European Open Science Cloud – BBMRI-ERIC Position Paper

Introduction

Open Science Cloud (OSC) is a long-term strategy to bring European research infrastructure ecosystem to the new level. Albeit big promises and visions, implementation of the OSC still lacks behind expectations. BBMRI-ERIC supports position of other European e-Infrastructures (Position Paper by EUDAT, LIBER, OpenAIRE, EGI, and GÉANT, attached on page 4), but also needs to add certain aspects that stem from the fact that BBMRI-ERIC is an infrastructure focused on medical and healthcare research. Since trust and privacy are of utmost concern for any infrastructure dealing with data of individual persons, the vision of the Open Science Cloud will require certain adaptations to be implemented in practice.

Trust and Privacy – Focus Points for Medical Infrastructures

Access to individual-level data may not be published and shared as entirely open data, in order to preserve privacy of participants, who are typically either patients of clinical facilities or donors for population cohorts. Typical workflow in BBMRI-ERIC is shown in Fig. 1, and needs to deal with the following aspects:

- The sample and data provided by a participant (a donor, a study participant, or a patient willing to share his samples/data for research) is accompanied by the informed consent. There is now ongoing discussion accompanying General Data Protection Regulation (GDPR), whether individuals should be granted right to provide broad informed consent, or whether only narrow/speciﬁc consent should be allowed. The narrow/speciﬁc consent would have serious negative impact on the medical research, as re-consenting of participants results in substantial decrease of response rate from participants and even raises ethical issues.

- Before being allowed to access any individual-level data, the researcher must provide a project proposal, which needs to undergo review by an ethical board. When reviewed favorably, the researcher can ask access to the data or samples. The custodian of the data/samples must ensure, that the project proposal is compliant with the informed consent for the given data/samples. The researcher must accept Data Transfer Agreement or Material Transfer Agreement before s/he can obtain access.

- Only anonymized or highly-aggregated data that have very low probability of re-identiﬁcation of individual participants can be shared beyond the infrastructure of the custodian.

Figure 1: Workflow of data collection, processing, and sharing in medical infrastructures.

• Because of necessary compliance between the informed consent provided by the participant and the research project, the infrastructure must solve the **multi-tenancy** problem. This means that a single researcher may be allowed to access dataset A for purposes of project 1 and dataset B for project 2. However, the user must not use neither dataset A nor dataset B in any other projects (unless explicitly approved), nor s/he may correlate dataset A and B without prior explicit permission.

• As discussed in many recent papers in high-profile journals\(^2\)\(^3\), the biomedical research suffers from **reproducibility** issues. Therefore particular attention needs to be paid to establishing and following well-documented and broadly accepted procedures to ensure high quality of stored biological material as well as data generated from that material or accompanying it.

• Specifically when integrating data from clinical facilities across multiple countries, the medical research suffers from **fragmentation** of regulations on data formats and ontologies in clinical records.

Because of these limitations on data access and processing, the medical infrastructures speak about **fair access** to the data and samples, not about entirely open access. With rapid advances in new analytical methods, such as genomics, proteomics, or metabolomics, the amounts of data generated in the medical field are growing at exponential rate and require establishing of parallel and distributed processing infrastructures to store and analyze the data.


Contributions of BBMRI-ERIC

- Harmonization of European regulatory frameworks, including contributions to GDPR.

- Development of trust-supporting and privacy-respecting federated IT infrastructures specific for biobanks and databanks and processing of data stored therein. These infrastructures will also support secure data sharing complying with the privacy requirements, while dealing with heterogeneity, continuously changing data sources, systematic progression towards data intensive sources.

- Contributions to data format standards, data integration standards and workflows, as well as ontologies translation for relevant existing ontologies, in order to support reliable and meaningful data integration.

- Particular attention is paid to identifiability, quality and reproducibility, both when dealing with biological material, as well as generation of data and its processing. This includes documenting and following standard operating procedures (SOPs) as well as generating and processing provenance information for all the data.

- Interaction with all other relevant infrastructures in order to avoid duplication of development effort, while ensuring that outputs of those infrastructures are relevant and usable for BBMRI-ERIC purposes, or vice versa, that BBMRI-ERIC services and tools are usable for other infrastructures that plan to rely on BBMRI-ERIC.

Prof. Jan-Eric Litton
Director General of BBMRI-ERIC

Assoc. Prof. Petr Holub
Senior IT and Data Protection Manager
& CIO of Common Service IT of BBMRI-ERIC
Summary

As part of the Digital Single Market strategy, the Open Science Cloud will raise research to the next level. It promotes not only scientific excellence and data reuse but also job growth and increased competitiveness in Europe, and drives Europe-wide cost efficiencies in scientific infrastructure through the promotion of interoperability on an unprecedented scale. The Open Science Cloud offers researchers from all disciplines seamless, open access to the advanced digital capabilities, resources and expertise they need to collaborate and to carry out data- and computing-intensive science. Secure and trustworthy, the Open Science Cloud engages researchers in governing, managing and preserving resources for everyone’s benefit. The Open Science Cloud is an open, service-driven endeavour, inclusive of all stakeholders. Governed as a commons, it leverages two decades of public and private investment in e-infrastructures for the benefit of scientific research and innovation.

Background

Science is changing, both in the way it is performed and the way it is communicated. Driven by remarkable advances in information and communication technologies, today’s scientific infrastructures offer researchers unprecedented access to data sources, data-intensive sensors, and increasingly comprehensive analysis and simulation facilities that have revolutionized scientific methods in a remarkably short space of time. Research services, processes and outputs are becoming accessible to all levels of society. Enormous amounts of data are being generated, bringing extraordinary new opportunities for their innovative reuse in novel scientific, commercial, and citizen-science contexts. This is Open Science.

Open Science is a key driver, not only of scientific progress, but also of economic and societal innovation. To harness its full value and reap the fruits of public and private investment, Europe needs to foster an open, collaborative platform for the management, analysis, sharing, reuse and preservation of research data on which innovative services can be developed and delivered. For this, Europe can build on decades of public investment in scientific infrastructures—experimental facilities, networking, high-performance and high-throughput computing, cloud services, scientific software and institutional and community data repositories—by connecting national and international infrastructures and services. The Open Science Cloud is the vehicle to achieve this vision. Below we articulate the eight essential elements it needs to succeed.

Many of the resources and services needed for the Open Science Cloud already exist; while technical challenges remain, most of the barriers are ones of policy and concern funding, lack of interoperability, access policies and coordinated provisioning. The Open Science Cloud will address these issues and enrich and further advance the portfolio of resources and services to make the entire scientific lifecycle more open and transparent. To this end, governance of the Open Science Cloud will be modelled after the governance of the Internet, conducted by a decentralized, international group of stakeholders drawn from across research and civic society, from both public and private sectors. The Open Science Cloud’s governance will hold custody of the shared services, policies and standards that maintain its persistency, its global interoperability and its adherence to the Open Science vision. By involving all the relevant stakeholders who support today’s research—funding agencies, policy makers, research infrastructures, e-Infrastructures, libraries, data providers and service providers—the Open Science Cloud will significantly impact the way research is done in Europe and will put European research at the forefront of Open Science globally.

The Open Science Cloud: Eight Elements for Success

I. Open: This is the driving principle of the Open Science Cloud: openness in design, in participation and in use. The Open Science Cloud will be based on open access and promote the development and adoption of open standards, enabling collaborative environments with no artificial barriers to participation or resource-sharing by any stakeholder. It will enable accessibility, transparency, and reproducibility in all stages of the research life-cycle. Having a flexible open design, the Open Science Cloud will foster public-private partnerships, turning all investment into economic growth.

II. Publicly funded & governed: A publicly funded and publicly governed Open Science Cloud will guarantee persistence and sustainability, and ensure that outcomes are driven by scientific excellence and societal needs rather than profit. This "commons" approach, welcoming partnership with private-sector actors while driven by the public good, will encourage the development of innovative services that are conducive to the future of Open Science, while guaranteeing the long-term, persistent care of resources.

III. Research-centric: Following the true spirit of agile co-design and participation, researchers and research communities—including those from the private sector—will be fully engaged in the design of the Open Science Cloud, to ensure the development of services responsive to their needs.

IV. Comprehensive: The Open Science Cloud will be universal, specific to no single scientific discipline or research field. It will promote inter- and multi-disciplinary science and encourage innovation and integrated knowledge creation among all research communities, also capturing the long tail of science and citizen science.

V. Diverse & distributed: The Open Science Cloud will leverage the richness of Europe’s distributed e-infrastructures, encompassing a resilient network of actors, resources and services organized nationally and at the European level. Embracing diversity through openness, the Open Science Cloud will drive a more efficient use of ICT investments across infrastructures and communities, addressing the digital divide and lowering the barriers to adoption for institutions and researchers.

VI. Interoperable: Through the promotion and adoption of common standards and protocols for all resources and digital services, the Open Science Cloud will connect networks, data, computing systems, software, tools and services for research as seamlessly as the Web connects information.

VII. Service-oriented: The Open Science Cloud will be protocol-centric and service-oriented. It will provide services that address the full research lifecycle, including data gathering, management, analysis, sharing and discovery. The Open Science Cloud will be the framework and testing environment for new, innovative methodologies and services that further advance research in the Open Science context.

VIII. Social: The Open Science Cloud will be a socio-technical endeavour that connects diverse communities and promotes the development of human networks. By adopting community-based rules and procedures with incentives for sharing and responsible use, it will enable the sharing of knowledge and facilitate the embedding of Open Science practices into researchers’ everyday workflows. This will require a strong social dimension of consultation, outreach, advocacy, training and support, in an ecosystem of local, national and international programmes.

Dr. Kimmo Koski, Project Coordinator, EUDAT
Kristiina Hormia-Poutanen, President, LIBER
Prof. Mike Chatzopoulos, Project Coordinator, OpenAIRE
Yannick Legré, Director, EGI
Dr. Bob Day MBE, Chief Executive, GÉANT