European Commission

Research and Innovation

Strengthening European sovereignty in data for research

OPINION PAPER

by the EOSC Steering Board expert group (E03756)

Opinion paper on Strengthening European sovereignty in data for research by the EOSC Steering Board expert group (E03756)

European Commission
Directorate-General for Research and Innovation
Directorate A – ERA & Innovation
Unit A.4 – Open Science and Research Infrastructures
Contact Stefan Liebler
Email RTD-EOSC@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu
European Commission
B-1049 Brussels

Manuscript completed in December 2025
First edition

PDF	ISBN 978- 92-68- 35044- 7	doi: 10.2777/8989807	KI-01-25- 246-EN-N

Luxembourg: Publications Office of the European Union, 2025

EUROPEAN COMMISSION

Strengthening European sovereignty in data for research

Opinion paper
by the EOSC Steering Board expert group (E03756)

2025 Directorate-General for Research and Innovation EN

Contents

1. Executive Summary

European research is at the forefront of open-science practices fostering collaboration and transparent access to data, metadata and analytical facilities including data curation, advanced algorithms and High-Performance Computing (HPC). Ensuring sustainable access to research data, implementing quality control on FAIR archives and services, protecting the European Open Science Cloud (EOSC) from intrusion of faulty contents requires appropriate sovereignty control and investment.

EOSC implements Europe's scientific sovereignty by design. Recent developments have demonstrated that access to critical datasets can be disrupted by political, commercial, or regulatory decisions outside the European Union's (EU) jurisdiction. Such "data suppression or loss-of-access scenarios" expose the dependency of European research on infrastructures and platforms hosted abroad. EOSC will secure the long-term accessibility, interoperability, and trustworthiness of research data within the European framework.

2. Principles for European Data Sovereignty

European data sovereignty for research rests on key interlinked principles:

1. **Openness with responsibility** – Upholding Open Science as a foundation for trust, collaboration, and transparency, while ensuring that ethical, legal, and security issues are effectively addressed.
2. **Autonomy through federation** – Building and maintaining interconnected European data infrastructures in a network-of-networks scheme under EU jurisdiction promoting decentralized governance and shared stewardship.
3. **Control of data and workflow quality** – Ensuring that the lifecycle of FAIR data, from acquisition to availability for reuse and long-term archiving, is traceable and transparent as a guarantee of persistent data quality.
4. **Portability and interoperability** – Adopting open standards and interfaces to prevent vendor lock-in and facilitate cross-border data flows.
5. **Accountability and trust** – Embedding community-based governance, transparent review and monitoring processes, and clear accountability mechanisms for data quality and integrity.
6. **Resilience and preparedness** – Enhancing data infrastructures and services suitable to support the full deployment of EOSC and capable of withstanding geopolitical, technical, or commercial disruptions.

3. EOSC as an Enabler of Sovereignty

The European Open Science Cloud (EOSC) embodies Europe's vision of data sovereignty by design. It is founded on the FAIR principles – Findable, Accessible, Interoperable, and Reusable – and guided by the principle of being "as open as possible, as restricted as necessary." Their application ensures that openness and sovereignty are not opposing goals, but complementary dimensions of a trustworthy European research ecosystem.

Since its inception, EOSC has been conceived as a federated infrastructure underpinned by shared governance, ethical stewardship, and European jurisdiction. Data sovereignty has therefore been an integral component of EOSC's strategic thinking, as reflected in 2022 "Opinion Paper on FAIR Data Sovereignty in EOSC by the EOSC Steering Board Expert Group" [1].

As the build-up phase of the EOSC Federation is currently progressing by interconnecting national, thematic and e-infrastructure EOSC Nodes and the EOSC EU Node, the FAIR principles are being put into practice, shaping how data, services and workflows, that connect them, are organised and linked. EOSC does not merely protect data – it ensures that European science remains open, innovative, and resilient in the face of the evolving global challenges.

4. Context and Risks

Access to and preservation of scientific data are increasingly affected by extraterritorial policies, commercial shifts, and evolving digital regulations. Situations in which publicly funded datasets become unavailable – whether through political decisions, decommissioning of repositories, or commercial withdrawal – highlight the fragility of current global research dependencies.

Today European research institutions and infrastructures rely significantly on digital services operated outside EU jurisdiction. This dependency undermines Europe's capacity to ensure continuity in research, reproducibility, and competitiveness, particularly in strategic domains such as health, climate, food, advanced materials, and artificial intelligence.

At the same time, creating redundant copies of large datasets without coordination could lead to inefficiency and fragmentation. The goal should therefore be EOSC`s resilience—ensuring that European research can operate autonomously, while remaining open and globally connected.

As highlighted in the 2022 Opinion Paper on FAIR Data Sovereignty, data sovereignty on contents and services of EOSC can only be based on its ability to control quality, monitor, and, when necessary, withdraw contents, through rules and mechanisms for FAIR quality assurance, traceability, and integrity. Data quality validation therefore constitutes the prerequisite through which sovereignty over data is exercised, and faulty or fraudulent contents detected and rejected.

Additional dimensions of data sovereignty must be taken into consideration with respect to incipient Artificial Intelligence service propositions. The ultrafast operation of AI on HPC resources makes it possible to explore large amounts of data and the issue of controlling the quality of the training dataset used by the algorithms is crucial.

The issue of FAIR data productivity, assessed in the 2024 Opinion paper on FAIR data productivity [2] by the EOSC Steering Board expert group, calls for investments to enable the timely availability of genuine research-datasets for direct analysis as well as for AI algorithm training. The risk of using cheap artificial training data sets for AI is very high, and it would represent a serious limit to scientific sovereignty.

Finally, recent international developments require that additional and new dimensions of data sovereignty be taken into consideration.

5. Possible Venues for Actions

To strengthen Europe's scientific sovereignty, it is essential to engage research communities across the continent in the active monitoring, improvement and curation of data accessibility.

Scientific sovereignty implies substantial investment addressing the need of acquiring adequate volumes of quality-data sets in all fields of research as well as strengthening the data analysis resources and identifying trustworthy FAIR datasets for AI training. Hardware and software resources must be upgraded and strengthened to cope with the emerging dimensions of sovereignty as the risk of dependency on unreliably trained AI services (degenerative-AI) is in front of us.

Scientific sovereignty also requires identifying critical datasets, e.g. time series and worldwide monitoring, that are needed by EU researchers and ensuring that these data are securely hosted within Europe.

The EU shall reduce its dependence on foreign technology platforms for storage, access and operations, thereby safeguarding long-term autonomy in the governance of scientific knowledge. Scientific data and services should remain under the stewardship of the scientific community. Public funding must never translate into political control or arbitrary restrictions.

Actions at EU Level

- **Building sovereign services beyond storage** – Strengthen Europe's control over analytical, computational, and governance layers to ensure that services critical to science operate within European jurisdiction. Invest in supporting access, promoting EOSC uptake and sustainability.
- **Coordinating with EU data spaces** – Aligning EOSC's principles and governance with relevant EU data-space initiatives (health, environment, transportation, industry, etc.) to ensure coherence, interoperability, and complementarity, avoiding duplication of infrastructures and efforts, favouring innovation in all fields.
- **Enforcing Sovereignty with openness** – Ensuring that data and data-service sovereignty support, rather than limit, global scientific collaboration implies that Europe must remain an active partner in tackling shared challenges such as climate change, public health, food, energy and advanced materials, through trusted, rules-based frameworks that enable responsible and reciprocal data sharing.
- **Protecting the foundational values of Open Science** – Enforcing transparency, academic freedom, reliability and reproducibility as core to a trustworthy European data ecosystem includes to effectively protect the EOSC against intrusions of malicious contents or services, via data quality/integrity control and cybersecurity solutions. Sovereign control should serve openness also by eliminating hidden dependencies and ensuring fairness in access and participation.
- **Resilience and autonomy of Research Infrastructures** – Strengthening the resilience and autonomy of European research and technology infrastructures as providers of massive quality-assessed research FAIR datasets, in line with ESFRI (European Strategy Forum on Research Infrastructures) strategies, will safeguard continuity of service, security, and competitiveness in the global research landscape.
- **Implementing a legal and funding framework** – Defining the lawful bases for dataset mirroring and transfer (including licenses, consent, and database rights) and developing a robust cost-analysis for hosting and repatriation of critical data within Europe (see Annex for technical considerations and recommendations) are key elements to monitor sustainability of EOSC.

Actions at National Level

- **Coordinating national data strategies** – Aligning national data policies with EU principles of openness, sovereignty, and interoperability, also by supporting the EOSC Federation with EOSC Nodes playing a structural effect at national level.
- **Enforcing Public data as a national asset** –Treating scientific and administrative data as public goods, protected from privatization, political interference, or misuse as a sovereignty and research capital protection measure.
- **Investing in infrastructure and skills** – Investing in secure, interoperable repositories and in capacity building for data management, cybersecurity, and FAIR principles. Creating data management and curation training open services.

Actions at Institutional Level

- **Enforcing Data redundancy and protection mechanisms** – Critical datasets, particularly those held by public administrations, should be mirrored, when needed, in trusted institutional repositories to guarantee long-term preservation regardless of political or administrative change.
- **Establishing Offline and cold storage for critical assets** – Where appropriate, institutions should maintain offline or cold storage copies of essential data to guard against accidental loss, cyber threats, or deliberate censorship.

- **Enforcing Governance and accountability** – Universities, research centres, and archives should adopt transparent governance models for data management, ensuring accountability, community oversight, and adherence to scientific and ethical standards, also linking to relevant EOSC Nodes in the EOSC Federation.

6. Legal and operational Issues

FAIR data sovereignty has legal implications, e,g. mirroring or transferring datasets from different institutional archives must comply with:

- Data protection and consent frameworks (General Data Protection Regulation (GDPR) where relevant),
- Intellectual property and database rights, ensuring lawful replication and reuse,
- License compatibility and provenance chains, with documented metadata and auditability,
- Notice-and-takedown procedures, ensuring responsiveness to rights holders.

Operationally, sovereignty implies building a prioritization rubric for critical datasets based on:

- Scientific and societal relevance,
- Uniqueness or irreplaceability,
- Rights of duplication of datasets,
- Community demand,
- Technical feasibility and cost (see Annex for technical considerations and recommendations).

Scientific knowledge is increasingly embodied in software, code, and workflows. Dependencies on commercial code-hosting platforms outside EU jurisdiction (e.g., GitHub, DockerHub) do represent sovereignty risks.

Europe should therefore:

- Support mirroring of key software artefacts into trusted repositories such as Software Heritage or Zenodo,
- Encourage container and workflow preservation to guarantee long-term reproducibility,
- Promote standard metadata and persistent identifiers (PIDs) for both data and software components,
- Integrate code and data preservation strategies under EOSC governance and monitoring frameworks,
- Invest in development of novel independent and public AI technologies.

7. Conclusions

European sovereignty in research data is all but isolation. It requires the ability to maintain open, fair, and trustworthy scientific collaboration, grounded in resilience, interoperability, and shared values.

By reinforcing the EOSC Federation, aligning national and EU initiatives, and embedding legal and operational clarity, Europe can ensure that its scientific data and data-services – and the knowledge they represent – remain permanently accessible, verifiable, and secure. The EOSC Steering Board is committed to advancing this vision: a European Research Area where data sovereignty enables openness, innovation, and global scientific leadership.

Ultimately, data sovereignty is a strong pillar of Europe's broader digital and technological sovereignty, closely linked to capabilities in data analytics, AI technologies and the underlying infrastructures.

[1] Opinion paper on FAIR data sovereignty in EOSC, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/32361

[2] Opinion paper on FAIR data productivity, Publications Office of the European Union, 2024, https://data.europa.eu/doi/10.2777/49194

Annex

Technical considerations and recommendations on data transfer and storage costs

Data transfer

- Transferring large datasets (e.g. Petabyte-scale) e.g. across the Atlantic is not instantaneous.
- The actual transfer time is limited by the slowest link in the data path.
    - Example: GÉANT (pan-European research network) can achieve several hundred Gigabit/s across the Atlantic, allowing 1 Petabyte to move in less than 10 hours under ideal conditions.
    - Many European national or local networks are limited to just 20 Gigabit/s or less, stretching the same transfer to a week or more.

Recommendation: Assessing and addressing national and institutional network bottlenecks when planning data repatriation.

Storage costs

- Storing research data within Europe involves both hardware and significant operational costs.
    - Cold storage (e.g. magnetic tape): lower cost and electricity use, best for data accessed infrequently.
    - Hot storage (e.g. hard drives): needed for regular scientific workflows; higher costs due to equipment, electricity, cooling, and staff.
- Interfaces and services to make data usable (search, select, access) add further costs, particularly in human resources.
- For reference, hardware costs represent only about a third of the total annual running costs.
- Example: Storing 1 Petabyte of data at one of the large European data centers costs about 5,000 Euros on tape and 60,000 Euros on disk per year, not including human resources and building costs.

Recommendation: Calculating total cost of ownership, including hardware, operations, human resources, and the creation of usable interfaces.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union.
You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

EU open data

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

In this opinion paper, the European Open Science Cloud (EOSC) Steering Board reflects on how Europe can ensure that its scientific data and data-services – and the knowledge they represent – remain permanently accessible, verifiable, and secure, by reinforcing the EOSC Federation, aligning national and EU initiatives, and embedding legal and operational clarity. Thereby, European sovereignty in research data is all but isolation. It requires the ability to maintain open, fair, and trustworthy scientific collaboration, grounded in resilience, interoperability, and shared values.

Studies and reports

Publications Office
of the European Union