Concept Paper
Research Data Infrastructures: a distinct characteristic in research infrastructures

Version of October 31, 2023. This document is designed by the StraCo to be a living draft that can be adjusted and refined to reflect the ongoing development of ideas and discussions.

The ORD National Strategy Council (StraCo) is a governance body formed by the four nationally funded education, research, and innovation (ERI) actors in Switzerland (the SNSF, the ETH-Domain, the Swiss Academies for Arts and Science, and swissuniversities).

Abstract

The StraCo proposes a conceptual framework to inform policy and strategic thinking on Research Infrastructure (RI), in light of new trends in relation to data-intensive or data-first research and Open Science. The framework considers differences in RI when they have characteristics of research facility infrastructure (RFI) and of research data infrastructure (RDI), which have important implications for strategic planning, evaluation, funding, and portfolio management. Better consideration of these differences is necessary to ensure that the development of new RI meet the needs of researchers and society, and to ensure the competitiveness of research in Switzerland.

Audience

This paper is intended for stakeholders involved in policy development, coordination, and governance of research infrastructures (RI) and for the broader research community in Switzerland and beyond.

1 Introduction 2
1.1 Purpose of the paper 2
1.2 Why are Research Data Infrastructure Important? 3
1.3 Support for RDI in the EU and Other Countries 4
1.4 RDI and the competitivity of Swiss research 5
2 Characteristics of research infrastructure 6

2.1 Archetypes of RI characteristics ... 6
2.1.1 Research Facility Infrastructure (RFI) ... 6
2.1.2 Research Data Infrastructure (RDI) ... 7
2.1.3 Coordination platforms ... 7
2.2 Comparative Overview of RI Characteristics ... 8
2.3 RI combining the characteristics of RFI and RDI ... 10
3 Integrating RDI needs in policy ... 12
3.1 Strategic planning at the level of clusters ... 12
3.2 Interface between bottom-up innovation and top-down planning... 13
3.3 Long-term funding and sustainable governance ... 13
3.4 Integration in international networking and interoperability ... 14

1 Introduction

Research Infrastructures (RI) bring together the financing, scientific resources, and expertise needed to build world-class research assets. In addition, they foster connections within research communities, thereby reinforcing their ability to tackle complex research questions, teach and train new experts, transfer knowledge to society, and engage with the public. By investing in RI, Switzerland leverages scientific advancement, supports economic development, and ensures its ability to address global challenges.

1.1 Purpose of the paper

The ORD National Strategy Council (StraCo) introduces a new conceptual framework for research infrastructure (RI) in Switzerland. It emphasises contemporary and future research trends, particularly the increasing significance of research data in science.

In this framework, RI initiatives are considered through the lens of the characteristics they can pursue: Research Data Infrastructure (RDI), Research Facility Infrastructure (RFI), and coordination platforms. These characteristics are not mutually exclusive, and RI initiatives can pursue one or several of them. This approach allows for a better integration of the specific needs of research data infrastructure (RDI)[1], which this paper focuses on.

[1] RDI are established around a resource, which is data itself, or a capacity to work with data. They include data repositories and collections, research software, analytical and visualization platforms and services, as well as statistical, AI and large language models (see definition in part 2).

The StraCo sees a need to support the development of high-quality RDI by better accounting for their specific attributes in policy instruments, funding, and institutional action. In this paper, the StraCo proposes a conceptual framework and policy considerations to better structure the support for RDI.

1.2 Why are Research Data Infrastructure Important?

Two long-term trends in research practice and culture explain the growing importance of RDI. Neither of them is new, but they have come in sharp focus in recent years.

Trend 1: Research as exploitation of data.

The **conjunction of abundant and high-quality data and the availability of powerful new data-intensive research techniques** means that more research activities can be conceived primarily around the exploitation of data. This double trend enables important new research lines and can also facilitate cost-effective research. This trend influences most disciplines, although the maturity of practices varies greatly. RDI are sometimes a necessity and almost always an accelerator. They foster more effective management of abundant data and help communities organise and attract the resources and expertise needed to provide data tools tailored to specific fields. As a result, more and more RDI are being established in disciplines where they did not previously exist, and small-scale research data practices are being transformed into full-scale RDI[2].

In climate science for instance, a plethora of small-scale data collection facilities are being networked and developed into a large-scale infrastructure: the Integrated Carbon Observation System (ICOS). It was set up in 2015 to give access to standardised open carbon data from across Europe[3] and is transforming how data-intensive research is conducted in this domain. Similar examples can be found in a wide and growing range of disciplines, from health and life sciences to the social science and humanities[4].

Trend 2: Open Science.

In parallel and not unrelated to the previous trend, the **global push for Open Science and for more stringent reproducibility requirements** have spurred the development of RDI dedicated to knowledge sharing and collaboration. The

[2] p. 111. National Academies of Sciences, Engineering, and Medicine et al. *Reproducibility and Replicability in Science.* Washington (DC): National Academies Press (US); 2019 (doi.org/10.17226/25303).
[3] Integrated Carbon Observation System (ICOS), “Our mission”, August 2023, www.icos-cp.eu/about/icos-in-nutshell/mission.

[4] *Ibid. 5.*

Findable, Accessible, Interoperable, and Reusable (FAIR) principles[5] set standards for dedicated data sharing platforms as well as existing RDI.

CERN, for example, has significantly adapted its infrastructure to better capture and accommodate computation details and data for reuse by the entire high-energy physics community. It is exemplified by services such as CERN Open Data and CERN Analysis Preservation[6]. With the increasing prevalence of data-driven science across various domains, the automatic capture of data processes and record-keeping is becoming a common practice[7].

As a result of both trends, RDI will play an increasingly important role in enabling and defining scientific excellence[8]. They are infrastructures of systemic importance, and key drivers of scientific competitiveness. RDI are enablers for collaboration, including interdisciplinary cooperation, which is essential in addressing current and future multifaceted global challenges.

1.3 Support for RDI in the EU and Other Countries

The European Union (EU) has long considered infrastructures for research data and for open science to be important assets, a position it recently reaffirmed in the *Lund declaration on Maximising the benefits of research data*[9]. It invested significantly in e-infrastructure and digital infrastructure and has policies to strengthen FAIR and open principles across its RI portfolio. As a result, RDI-type characteristics figure extensively in the ESFRI landscape, both in the form of dedicated data RI, and in the data-intensive characteristic of thematic RI[10]. In addition to continuing investments, the European Commission is setting up a new instrument called EDIC (European Digital Infrastructure Consortium) to facilitate multinational data infrastructure[11] and is investing significant resources—funding and coordination—in the European Open Science Cloud (EOSC).

---

[5] See www.go-fair.org

[6] See Fig. 1 p.144 in Chen, X., Dallmeier-Tiessen, S., Dasler, R. *et al.* Open is not enough. *Nature Physics* **15**, 113–119 (2019). doi.org/10.1038/s41567-018-0342-2.

[7] National Academies of Sciences, Engineering, and Medicine et al. (2019) *Reproducibility and Replicability in Science*, p. 111, Washington (DC): National Academies Press (US).

[8] Nature Methods 20, 471, « Data Sharing is the Future », 12 April 2023 www.nature.com/articles/s41592-023-01865-4.

[9] Swedish Presidency of the Council of the European Union, Ministry of Education and Research, Division for Research Policy, Lund Declaration on Maximising the Benefits of Research Data (2023). https://swedish-presidency.consilium.europa.eu/media/5wehfvzx/2023-06-20_eu2023_maximising-the-benefits-of-research-data_declaration.pdf

[10] ESFRI Strategy Roadmap on Research Infrastructure 2021, part 2: Landscape Analysis. https://roadmap2021.esfri.eu/media/1251/rm21-part-2.pdf

[11] https://digital-strategy.ec.europa.eu/en/policies/edic.

Many individual countries in the EU and beyond have established strategic planning bodies or infrastructure consortia dedicated to research data. Germany has set up its National Research Data Infrastructure (NFDI) in 2018[12], Australia has started working on a National Research Data Infrastructure (NRDI) Strategy in July 2023[13], China is looking into integrating its China Science and Technology Cloud (CSTC) into a Global Open Science Cloud (GOSC)[14], and the UK uses the concept of digital research infrastructure in its national RI governance[15].

The policy frameworks and instruments developed in the EU and other countries provide models that Switzerland should consider in its own RDI development.

1.4 RDI and the competitiveness of Swiss research

Swiss research and innovation benefit greatly from their international connectedness and openness. As a result, the StraCo puts a high priority on ensuring that investments in RDI in Switzerland are coherent with international activities, notably the EU framework. This was and will continue to be an important factor for the scientific and industrial competitiveness of Switzerland.

At the same time, investments in RDI must be recognised as a strategic instrument to support the global leadership of research in Switzerland. RDI support researchers' ability to shape research agendas, practices, and standards, enables them to develop cutting-edge research techniques that maximise their scientific impact, gives them influence in world-class research networks, and attracts global talent. It helps to manage undesired privatisation of RDI, and the complex geopolitical context in which academic collaboration is taking place.

---

[12] https://www.nfdi.de/association/?lang=en.

[13] Australian Government, National Research Infrastructure Advisory Group (2023). *Update from the chair* (https://www.education.gov.au/national-research-infrastructure/resources/chair-update-26-july-2023).

[14] Zhang, L.L., Li, J.H., Uhlir, P.F., et al. (2023) Research e-infrastructures for open science: The national example of CSTCloud in China. Data Intelligence 5(2), 355-369. doi: 10.1162/dint_a_00196

[15] https://www.ukri.org/what-we-do/creating-world-class-research-and-innovation-infrastructure/digital-research-infrastructure/

2 Characteristics of research infrastructure

Research infrastructures (RI) serve a community of researchers—as well as in some instances industry or public users—by providing a research resource at a scale and complexity beyond what can practically be achieved by individual researchers. This broad definition corresponds to international definitions of RI, such as the one used in the EU's Horizon 2020 framework[16].

It is important to appreciate that RI usually encompass several functions in addition to the provision of a research resource. They provide access to expertise and other support services, manage interconnection with other RI, define research standards, and provide governance service for the community formed by its users and contributors.

This definition, however, does not account for differences in the characteristics of RI that have important implications for research strategy, policy, and funding. A key distinction should be made between the characteristics associated with research data infrastructure (RDI) and those associated with research facility infrastructure (RFI). Both are defined below, although the focus of this paper is on RDI.

RDI and RFI are *archetypes* of RI, distinguished by the type of research resource they provide. They exhibit properties that differ in important aspects. While organisations operating as a single RI can exhibit the characteristics of both RFI and RDI, the RDI/RFI framework nevertheless provides an essential structure to evaluate purpose, organisation, and management[17] in a manner appropriate to their respective characteristics.

2.1 Archetypes of RI characteristics

Research Facility Infrastructure (RFI)

RFI are RI established around a tangible research resource. Examples of RI strongly defined by their RFI characteristics include particle accelerators, high-performance computing centres, or an experimental system or specialist laboratory.

A key characteristic of RFI is that the number of users is limited by the available resource, which must be allocated: beam time, computation, access to the

---

[16] "Research infrastructures are facilities, resources and services that are used by the research communities to conduct research and foster innovation in their fields". Article 2 (6) of the Regulation (EU) No 1291/2013 of 11 December 2013: `Establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014- 2020).

[17] In connection to the revision of the Swiss Roadmaps for RI, it is proposed that coordination instruments (or platforms) have characteristics that should be further distinguished from RDI/RFI. This is not further discussed in this paper, but the propositions described are compatible with this.

system or laboratory, etc. RFI serve researchers and their projects in priority: they have a direct use for the individual researcher. As a result, users typically apply competitively for access, then use the resource to produce or process the data they want.

Research Data Infrastructure (RDI)

RDI[18] are RI established around a resource, which is data itself or a capacity to work with data. Examples of RI strongly defined by their RDI characteristics include data repositories and collections, research software, analytical and visualisation platforms and services, as well as statistical, AI and large language models.

A key characteristic of RDI is that the resource is essentially unlimited, and users are not competing for access. They grow their impact and usefulness by increasing the number of users and contributors. RDI typically become more significant as datasets expand, enabling a wider range of research possibilities. Examples include clinical cohorts and social survey data collections. As a result, RDI are most successful when they emphasise community-building and operate according to Open Research Data (ORD) principles.

Coordination platforms

Coordination platforms are RI whose primary function is to focus on coordination or institutional representation (such as within international organisations or infrastructures). For instance, the Swiss institute of Particle Physics (CHIPP)'s mission is to enhance Swiss engagement in international projects and committees, coordinate research and teaching activities in the field in Switzerland and raise public awareness. In the context of this paper, we will not delve further into this RI characteristic as it is still being developed in the context of the discussions on the future Swiss Roadmap RI process and the focus is here on RDI, with RFI being used for contrast.

The table on the next page details the differences in the characteristics of RDI and RFI, while the next section further explores why some RI combine the characteristics of RDI and RFI, and how they overlap.

[18] The Swiss ORD Strategy uses the terms "ORD infrastructure and services" and "Research Data Management (RDM) infrastructure": RDI encompasses both. The terms Digital Infrastructures (DIS) or e-infrastructure are sometimes used internationally. However, the reference to "digital" or "electronic" may become confusing over time as most infrastructures have an increasingly digital component. We therefore deem the term RDI to be more appropriate.

2.2 Comparative Overview of RI Characteristics

	RFI—Research Facility Infrastructure	RDI—Research Data Infrastructure
Examples	Particle accelerators, microscopy facilities, space probes and survey vessels, experimental or translational systems, laboratories, or locations, etc.	Data repositories and platforms, analytical and visualisation platforms, research software, LLM, statistical and AI models, biobanks, open-source software communities, *diamond* publication platforms.
Type of resource	Physical research equipment or another tangible resource located in a single facility or distributed across facilities.	Data as a resource, and by extension a capacity to produce and do research on data. Data is used in the broadest sense, as the examples demonstrate.
Usage model	A limited availability resource is allocated to users, usually by requiring them to apply for it in a competitive process.	An unlimited, or low-cost, resource is available (common good). The utility of the resource grows when more users and contributors engage with it. It can also grow with increasing social interest in a research topic, irrespective of the number of users.
Beneficiaries	Directly benefits the individual research projects of the users. As a result, it is usually intended to serve researchers from participating institutions or countries.	Benefits communities or research fields. As a result, it usually operates as an open infrastructure serving researchers globally within the field.

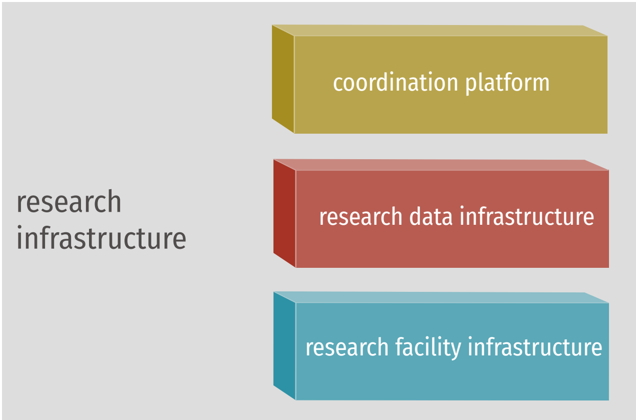
	RFI—Research Facility Infrastructure	RDI—Research Data Infrastructure
Life cycle, impact, and prioritisation	RFI are usually developed by an established community or organisation, whose research objectives and needs define the development of the instrument. Prioritisation based on project evaluation is well suited. RFIs can be assessed based on the impact of the research they will enable, accounting for future potential, and prioritised accordingly.	RDI are usually developed for the purpose of establishing a community and making it possible for new research practices to emerge. Their impact is the result of a community's eventual engagement and should be evaluated as such. It is common for RDI to originate as experiments started by a single researcher or group[19]. Data resources naturally grow over time: more data is added to a database, an analysis software is extended with more advanced features. Because RDI operate as an ecosystem more than individual RI, strategic planning and prioritisation at the level of clusters are needed.
Cost model (OPEX, operating expenditure vs. CAPEX, capital expenditure)	Driven by CAPEX during up front development or upgrades. OPEX is relatively stable once operational.	Driven by OPEX. CAPEX expended over lifetime of RDI (development of features with growth of community).
End-of-life	RFI are eventually upgraded or decommissioned.	RDI may hold patrimonial value, requiring consolidation into another RDI, operated long-term, or as an archive.

[19] For example, UniProt/SwissProt is the primary database of protein sequences in the world, managed by a consortium involving the Swiss Institute of Bioinformatics alongside major international partners. This now-essential RDI originated in 1986 as a side-project of then-PhD student Amos Bairoch.

2.3 RI combining the characteristics of RFI and RDI

Characteristics can be thought of as building blocks or components: RI initiatives can pursue single or multiple characteristics, with different importance for each[20]. A notable and important trend is for organisations operating an RFI to also operate an RDI.

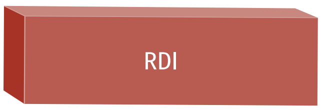
Fig 1. The different characteristics of research infrastructure



The integration of different characteristics and needs in one RI is an important factor to consider in policy. The following examples help understand how RI initiatives can combine RDI and RFI characteristics.

LiRI: a pure RDI

LiRI, the Linguistics Research Infrastructure in Switzerland, is a dedicated platform that supports linguistic research in Switzerland. It acts as a pure RDI by providing tools, resources, and collaborative spaces for researchers to collect, analyse, and share linguistic data.

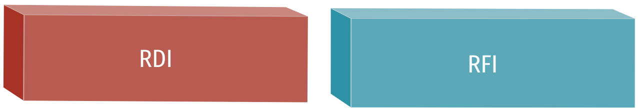


ESO: an organisation operating RFIs, with a dedicated RDI to connect them.

The European Southern Observatory (ESO) automatically deposits the data produced by the users of its various RFI such as the Very Large Telescope (VLT) in ESO's Science Archive Facility. It is then available to the community at large after an embargo period. In 2022, 24% of all ESO data papers were produced

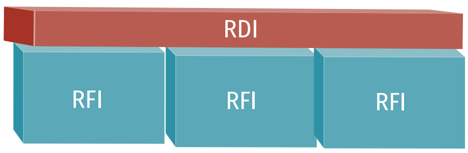
[20] In line with the focus of this document, the coordination platform characteristic is not considered here, but is often important when considering RI linked to larger organisations.

exclusively with data from the Archive—ESO's RDI—without use of the instruments themselves—the RFI components[21].



NCCR Marvel's Materials Cloud and the Swiss National Supercomputing Center: a RDI delegating tasks to an RFI.

In computationally intensive domains, RDI may not provide the facilities to process data or run analytical models as an unlimited resource. For example, simulation models available on NCCR Marvel's Materials Cloud platform require high-performance computing (HPC) resources to run. Users are expected to obtain them separately, such as from the Swiss National Supercomputing Center (CSCS). Materials Cloud and CSCS partner to ensure interoperability, so that users can easily operate across the RDI and RFI. A similar situation arises when users require personalised support, for example, data science expertise by the staff of the Swiss Data Science Center, or the time of statisticians who evaluate data in medical cohorts.



Planning the interconnection and superposition of RDI and RFI characteristics are an important policy consideration. It has a significant impact on the scientific utility of RI. Poor interconnection of RDI and RFI characteristics within or across RI initiatives can significantly hinder data exploitation and slow the uptake of scientific approaches necessary to solve global challenges, in a cost-effective manner.

---

[21] European Southern Observatory (ESO) (2023), Publication Digest, *ESO Annual Report 2022*, p.23 (https://www.eso.org/public/archives/annualreports/pdf/ar_2022.pdf).

3 Integrating RDI needs in policy

RDI and RFI characteristics entail distinct needs in terms of strategic planning, funding, and governance. Existing policy instruments that do not consider these characteristics fall short in addressing the challenges and opportunities posed by RDI, or by various models for the combination of RDI and RFI.

3.1 Strategic planning at the level of clusters

A focus on project-based evaluation and funding leads the operators of RDI to compete for funds and users, often developing separately the same resources or services. It fractures research communities and leads to inefficient use of resources. To better serve researchers, a strategic planning and funding framework for RDI should encourage the development of clusters of interconnected RDI.

While science policy instruments such as the Swiss Roadmap for RI can provide some guidance—for example it defines FORS and DASCH as the key infrastructure in social sciences and the humanities respectively—, it does not currently offer a full framework for RDI coordination and strategic planning. The development by the EU of new policy instruments, for example, demonstrates the necessity to develop innovative frameworks for this purpose.

In the prioritisation and evaluation of RDI, the following should be considered:

- Opportunities for the development of strategic RDI initiatives in communities where Switzerland already holds or aims to attain a position of leadership and excellence should be identified and supported. Additionally, RDI efforts necessary for advancing new research methods and practices, particularly those related to AI, should be prioritised.
- A notion of National Importance relevant to RDI should be used. The principle of a minimum cost is not adequate, as it prevents opportunities linked to certain emerging technologies or in certain disciplines.
- The principles of the National ORD Strategy and Action Plan, which has already received approval from the key ERI actors in Switzerland, should be fully integrated in the funding and governance frameworks used to manage the Swiss RDI portfolio.

The StraCo's Blueprint as a strategic planning instrument

The StraCo develops a strategic planning instrument, currently in prototype phase for the health and life sciences cluster. It utilises landscape analysis to craft a blueprint, i.e., strategic options for the development and coordination of the cluster. This may include development of new areas, extensions, closure, or merger of infrastructures. The blueprint is meant as a guidance instrument, providing a framework for long-term planning to which funding decisions can refer to, but it is in no case a decision-making instrument. The blueprint could provide a dedicated RDI-specific instrument within the Roadmap for RI process, or provide a parallel resource to inform the process.

3.2 Interface between bottom-up inventiveness and top-down planning

Experiments with data resources are an expression of scientific creativity and inventiveness which should be encouraged. The National ORD Action Plan includes a mechanism for the bottom-up, non-strategic development of data resources (Action Area A)[22]. However, there is no framework to transition these experiments into full RDI when their importance warrants it, or to consolidate them with existing RDI. Switzerland requires such a framework, building on a clear vision for the development of RDI.

New criteria should be established to effectively measure the value and efficiency of RDI in evaluation processes. It is crucial that these criteria align with the intrinsic characteristics outlined in section 2.2.

3.3 Long-term funding and sustainable governance

For an RDI to be internationally credible and scientifically useful for its community, it often needs to provide at least its core data resource for free. This requires long-term, sustainable funding for operating expenditure (OPEX) and a robust community governance model (a model called *diamond* in open access publications)[23]. This does not prevent the use of cost-recovery models for other

---

[22] « [The] Action Area A [aims to] support researchers and research communities in imagining and adopting ORD practices. Action Area A is driven by the demands of researchers who take a collaborative approach to research and whose overall research culture is shaped by developing and adopting novel ORD practices. As such, Action Area A is demand-driven, whereas the other action areas are geared to supporting researchers through services, infrastructure, information, training, and legal advice." (See p. 17, National Open Research Data Action Plan, November 2021, https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/ActionPlanV1.0_December_2021_def.pdf).

[23] Fuchs, C., & Sandoval, M. (2013). The Diamond Model of Open Access Publishing: Why Policy Makers, Scholars, Universities, Libraries, Labour Unions and the Publishing World Need to Take Non-

functions the RDI may provide. Flexible funding for capital expenditure (CAPEX) is also needed, to support incremental investments in the development of an RDI. It should connect to the expansion of the community and of its research practices, or to its relevance, and not to the development of the data infrastructure itself. This simultaneously incentivises RDI to engage with their users and facilitates the management of funding risk. RDI funding should encourage the consolidation of infrastructures that aim to serve the same community to avoid fragmentation. This can imply closing or merging operations of RDI.

While existing funding instruments in Switzerland can respond to the specific financial needs of RDI, there is no appropriate framework to plan, request, and coordinate the funding.

3.4 Integration in international networking and interoperability

To foster an efficient development of RDI, it is crucial to recognise that foreign users of Swiss RDI also make valuable contributions to the Swiss research system. It is important to identify Swiss RDI initiatives that compete with well-established international counterparts and discourage the further development of services that do not complement or integrate with these global networks. The use of isolated RDI initiatives should not be incentivised, as this could push Swiss researchers away from their international peers.

Commercial, Non-Profit Open Access Serious. TripleC: Communication, Capitalism & Critique, 11(2), 428–443. Retrieved from http://triplec.at/index.php/tripleC/article/view/502

The StraCo's position on the Swiss Roadmap for Research Infrastructure

The StraCo contributes to the ongoing SERI-led discussions on the adjustments to the Roadmap Process. The president of the StraCo participates in the high-level roundtable meetings convened by the Secretary of State. Experts from StraCo's coordination group contribute to the working group tasked with informing the roundtable with options for the future of the Roadmap process.

The framework developed in this paper, translates into the positions of the StraCo takes regarding the Swiss Roadmap for RI. These positions should also be considered in relation to other federal and cantonal funding and governance frameworks concerned with RI, and the multi-annual planning of institutions.

The position of the StraCo is driven by the following principles:

1. The Roadmap process should **consider the distinct specificities of RDI** in strategic planning, evaluation, and funding.
2. The Roadmap process should include a **strategic vision for RDI** in Switzerland. It should guide ERI actors and institutions and facilitate their coordination, without oversteering them. It should be developed by a **national body** involving all ERI actors.
3. **Transparency** should be a key feature of the Roadmap process, with efficient information flows among all ERI actors involved. Responsibilities should be clearly allocated to specific ERI actors, including for information collection.