

## Informal meeting of Ministers responsible for Higher education, Research and Innovation

Stockholm 7-8 February 2023

### Policy debate – Briefing note on research infrastructures in the digital transition – maximizing the benefit of research data

#### Research infrastructures and the data overload

Research infrastructures is a top priority for the current Trio of French, Czech and Swedish Presidencies. During the last years, Europe has invested large sums in research infrastructures (RIs), state-of-the-art experimental facilities that reinforce Europe's research and innovation system by providing researchers from academia and industry with instruments and technical resources. These facilities generate extensive collections of data sets of various sizes and types which require management, computing resources and storage. Major investments have been made in digital infrastructures like high-performance computers and connectivity networks (e-infrastructures), to process and disseminate data generated by research. The Council Conclusions on Research infrastructures of 2 December 2022 addressed inter alia the development of the European infrastructure for scientific data and digital transition.

During the COVID-19 pandemic, it has become apparent that such investments in infrastructure pay off. For example, in response to the outbreak of COVID-19 early 2020, a data-sharing platform was rapidly set up in April 2020 as a priority pilot of the European Open Science Cloud (EOSC). Clinical and genomic data regarding the virus could be efficiently

collected from member states through the European COVID-19 Data Platform and re-used by researchers, allowing them to extract key findings from an interconnected pool of data, to accelerate new discoveries and testing and ultimately the delivery of certified vaccines.

This example clearly demonstrates the possible benefits for the whole of society of making available and effectively disseminating data generated by national, European and international RIs. With open access and FAIR management of research digital outputs, including publications, data and software, the results of research can be used by others – both within and outside the research community to help meeting societal challenges or strengthening the competitiveness of European industry. Access to research outputs benefits both research and a knowledge-intensive society.

Moreover, when partners from across academia, industry, public authorities, and citizen groups are involved in the research and innovation process, trust in science increases and new ideas arise that can develop our society or be commercialized and contribute to competitiveness and sustainable growth in Europe. Thus, the deployment of data and accelerating the digital transition are important factors to boost competitiveness and productivity, as set out in the European Council conclusions from 15 December 2022.

#### “What do we do with all these data?”

Since 2019, the European directive on open data and the re-use of public sector information (1024/2019) includes research data, in recognition that the volume of research data generated is growing each year and has potential for re-use within and beyond the scientific community. Research data in this context includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. It also includes metadata, specifications, and other digital objects. However, to be re-usable, research data also must be findable, accessible, and interoperable (**FAIR**). These attributes are essential to extract the full value from data resources and enable for machine-driven analysis. This is a major challenge today when too many publicly-funded research data remain on local storage including the scientists’ personal hard disks<sup>1</sup>.

---

<sup>1</sup> *European Research Data Landscape*, study commissioned by the European Commission, 2022

Many RIs are leaders in data-driven science and are at the forefront of establishing good community practice in relation to FAIR and open data science. Universities also play a major role in data generation, storage and curation, as they often channel the scientific productivity of their researchers. The combination of several factors such as bigger data volumes, higher data heterogeneity and greater demand for multi-disciplinary data integration across several research communities call for improved FAIR data management practices in order to enable real data reuse. Clearly defined **data policies, guidelines and tools** to govern the generation, management and sharing of research data are often lacking.

Another important factor is **interoperability frameworks** that define community practices for data sharing, data formats, metadata standards, and infrastructure. Such common frameworks need to recognise the objectives and cultures of different research communities, support FAIR across traditional discipline boundaries and in the context of high priority interdisciplinary research areas. Thus, such frameworks must be established on a European or international level. As an example, astronomy has been a pioneer of data sharing, and remains at the forefront. Jointly using open data from different instruments or gathered at different times is at the core of the discipline's science process, another driver being to optimize the science return of investments in the observatories. The disciplinary interoperability framework is defined at the international level and widely used by data providers world-wide.

Increased **technical support** as well as career possibilities for data managers and support functions at universities and RIs facilitate open access in practice for users and lead to the establishment of best practices. Common solutions and support to researchers should be sought. Establishing RIs that combine scientific and operational excellence and allow open, transnational access is challenging. Implementing the 'FAIR-by-design' principle for new data generated by RIs requires significant investments. Making the already existing data FAIR would be even more costly. Therefore, it is important to put in place beneficial **incentives** for RIs to ensure the realisation of FAIR data and open access to research data.

In the ERA Policy Agenda action 1, the EU has set up clear goals regarding sharing and re-using research outputs. The cost for the European economy of not having FAIR research data has been estimated to more than 10 billion

Euros every year<sup>2</sup>. Europe thus has a lot to gain by getting scientific communities and research infrastructures to accelerate efforts to make data FAIR and to share data following the principle “as open as possible, as closed as necessary”.

Research infrastructures can be instrumental in driving investments in making data FAIR, promoting standards and interoperability within and across disciplines, both at EU and global level.

### Questions for the exchange of views

In the light of the above, ministers are invited to give their views on the following questions:

1. The share of high-quality research data that can be re-used and have scientific, economic, and social impact needs to increase. How to deal with the related investment challenge, and what policy measures would you propose to move forward?
2. Implementing FAIR principles to all research outputs and across disciplines is a significant undertaking for Europe and requires changes in terms of research culture and infrastructure. These changes also require alignment with international partners in like-minded regions or countries. Should the EU take action to accelerate this work?
3. A commonly accepted principle is that data should be “as open as possible and as closed as necessary”. “Open” in order to foster the reusability and accelerate science. “Closed” means that open access is based on applicable legislation and ethical aspects – for example IPR, to safeguard security or the privacy of the subjects – not all FAIR data can be open data. Geopolitics adds a security dimension to this. What additional measures should be taken to fulfil the ambition of data that is “as open as possible and as closed as necessary”?

---

<sup>2</sup> *Cost-benefit analysis for FAIR research data*; PwC report commissioned by the European Commission, 2019