swissuniversities

**swissuniversities**
Effingerstrasse 15, PO Box
3001 Bern
www.swissuniversities.ch

# National Open Research Data Strategy

## Analysis Report based on Survey and Workshop Panels

swissuniversities

**Impressum**

| | |
|---|---|
| Contractor | Jean-Marc Piveteau, President of the Open Science Delegation<br>*This report serves as background analysis for the Open Research Data Project Group in the preparation of the ORD National Strategy and Action Plan developed on the basis of a SERI mandate.* |
| Project Manager | Dr. Martin Jaekel |
| Report Version | 29.11.2020 – V.1.3 |
| Report Author | Dr. Martin Jaekel, Dr. Claudia Biniossek, Dirk Betz |

swissuniversities

**Content**


1. Introduction: Concepts and Developments 7
   1.1. Basic Concepts on Open Research Data 7
      1.1.1. Comprehensive Capture of Digital Objects during the Research Data Cycle 7
      1.1.2. Research Data Centers 8
      1.1.3. Repositories 8
      1.1.4. Dimensions to Classify Repositories and Data Centers 9
      1.1.5. International Developments 10
2. ORD at Swiss Institutions (Analysis of ORD-Survey) 10
   2.1. Introduction and Methodology 10
   2.2. Participating Institutions 11
   2.3. Policies on Open Science 11
   2.4. Monitoring of Open Science 15
   2.5. Institutional Data Storage Infrastructure 16
      2.5.1. Existing Data Storage Infrastructures 17
      2.5.2. Planned Data Storage Infrastructures 18
   2.6. Institutional Open Access Infrastructures 18
      2.6.1. Existing Publication and Open Access Infrastructures 18
      2.6.2. Planned Publication and Open Access Infrastructures 19
   2.7. HPC and Cloud Computing Infrastructures 20
      2.7.1. Existing HPC and Cloud Computing Infrastructures 20
      2.7.2. Planned HPC and Cloud Computing Infrastructures 20
   2.8. Repositories 21
      2.8.1. Existing Repositories 21
      2.8.2. Planned Repositories 22
   2.9. Institutional Services Landscape: Current and Planned E-science Services 22
      2.9.1. Data Management Plan Services 22
      2.9.2. Publication and OA Services 23
      2.9.3. General RDM Services 23
      2.9.4. Scientific Computing / Reproducibility Services 24
   2.10. General Comments 24
   2.11. Conclusions of Survey Analysis 24
3. Future ORD Landscape for Switzerland (Analysis of the Workshop Panels) 25
   3.1. Concept of the ORD Panels and Objectives 25
   3.2. Results 26

3.2.1. Best Practices 26
3.2.2. Diversity: Need for Differentiation by Research Purpose, Discipline and Data Type 27
3.2.3. How to Support a New Data-sharing Culture? 28
3.2.4. Do we Need Central Swiss Solutions? Need for Discipline-specific/Inter-institutional Network Collaborations versus Central Solutions 29
3.2.5. Efficiency and Financing 30
3.2.6. Challenges and other Remarks 30
3.3. Summary and Conclusion of the Chapter 30
3.3.1. Summary of the Workshop Notes 30
3.3.2. Conclusion of the Chapter 31

swissuniversities

**Appendix**

Annexe 1: List of Research Data Infrastructures

**Glossary**

| | |
|---|---|
| DFG | Deutsche Forschungsgemeinschaft |
| DMP | Data Management Plan |
| EOSC | European Open Science Cloud |
| EUDAT CDI | EUDAT Collaborative Data Infrastructure |
| GWK | Joint Science Conference (GWK) |
| NASEM | U.S. National Academies of Science, Engineering and Medicine |
| NFDI | Nationale Forschungsdateninfrastruktur |
| OA | Open Access |
| ORD | Open Research Data |
| OS | Open Science |
| PID | Persistent identifier |
| RatSWD | German Data Forum |
| RDA | Research Data Alliance |
| RDC | Research Data Center |
| RDM | Research Data Management |
| SRDD | Swiss Research Data Day |
| Swiss-ORD | National Strategy Open Research Data |

swissuniversities

## Executive Summary

This analysis report accompanies the Swiss National Open Research Data Strategy (ORD-Strategy); it has been written at the request of the SERI to clarify concepts underlying the ORD-Strategy as well as the current state of the ORD landscape. It is based on a survey among Swiss higher education institutions and three expert workshop panels. To **distinguish different types of ORD infrastructures**, the report suggests three dimensions. The first dimension is the **quality** of the infrastructure as defined, for example, by the quality of the digital object, the quality of metadata, the technical quality of the infrastructure, and its compliance with the FAIR data principles (e.g., as measured by the FAIR maturity indicator). The second dimension is the **coverage and comprehensive capture** of digital objects during the research data cycle, including process data, algorithms, and used software. The third dimension is the **reach**, measuring the amount and size of deposited digital objects and the frequency of re-use of these digital objects. This third dimension also defines the efficiency and costs of infrastructures. This report will not classify any existing or planned infrastructures but defines possible frameworks for such classifications.

The **Swiss ORD landscape** currently has the following attributes:

- A majority of institutions have already implemented open science (OS) **policies**. Particularly open access (OA) policies are well established, while policies regarding ORD, DMPs, long-term data availability, and compliance with the FAIR principles are established by less institutions. Especially umbrella OS policies are not yet widely implemented.
- Most institutions have established **monitoring** mechanisms for OA publications, while monitoring of research data management (RDM) is less well established.
- Most institutions reportedly use internal **OS infrastructures** with a future trend towards using more external infrastructures. For **OA publications**, a majority of institutions have built in-house solutions, which they will also continue to use for future developments. About one-third of the institutions use internal **data repositories**. Another third relies on external OS solutions, like FORSbase, SWISSUBase, openBiS, Zenodo, or OSF. This heterogeneous picture is likely to persist in the future. The European Open Science Clout **(EOSC)** is not often considered for infrastructural solutions.
- The **HPC-infrastructure** landscape is heterogeneous with internal, external, and hybrid infrastructures currently used; future trends point towards an extension of existing solutions and the expansion of connections to external HPC providers.

**Institutional OS services** are available at most institutions. Especially DMP and publication/OA services exist at most institutions. Interestingly, almost half of the institutions focus on individual support and consultation for researchers. General RDM services and scientific computing/reproducibility services are offered to a lesser extent.

The assessment of **investment and operational costs** proved to be difficult since performance, size, and quality of services (single repositories versus RDM departments) have to be considered to measure, e.g., costs per archived digital object. For all infrastructures, most institutions tend to use on-premise instead of in cloud services, which have to be considered in future national solutions.

**The workshop panels testify to the current landscape's broad diversity.** Guidelines, best practices, metadata schemes, teaching and training, and standards on a national level are preferred to be provided centrally. In contrast, the experts discussed a national data repository/data center solution very controversially. A central solution seems challenging, given the diversity of disciplines and data types. Hence, according to the participants, a diversity-driven balance of central and local solutions and heterogeneous scientific communities' involvement is needed. Simultaneously, the efficiency and the use of existing tools and services need to

be considered. In summary, the workshops demonstrate the benefits of involving various disciplines in developing the future Swiss-ORD strategy.

swissuniversities

swissuniversities

## 1. Introduction: Concepts and Developments

In January 2020, swissuniversities and the State Secretariat for Education, Research and Innovation (SERI [en] / SEFRI [it & fr] / SBFI [de]) signed the agreement to develop a **National Strategy Open Research Data (Swiss-ORD**) and a corresponding action plan. In addition, swissuniversities is to deliver an analysis report providing necessary background information. In particular, the analysis report´s objective is to involve all Swiss academic institutions and other relevant parties to assess their current status and needs concerning open research data (ORD). Therefore, this report is based on a survey and workshop panels focusing on which future data infrastructure for repositories, data centers, and services fits the current and future needs. Swiss-ORD has run a **Swiss Open Data survey.** The data entry forms of the conducted survey were sent to all Swiss universities, ETHs, universities of applied sciences, the 4 academies, and 19 other research organisations. This sums up to **58 contacted institutions**. The questionnaire mainly focused on the status quo concerning used infrastructures, services provided, and policies implemented. Furthermore, it asked about planned and required infrastructures and services to identify the future ORD-strategy's needs. In addition to the survey, several **workshop panels** were consulted on the development of a future ORD landscape. Feedback and comments were documented for qualitative data analysis.

### 1.1. Basic Concepts on Open Research Data

Efficient data-driven research will depend on the early and stable availability of data and other digital artifacts (code, workflows etc.). Such early and stable availability requires simple but difficult to achieve measures: (1) Data needs to be stored, curated, and made accessible by trustworthy repositories for long periods of time. (2) Data needs to be findable by globally unique, persistent, and resolvable identifiers (PID). (3) Data needs to be associated with suitable and comprehensive metadata. These pillars are the base of the FAIR data principles[1] and make up what the globally recognized concept of (FAIR) Digital Objects implies.

The ORD concept is part of the international open science (OS) movement, which also includes concepts such as open software, open access (OA), open review, and open method. OS seeks to make scientific research and the data it produces publicly accessible to all (amateurs, researchers, and professionals alike). To achieve this, OS relies heavily on using the open Internet, the web, open data, collaborative work tools, online training, and the social web. Because it is voluntarily and actively open, it can also foster multidisciplinary research and possibly a multilingual character by considering science and data as a "common good." There is a general understanding that efforts on an institutional level might not be enough to fully realize the potential of OS. In addition, there ought to be efforts on a national (e.g, the program Nationale Forschungsdateninfrastruktur [NFDI] in Germany) or even on a supra-national level (e.g., the European Open Science Cloud [EOSC]). However, these national and supra-national endeavours need to involve the relevant partners in a bottom-up approach. In this spirit, swissuniversities included all institutions in the development of the Swiss national ORD-strategy.

#### 1.1.1. Comprehensive Capture of Digital Objects during the Research Data Cycle

It is one key objective for any national strategy to clarify in the universe of scientific digital objects which data and data types shall be included in the national ORD-strategy. Figure 1 illustrates that the vast majority of research data is not yet published in repositories or data centers.

1 Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038

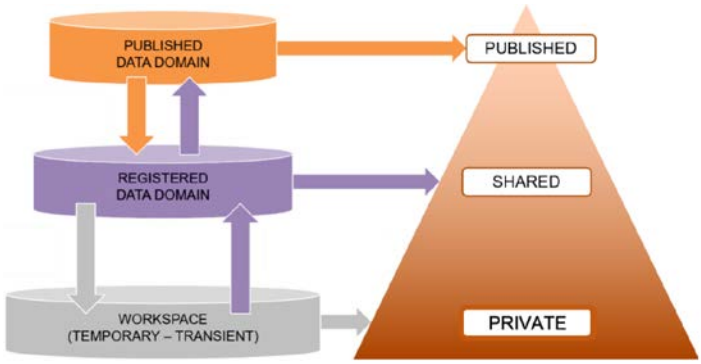


**Figure 1:** Data Domains defined by EUDAT Collaborative Data Infrastructure (CDI)[2].

The survey, as well as the workshop panels, reflects the importance of non-published data. This includes data that cannot be published (because of legal issues, anonymity, dual-use, etc.) or is not subject to publication, according to a DMP. Figure 1 illustrates the volume of non-published data within registered data domains and the workspace. In addition, many disciplines still lack existing data repositories with specialized metadata standards and currently publish data in unspecific (commercial) infrastructures (e.g., figshare, Mendeley Data).

### 1.1.2. Research Data Centers

The agreement between SERI and swissuniversities distinguishes between the concept of repositories (Repositorien/ dépôts de données de recherche/ repository di dati di ricerca) and research data centers (RDCs) (Daten-Zentren/ Centres de données/ Centri per i dati scientifici). In Switzerland, the term *data center* is often used in a more institutional and less technical sense. Hence, RDCs sometimes cover data processing and service functions for research purposes. In contrast, the term RDC is used in Germany for digital data sources with minimal functions. There are 38 such RDCs[3] related to the German Data Forum (RatSWD) with a broad heterogeneity of functions. Minimal requirements to get listed as RDC are: (1) providing at least one data access path, (2) providing sufficient data documentation, (3) ensuring the long-term availability of the data. Contrary to both the Swiss and the German definition, the term RDC is sometimes used synonymously with the term repository. From this perspective, data centers (in the sense of repositories) need to take a much more active role in making steps towards certification and FAIRness and guiding their customers – the individual researchers. In recent literature, this includes, in particular, to foster modular and generic data management workflows[4].

### 1.1.3. Repositories

A repository, often equated in everyday language with the database concept, is defined as a digital archive in which research data is stored, described, and managed. Repositories contain Digital Objects, as initially defined by Kahn & Wilensky[5], extended by Research Data Alliance (RDA) based on many use cases[6]. The managed objects within a repository can be, for example, programs (software repository), publications (document server), data models (metadata repository), or business management procedures. Often a repository also contains

2 De Smedt, K., Koureas, D., & Wittenburg, P. (2020). FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications, 8(2), 21.
3 https://www.konsortswd.de/en/datacentres/all-datacentres/
4 https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda
5 R. Kahn,·R. Wilensky: A framework for distributed digital object services; https://www.doi.org/topics/2006_05_02_Kahn_Framework.pdf
6 [4] RDA DFT: DFT Core Terms and Model; http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318

functions for version management of the managed objects. The Smithsonian Libraries[7], for example, distinguishes between specialty repositories, general-purpose external repositories, or local Smithsonian Institute repositories, defines quality measures according to the FAIR principles, and recommends to make sure that the services and terms offered to fit the needs of the data. The SNSF assesses the *FAIRness* of a data repository as FAIR compliant if the repository provides "sufficient information to be clearly verified as compliant with the SNSF criteria"[8] (p.6), which is measured by a set of minimum criteria that repositories have to fulfill", e.g., it allows publishing of FAIR data and is non-commercial[9]. One example of providing sufficient information is the Seal for Trustworthy Digital Archives[10].

swissuniversities

The FAIR data principles, as first described by Wilkinson et al. (2016)[11] give a guideline on how to proceed with data to meet a high level of good scientific practice. The FAIR principles mark the beginning of a global understanding of data. It is widely agreed that trustworthy and FAIR compliant repositories are the pillars of a stable and efficient data landscape[12]. One approach to measure FAIRness is the FAIR Data Maturity indicator[13]. However, established data managing practices are continued, and a large variety of software stacks (files, clouds, databases) is being used. In these cases, internationally accepted assessments for trustworthiness, such as CoreTrustSeal[14], are not applied. Only stricter practices will finally help improving data management[15].

#### 1.1.4. Dimensions to Classify Repositories and Data Centers

The agreement between SERI and swissuniversities distinguishes between "local repositories" and the use of "nationally" and "internationally established repositories." Criteria for evaluating existing or planned repositories and data centers can, for example, be assigned to the following three dimensions:

(A) Quality:

- The relevance of the digital objects contained (e.g., data collections corresponding to articles published in journals with a high impact factor)
- Quality of the digital objects itself (e.g., description with sufficient, relevant, and standardised metadata)
- Quality of the repository (e.g., quality control mechanisms for data uploads)
- FAIR compliance(e.g., as measured by the FAIR maturity indicator)

(B) Coverage: comprehensive capture of digital objects during the research data cycle, including process data (see section 1.1.1)

(C) Reach:

- Regional, national or international use of and reference to the RDC / repository for data upload

7 https://library.si.edu/research/data-repositories
8 https://zenodo.org/record/3618123
9 http://www.snf.ch/SiteCollectionDocuments/FAIR_data_repositories_examples.pdf 10https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html;jsessionid=9CE1F5394D08E21AB0D4CA9C70945022.internet281
11 Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038
12 Turning FAIR into reality - Final report and action plan from the European Commission expert group on FAIR data; https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283
13 https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines
14 Core Trust Seal; https://www.coretrustseal.org/
15 Jeffery, K., Wittenburg, P., Lannom, L., Strawn, G., Biniossek, C., Betz, D., & Blanchi, C. (2020). Not Ready for Convergence in Data Infrastructures.

- Project internal, community internal, or transdisciplinary use of / reference to the RDC / repository for data download

This report will not classify any existing or planned infrastructure but shows possible frameworks for such classifications.

#### 1.1.5. International Developments

The OS movement has evolved quickly and continues to do so on an international stage. OS adoption and success is driven foremost by statements like the Leiden Manifesto[16], the Amsterdam Call for Action on Open Science[17], DORA[18], and contributions of the Group of European Data Experts (GEDE) in RDA[19], to name only a few. UNESCO has drafted the Recommendation on Open Science[20] for a global understanding of OS. With the FAIR Principles[21], researchers have a blueprint for making their research OS compliant.

swissuniversities

Current challenges in OS still arise in the field of OA publications from the commercial interests of publishers. In Germany, the DEAL project, a consortium under the leadership of the German Rectors' Conference, negotiated for the benefit of all German academic institutions OA agreements with the three largest commercial publishers of scholarly journals (Elsevier, Springer Nature, and Wiley). The negotiations were successful with Wiley in 2019 and Springer Nature in 2020 but not with Elsevier. The second major challenge remains the implementation of OS in the everyday routine of researchers. The U.S. National Academies of Science, Engineering, and Medicine (NASEM) Roundtable on Aligning Research Incentives for Open Science[22] aims to overcome this hurdle by incentivizing researchers for the practice of OS in their work. For this, NASEM brings together research funders, academic leaders, and researchers to implement new incentive structures and disperse them in the research ecosystem. Europe has chosen a slightly different path to tackle this issue with the EUDAT Collaborative Data Infrastructure[23] that supports OS research by offering heterogeneous research data management (RDM) services and storage resources for a widespread field of disciplines. For example, with B2SHARE[24], the EOSC-hub offers the possibility to store, manage, and reuse research data according to OS. Similarly, Germany started implementing a national research data infrastructure (NFDI)[25] with the first of three funding rounds in 2019. The NFDI intends to systematically manage scientific and research data, provide long-term data storage, backup, and accessibility. The stakeholders jointly involved are the Deutsche Forschungsgemeinschaft (DFG), the Joint Science Conference (GWK), and groups of users (consortia) to improve access to and use of research data in their respective fields. The funded groups of users in the first funding round largely correspond to different disciplinary communities. In both the European and German contexts, the advent of FAIR digital objects[26] could facilitate OS implementation and could thus influence the design of future data infrastructure.

## 2. ORD at Swiss Institutions (Analysis of ORD-Survey)

### 2.1. Introduction and Methodology

The swissuniversities ORD-survey was designed to gather data on the current and planned ORD activities at Swiss academic institutions and their needs concerning ORD. After discussions within swissuniversities, it was agreed that a high response rate would be achieved if

16 http://www.leidenmanifesto.org/
17 https://www.government.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science
18 http://www.snf.ch/en/theSNSF/research-policies/dora-declaration/Pages/default.aspx
19 https://www.rd-alliance.org/group/gede-group-european-data-experts-rda/wiki/gede-digital-object-topic-group
20 https://en.unesco.org/science-sustainable-future/open-science/recommendation
21 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.
22 https://www.nationalacademies.org/our-work/roundtable-on-aligning-incentives-for-open-science
23 https://www.eudat.eu/eudat-cdi
24 https://www.eosc-hub.eu/services/B2SHARE
25 https://www.dfg.de/en/research_funding/programmes/nfdi/index.html
26 De Smedt, K., Koureas, D., & Wittenburg, P. (2020). FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. Publications, 8(2), 21.

all Swiss academic institutions and associated initiatives and services were addressed and this fact would be communicated within the survey form; the survey was sent to 58 institutions. Mainly open questions were asked to obtain a broad and undistorted picture. A list of all infrastructures mentioned in the survey can be found in the appendix.

Our analysis of the survey follows the order of its sections and questions:

1. we give a short overview of the survey participants,
2. we summarize the various implemented OS policies and associated monitoring systems.
3. we present the institutional infrastructures landscape (including OA infrastructures, HPC, cloud infrastructures, and repositories)
4. we present the current institutional service landscape

swissuniversities

### 2.2. Participating Institutions

For the ORD-survey, swissuniversities consulted the following institutions: universities, FHs, PHs, and ETHs (including Eawag, EMPA, ETH Council, PSI, WSL); Academies; Blue Brain Initiative; CSCS; DaSCH; FORS; Innosuisse; NICT; SDSC; SHARE; SIB; SLiNER; SNF; SPHN; Swiss Biobanking Platform; SWITCH. In total, 58 institutions were contacted; 44 participated in the survey and returned the questionnaire resulting in a response rate of 75.9%.

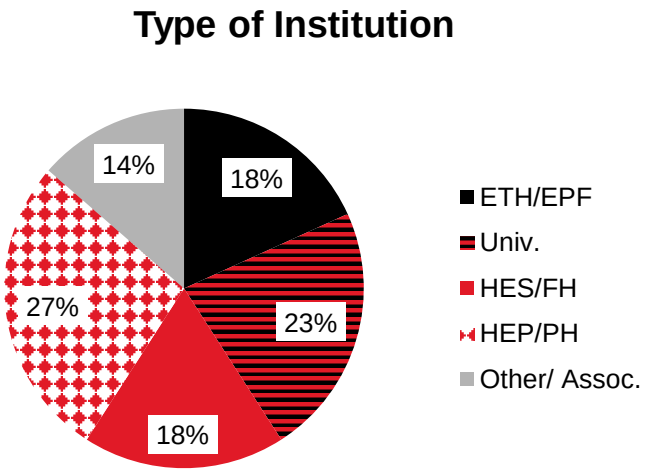


**Figure 2:** Participating institutions by type (*N*=44).

*The responding institutions were primarily members of swissuniversities (86.4%), consisting of 27.3% HEP/PH, 22.7% universities, 18.2% HES/FH, and 18.2% ETH/EPF.*

### 2.3. Policies on Open Science

The initial part of the questionnaire refers to policies related to OS at the institutional level. It covers umbrella-OS policies, OA-policies, and ORD-policies within the respective institution. Additional policies of interest regarding OS include RDM, long-term research data availability, and compliance with the FAIR principles.

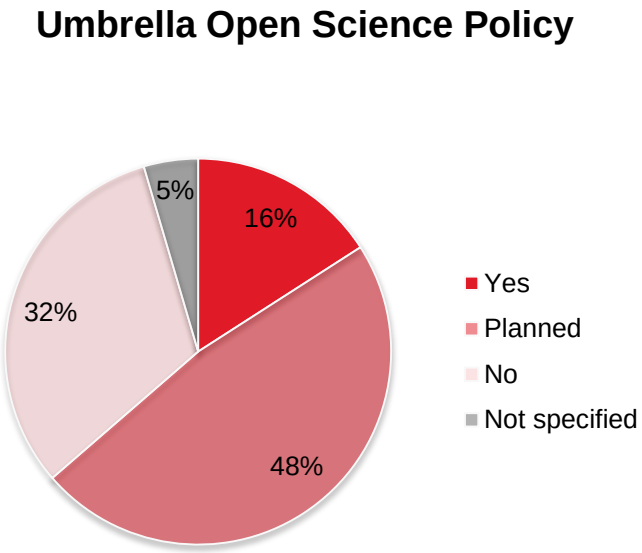


**Figure 3:** Existing policy on Umbrella-OS (*N*=44).

*Nearly half of the responding institutions (47.7%) have planned to develop an Umbrella-OS policy, while 15.9% already have such a policy. Almost a third does not have a policy nor currently plans one. In 3 of 6 institutions with an existing Umbrella-OS, a differentiated bundle of concrete measures is available.*

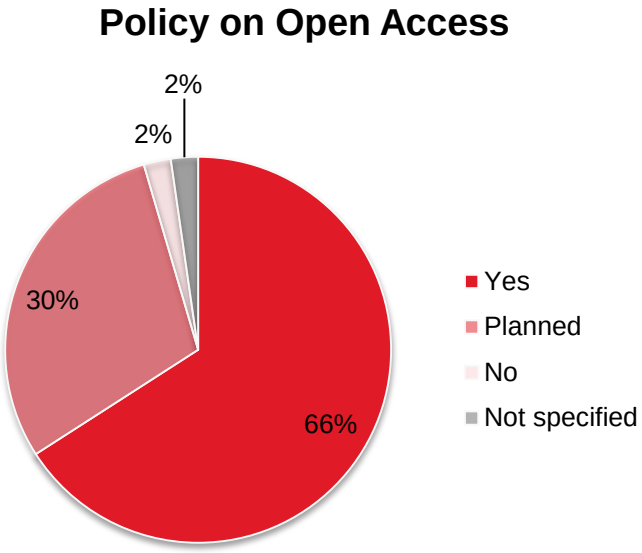


**Figure 4:** Existing policy on Open Access (*N*=44).

*The majority (65.9%) of the 44 institutions already have a policy regarding OA publications and 29.5% plan to develop it.*

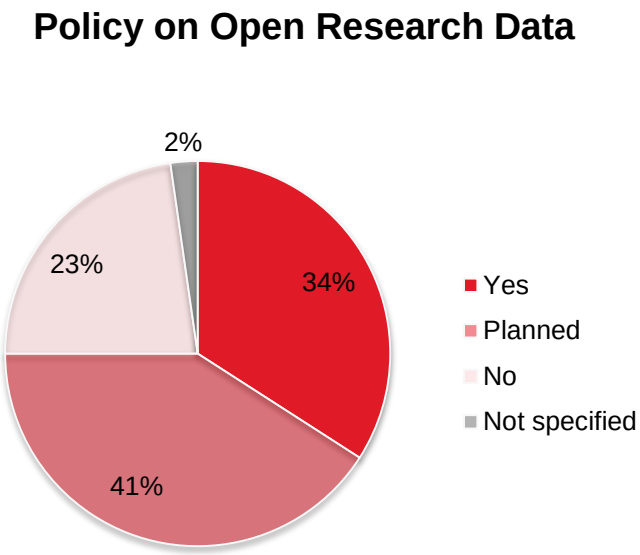


**Figure 5:** Existing and planned policy on open data (*N*=44).

*75% of all institutions see the need for a policy on ORD: 34.1% already have an ORD-policy while 40.9% currently plan one.*

The following policies complement the scope of the central institutional policies above and are also relevant regarding OS.

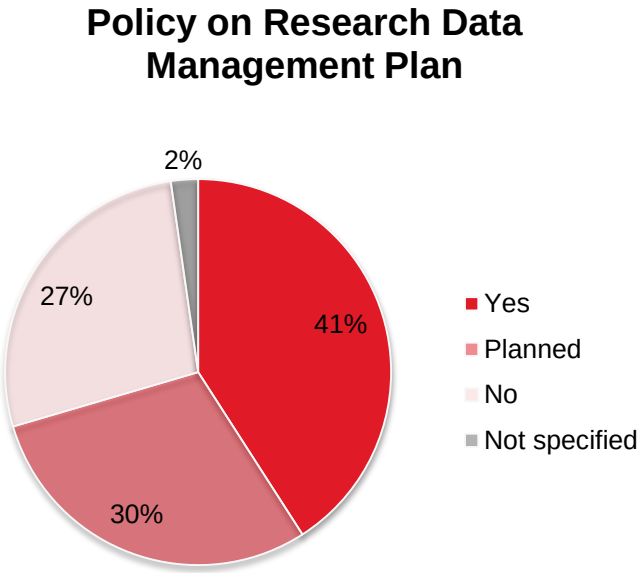


**Figure 6:** Policy on data management plans (*N*=44).

*A policy on research data management plans (DMP) is in place (40.9%) or planned (29.5%) in the large majority of institutions.*

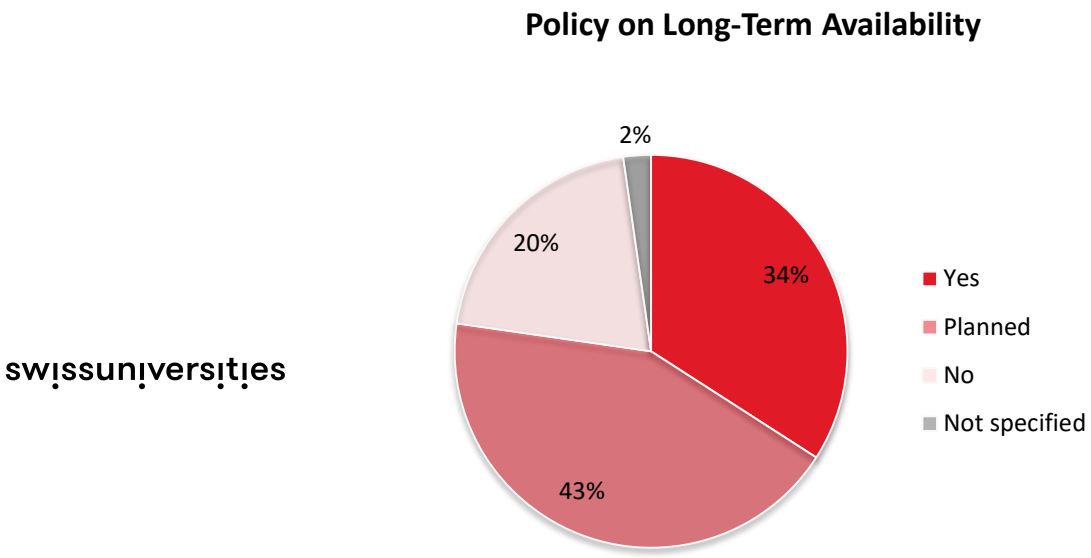


**Figure 7:** Policy on long-term archiving (*N*=44).

*Regarding the policies on long-term archiving, figure 7 shows a heterogeneous picture. 43.9% plan policies regarding archiving. 34.1% do have one, while 20.5% do not. 19 institutions plan to have a policy on long-term archiving; 7 of these 19 intended to implement a policy by 2021. Note: In the data science and data archiving communities, the expression long-term archiving is understood as archiving for at least ten years. Professional long-term archiving includes format migration (if necessary) to make data FAIR for generations to come. However, in the qualitative survey, some institutions use the term "long-term archiving" for storing data for five years in local data silos that are not FAIR compliant.*

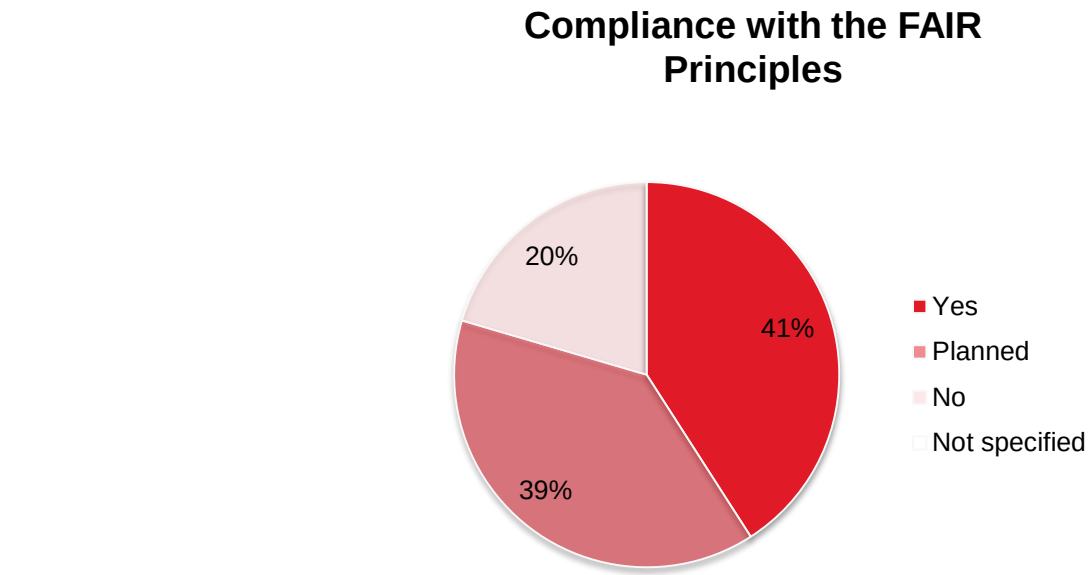


**Figure 8:** Self-reported compliance to the FAIR data principles (*N*=44).

*Policies regarding the compliance with the FAIR principles are implemented (40.9%) or are planned (38.6%) to be implemented by a large majority of institutions.*

In summary, most institutions claim to follow the FAIR principles and to have or to develop a long-term archiving strategy. On the other hand, the existing situation is far from FAIR compliant. Against this background, it still needs to be clarified how to translate the FAIR principles further into practice while certain elements are already there (e.g., the SNSF[27] assessment of DMPs and the set of minimum criteria that repositories have to fulfill[28], Seal for Trustworthy Digital Archives[29], the FAIR Data Maturity indicator[30]). The FAIR data principles, as first described by Wilkinson et al. (2016)[31] give a guideline on how to proceed with data to meet a high level of good scientific practice.



### 2.4. Monitoring of Open Science

The following figures refer to the survey question on the monitoring of OS within the institutions.

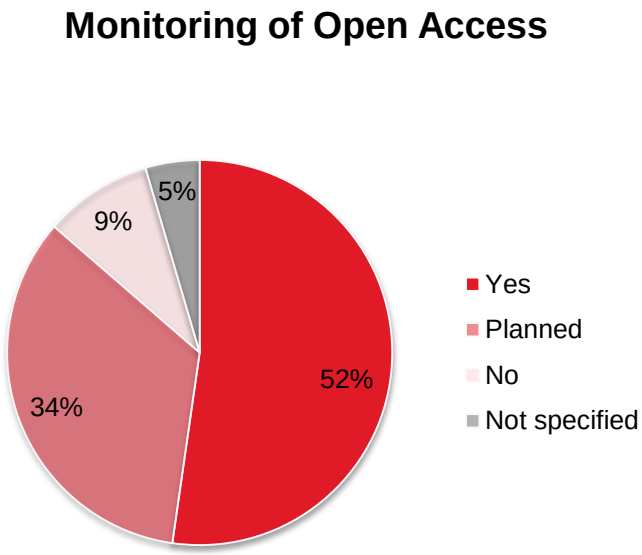


**Figure 9:** Monitoring of Open Access (*N*=44).

*Many institutions monitor (52.3%) or plan to monitor (34.1%) OA Publications.*

*Regarding the monitoring's exact procedure, 6 institutions report monitoring once a year, one institution twice a year. 11 (25%) institutions plan to integrate monitoring of OA in their institutional report system by the end of 2021 or earlier. 1 institution plans to implement monitoring at the beginning of 2022.*

27 https://zenodo.org/record/3618123
28 https://zenodo.org/record/3618123
29https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html;jsessionid=9CE1F5394D08E21AB0D4CA9C70945022.internet281
30 https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines
31 Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038

Compared to the monitoring of OA publications, monitoring of RDM is on a lower level.

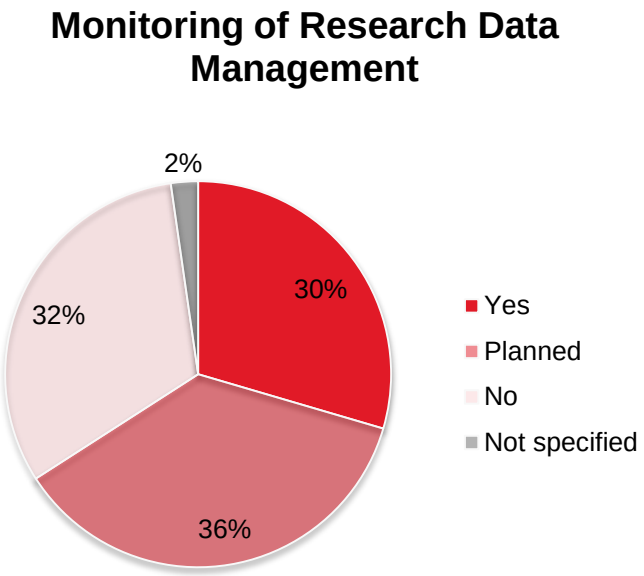


**Figure 10:** Monitoring of research data management (*N*=44).

*Almost a third of the institutions monitor RDM while another third does not. 36.4% do plan to implement monitoring of RDM, half of them at the latest until the end of 2021.*

*One reason for not implementing RDM monitoring is that RDM falls under the researcher's responsibility. In particular, for third-party funding, researchers are directly responsible for the DMP and the related RDM (the institution is only indirectly responsible). Another reported reason for not implementing unified institution-wide monitoring is different approaches to RDM across the disciplines and scientific communities.*

### 2.5. Institutional Data Storage Infrastructure

This block of the survey aims at assessing already existing infrastructures within the participating institutions.

### 2.5.1. Existing Data Storage Infrastructures

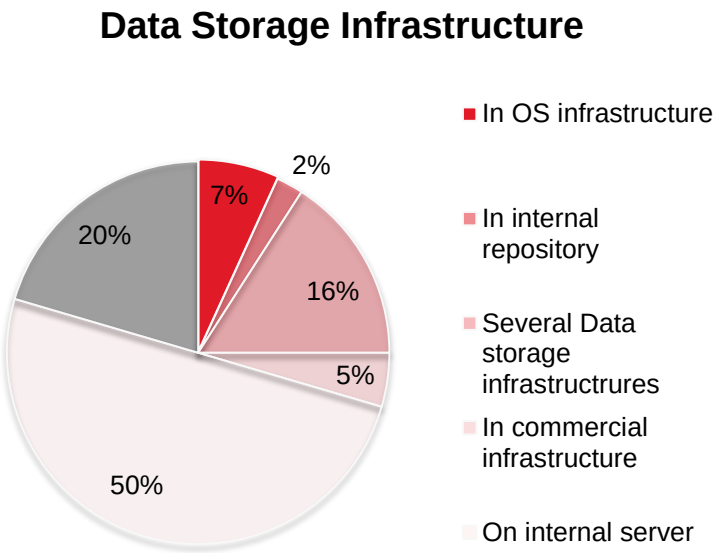


**Figure 11:** Type of existing data storage infrastructures (*N*=44).

*Asked for a short description of data storage infrastructures, most institutions (50%) state that they store their data on internal servers. Another 15.9% describes a two-step process concerning data storage: first, the data is stored on internal servers and then subsequently stored in an internal repository (e.g., ETH Research Data Hub, SENSA) or in an OS infrastructure like DaSCH, FORS/ FORSbase, LORY (Zenodo), RENKU (SDSC), YARETA, SWITCHdrive/ SWITCHengines, BIOMedIT(SPHN.ch), OpenRDM.swiss, DLCM. A small proportion of institutions does solely use an internal repository (2.3%), a commercial infrastructure (4.5%), or exclusively an OS infrastructure (6.9%) as data storage infrastructure. 20.5% do not provide information about their data storage infrastructure.*

**Costs (Investment and Operation)**

Concerning the costs of the data storage infrastructures regarding the infrastructure's investment and operation, nearly half of the institutions (47.7%) do not provide any information. About one-fifth (18.5%) cannot ascertain the data storage infrastructures' cost because they are part of the general IT services and therefore make a concrete break down difficult.

In five of the institutions (11.4%) that provide the costs for data storage per annum, there is a span from 100.000 CHF to 800.00 CHF per year. The average yearly costs are 420.833 CHF. Another six institutions (18.2%) give a more detailed overview of data storage costs in their institutions. If the annual costs are split between investment costs and operational costs, then investment costs range between 175.000 CHF and 5 Mio CHF, with an average of 1.525.000 CHF. Operating costs vary between 100.000 CHF and 400.000 CHF with an average of 201.200 CHF. The staff that is required can vary between 1 full-time equivalent (FTE) and 3.5 FTE.

**Is an EOSC Connection Planned?**

Half of the institutions do not plan to connect to EOSC, while only 6.8% plan to connect. Six institutions (13.6%) have not decided yet. The remaining institutions did not answer the question.

### 2.5.2. Planned Data Storage Infrastructures

Concerning future plans for data storage infrastructures, 27.3% of the institutions do have concrete blueprints. The following infrastructures were mentioned: openBIS, ACQUA, SWITCHengines, SWISSUbase, SwissFEL, SELVEDAS, SwissEGA, RoTaBio, CLARIN. A quarter of the institutions that responded to the survey have unspecified plans regarding enhancing the existing infrastructures, long- versus short-term archival solutions, and on-premise versus cloud infrastructures. 15.9% plan activities regarding data storage infrastructure but have not yet decided on the scope of the solution. 2.3% do not have plans, and 29.3% do not give any specifications.

swissuniversities

**Estimated Costs (Investment and Operation)**

Asked about the estimated costs for the planned data storage infrastructures, most institutions (61.4%) did not provide information. Another 22.7% stated that they could not provide information about the costs because those are unknown or will have to be specified by SWITCH. The estimated per annum costs provided by only 11.4% of the institutions range from estimates of 40.000 CHF up to 5 Mio CHF depending on the planned scope of the data storage infrastructure. Only 4.5% gave an estimation of the costs for investment and operation. For investment, the average costs are 750.000 CHF, whereas the average annual costs for operation are 150.000 CHF.

**Is an EOSC Connection Planned?**

A majority of institutions did not provide an answer. An EOSC connection is planned by 13.6% of the institutions, while 27.3% do not plan an EOSC connection.

**The Planned Date of Launch**

While most respondents did not give any specifications on the planned launch date, 22.7% state that they plan to launch their future data storage infrastructures by the end of 2021. 9.1% plan to launch by the end of 2022 and 2.3% after 2022.

## 2.6. Institutional Open Access Infrastructures

### 2.6.1. Existing Publication and Open Access Infrastructures

In this section of the questionnaire, the 44 institutions provided a short description of their existing publication/ OA infrastructure.

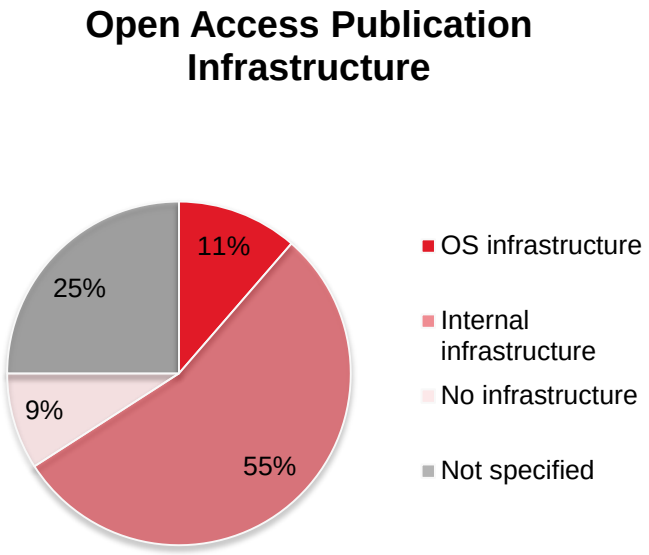

**Figure 12:** Description of the publication OA infrastructure (*N*=44).

*The majority (54.5%) state that they have an internal infrastructure like repositories or publication platforms, for example, SERVAL, HOPE, BOPserials, ZHAW digitalcollection, Instory, ORFEE, and edoc. Some of the mentioned infrastructures a used jointly by several institutions like RERO DOC or DORA, INFOSCIENCE, and Lib4RI in the ETH-domain. A minority (11.4%) uses OS infrastructures like LORY or ZOPAR, maintained by ZENODO, ZENODO itself, or Open Science Framework (OSF).*

*9.1% do not have any Publication/ OA infrastructure. A quarter of the institution did not provide any information.*



**Costs (Investment and Operation)**

45.5% of the institutions did not specify their costs for a publication/OA infrastructure or do not have expenses for these infrastructures. 27.3% of the infrastructures give their budget per annum. This budget varies from 10.000 CHF to 11.2 Mio. CHF. Some of the respondents give FTE costs, which vary between 0.4 FTE and 1.1 FTE per year.

**Is an EOSC Connection Planned?**

Half of the institutions did not provide any information if they plan an EOSC connection or not. 38.6% do not plan any EOSC connection for their publication/OA infrastructure. Some infrastructures are connected via OpenAIRE.

### 2.6.2. Planned Publication and Open Access Infrastructures

27.3% of the institutions plan to implement internal OA infrastructures. The following infrastructures were mentioned: New DORA, Infoscience3, ORFEE, ArODES, ORCIDintegrator, Lib4RI (with New DORA in the ETH-domain), BOPbooks, BOPdissertations, Alexandria, improvement of ZHAW digitalcollection. Cooperation with external OA infrastructures like SONAR or OJS is planned by 11.4%.

**Estimated Costs (Investment and Operation)**

While most institutions (65.9%) did not give any specifications about the estimated costs of future OA infrastructures, 9.1% stated the estimated costs per annum, which lie between 30.000 CHF and 150.00 CHF. However, it remains unclear if this covers investment costs, operational costs, or both. The few institutions distinguishing between investment and operational costs estimate a minimum of 200.000 CHF and a maximum of 400.000 CHF investment costs (the latter for an infrastructure covering the whole ETH domain). Estimated operational costs at these institutions are between 4 FTE (no monetary equivalent was stated) and 600.000 CHF per year.

**Is an EOSC Connection Planned?**

The majority (70.5%) of the institutions does not specify their answer. 18.2% do not plan to connect to EOSC concerning publication / OA infrastructure, while 11.4% plan to do so.

**The Planned Date of Launch**

18.2% of institutions estimate that the future publication infrastructure will be launched by the end of 2021. Two institutions (4.5%) have their planned infrastructures already established(OJS platform and University Journals platform). 9.1% plan the launch for the end of 2023, and 2.3% cannot define a date at this point. The majority thus did not provide any specific date.

## 2.7. HPC and Cloud Computing Infrastructures

### 2.7.1. Existing HPC and Cloud Computing Infrastructures

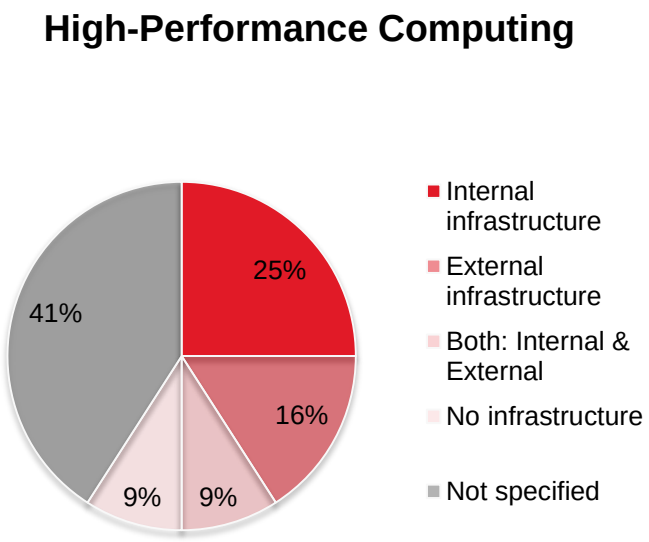


**Figure 13:** Description of the institutional HPC landscape (*N*=44).

*40.9% of the respondents did not provide any information concerning the current HPC landscape at their institution. A quarter of the institutions use internal HPC infrastructures like, e.g., sciCORE or Hyperion. Except for the cases that do not use any HPC infrastructure (9.1%), the rest uses external solutions or a mix of internal and external HPC infrastructures. Examples for used infrastructures are CSCS (ETH domain), SCITAS, RENKU instance, SWITCHdrive/SWITCHengines, BAOBAB, SENSA. Among the external infrastructures are also a few commercial providers like AMAZON cloud and web service and MICROSOFT AZURE.*

**Costs (Investment and Operation)**

Concerning the costs for HPC/ cloud computing infrastructures in their institutions, most (59.1 %) do not provide an answer. 6.8% give an amount per annum for the investment and the operational costs. The sum spreads from 100.000 CHF to 1.5 Mio CHF with a mean of 225.000 CHF. Costs divided by investment costs and operations costs spread for investment costs from 300.000 CHF to 3 to 4 Mio CHF with an average of 1.25 Mio CHF. The operational costs per annum lie between 60.000 CHF and 450.000 CHF. The staffing requirements are between 1 and 11 FTEs.

**Is an EOSC Connection Planned?**

Regarding an EOSC connection for HPC, a majority (56.8%) did not give any specification, while 40.9% say that they do not have any EOSC connection. Only one institution (2.3%) says that they do have an EOSC connection for their HPC.

### 2.7.2. Planned HPC and Cloud Computing Infrastructures

HPC development follows mainly two patterns: The extension of already existing (internal) HPC clusters (25%) and the extensions of connections to external HPC providers (6.8%) such as CSCS, RENKU, RoTaBio, SENSA, SwissFEL, or SWITCHengines. 13.6% have no plans regarding their HPC infrastructure, while 54,6% do not provide any information.

**Costs (Investment and Operation)**

Most institutions (77.3%) do not plan to extend their HPC capacities. Some institutions estimate that they have to invest around 2.5 Mio CHF for HPC plus a minimum of 80.000 CHF (up to 2 Mio CHF) yearly running costs.

**Is an EOSC Connection Planned?**

Concerning the statements for a future connection of HPC to EOSC, most (75%) do not provide any information. 18.2% state that they do not plan a connection, 4.5% have not yet decided, and for 2.3%, the decision depends on policies, OS strategy, and costs for the EOSC.

**The Planned Date of Launch**

A large majority (75%) of institutions did not provide an answer. 6.8% of institutions state that they do continuous (yearly) updates on their infrastructures. Only 13.6% of the responding institutions could give a date for the launch: 9.1% plan to launch in the next two years, 4.5% will launch between 2023 and 2025.

### 2.8. Repositories

#### 2.8.1. Existing Repositories

In this section, the 44 institutions that responded to the survey give a broad overview of repositories they are using.

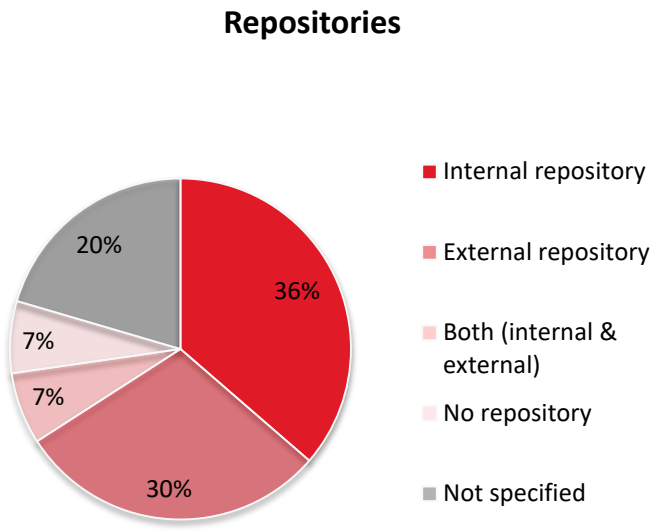


**Figure 14:** Type of repositories used (*N*=44).

*36.4% do have an internal repository, e.g., Alexandria, Archive ouverte, ArODES, BORIS, C4science, ERIC/intern, ERIC/open, edoc, EnviDat, Instory, IRF, Medienarchiv, PHIQ, SciCat, Tesi, ZHAW digital Linguistics – Corpus resources, ZORA.*

*29.5% of the institutions do make use of external repositories. The following repositories were mentioned: DaSCH, FORSbase, LORY, OpenDOAR, OSF, PHZH (by ZENODO), RERODOC, SWISSUBase, Yareta, ZENODO. 6.8% do use a mix of internal and external repositories in their institution. Another 6.8% do not have a repository, while 20.5% do not provide an answer.*

**Costs (Investment and Operation)**

The institutions quantify the costs of investment and operation as follows: The general costs per annum vary between 3.000 CHF and 500.000 CHF, giving an average of roughly 125.000 CHF for the eight institutions that responded. When the investment costs are listed separately, a minimum of 20.000 CHF versus a maximum of 3 Mio CHF is mentioned. Regarding the costs for operation, they lie between 0.4 FTE and 6.4 FTE. 9.1% cannot assign the costs to one specific account, 2.3% have zero costs, and 50% do not provide any information.

**Is an EOSC Connection Planned?**



31.8% of the respondents state that they do not plan an EOSC connection of their already existing repositories, while 15.9% do have or plan an EOSC connection mainly via OpenAIRE or CLARIN. The majority (50%) does not respond to the question.

### 2.8.2. Planned Repositories

Regarding planned repositories, the answers are quite heterogeneous. 15.9% want to establish internal repositories while 11.4% plan to enhance existing ones, e.g., EnviDat, or continue working with the yet established repositories (4.5%) like PHZH (via ZENODO). Another 18.2% intent to cooperate with external infrastructures most likely to meet OS criteria like SONAR, openBIS (ZENODO), SWISSUbase, ETHZ research collection (via SWITCH). 11.4% of the institutions stated that they evaluate possibilities or have work in progress. 31.8% do not provide an answer or state that this question is not applicable to their institution.

**Estimated Costs (Investment and Operation)**

Most institutions (61.4%) did not estimate costs for planned repositories or feel the topic not applicable to them. Some institutions estimate that their expected costs per annum would be between 5.000 CHF and 200.000 CHF plus 1.1 FTE. Some other institutions provide estimates of investment and operational costs; those range from 30.000 CHF investment and 50.000 CHF operational costs per annum to 2 Mio CHF investment and yearly 700.000 CHF operational costs.

**Is an EOSC Connection Planned?**

20.5% plan an EOSC connection for their planned future repositories. Therefore, for future repositories, a connection to the EOSC is planned to a much higher degree than for current repositories. Still, a large majority (61.4%) did not provide any information regarding this topic or felt it is not applicable to them.

**The Planned Date of Launch**

59.1% provided no information or did not feel the topic applicable to their case. The planned repositories in development should be launched by the end of 2021 (13.6%), 2022 (6.8%), or a later date (2.3%).

## 2.9. Institutional Services Landscape: Current and Planned E-science Services

This chapter summarises the institutions' feedback on the current and planned institutional service landscape for OS.

### 2.9.1. Data Management Plan Services

In a large majority (65.9%) of the institutions, data management plan services (DMP services) exist. 15.9% do plan them or plan to discuss them (2.3%). Three of the institutions (10.3%) plan to expand their DMP services due to researchers' increased demand. 15.9% do not provide any information.

The large majority (75%) of the respondent institutions give short descriptions or goals concerning DMP services. The concrete measures are heterogeneous; therefore, we provide here

a short overview of these measures. The majority of the institutions that provide DMP services state that they offer support and counseling on a general level. Some institutions offer guidelines, detailed (formalized) template guides, training and workshops, machine-actionable DMP, and DMP writing labs. Some institutions use tools like e-DMP generation via RENKU lab, the DMP Canvas Generator. Cooperation with external partners (training by external experts, collaboration with grant offices, partner institutions like FORS, PLATEC, DaSCH, BiUM, SIB CORE IT) is also referred to.

The localization of the DMP services within the respective institutions is heterogeneous: DMP services are sometimes located in the library or/and IT department while some institutions (12.1%) have a dedicated unit for DMP services. 15.1% of the institutions that do offer DMP services integrated those in higher-level units like the R&D department, Center for Scientific Competence, Prorectorate for Research, or they include research correspondents from the faculties. 18.2% did provide any information or did not feel the subject applicable to their institution (4.5%).



**Operation Costs/ Business Model for Data Management Plan Services**

20.5% of institutions state that these costs are part of the institutional budget. 13.7% give a more detailed overview of the exact costs that vary between 0.1 FTE to 2.8 FTE or Zero CHF to 10.000 CHF per annum. Nearly half of the institutions (47.7%) did not specify their costs or classified this question as not applicable.

### 2.9.2. Publication and OA Services

A majority (56.8%) of the institutions state that they offer publication and OA services. 6.8% state that they have an existing service structure that will be extended due to researchers increasing demand. 6.8% plan to install publication and OA services at their institution. 29.5% did not provide an answer.

Libraries play a significant role in providing these services. Support and consulting are the primary content (57.6%) of publications and OA services supplemented by workshops and training. Regarding archiving solutions for publications, the following infrastructures were named: Lib4RI for the ETH domain, ReroDoc, ORFEE, RENKULab (interoperable with, e.g., ZENODO, The Dataverse), SWITCH. The third category of support regards funding for OA publications (15.2%).

**Operation Costs/ Business Model for Publications and OA Services**

The costs for publications and OA services are reported by 22.7% of the institutions and cover a broad spectrum. Reported resources vary from 0.1 FTE up to 4.5 FTE; reported costs vary from 40.000 CHF for OA funding to 200.000 CHF designated for the transition to OA. However, these figures conceal very different publications and OA services. 20.7% state that the institutional budget covers the operational costs for publications and OA services.

### 2.9.3. General RDM Services

11.4% of the 44 institutions have RDM services and plan further extensions. 38.6% do provide RDM services but currently do not plan to extend them. 18.2% plan to offer RDM services in the future. 31.8% did not provide an answer.

A large majority (61.3%) of the respondent institutions give a brief description of their existing or planned RDM services. There is a focus on individual support and counseling (66.6%), while formalized guidelines, training, and workshops are also offered. Another focus lies in deploying data infrastructures/portals. Infrastructures that were named in this context include AGATE, openBIS, SWISSUbase, RENKU, ZHAW digitalcollection. In addition, the usage of various data analysis tools was also mentioned. These services are provided mainly by libraries, IT services, the R&D department, and research correspondents. Some institutions like ZHAW have a central drop-in for RDM (and OS in general, including OA, DMP, etc.), bringing together the library, IT, and R&D department in a new unit.

**Operation Costs / Business Model for RDM Services**

13.6% of the institutions state that their RDM service costs are part of the institutional budget or that the RDM service is spread across the budget of several institutional units (2.3%) and thus cannot be calculated. Another 13.6% give an overview of their costs for RDM services. They vary between 0.2 FTE and 108.000 CHF investments and 4.7 FTE. 61.3% did not provide an answer or indicated the question as not applicable.

### 2.9.4. Scientific Computing / Reproducibility Services

Scientific computing and reproducibility services are existent in 18.2% of the respondent institutions and planned by 20.5%. Another 11.4% of the institutions state that they have existing computing and reproducibility services and plan additional services. Half of the institutions state that they do not have this kind of service (4.5%) or did not provide any information on this subject (45.5%).

The institutions indicate a broad range of measures concerning the existing or planned services, including general support, training, providing RDM tools or offering a secure IT environment for sensitive data and encrypted information technology infrastructure. The department responsible for handling these tasks is commonly the IT service, though, as mentioned above, some institutions have a central drop-in for all the services mentioned above (OA, RDM, scientific IT services). Regarding scientific computing and reproducibility services, the responding institutions seek advice and support from external partners. Here, SDSC, Openstack, DAViS, FORSbase, SWISSUBase, or SENSA are mentioned.

**Operation Costs/ Business Model for Scientific Computing/ Reproducibility Services**

Many institutions (52.3%) do not provide information or state that the answer cannot be provided because exact information is not available. Some institutions state that industrial partners or grants cover the costs. In 15.9% of the institutions, the costs are part of the institutional budget.

### 2.10. General Comments

The survey allowed for open comments capturing issues that the institutions consider particularly crucial for the future Swiss-ORD strategy. These comments include:

- Some institutions state that they are in the middle of a planning period and are still in the process of implementing the elements of an OS strategy.
- Other institutions express their wish for a **centralized Swiss Platform on Open Science** and/or general services such as OLOS, SDSC, or SWISSUbase instead of institutional repositories as an opportunity for national collaboration.
- Further remarks relate to a national ORD-strategy's financial impact,
- stressing out that both operating costs and investments for infrastructures are substantial.
- Other institutions promote a demand-driven OS vision, pointing out that infrastructures should be research and researcher driven instead of the existing, inefficient supply-driven infrastructures.
- Another stated concern targets that current ORD efforts would negatively affect research with data that cannot be easily anonymized.
- A need to further clarify rules and regulations concerning data is also mentioned in the comments.

### 2.11. Conclusions of Survey Analysis

The survey responses illustrate the heterogeneous picture of OS in the Swiss research landscape. **Policies** are implemented or planned to be implemented by a majority of institutions.

Particularly policies regarding OA are well established (66 %). In contrast, umbrella-OS policies addressing ORD, RDM, long-term availability, and compliance with the FAIR principles are mostly still in development. Regarding monitoring of OS, most institutions have established the monitoring of OA but monitoring of RDM is only established in one-third of institutions. Institutions suggested in the survey that researchers are mainly responsible for RDM themselves.

According to the questionnaire, most institutions report using mainly internal data storage **infrastructures**. A trend towards using (external) OS infrastructures or just enhancing existing internal infrastructure can be observed. For OA publications a majority of institutions have built in-house solutions which they will also continue to use for future developments. The landscape of HPC-infrastructures is more heterogeneous landscape with internal, external and mixed infrastructures. The future development of HPC infrastructures follows mainly two directions: the extension of already existing solutions and the expansion of connections to external HPC providers. Nearly three-quarters of the institutions use some sort of data repository: a weak majority uses internal repositories while the rest uses external repositories or a mix of internal and external infrastructures. This heterogeneous picture is likely to persist in the future according to the questionnaire.

swissuniversities

I**nstitutional OS services** are also available at most institutions (66%). Especially DMP services exist in the large majority of institutions; however, the concrete support measures and the service providers differ between institutions involving libraries, IT departments as well as units responsible for R&D at an institutional level. Aside from DMP services, almost half of the institutions focus on individual support and counseling for general RDM services. Publication and OA services are provided by most institutions and mainly by their libraries. Scientific computing and reproducibility services are provided or planned by fewer institutions. Mainly IT departments procure a wide range of such services.

The assessment of **investment and operational costs** proved to be difficult since, in some cases, the costs are part of the global institutional budget while in other cases, the institution set up a service point responsible for OS and all related services and facilities. Costs for such a service center cost are not easily comparable with operating costs of a single repository. Furthermore, performance, size, and quality of services should be considered, e.g., to measure costs per archived digital object. The interpretation of the results above should consider these limitations.

For all infrastructure, most institutions tend to currently use **on-premise** instead of in **cloud** services. In future developments, there is a weak trend towards increased use of cloud infrastructures. Finally, the **EOSC** is only moderately considered for infrastructural solutions.

## 3. Future ORD Landscape for Switzerland (Analysis of the Workshop Panels)

### 3.1. Concept of the ORD Panels and Objectives

The workshop panels build on the ORD survey and were designed to further clarify the Swiss ORD landscape's future development. The panels concept and objectives were communicated in the description text for the Eventbrite registration tool:

"*The main goal of the panels is to identify, thanks to the expertise of the participants, existing practices and needs in terms of Open Research Data.* ***What would be, according to them, an ideal ORD landscape in Switzerland?*** *The panels will also allow collecting feedback on the strategy itself as well as on the results of the survey. […]*"

**Participants:**

All panels were open to all invited participants. To ensure the necessary expertise, selected experts from higher education institutions and ORD relevant organizations were invited. Additionally, delegates of institutions were invited to guarantee the involvement of all academic organizations (two persons per institution, each from a different field). The participants were

distributed across subgroups of the panel to ensure the presence of different competencies and backgrounds in all discussions. The maximum number of participants per panel was restricted to 60 persons, summing up to a total of 180 participants over 3 workshop panels. The actual number of participants was 105.

**All three online workshops followed the same program and reflected the invitations´ objectives:**

- **09:00 – 09:10:** Welcoming words/program of the day/ Introduction topic (SERI mandate ...)
- **09:10– 09:40:** Break Out Sessions 1: Research Data Management - what does the current landscape look like? Best practices, needs and challenges?
- **09:40- 10:10:** Plenary Session: return/discussion
- **10:20– 10:50:** Break Out Sessions 2: What would be the ideal ORD landscape in Switzerland a national strategy should aim at?
- **10:50– 11:20:** Plenary Session: return/discussion
- **11:30– 11:50:** Plenary Session: Discussing Vision and Objectives of the draft of the national Open Research Strategy
- **11:50 – 12:00:** Conclusion and Acknowledgements

**Moderation and Animation of the panels:**

- 7 moderators for the panel of 09.09.2020: Christine Choirat, Isabel Bolliger, Silke Bellanger, Christian Lovis, Henry Luetcke, Mark Robinson, Axel Marion
- 5 moderators for the panel of 14.09.2020: Christine Choirat, Isabel Bolliger, Silke Bellanger, Christian Lovis, Axel Marion
- 7 moderators for the panel of 16.09.2020: Christine Choirat, Isabel Bolliger, Christian Lovis, Martin Jaekel, Henry Luetcke, Mark Robinson, Axel Marion

Moderators documented the panels in the form of short summaries and bullet points. These notes were visible to all participants, not just the moderator(s). Panel results (notes) were uploaded to SWITCHdrive. The following descriptive sections report bullet points within different topics.

### 3.2. Results

#### 3.2.1. Best Practices

This section documents statements from workshop participants on best practice examples, on the helpfulness and disadvantages of existing infrastructures (repositories, data centers) and ORD initiatives.

**Best Practice Policies**:

- The DMP set up by the SNF is seen as a useful guideline for best practice.
- Through workshops and in PhD programs, train researchers on how to work with RDM infrastructures efficiently
- Supporting translation of strategies into practice through data councilors and data stewards

***Modular Aspects on how to Improve the Fair Maturity Indicator and Efficiency***

- National bibliography of ORD metadata would be useful (e.g., NARCIS: https://dans.knaw.nl/en/about/services/narcis), but only useful if the community is empowered to use it

***Data Repositories:***

- Data Repository - NASA: Advantages: gives an idea of quality and provides different levels of data Disadvantages: need a good understanding to use it; is project-specific,
- Genetics: sequence data is in discipline-specific databases (ENA, NCBI Genbank, etc.). Advantages: very structured, good practice, quality control of uploaded data possible.
- BORIS (institutional repository of UniBE): Possible / Potential Disadvantage: – Will this data be findable? (even if possible to get DOIs)



***Data Infrastructures Similar to Repositories Covering the Workspace:***

- Materials Cloud: Advantage: Preprint Archive with citable DOIs, stored data sets easily accessible.

***Services:***

- DaSCH services. Advantages: heavily used. Problem with financing the services, changed by long-term financing of FORS and DaSch; this is the first step by SNF
- ESA (European Space Agency): Disadvantage: not user-friendly and is more for raw data

***National and International ORD-Initiatives:***

- German NFDI (National Research Data Infrastructure) funding program: provides funding for consortia that have strong disciplinary roots https://www.nfdi.de/, https://www.nfdi.de/konsortien-2

**Conclusion Best Practices:**

A high level of satisfaction with ORD-practices is evident in data-intensive disciplines using discipline-specific national and international research data repositories and infrastructures. In contrast, in fields without community-driven ORD services, local, national and international repositories and unspecific infrastructures do not fill the gap.

### 3.2.2. Diversity: Need for Differentiation by Research Purpose, Discipline and Data Type

**Need to Define Disciplinary-Specific Best Practices on how to Translate the FAIR Principles into Practice:**

- Definition of data sharing: Important to have a discipline-specific understanding of what data sharing means. Need to find ways (discipline-specific) to agree on the standards; FAIR does not need to be the same for all disciplines.
- Discipline-specific guidelines are needed to indicate to researchers what they should publish (everything, raw data, metadata, results, etc.) and at what level of maturity should the data be published. Maybe a common methodology can be developed to define specific guidelines for disciplines.
- Consideration of qualitative research methods: some data generated by such methods cannot be meaningfully anonymised / pseudoanonymised (e.g., biographical data).
- Datasets are organized in specific ways, usually geared to more quantitative methods and interpretation; there is a risk of losing a lot of the more hermeneutical insights.
- Interface must be more interoperable: There is a need for a link between data centers and researchers. There is a need for standardization, but it is discipline-specific.

- Data sharing in medicine: lobbies and privacy requirements make it difficult to share (meta-)data. There is a need to come to an understanding of how to share meaningful data.

**Need to Define Standards: Which Digital Object (Literature, Data Publication, Unpublished Data from the Workspace) Shall Be Stored in which Infrastructure (Repository, Hardware/Cloud Services, Long-term Storage in Data Archive)?**

- Clarify further what data is in a publication and what is in the repository.
- Need to differentiate between types of infrastructure (repositories, hardware/cloud services etc.) and their roles.
- Data appraisal: Clarify what types of data need to be preserved as they are valuable for the communities. Repositories face the task of deciding which data to preserve and what to reject (difficult decision that needs a pragmatic approach).
- The national ORD-strategy could also address public data in general.
- Long-term data storage: Clarify how long data should be preserved. Long-term preservation and curation needs long-term funding (cannot come from project money)
- Proposal for a two-stage national strategy.

  1. In the first stage, researchers must be supported in carrying out their research activities via high-performance sharing platforms.

  2. In the second stage, platforms must be provided for the permanent archiving of data and also for the efficient dissemination of data in other words, for the sharing of information and the dissemination of results.

**The Challenge of Diversity**

- Consider discipline-specific needs to establish services/infrastructures. Need to establish specific solutions for different areas (which are not at the same stage of development at ORD level and therefore have different requirements).
- All disciplines must be taken into account (leave none behind). Everyone needs to be involved
- Manage the wide variety of data/metadata types across domains
- Consider the diversity of institutions: it is difficult for smaller institutions to provide all the support and infrastructure

### 3.2.3. How to Support a New Data-sharing Culture?

**Need for Awareness-raising and Culture Change on ORD:**

- Need to increase awareness of the value of data and ORD services.
- Finding a new way to evaluate the work and career of researchers.
- Depending on the discipline, researchers see ORD as an obligation and not as a source of benefit.
- Communities need to organise themselves, but they have to be pushed to collaborate.
- Need to develop a data culture: You cannot force people to share data: it is a slow cultural change. We have to give everyone the freedom to move at their own speed.
- As soon as institutions offer services, researchers will practice ORD.

**Need for more Support, Teaching, and Training on the Local versus the National Level:**

- Setting up the necessary support for the implementation of good practices at the national level.
- Various bodies are waiting for swissuniversities (or national-level) policy guidance.
- For ORD-trainings, it is important to include disciplinary know-how. Courses could, for instance, also be organised at the national level for individual disciplines.
- ORD services have to be close to the researchers and understand their needs. As it is the responsibility of the researchers to manage the data correctly. ORD Services should cover the whole data life cycle.
- Close cooperation of researchers with local IT-Services is necessary researches have a (strong) need for support and advice regarding available tools (e.g., in humanities).
- ORD services should be connected in a national network (creating synergies in training, guidelines, support, establishing best practices, etc.).

#### 3.2.4. Do we Need Central Swiss Solutions? Need for Discipline-specific/Inter-institutional Network Collaborations versus Central Solutions

**Statements in favour of federated networks:**

- Need to fund discipline-specific national networks capable of setting specific standards (should define know-how and educate researchers, discuss technical issues, raise awareness and compliance, etc.).
- Collaboration is a key in developing the tools for ORD.
- Need for coordination and federation of activities: need to bring the communities to structure themselves. Currently, there is a lack of coordination.
- Need for good connections between databases. Different infrastructures should be able to interoperate via common standards. Ideally, focus on established international standards (for example International Image Interoperability Framework, IIIF). Interoperability can be achieved via open APIs.
- Leading role of the universities and SNSF (not to a single institution/university): One possible approach is to develop discipline-specific solutions at one university and then open up for the whole community (different specializations at different universities)

**Balance between a national center and other existing solutions:**

- A Combination of federated structure and centralised structures is necessary as, for example, practiced by the GeoSciences in SCNat. There is a need for a central aspect to connect data
- There is a strong need to establish coordination for hardware/cloud services (Swiss equipment is of good quality, but coordination is lacking).

**Statements pro central/national structures:**

- From the library perspective, normalisation of research data is important to make it better accessible to the public and to connect it.
- Some disciplines (e.g., humanities) would benefit from centralised structures.
- A pilot project “connectome” seeks to connect data from various disciplines (through a linkage layer to other datasets). Such a connectome could represent a national layer on top of the discipline-specific repositories.

**Global perspective**

- Similar challenges are also faced in other countries (How is data curation handled? How is long-term funding secured?): CH should learn from experiences on an international level. EOSC, CERN, etc., show that it is necessary to balance the tasks of local and central levels.
- The ORD-strategy needs to ensure alignment with international standards (e.g., metadata), repositories, and international regulations (data security). But also a connection to the national ecosystem (e.g., BFS, Open Data initiatives)

#### 3.2.5. Efficiency and Financing

swissuniversities

**Need to Increase Efficiency and Reduce Costs by Improving Existing Services and Tools:**

- Improve existing databases, rather than reinvent the wheel.
- Convincing disciplines that have developed large projects with different technologies to merge them to reduce costs in the long-term.

**Opportunity-costs if there is no Swiss ORD Strategy**

- Private sector taking the lead on the ORD theme
- Inefficiencies in generation and use of research data will persist

**Need for Financial Investments**

- There is a strong need for long-term, sustainable funding of data infrastructures and qualified staff for ORD-services.
- Communities and institutions could help to identify costs
- Funding is needed to ensure connectivity nationally and internationally.
- Funding for a governance is needed to ensure coordination among institutional and national level and among institutions.

#### 3.2.6. Challenges and other Remarks

- Libraries, IT, and R&D departments face a big change in their service tasks.
- Similar process to open access would be nice; research communities need to be taken on board early
- The type of deposit used does not matter for some people
- Support funds, tools but no clear process, no regulation – the gap between high level and tools available

### 3.3. Summary and Conclusion of the Chapter

#### 3.3.1. Summary of the Workshop Notes

The participants' statements sketch a heterogeneous picture of the future ORD landscape depending on the contributor's professional and institutional background. The diversity of participants provides insight that complements the survey.

In a total of 106 session summaries (written by session moderators), 83 (78.3%) refer to the overarching question of the Swiss ORD strategy and the functions of future national RDCs:

Solutions that have to be provided centrally include guidelines, best practices, metadata schemes and standards on a national level. In addition, the need for central teaching and training activities is mentioned. A national data repository / data center solution is discussed very controversially. In particular, given the diversity of disciplines a central solution is challenging. Hence, a diversity-driven balance of central and local solutions is needed. Improving

and connecting existing tools and solutions could contribute to cost-efficiency in certain disciplines and cases. The relevance of efficiency and generic solutions is questioned.

### 3.3.2. Conclusion of the Chapter

Overall, the necessity for a balance between central functions and the diversity of disciplines, data types, and local solutions was the main discussion point. Efficiency and the use of existing tools are also important. In summary, the workshops demonstrate the great benefits of involving experts from various disciplines and institutions in developing the future Swiss-ORD strategy. Individual communities have potentially so far not been sufficiently been involved.

swissuniversities

**Appendix: List of Research Data Infrastructures**

Find below an alphabetical overview of the research data infrastructures mentioned in the survey.

AGATE

Airspace ELN (STI school)

Amazon Cloud Services

Amazon web services

ARBOR

Archival servers ZHdK IT infrastructure

Archive ouverte

ArODES

Baobab 2

BOP Serials. Application: Open Journal Systems

BORIS

C4science

CSCS

DaSCH

DAViS

Discipline-specific repositories / data E.g. ZHAW digital Linguistics – Corpus resources

DLCM

DORA

ELIXIR

EnviDat

EPFL Academic Output Archive (ACOUA)

EPFL Data Repository: ORD platform connected to long-term research data storage (expansion of ACOUA infrastructure)

Euler: On-premise (CSCS, Lugano)

FORS

FORSbase HOPE

INFOSCIENCE

Institutional repository (Instory)

Leonhard : On-premise (Zurich)

Lib4RI

LORY (ZENODO)

Medienarchiv

Microsoft Azure for HPC

New DORA

OLOS

Open Access publication platform based on OJS

Open Access repository edoc (Green Open Access)

Open Publishing Platforms for different publication types, emono/eterna (Gold Open Access)

openAIRE

openBIS (ZENODO)

openRDMswiss

ORFEE

OSF

PetaByte-Archive

PHIQ

Projektordner auf Microsoft Sharepoint

RDH (ETH Research Data Hub)

RENKU platform

RenkuLab

Repositorium PHZH

repository ZuRo

RERODOC

Research platform Alexandria

Resolos

SciCat

sciCORE

ScientificIT

SCITAS

SENSA

SERVAL

SONAR

SwissEGA (planned)

SWISSUbase

SWITCH

SWITCHdrive

SWITCHengines

Tesi

The Dataverse

Yareta

Zenodo

ZHAW digitalcollection

ZOPAR

ZORA