EOSC Sovereignty on FAIR Data

Giorgio Rossi, Dipartimento di Fisica dell'Università di Milano
Italian member of EOSC-Steering Board, co-chair of Policy sub-group
NFFA-Europe/NEP coordinator

EOSC

- The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with **a federated and open multi-disciplinary environment** where they can publish, find and reuse data, tools and services for research, innovation and educational purposes.

- EOSC is not a single monolithic organisation or resource provider but is rather a federation built out of many independent organisations and resource providers as in a system of systems approach. As such, **it ensures the independence and autonomy** of resource providers. (SRIA V1.0, June 2021).

- The EOSC Steering Board (EOSC-SB) has initiated an overarching analysis of the **main disruptive concepts and practices** connected to the construction and future operation of the EOSC.

EOSC Policy Areas

The foreseen benefits of EOSC for European research and innovation require transformative practices in several domains covering from **FAIR-by-design data collection technologies** to **FAIR data and service exploitation skills**.

The Specific issue of **Sovereignty** on FAIR Data was chosen by EOSC-SB based on the perceived urgency to clarify all scientific, legal and operational implications of the "**as open as possible, as restricted as necessary**" paradigm.

An Opinion Paper has been drafted by the Policy Subgroup, elaborated by EOSC-SB level, and then proposed for discussion and criticism to the SB delegations and their national tripartite conferences, like today's one. Subsequently it will be refined and finally **adopted at tripartite level** to become a **living reference for the definition of EOSC policies.**

EOSC Sovereignty on FAIR Data (SFD)

- ***The general idea is that the research FAIR data space is subject to rules, not reducing openness unless necessary, and protecting value of use and reuse.***
- *Data sovereignty is a broad concept having to do with technical and legal frameworks that impact the data, their usage conditions and ultimately their* ***effective fruition as common goods****.*

- Implementing the EOSC principle on access to data "**as unrestricted as possible**" has implications on legal, IPR, and institutional levels that concur to formulating a **pragmatic concept of "Data Sovereignty".**
- The EOSC-SB has elaborated an opinion on the specific *EOSC sovereignty on FAIR research data* by **attributing to *EOSC* and to the participating data-providers the responsibility to protect the value of the FAIR datasets and related data services throughout the whole lifecycle of data exploitation.**

SFD

- EOSC sovereignty on FAIR data (EOSC-SFD) implies **governance and practices oriented at maintaining the value of the FAIR data and services**, by knowing their usage and by protecting the original quality of the FAIR data and its persistence during and after reuse, and by applying effectively the restrictions when needed, including the rights of data providers to withdraw FD sets recognised as faulty or severely incomplete.

- A substantial investment of human and financial resources is required to implement a raw-data set into a FD-set. This effort is typically supported by the data producer being a research infrastructure, and RPO or, seldom, a single user group.

- The defining points of **EOSC-SFD** are:

- the notion that FD are **common goods** governed by rules that must be known and understood;

- the responsibilities to protect FD value of use and reuse **do exist**, and these need to be clearly defined.

Specifications of SFD

a) the protection of rights of the data owners/communities/funders (IPRs, academic freedom, institutional sovereignty, geo-localization, *possession, custody or control* over the data stored in the EU by an international company or by a company with a parent company out of the EU, quality protection) to safeguard the value and investment in making FAIR datasets and in developing/operating FAIR data services;

b) the safeguard of a minimum autonomy/freedom by the research community regarding access to critical services along the data life cycle (including digital services from non-profit public actors);

c) the transparency on data provenance and of how the data have been processed;

d) the right of data providers, according to agreed rules of procedures, to withdraw a FD set based on scientific reasons;

e) the elaboration of appropriate rule settings targeting public or private providers and other users from non-EU and associated countries.

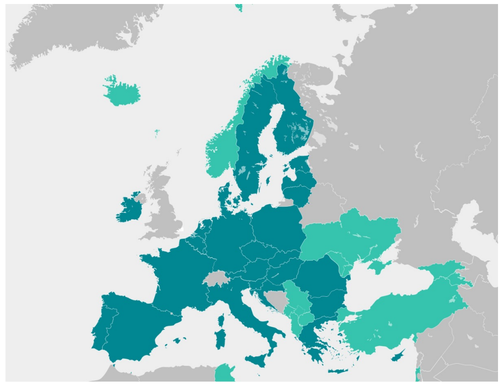
SFD is not a hurdle to open science

On the contrary it is meant to augment the notion of openness, persistent quality, and interoperability of FAIR research data with:

- ownership, in particular for personal or sensitive data;
- effective/adequate control of the data flows and usage monitoring including of where the data are being used and under which laws (national or extraterritorial), where computation on the data is being performed, and where the newly generated data and metadata, e.g., from analysis of FD by users, are stored and become accessible resources;
- applicability of legal rules over data infrastructures, in particular when they are operated in cloud environments operated by public or commercial organizations, including relevant data geo-localisation requirements and specific international agreements on mutual sharing of data, when appropriate;
- transparent relationship to the commercial sector; avoidance of vendor locking solutions and appropriated community ownership of key research infrastructures.

SFD

- The concept of a Single Market for Data would ensure that data from the public sector (including publicly funded research data), businesses and citizens can be accessed and used in the most effective and responsible manner possible, while businesses and citizens keep control of the data they generate, and investments made into their collection are valorised.



- EOSC-SFD should by 2024 become a common practice of:

1. the EOSC Tripartite Governance and all related levels at EC and MS/AC levels;
2. the research communities including RIs, RPOs, learned societies; and
3. the developers and operators of the EOSC platform and the EOSC federation.

SFD recommendations: EOSC level

1. the "EOSC Rules of Participation" must be phased in time to adapt to the evolving quality standards in discipline and research communities; the EOSC-SFD shall be acknowledged in the RoPs.

2. The EOSC Partnership will ensure that the IT-architecture and services shall cope with EOSC-SFD needs;

3. The EOSC Tripartite Governance shall ensure that EOSC-SFD elements are implemented:

a. Update Rules of Participation and onboarding processes;

b. Tag data quality and persistent integrity as a key asset in the EOSC ecosystem;

c. Meta-geo-localization in the EU and ACs of FD services and FD repositories;

d. Ownership of data and services by EOSC partners in a cloud that is overall "known";

e. Promotion of appropriate licensing, preservation of IPR and recognition of authorship of research FD in the EOSC;

f. Implementation of the EOSC Interoperability Framework (including legal and technical interoperability) supporting full FAIRification of 'relevant' licensed data sets;

g. Definition of common guidelines for quality control, skills and stewardship.

SFD recommendations: national level

4. the national contributors to EOSC shall be clearly identified (geo-localized, fully described for capacity, technology, policy, and legal constraints);

5. the list of national EOSC-ready nodes (computing, storage, services compliant with EOSC rules of participation) shall be constantly updated and communicated to EOSC;

6. Investments in skills and data stewardship by national funders should be proportional to the expected growing volume of the FAIR data market;

7. national policies and EOSC-SFD implementation strategies should be regularly evaluated and updated;

8. engagement (dedication) through European and national tripartite events should be accelerated;

9. EOSC-SFD for full compliance w/r national laws and European common practices should be refined and complemented.

SFD recommendations: institutional level (RIs, RPOs, …)

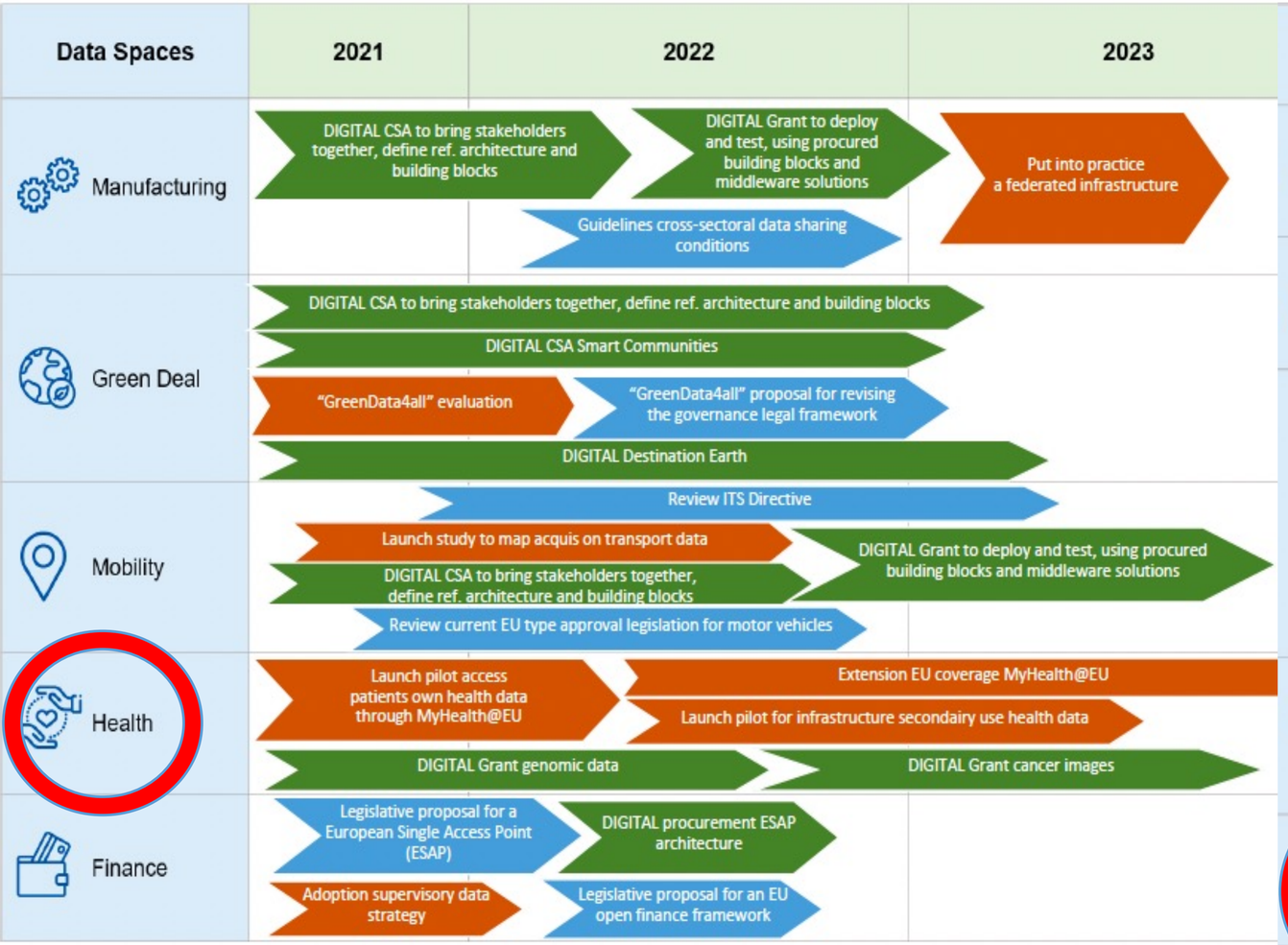
10. the readiness to contribute with contents (FD sets) and services shall be qualified, quantified and made public.

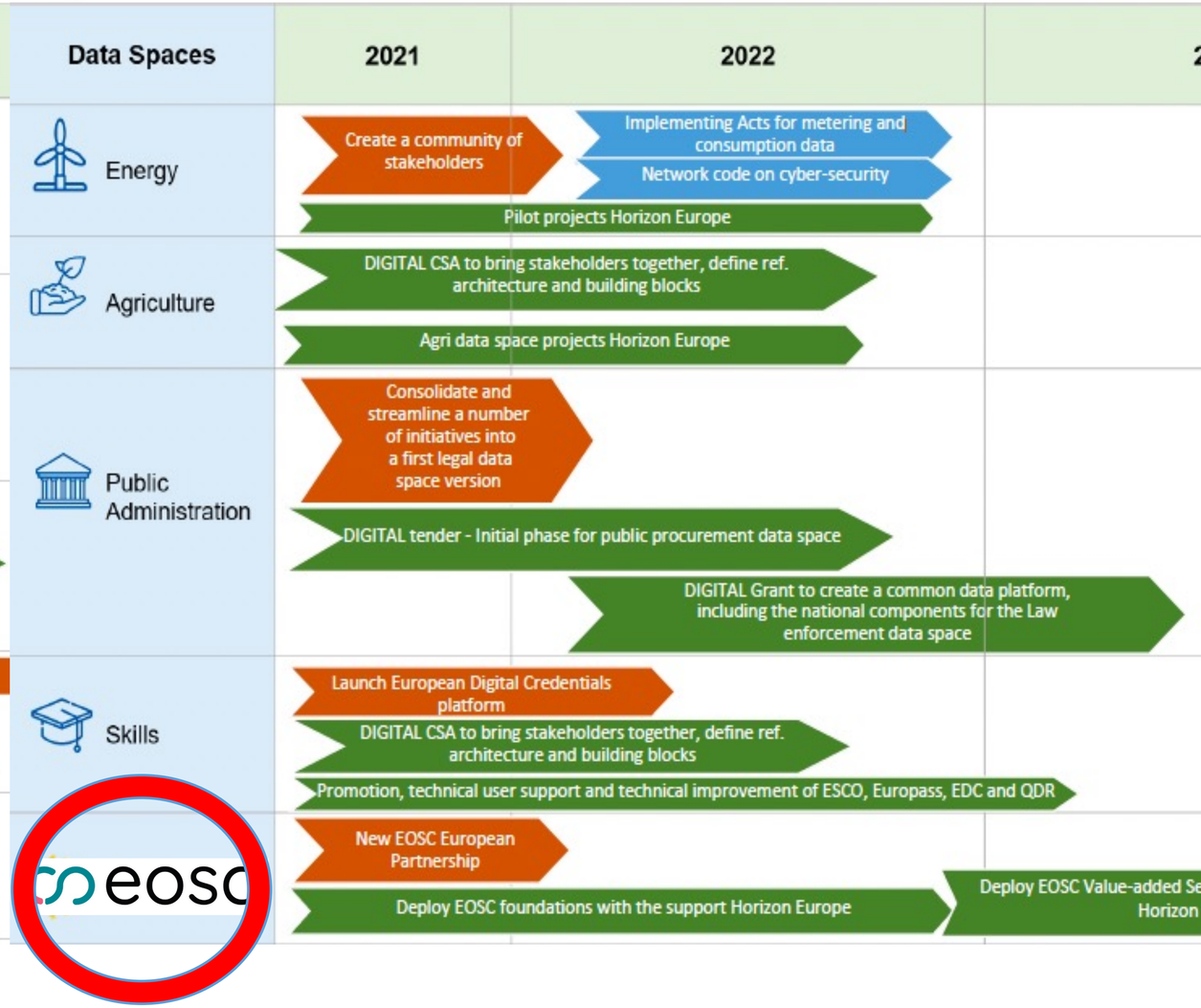
11. sustainable efforts shall be described to cope with FAIR principles, to adopt FAIR-by-design solutions, to invest in data stewards.

12. Institutes shall interact both at the national level and directly with EOSC. (EOSC-A, national mandated org., RIs…).

The **EOSC Steering Board** shall in 2024 evaluate the effectiveness and efficiency of the EOSC-SFD objectives and recommendations.

European Data Spaces: Commission Staff Working Document on Common European Data Spaces





SFD shall apply to all Data Spaces, at least for their FAIR-Data content enablig research through EOSC

Tripartite parties

European Commission

Representing the European Union

EOSC Association

Representing the European research community

EOSC Steering Board

Representing EU Member States and Associated Countries