

Open Infrastructures

Silvio Peroni

silvio.peroni@unibo.it – <https://orcid.org/0000-0003-0530-4305> – [@essepuntato](https://twitter.com/essepuntato)

Open Science (A.Y. 2020/2021)

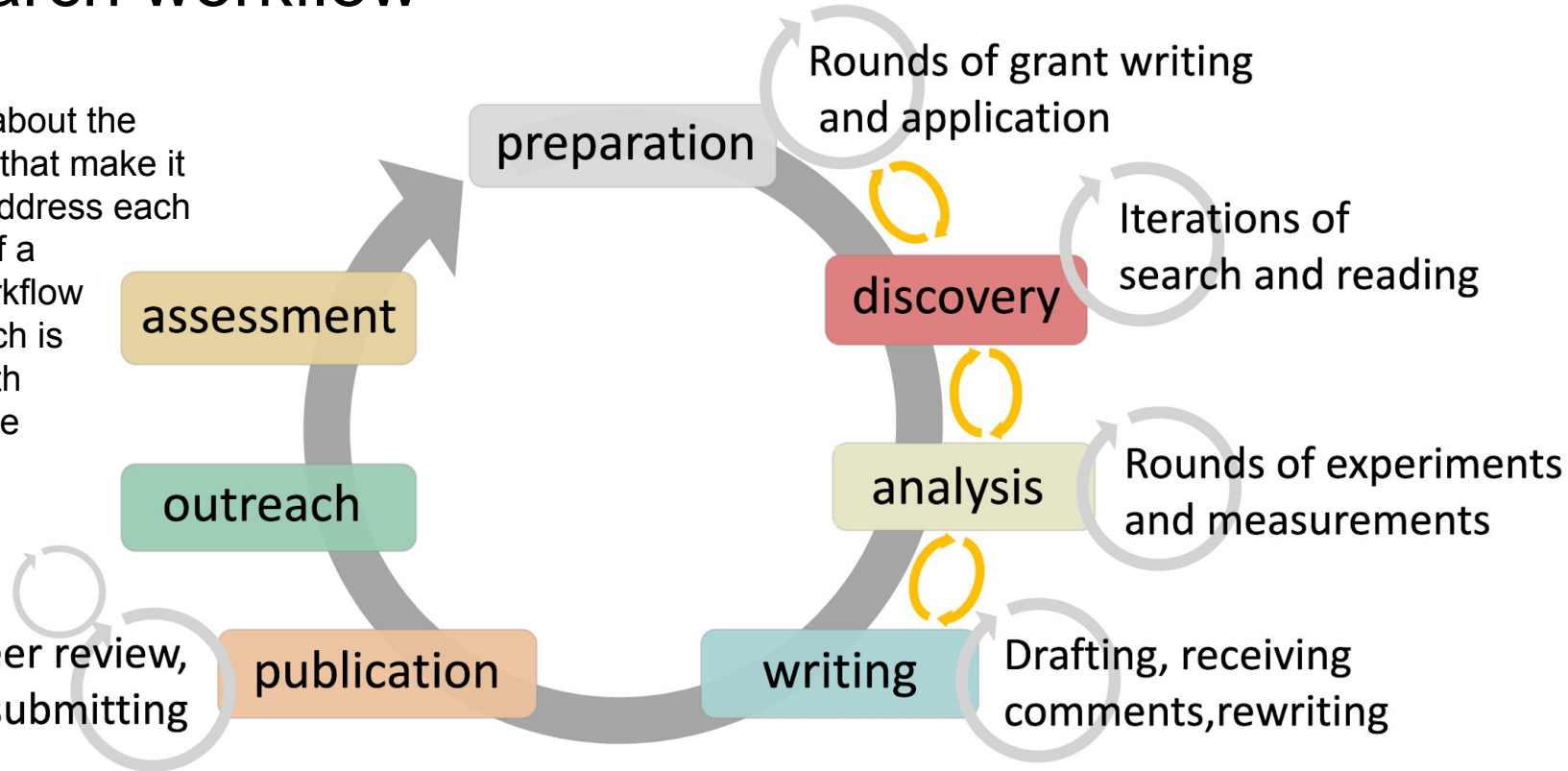
Second Cycle Degree in Digital Humanities and Digital Knowledge

Alma Mater Studiorum - Università di Bologna



Research workflow

We discuss about the components that make it possible to address each single step of a research workflow in a way which is compliant with Open Science practices



The commercial side of the force

[4TU.ResearchData](#) is an international repository for research data in science, engineering and design

It has used [Fedora](#) (an open source repository system) in the past 10 years to host data, but in 2020 decided to migrate a significant part of its technical infrastructure to a commercial solution offered by [figshare](#) (which is a hosting solution similar to Zenodo)

They provided a justification about why they decided not try with open source repository solutions:

“We tried hard. We were in discussion with Zenodo (who are working on the Invenio out of the box repository solution), but the product was still at the pilot stage when we had to start our tender. We had discussions with Dryad, but Dryad’s offering at the time did not give us the functionality we required.”

They set up a tender but they did not receive any interest from other open source repository tools providers, only by commercial entities

How the open replied

[Zenodo](#) and [Dryad](#) decided to reply to that message clarifying some aspects

In the 4TU.ResearchData post, the original authors stated that the decision about who to choose was made around features and capabilities; however Zenodo and Dryad clarified that the **structure of the process excluded non-commercial open source solutions**, due to the huge and unmanageable bureaucracy of the tender process which presented a number of challenges that inherently **favor commercial entities** that are well-suited to go through the process

Zenodo and Dryad met briefly with 4TU.ResearchData (a single 1-hour meeting per entity), and they understood that the tender process was not one in which they could be able to compete, so they did not continue conversations – thus, **the decision was not made** because of features, pilot-phases, or other product judgements

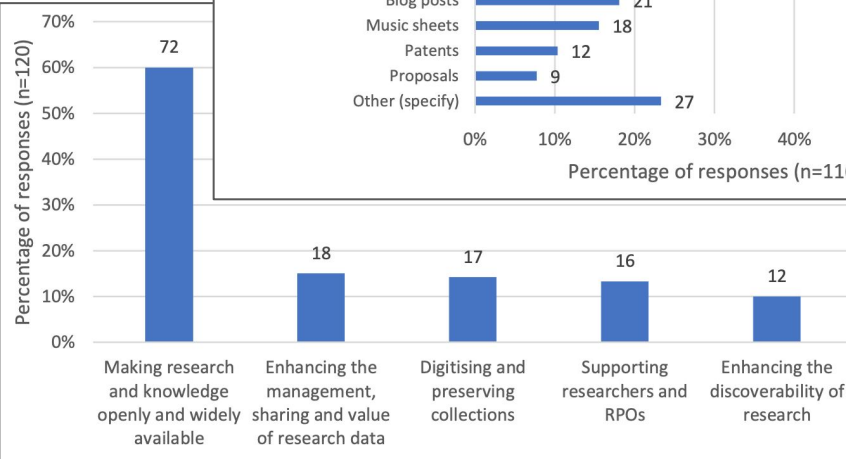
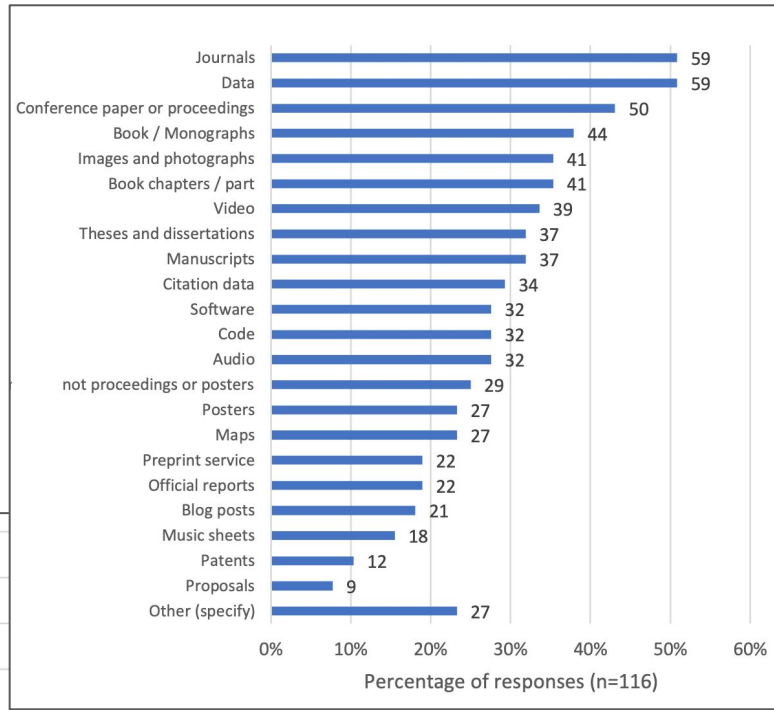
Processes that disfavour non-commercial platforms and communities:

1. will continue to feed this cycle of **questioning the sustainability of our well-adopted and researcher-supported platforms**
2. illogically **promote belief that commercial solutions are more sustainable** and well suited to meet researcher needs

Sustainable, open source alternatives for open research infrastructure not only exist but also thrive

Are they right?

Survey of infrastructures that are part of the European Open Science Infrastructure (OSI) landscape – total respondent: 120 infrastructures



Open Access is not enough

Even if Open Access (OA) publication could become the default in the sciences within the next several years, a particular concern for the academic community is the **growing ownership of data analytics, hosting, and portal services by large scholarly publishers**

The healthy functioning of the academic community requires that the global marketplace for data analytics and knowledge infrastructure be kept **open to real competition**, but some publishers are seeking to protect their profitability by accelerating investment in research infrastructure and data analytics (e.g. by acquiring existing commercial and nonprofit technology companies), and by bundling these and other offerings with journal access

Without investments, the academic community may find itself **beholden to a small number of vendors** for managing communities, data flows, research assessment, and learned society communications, all within digital silos that could hinder the growth of cross-disciplinary collaboration and discovery

First step: **engage in efforts to model consortial funding for and ownership of non-for-profit and non-commercial platforms** – e.g. universities should invest in home-grown research infrastructures and cross-institution consortia, with the goals of establishing competition, sustaining best-in-breed open alternatives, and perhaps eventually providing a suite of services that can substitute for all-in-one commercial workflow solutions

Principles of Open Scholarly Infrastructures

Infrastructure at its best is invisible – we notice it when it fails but, if successful, it is stable, sustainable, trusted and relied on by the broad community it serves

Trust must run strongly across

- **Governance**, i.e. running the infrastructure
- **Sustainability**, i.e. funding it
- **Insurance**, i.e. preserving community ownership of it (insurance)

The authors of the **Principles of Open Scholarly Infrastructures** (POSI) have drafted a set of design principles we think could support the creation of successful shared infrastructures

POSI: Governance

Coverage across the research enterprise. To transcend disciplines, geography, institutions and stakeholders

Stakeholder Governed. Board-governed, drawn from and driven by the stakeholder community

Non-discriminatory membership. Any stakeholder group expressing interest should be welcome

Transparent operations. Push transparent processes and operations in general

Cannot lobby. The community, not infrastructure organisations, should collectively drive regulatory change

Living will. Describe a plan addressing the condition under which an organisation would be wound down

Formal incentives to fulfil mission & wind-down. Have incentives to deliver on the mission and wind down

POSI: Sustainability

Time-limited funds are used only for time-limited activities. Day to day operations should be supported by day to day sustainable revenue sources

Goal to generate surplus. Need financial resources beyond immediate operating costs

Goal to create contingency fund to support operations for 12 months. Generate a contingency fund that can support a complete, orderly wind down, separate from those allocated to covering operating risk and investment in development

Mission-consistent revenue generation. Potential revenue sources should be considered for consistency with the organisational mission and not run counter to the aims of the organisation

Revenue based on services, not data. Revenue sources might include value-added services, consulting, API Service Level Agreements or membership fees

POSI: Insurance

Open source. All software required to run the infrastructure should be available under an open source license

Open data. To be forked it will be necessary to replicate all relevant data – CC0 waiver suggested

Available data. Implement practical ways to obtain data – underlying data should be made easily available via periodic data dumps

Patent non-assertion. Commit to a patent non-assertion covenant, i.e. using patents to prevent the community from replicating the infrastructure is forbidden

Values and Principles Framework

The assessment checklist of the [Values and Principles Framework](#) is designed to help scholarly communications funders, organizations, and communities to identify and align their operations with academic values and principles

- **Diversity, Equity, and Inclusion** – attract stakeholders from diverse experiences and viewpoints with distinct needs and contexts, equitable treatment and opportunity, safe environment, adhere to standards for accessibility, select suppliers on their explicit adherence to equitable standards and inclusive practices
- **Transparency** – publicly documents its internal operations, finances, contract terms and pricing, protect the confidentiality and privacy
- **Openness and Interoperability** – favor open technologies and standards, make it interoperable with others, encourage reuse, provide its members control content and code
- **Access to Knowledge** – support a broader network of others contributing to reducing exclusions from knowledge production and sharing, commit to the development of a robust knowledge commons, prioritize access to knowledge,
- **Representative Governance** – stakeholders with decision-making roles, encourages participation, engage representatives from active stakeholder communities and ensure that representative voices change over time
- **Financial and Organizational Stability** – actively plan for the future, community maintains control over and monitors its own finances and documents its succession/dissolution plans

The beauty of building open scholarly infrastructures

- In May 2015, the European Commission (EC) published its [Communication on the Digital Single Market](#) which contained 10 key priorities of the then new president Juncker, including one of the first policy references to **open research data and cloud** to be found in EC policy making and the announcement of the **launch** of a cloud for research data, i.e. the “research open science cloud”
- In 2016, because judged so important by the [Council Conclusions on OS in Amsterdam](#), a separate [Commission Communication](#) was published on the European Cloud initiative in which **European Open Science Cloud (EOSC)** got first fleshed out
- In the first half of 2018, the Commission Services to come up with a so-called [Staff Working Document \(SWD\) on implementing EOSC](#), acting as an unofficial roadmap for EOSC, was largely inspired by the work of a [European expert group](#)
- In May 2018, the roadmap was agreed in the usual policy gremia of European decision making and new Council Conclusions on EOSC allowed to **officially launch in November 2018** under the Austrian Presidency – which then resulted in having published the [governance structure of EOSC](#)
- All of what seems to be logical steps demonstrate a coherent picture: it fits the formal account of European Open Science (OS) and Open Data (OD) policies
- However, it only documents the **final outcome** of what was a much less logical decision process: the making of EOSC as known today, is also the result of a process of fuzzy logic which can best be read as **a Greek tragedy**

The issues in building open scholarly infrastructures

- The EC started working on it in 2014 when observing the impact of digital technologies on science, and the need of a new modus operandi for science named **Science 2.0**, referred to [earlier work](#) which analysed why Europe could well (and did) miss the boat in the Web 2.0 revolution in the economy, industry, education and social and in science
- In 2014, several workshops were organised and all identified the need of having **non-discriminatory** access to the data, **interoperability** across disciplines and making sure these **data are managed in respect of European sensitivities** (a.k.a. GDPR)
- Several scientists presented their fears that GDPR would make their work impossible as data scientists and that European researchers should not be inspired, to bypass the restrictions of GDPR, to **entrust all their data to foreign owned/registered data servers** to bypass European laws
- Solution: **Europe should manage the data** resulting from its public research system itself and, for that, a science cloud for the EU was needed
- Most of the specialists and key stakeholders supported the baseline that EOSC as a **federated approach to European research data management** in an area of data-driven science, but a lot of **scepticism by policymakers** still needed to be faced
- The Communication describing EOSC (mid-2016) needed **126 iterations of the initial document** in order to get to a compromise all involved EC services could defend and accept – and that **compromise text was heavily criticized by the stakeholders for the confusion it created** as it joined up two rather different technical projects within EC
- The more visible support EOSC got, the more the wider European **ICT community began to pull in the existing schemes as the solution** for what was to come – there was need of **involving three undisputed wise men in the world of European DS to convince top policy people in EC** and explain to them why EOSC must be implemented as described in the Communication of 2016
- Not foreseen **support came from the private sector**: if EOSC will exist, the private players will be asked (via subcontracting) to supply capacity (hosting or services) as they were amongst the only ones able to provide it
- All of the research infrastructures and e-Infrastructures that were also interested to play a role in EOSC – but **several member states MSs (that were not involved extensively) rejected the tender idea to support EOSC** since money was needed and that they agreed on initially when the work program was written, would have been used to do something they did not agree years ago
- EU **needed 4 years** to make a grand idea from scratch into a reality launched in 2018

SCOSS

The [Global Sustainability Coalition for Open Science Services \(SCOSS\)](#) is a network of organisations formed in 2017 and committed to helping secure OA and OS infrastructure well into the future

Goal: provide a new co-ordinated cost-sharing framework that will ultimately enable the broader OA and OS community to support the non-commercial services on which it depends



IOI

[Invest in Open Infrastructure \(IOI\)](#) is an initiative created in 2018 dedicated to improving funding and resourcing for open technologies and systems supporting research and scholarship

Two core premises:

- Open, community-owned infrastructure is necessary for research to thrive
- The way we fund and resource open projects we rely on is insufficient, and working against our aims to build a healthy, collaborative ecosystem

[JROST Rapid Response Fund](#) was launched in 2020 to create a means to give back to the open infrastructure and technology community, awarding a range from 5,000 to 10,000 USD dedicated for activities that are necessary and that would not be possible, or would be in jeopardy, without them



PREREVIEW

sktime



2i2c

HUMANITIES
COMMONS



End

Open Infrastructures

Silvio Peroni

silvio.peroni@unibo.it – <https://orcid.org/0000-0003-0530-4305> – [@essepuntato](#)

Open Science (A.Y. 2020/2021)

Second Cycle Degree in Digital Humanities and Digital Knowledge

Alma Mater Studiorum - Università di Bologna

