10 December 2020

National Centres of Competence in Research (NCCRs):

Guidelines regarding the SNSF Open Research Data Policy in the NCCRs of the series 5

Managing and sharing research data as openly as possible is one of the principles of good scientific practice. The SNSF adheres to this principle and values research data sharing as a fundamental contribution to the impact, transparency and reproducibility of scientific research. The SNSF has therefore set out the criteria it expects funded researchers to meet in its Open Research Data Policy statement. Pursuant this policy, funded researchers are required to (i) elaborate a plan regarding the life cycle of the data to be generated by funded research and (ii) share the data underlying resulting publications onto existing and publicly accessible digital databases.

In this context, the NCCRs are expected to expose their plans in terms of data management and data lifecycle in a **Research Data Management (RDM) Strategy**, whose aim is to outline:

- The global organisation of data management within the NCCR.
- The planned life cycle of the data that will be generated, collected or used during the phase.

In addition, the NCCR researchers will be expected to fulfil the SNSF data sharing requirements, i.e. to deposit at least the data underlying their publications on existing publicly accessible repositories provided that there are no legal, ethical, copyright or other issues (more details in the "NCCR Data Sharing" section).

The best way of managing and sharing data depends on the research field and on the organisation of the NCCR. The SNSF provides general guidelines, coversheet and explanations for the elaboration of the RDM strategy. These regulations provide sufficient flexibility to adapt the strategy to each NCCR's organisation and to discipline-specific standards and practices regarding management and sharing of data.

The NCCR Leadership has the responsibility to define the data management strategy for the Centre. Every NCCR also has to designate a "Data Manager", responsible for the coordination of data management within the network, and for establishing and maintaining a "dataset index" listing all published datasets resulting from the NCCR research.

The following sections of this document provide general guidelines to the NCCRs for the elaboration of the RDM strategy and for data sharing.

NCCR Research Data Management Strategy

Content

An NCCR **RDM strategy** comprises one <u>general section</u>, and as many scientific <u>unit-specific sections</u> as there are scientific units (see definition below) in the NCCR:

A. General section – description of the global organisation of data management in the NCCR.
- o **Data manager, data management organisation and budget**
- o **Intellectual property rights and copyright**
- o **Internal data sharing policy**

B. Scientific unit-specific sections - specific information about research data management (equivalent to a project's DMP) for each scientific unit (see definition below).
- o **General information**
- o **Data collection and documentation**
- o **Ethics and security issues**
- o **Data storage and preservation**
- o **Public data sharing**

Definition of "scientific unit": Depending on the scientific organisation of the NCCR, "scientific units" can be (groups of) work packages, (sub-)projects, or a combination thereof (see figure 1).
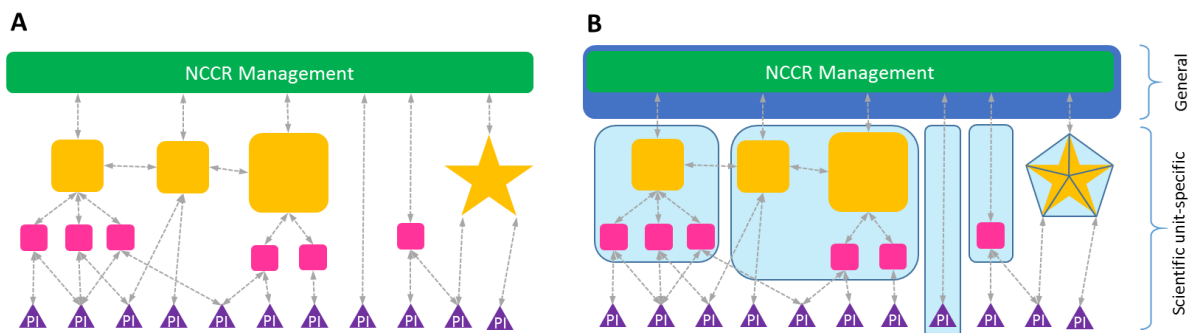

Figure 1: A. Fictive representation of the organisation of an NCCR. Yellow and pink squares and stars can represent work packages, (sub-)projects, platforms, cohorts,… Purple triangles represent PIs. B. Representation of a RDM strategy. The dark blue box represents the general section. Light blue boxes show the designated scientific units, each of which elaborate a scientific unit-specific section for the RDM strategy.

It is the responsibility of the NCCR Leadership to define the relevant scientific units. Each scientific unit must complete a "scientific unit-specific" section of the RDM strategy. Every research activity taking place in the frame of the NCCR must be part of a scientific unit. Delineation of scientific units can be driven by similarities of (sub-)projects in terms of data management rather than in terms of scientific content or organisation. It is recommended to set the scientific units at the highest possible organisational level of the NCCR (i.e. one scientific unit per projects or workpackage and not per PI). If the NCCR defines a common Data Management Strategy for the whole network, this implies only one scientific unit-specific section to be submitted. In exceptional cases, and only if the NCCR organisation requires so, single PIs can form a scientific unit.

The overall RDM strategy must not conflict with the data management / sharing policies of the different affiliating institutions / departments / institutes.

Every PI must agree with the general strategy and the strategy of the scientific unit section(s) they are involved in. In conflictual cases, scientific units must be further divided (in exceptional cases, down to the single PI level if necessary) to reach agreement.

The RDM strategy coversheet and the corresponding explanation file are available for download in NIRA.

Timing and lifetime management

NCCRs of the 5th series will submit their RDM strategy for phase 1 together with the intermediate report of year 1 (Annex 4 of IR01, see corresponding coversheet).

The strategy can be updated and sent back to the SNSF at any time. If research sub-projects start or end within the NCCR, the RDM Strategy must be adapted accordingly and re-submitted to the SNSF at the latest with the next intermediate report.

Assessment

The RDM strategy will be assessed by the SNSF Administrative Office for its plausibility and adequacy with the SNSF policy on open research data and the NCCR specific guidelines. If there are shortcomings in the submitted document, these will be communicated in the course of the formal control and must be addressed. However, the RDM Strategy is not part of the scientific/structural evaluation of the NCCR. The review panel and Members of the Research Council have access to the NCCRs' RDM strategies but will not evaluate them.

Procedural aspects

Submission of a plausible RDM strategy and of its potential required updates (see "timing and lifetime management") is a contractual obligation and is necessary for the liberation of the full yearly tranches. After RDM strategy submission, potential shortcomings will be communicated by the SNSF Office and must be adequately addressed for approval.

Publication

All PIs of the network must be informed of the final NCCR RDM strategy maximum one month after approval of the strategy by the SNSF Office, or one month after submission of a revised version.

The RDM strategy must be published on the NCCR website, latest by the end of the phase in which it was submitted to the SNSF.

NCCR Data Sharing

The NCCR researchers are expected to fulfil the SNSF data sharing requirements for the data produced in the frame of the NCCR. The SNSF expects that researchers share at **least the data underlying their publications, but only to the extent to make the published results reproducible and provided there are no legal, ethical, copyright or other issues**. This data should be shared as soon as possible, but at the latest together with the relevant scientific publication.

Best practices

To facilitate the discovery, access, re-use and citation of datasets, it is important that the publication of research data follows a set of clearly defined and broadly applicable best practices. The FAIR Data Principles define a range of qualities a published dataset should have in order to be Findable, Accessible, Interoperable and Reusable (see Explanation of the FAIR Data Principles).

The SNSF expects researchers to **share their data according to the FAIR Data Principles on existing, publicly accessible, digital repositories**. The SNSF does not expect NCCRs to create their own repositories. It is also important to note that the FAIR Data Principles do not require researchers to share all their data without any restrictions. Rather they advocate applying a standard procedure when sharing research data for reuse, so that humans and computer systems can easily find, interpret and use them under clearly defined conditions.

Datasets must always be carefully documented with associated metadata, such that other researchers understand how the data was collected, as well as under which conditions and how it can be re-used. The provided data and documentation (metadata) must be sufficient to ensure their reusability. The RDM strategy should explicitly state wherever these requirements cannot be met. For more details, please refer to the ORD section of the SNSF website.

Eligible costs

The costs of enabling access to research data that is collected, observed or generated within the NCCR are eligible for the SNSF and Home Institution funding if the research data is deposited in recognised scientific, digital data archives (data repositories) that meet the FAIR principles and do not serve any commercial purposes. It is permitted to upload data to commercial repositories, but only the data preparation costs will be eligible.

Costs resulting from data management (preparation, deposition on repository) should be entered in NIRA under the corresponding project. Salary of a potential data manager can be entered in NIRA in Management (Office).

NCCR dataset index

To facilitate the access to the data generated by the funded research, NCCRs are required to publish a "**dataset index"** on their website. In analogy to a publication list, the dataset index lists all published datasets resulting from the NCCR research and provides a valid link to it (URL).

The NCCR dataset index should catalogue at least the following information: Dataset title / Dataset persistent identifier / URL.

NCCR series 5: Research Data Management (RDM) Strategy Coversheet and Explanations

Coversheet

Title of the NCCR	
NCCR Director Name, first name Institution	
Funding Phase	
Version (Date of submission)	

A. General section

A1. Data manager, data management organisation and budget

      A1.1. Data manager

      A1.2. Internal organisation, roles and responsibilities

      A1.3. Definition of the NCCR's "scientific units"

      A1.4. Data management budget

A2. Intellectual property rights and copyright

Annex A. Internal data sharing policy

B. Scientific units-specific sections

BX. Scientific unit X

BX.1. General information

      BX.1.1. Person responsible

      BX.1.2. PIs concerned

BX.2. Data collection and documentation

      BX.2.1. Description of the data collected, observed, generated or reused

      BX.2.2. Documentation and metadata provided

      BX.2.3. Data preservation plan

BX.3. Ethics and security issues

       BX.3.1. Handling of ethical issues

       BX.3.2. Management of data access and security

BX.4. Data storage and preservation

       BX.4.1. Data storage and back-up during the research

       BX.4.2. Data preservation plan

BX.5. Public data sharing

       BX.5.1. Public data sharing location

       BX.5.2. Data sharing constraints

Explanations

General remarks

The NCCR RDM Strategy must contain:
- One general section
- One scientific unit-specific section per scientific units

submitted to the SNSF as a single document.

If the NCCR defines a common Data Management Strategy for the whole network, this implies only one scientific unit-specific section to be submitted.

Every heading (sub-section) listed in the coversheet must appear in the submitted RDM strategy. If a heading does not apply for your organisation / discipline / scientific unit, please include it and briefly explain why it cannot be filled.

A. General section

This section must be elaborated and filled by the NCCR Leadership.

A1. Data manager, data management organisation and budget

A1.1. Designate a "data manager". The data manager is responsible for the coordination of data management within the network. Indicate the %FTE that the data manager will devote to the NCCR data management.

A1.2. Outline rights/duties, internal procedures incl. quality control/self-assessment. Describe how data management will be coordinated between the groups of different institutions. Please assign responsibilities in terms of:

- Information and training of the NCCR members regarding SNSF/NCCR data management policy and requirements
- Maintenance of the data management infrastructures, data backup
- Preparation, curation, and documentation of datasets
- Timely submission of datasets on repositories
- Maintenance and update of the NCCR Dataset Index

The tasks can be distributed between the data manager, members of the network and external services. If more than one person is designated for a task, the individual responsibilities should be clearly delineated. The data manager is responsible to supervise the execution and overall fulfilment of the SNSF ORD policy requirements.

A1.3. Describe the scientific units of the NCCR used as basis for data management. The description can be provided as a scheme or as a table (see Table 1).

Depending on the scientific organisation of the NCCR, "scientific units" can be (groups of) work packages, (sub-)projects or a combination thereof. Delineation of scientific units can be dictated by data management coherence (i.e. a set of (sub-)project having the same data management strategy / procedures) rather than scientific content or organisation. All research activity performed in the frame of the NCCR must be part of a scientific unit (even if it is not generating or reusing data). The scientific units must be numbered and a "scientific unit-specific section" of the RDM strategy must be provided for each of them. It is recommended to set the scientific units at the highest possible organisational level of the

NCCR (i.e. one scientific unit per projects or work packages and not per PI). In exceptional cases, and if the NCCR organisation requires so, single PIs can form a scientific unit. In case of questions regarding the definition of the scientific units for your NCCR, please contact the SNSF NCCR Office. For each unit, a responsible person is designated for the elaboration of the document. This can be the data manager or other members of the NCCR.

Table 1: Description of the NCCR scientific units

Scientific Unit No	(Sub-)project / work package / platform,… name / number	Responsible*
1	e.g. project 1	
2	e.g. subproject 2.4, 2.5, 4.2 and 5.11	

* Person responsible for the submission of the corresponding "scientific unit-specific section"

A1.4.   Provide the approximate budget planned by the NCCR for data management for each year of the phase. Please categorize between salaries, data preparation and data reposition. Eligible costs are described in the NCCR Budget Guidelines.

A2. Intellectual property rights and copyright

A2.1.   Outline the owners of the copyright and Intellectual Property Right (IPR) of the data collected and generated in the frame of the NCCR, including the license(s). Indicate any IPR ownership agreement for the consortium.

Annex A. Internal data sharing policy

Provide the NCCR's internal data sharing policy as an annex. The internal data sharing policy should describe the access of the NCCR fellows to the data produced in the frame of the NCCR at any time (i.e. before the data is made public, between data collection and dataset submission). The policy should clearly define who has access to what data, at what time point and how. The treatment of sensitive data in this context should be addressed.

B. Scientific units-specific sections

Every scientific unit listed in A1.3. must provide a "scientific unit-specific section" (e.g. if the NCCR comprises 3 scientific units, the RDM Strategy must contain 3 scientific unit-specific sections). The scientific unit-specific sections are similar to the Data Management Plans (DMP) required with the submission of proposals for "Projects" to the SNSF.

The data management strategy and information provided in these sections must be approved by all the PIs concerned.

Some scientific units may not produce or reuse any data. For these cases, please complete part BX.1.1 and BX.2.1. only.

BX.1.[1] General information

BX.1.1. Name of the person responsible for the setup of this scientific unit-specific section (as listed in Table 1).

BX.1.2. List all the PIs concerned by this scientific unit-specific section (i.e. the PIs participating in work packages and/or (sub-)projects comprised in this scientific unit).

---

[1] For subchapter numbering, please replace the « X » by the number of the corresponding scientific unit as described in Table 1 (e.g. for scientific unit No 2, « BX.1. » should be « B2.1.»)

BX.2. Data collection and documentation

BX.2.1. Briefly describe the data collected, observed, generated or reused. The descriptions should include the type, format and content. If no data generation, collection or reuse is expected in this scientific unit, please explain why.

BX.2.2. Describe the documentation that will be provided to enable secondary users to understand and reuse the data. Metadata include at least a name and a persistent identifier for each file, the name of the person who collected or contributed to the data, the date of collection and the conditions to access the data. If specific tools are needed to re-use the data, this needs to be documented and, if possible, the tools made available.

BX.3. Ethics and security issues

BX.3.1. Describe potential ethical issues related to the research performed or the data collected within the scientific unit. Describe how these will be addressed or handled. Methods to manage ethical concerns may include: anonymization of data; gain approval by ethics committees; formal consent agreements,…

BX.3.2. Describe potential personal or other sensitive data to be collected within the scientific unit and the corresponding level of risks. If such data will be collected, outline the security measures for data protection and how data access and security will be managed. Describe the main processes or facilities for storage and processing of personal or other sensitive data.

BX.4. Data storage and preservation

BX.4.1. Describe how/where the data will be stored during research (e.g. dedicated server, institutional IT service, laptops,…). Describe the back-up procedures (frequency of updates, responsibilities, automatic/manual process, security measures, etc.). If different groups involved in the scientific unit use different storage strategies, please list accordingly.

BX.4.2. Specify which data will be retained, shared and archived after the completion of the NCCR and the corresponding data selection criteria (e.g. long-term value, potential value for reuse, obligations to destroy some data, etc.). Please outline a long-term preservation plan for the datasets beyond the lifetime of the NCCR. Comment on the choice of file formats and the use of community standards. If different groups involved in the scientific unit use different strategies, please list accordingly.

BX.5. Public data sharing

BX.5.1. Define on which repository the data will be made available. The repository chosen **must be conform to the FAIR Data Principles** and must be maintained by a non-profit organisation. If these conditions cannot be fulfilled, please explain why. If different groups involved in the scientific unit use different repositories, please list accordingly. Note: the NCCRs are not expected to create their own repositories, but rather to use existing structures.

BX.5.2. Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output. If some data cannot fulfil this principle due to legal, ethical, copyright, confidentiality or other issues, please list them and explain why. Describe under which conditions these will be made available (timing of release,…).