Nakayama J, Nelson H, Nemecek E, Nguyen R, Nizam A, Nohria A, Nuzzo PV, Ohri N, Olszewski A, Owenby S, Painter C, Palmer J, Panagiotou O, Park C, Pasquinelli M, Patel J, Patel K, Peddi P, Pennell N, Peters S, Pilar C, Pillainayagam C, Puc M, Ramirez A, Rathmann J, Ravindranathan D, Reid S, Reuben D, Revankar S, Reynolds K, Rho YS, Rhodes T, Rice R, Riess J, Rini B, Rink C, Rosen L, Rosenstein L, Rosovsky R, Routy B, Rovito M, Rubinstein S, Saif MW, Salazar M, Santos Dutra M, Schapira L, Schmidt A, Schroeder B, Schwartz G, Schwartz C, Schweizer M, Serrano O, Shafer D, Shah P, Shah D, Shah M, Shah S, Shah C, Shaw G, Shaya J, Shyr Y, Slosky D, Smits M, Solorzano C, Stauffer K, Stockerl-Goldstein K, Stover D, Stratton J, Stratton C, Streckfuss M, Subbiah S, Tachiki L, Tadesse E, Thompson M, Topaloglu U, Tucker M, Van Allen E, Van Loon S, Vega-Luna K, Venepalli N, Verma A, Vikas P, Vinayak S, Vinh D, Wagner M, Wall S, Wang L, Warner J, Wehbe F, Weinstein P, Weiss M, Weissmann L, Wildes T, Williams N, Wise-Draper T, Wood W, Wu JT, Wulff-Burchfield E, Xie Z, Xu W, Yeh A, Yu I, Yu PP, Zacks R, Zaman QU, Zaren H, Zhang T, Zhou A, Zhu H, Zon R, Zubiri L (2020) A Systematic Framework to Rapidly Obtain Data on Patients with Cancer and COVID-19: CCC19 Governance, Protocol, and Quality Assurance. Cancer Cell 38 (6): 761-766. https://doi.org/10.1016/j.ccell.2020.10.022

- Agate N, Long C, Russell B, Kennison R, Weber P, Sacchi S, Rhody J, Thornton Dill B (2022) Walking the Talk: Toward a Values-Aligned Academy. Humetrics HSS. DOI: 10.17613/06SF-AD45. URL: https://works.hcommons.org/doi/10.17613/06sf-ad45
- Allen L, O’Connell A, Kiermer V (2019) How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. Learned Publishing 32 (1): 71-74. https://doi.org/10.1002/leap.1210
- AmeliCA (2019) AmeliCA Principes and Values. URL: http://amelica.org/index.php/en/principles-and-values/
- Amorim RC, Castro JA, Rocha Da Silva J, Ribeiro C (2017) A comparison of research data management platforms: architecture, flexible metadata and interoperability. Universal Access in the Information Society 16 (4): 851-862. https://doi.org/10.1007/s10209-016-0475-y
- Andreasen MM, Hein L, Andreasen MM (1987) Integrated product development. IFS (Publ.), Bedford, UK.
- ARDC (2020) National Data Assets \textbar ARDC. Australian Research Data Commons. URL: https://ardc.edu.au/program/national-data-assets/
- Badolato A (2019) Criteria for the eligibility of projects for funding by the National Fund for Open Science. URL: https://www.ouvrirlascience.fr/criteria-for-the-eligibility-of-projects-for-funding-by-the-national-fund-for-open-science
- Baldwin C, Clark K (2014) Design Rules: Volume 1. The Power of Modularity. The MIT Press, Cambridge, Massachusetts London.
- Bijl-Brouwer Mvd (2017) Designing for Social Infrastructures in Complex Service Systems: A Human-Centered and Social Systems Perspective on Service Design. She Ji: The Journal of Design, Economics, and Innovation 3 (3): 183-197. https://doi.org/10.1016/j.sheji.2017.11.002
- Bilder G, Lin J, Neylon C (2020) The Principles of Open Scholarly Infrastructure. Websit https://doi.org/10.24343/C34W2H
- Bladek M (2014) DORA: San Francisco Declaration on Research Assessment (May 2013). College & Research Libraries News 75 (4): 191-196. https://doi.org/10.5860/crln.75.4.9104
- Bollier D, Helfrich S (Eds) (2015) Patterns of Commoning. Off the Commons Press, Amherst. URL: https://patternsofcommoning.org/ [ISBN 9781937146832]

- Boothroyd G, Dewhurst P, Knight W (2011) Product design for manufacture and assembly. 3. ed. CRC Press, Boca Raton, Fla..
- Brown C, Simons N, Bangert D, Sadler S (2023) RDA National PID Strategies Guide and Checklist. Zenodo https://doi.org/10.15497/RDA/00091
- Brown T (2008) Design Thinking. Harvard Business Review 86 (6): 84-92. URL: https://hbr.org/2008/06/design-thinking
- Carroll SR, Herczog E, Hudson M, Russell K, Stall S (2021) Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data 8 (1). https://doi.org/10.1038/s41597-021-00892-0
- CHAOSS (2025) Community Health Analytics in Open Source Software. URL: https://chaoss.community/
- Chesbrough H (2003) Open innovation: the new imperative for creating and profiting from technology. Harvard Business School Press, Boston, Mass.
- COAR (2024) COAR Next Generation Repositories: Principles. Coalition of Open Access Repositories. URL: http://ngr.coar-repositories.org/principles/
- CoARA (2025) WG Towards Open Infrastructures for Responsible Research Assessment (OI4RRA). https://coara.eu/working-groups/working-groups/wg-towards-open-infrastructures-for-responsible-research-assessment-oi4rra/. Accessed on: 2025-8-11.
- Collister L (2024) Infra Finder Spotlight: OAPEN and DOAB. URL: https://investinopen.org/blog/infra-finder-spotlight-oapen-doab/
- Costa Kohwalter T, Gresta Paulino Murta L, Walter Gonzalez Clua E (2017) Capturing Game Telemetry with Provenance. 2017 16th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). [ISBN 978-1-5386-4846-9]. https://doi.org/10.1109/SBGames.2017.00016
- COUNTER (2023) COUNTER Code of Practice Release 5.1 — COUNTER Code of Practice Release 5 5.1 documentation. URL: https://cop5.countermetrics.org/en/5.1/index.html
- DataCite (2025a) DataCite Commons. Dataite. URL: https://commons.datacite.org/about
- DataCite (2025b) DataCite Usage Tracker. DataCite. URL: https://support.datacite.org/docs/datacite-usage-tracker
- DataCite (2025c) DataCite Fabrica. DataCite. URL: https://doi.datacite.org/
- DPSC (2018) Digital Preservation Declaration of Shared Values. Digital Preservation Services Collaborative. URL: https://dpscollaborative.org/shared-values_en.html
- Edwards K, Bellotti V, Dey AK, Newman M (2003) Stuck in the middle: The challenges of user-centered design and evaluation for middleware. In: CHI (Ed.) Stuck in the middle: The challenges of user-centered design and evaluation for middleware., 2003. CHI 2003, Ft. Lauderdale, 5-10 April 2003. ACM Proceedings of CHI, 5, 6425611-642664 pp. [ISBN 1-58113-630-7/03/0004].
- Flanagan H, Haak L, Paglione LD (2021) Approaching Trust: Case Studies for Developing Global Research Infrastructures. Frontiers in Research Metrics and Analytics 6 https://doi.org/10.3389/frma.2021.746514
- FORCE11 Scholarly Commons Working Group (2017) Principles Of The Scholarly Commons, Version 0.1.1. Website https://doi.org/10.5281/ZENODO.569952
- Gagliardi A, Chen RC, Boury H, Albert M, Chow J, DaCosta R, Hoffman M, Keshavarz B, Kontos P, Liu J, McAndrews MP, Protze S (2023) DORA-compliant measures of research quality and impact to assess the performance of researchers in biomedical institutions:

Review of published research, international best practice and Delphi survey. PLOS ONE 18 (5). https://doi.org/10.1371/journal.pone.0270616
- Gagnon D, Swanson L (2023) Open Game Data: A Technical Infrastructure for Open Science with Educational Games. In: Haahr M, Rojas-Salazar A, Göbel S (Eds) Serious Games: 9th Joint International Conference, JCSG. 14309. Springer Nature [ISBN 978-3-031-44750-1 978-3-031-44751-8]. https://doi.org/10.1007/978-3-031-44751-8_1
- GIDA (2023) CARE Principles. URL: https://www.gida-global.org/care
- GO FAIR (2016) FAIR Principles. URL: https://www.go-fair.org/fair-principles/
- Google (2025) Analytics & Data Analysis Features List - Analytics. Google. URL: https://marketingplatform.google.com/about/analytics/features/
- Group F1SCW (2025) Charters and principles in scholarly communication. FORCE 11. URL: https://docs.google.com/spreadsheets/d/1-aRXFiRg-VL9hpLpxoJqX6-OC-A0R2oCogHflx52Nug/edit?gid=956616118
- Haak L, Baker D, Ginther D, Gordon G, Probus M, Kannankutty N, Weinberg B (2012) Standards and Infrastructure for Innovation Data Exchange. Science 338 (6104): 196-197. https://doi.org/10.1126/science.1221840
- Haak L, Meadows A, Brown J (2018) Using ORCID, DOI, and Other Open Identifiers in Research Evaluation. Frontiers in Research Metrics and Analytics 3 https://doi.org/10.3389/frma.2018.00028
- Haak L, Greene S, Ratan K (2020) A New Research Economy: Socio-technical framework to open up lines of credit in the academic community. Research Ideas and Outcomes 6 https://doi.org/10.3897/rio.6.e60477
- Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, Payne PRO, Pfaff ER, Robinson PN, Saltz JH, Spratt H, Suver C, Wilbanks J, Wilcox AB, Williams AE, Wu C, Blacketer C, Bradford RL, Cimino JJ, Clark M, Colmenares EW, Francis PA, Gabriel D, Graves A, Hemadri R, Hong SS, Hripscak G, Jiao D, Klann JG, Kostka K, Lee AM, Lehmann HP, Lingrey L, Miller RT, Morris M, Murphy SN, Natarajan K, Palchuk MB, Sheikh U, Solbrig H, Visweswaran S, Walden A, Walters KM, Weber GM, Zhang XT, Zhu RL, Amor B, Girvin AT, Manna A, Qureshi N, Kurilla MG, Michael SG, Portilla LM, Rutter JL, Austin CP, Gersing KR, the N3C Consortium (2021) The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. Journal of the American Medical Informatics Association 28 (3): 427-443. https://doi.org/10.1093/jamia/ocaa196
- Henfridsson O, Bygstad B (2013) The Generative Mechanisms of Digital Infrastructure Evolution. The MIS Quarterly 37 (3): 907-931. https://doi.org/10.25300/MISQ/2013/37.3.11
- Hicks D, Wouters P, Waltman L, De Rijcke S, Rafols I (2015) Bibliometrics: The Leiden Manifesto for research metrics. Nature 520 (7548): 429-431. https://doi.org/10.1038/520429a
- Hrynaszkiewicz I, Simons N, Hussain A, Grant R, Goudie S (2020) Developing a Research Data Policy Framework for All Journals and Publishers (Correction). Data Science Journal 19 https://doi.org/10.5334/dsj-2020-017
- IOI (2024) About Infra Finder. URL: https://investinopen.org/data-room/about-infra-finder/
- IxDF (2023) The Principles of Service Design Thinking - Building Better Services. Interaction Design Foundation URL: https://www.interaction-design.org/literature/article/the-principles-of-service-design-thinking-building-better-services
- Jacobsen A, De Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, Courtot M, Crosas M, Dumontier M, Evelo C, Goble C, Guizzardi G, Hansen KK, Hasnain A, Hettne

K, Heringa J, Hooft RW, Imming M, Jeffery K, Kaliyaperumal R, Kersloot M, Kirkpatrick C, Kuhn T, Labastida I, Magagna B, McQuilton P, Meyers N, Montesanti A, Van Reisen M, Rocca-Serra P, Pergl R, Sansone S, Da Silva Santos LOB, Schneider J, Strawn G, Thompson M, Waagmeester A, Weigel T, Wilkinson M, Willighagen E, Wittenburg P, Roos M, Mons B, Schultes E (2020) FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence 2 (1-2): 10-29. https://doi.org/10.1162/dint_r_00024
- Leonelli S (2023) Philosophy of Open Science. Cambridge University Press, Cambridge. [In English]. https://doi.org/10.1017/9781009416368
- Lin Y, Maruping L (2021) Open Source Collaboration in Digital Entrepreneurship. Organization Science 33 (1): 212-230. https://doi.org/10.1287/orsc.2021.1538
- Lippincott S, Skinner K (2022) FOREST Framework for Values-Driven Scholarly Communication. Zenodo URL: https://zenodo.org/records/6557302
- Miceli M, Yang T, Garcia AA, Posada J, Wang SM, Pohl M, Hanna A (2022) Documenting Data Production Processes: A Participatory Approach for Data Work. arXiv. https://doi.org/10.1145/3555623
- Miettinen S, Sarantou M (2017) Social design for servcies framework: Capturing service design for development framework. . Design Management Association
- Mons B, Neylon C, Velterop J, Dumontier M, Da Silva Santos LOB, Wilkinson M (2017) Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services and Use 37 (1): 49-56. https://doi.org/10.3233/ISU-170824
- National Academies of Sciences, Engineering, and Medicine (2018a) International Coordination for Science Data Infrastructure: Proceedings of a Workshop". National Academies Press, Washington, D.C.. URL: https://www.nap.edu/catalog/25015
- National Academies of Sciences, Engineering, and Medicine (2018b) Open Science by Design: Realizing a Vision for 21st Century Research. National Academies Press, Washington, D.C.. URL: https://www.nap.edu/catalog/25116
- NGLP (2022) Next Generation Library Publishing Software. URL: https://www.nextgenlibpub.org/software
- OANA (2016) Vienna Principles a vision for scholarly communication. URL: https://viennaprinciples.org/
- OECD (2019) Reference framework for assessing the scientific and socio-economic impact of research infrastructures. DOI: 10.1787/3ffee43b-en. URL: https://www.oecd.org/en/publications/reference-framework-for-assessing-the-scientific-and-socio-economic-impact-of-research-infrastructures_3ffee43b-en.html
- OECD (2025) Steering AI's future: Strategies for anticipatory governance. DOI: 10.1787/5480ff0a-en. URL: https://www.oecd.org/en/publications/steering-ai-s-future_5480ff0a-en.html
- Osagie E, Waqar M, Adebayo S, Stasiewicz A, Porwol L, Ojo A (2017) Usability Evaluation of an Open Data Platform. Proceedings of the 18th Annual International Conference on Digital Government Research. [ISBN 978-1-4503-5317-5]. https://doi.org/10.1145/3085228.3085315
- Ostram E (2009) A general framework for analyzing sustainability of social-ecological systems. Science 325 (5939): 419-422. [In English]. https://doi.org/10.1126/science.1172133

- Platts KW, Gregory MJ (1990) Manufacturing Audit in the Process of Strategy Formulation. International Journal of Operations & Production Management 10 (9): 5-26. https://doi.org/10.1108/EUM0000000001264
- Reason B, Løvli L, Brand Flu M (2015) Service design for business: A practical guide to optimizing the customer experience. John Wiley & Sons [In English].
- Rossi B, Russo B, Succi G (2012) Adoption of free/libre open source software in public organizations: factors of impact. Information Technology & People 25 (2): 156-187. https://doi.org/10.1108/09593841211232677
- Rytilahti P, Miettinen S, Vuontisjärvi HR (2015) The theoretical landscape of service design. In: Aaron M (Ed.) Design, user experience, and usability: Design discourse. Springer, Berlin, 86-97 pp. [In English]. https://doi.org/10.1007/978-3-319-20886-2_9
- Sangiorgi D (2011) Transformative Services and Transformation Design. International Journal of Design 5 (2): 29-4.
- Schultes E, Magagna B, Hettne KM, Pergl R, Suchánek M, Kuhn T (2020) Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: Grossmann, G., Ram, S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science, vol 12584. Springer. https://doi.org/10.1007/978-3-030-65847-2_13
- Seffah A, Donyaee M, Kline R, Padda H (2006) Usability measurement and metrics: A consolidated model. Software Quality Journal 14 (2): 159-178. https://doi.org/10.1007/s11219-006-7600-8
- Skinner K, Wipperman S (2020) Living Our Values and Principles: Exploring Assessment Strategies for the Scholarly Communication Field. URL: https://educopia.org/wp-content/uploads/2024/09/20201105_NGLP_PrinciplesValues_FinalPublication.pdf
- Souheimo M, Fidos M, Kuronen M, Lee S (2025) Blending Boundaries: A thorough exploration of systems-oriented design and service design integration. In: Souheimo M, Jones P, Lee S, Sevaldson B (Eds) Systemic Service Design. Routledge, New York, 17-40 pp. [ISBN 9781032817200]. https://doi.org/10.4324/9781003501039-3
- SPARC (2019) Good Practice Principles for Scholarly Communication Services. URL: https://sparcopen.org/our-work/good-practice-principles-for-scholarly-communication-services/
- Suraj V, Del Vecchio Fitz C, Kleiman LB, Bhavnani SK, Jani C, Shah S, McKay RR, Warner J, Alterovitz G (2022) SMART COVID Navigator, a Clinical Decision Support Tool for COVID-19 Treatment: Design and Development Study. Journal of Medical Internet Research 24 (2). https://doi.org/10.2196/29279
- UCOLASC (2018) Declaration of Rights and Principles to Transform Scholarly Communication. University of California. URL: https://senate.universityofcalifornia.edu/_files/committees/ucolasc/scholcommprinciples-20180425.pdf
- UNESCO (2023) Recommendation on Open Science. UNESCO. URL: https://www.unesco.org/en/open-science/about?hub=686
- United Nations (2015) Transforming Our World: 2030 Agenda for Sustainable Development. United Nations. URL: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf
- Vora P, Dolan J (2022) What is 'good' digital infrastructure? . Brookings Global Working Paper 167 URL: https://www.brookings.edu/wp-content/uploads/2022/02/Good-Digital-Infrastructure.pdf
- Whitney D (1988) Manufacturing by Design. Harvard Business Review 66 (4): 83-91. URL: https://hbr.org/1988/07/manufacturing-by-design

- Wilkinson M, Dumontier M, Jan Aalbersberg I, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, Da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, Hoen PC', Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, Van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, Van Der Lei J, Van Mulligen E, Jan Velterop, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2019) The FAIR Guiding Principles for scientific data management and stewardship (Addendum). Scientific Data 6 (1). https://doi.org/10.1038/s41597-019-0009-6
- Womack J (2007) The Machine That Changed the World: The Story of Lean Production. Free Press, New York.
- Wynn DC, Eckert CM (2017) Perspectives on iteration in design and development. Researh in Engineering Design 28: 153-184. https://doi.org/10.1007/s00163-016-0226-3
- Zoeller G (2013) Game Development Telemetry in Production. In: Seif El-Nasr M, Drachen A, Canossa A (Eds) Game Analytics. Springer, London. [ISBN 978-1-4471-4768-8 978-1-4471-4769-5]. https://doi.org/10.1007/978-1-4471-4769-5_7

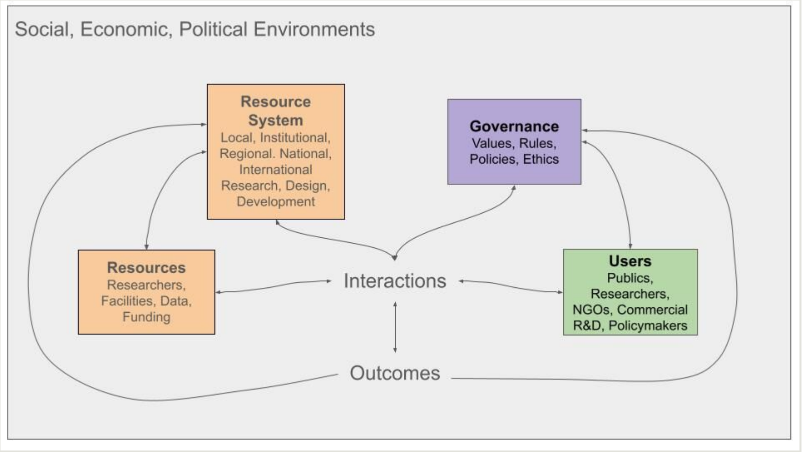

Figure 1.

The socio-economic framework of open science.

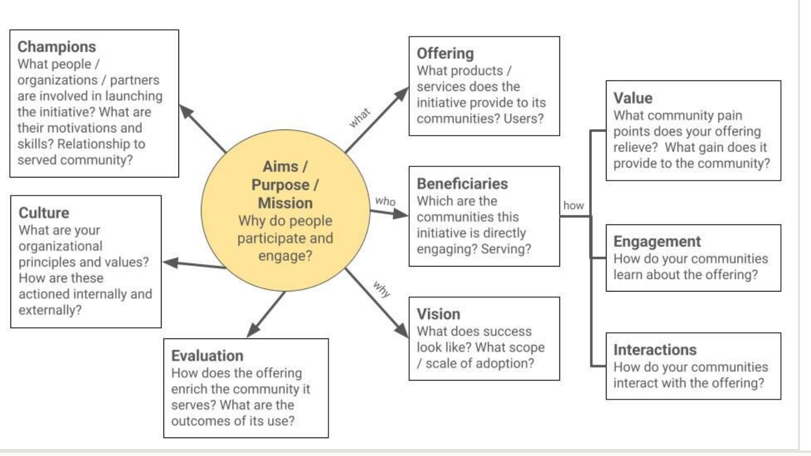

Figure 2.

Systems Service Design: Mapping stakeholders and user needs.

Figure 3.

Open by Design evaluation framework.

Table 1.

Open infrastructure principle frameworks: audiences and assessment features.

Framework	Target audience	Assessment	Content
POSI	Scholarly infrastructure organizations, geared toward leadership	Self-audit of practices. Self-reporting. Non-exclusionary, open to different business types	Governance, Sustainability, and Insurance
FAIR	Data and machine actionability; geared towards researchers and analysts/curators working with data	Self-audit of machine readability. Self or community reported; GOFAIR implementation networks	Findability, Accessibility, Interoperability, Reusability
CARE	Data collectors and curators; geared toward communities, researchers and analysts/curators working with Indigenous data	Self-audit of data ethics; consultation with Indigenous communities; labels and notices to indicate compliance	Collective Benefit Authority to Control Responsibility Ethics
FOREST	Scholarly communications organizations, publishers, libraries, archives; geared towards leadership and staff and towards those adopting open tools or open infrastructures	Self-audit of organizational practices; Adopter community audit of organizational practices	Financial and Organizational Sustainability Openness Representative Governance Equity, Accessibility, and Anti-Oppression Sharing of Knowledge Transparency

Table 2.

Design thinking criteria applied to open research infrastructures.

Technical	**Design Questions**	**Behavioral**	**Design Questions**
Modular architecture	*What standard (open) tools are you using?*	Cross-disciplinary or locational collaboration	*How does your product support collaboration? Data sharing? Integration?*
Standardized interfaces	*What standards, schema, persistent identifiers are you using?*	Stakeholder engagement	*What pain or gain does the infrastructure solve or support? Who is/are the direct beneficiaries? What is the strategy for collecting community feedback to ensure the infrastructure is and continues to solve this problem?*
Sustainability	*How will your product be supported over time, both in terms of user support and financial support?*	Strategy development	*What is your key innovation(s) or differentiator that enables the product to solve the pain/gain? How does the infrastructure enable cooperation with other products and services in research workflows? How are you engaging stakeholders in strategy development and decisions? How do you ensure longevity for your product?*
Operational effectiveness	*What are your terms and conditions of use? How are you ensuring reliable service? What policies and practices are in place to ensure ethical data collection, management, and sharing?*	Transparency in knowledge dissemination	*Who are your stakeholders? What are your product evaluation criteria? How are you collecting data to support evaluation of infrastrucure effectiveness and impact for your stakeholders? How are you ensuring that core data are accessible over time?*

Table 3.

Infrastructure "Open by Design" product features aligned with SDG 9, FAIR, CARE, POSI, and FOREST principles.

Sustainable Development Goal 9: Build resilient infrastructure, promote inclusive and sustainable industrialisation, and foster innovation. Aligns with OECD RI Core Impact Indicators: Technical Impact and Economic Impact.					
Open By Design	**Feature/ Function**	**FAIR**	**CARE**	**POSI**	**FOREST**
Modular Architecture (T)	Extract- Transform- Load Processes	Interoperable			Openness (O2)
Modular Architecture (T)	Data De-Identification		Responsibility		Transparency (T4)
Modular Architecture (T)	Standard Data Aggregation	Interoperable			Sharing of Knowledge (S3) Transparency (T4)
Modular Architecture (T)	(Secure) Data Warehouse	Accessible	Authority to Control	Insurance	Transparency (T4)
Standardized interfaces (T)	API	Accessible, Interoperable			Openness (O2)
Standardized interfaces (T)	Web and Data Access Protocols	Accessible	Responsibility	Insurance	Openness (O4) Equity (E4)
Standardized interfaces (T)	User Interface	Accessible			Equity (E4)
Sustainability (T)	Business Structure, Business Model	Accessible	Ethics	Sustainability	Financial and Organizational Resilience (F3) Transparency (T3)
Operational Effectiveness (T)	Leadership and Ownership Succession Policies, Team Handbook		Responsibility	Governance	Financial and Organizational Resilience (F4)
Operational Effectiveness (T)	Operational Policies and Practices				Financial and Organizational Resilience (F2)
Operational Effectiveness (T)	Personnel Policies (Governance, Team, Participants)				Equity (E1-E3)
Transparency in Knowledge Sharing (B)	Open Source Protocols, Code, Tools			Insurance	Openness (O1)
Transparency in Knowledge Sharing (B)	Data Dictionary, Data Model	Accessible, Interoperable			Openness (O2)
Transparency in Knowledge Sharing (B)	Data Governance		Responsibility	Governance	Openness (O4)

Transparency in Knowledge Sharing (B)	Licensing, Consent, Privacy		Consent, Authority to Control, Responsibility, Ethics	Insurance	Openness (O3) Transparency (T4)
Transparency in Knowledge Sharing (B)	Communication Policies and Practices				Transparency (T2)
Stakeholder Engagement (B) Strategy Development (B)	Stakeholder- Engaged Infrastructure Governance		Responsibility	Governance	Financial and Organizational Resilience (F1) Responsible Governance (R1-4) Transparency (T1)

Table 4.

Infrastructure “Open by Design” product features aligned with SDG 4, FAIR, and CARE, and POSI.

Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. Aligns with OECD RI Core Impact Indicators: Training and Education Impact.					
Open by Design	**Feature/ Function**	**FAIR**	**CARE**	**POSI**	**FOREST**
Cross- Disciplinary Collaboration (B)	File Download	Reusable			Sharing of Knowledge (S3)
Stakeholder Engagement (B)	Outreach Events	Accessible, Reuseable	Responsibility		Sharing of Knowledge (S1)
Stakeholder Engagement (B)	Help Desk, Curation Service	Accessible	Responsibility		Sharing of Knowledge (S2)
Transparency in Knowledge Sharing (B)	Open Source Data Model	Interoperable, Reusable	Responsibility	Insurance	Sharing of Knowledge (S3)
Transparency in Knowledge Sharing (B)	Open Source Tools and Training	Interoperable, Reusable	Ethics	Insurance	Sharing of Knowledge (S3)

Table 5.

Infrastructure “Open by Design” product features aligned with People & Planet, FAIR, CARE, POSI, and FOREST.

People and Planet: Select SDG(s) that aligns with infrastructure purpose and audience. Aligns with OECD RI Core Impact Indicators: Scientific Impact and Social and Societal Impact.

Open by Design	**Feature/ Function**	**FAIR**	**CARE**	**POSI**	**FOREST**
Modular Architecture (T)	PIDs for datasets, code, workflows, etc.	Findable, Reusable		Insurance	Sharing of Knowledge (S1)
Modular Architecture (T)	PIDs for people, organizations, communities, funding	Findable, Reusable	Authority to Control	Insurance	Sharing of Knowledge (S1)
Modular Architecture (T)	Data Linkages	Findable, Accessible, Interoperable			Sharing of Knowledge (S3)
Standardized Interfaces (T) Cross- Disciplinary Collaboration (B)	(Secure) Modelling Workspace, Research Layer	Reuseable	Responsible		Equity (E4)
Standardized Interfaces (T) Cross- Disciplinary Collaboration (B)	Visualisation Tools	Reusable	Responsible		Equity (E4)
Standardized Interfaces (T) Cross- Disciplinary Collaboration (B)	Data Integrated into External Portals	Findable, Accessible, Interoperable, Reusable		Insurance	Sharing of Knowledge (S3)
Standardized Interfaces (T) Strategy Development (B)	Indicators and Trends	Reusable	Responsible		Transparency (T1)
Transparency in Knowledge Sharing (B)	Discovery Catalogue, Inventory	Findable			Sharing of Knowledge (S3)

Table 6.

Impact drivers, measures, and PID strategies for the Infrastructure Impact Area. Font Key: Narrative, *Quantitative*, User Engagement.

SDG 9: Infrastructure			
Are infrastructure data being curated using data models	**Do we see broader adoption of infrastructure data models and/or governance protocols**	**Is the data governance framework maintained?**	**Are users able to access the data or metadata?**
# new infrastructure datasets curated using infrastructure data model / Total # new infrastructure datasets (Backlog) *Requests to infrastructure team for curation assistance and/or Availability of curation assistance* *Workshops for engaging community in discussion on standards* Data sharing standards, principles and documentation available on infrastructure websites and/or Federation service	Data sources providing data aligned with data model vs. infrastructure creating a layer from non-homogeneous data Are related e-infrastructures adopting the data model used by the infrastructure (including international standards) Infrastructure team participation in (inter)national data sharing methodologies and standards Open data endpoints for sharing infrastructure data between e-infrastructures	*# endpoint harvest enhancements | endpoint enhancements up-to-date with data model* Infrastructure Webpage with information on governance group members and activities, updated annually Update processes for data models and metadata in place Participation/Contribution to international data standards working group (Advisory Committee) for maintaining metadata standards Active outreach by infrastructure team to encourage adoption and gather feedback	*# users able to access data / total # requests* *# HelpDesk requests / total # data requests* *# of data access requests per dataset: change over time, total and median* User survey to determine user type, field / sector # infrastructure team points of contact, turnover
Increased understanding of data access needs drives data collection design	***(Inter)national guidance on FAIR-CARE sharing enhances data interoperability***	***Active maintenance of metadata standard supports community trust and adoption***	***Infrastructure data sharing leads to improvements in data access processes for the infrastructure and adjacent research communities***
PID Strategy	1. Integrate authenticated ORCID collection into account creation / sign-in 2. Include PIDs in metadata model 3. Use DOIs or ARKs for component datasets and vocabularies when possible 4. Citation advice on infrastructure landing page and in user documentation for how to cite infrastructure derived-data using PIDs.		

Table 7.

Impact drivers and measures for the Education Impact Area. Font Key: Narrative, *Quantitative*, User Engagement.

SDG 4: Education		
Are educational materials provided by the infrastructure or its Federation service?	**Is training provided by the infrastructure or its Federation service?**	**Is open source code available via the infrastructure or its Federation service?**
Curation documentation available on infrastructure host and/or Federation service websites Metadata schema openly available on infrastructure host and/or Federation service websites Citation guidance available on infrastructure websites and/or Federation service websites User survey: field / sector info	*Availability of training (via infrastructure host or Federation service)* *Completed training modules / # registered users (via infrastructure host or Federation service)* Workshops/ Number of registered users (via infrastructure host or Federation service) Participant survey: field / sector info	*Accesses and/or contributions and/or code forks to the open source code repository?* Is there an open source repository (data model, API, analysis code, etc.)? Codefests or hackathons / Participants / Number of accesses to the code base Participant survey: field / sector info
Materials increase understanding and use of infrastructure resources	***Training increases data sharing and reuse activities***	***Training increases data sharing and reuse activities***
PID Strategy	1. Integration of PIDs into project educational activities and resources (training materials, source code, etc.) 2. Project documentation includes citation advice (e.g., how to cite using PIDs)	

Table 8.

Impact drivers and measures for the People & Planet Impact Area. Font Key: Narrative, *Quantitative*, User Engagement.

People & Planet		
Are infrastructure data findable?	**Does the infrastructure have users?**	**Is there new research based on infrastructure resources?**
# of infrastructure DOIs (or accession numbers) cited / infrastructure DOIs accessed. *# of infrastructure datasets with DOI / total infrastructure datasets* *# of infrastructure datasets embedding PIDs for data, people, organisation, funding / total number of infrastructure datasets* Are data contributions attributed and acknowledged? Use of search tools - which keywords, datasets, user sector	*# infrastructure data PID resolutions / # infrastructure repository accesses* Use of analytics tools - which tools, what questions, which users Community engagement activities (Number, topics, participation) User survey to determine user type, field / sector. Do users map onto expected beneficiaries? Commencement and completion survey for users of platform(s)	*infrastructure PID citations: total, timeseries, topics* *Citation (PID) of an infrastructure or infrastructure dataset in an article, report, grant, patent application or policy guideline* Analysis code citing infrastructure data (PID) *Data Management Plans (DMPs) citing infrastructure (PID) / total machine readable DMPs (PIDs). Interim metric: Letters of Support for grant writers.* Are users able to use (or create new) indicator tools and summaries?
Attribution of infrastructure data drives data sharing	***Improved accessibility of infrastructure data drives community interest***	***Improved accessibility of infrastructure data drives community trust in shared data***
PID Strategy	1. Implementation of Fabrica, DataCite Commons queries, COUNTER, or similar tool to track citation of infrastructure PIDs or hits/downloads of resources	