NCCR Robotics Research Data Management Strategy: A Workflow Application

Valeria Di Cola
NCCR Robotics - EPFL
Lausanne, Switzerland
valeria.dicola@epfl.ch
0000-0003-2925-8366

Abstract— *In compliance with the SNSF requirements, a Research Data Management strategy has been developed for NCCR Robotics in synergy with the EPFL Library, and the EPFL Legal and Ethics Departments. The strategy includes guidelines and a workflow for making research data related to NCCR Robotics journal publications and proceedings publicly available, including the special cases of sensitive and protected data.*
The article focuses on the coaching program that was established to support NCCR members in the implementation of this strategy. Concrete resources and tools – such as tailored presentations answering the specific needs of each laboratory – have been developed in order to offer practical solutions, focusing on three main aspects of research data management (RDM): (1) Data lifecycle and FAIR principles; (2) A proposed practice within NCCR Robotics, with emphasis on Zenodo as a recommended data repository; and (3) General Data Management Best Practices.

Keywords—RDM Strategy, FAIR, data repository, workflow.

I. INTRODUCTION

The National Centre of Competence in Research (NCCR) Robotics is a Swiss nationwide organization funded by the Swiss National Science Foundation, bringing together more than 100 top researchers from all over the country with the objective of developing new, human-oriented robotic technology for improving quality of life. The Centre binds together experts from seven world-class research institutions; École Polytechnique Fédérale de Lausanne (EPFL) (leading house), Eidgenössische Technische Hochschule Zürich (ETH Zurich) (co-leading house), Universität Zürich (UZH), Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA) Lugano, University of Bern (UNIBE), the Swiss Federal Laboratories for Materials Science and Technology (Empa) and the University of Basel (UNIBAS). In addition to standard research data collected, the research units that are part of the consortium produce sensitive personal data as well as protected data, which may include - but is not limited to - information relating to data received by a third party under confidentiality or a specific agreement; data being subject to be protected via a patent or any other intellectual property title or subject to be licensed to a third party for commercial purposes (e.g. start-up).

Studies have shown that keeping research data freely available is crucial for open science — and the research funding could depend on it (Schiermeier, 2018). In 2019 we developed a Research Data Management Strategy (RDMS) as part of our contractual obligation with SNSF in close collaboration with related offices of the home institutions (EPFL, ETHZ and UZH Libraries, EPFL Legal and Ethics Departments).

The RDMS aims at contributing to recognize research data as valuable academic resources that need to be managed, shared and preserved to foster research and science. It provides relevant tools and guidance to better manage the data that our scientists handle throughout its whole life-cycle, from the planning stage of the project to the long-term preservation strategies. Hence, contribute to making FAIR Science a reality as well as help researchers be more productive for themselves and their collaborators (ELIXIR, 2021).

The strategy includes guidelines and a workflow for making research data related to NCCR Robotics journal publication and proceedings publicly available, including the special cases of sensitive and protected data. Here, we present the workflow application of the implementation of the RDMS over the 28 Laboratories that are part of our Centre of competence.

II. General section

A. Internal Organisation, Roles and Responsibilities

All information and training activities for NCCR members regarding SNSF/NCCR data management policy and requirements are managed by the NCCR Robotics RDM Officer, supported by the EPFL Library.

Maintenance of the data management infrastructures, data backup and preparation, curation and documentation of datasets, submission of datasets on repositories are under the responsibility of each individual laboratory.

B. General Strategy for Data Storage

Each laboratory defines its own data storage and preservation strategy based on its needs and safety concerns. Support is provided by the IT services related to the laboratory, that install and support the data storage of each laboratory according to its needs.

However, these common principles must be followed:

• The data (or at least a copy) must be stored on the institutional hard-drive; cloud-based storages (such as Google Drive, Dropbox, or Switch Drive) are not considered institutional hard-drives.

• In case of sensitive data, the storage must guarantee that the data are well protected and only authorized people can have access to them.

C. Public Data Sharing

Only data for journal publications and proceedings that resulted from activities that have been funded through NCCR Robotics are concerned by this strategy.

NCCR Robotics recommends the use of Zenodo (zenodo.org) , a data repository that respects the FAIR Data Principles and is maintained by a non-profit organisation (CERN). Every publication made on Zenodo is publicly available and has its own DOI for unique identification and citation.

In case a NCCR member does not want to use Zenodo, they have the freedom to use another data repository. In this case they have to make sure the chosen data repository respects the SNSF requirements (it must respect the FAIR Data Principles and must be maintained by a non-profit organisation), and also has to communicate manually the DOI of the data publication to the NCCR Robotics RDM Officer.

The workflow application of the RDMS (with Zenodo, or with another data repository) is summarized in Fig. 1 in order to help NCCR Robotics members with the data publication. The cases of sensitive and protected data are also included in the workflow.

Within every published dataset, a README text file should be included that contains at least the following meta-data (as required by the SNSF FAIR Principle) :

- a title (a name given to the dataset or the research project that produced it);
- an abstract of the project (description);
- creator (the name of the person who collected or contributed to the data) identified by ORCID;
- the date of collection;
- a short description of each file;
- a persistent identifier (ISBN, or DOI);
- the license;
- If unique tools or proprietary software are used, this will also be documented in the metadata when appropriate. If possible, the tools or links to the provider will be made publicly available.

In the case of developed code or scripts, they are considered as data and should be also published:

- the programming language used
- the versions used (libraries, compiler, packages, etc.)
- the machine used
- the license of the code

In case the data has no license already attached and there is no claim by any third party on such data, NCCR Robotics recommends the application of the CC-BY license to the published data. The CC-BY license authorizes the access, usage (commercial or not) and the modification of the published data but requires the citation of the author by the user of the data.

III.   CONCLUSION

It is clear that keeping research data available is essential for open science. Not only does it ensure transparency and reproducibility, increasing data visibility and number of citations, but fulfilment of funders' requirements. Following best research practices leads to saving time and avoiding risks of data loss. Lastly, comprehensive RDM allows our researchers not only to produce new knowledge and make more discoveries just by re-using data, but also to archive, retrieve and re-use their own data. Nowadays, we are witnessing an increase in demand and use of big data, so having an up-to-date RDMS is part of maintaining modernity with the world scale digital research.

With our RDMS we took the first steps towards a paradigm shift in providing FAIR RDM together with training and education for our consortium on this topic. The strategy was implemented virtually in 2020 during 4 months, over 28 Labs through one-to-one trainings) and reached more than 100 NCCR Robotics researchers.

A. Final recommendations

- Share any data that is relevant for re-use.

- Data underlying publications must be made available at the time of the publication.

- Wider data can be made available after the project ends.

ACKNOWLEDGMENT

REFERENCES

ELIXIR (2021) Research Data Management Kit. A deliverable from the EU-funded ELIXIR-CONVERGE project (grant agreement 871075). URL: https://rdmkit.elixir-europe.org

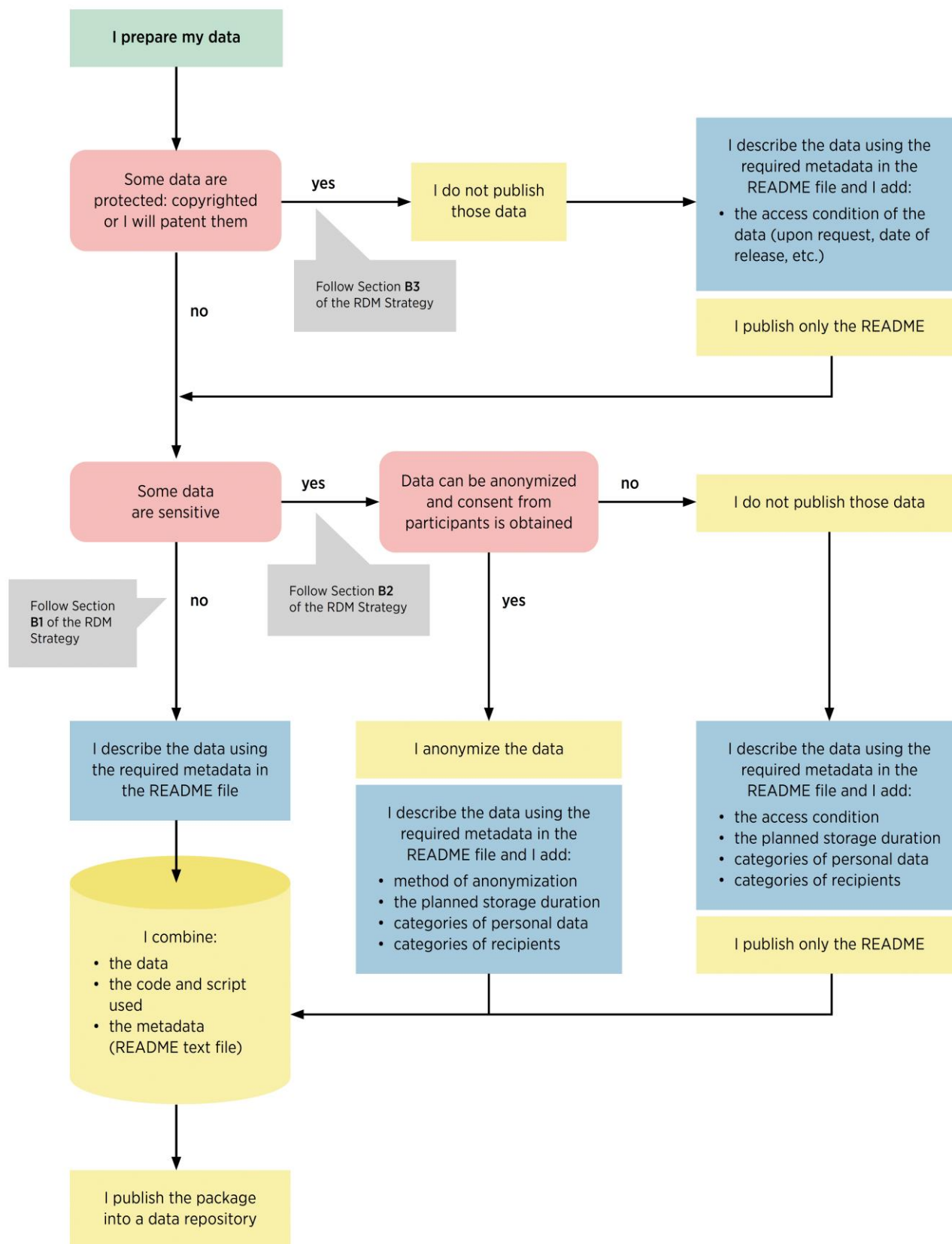Schiermeier, Q. (2018). Data management made simple. *Nature, 555* (7696).

Fig. 1. Workflow application of the NCCR Robotics RDMS