# OpenAIRE: A European e-Infrastructure for Open Science

Yannis Ioannidis

Video link: https://www.youtube.com/watch?v=tv9sX_iFlAM&t=42m58s

## Main Mission

Open Science needs pragmatic, participatory infrastructures to work. The diverse and culturally varied research communities of Europe will accept no 'one-size-fits-all' solutions. While some see variety and diversity as insurmountable barriers, for OpenAIRE they are the foundation stones upon which we build. Embedded in a global network of regional repository networks, OpenAIRE is an exemplar for other regions in the world, placing Europe at the forefront of Open Science developments.
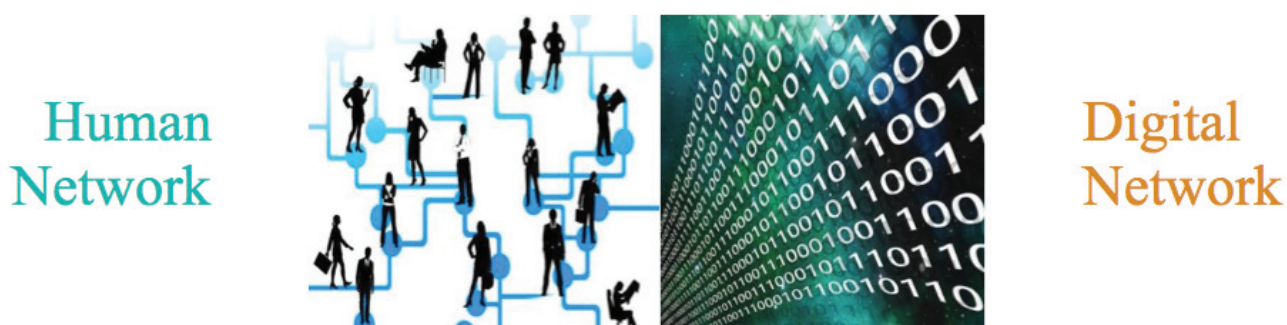


Fig. 1. OpenAIRE fosters the social and technical links
that enable Open Science in Europe and beyond.

OpenAIRE is an infrastructure that implements the policies of the European Commission regarding its Open Access [OA] policies. It has been funded through a series of European projects (OpenAIRE, OpenAIREplus, OpenAIRE2020), with approximately fifty partners from all over the European Union and beyond, including data centres, universities, libraries and repositories (fig. 1). OpenAIRE is scheduled to become a legal entity in 2018. It has been operational on a 24/7 basis since December 2010 and is considered as one of the five or six key electronic infrastructures of the EU.

Above all, OpenAIRE is a socio-technical network that supports the *implementation* and *monitoring* of Open Science policies, including Open Access to publications and research data:

- **Implementation** is enabled by a pan-European network of Open Access / Open Science experts – the National Open Access Desks [NOADs], organised and present in every EU country and beyond, and partitioned geographically into four European regions, as in figure 2. The NOADs

work together to align national policies and to define shared solutions and best practices, and coordinate outreach and advocacy activities through a range of targeted training events and support materials.
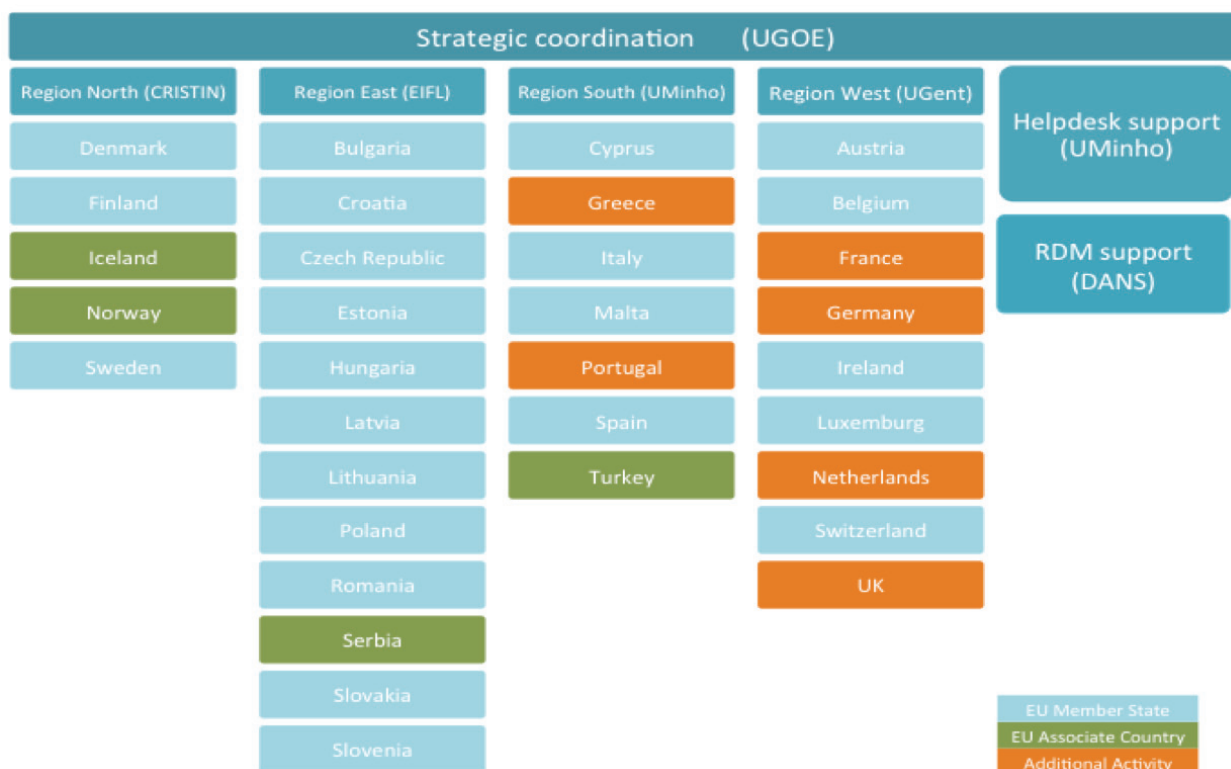


Fig. 2. National Open Access Desks.

- **Monitoring** is achieved by means of an advanced data infrastructure (see fig. 3) consisting of a decentralised network of data sources, namely publication repositories, data repositories and current research information systems, established by research institutions, individual scientific communities and publishers. By harnessing the contents of 'compatible' publications, data, software and method repositories (both institutional and disciplinary), linking them to other research entities (researchers, institutions, projects) and building a broad spectrum of services on top of the resulting 'research information' graph, OpenAIRE produces a 360˚ picture of the impact of European research funding.
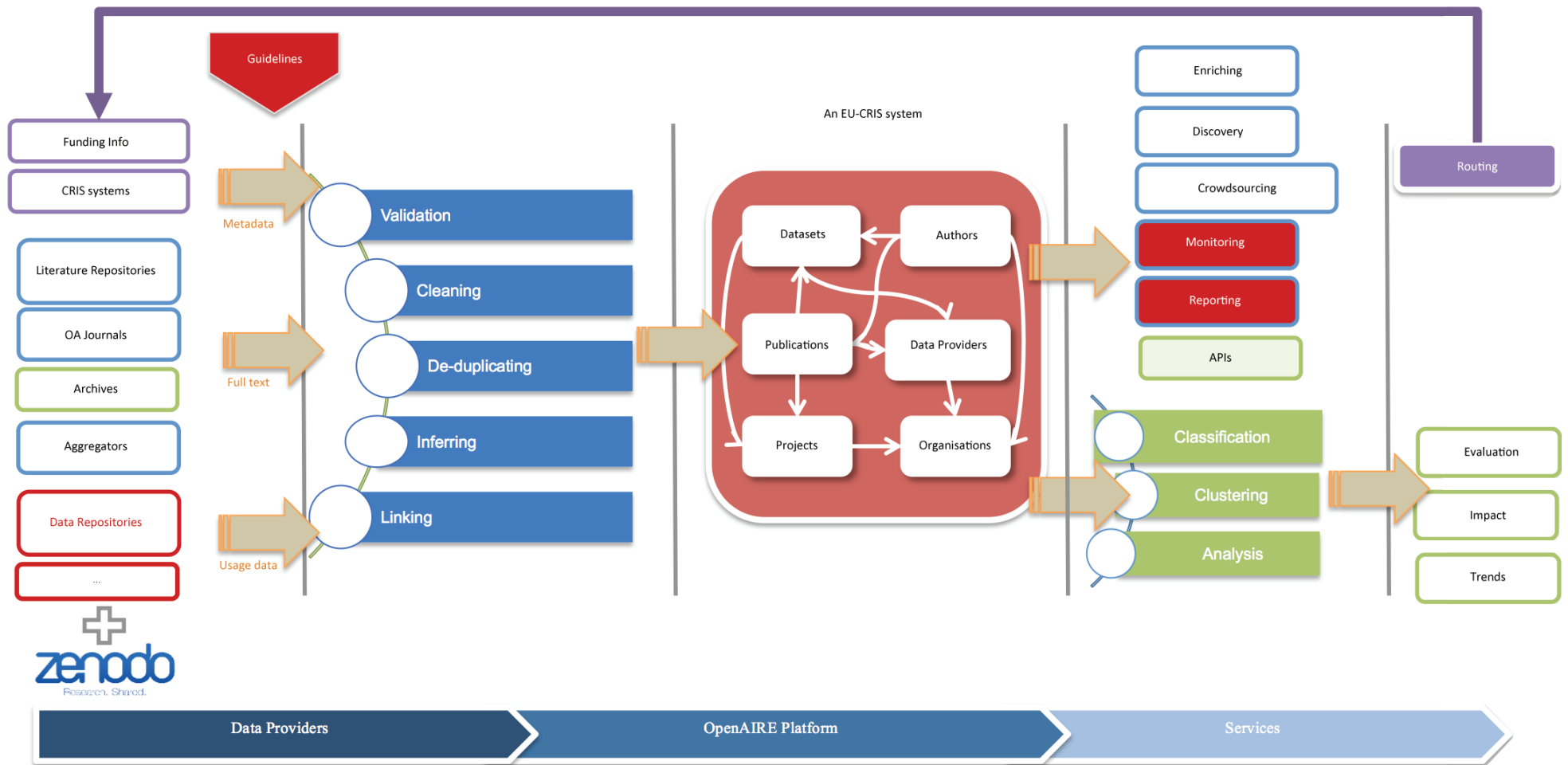
Fig. 3. Advanced data infrastructure.

Research Infra
Operators

National Infra
Operators

Principles
Investigators

Project Coordinators

**3rd Party Service
Providers**

Researchers

Research
Communities

Data
practitioners

**Citizen Scientists**

Research Officers

Funders

Policy

Repository
Managers / Libraries
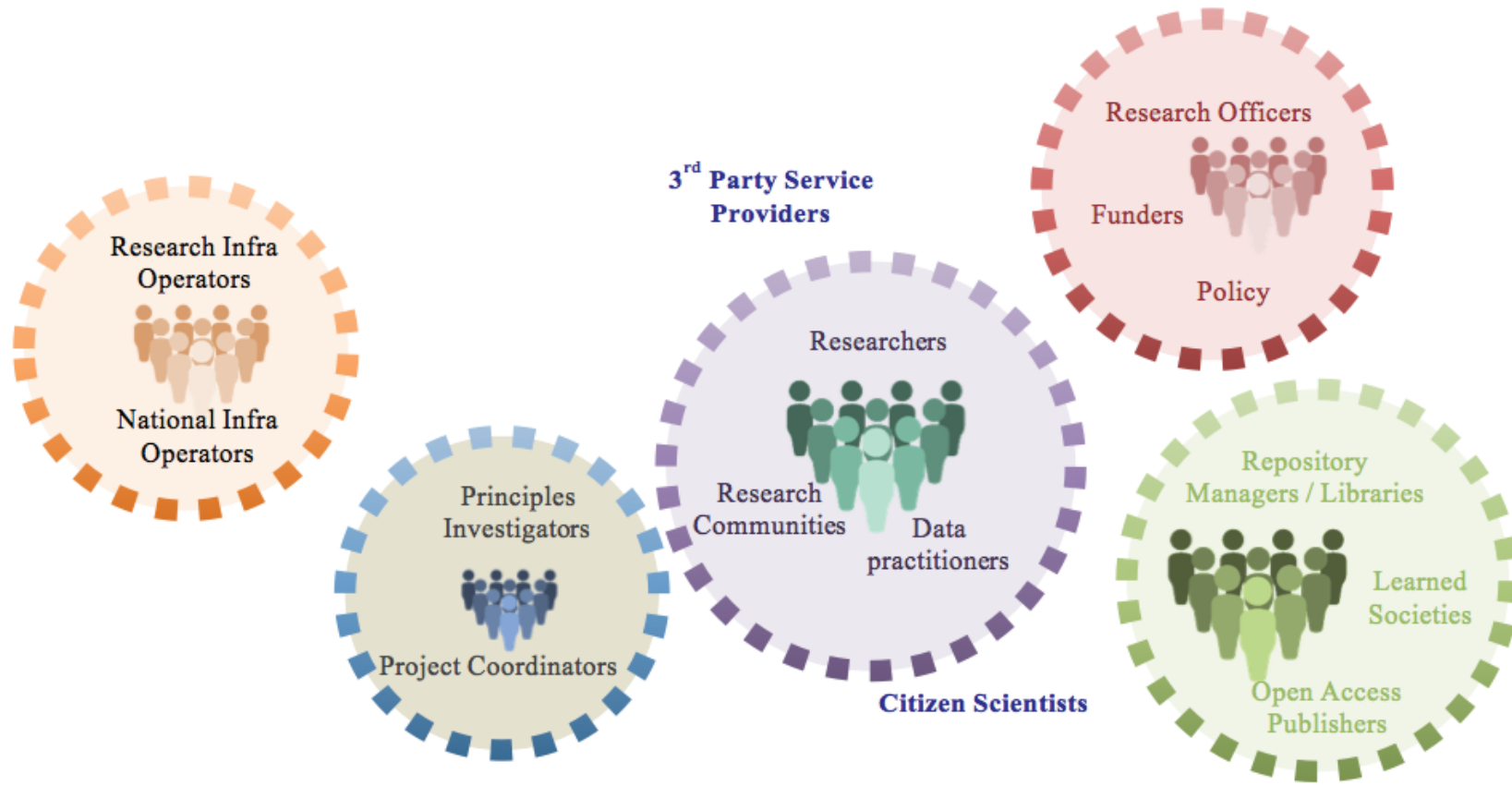
Learned
Societies

Open Access
Publishers

Fig. 4. Target users.

**Target Users**

OpenAIRE addresses a variety of stakeholders (as indicated in fig. 4), each with different perspectives and specific requirements on several aspects of scholarly communication:

- *Researchers* – who want to discover the work of their peers and the context in which it took place; who are looking for having easy-to-use services embedded into their research workflows to allow them to collaborate with their peers (data sharing and reuse);

- *Research communities (data initiatives / labs / groups)* – who are in need of a one-stop shop for their Open Science needs regarding publications, data and other research artefacts: opening them up, sharing and reusing them, linking and putting them in context, monitoring their outputs;

- *Repository managers and university libraries* – who are keen to implement Open Science policies and are at the forefront of support for Open Science;

- *Principal investigators and project coordinators* – who are responsible for monitoring and reporting all project research outcomes to their funders, their peers and the public, while at the same time complying with the funder policy mandates;

- *Research administrators in research performing organisations* – who look for ways to discover the impact of their organisations and see how this compares to others; who are looking into the practical issues of the implementation of Open Science policies as they are key in influencing decisions taken at the organisational level;

- *Open Access publishers* – who want to keep up with the latest policies and researcher behaviours; who need to follow the latest trends of Open Science, for example linking publications to data and other research artefacts; who want to be more visible to more researchers;

- *Learned societies* – who want to keep up with the latest Open Science trends and implement them for their research constituencies;

- *Open Science practitioners* – who need a knowledge hub and a one-stop shop for all aspects of openness in science: legal, organisational, technical;

- *National infrastructure policy makers and operators* – who need to follow the latest Open Science policies and best practices so that they align with EU and global networks;

- *National funders and other policymakers* – who need the community interaction and feedback to shape Open Science policies; who, once they issue the policies, need to monitor them and adapt them; who must employ advanced research analytics to support future decisions.

**Factual Library Data**

OpenAIRE is an initiative that has been funded through a series of three EU projects (OpenAIRE, OpenAIREplus, OpenAIRE2020) from FP7 and H2020. It started in 2009, and the current project will end in June 2018. It has 50 partners from 33 European countries, bringing together broad expertise on Open Science policies and practices and their implementation. The human capital of OpenAIRE includes librarians, repository managers, OA experts, legal experts, e-Infrastructure developers including data centre operators, computer and data scientists, and domain discipline experts.

Through its advanced and rigorous operations, which include harvesting, homogenising, cleaning, transforming, text mining, enriching and de-duplicating, OpenAIRE has created an enriched graph of research and scholarly communication entities which is continuously populated by data from new data sources and updated and enriched with new types of relationships. Presently, OpenAIRE provides access to the following:

- 760+ validated data sources

- 17.3 million unique publications (with 360,000 publications linked to 6 funders)

- 28,000 data objects linked to publications or funding

- 370,000 publications linked to projects from 6 funders

- 3500 links to software repositories

- 33,000 organisations

Furthermore, OpenAIRE provides an extensive suite of services, which is continuously updated as Open Science requirements and new functionalities emerge (see table 1).

**Old Terms versus New / Shifting Concepts**

Based on emerging trends in scholarly communication and recent relevant innovative proposals, it is evident that the research community has moved beyond Open Access in order to embrace a more holistic view of Open Science. Furthermore, as the research paradigm moves to data driven and shared economy practices, Open Science becomes a vehicle for collaboration and innovation.

Open Science touches upon all aspects of the research life cycle, such as access to research facilities (e.g. physical spaces, equipment, lab instruments), storage, stewardship and processing of data (e.g. proper identification, metadata documentation, repository certification, distributed and shared cloud system use) and novel forms of publication of research results, including submitting to new types of journals as these emerge from various communities or universities, employing open (peer) review methods, incorporating innovative ways to disseminate science to the broad public (societal benefits) and adopting new ways to measure impact.

| Service | Description |
|---|---|
| OA Depositing | Promotion of the use of institutional repositories as a means to comply to funder OA mandates |
| OA Publishing [APCs] | Support for the FP7 post-grant pilot, monitoring the use of APCs. *The system developed can be used as a cloud service by funders / institutions.* |
| Interoperability: Guidelines and Validator | Common standards (guidelines) for literature, data repositories, aggregators, OA journals, CRIS systems. Validator of guideline compliance as a web service or a standalone tool. |
| (Intelligent) Discovery / Access | Search and browsing capabilities over a catalogue of Europe's interlinked research results (and beyond) |
| Monitoring / Reporting | Off the shelf monitoring of OA for funders. Reporting to the EC back end |
| Linking | Author, publication, data, project and other research entity (e.g. bioentities) linking that produces a comprehensive graph capturing all elements related to scholarly communication |
| Brokers | (Meta)data exchange among different data providers |
| Resolvers | Data-data and data-publication resolver (DLI service) |
| Enriching | Similarity of documents and references / citations (publication, software and data citation) |
| Knowledge Extraction | Author affiliation, project funding, classification |
| Usage Analytics | Impact monitoring based on repository use |
| Research Analytics | Clustering (hidden relationships), correlations, trends (advanced APIs and visualisations) |
| Helpdesk Consulting and Training | Open Access implementation support, including research data management [RDM], data management planning [DMP] tools and best practices |
| Storing | OpenAIRE's Zenodo (hosted by CERN) has become a well-recognised catch-all repository used for research artefacts from all over the world |
| Data Provision to 3rd-Party Services | Standard protocol offerings: OAI-PMH, REST APIs, LOD |

Table 1. OpenAIRE services.

These new practices require advanced knowledge and skills. In particular, data skills are important in all phases of the research process, and in the case of publishing, where OpenAIRE comes into play, the libraries play a key role in supporting and improving these practices. The new generation of 'data librarians' needs to be educated to guide and train researchers on a variety of topics, including the following:

- availability and use of tools and services from commercial or public e-Infrastructures (whether these are national / EU / global thematic or more generic e-Infrastructures) to be embedded into the library processes;

- promotion and use of open tools (e.g. Jupyter Notebook) and their effective integration into research processes;

- different modes of publishing for articles and data (electronic publishing) and the accompanying costs and benefits;

- support of data curation and stewardship and relevant best practices.

**Two Positive Perspectives on the e-Infrastructure Model**

1) E-Infrastructures achieve economies of scale as far as service provision is concerned (e.g. monitoring and assessment) and affect the behaviour and practices of researchers regarding Open Science. OpenAIRE's proven capabilities to deliver advanced, production-quality technical services used by a wide range of stakeholders has exactly shown this (e.g. increasingly more funders and research performing organisations rely on OpenAIRE's data and off the shelf services for research analytics regarding their constituencies). Significantly, OpenAIRE's active network of National Open Access Desks has a multiplication effect, while using shared resources for support and training.

2) If well-designed and operated, e-Infrastructures are community driven and reveal real researchers' needs more intimately than commercially driven e-Infrastructures or services (e.g. publishers). Engaging the wider community via its representatives (in OpenAIRE's case, the libraries), e-Infrastructures can easily adapt to new trends, adopt new services as they come along and, more importantly, help invent new scholarly communication mechanisms.

**Two Negative Perspectives on the e-Infrastructure Model**

1) E-Infrastructures are hard to grasp, as they provide several soft and low-level services that are not well understood by the different stakeholders; things change only when value added services appear and are embedded in researchers' daily activities. In this sense, e-Infrastructures require an effective coordinated effort that overcomes these barriers and brings to the same level of understanding people from different regional backgrounds and technical expertise. Such coordination may be best left to appropriate national organisations (e.g. national data services).

2) As with all other infrastructures, it takes time for researchers to change behaviours and practices and allow for the value of e-Infrastructures to be noticeable, particularly via advanced services that are in the critical path of research. Therefore, they require continuous investment in people and money and long-term commitment from funders around the world (research is a global endeavour, and e-Infrastructures only succeed if they address the challenges that arise at this level). Naturally, in doing the above, another layer of complexity arises, as these global efforts require alignment.