open research data National Strategy Council

Enhancing Open Research Data in Switzerland

Analysis and recommendations from the ORD Sounding Board of Service Providers on "Data Archiving & Sharing", "User Access", "Technical Interoperability", "Reuse" and the European Open Science Cloud

Mandator
Swiss National ORD Strategy Council

ORD Sounding Board of Service Providers
Christoph Witzig, chair, Switch
Abdel Benhauresch, NICT
Christophe Dessimoz and Lucy Poveda, SIB
Chiara Gabella and Clemens Trautwein, SLiNER
Georg Lutz and Bojana Tasic, FORS
Oksana Riba Grognuz and Olivier Verscheure, SDSC
Rita Gautschy and Ivan Subotic, DaSCH
Thomas Geiger, SPHN

Date
August 2024

1 Content

2 Executive Summary

Based on the mandate from the ORD Strategy Council of 23.9.2023, the Strategy Council Coordination Group has identified the following four areas of inquiry to be explored by the Sounding Board of Service Providers (SB SP): “Data Archiving & Sharing”, “User Access”, “Technical Interoperability” and “Reuse”. The identified areas are intended to contribute to on-going discussions and upcoming decisions of the Strategy Council. On 7.5.2024, the Strategy Council issued a mandate to obtain a view on the European Open Science Cloud (EOSC) from the Sounding Board of Service Providers.

Given the strong interconnectivity of these five topics, the SB SP has decided to address them in a single report in five chapters that can be read individually.

Data archiving and data sharing are key elements for putting FAIR principles into practice. Having reliable and high-quality data is not only a requirement from funders, who want to make sure that publicly funded data collections are a public scientific good that can be used by the research community at large. It is also essential to ensure replicability, reuse, and innovation, and it is more and more also the basis for high-quality AI technologies. Many different elements need to be in place to allow for broad archiving and sharing of research data. Clear and common policies addressing which data requires archiving and how it should be shared would be beneficial to the scientific community in Switzerland. Such policies need to provide solutions that allow data access while respecting data privacy and protection of intellectual property rights. Repositories that ensure high-quality data and support services must be operational with sustainable funding and should be able to cover discipline-specific needs. Repositories need to follow best community of practice standards and FAIRification processes, to enable interoperability and allow the entire data repository system to be FAIR and interconnected. Finally, there need to be better incentives and rewards for researchers to comply with data archiving and sharing requirements, as well as support in research data management and sharing to help them.

User access is meant as giving access to all national and international users while respecting the fact that not all data can be open (“be as open as possible and as restricted as necessary”). Requiring that metadata must be freely available while the data itself may not, was identified as an important lever to bring restricted data into balance with FAIR principles.
Access was analyzed along five criteria: a) strategic issues and policies, b) barriers to access with corresponding countermeasures, c) incentives for access, d) technologies for access and e) cost models. Key findings include that various policies need to be harmonized and aligned nationally as well as internationally. Various barriers, among them lack of findability, cumbersome access controls, legal and cultural constraints and inadequate cost models were identified. On the other hand, incentives for making data accessible exist and should be exploited, such as providing professional and sustainably operated repositories with proper support, monitoring and publicly sharing information about data access as well as acknowledging researchers who make their data accessible.

Technical interoperability is understood as the ability to exchange, combine, and interpret research data stored in different systems. It is crucial for enabling researchers to utilize multiple complementary research data infrastructures across the data lifecycle. The analysis examined interoperability across the following three levels: a) metadata level, b) data level and c) service level. At the metadata level interoperability is absolutely necessary and has to be in line with international standards like those elaborated by EOSC. At the data level, interoperability should mainly be striven for at the semantic level, with improvements on technical interoperability, which will provide additional benefits. The service level is crucial for supporting researchers such that they can focus on their individual research interests without compromising the benefits for the research community at-large.

Reuse in the context of this work is understood as the ability for hardware, back-end services, software components, as well as full-stack services to be shared between more than one research data infrastructure. Reuse is not meant as reuse of data, which in itself is an important topic. Reuse was analyzed in four different layers: At the IT infrastructure layer, the application layer, data layer and business/service layer. It was found that current funding model, the lack of knowledge and collaboration, the fear of reusing a non-sustainable component or service as well as the "not-invented-here syndrome" are the main obstacles for reuse.
While a diverse array of services exists, consolidating them into a common shared system is not feasible. On the other hand, embracing an ecosystem approach could transform the current landscape. A collaborative culture, adherence to best practices and standards, and harmonized processes are vital to enhancing the reuse of services and components. Transitioning to national-level services can break down institutional silos, leading to substantial cost savings, improved service reuse, and greater sustainability. Realizing this potential, however, demands funding models that align with these goals.

With the **European Open Science Cloud** and the corresponding EOSC Association, the European Commission has initiated a highly ambitious large program to advance open science in Europe. Participating actively in the building and development of EOSC and related initiatives at the European level such as e.g. the Common European Data Spaces is important from the SP's view. The ORD activities in Switzerland should align with these European activities as science does not stop at national borders. However, it should also be acknowledged that the SP and other Swiss organizations have already been participating for several years in typically domain-specific European Research Infrastructures. The collaboration and alignment with EOSC should be building on these existing international activities to avoid duplication of efforts. Therefore, the SB SP recommends adopting a cautious but collaborative approach to align the Swiss ORD efforts with EOSC, particularly since many concepts, organizational structures and services of EOSC are still being developed and likely to evolve significantly over the coming years.

Recommendations were obtained based on the analysis of the five topics archive and sharing, access, interoperability, reuse and EOSC. In our view, the key recommendations concern a) common minimal standards for repositories, b) harmonizing governance and funding models, c) ensuring common policies, d) aligning mandates for national service providers along the entire research data lifecycle and e) a common national approach towards EOSC while building on already existing international efforts of the Scientific Communities. We refer to section 3 for the full text of the recommendations.

3 Recommendations

Based on the analysis of the situation the ORD SB SP makes recommendations which can be divided into the following thematic sub-groups: a) General ORD statements, b) Governance and Funding, c) Policies, d) Incentives and Professional Support, e) EOSC and f) Topic-specific Recommendations.

3.1 General ORD Recommendations

1. **Open repositories**
   Whenever possible, data must be deposited in open repositories, particularly those tailored for particular data types, as these tend to maximize FAIR prospects. A list of repositories is provided by SNSF[1] as well as other institutions[2], but could be maintained on the national level by the ORD Strategy Council in the future according to guidelines issued by the Strategy Council.

2. **Standards for repositories**
   ERI actors should ensure that repositories provide professional services and adhere to common minimal standards for:
   a) **Professional data curation:** Data repositories without active and professional data curation and the respective support services are not desirable, because only well-documented data with rich metadata are FAIR. Data curation and thorough documentation are key for ensuring minimum quality of the data and making them valuable and usable for the broader research community.
   b) **Metadata standards:** A core set of standards for rich metadata across disciplines and for data documentation is essential and must be defined. This set of standards should align with the requirements of the EOSC ecosystem, ensuring interoperability and consistency. Clear guidelines would help maintain the quality and accessibility of data, facilitating its effective use within the broader research community.
   c) **Standards for components**: The adoption of standards for both, the Application Layer and the Data Layer components (see chapter Reuse) should be fostered. Emphasizing the use of standardized web components and modular architectures based on microservices should be a key objective. The sustainability of these external components, which are often developed and maintained by other institutions, is essential for effective reuse. Ensuring that these standards are robust and widely adopted will enhance the interoperability and sustainability of the systems built upon them.
   d) **Interoperability:** Repositories should be interoperable not only within Switzerland but also with the broader European research data ecosystems, such as EOSC, the Common European Data Spaces and other European and international research data infrastructures. This will facilitate the efficient exchange and reuse of data across national and international boundaries, fostering collaboration and innovation in research and industry.

   The SB SP would welcome if the Strategy Council supports common core standards for repositories and initiates a task force charged to develop them.

[1] https://www.snf.ch/en/WtezJ6qxuTRnSYgF/topic/open-research-data-which-data-repositories-can-be-used
[2] See for example https://www.epfl.ch/campus/library/services-researchers/data-publication/data-repositories-and-related-platforms/ or https://www.nature.com/sdata/policies/repositories

3. **Free access for researchers to repositories**
Given that the real value of open data lies outside the data producer, it is essential that depositing and accessing curated data is free of charge for researchers, unless specific services are needed. This requires that the national research data infrastructures and services are sufficiently financed on a sustainable level (see below). Researchers, possibly with support of RDM professionals such as data stewards, bear the responsibility to prepare and document their data, so they can be archived and shared. To avoid that this critical yet often undervalued task gets overlooked and neglected by researchers or goes at the expense of the research project, research grants must include adequate funding and monitoring to ensure research data is made FAIR. Researchers and data archiving/repository organizations need to work hand in hand to ensure that data can be archived and shared according to FAIR principles.

4. **Exchange forums**
Permanent exchange forums between service providers of national importance such as the SB SP should be established to identify low-hanging fruits for facilitating transfer between platforms and to systematically make infrastructures and services interoperable. We envision an extended group of service providers to be part of such a forum.

5. We further recommend data reuse to be investigated in a separate mandate in collaboration with the Sounding Board of Researchers (as outlined in the section Scope in the chapter Reuse).

3.2 Governance and Funding

1. **Common governance model**
In order to provide sustainability and thus stable funding for the national data infrastructures and services, ERI actors should agree on a common governance model for all national services.
    a) Such a model should follow the principle that those who fund research should also jointly fund and govern national data infrastructure and national services. A joint legal basis for steering and funding of national data infrastructures and national services including all ERI actors at the national level should be implemented.
    b) There should be clarity on who is expected to pay what when it comes to data storage and services around data archiving.
    c) Research data infrastructures should not be built project-based through open calls for higher education institutions. Existing national service providers should be eligible for national funding programs such as federal project contributions[3] under the same conditions as higher education institutions, as they are most suited to provide national services.
    d) Where feasible, existing data also from non-research sources ("Real World Data", e.g. from routine healthcare or public administration) should be mobilized for research projects instead of generated anew (once-only principle). The necessary data curation efforts by data providers and researchers and its maintenance must be eligible and adequately reflected in research grants. In the long term, this will reduce costs and improve data representativeness over generating research data from scratch.

2. **Funding model fostering reuse of data infrastructure and services**
A funding model appropriate for fostering reuse with incentives and rewards for reuse and collaborative efforts between institutions, promoting national services should be set up by:

[3] https://www.sbfi.admin.ch/sbfi/de/home/hs/hochschulen/finanzierung-kantonale-hochschulen/projektgebundene-beitraege.html

a) prioritising national services over institutional ones to ensure broader access and consistency,
b) requiring detailed explanations and impact on reuse of data in proposals,
c) promoting funding mechanisms for OpEx vs. CapEx (Operating Expense vs. Capital Expense) to ensure long-term sustainability,
d) demanding adherence to common standards, coordination with existing RDIs and interoperability in proposals.
e) Requiring that funding for a new infrastructure only becomes available after it has been shown that it adds real value and that existing infrastructures don't meet the need. In addition, it should demonstrate convincingly that it can be operated subsequently on a sustainable basis.

The SB SP believes that governance and funding models are key for the success for ORD. It recommends to the Strategy Council to initiate a corresponding process for devising these models, fostering the reuse of data infrastructures and services at the national level and adhere to these principles in further strategic recommendations.

3. **Monitoring of impact and potential for reuse**
   a) ERI actors should ensure monitoring the impact and use of data infrastructure and services and eliminate redundancies.
   b) Solutions for transversal activities (e.g. data science, data archiving) should be identified, assessed, and reused.

3.3 Policies

1. **Joint policies and guidelines**
   All Swiss ERI actors should have joint policies and guidelines in place that clearly outline requirements for data management, data archiving, data sharing, data curation and documentation. These must accommodate discipline-specific needs and standards to ensure that data is consistent, usable and accessible across different research institutions and domains.

2. **National repositories**
   Repositories should have a national and international scope and relevance avoiding those serving institutional or local needs. Selected national repositories are required due to the nature of the data and/or the service they provide, though their number should be limited to prevent duplication of investments and operating costs. Criteria should be formulated in order to guide allocation of national resources, as outlined in recommendation 3.1.

3. **Clear mandates and responsibility assignment for interoperability**
   ERI actors should jointly identify and assign responsibility and clear mandates to those organizations designated to be in charge of interoperability. These organizations are responsible for bringing together the key actors to find consensus on minimal requirements and common standards essential for data sharing and collaboration.

4. **Alignment with European Data Space and EOSC ecosystem**
   The above proposed joint policies and guidelines should be aligned to the Common European Data Spaces and EOSC ecosystem to facilitate the interchange with the wider scientific community. Especially since achieving interoperability within EOSC is pivotal to provide added value for service users through the federation of services that will compose the EOSC landscape.

5. **Making existing non-research data accessible for research**
   Commercial companies as well as governmental agencies also provide access to open data[4] but non-research data is often difficult to access for research purposes. ERI actors should engage to promote and improve access to existing public and private data for research, including the participation in-going legislative efforts for secondary use of data.

The SB SP recommends that the Strategy Council initiates a process for devising common policies at the national level.

3.4 Incentives and Professional Support

Acknowledging the current ORD action plan measures C and D, we propose the following recommendations to reduce barriers to ORD:

1. **Recognition in career and project funding decisions**
   Incentive structures for individual researchers with respect to data (incl. software) sharing and reuse through research assessment need to be put into practice, which means a *de facto* recognition in career and project funding decisions. This includes recognizing qualitative and quantitative evidence of data reuse, e.g. through citations in academic publications, to give credit to data creators and to encourage more researchers to share their data.

2. **Investment in interoperability**
   Buy-in from researchers should be increased with incentives to invest resources in interoperability. This includes a recognition of such attempts for career promotion and funding of attributable costs for making data FAIR.

3. **RDM training and support at the institutional level**
   Training and support to raise awareness for research data management and data archiving should especially be provided to early career researchers. Training and support for researchers and data managers concerning best practices for data management and sharing should be offered at the institutional level in line with national policies and standards. Data stewardship should be a recognized and promoted career pathway at HEIs and organizations hosting repositories. Furthermore, the existing professional profiles should be expanded and investments in capacity building taken, e.g. for semantic experts and programming support.

3.5 European Open Science Cloud

International relations are key for science. The European Open Science Cloud (EOSC) is an effort at the European level that is relevant for establishing an ORD landscape in Switzerland. Hence, we would like to put forward the following recommendations with respect to aligning these national and European initiatives:

[4] e.g. cloud providers such as amazon (https://aws.amazon.com/opendata/?wwps-cards.sort-by=item.additionalFields.sortDate&wwps-cards.sort-order=desc) or governmental efforts such as l14y (https://www.i14y.admin.ch/en/home)

1. **Recommendation for a national effort to position the Swiss landscape towards EOSC:** The SB SP sees EOSC as essential to structure and build the European Research Data Space and the Swiss government should ensure that Switzerland is fully integrated in this data space in the future. EOSC is therefore of high strategic relevance for the Swiss research community. A coordinated national effort is needed to comprehensively map out the flourishing and evolving Swiss EOSC landscape. This effort should
    a) include the identification and engagement of all relevant national actors, such as leading research institutions, universities, libraries, higher education institutions, research infrastructures and specialised data centres from the different scientific domains or clusters.
    b) distinguish between the roles of these actors and the specific initiatives or projects they are involved in, such as data management projects or collaborative platforms.
    c) be facilitated by a mandated 'working group' or some national body such as the ORD Strategy Council, which could lead the process of addressing any major gaps in the national EOSC landscape and ensure a cohesive national strategy.

2. **The role of a (potential) national EOSC node for Switzerland:**
   The SB SP is of the opinion that at the European level the concept of an EOSC node is evolving and relations and responsibilities between national and the thematic nodes are to be defined. Therefore, before taking any rushed decisions, a national discussion on the strategic value of a national node should be held first with the corresponding actors to assess the value, role and shape of a potential Swiss EOSC node. Thus, targeted actions to create a national node should only start once the tasks and added value of such a national node are clearly defined and understood, which is not yet the case. Proceeding prematurely risks duplicating efforts with minimal benefit.

   In the meantime, the SB SP recommends adopting a collaborative approach to developing a national node, comprising all Swiss EOSC stakeholders as equal partners with national focus. These stakeholders should work to better align their services at the national level, follow the developments at the EU level and identify gaps that cannot easily be addressed. To identify these gaps, a multi-faceted approach could be employed comprising of surveys, stakeholder consultations and comparative analyses of existing infrastructures and international relations. These gaps ultimately highlight the areas where a national node can truly add value.

3. **Representation of Swiss Interests towards EOSC:**
   First and foremost, a full association of Switzerland not only to Horizon Europe, but also to Digital Europe would be the best way to support EOSC related activities of the SP. Lacking this association, we recommend the following measures:
    - A stronger coordination role for SERI with respect to Swiss nodes (ERIC, ESFRI, …)
    - A mandate from SERI to participate stronger in EOSC activities with possible funding support
    - A visible alignment and coordination between the national ORD effort and EOSC activities. As first step, the national landscape could be mapped to EOSC activities to identity strengths and gaps (see recommendation 1).

3.6 Topic-specific Recommendations

1. **Access: Sensitive data**
   Access to sensitive data is a special case for ORD. It is relevant not only for health data but also social sciences. Common policies, infrastructures, services and standards should be in place to

facilitate access, archiving and (re-)use of sensitive data.

2. **Reuse: Developer platform and community of providers**
   A dynamic developer platform to facilitate the sharing, discovery, and plug-and-play reuse of components and services should be set up. Furthermore, a "community of service providers" to leverage this platform as a collaborative hub, fostering engagement and collaboration between service providers and researchers should be established and promoted for regular exchange and share best practices, infrastructure components and experience.

4 Introduction

Based on the mandate from the ORD Strategy Council of 23.9.2023, the Strategy Council Coordination Group has identified the following four areas of inquiry to be explored by the Sounding Board of Service Providers (SB SP): “Data Archiving & Sharing”, “User Access”, “Technical Interoperability” and “Reuse”. The identified areas are intended to contribute to on-going discussions and upcoming decisions of the Strategy Council. On May 7, 2024, the mandate was given to the SB SP to express its views on various aspects of the European Open Science Cloud and its impact on the national ORD scene.

Given the strong interconnectivity of the initial four topics, the SB SP has decided to address them in a single report. The four topics are treated as independent, self-contained chapters that can be read individually. These chapters have the same structure: an introduction, some comments on the scope of the topic, the analysis followed by conclusions. The EOSC report was subsequently added as the fifth chapter in order to have a comprehensive report. Finally, the SB SP issues its recommendations for all five topics of this report in the third section of the report, following the executive summary.

The sounding board proceeded as follows in establishing this report. The group first divided into working groups, each with a designated chair. Each working group was responsible for drafting different sections and proposing recommendations. The groups met approximately twice to prepare the initial drafts. These drafts were then shared for comments and discussed with the entire sounding board to assess whether there was broad consensus on the covered topics. After discussion, the sections were revised and merged into one document. Subsequently, the entire document and the joint recommendations were consolidated into a single comprehensive report, including an annex with the terminology used. The report was revised once comments and questions from the ORD Coordination Group and Strategy Council have been received.

Members of the sounding board were permitted to share the drafts with colleagues within their respective organizations if they sought additional expertise. However, the document was not shared outside the sounding board, as it was believed that there was sufficient expertise within the group. Overall, there was a broad consensus among the sounding board members on the various topics discussed.

In the rapidly evolving landscape of scientific research, Open Research Data (ORD) has emerged as pivotal, essential for enhancing the impact of research and fostering scientific innovation through the reuse of existing data. It also contributes to the transparency and reproducibility of the research process, thereby building trust in scientific results.

This document focuses on the infrastructural and services-related challenges and strategies associated with ORD. It acknowledges the interconnectedness of data access, interoperability, reuse, archiving and sharing, addressed in subsequent chapters in detail. All these aspects are crucial to making research data FAIR (findable, accessible, interoperable, reusable) in practice.

Given the topic's relevance, the main Education, Research, and Innovation (ERI) actors have formulated policies to mandate archiving and sharing of research data, marking a cultural shift towards open science. For example, the SNSF requests researchers to provide Data Management Plans (DMP) that include how data will be made accessible at the end of the project and "expects all its funded researchers

- to store the research data they have worked on and produced during the course of their research work,

- to share these data with other researchers, unless they are bound by legal, ethical, copyright, confidentiality or other clauses, and
- to deposit their data and metadata onto existing public repositories in formats that anyone can find, access and reuse without restriction."[5]

Similar policies are in place at swissuniversities, the Swiss Academies of Arts and Sciences, many higher education institutions and at the European level for Horizon Europe projects or in the realm of the European Open Science Cloud (EOSC). Increasingly, journals also request or recommend that replication material be available for published articles and/or the review process.

ORD principles expand the research data lifecycle, enabling datasets to be reused by third parties beyond the original data producers for new purposes, such as additional research projects.

The researcher must already plan and prepare to make the deliverables of a project compliant to FAIR principles before and during the execution of his project. In the final stages of the project, it encompasses the management, storage, preservation, sharing, and publishing of data. In addition, it also involves all measures to make sure this data can be discovered, reused and properly cited.

With regard to the Research Data lifecycle, several key requirements are identified across different stages. The FAIR principles define the requirements for data, metadata and data documentation, and infrastructures to ensure machine-actionability, meaning the ability of computational systems to find, access, interoperate, and reuse data with limited human intervention. For data access, the governance, and technical requirements need to be transparent, clear, and conform to FAIR standards and common practices. Data integration requires semantic and technical interoperability and is facilitated by harmonization of data and metadata and documentation standards. Demands for comprehensive documentation and rich metadata, need, however, to be balanced with the burden for researchers to comply with such standards. The analysis stage emphasizes scientific collaboration without undue governance or technical constraints. Publishing results mandates linking to source data with unique identifiers and ensuring datasets are FAIR compliant. Finally, when depositing data in a repository, it must be verified for FAIR compliance, findable through open catalogs, and maintained with appropriate infrastructure and processes for long-term data accessibility.

As depicted in Figure 1, the Research Data Lifecycle exhibits notable complexity, emphasizing the need for interoperability at various levels to ensure (meta)data are FAIR. Integration and support by data stewards, IT and other support staff is crucial to facilitate a smoother and more efficient journey for researchers navigating the various stages of the data lifecycle, thereby enhancing their capability to manage and effectively utilize diverse infrastructures.

(Open) Research Data Infrastructures (RDIs)[6] support this extended data lifecycle by providing researchers with not only necessary infrastructures, but also a diverse set of surrounding services and expert knowledge beyond researchers' domain-specific expertise (e.g., on data privacy and security, data (access) management, curation, analysis, etc.).

[5] https://www.snf.ch/en/dMILj9t4LNk8NwyR/topic/open-research-data

[6] (Open) Research Infrastructures must clearly be distinguished from research infrastructure facilities where data is initially generated by various scientific instruments (such as synchrotrons, microscopes, etc). See also the ORD Concept paper on Research Data Infrastructures at https://openresearchdata.swiss/wp-content/uploads/2024/07/Concept-Paper-StraCo_V4_2023-10-23-5.pdf

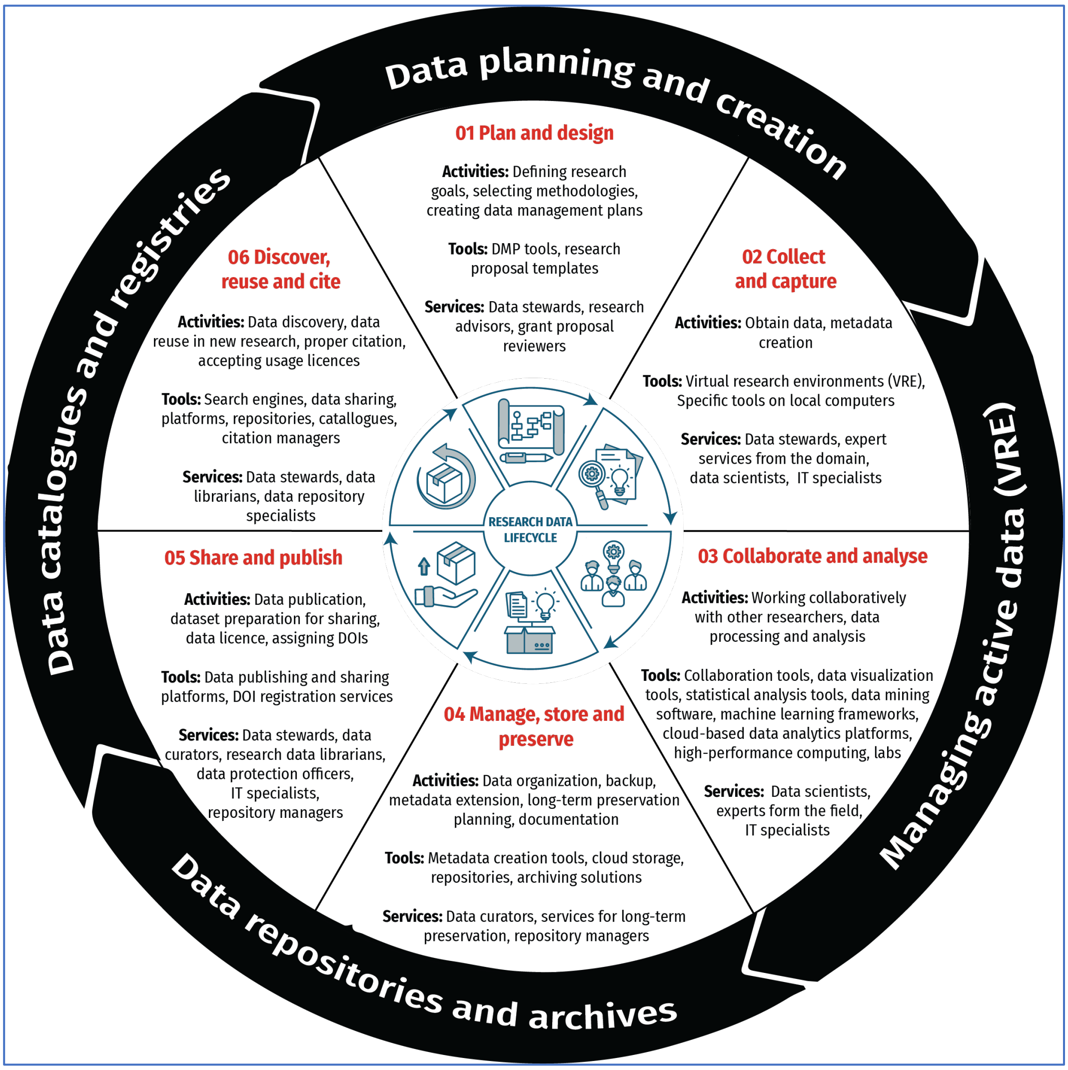

Figure 1 Research Data Lifecycle Ecosystem[7]

An efficient ORD ecosystem must have the capability to exchange, combine, and interpret research data stored in various systems and the ability of researchers to use multiple, complementary RDIs for different parts of their data management cycle with minimal friction in data transfer, process automation, and overall user experience.

This paper first outlines the requirements for archiving and sharing data, accessing data, and discusses issues around interoperability between different stages of the data lifecycle. It also addresses interoperability within the same level, as well as the reuse of existing tools and infrastructures. Open Science is an international endeavor. With the European Open Science Cloud, the European Commission has initiated an effort at the EU level, which must be taken into account at the national ORD level. The last chapter is dedicated to the alignment of ORD with EOSC.

[7] All figures by B.Tasic

The SB SP is aware of the ORD Action Plan and understands that some (but by far not all) of the issues raised in this report are related to elements of that plan and its corresponding call for projects. References are made in the text where appropriate.

This report has two appendices: a list of abbreviations used and a glossary of terms with an explanation how they are used within this paper.

Last but not least, it must be pointed out that this report reflects the views of the ORD Sounding Board of Service Providers.

5 Archiving & Sharing

5.1 Introduction

In the rapidly evolving landscape of scientific research, the archiving and sharing of research data have become pivotal and are seen as essential not only to enhance impact of research and foster scientific innovation through reuse of existing data, but also to contribute to transparency and reproducibility of the research process and hence to build trust in scientific results.

The fast speed of technology advancement requires constant efforts to preserve digital data in a proactive way to ensure that data remains available. Without these efforts, valuable research data risks being lost or becoming inaccessible, highlighting the importance of forward-thinking strategies in archiving and sharing research data.

When talking about research data archiving, we often think first and foremost about solutions such as reliable digital storage and backups. However, as shown in Figure 2, the storage is only one small part of the comprehensive Data Archiving System. Archiving of research data encompasses a range of activities designed to ensure that digital research data and information remain accessible and understandable for future generations. It involves strategies and actions to protect digital materials from obsolescence, degradation, and loss, guaranteeing long-term access to data.

Besides archiving the original research data, it is also important to capture information about its context (rich metadata), and any relevant documentation and materials, as this will enable future researchers to use and understand the data effectively. Therefore, an appropriate archiving approach requires collaboration between the data curators and the data producing researchers. This collaboration often necessitates that data curators possess a deep understanding of the specific research domain to conduct accurate and thorough data quality evaluations.

Additionally, archiving research data extends beyond technical solutions, encompassing organizational and resourcing challenges over time, such as risk assessment, sustainability planning, and technology development that requires skills development and change management.

This proactive approach safeguards the cultural, scientific, and historical value encapsulated in digital formats, ensuring their availability for ongoing and future research, education, and cultural enrichment.

5.2 Scope

Archiving and sharing of research data is an active process that requires managed systems and services that support researchers in depositing data and others to access the data. Archiving and sharing of research data need to be thought through in parallel, since archiving for the sake of archiving without the perspective of others reusing the data is not consistent with ORD principles. Therefore, in this paper, we focus on both archiving and sharing to emphasize the importance of holistic approaches to research data management.

Repositories are established at different levels – some are institutional (e.g., at universities), some are national, and some are European or International. Repositories also vary with respect to their scope: some are generalist, some are discipline-specific, and some are a combination of the two, with a generalist model but discipline-specific adaptations. Typically, funders recommend data repositories that are commonly used by the Swiss research community and fulfil the Open Research Data (ORD) criteria.[8]

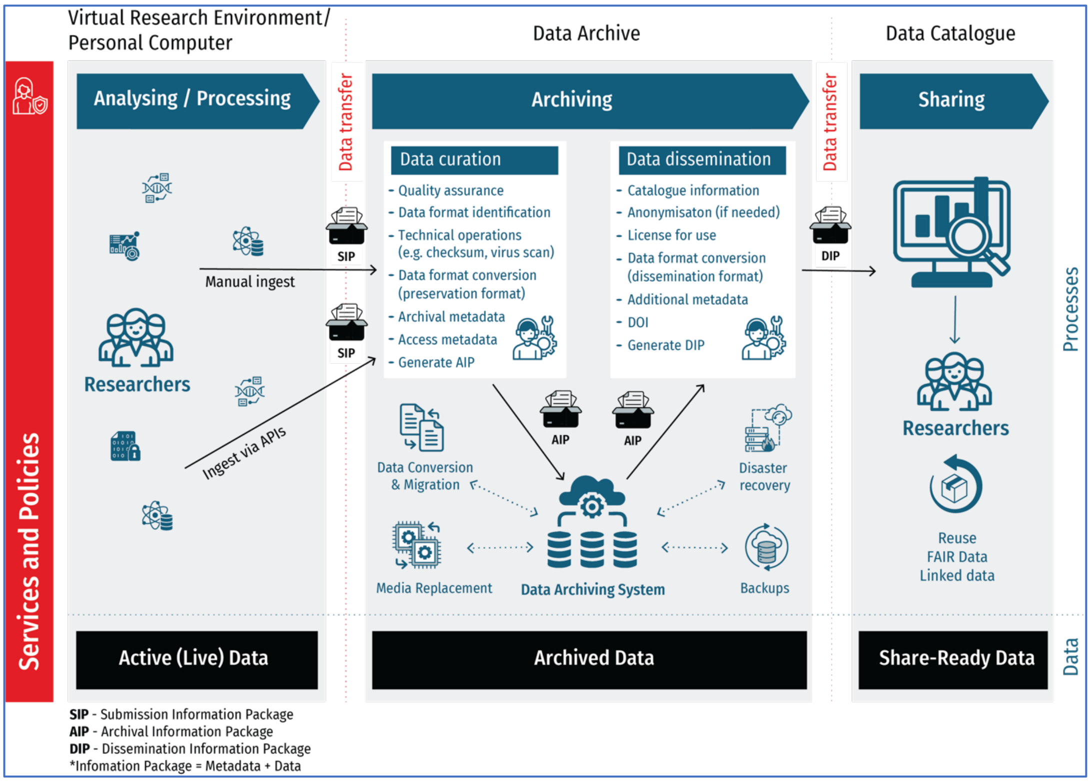

Figure 2 Data Archiving and Repository Systems

Data Archiving and Repository Systems typically organize three key processes (Figure 2). They first of all allow researchers to deposit their data. Researchers upload their data manually or data is ingested via an API. Archives also request researchers to provide metadata and documentation in order to make the deposited data reusable by others. In a second step, repository managers curate the data to ensure quality, accessibility, and preservation. They then also prepare the data for dissemination. Data archives often also make sure that data can be used long term when data and documentation formats change over time. In the last step, data can be searched in publicly available data catalogues and researchers can download the data.

This chapter presents an analysis of key issues related to the archiving and sharing of research data and provides guidelines and recommendations, particularly for the Swiss context. It aims to underline the significance of the FAIR data principles and what it concretely means to put them into practice. Having effective technical solutions and services in place that allow researchers to deposit data and other researchers to access that data is a key part of making research data FAIR.

5.3 Analysis, Key Dimensions of Archiving and Sharing

Financial and governance issues. The current Swiss data archiving landscape lacks a comprehensive funding and governance model that is accepted by all stakeholders. There is no clear consensus on how data archives and related services, including the support to researchers, shall be funded and who carries the cost of archiving and making data available. Adequate sustainable financing is essential for

[8] E.g. see the SNSF reference https://www.snf.ch/en/WtezJ6qxuTRnSYgF/topic/open-research-data-which-data-repositories-can-be-used

maintaining long-term accessibility of research data archives and for ensuring that future researchers can reliably access both data and associated services.

As open as possible, as restricted as necessary. The approach to sharing research data often needs to find the right balance between openness and data protection and copyright restrictions, tailored to maximize the benefits of data availability while safeguarding sensitive information and intellectual property rights. Open sharing promotes transparency, reproducibility, and collaborative advancements in research by allowing unrestricted access to datasets. However, restricted access is necessary in cases involving personal data, proprietary information, or security concerns, where data privacy and ethical considerations take precedence. This balance ensures that while the research community and the public can access and build upon shared data, necessary protections are in place to address privacy, ethical, and competitive concerns.

Standards for data archiving. The process of how research communities and institutions define and agree on standards for data archiving is currently not developed in a systematic way. Internationally known standards for data archiving (i.e. Open Archival Information System Reference Model (OAIS) - ISO 14721) or archiving best practices and repository certifications (i.e. CoreTrustSeal (CTS)[9]) exist, but they are not yet applied uniformly[10]. Some disciplines also have developed standards for trusted data sources and list such sources, like in life sciences the Global Core Biodata Resources or the ELIXIR Core Data Resources.

The **purpose of archiving and sharing data** can vary. Many projects produce datasets that according to requirements of funders or academic policy makers and institutions must be available with the respective documentation and descriptive metadata to make sure others can use the data and build on these. This is by now the standard for publicly funded research in Europe. An increasing number of journals also require that researchers make replication material available that is used in scientific articles. This replication material is, however, usually only a subset of the data that is collected in a research project and may instead include additional material on how the data have been processed and analyzed. Some publishers and also some data archives have systems in place that allow making replication data and other material available.

Incentives for researchers to archive and share data[11]. While incentives for article and book publications openly exist, equivalent incentives for data archiving and sharing are lacking. Researchers prioritize publishing their findings in articles or books without considering the importance of sharing the underlying data, leading to missed opportunities for validation, reuse, and further long-term exploration of research outcomes. Implementing incentives for data archiving and sharing, such as professional recognition and rewards, funding opportunities, and compliance requirements from funding agencies and journals, can encourage researchers to make their data available.

Active data curation and support services[12]. Technical platforms that employ self-deposit and that lack archiving, services, and standards for metadata and documentation lead to a "deposit and forget" scenario. Further, such solutions are not FAIR compliant. Many researchers need support and guidance in data management and data archiving practices. Active data curation also includes quality control of the metadata, data and documentation, as well as anonymization in case of sensitive data. As part of data curation, Persistent Identifiers (PID) mostly in the form of a DOI are attached to facilitate findability

[9] https://www.coretrustseal.org/

[10] For example, currently, only two repositories in Switzerland hold the CoreTrustSeal certification (an international accreditation system for trustworthy data repositories)

[11] See also ORD Action Plan, measure D1.

[12] See also ORD Action Plan, measure B5

of data and data citation. Data stewards[13] and data curators, as new emerging professions, play an important role in supporting researchers with all aspects of archiving and sharing data. Additionally, sound research data management practices are key and need support from higher education institutions.

Rich metadata and comprehensive documentation. Rich metadata and comprehensive documentation according to international and disciplinary standards are essential to ensure that data can be re-used, and easily found and discovered: this requires machine-readable rich metadata and documentation, search engines, and in the future knowledge graphs. Also, researchers who want to re-use existing data will only be able to do so if good and complete documentation exist. Standards, especially for metadata, are also essential to make archiving platforms interoperable and create an entire FAIR ecosystem of national and international archival solutions. Creating rich metadata and comprehensive documentation may nevertheless need to be balanced with the related workload for researchers.

The archiving and sharing of data need to be thought through at the beginning of a data life cycle. Researchers often think about data archiving only at the end of a project. This makes data preparation and documentation burdensome and complicated. Many complications could be avoided if requirements for data sharing (e.g., informed consent, anonymization, and documentation), are built into the entire data collection and preparation process. Data management plans are a crucial element for this, and researchers need training and support in establishing such plans and to implement them appropriately, especially during the early career phase. Data stewards have an important role here, helping researchers develop practical skills and supporting them in planning for sustainable archiving and sharing strategies.

Variations in needs and possible solutions exist with respect to different scientific disciplines. Different disciplines need different metadata, different documentation standards and sometimes also different technical solutions and services to make their data FAIR. Needs and standards must be developed and defined by each research community/discipline, where they don't yet exist already.

International platforms, national repositories and interoperability. Data archiving and sharing platforms do not need to be national. For some disciplines international solutions exist that are widely used and accepted by the research community, and these should be prioritized to maximize reuse prospects. Repositories and data catalogues need to be interoperable and embedded into data spaces, such as EOSC. This involves implementing technical standards that are widely shared internationally, e.g. with regard to file formats or metadata.

National repositories exist for example when they have been established some time ago to make research data available or when international infrastructures are providing decentralized country-based platform and services, which is the case for example in the social sciences within CESSDA ERIC or CLARIN ERIC. National repositories may also be required to make sensitive personal data or data with intellectual property right protection – data that cannot be fully open – accessible. Such data needs to make available following national regulation and it requires active data curation as well as access control. Also funding and governance of repositories and services is often easier to organize at the national level than through cross-country collaborations.

Some higher education institutions have developed in the past their own archiving solutions. Those solutions, however, are costly to develop, to maintain, and are not necessarily in the interest of researchers, since the findability of such data is usually limited.

[13] See also ORD Action Plan, measure B5

Having **national common policies in place on what data should be archived and for how long** is important, especially since data volumes are growing rapidly in many disciplines. What needs to be kept and for how long may vary because of:

- Legal or regulatory reasons for making data available for a certain period of time.
- High scientific relevance of the data, which includes the added value of time series, since some measures are time sensitive and cannot be repeated (examples: climate and environmental data, social data from surveys), and some data may be of historic value. Decision about which data is of high scientific relevance has to be taken by the respective scientific community in dialogue with those who fund data repositories.
- Re-use value and potential of a dataset. Not all data have a high re-use potential. Some data are collected to answer a specific research question and have little value to others except for replication purposes. Some data has high re-use potential, cover a wide range of different analytical purposes that is usually not fully exploited by the data collectors and therefore allow other researchers to build on that.
- Replication of material with journal articles is increasingly required by journals.

The relevance of the data may also impact the level of data curation. A cost/benefit analyses may be conducted, since archiving requires resources from the researchers and the data archives. For large datasets, physical storage costs as well as data curation costs may be substantial, and storage is also energy intensive and therefore raises environmental concerns. Periodic assessment on policies and what data should be kept should be in place.

Clear **technical and practical guidelines are still missing**:

- There should be more clarity on the technical requirements for data archives in terms of required standards, availability, speed of access (for different types of data), backup requirements, etc.
- There is the unresolved issue of how to deal with large volumes of data. Long-term storage costs are a problem especially for small universities. And the exponentially growing production of data further increases the storage and backup needs. Storage not only creates financial costs, but also creates a large carbon footprint.
- Dealing with different file formats is an important issue, especially as technology evolves. Variation in formats is needed, because file formats are domain specific. They are, however, a burden for the interoperability of data, and in addition the usability of different formats needs to be assured over time. As technology progresses, files become obsolete, and new file formats and tools emerge. This evolution necessitates the ongoing migration of data to new formats if existing formats become obsolete or if communities adopt new standards. This is a standard data archiving process, and it ensures that data remain accessible and usable across different platforms and by future technologies.

Artificial Intelligence (AI). The importance of archiving and sharing research data and enriching it with detailed metadata is critical for the advancement of AI technologies. Such measures ensure that essential, high-quality research data for AI model training and validation is organized, accessible, and preserved for future use. Rich metadata enhances AI's ability to interpret data, identify relevant sources, and draw meaningful insights by providing necessary context and structure. AI's significant impact across various academic fields is evident in its ability to analyze extensive datasets, revealing patterns and insights beyond human capability. In-depth data management, enriched metadata, effective archiving, and providing access to high-quality research data are crucial for the progress of AI-driven research, supporting the development of scientific knowledge and innovation.

Last but not least, **barriers** to openness still exist:

- Many researchers still lack knowledge on good data management practices and what it requires to archive and share data in an easy way, and they don't look for support early enough (at earlier stages of the research data lifecycle). Research data management is not a systematic part of training especially at the early career stage.

- Many researchers are not sufficiently aware of the multiple benefits of data sharing. A lack of a culture of data sharing still exists at the individual, institutional (university), and disciplinary level.
- Dealing with data protection for sensitive data is an issue in some disciplines that rely on personal data. Concerns and confusion about what is and is not possible under current data protection legislations hinders some researchers in making their data available. Researchers in some disciplines are also concerned with ethical issues. Many universities are in the process of setting up ethics committees, which in turn requires resources.
- Copyright or intellectual property rights may hinder data sharing as well in some cases.
- Incentives for researchers to share their data are still missing. While key policies exist to make data sharing also part of research assessment[14], they are not yet sufficiently relevant in --day-to-day career or project funding decisions.
- Policies by the different actors vary a lot in detail and are also at times contradictory, which leads to confusion among researchers on what is required. For example, some journals only require replication material to be published, some funders, however, entire data sets. Some repositories require data contracts to deposit and download data especially for personal data based on the current data protection regulation, while others have no such requirements in place.

5.4 Conclusion

Data archiving and data sharing are key elements for putting FAIR principles into practice. Having reliable and high-quality data is not only a requirement from funders, who want to make sure that publicly funded data collections are a public scientific good that can be used by the research community at large. It is also essential to ensure replicability, reuse, and innovation, and it is more and more also the basis for high-quality AI technologies. Many different elements need to be in place to allow for broad archiving and sharing of research data. Clear and common policies addressing which data requires archiving and how it should be shared would be beneficial to the scientific community in Switzerland. Such policies need to provide solutions that allow data access while respecting data privacy and protection of intellectual property rights. Repositories that ensure high-quality data and support services must be operational with sustainable funding and should be able to cover discipline-specific needs. Repositories need to follow best community of practice standards and FAIRification processes, to enable interoperability and allow the entire data repository system to be FAIR and interconnected. Finally, there need to be better incentives and rewards for researchers to comply with data archiving and sharing requirements, as well as support in research data management and sharing to help them.

[14] https://coara.eu/

6 Access

6.1 Introduction

Following the mandate as given by Strategy Council, user access is understood as “the ability for all researchers across types of institutions, as well as their research partners globally and their partners in other domains (e.g. industry), to easily and equally access and utilize open research data infrastructure, taking into account technical barriers (e.g. authentication), legal barriers (e.g. conflicting policies in institutions), and financial barriers (e.g. open cost model)”.

This section focuses on the challenges and strategies associated with user access to Open Research Data (ORD) but also considers other data that researchers need increasingly access to. It acknowledges its interconnectedness with interoperability, reuse, and archiving and sharing that are comprehensively detailed in separate chapters. This chapter is organized as follows: After a short remark on the scope of access, we present in the second section an analysis of the different aspects of access to research data in terms of “strategic issues and policies”, “barriers to access”, “incentives for access”, “technology” as well as “cost models”.

6.2 Scope

In setting the scope, we refer to the fact that FAIR data is meant to be “as open as possible, as restricted as necessary”[15] and does therefore not imply unrestricted access.
In alignment with the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, it's crucial to emphasize the importance of data being findable as a prerequisite for access. This means that before data can be accessed, it must first be catalogued and indexed to allow potential users to discover it (as outlined in section 5).

6.3 Analysis

The analysis of Access is given along five main criteria, on which we comment in the following subsections (see Figure 3).

6.3.1 Strategic Issues and Policies

Today, open research data is being promoted, even requested by funding agencies and institutions on a strategic level and required by corresponding policies and funding criteria. On the other hand, on the legal level there are also constraints on accessing sensitive / personal data that hinder complete openness. These barriers sometimes present conflicting requirements with the push for openness, necessitating a careful balance between data accessibility and protection.

The picture is further complicated by the perception among many researchers that the data is their own good, and little incentive to invest the additional effort to make it available beyond the publication. Hence, the key question is whether and how the rights and responsibilities of the data are defined in the academic sector and to what extent the owner has a role in deciding whether the data should be findable and accessible. These points are of strategic importance for education and research, necessitating the corresponding policies that clearly address ownership and accessibility issues.

The effectiveness of policies promoting open research data is contingent on their completeness and consistency. This raises questions whether existing policies are complete and consistent, who should be responsible for making them consistent if they are not, and whether and how they should be enforced.

[15] See also section 5.3

In addition, while research has an international focus, funding and thus funding requirements are mostly national. Therefore, national policies should also be consistent with international ones. Alignment with EOSC is of particular importance for open research data. Clearly, an effort to harmonize policies and making them enforceable is needed.

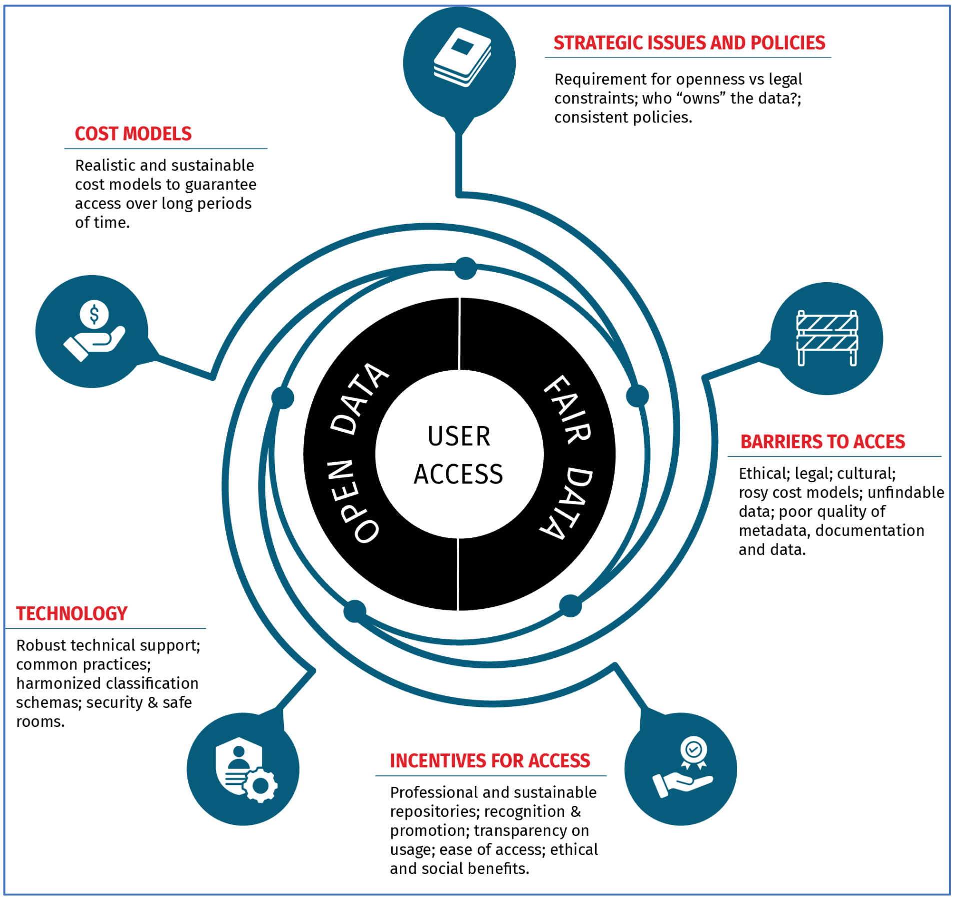

Figure 3 Open / Fair data: Access

The scope of open research data extends beyond academia. There is a growing need to make academic data accessible to the broader public, aligning with the principles of citizen science. Conversely, more and more data is being used in research that has not been produced by researchers but from the public and private sector. Governments increasingly offer open access to their data, however, such data is even less FAIR than research data. Hence, it is essential to assess whether current policies adequately cover these broader implications of data access.

One way to remedy the situation is to require that the metadata (or at least that part of the metadata that is non-sensitive) must be freely accessible, while the data itself may not. The metadata must also include transparent criteria under which conditions the data may be accessed. This approach needs a robust classification system to distinguish between different cases. Additionally, monitoring and enforcing adherence to these policies is crucial to ensure effective implementation.

6.3.2 Barriers to Access

The following barriers to access were identified:

- **Legal and ethical barriers**[16]: sensitive data or data protected by copyright and intellectual property rights pose significant legal and ethical barriers. These concerns are often coupled with

[16] See also ORD Action Plan measure D2

fears of legal or reputational risks and loss of trust. This barrier cannot be eliminated completely and must be managed with proper processes.
Countermeasures:
  - Harmonization of policies and alignment with a) the law and b) international regulations. This includes promoting open data policies also at the political and societal level, in particular a legal framework for secondary use of data. Additionally, implementing a comprehensive system to manage and coordinate data access policies among different stakeholders is crucial. Such a system would enable clearer and more efficient adherence to diverse data access needs while maintaining compliance with overarching legal and ethical standards.
  - Legal experts should be included in the teams early from the beginning.
- **Cultural barriers** as outlined in section 5.3
- **Access control** (where required): implementing access control is often a cumbersome process that requires substantial expertise and resources. Existing AAI[17] framework such as edu-ID, while powerful may not adequately support all those needs.
  Countermeasures:
  Missing functionalities such as support for including non-academic users, or mechanisms for managing authorization of user groups should be identified, prioritized, implemented and disseminated.
- **Findability of Data:** data scattered in many local repositories may be accessible in principle, but practically not known to exist (i.e. not findable).
  Countermeasures:
  a) Comprehensive, complete and linked metadata catalogues, b) support for federated searches and c) simply a consolidation of repositories to fewer well-maintained ones. In addition, there is a need for effective communication strategies to increase awareness of available data resources.
- **Inadequate Cost Models:** existing cost models often fall short in providing the necessary financial support for sustainable services needed to make data accessible.
  Countermeasures: see section below.
- **Rich Metadata, Standardized Documentation and Contextual Information: A** lack of documentation, context and meaning of the data impedes users' ability to locate and identify relevant data effectively, making it challenging to even begin the process of data utilization.
  Countermeasures: Establishing standards for documentation and requiring comprehensive metadata and context information with data sets can help mitigate this issue. It is essential to consider the need for providing documentation in universally understood formats and include generic discipline agnostic descriptions.
- **Data Quality:** concerns about the accuracy, completeness, and reliability of data can hinder its access and use.
  Countermeasures: to address these concerns it is essential to establish mechanisms to ensure and communicate data reliability and quality.

6.3.3 Incentives for Access

The following incentives[18] for access were identified to promote the accessibility of data in repositories:
- Providing professional and sustainably operated repositories with support, proper curation services and adequate tools are a strong incentive for researchers to make their data accessible. These facilities reassure researchers that their data will be actively used and maintained, rather than being neglected or lost in a data graveyard.

[17] AAI = Authentication and Authorization Infrastructure
[18] See also ORD Action Plan measure D1

- Funding agencies and higher education institutions can create incentives for researchers to make their data accessible through recognition and promotion and/or taking previous data sets shared into account in funding and career decisions.
- Monitoring and publicly sharing information about data access and usage can encourage researchers by demonstrating a real-world impact of their shared data. This transparency not only promotes better access but also creates an incentive to make more data accessible.
- Developing a system to quantify and display data quality metrics in repositories can be a powerful incentive for researchers.
- Simplifying the process of making data accessible, through user-friendly interfaces and streamlined procedures, can reduce the perceived burden and thus incentivize researchers to share their data.
- Promoting the ethical and social benefits of accessing shared data for research can serve as a moral incentive, especially in fields with a direct societal impact.

Although incentives are recognized, however, their implementation must be supported by institutions.

6.3.4 Technology for Access

Ease of access to open research data is of paramount importance for user acceptance. This requires robust technological support, particularly for scaling effects coupled with clear and uniform access rules to yield the most benefit. Scaling effects can be exploited with common infrastructures, uniform access can be provided by standardized tools. For both areas there are good individual examples today, but much remains to be done.

Access to sensitive data requires different measures than access to non-sensitive data. There are currently no common practices, e.g. how non-academic individuals are identified, when requesting access, or need to give permission to grant access (e.g. if the data owner must be involved in the process). Establishing rigorous and harmonized classification schemas is key for appropriately categorizing data based on sensitivity and determining access levels. Enhanced authentication, authorization and access governance mechanisms are needed to ensure secure and appropriate access to sensitive data. Moreover, integrating technologies that evaluate the privacy sensitivity of data can further refine and enhance the access decision-making process. These technologies employ advanced algorithms to assess the level of sensitivity and privacy risk associated with specific datasets. By doing so, they provide a dynamic layer of analysis that aids in determining appropriate access levels. In addition, the automation of data access policies plays a crucial role in facilitating and securing data access processes. Automated systems can be integrated to automatically grant or restrict access to data based on the user's credentials and the data's classification, ensuring compliance with predefined access policies.

Accessing data in this context is typically understood as downloading the data. However, alternative technical approaches exist, such as infrastructures that permit to analyze the data without downloading it. For instance, 'safe rooms' or secure virtual environments with remote access allow researchers to analyze data without downloading it, thus mitigating data transfer and storage risks. Similarly, federated learning approaches enable data analysis across multiple repositories without moving the data itself, preserving privacy and reducing data transfer loads.

Lastly, the risk of adopting standards from private or public cloud providers should be mentioned. While these standards may be initially openly available and drive communities, change of policies or contractual conditions may lead to them becoming unsupported or only available at unexpected fees.

6.3.5 Cost Models

Making access available over long period of times requires realistic cost models for the corresponding services. They latter should not only cover preparing (meta-)data such that data can be found and accessed, but also operate metadata catalogues and processing data access requests with a professional service level. The cost model situation presents therefore a particular challenge.

The research community as well as many institutions view user fees for access (not necessarily services) as the wrong path, which leaves project-based central or institutional funding.

Project-based level has turned out not to work particularly well as the long-term nature of providing access is not compatible with project funding. Some scientific communities (e.g. high-energy physics or astronomy) manage their (huge) data through domain-specific, international infrastructures. However, this approach does not work for most disciplines. In addition, not all international repositories add a service layer to their activities which is relevant for data access. This leaves central or institutional funding as the remaining options.
Today, while SNSF supports repositories for social sciences and humanities, some institutions have built or are building their own repositories. This leads to differences between institutions, scattered resources, non-interoperable solutions and non-unified access.

6.4 Conclusion

User access is meant as giving access to all national and international users while respecting the fact that not all data can be open ("be as open as possible and as restricted as necessary"). Requiring that metadata must be freely available while the data itself may not, was identified as an important lever to bring restricted data into balance with FAIR principles.
Access was analyzed along five criteria: a) strategic issues and policies, b) barriers to access with corresponding countermeasures, c) incentives for access, d) technologies for access and e) cost models. Key findings included that various policies need to be harmonized and aligned nationally as well as internationally. Various barriers, among them lack of findability, cumbersome access controls, legal and cultural constraints, inadequate cost models, were identified. On the other hand, incentives for making data accessible exist and should be exploited, such as providing professional and sustainably operated repositories with proper support, monitoring and publicly sharing information about data access as well as acknowledging researchers who make their data accessible.

7 Technical Interoperability

7.1 Introduction and Scope

This chapter explores (technical) interoperability[19] between Research Data Infrastructures (RDIs), which is crucial for enabling researchers to utilize multiple, complementary RDIs across their data management cycle as seamless as possible to transfer data, to automate processes, and to have a good overall user experience. Achieving (technical) interoperability is essential for enabling researchers to efficiently retrieve, combine, reuse and share data across various platforms and systems.

For the purpose of this report, (technical) interoperability of RDIs is understood as the ability to exchange, combine, and interpret research data stored in different systems.

EOSC distinguishes 4 layers of interoperability: semantic—technical—organizational—legal. While the original scope of the questions from the ORD Strategy Council mentions technical interoperability only, the issue of interoperability is very complex. Being able to technically exchange data will not guarantee that data can be used in a shared way. Therefore, we here discuss technical interoperability together with semantic interoperability, and furthermore point at the services needed for all aspects of interoperability, emphasizing the links and dependencies to the technical aspects.

We examine interoperability across three distinct levels (see **Figure 4**)

1. **Metadata level** – refers to standardizing the information that explains the meaning of the data, how the data were generated, and how the data can be further used.
2. **Data level** – refers to the ability to combine, integrate and use data from different sources or areas of study.
3. **Service level** – refers to having the right support and skills available to use different research data systems effectively.

This chapter illustrates the potential and limitations of (technical) interoperability. We highlight the bottlenecks and gaps that currently impede interoperability and propose targeted recommendations to ensure (technical) interoperability of different infrastructures. Current RDIs have evolved independently, leading to a diverse technological landscape. This diversity—while beneficial in some aspects—poses significant challenges for interoperability. Therefore, our recommendations consider the practical aspects of integrating these disparate systems and the realistic extent to which interoperability can be achieved.

Technical interoperability is a key factor for a potential alignment to European Commission (EC)-funded initiatives such as the Data Spaces initiative or taking part in an operational European Open Science Cloud (EOSC) platform that can harbor a federation of services belonging to multidisciplinary RDIs, researchers, innovators and companies.

[19] See also ORD Action Plan measure B4

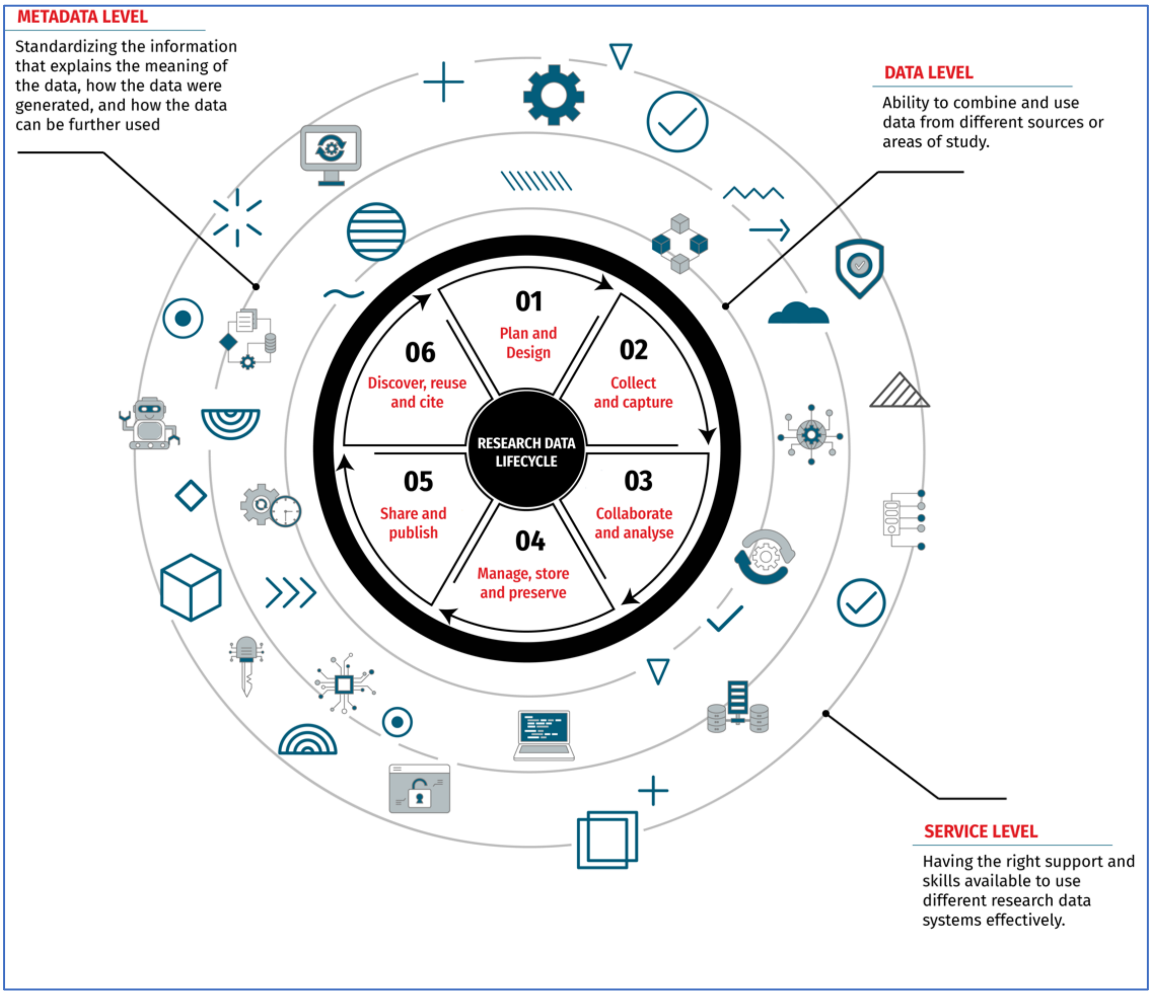

Figure 4: Aspects of interoperability

The EOSC Interoperability Framework (IF) aims to promote interoperability across the entire research data lifecycle within EOSC. It serves as a guiding framework that underpins the interoperability goals of EOSC and its ecosystem, facilitating seamless integration, collaboration, and data exchange across diverse scientific domains and infrastructures. The IF is integrated into the core principles and strategies of EOSC, providing guidelines, standards, and best practices to ensure that data, services, and tools can effectively inter-operate within and across different components of the EOSC infrastructure. These interoperability guidelines at a trans-domain level provided by EOSC IF should enable national research communities and RDIs to align with European standards, facilitating interoperability not only within their respective countries but also across the broader European research landscape. This alignment aims to ensure coherence, transparency, and accessibility, ultimately strengthening collaboration, fostering innovation, and advancing scientific knowledge both at the national level and within the larger European research ecosystem. In addition to metadata interoperability, the framework also emphasizes technical interoperability, semantic interoperability, and organizational interoperability. These aspects collectively contribute to creating a cohesive and interoperable environment where researchers should easily discover, access, share, and reuse data and services across different domains and disciplines.

EOSC emphasizes the importance of aligning data management practices with FAIR (Findable, Accessible, Interoperable, Reusable) principles to enhance data discoverability and reuse. By meeting these requirements, EOSC endeavors to create an integrated and interoperable environment that supports seamless data exchange and collaboration across the European research landscape.

A careful evaluation is needed to assess how Switzerland's current interoperability landscape can enable participation in EOSC (see section 9) taking into account the unique institutional, infrastructural, and regulatory context of Switzerland. This involves understanding both the opportunities and constraints presented by EOSC guidelines and requirements, as well as identifying potential areas for alignment and collaboration to enhance Switzerland's integration into the broader European research ecosystem[20].

7.2 Analysis

7.2.1 Interoperability of Metadata:

Rich metadata describing what information is contained in the data and how data can be accessed and used is of utmost importance for ORD. Hence, metadata must be globally accessible, comprehensive and descriptive.

A core set of metadata requirements on content and their findability needs to be formulated at the national level, independent of the specific research domain, but based on a joint understanding by the research community of ORD best practices. These requirements need to follow international standards such as those defined by e.g. EOSC and transversely applicable to all domains, to enable open research globally. They have to be embraced and fostered by the ERI actors (higher education institutions, funders) and the research community at large and subsequently be incentivized and implemented in RDIs.

Most data are used within their own research domain and hence additional sets of domain or domain cluster specific metadata need to be defined. The definition of cluster- or domain-specific metadata is a long process that requires resources not readily available in ERI institutions. There are few incentives for researchers and single research institutions to take responsibility and invest in building community consensus on interoperability requirements. Agreeing on common standards, based on community of practice and best practices, is a long and complex process that is currently often not conducted in a coordinated fashion with a clear lead and structure. Community of practice involves an outlook view and discussion on those requirements used by the domain/field-specific international communities, ERICs, EDICs and ESFRI Roadmap Research Infrastructures that are key actors in EC Data Spaces. In practice, some metadata are established *de facto* by leading databases for particular kinds of data: for instance, the *Cellosaurus* database's persistent and unique identifiers for referencing cell lines, making it possible to retrieve the rich meta-data associated with these cell lines.

To establish consensus on national metadata standards, the national ORD strategy needs to clarify the roles and responsibilities of the different actors concerned with interoperability for ORD in Switzerland, and to support respective organizational structures.

EOSC emphasizes achieving metadata interoperability through the adoption of common standards, harmonization of metadata practices, and promotion of semantic interoperability. It encourages the use of widely accepted metadata schemas and vocabularies/ontologies while advocating for the semantic enrichment of metadata to facilitate cross-domain discovery. EOSC also promotes the development of metadata mapping mechanisms and emphasizes the importance of ensuring metadata quality and consistency. By adhering to these principles, EOSC aims to foster collaboration and enable seamless data discovery, access, and reuse across diverse scientific communities and infrastructures. Key requirements include the adoption of common metadata standards, such as Dublin Core, DataCite and Schema.org, to facilitate consistent data description and discovery.
Organizations and communities often select and adopt specific standards based on their requirements and objectives. Therefore, there is a need for a careful assessment that encompasses general, national,

[20] See ORD Action Plan measure B3.1.

domain, community or organization specific requirements and their alignment to the metadata interoperability proposed by the EOSC IF.

7.2.2 Interoperability of Data

In consideration of the existing situation with a considerable number of infrastructures having built their own tailored systems over many years, full technical interoperability on the data level is not realistic and also not worth striving for. This is especially true for semantic interoperability, which is highly context dependent and requires specialist knowledge i.e. expert curation. This should be in the responsibility of designated RDIs within the different research domains. These RDIs should act as competence centers and facilitators for the exchange of data within but also across research domains.

Due to the efforts and demands of funders to develop complementary infrastructures and services within domain clusters, however, infrastructures from different domains may be better interoperable from a technical perspective than multiple infrastructures within the same domain. Technical interoperability within the same domain may primarily be achieved by using the same standards and norm data, and to agree on a few common open file formats in which the data can be exported. In addition, interoperability of certain data types (like e.g., images) could be addressed also across domains. Of note, technical interoperability not only concerns the data themselves but also the tools for those data, and models and ontologies building on those data.

In addition, researchers often use different tools at various stages of their work, so it is vital that data can automatically move between these tools throughout the data lifecycle. This means data must be easily transferable between platforms, from collection to analysis and storage, while also ensuring that metadata is transferred alongside the data to avoid information loss. Smooth data transfer is crucial for efficient research. It simplifies the process, letting researchers focus on their work without worrying about data compatibility issues.

Regarding interoperability of data, EOSC promotes semantic interoperability through the use of controlled vocabularies and ontologies, thereby enabling meaningful connections between related datasets. Technical interoperability is facilitated through the implementation of interoperable APIs (Application Programming Interfaces) and data exchange mechanisms.

7.2.3 Interoperability of Services

The service level, while not strictly technical, constitutes an indispensable aspect of interoperability. If researchers and RDI staff lack the requisite skills and if appropriate services are unavailable, the technical interoperability of infrastructures becomes of little use.

Interoperability of services is crucial in supporting researchers through their research data lifecycle, as different stages demand specific, interconnected services that effectively communicate and exchange information. Services ensure that researchers are mindful of the requirements for later stages, particularly when it comes to depositing data into repositories who come at the end of the data lifecycle. Data stewards play a crucial role in this ecosystem. Data stewards act as advocates for FAIR data principles, providing guidance and support to researchers in managing and curating their data throughout its lifecycle. By integrating the guidance of data stewards and ensuring interoperability between services, the research process becomes more efficient and aligned with the FAIR principles, ultimately enhancing the value and impact of the research within the scientific community. The exchange between data stewards, researchers, and infrastructure services streamlines the research process from data collection to final data preservation and sharing, simplifying tasks for researchers and ensuring compliance with ORD principles and best practices.

Currently, interoperability cannot be established in a fully self-explanatory and automated way, but needs to be supported by expert services to bridge the researchers' needs with the available data. Since no single institution and researcher derive sufficient benefit from interoperability to justify the necessary investments, these services must be provided either as a community effort, or through an "honest broker"—a role often taken by leading databases for a kind of data[21]. Expert services also ensure the flexibility needed in research. For example, it would be not feasible to establish full technical interoperability between e.g., life sciences data and humanities data to generate new insights and possibilities to such an amount that it would justify the necessary efforts. Instead, the services ensuring interoperability within those domains must be able to interact in an efficient manner. This requires that there are mechanisms established for exchange of knowledge and coordination of processes between service providers.

7.2.4 Further Needs

It is conceivable that a limited number of professional RDIs is more efficient for establishing and maintaining interoperability than many local RDIs. However, consolidation of infrastructures and services needs to be balanced with the specific requirements of the various research domains. Hence, the ERI actors should designate national RDIs in all relevant domains and provide them with clear mandates and sufficient resources for defining, establishing and maintaining interoperability of metadata, data and services. Interoperability and respective national RDIs cannot be built through bottom-up projects and funding instruments targeted at higher education institutions and researchers but need to be coordinated and supported nation-wide with a careful look at the national environment.

At the same time, the buy-in from the research community to support interoperability is critical. Commitments and investments from researchers to support interoperability must be honored and rewarded in career advancement and research funding. It must further be ensured that the costs of interoperability services are not diminishing research grants but are included by funders in attributable costs. Stable core financing for interoperability services of RDIs should keep user fees low for academic research.

A substantial investment in skills development and capacity building is urgently required. For example, there are very few semantics experts in Switzerland. We need to train such people and provide them attractive career perspectives to be able to scale our ORD efforts. Similar to the newly established Data Stewards, we should expand the professional profiles in the realm of data science to semantic experts, research software engineers, etc.

Ultimately, a careful mapping and evaluation of the Swiss interoperability landscape and all of its components would be key to assess the current alignment to European and Global initiatives such as EOSC or the driver community of the Research Data Alliance. This would enable to take informative decisions on the amount and place of allocation of resources, avoid duplication efforts so that Switzerland can work towards a national strategy that encompasses all relevant stakeholders in a coordinated fashion. The results would be ideally based on consensus adoption by the relevant stakeholders with a track-record and extensive know-how.

Even if there are requirements and standards recommended by the EOSC IF, EOSC is an evolving structure in real-time at present. Therefore, it is important for Switzerland to participate in the discussion, share experiences, challenges to generate shared solutions, formats, procedures and standards.

---

[21] Several quality seals for reference databases have been established, e.g. Global Core Biodata Resources, ELIXIR Core Data Resources, or CoreTrustSeal.

In the context of research data interoperability, it is vital to consider also legal and regulatory factors, especially those related to sensitive data. Balancing data interoperability with data protection can be a complex task. It involves sharing data across different systems while ensuring security measures and consent, which can be challenging, especially when dealing with varying data formats, security protocols, and levels of trustworthiness among different platforms.

7.3 Conclusion

Interoperability at the metadata level is absolutely necessary and has to be in line with international standards like those elaborated by EOSC. At the data level, interoperability should mainly be striven for at the semantic level, with improvements on technical interoperability (e.g. between completely different disciplines via APIs that understand each other since they are based on the same standards) which will provide additional benefits. The service level is crucial for supporting researchers such that they can focus on their individual research interests without compromising the benefits for the research community at-large.

Clear mandates and sustainable financing of national Research Data Infrastructures are needed. Currently, national infrastructure providers are not sufficiently involved in the roadmap process for national infrastructures and are largely excluded from grants for ORD infrastructures by swissuniversities — but they are expected to play a national role.

8 Reuse

8.1 Introduction

Reuse is understood to mean the ability for hardware, back-end services, software components, as well as full-stack services to be shared between more than one research data infrastructure, thereby lowering cost and improving interoperability. It implies the ability to standardize these shared resources, to engage with research data infrastructure (RDI) owners, and the availability of appropriate cost-sharing models.

After setting the scope of this document, we analyze prospects for reusing components and services in a four layered model of research data infrastructures from which we derive our conclusions.

8.2 Scope

We take this mandate in a narrow sense and focus on reuse of technical solutions (e.g. hardware, back-end microservices, software modules, full-stack solutions) as well as comprehensive data management services and know-how such as anonymization of sensitive data, archiving, data science, or legal support. However, we point out that reuse of data, although excluded here, is also a topic of paramount importance in ORD[22]. We suggest that the topic of data reuse could be handled separately, in collaboration with the Sounding Board of Researchers, while acknowledging that some research data infrastructures and the Higher Education Institutions already offer services related to data sharing and reuse.

The scope is to illustrate the requirements and impact of reuse. At the Sounding Board of Service Providers, we believe in general that the reuse of technical solutions and comprehensive data management services is crucial for several reasons. It enables a more efficient allocation of resources, reducing costs and avoiding the need for developing similar tools and services from scratch. By building upon existing, proven research data infrastructures and expertise, it also accelerates innovation, allowing other infrastructures to focus on advancing their specific important areas of expertise rather than on the development of basic infrastructure. Furthermore, the advantages of reuse include standardization of components, potentially lower costs, improved interoperability of services and their components and sustainability.

The scope extends beyond a narrow focus on reusing research data infrastructure components, advocating for principles such as modularity, a collaborative culture, common policies, and adherence to standards and best practices to be embedded at the core of such infrastructures. This means that not only should the components be designed with reuse in mind, but the environment should also foster collaboration and dialogue as the norm, rather than the exception.

8.3 Analysis

We first consider reuse in distinguishing a four layered model of research data infrastructures (see Figure 5) and then continue with general observations on reuse. The model has the following layers:

1. IT Infrastructure Layer - hardware and lower level (virtualization) services
2. Application Layer
3. Data Layer - presentation and data analysis
4. Business/Service Layer

[22] The ORD Action Plan mentions "reuse" mainly under the prospect of "reuse of data".

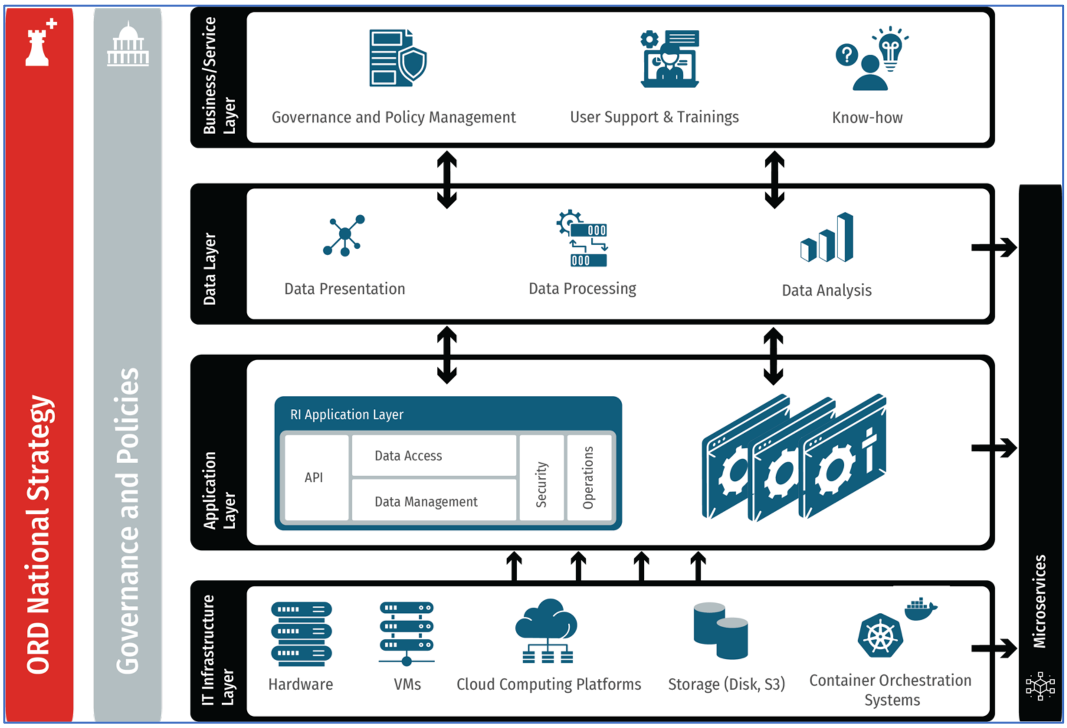

Figure 5 Research Data Infrastructure Layers

Concerning the **IT Infrastructure Layer** there are several suppliers offering services such as virtual machines (VMs), container orchestration systems (e.g. Kubernetes) or other supporting services (GitLab, database servers). These services are available for acquisition through diverse delivery models, including, but not limited to, cloud-based solutions. Examples are institutional service offerings, national cloud providers such as Switch or the Swiss National Supercomputing Centre, as well as hyperscalers such as AWS, Google or Microsoft. The current funding model does not support relying on third party infrastructure providers as this would require a shift from a capital expense to an operating expense model (CapEx vs. OpEx). In addition, we would like to point out that the potential for reuse in hardware and lower-level services is significant, but frequently unexploited. It would necessitate a shift towards perceiving these services as cornerstones of a flexible, cost-efficient, and reusable research data infrastructure ecosystem, adaptable to diverse research settings. More than technical fit, this shift would demand a collaborative culture where stakeholders, despite varied frameworks, actively work together to harmonize services. Data security is guaranteed by **storage in multiple geographic locations,** as well as by additional offline data storage.

Concerning the **Application Layer** each research data infrastructure has built and continues to build its own solution depending on their specific needs, leading to minimal common practice in reuse. Reuse always involves considerable efforts to adapt an existing software to new requirements and uncertainties regarding long-term sustainability, especially given the current funding schemes. While adherence to standards and well-established software tools and practices should be actively promoted to help to build the base for more potential reuse in the future, the challenges at this layer are more than merely technical. These challenges often stem from the highly specialized nature of application layer systems. To address this, fostering a culture of collaboration and dialogue is essential. Such a culture enables stakeholders to share insights and strategies, thereby identifying common needs and solutions that transcend individual projects. This approach, coupled with a commitment to modular design following standards and best practices, paves the way for building application components that can be adapted or extended for reuse in various contexts, enhancing the overall efficiency and innovation in the research data infrastructure. Additionally, the assessment of solutions dealing with transversal activities (e.g., data science, data archiving) should be evaluated for their potential (re)use as a complete solution, ensuring a unified approach to meeting the research community's needs.

The **Data Layer** (presentation, processing and data analysis) offers considerable potential for reuse, particularly when employing web components with well-established standard APIs or tools like Helm and Kubernetes for deploying and managing applications. However, realizing this potential requires an ecosystem where providers of different infrastructures are aware of and actively engaged in the exchange of components and ideas. Crucially, this ecosystem should support close collaboration between infrastructure providers and researchers, which is often lacking today. This support is critical for enabling researchers to package their work as reusable components or format their research outcomes for integration with services offered by other providers. Such an ecosystem approach enhances the reuse of research outputs and infrastructure components within the presentation and data analysis layer across various infrastructures.

In the **Business/Service Layer**, each solution comes with its domain-specific services and know-how offering a wealth of knowledge that can be shared and leveraged. It is essential to reuse these existing services and know-how and to build upon them, creating more comprehensive solutions. Unlocking the full potential for reuse at this layer requires more than just making these services accessible. It involves creating a collaborative environment where knowledge, best practices, and insights are shared freely across institutional boundaries. By doing so, not only can services be repurposed and adapted to new contexts more efficiently, but the collective expertise can also lead to the evolution of these services, making them more robust, adaptable, and aligned with the evolving needs of the research communities. In the Business Service Layer, pooling of knowledge across institutions is not just a strategy; it is necessary to enhance the quality and efficiency of research infrastructures. By collaboratively addressing best practices, particularly in areas such as sensitive data management, legal compliance, and even the maintenance of common helpdesk services, institutions can leverage collective expertise to overcome common challenges. This collaborative approach can lower costs and improve the quality of service delivery.

Beyond the four layered model of research data infrastructures, the key to unlocking the full potential of any reuse strategy lies in the sustainability of all externally obtained components. The interdependence of reuse and sustainability is profound, influencing the operational efficacy and the funding model underpinning research data infrastructures. It is critical to ensure that reused components remain viable, reliable, and efficient over time, aligning with a funding model emphasising long-term viability and adaptability. Therefore, as we advocate for reuse across various layers of the research data infrastructure, we must also champion sustainable practices and supportive funding models.

Alongside the above-mentioned horizontal layers (IT Infrastructure Layer to the Data Layer) **microservices** are critical components for building reusable research infrastructures. Microservices are small, self-contained services that can encapsulate all layers, but with very narrow functionality. They are designed to do one thing very well, embodying the principles of modularity and single responsibility. Different microservices operated by the same or different institutions can communicate with each other through well-defined interfaces and protocols, enabling them to be connected and orchestrated to form entire infrastructures. Examples of such microservices are Transkribus (handwriting recognition) or ChatGPT. Microservices promote a culture of focused expertise and best practices while aligning with the broader goals of efficiency, reuse, and collaboration in the research data infrastructure.

Beside institutional services there are also national ones and services on the European level, e.g. repositories. There are obstacles for reuse across institutional boundaries leading to duplications due to a lack of awareness and cooperation, and a funding model that supports such a development. Institutional repositories may be integrated too closely into the IT infrastructure of the institution.

8.4 Conclusion

The current funding model, the lack of knowledge and collaboration as well as the fear of reusing a non-sustainable component or service and the “not-invented-here syndrome” do not favor reuse. While a diverse array of services exists, consolidating them into a common shared system is not feasible. The extension of existing services to other communities is time-consuming and entails the process of community building which is slow. Moreover, services designed at an institutional level frequently fail to meet the requirements of inter-institutional collaboration.

However, embracing an ecosystem approach could transform the current landscape. A collaborative culture, adherence to best practices and standards, and harmonized processes are vital to enhancing the reuse of services and components. Transitioning to national-level services can break down institutional silos, leading to substantial cost savings, improved service reuse, and greater sustainability. Realizing this potential, however, demands funding models that align with these goals.

9 European Open Science Cloud

Note: EOSC and its services are still under development. This chapter reflects the current (August 2024) situation as the SB SP understands it. Given the rapidly evolving European and national landscape, certain statements may not hold in the future and will likely need to be revised.

9.1 Introduction: The EOSC Landscape

The **European Open Science Cloud[23] (EOSC)**, European Research Infrastructure Consortia (ERICs), international Research Infrastructures (RIs), and Science Clusters offer significant opportunities for Switzerland (CH) in terms of international relations, collaborations, alignment, and innovation. Science clusters are beneficial for ensuring better coordination within scientific disciplines. ERICs and other RIs play a crucial role in uniting communities of practice in the science clusters and building capacity in the national nodes/countries. Many RIs are well-established and have long been working on FAIR (Findable, Accessible, Interoperable, and Reusable) and open research data initiatives at both national and international levels. Some of these RIs are outlined in the ESFRI Roadmap[24] and can include ERICs, other RIs, and potentially digital or e-infrastructures such as EDICs, all of which may serve domain-specific roles within the broader vision of Common European Data Spaces (CEDS).

The **Common European Data Spaces (CEDS) initiative** aims to create a single market for data that ensures more accessible, interoperable, and secure data sharing across various sectors and borders within the European Union. This European Commission (EC) initiative aims to foster innovation, enhance economic growth, and address societal challenges by enabling the free flow of data among businesses, governments, and citizens while maintaining high standards of data privacy and security. Whilst EOSC is part of the CEDS initiative under the Research and Innovation domain, it is not the sole contributor. Additionally, EOSC may also likely participate in various data spaces such as the European Health Data Space (EHDS). The goal of the EOSC initiative is highly ambitious, aiming to create a seamless and interoperable environment for sharing and reusing research data and tools, feeding into the broader goals of CEDS. Other projects and initiatives also play a role in this framework, some of which are driven by Horizon Europe and Digital Europe funded projects. Switzerland's non-association impacts Swiss participation in these programs in various ways as described in the subsequent sections. For instance, two out of the four projects listed in the European Health Data Space fall under Digital Europe, one project under Horizon Europe, and the fourth project under the EU4Health program.
Regarding the EOSC ecosystem, there is the EOSC initiative, an effort driven by the EC to advance the EU's digital agenda, and the EOSC Association (EOSC-A), a non-profit international organization that gathers stakeholders to drive the EOSC initiative forward.

The **priorities for building EOSC** are set jointly by the EC and the EOSC Association through the co-programmed European Partnership, known as the EOSC Partnership (INFRAEOSC calls), outlined in the Strategic Research and Innovation Agenda (SRIA). Horizon Europe continues to support the development of EOSC, building on the foundation established during the Horizon 2020 program. Horizon Europe funds the development of the next generation of EOSC functionalities, networks, and services, while also providing a fully operational EOSC infrastructure node (referred to as the EOSC EU Node) as the result of a public procurement action launched in 2022. The EOSC EU Node is currently under development by contractors and governed by the EC. The ownership of the EOSC EU node is with DG CNCT. The EOSC EU node is meant to be just the first node of the federation, with the aim that other EOSC nodes will join the federation so that the EOSC node federation can start to develop the first set of participation rules. The drafting of an EOSC Federation Handbook currently takes place together with

[23] Also sometimes called European Open Science Commons
[24] https://landscape2024.esfri.eu/

community involvement from up to 10 potential candidate/prototype Nodes. National, regional, and thematic service providers are encouraged to connect to the federation via established interoperability frameworks and policies. These nodes may serve as access points for local or regional resources, ensuring that these are connected and interoperable with the central EOSC platform. As of now, the managed services of the EOSC EU Node are owned by the EC until at least 2027. The EOSC federation of nodes is intended to be governed by the EOSC Tripartite Governance structure, which includes the European Commission, the EOSC-A, and the EOSC Steering Board with representatives from Member States and Associated Countries to Horizon Europe. This can be taken as a sign that the EC is determined to guide these developments stronger than with past programs.

The EOSC-A structure is slowly developing and is expected to change as of the new roadmap gets developed. Therefore, there is an uncertainty regarding the EOSC-A structure as it may evolve after 2027, likely shifting from a membership/observer model (with current fee payments) to a new federated node structure. This new federated node structure is currently evolving, some first insights on the concept can be found in GÉANT's EOSC Nodes position paper[25] on behalf of the NREN community and the EOSC-Association Board's position paper[26] as well as the Science Clusters position paper[27]. The so-called Tripartite Group shared ahead of the EOSC General Assembly on 27th May 2024 the white paper ("Building the EOSC Federation: requirements for EOSC Nodes"[28]), which intends to help with starting the process of identifying potential Candidate/Prototype Nodes by the Tripartite Group to join the federation through a minimum set of requirements. These may evolve in the long term given the experiences of the first wave of nodes in the EOSC node federation. They also do not include interoperability requirements, but these will likely become important in a second step. The first results from the EOSC EU node as a platform are expected to be presented at the EOSC symposium in October 2024 with running services by 2025. It is worth noting, that the future of the EOSC EU node beyond the procurement and likely extension period is unknown. EOSC-A together with the Tripartite Group are currently trying to identify potential candidate EOSC nodes and will launch amongst its members an Expression of Interest in the coming months, as they aim to have a running federation by the end of 2025. Furthermore, the EOSC-A is planning work with EOSC-A member volunteer to draft a Handbook that contains guidelines and recommendations on how the federation and nodes can/will operate. This is a similar document to the ELIXIR's Handbook of operations[29].

In the following subsections, we first outline the strategic vision of SB SP towards EOSC, followed by an overview of the current initiatives and integration into EOCS Ecosystem and close with assessing data interoperability with EOSC. It must be pointed out that the focus lies on the activities of the organizations represented in the SB SP and the views reflects the opinions of the SB SP.

9.2 Strategic Vision of SP

9.2.1 Recent EOSC, ERICs and Science Clusters Development and the EU Framework of RDI Coordination

Participating actively in the building and development of Common European Data Spaces and EOSC is important from the SPs view. Working towards FAIR and as open as possible data is a desirable initiative that Switzerland should pursue at a national and international level to align with the EU. The EOSC concept is promising, but it needs to leverage existing scientific domains (aka science clusters) and their

[25] https://eosc.eu/news/2023/09/geant-publishes-eosc-nodes-position-paper-on-behalf-of-nren-community/
[26] https://eosc.eu/wp-content/uploads/2023/11/20231112-Short-paper-on-the-EOSC-Federation-draft-v3.pdf
[27] https://zenodo.org/records/10732049
[28] https://eosc.eu/wp-content/uploads/2024/05/EOSC-A_GA8_20240527-28_Paper-G_Update_EOSC_Nodes_requirements-DRAFT-v240524.pdf
[29] https://elixir-europe.org/system/files/elixir-handbook-operations-v7-202405.pdf

infrastructures to be fully effective and avoid duplication of efforts. The integration of existing RIs into the overall EOSC framework will entail a multi-faceted process that will likely involve the alignment of technical, organizational and policy aspects. Currently, the engagement of RIs in the EOSC landscape is an ongoing discussion. There is an ESFRI-EOSC task force, that regularly meets with the EOSC Steering Board and European Commission and ESFRI RI/landmark to co-create the vision of EOSC, promoting Open Science and FAIR policy to RI and connecting them to EOSC. Identifying which RIs with a Swiss node are taking part in the ESFRI-EOSC Working Group can help assess and understand the impact and feedback into the country. It is important to emphasize that any over-harmonized structure that EOSC seeks to implement, should not go at the expense of established, running and successful infrastructures, to avoid re-inventing the wheel and work needs to be done synergistically. Works within RIs and science clusters and domains are based on community coordination and likely to be more effective as a starting point.

As national service providers the natural entry point and connection to EOSC at the moment is through ERICs and other international RIs where they exist. (For example, FORS connects through CESSDA, and SIB through ELIXIR.) This is key as much is happening through existing organizations, which have been working within their domain in FAIR and open research data initiatives. It is worth pointing out that there have been national initiatives to harmonize within some domains, for instance in the Social Sciences and Humanities, SWISSUbase is now the joint national archiving platform for the nodes of CESSDA ERIC (through FORS), DARIAH ERIC (through DaSCH), and CLARIN ERIC (through LaRS) and ensures connectivity to the respective ERIC platforms. In Life Sciences, within ELIXIR including SIB, there is close alignment and collaboration internationally to drive FAIRification processes, archiving, and reuse of data in the domain through open resources, research data and software management. SPHN is engaging in the Federated European Genome-Phenome Archive (FEGA), a federated and interoperable network to enable discovery and sharing of human genetic data across Europe and internationally. It is also working with health data initiatives in Germany and the Netherlands in view of interoperability with the European Health Data Space. The Swiss Biobanking Platform (SBP) is the national BBMRI-ERIC node. Within the context of its Connectome project and ORD activities, Switch has become involved in EOSC working groups.
Therefore, EOSC should be largely seen as a mean to build on existing scientific domains and their infrastructures at national and international levels and bring them together rather than aiming to build a new parallel structure. The actual services and platforms should be built as much as possible on existing national and European institutions and services, which, however, also must be willing and able to adapt their services and to coordinate with others.

Additionally, libraries have a big role to play within the EOSC ecosystem as harmonizers across the different domains, although their readiness for participating in EOSC and related initiatives varies across the fields. Libraries should be included in general collaborative development of standards as part of the EOSC actors to implement agreed-upon standards. Libraries can provide a valuable and beneficial contribution to the coordination of services, infrastructures and standards. Libraries across Europe are already participating in numerous EOSC projects and Task Forces.

Overall, Switzerland should approach recent EOSC developments with cautious optimism. It should recognize their potential benefits while emphasizing the need for careful evaluation of the alignment between EOSC initiatives and existing Swiss infrastructures as well as the financial implications and resource allocation required to integrate these developments effectively. Informed decisions are difficult to make as long as the structure of the future node(s) and federation remains largely under discussion and unclear. A coordinated national effort is needed to comprehensively map out the flourishing and evolving Swiss EOSC landscape. This effort should

- include the identification and engagement of all relevant national actors, such as leading research institutions, universities, higher education institutions, libraries, research infrastructures and specialised data centres from the different scientific domains or clusters.

- distinguish between the roles of these actors and the specific initiatives or projects they are involved in, such as data management projects or collaborative platforms.
- be facilitated by a mandated 'working group' or some national body such as the ORD Strategy Council, which could lead the process of addressing any major gaps in the national EOSC landscape and ensure a cohesive national strategy.

Furthermore, being associated with Horizon Europe and Digital Europe is key to support Switzerland's role in EOSC, especially given that the Horizon Europe INFRAEOSC calls are building EOSC functionalities, networks, and services through multiple projects and that Digital Europe projects populate also the European Data Spaces. The non-association of Switzerland with Horizon Europe and Digital Europe is clearly a disadvantage and hampering if Switzerland aims to be more present in the overall EOSC landscape in the close future, especially given the very low number of EOSC projects where Swiss entities are participating. It is worth noting, that the non-association of Switzerland to the Horizon Europe program has impacted the participation of Swiss institutions in certain INFRAEOSC calls in the past.
Ideally, it would be desirable that ultimately EOSC, even if driven by the EC and EOSC-A, becomes an inclusive environment with membership being open beyond the EU member states.

9.2.2 How Should Swiss Actors Adapt EU Framework of RDI Coordination?

Overall, SB SP widely recognizes the potential of EOSC to unify research infrastructures, building upon the existing nodes, ERICs, and RIs, in alignment with Switzerland's goals of promoting open science and data FAIRness. EOSC offers a platform to enhance coordination and interoperability among these infrastructures, thus facilitating international collaboration and strengthening Switzerland's position in the European research landscape. SPs emphasize the importance of a community-driven approach as well, where the development of national, thematic and institutional nodes within EOSC is guided by the specific needs of the scientific communities they serve.

However, there's no clear consensus on the specifics of Swiss involvement in EOSC. The lack of full Swiss participation in Digital Europe and Horizon Europe programs hinders the development of expertise and participation in the EOSC ecosystem. This lack of full participation, coupled with uneven engagement and resource constraints, poses challenges for developing a shared vision and implementing EOSC initiatives. Incentivizing active participation in EOSC projects, events, and task forces is crucial to foster a cohesive approach. Additionally, careful coordination is required to align and integrate a potential national EOSC node with existing efforts, such as those of Strategy Council, but also those that are taking part of RI or ERICs. While recent initiatives like the European Commission's interoperability strategy through Smart middleware (SIMPL) hold promise, their early stage of development may introduce complexities that need to be navigated.

In summary, while the outline of a shared vision exists among Swiss SPs, further efforts are needed to solidify and clarify this vision as the full scope of EOSC becomes clearer. A coordinated national effort involving diverse stakeholders and incentivizing broader participation will be instrumental in developing a concrete vision, uncovering existing gaps, and ultimately ensuring a more effective and impactful Swiss engagement in EOSC.

9.2.3 Relation and Division of Responsibilities between Thematic Nodes and a National Node in Switzerland

As described above, most of the work in the past years has happened within discipline-specific research infrastructures, such as ELIXIR, CESSDA, DARIAH, CLARIN and others. EOSC envisions these infrastructures as thematic nodes that should be connected and integrated into the EOSC federation. The EC took an active role in shaping this approach by having initiated the construction of a central EOSC EU node following a procurement process. In addition, partner countries are encouraged to build national nodes that could connect to the EOSC EU node and become part of the EOSC federation.

However, the key question of how these national nodes will interact with existing thematic nodes and the added value they would bring is not clearly defined in the evolving EOSC landscape. Furthermore, national nodes may be shaped very differently from one country to another.

Recommendation: The SB SP is of the opinion that at the European level the concept of an EOSC node is evolving and relations and responsibilities between national and the thematic nodes are to be defined. Therefore, before taking any rushed decisions, a national discussion on the strategic value of a national node should be held first with the corresponding actors to assess the value, role and shape of a potential Swiss EOSC node. Therefore, targeted actions to create a national node should only start once the tasks and added value of such a national node are clearly defined and understood, which is not yet the case. Proceeding prematurely risks duplicating efforts with minimal benefit.

In the meantime, the SB SP recommends adopting a collaborative approach to developing a national node, comprising all Swiss EOSC stakeholders as equal partners with national focus. These stakeholders should work to better align their services at the national level, follow the developments at the EU level and identify gaps that cannot easily be addressed. To identify these gaps, a multi-faceted approach could be employed comprising of surveys, stakeholder consultations and comparative analyses of existing infrastructures and international relations. These gaps ultimately highlight the areas where a national node can truly add value.

9.3 Current Initiatives and Integration into EOSC Ecosystem

9.3.1 Initiatives of Service Providers on EOSC Nodes: Current Situation, Capacities and Opportunities

The following members of the SB SP are already the Swiss node for an existing thematic node:

a) DaSCH: DARIAH
b) FORS: CESSDA (and also participation in ESS ERIC, SHARE ERIC and GGP that have as an objective to collect data in many different countries)
c) SIB: ELIXIR

In addition, SB SP is aware of the following Swiss nodes:

a) CLARIN-CH-LaRS: CLARIN
b) PSI: PaNOSC
c) PREMOTEC: METROFOOD-RI
d) SBP: BBMRI

At this point, the SB SP is not aware of any other nodes in use or in the planning stage in Switzerland. It should be pointed out, that the time available to deliver this report did not allow to obtain a comprehensive overview and/or other initiatives are in the making that the SB SP is currently not aware of. Hence, this list may be incomplete.

At the national level, a few meetings on a national EOSC node comprising several stakeholders, including members of the SB SP, has taken place over the past year. These meetings should continue and evolve into the above-mentioned collaborative approach in defining the role and value of a national EOSC node.

In the view of the SB SP, this process is in too early a stage, such that valid statements on capacities and opportunities can be given.

9.3.2 Current Participation of Service Providers in the EOSC/ERIC Ecosystem

Besides joining a national or thematic node, SP can integrate into the EOSC/ERIC/RI/Scientific Domain ecosystem at three distinct levels:

1. Membership in the EOSC Association: the following members of SB SP have opted for this participation: ETH Zurich (Mandated Organization by SERI), SIB, and Switch.
2. Membership in EOSC task forces[30], working groups and future Opportunity Area Groups. In spring the EOSC Task Forces were reconstituted. Currently, there are four task forces in which the following Swiss institutions participate:
    a. Technical and Semantic Interoperability: PSI, Switch,
    b. FAIR Metrics and Digital Objects: ETHZ, Switch,
    c. Health Data: ETHZ
    d. Long-Term Data Retention: ETHZ
3. Participation in EOSC related projects: Several SP are actively engaged in several EOSC related projects:
   DaSCH as the Swiss DARIAH ERIC node will be part of the recently granted INFRAEOSC project EOSC Data Commons which will start in April 2025.
   Switch is a partner in the INFRAEOSC projects EOSC Gravity and EOSC Data Commons.
   Libraries are also a contributor to the INFRAEOSC proposal EOSC Data Commons.
   SIB is already participating in the INFRAEOSC BY-COVID project, where it co-leads a work package aimed at establishing a framework to make infectious disease data openly accessible. This cross-domain INFRAEOSC project focuses on enhancing metadata interoperability for advanced interdisciplinary research, exemplifying close collaboration between ELIXIR and CESSDA.
   Additionally, SIB has recently contributed to three INFRAEOSC proposals – EOSC-FIDELIS, EOSC-EDEN and EOSC Data Commons – each of which has been successfully awarded by the EC. These projects address key EOSC priorities, including the development of a framework for trustworthy repositories, the long-term preservation of data with the establishment of a cross-domain network of curation experts, and the implementation of a search engine within the EOSC framework. These initiatives closely align with SIB's commitment to open research data, leveraging its open biodata resources and advancing national efforts in infrastructure interoperability.

This multi-tiered integration ensures that SPs play a pivotal role in shaping and enhancing the EOSC ecosystem.

9.3.2.1 Representation of Service Providers in EOSC/ERIC/ESFRI Governance Bodies

The SB SP is comprised by a mix of SP, and hence their representation differs in these governance bodies. On the one hand, the long-standing, active and coordinated involvement in certain ERICs and RIs has ensured that Swiss interests and priorities have been well-represented over many years in those specific governance bodies.
However, Switzerland's removal from ESFRI in 2022, has significantly impacted the ability to influence research infrastructure policy. This expulsion means that Switzerland is no longer able to participate in the ESFRI Forum and the thematic ESFRI Strategy Working groups (SWG) limiting its role in shaping research infrastructure policy. Switzerland is currently working toward reintegration into these crucial decision-making bodies.

Furthermore, in other governance bodies, particularly the EOSC-A, Swiss entities can become members or observers of the EOSC-A through an annual fee and consequently be part of the general assemblies that take part twice a year. Recently and to add further burden to our non-association to Horizon and Digital Europe funding programs, Switzerland has lost its membership in the e-Infrastructure Reflection Group (e-IRG).

[30] In spring 2024 the EOSC task forces were reconstituted, and their new composition has not yet officially been announced on the EOSC website. See https://eosc.eu/eosc-task-forces/

At a national level, there are monthly EOSC meetings organised by the mandated organisation (ETHZ) that some of the SP members regularly attend. These provide a valuable national EOSC forum to share ideas and information amongst the Swiss stakeholders. Additionally, last September, a workshop was organised by the mandated organisation (ETHZ) and SWITCH with some of the national stakeholders to start the discussions towards the EOSC node concept and inform about the EOSC federation. Quarterly Swiss EOSC coffee sessions are organised to promote and inform the national stakeholders of ongoing projects or activities that are happening at the EOSC European level.

9.3.2.2 Representation of Swiss Interest: A Recommendation

First and foremost, a full association of Switzerland not only to Horizon Europe, but also to Digital Europe would be the best way to support our activities.
Lacking this association, we recommend the following measures:

a. A stronger coordination role for SERI with respect to Swiss nodes (ERIC, ESFRI, …)
b. A mandate from SERI to participate stronger in EOSC activities with possible funding support
c. A visible alignment and coordination between the national ORD effort and EOSC activities. As first step, the national landscape could be mapped to EOSC activities to identity strengths and gaps.

9.4 Assessing Data Interoperability in Switzerland: Enabling EOSC Participation and Exploring EOSC's Potential as a Driver for Improvement

A comprehensive assessment of Switzerland's data interoperability situation, including the evaluation of latest developments such as those funded by ORD measures, requires more time than was available to write this report. However, the following promising examples at the national level highlight Switzerland's potential for engagement and contribution to EOSC.

Swiss National Examples of Data Interoperability (non-exhaustive):

- **Metadata Level**: Swiss libraries actively implement international metadata standards, such as Dublin Core, to ensure the findability and accessibility of research outputs. The SWISSUbase platform implements standardized metadata schemas for archiving and disseminating data across diverse domains. The open government data platform, opendata.swiss, provides standardized metadata for its datasets, enhancing discoverability and reuse across various sectors. EnviDat, the WSL's environmental research data platform, utilizes Dublin Core for comprehensive and compatible metadata. Switch provides data services for discovery and scientific reuse (APIs) of Open Data by aggregating and connecting metadata of ORD (scientific publications,datasets and projects) and non-scientific Open Data (GLAM, NGO, GOV).
- **Data Level**: The SIB, through its contributions to ELIXIR, ensures interoperability and reusability by implementing FAIR data principles and promoting standardized ontologies and formats for biological data. SIB has also developed RDM tools and FAIRification tools, such as the FAIR Cookbook, to guide researchers in adopting FAIR practices. In the health domain, the SPHN develops the SPHN Semantic Interoperability Framework, enabling secure and standardized health data exchange across national and international institutions and systems. Similarly, EnviDat adheres to FAIR principles and supports standard data formats along with RESTful APIs for data access, enhancing data-level interoperability.
- **Service Level**: SIB, as the Swiss ELIXIR node, actively participates in federated data analysis and collaborates with other ELIXIR nodes, harmonizing services and platforms for data exchange and analysis. Libraries have been instrumental in driving service-level interoperability by operating interoperable systems like BORIS, BORIS Portal, BORIS Theses, and Bern Open Publishing (BOP). The SDSC is focused on developing standards for semantic and technical interoperability, such as APIs and data transfer protocols, contributing to efficient data sharing and collaboration.

These efforts demonstrate Switzerland's dedication to data interoperability and its potential to actively contribute to the EOSC ecosystem.

Examples of Interoperability through European and International Collaborations (non-exhaustive):

- **European Level**: SIB's involvement in the ELIXIR infrastructure, facilitates data sharing and interoperability across Europe. Swiss libraries' integration with OpenAIRE is another example of European collaboration. SPHN is engaged with FEGA, planning to establish a Swiss node and repository of human genomic data.
- **International Leve**l: SIB's contributions to the Global Biodata Coalition (GBC) and the Global Alliance for Genomics and Health (GA4GH) showcase Switzerland's commitment to data interoperability on a global scale. The Swiss node of Global Biodiversity Information Facility (GBIF) is an excellent example of international collaboration on the metadata level, aggregating and sharing biodiversity data. In the context of SPHN, Switzerland participates in the Personalized Health Train project on federated analysis of health data between 22 countries worldwide. For the cataloguing of Swiss cohort meta data, a collaboration with the Maelstrom Research of McGill University was established. CESSDA has a joint data catalogue that include all national social science data archives, including SWISSUBase.

In summary, there are national as well as international examples which clearly show the value of the work done by Swiss institutions. It should be noted that most of these efforts were done within existing thematic clusters. The SB SP does not have a comprehensive overview of all these efforts and is not aware of one ever being done within Switzerland. We also refer to the above recommendation of mapping the current national landscape to EOSC activities to identify strengths and gaps.

9.4.1 Swiss Specificities and EOSC Interoperability

Following the FAIR principles, the Swiss standards and norms in ORD should be “as international as possible and as national as necessary”. Swiss specificities and legal norms must be considered to ensure successful EOSC integration. They include data protection laws, linguistic diversity and cultural heritage considerations. The Human Research Act sets specific conditions and processes for sharing personal data for research. They may also require adaptations to ensure compliance with Swiss sensitivities, but overall should be kept to a minimum. An example of such adaptation is the use of DCAT-CH, a Swiss-specific profile of the international DCAT standard for data catalogues.

10 Appendix I: Abbreviations

Abbreviation	Meaning
AAI	Authentication and Authorization Infrastructure
AI	Artificial Intelligence
API	Application Programming Interface
CEDS	Common European Data Spaces
CTS	Core Trust Seal
DMP	Data Management Plan
EC	European Commission
EDIC	European Data Infrastructure Consortium
EHDS	European Health Data Space
EOSC	European Open Science Cloud (Common)
ERIC	European Research Infrastructure Consortium
ERI	Education, Research and Innovation
ESFRI	European Strategy Forum for Research Infrastructures
HEI	Higher Education Institutions
IF	(EOSC) Interoperability Framework
OAIS	Open Archival Information System
ORD	Open Research Data
RDI	Research Data Infrastructure
SB SP	(ORD) Sounding Board Service Providers
SNSF	Swiss National Science Foundation
VM	Virtual Machine

11 Appendix II: Terminology

1. **Access Control**: The method by which access to data by users is managed and restricted. In research data archiving, this is important for protecting sensitive information and intellectual property.
2. **Data:** Used in a broad sense encompassing raw measurements, processed information, software, ontologies and models.
3. **Data Archiving:** At its core, archiving refers to the process of storing data that is no longer actively used in a secure and accessible manner for future reference or compliance needs. In a broader sense, however, archiving of research data refers to the process of systematically organizing and preserving digital research data to ensure it remains accessible and understandable for future use. This involves protecting the data from obsolescence, degradation, and loss, and includes strategies such as reliable digital storage, backups, and capturing rich metadata and documentation.
4. **Data Curation**: The active and ongoing process of managing data to ensure its quality and relevance over time, including updating metadata and ensuring accessibility.
5. **Data Governance**: The management of data's availability, integrity, and security, including policies, processes, and standards governing data use.
6. **Data Integration**: Combining data from multiple sources to provide a cohesive, unified view, ensuring that the integrated data maintains its quality and integrity.
7. **Data Lifecycle**: The series of phases that data undergoes from initial generation or collection to eventual archiving and reuse.
8. **Data Management Plan (DMP)**: A formal document detailing how data will be handled, stored, and shared during and after a research project. It outlines data collection, preservation, and access strategies.
9. **Data Preservation**: The process of maintaining data over time in a manner that ensures its ongoing accessibility and usability.
10. **Data Protection**: The process of safeguarding data against unauthorized access or corruption and ensuring its confidentiality and integrity.
11. **Data Publication**: The release of research data, often with accompanying metadata and licenses, to make it publicly accessible and usable.
12. **Data Quality**: The degree to which data is accurate, complete, reliable, and suitable for its intended use.
13. **Data Repository**: A storage location where data and/or metadata is kept and managed. In research, data repositories often have functions for archiving, cataloguing, and preserving data as well as making data available to other researchers and/or the public.
14. **Data Security**: The protection of data from unauthorized access, alteration, or loss, ensuring its confidentiality and integrity.
15. **Data Sharing**: The act of making data available to others, including the public or specific research communities, for further analysis and interpretation.
16. **Data Stewardship**: The management and oversight of data to maintain its quality, accessibility, and security throughout its lifecycle.
17. **Data Storage:** refers to the use of recording media to retain digital information.
18. **Digital Preservation**: The series of managed activities necessary to ensure continued access to digital materials for as long as necessary. This goes beyond storage and includes maintaining the usability of data.
19. **Ethical Data Use**: The use of data in ways that respect ethical considerations, privacy, and confidentiality, and comply with all relevant laws and guidelines.

20. **FAIR Principles**[31]: Principles that advocate for data to be Findable, Accessible, Interoperable, and Reusable, facilitating knowledge discovery and innovation.
21. **Interoperability**: The capacity of diverse systems and organizations to work together, allowing for the seamless exchange and reuse of data across different formats and platforms.
22. **License for Data**: Legal agreements specifying the terms and conditions under which data can be used, shared, or modified.
23. **Metadata**: Descriptive information about data that aids in its discovery, use, and management, ensuring its findability and accessibility.
24. **Open Research Data**: The practice of making research data publicly available online, ensuring that it is accessible, reusable, and as open as possible.
25. **Reuse**: The ability for hardware, services, and software components to be shared among research data infrastructures, reducing costs and improving interoperability.
26. **Sensitive Data**: Data that must be handled with extra care due to its confidential, personal, or sensitive nature.

[31] https://www.go-fair.org/fair-principles/