



Achieving Open Science in the European Open Science Cloud

Setting out OpenAIRE's vision and contribution to EOSC

Position Paper – September 2019

Authors and Contributors

Natalia Manola, OpenAIRE Managing Director (Athena Research & Innovation Centre)

Najla Rettberg, OpenAIRE Network and Outreach Director (Univ. of Göttingen)

Paolo Manghi, OpenAIRE Technical Director (CNR-ISTI)

Mike Mertens, Univ. of Göttingen

Birgit Schmidt, Head of Knowledge Commons, Göttingen State and University Library

Tobias Steiner, OpenAIRE Scientific Advisor, Göttingen State and University Library

Prodromos Tsiavos, OpenAIRE Policy and Legal Advisor (Athena Research & Innovation Centre)

Eloy Rodrigues, Director Documentation Services, University of Minho

Nina Karlstrom, Norwegian Directorate for ICT and Joint Services in Higher Education and Research (UNIT)

Mojca Kotar, University of Ljubljana

Version 1.3

26 September 2019

About OpenAIRE

OpenAIRE is a socio-technical infrastructure for scholarly communication and Open Science. For over ten years it has been supporting Open Science at national levels via its network of experts from key national organizations (National Open Access Desks – NOADs) who support policy development for Open Science within the research realm. Together they train and guide today and tomorrow's researchers in the many aspects of Open Science.

At a technical level, OpenAIRE has developed a range of innovative services, powered by its Open Science Graph - a unique and vast body of knowledge, an interoperable dataset of research results that promotes discoverability and relationships between research artefacts. OpenAIRE's B2B (business to business) and B2C (business to client) services consist of a services layer, a data interoperability layer and an access interoperability layer. By gathering knowledge from all its different data sources, OpenAIRE also plays a role of promoting interoperability - via its technical guidelines - which in turn facilitate the integration of diverse datasets for discovery in EOSC.

Given this backdrop, OpenAIRE is now in a position to bring established and future forms of scholarly communication to EOSC and to move from an abstract, technological infrastructure closer to EOSC being meaningful for every individual researcher in Europe. OpenAIRE is an infrastructure in itself and, by this nature, primarily focuses on B2B processes behind the scenes. But OpenAIRE takes the human part of infrastructure and the local roots of research very seriously and – with topics such as publishing, reputation and assessment -- addresses questions that are of utmost importance to individual researchers.

OpenAIRE AMKE, a not-profit civil partnership company founded in 2018 with currently 16 members, is the legal entity behind OpenAIRE organization overseeing the operations of the OpenAIRE infrastructure, network and services.

OpenAIRE-Advance is a project funded by the European Union (Grant Agreement No 777541). This document is openly available under the CC-BY license.



Table of Contents

EXECUTIVE SUMMARY	5
ACRONYMS	4
1 INTRODUCTION	5
2 OUR VISION FOR EOSC	6
3 OPEN SCIENCE IN EOSC	7
3.1 RESEARCHERS IN THE CENTER	7
3.2 TRANSFORMING AND EMBEDDING OPEN SCIENTIFIC COMMUNICATION IN EOSC	9
4 EOSC DECONSTRUCTED	12
4.1 THE TECHNOLOGICAL LAYER	12
4.2 WHO IS THE HUMAN LAYER?	16
4.3 EOSC - A MULTI-LEVEL ENDEAVOUR	17
5 OPENAIRE AS A KEY PILLAR FOR OPEN SCIENCE IN EOSC	20
5.1 POLICIES AND GOVERNANCE AT THE NATIONAL LEVEL	21
5.2 INFRASTRUCTURES AND SERVICES FOR FEDERATING DATA, FOR FEDERATING OPEN SCIENCE, FOR MONITORING OPEN SCIENCE	22
5.3 COORDINATED TRAINING AND SUPPORT AT THE EU LEVEL	24
6 OPENAIRE CURRENT OUTLOOK AND SUSTAINABILITY WITHIN EOSC	26
6.1 OPENAIRE LEGAL ENTITY	26
7 CONCLUSIONS	28
8 APPENDIX – OPENAIRE ACTIVITIES FOR EOSC IMPLEMENTATION	30
9 REFERENCES	34

Executive Summary

The European Open Science Cloud (EOSC) is Europe's undertaking to address the rising field of data-driven research in a more efficient way, to break down walls in existing siloed approaches for access to services and all scientific content (publications, data, software, methods), and to further evolve the ways in which researchers are able to collaborate. This paper has a twofold role: (i) to strengthen and reinforce Open Science as the *modus operandi* in EOSC, i.e., embedding openness by default in all stages of research, and (ii) to highlight the important role of *open* scientific communication and OpenAIRE participatory infrastructure as a vehicle to deliver Open Science in a trusted and reliable way.

EOSC as a “Research Commons”

EOSC is best envisioned to be a **co-created space**, in which research information is produced and transformed into knowledge. A “Research Commons” that is collaboratively developed and managed by Europe's research communities. Interconnecting both digital and human resources, EOSC adds *new* value to existing infrastructures, via **frictionless data flow, intelligent discovery and retrieval of scientific output**, as well as **homogenized and secure access to services** that transfer, store and analyse data.

Its key characteristics are:

Openness by default: A trusted infrastructure with Open Science practices for transparency, reproducibility and accountability embedded at every phase of the research life cycle.

A network of digital resources: A System of Systems, focussing on *business-to-business* sharing and access. A fit-for-purpose, lightweight *regulatory* framework, to ensure that different players have the best possible conditions to develop on their own, with freedom of contract as a cornerstone.

Researcher-centric: An abstraction layer for data manipulation, including reporting, workflows, parallelism, and persistence, enabling a reproducible ‘digital laboratory’ as a

standard way of working, allowing them to communicate with peers and non-peers alike.

Open governance: A shared arena, maintained and preserved by a range of players in collaboration, governed openly at a national level (via national nodes to strengthen EOSC structures), and guided by the EC.

Training tomorrow's data experts: A coordinated, professional and certified training infrastructure, embedded in institutions and research groups, as part of the European Digital Skills agenda.

Achieving Open Science in EOSC

For EOSC to fulfil its mission and vision, Open Science and specifically Open Access to research results should be a strategic priority. First, this would bootstrap EOSC uptake as it would achieve a critical mass of openly accessible content for researchers to experiment with, stimulating the demand. Second, it would be the key mechanism for linking research to the critically needed in the emerging AI ecosystem, by turning data into *smart and actionable data*: connecting, organizing, and packaging it into making insights available to all (accessible), easy to apply (understandable), and focused on the task at hand (actionable).

To achieve this, we consider three focus areas:

Holistic perspective on data flow: Equally addressing large scale science and scientific research carried out on a smaller scale, is critical for the success of open science in EOSC, both in terms of value and investments, and in terms of researcher uptake.

Researcher in the centre: From gaining traction (what we have now) to gaining widespread adoption of open science, EOSC should pay attention to:

Easiness: Offer services and tools *seamlessly embedded* in research workflows.

Reward: Help with restructuring incentives for data-driven science by keeping track of research production and use.

Skills: Coordinate targeted training and support, addressing all aspects and all levels of the research life cycle.

Commitment: Influence the efficiency of processes and increase sustainability of resources and business models for openness.

Scientific communication as a vehicle for Open Science: The research and cultural diversity of Europe will accept no “one-size-fits-all” solutions for open science. Scientific communication, a long-established research practice, is an apt choice as a common language for understanding and applying open practices across borders and disciplines, with all stakeholders in the loop.

EOSC in this regard should be the catalyst to ensure the delivery of required changes and to establish openness as the new norm in open Scientific Communication:

- Publishing all kinds of scientific products
- Publishing semantic links
- Publishing experiment products as digital packages of workflows

- Innovation in publishing and dissemination practices and methods
- Quality control for securing the quality, reproducibility, FAIRness of research results
- Assessment and rewarding by intelligently combining diverse/open/auditable metrics.
- Monitoring Open Science to ensure its implementation and adjust its aims.

EOSC ‘Deconstructed’

EOSC, similar to any other digital infrastructure, will require a seamless interaction between **technological** and **human** elements to ensure its effective operation and pervasiveness within researcher communities. As a multi-tier structure EOSC interconnects three main key players: **institutions** which implement policies, deliver services and support researchers; **national** structures which shape and align policies and coordinate their connection to EOSC; the **EC** who guides and coordinates efforts.

Technological layer: Key to EOSC architecture is its sustainability, which will *primarily* be driven by: **1-** a participatory design and governance to accommodate different needs and requirements, **2-** shared investments as they are being developed by member states (MS/AC), **3-** the ability to adapt to new technologies and foster innovation. It is therefore critical that EOSC architecture **avoids at any cost a monolithic and centralized approach**, and follow a “System of Systems” approach, where resources are brought together at different levels to deliver data and data services. Emphasis is therefore placed on a **business-to-business (B2B)** sharing (data, services, people) and access, with agreements on:

- 1- A shared policy compliance framework** (i) dictating and applying the rules of how the data

elements are published, shared and re-used, and (ii) implementing an interlinked data space where every research result comes with its context (related entities), provenance (full data and science path) and usage.

2- A uniform delivery of services and data starting from non-regulatory measures, such as recommended *standard contract terms* for services and data access, *specifications* for quality (certification), security and privacy (including trust and identity), accompanied by functions for *categorization* (to simplify browsing), *curation* (to deliver quality), *cataloguing* (to offer intelligent discovery), *crowdsourcing* (to engage researchers for active participation).

The human infrastructure reflects a **balance of policy, training and support implementation** to ensure take-up at MS/AC level and among research communities. It is crucial to highlight the importance of this network, which helps in shaping, adopting and adapting the *shared policies* on national and institutional levels, and ensures that the next generation of researchers is trained in Open Science practices.

OpenAIRE – a pillar for Open Science

OpenAIRE has been supporting Open Science since 2009. It embodies a participatory, service-driven infrastructure anchored on the triplet of **policies – services - training**, contributing the following assets in EOSC:

Policies & Governance: A network of **34 National Open Access Desks (NOADs)**, which comprises experts working on transferring and translating EU policies to a local level. NOADs and their organizations are the *de-facto national nodes for Open Science* in the majority of their countries, already highlighting the connection to EU policies, and ready to step up and become an integral part of **national EOSC structures**.

Infrastructure & Services: A **data infrastructure** enabling open scientific communication, connecting and federating repositories and services across institutions, national settings and RIs, via two key services:

The **OpenAIRE Guidelines for metadata exchange**, a common interface and a key access mechanism (RoP) used by content providers (literature, data, software, methods, workflows) from national infrastructures around the world.

The **OpenAIRE Research Graph**, a decentralized, trusted, interoperable dataset containing all interlinked EOSC resources: research results, who has produced them, under which grant, using which facility and e-science service, how open and fair are they, how has it been used.

Training & Support: An Open Science Helpdesk bringing coherence to the EOSC training and support landscape, by leveraging the unique potential and placement of NOADs to train stakeholders and build local support networks for researchers and data practitioners.

OpenAIRE AMKE, founded in 2018 (counts 16 members as of Aug 2019), is a legal vehicle to engage with MS/AC on committing to setting aside resources for Open Science coordination and activities, as an integral component of their national research infrastructures.

Key Focus Areas & priorities for EOSC

We have identified a set of key recommendations for core activities and functions for an effective implementation of open science and EOSC.

1 - Policies and Governance

EOSC will converge **national support structures** bringing all players together in a collaborative arrangement. No single organization is able to fulfil the Commons approach alone and implementation of Open Science requires

specific handling, as most of the barriers are cultural and organizational.

OPEN SCIENCE IN NATIONAL EOSC STRUCTURES:

Include authoritative open science experts, such as the OpenAIRE NOADs as integral components in national EOSC structures to formulate, align and implement policy: from macro to micro, from EU to local.

OPEN SCIENCE IN THE WIDER NATIONAL AGENDA:

Place open science policy for EOSC as part of a collective set of national innovation strategies and overall governmental agendas to gain: (i) stronger commitment as national agendas are translated into concrete initiatives by other innovation system actors, and (ii) better coordination among different national players, as science transcends ministries' jurisdiction.

2 - Infrastructure and services

For Open Science to succeed in EOSC, we need to: i) provide services for all stakeholders involved in the research life cycle, ii) ensure data federation for both small and big data to become an integral part of EOSC, iii) embed services in institutional settings, and iv) link to international infrastructures.

SCIENTIFIC COMMUNICATION FOR OPEN SCIENCE:

Identify commonalities across disciplines and develop / operate horizontal publishing and research rewarding services. Avoid duplication by finding the touching points of national vs. cross-RI data federation.

LINKED OPEN SCIENCE: Develop and maintain a curated interlinked graph of all research outcomes and make it a **core, public shared resource (asset) of EOSC**. A global and trusted research graph under which many efforts and services converge, and which can be used as an overarching EOSC catalogue used in contextual discovery and assessment of research.

OPEN ANALYTICS FRAMEWORK & INFRASTRUCTURE:

Embed an infrastructure for scientific reward in EOSC from early stages by collecting all types of usage data for all types of resources (citations, usage events for data, services, software). As part of the *Research Commons* and an international undertaking, this infrastructure should place emphasis on B2B aspects (specifications, formats, APIs) rather than indicators themselves.

3 - Support and training

EOSC should develop and rollout a robust and **pervasive training and support Infrastructure**. Grounded on a B2B approach, EOSC must dedicate efforts for **coordination** and **federation**, mixing national and thematic training activities (RIs).

CORE EOSC TRAINING OFFICE FOR OPEN SCIENCE:

Operate a coordination office for the production and collection of materials (courses, modules, curricula, etc.) to be distributed and adapted by members who will perpetuate/propagate the training.

FEDERATION OF LEARNING MATERIAL: Develop and operate a **Learning Material Registry** as a core supporting mechanism for both trainers and trainees in EOSC. Collecting "certified", classified and reusable (open and FAIR) training materials (OER), as these are produced across communities and countries.

TRAINING CERTIFICATION: Establish a credential mechanism that offers possibilities for official recognition (for both trainers and learners).

EOSC OPEN SCIENCE HELPDESK: Organize and operate a distributed and hierarchical structure with the country in the centre, addressing the technical-organizational and legal aspects of data and service Interoperability. Augment with RI/domain specific support.