Beginner's Guide for Data Stewards

Authors

Dagmar Hanzlíková (Charles University)

https://orcid.org/0000-0001-9287-399X

Matyáš Hiřman (Charles University)

https://orcid.org/0000-0003-0153-9414

This manual is based on an existing document at Charles University. The manual went through several rounds of reviews by experts from the Czech data steward community.

This is a living document that is expected to be updated regularly.

Current version: **1.0**

Changelog:

Introduction

Did you take a position in charge of research data management (RDM), or did you get the role of data steward (DS)? The following guide can help you identify key areas to focus on and create a plan to get started.

General tips for an efficient start

- **Get inspired**: do detailed research on how research data management support is handled at other institutions (both in the Czech Republic and abroad), get inspired and establish cooperation.
- **Involve colleagues**: collaborating with other data stewards, support and researchers will help you learn best practices.
- **Work gradually**: complex issues cannot be implemented at once. Focus on specific areas and proceed one step at a time. If you hit an impenetrable wall, focus elsewhere for a while and then return to it. Some things take time.
- **Stay informed**: keep up to date with open science (OS) news through conferences, webinars and communities.
- **Make an overview of existing OS policies**: it's good to be aware of OS/RDM policies at international, national and institutional levels. The best way to do this is to create and update a document that can be consulted at any time.

1. Where to start

If you are new to the topic, it is a good idea to start with a basic overview of open science, specifically research data management (RDM). Below you will find links to several resources where you can get such a basic overview (see also Appendix 1).

Introduction to the role of a data steward: eosc.cz/data-stewards

Data steward **online** courses:
- free online course for DS beginners: eosc-synergy.eu
- free online course for data stewards/curators: mantra.ed.ac.uk
- free online course on data management with certificate option: coursera.org

Data steward **attendance** courses:
- certified course at Charles University: uisk.ff.cuni.cz
- international certificate course at Uni Wien, Austria: univie.ac.at
- international certificate course at Tampere University, Finland: tuni.fi

Other courses focused on **OS** and **RDM**:
- fosteropenscience.eu – e.g., Introduction to open science, Managing and sharing research data, Data protection and ethics, Assessing the FAIRness of data and more
- openplato.eu
- openaire.eu
- nasa.github.io/transform-to-open-science

Open Science **Handbook**: handbook.gitbook.io/book
General RDM **Starter Kit**: go-fair.org
Handbook Turning Way: the-turning-way.netlify.app

General course on **FAIR** principles: howtofair.dk

CESSDA Data management expert guide: dmeg.cessda.eu

- a guide to research data management for the **social sciences**

RDMkit from Elixir Infrastructure: rdmkit.elixir-europe.org

- a research data management guide for the **life sciences**

RDM document from NLT: techlib.cz/research-data-management

FAIR principles document from the Czech community of DS: TBD

Recordings of **webinars and online lectures**:

- 4EU+ lecture series recordings: 4euplus.eu
- Open science training records: eosc.cz/training
- recordings of lectures from Open Science Week 2023: YouTube
- recordings of lectures from Open Science Week 2024: TBD

Websites of open science support **centers**: e.g., openscience.cuni.cz, openscience.upol.cz, openscience.lib.cas.cz

---

Familiarization with **the tools** (see Appendix 1 for more):

- National Open Science Platform:
  - link TBD
  - link to dictionary TBD
- Data Stewardship Wizard (**DSW**) – a tool for creating research data management plans (DMPs): ds-wizard.org
- **DMPonline** – other DMP creation tool: dmponline.dcc.ac.uk
- Open Science Framework (**OSF**) – open sharing platform: osf.io
- Data **repository** registry: re3data.org
- **FAIRsharing** – database of data and metadata standards: fairsharing.org
- Humanities.tools – database of **tools and resources for the humanities**: humanities.tools
- Electronic Laboratory Notebook (**ELN**) database: eln-finder.de

---

Engaging with **communities**:

- meet with **other DS and their agenda in the Czech Republic**: eosc.cz/data-stewards
- **AKVŠ** OS working group and how to get involved: akvs.cz/pracovni-skupina
- **implementation of the** European Open Science Cloud (**EOSC**) in our country:
  - joining **working groups**: eosc.cz/working-groups

TIP: Investigate **whether there are any other communities** in your field, both nationally and internationally, focused on research data management or OS issues. Such communities may also be emerging directly at your institution. If you discover such communities, **get involved** and actively engage in them. **Bottom-up** support is essential for the implementation of new topics and will help you, among other things, to establish contacts and close cooperation.

2. Getting to know the institution

After getting a basic overview of open science and research data management, it is important to orient yourself in your working environment. The insights you gain during this phase will help you better **understand the specific needs and challenges of researchers at your institution**.

To start with, find out how the RDM agenda has been handled **at your institution** so far and what **support is available to** researchers in managing research data. Contact people who may have provided such support or have dealt with RDM at your institution. Don't forget about your compiled **inventory of existing open science policies**.

Do your research on which departments **provide support to researchers** and what type of support they offer, for example:

- **Grant/project support**: find out about grant application processes and deadlines, get contacts to key people in the grant departments.
- **IT support and Core facilities**: find out what technologies and tools are available for data processing and storage, and what technical support options are available e.g. data storage, repositories, DMP tools, ELN software etc. Find out if there is also a cyber/security office at the institution.
- **Legal support**: learn about the rules and regulations on data protection, copyright and research data licensing. Find out who oversees these parts – e.g. the Data Protection Officer.
- **Education of PhD students**: connect with people who provide education for PhD students (and possibly others) at your institution and establish collaboration with them. Education at an early stage of research is crucial to integrate data management the research process.
- Get to know the departments, inform them about **the existence of your position** and the support you offer.
- Find out about **other** support **options** at your institution, e.g. for depositing and data processing at a central level or at a faculty or institute.
- Get to know the people who oversee **the open access** publishing **agenda** at your institution – you will often work together on open access publishing and data issues. Arrange meetings with researchers together.

Meet the **researchers** within your institution:

- Determine the **current state of RDM knowledge at the institution** among researchers:
  - Find out if **there are surveys into RDM** have been conducted at the institution or try to organize one.
  - Find out if **any training** on OS or RDM related topics **has been conducted** at the institution so far (contact trainers/get participation/feedback, etc.)
  - Along with grant/project support**, identify researchers who might have experience with RDM and arrange meetings with them**.
  - Try to arrange and organize **an event/education/seminar to promote RDM** at the institution and thereby get people interested in the topic at the institution.
  - Familiarize yourself with the way **PhD students are trained** and the possibilities to support their knowledge in RDM – e.g., whether any training in this area has already been/is ongoing
- **Actively** seek out **contacts** with researchers - depending on the type of institution, follow up with all researchers and research fields, even if you think they have no data.
  - **Find out and analyze** what type of data they work with, how they collect it, where they store it and what they do with it.

- o For meetings, you can use **the interview structure in** Appendix 2.
- o During the interviews, try to **help** the researchers **and inform** them about the support opportunities offered by your institution.
- o **Identify potential problems** (e.g., working with large volumes of data, personal data, ethical issues, access issues for large teams). Explain them to the researcher.

TIP: Keep careful **records of** all your findings **and document them in detail**. Such a comprehensive document is very valuable not only for you, but also for other data stewards or colleagues dealing with research data management issues at your institution.

3. More advanced knowledge of the institution

You have already gained an overview of open science and support opportunities at your institution. Let's start to look deeper into the relevant subject specifics that would be applicable.

Search for **ready-made DMPs in the fields that are** found at your institution:
- **LIBER** catalogue: libereurope.eu
- **DCC** directory: dcc.ac.uk
- published plans in **DMPonline**: dmponline.dcc.ac.uk
- **Cordis** (EU funded projects): cordis.europa.eu
- **Zenodo**: zenodo.org can be used to find already revised and commented DMPs

Explore what **repositories** researchers use and what the options are:
- information from interviews with scientists, questionnaires, or use of **portal**: re3data.org
- description and comparison of **general repositories**: harvard.edu

Find out if and what **domain specific standards** are used at your institution:
- information from interviews with scientists, or, e.g., via FAIRsharing.org

Analyze the usability of available **services** at **international** level:
- EU Node: open-science-cloud.ec
- Galaxy: galaxyproject.org
- and many others, see courses above or Appendix 1

Analyze the usability of available **services** at **national** level:
- services offered by the national data infrastructure: eosc.cz
- CESNET services: cesnet.cz
- services related to persistent identifiers (PIDs): identifikatory.cz
- National Data Repository: data.narodni-repozitar.cz

Finally, analyze services **beyond institutional policies**, established through involvement in service grants, transnational grants or as part of involvement in the **Large Research Infrastructure**: vyzkumne-infrastruktury.cz

TIP: To establish long-term cooperation and to get information from academics at the institution, you can try to organize regular community events (e.g., following the example of UST) and invite various experts to help you introduce different subjects at the institution. Alternatively, you can try to **organize an introductory training** and invite subject matter experts to help you introduce the issues at the institution. Use Appendix 3 for a training design.

4. Active project support

One of the essential parts of your job will be to **assist researchers in applying for research funding** and helping to deal with funders' requirements on research data management.

Strengthen cooperation with **project** support departments:
- Find out about **project calls that are** relevant to your institution.
- Get in touch with key people in the **project department**.
- Arrange for applicants to **be informed** about your position and the support you offer.
- You are not expected to write parts of the application or feasibility study in place of the researchers. However, you should be able to **advise** them on what should appear in the sections and point them to **relevant sources**.
- **Be on hand** if support applicants are dealing with issues such as where to store their data. You should make them aware of the options offered by the university or faculty, let them know which repositories are secure, etc.

Familiarize yourself **with the terms and conditions of the funders** you will be dealing with:
- Keep an eye on the **current terms and open science requirements** of providers. In some cases, conditions may change, and it is important to keep up to date.
- Make sure this information **is up to date** at your point of contact (e.g., website).
- With the help of the project office, find out **how many and which projects are currently in progress at the institution**, what **training sessions** are taking place **for future calls** and try to get involved with information on OS and RDM.

Help researchers with the creation of the **DMP**:
- It is important to explain **what the DMP is for** and why they should prepare it – e.g., it is a living document, so the information in it may change during the project, its basic purpose is to help the researchers to clarify issues regarding research data management.
- Inform researchers about **what the DMP should contain**, templates and tools that can be used. Help them explain concepts that are unfamiliar to them (e.g., FAIR data) and review the finished DMP. Suggest options for modifications or improvements if needed. You are not expected to create the DMP for them.
- If you identify problems in the DMP (e.g., the researcher states in the DMP that data will not be shared even though the provider's terms and conditions require it), bring this to their attention or **help** them **formulate an adequate justification**.

Support in post-application/DMP projects (depends on the definition of your position):
- Support for data management **throughout the data lifecycle**
- Process **automation** and communication with the whole project team
- Regular **DMP update and review**
- Recommendations for **domain specific and general repositories** for research data
- Support for **data storage** itself

TIP: Secure early **contacts** through the project support department **for principal investigators** who have funded projects where they must submit a DMP and proactively **approach** them **in advance** with an offer of support. This will ensure you have enough time to reach everyone.

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

National Repository Platfrom for Research Data (NRP)
CESNET, association of legal entities
Correspondence address: Generála Píky 430/26, 160 00, Prague
info@eosc.cz; www.eosc.cz

5

Appendix 1

List of interesting links on OS, RDM and other topics

Centre for Open Science and Open Science Framework: open scholarship environment
CODATA: committee on data for International Science Council
Committee on Publication Ethics: training and community
Community of Practice: discussion and networking for training coordinators
CoreTrustSeal: trusted digital repository certification
Creative Commons: open licenses for research objects
DOAJ: directory of Open Access journals
DORA: researcher assessment matrix
ELIXIR COMMUNITY: Life science RDM community
EOSC
EU Open Science: policy and resources
EUDAT: collaborative data infrastructure
F-UJI: tool for assessing FAIRness of data objects
FAIRAware: decision tool for FAIR data practices
FAIRsFAIR: project outputs with practical solutions for implementing FAIR research practices, including Open Science policy checklist
Force11: community for research communication and e-scholarship
FORRT: Training repository on Open and responsible research
GDPR decision tree
GEANT: networks and associated research infrastructures
goFAIR: FAIR community
IUPAC FAIR Cookbook: cookbook for machine-automatable data management
KNOWLEDGE CLIPS: short introductory videos about data management
National Research and Education Networks (NRENs)
Open Access directory: guide to Open Access
Open Data handbook
Open Science career assessment matrix
Open Science MOOC
Open Science training handbook
OpenAIRE: open access infrastructure for research - tools and services
OpenDOAR: directory of Open repositories
ORION Open Science: researcher checklist
OS CLIPS UGENT: short videos with tips and guidance for OS
Plan S: Open Access strategy for funding and publishing
RDA: Research Data Alliance community
The Carpentries: training on Free and Open-Source Software and data science sills
ThinkCheckSubmit: tools for research integrity
Turing Way: resources for reproducible, ethical and collaborative data science
Turning FAIR into reality: report
UNESCO Open Science: report and resources
Zooniverse: citizen science platform
5-star Open Data: guide to Open Data

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

National Repository Platfrom for Research Data (NRP)
CESNET, association of legal entities
Correspondence address: Generála Píky 430/26, 160 00, Prague
info@eosc.cz; www.eosc.cz

6

Appendix 2 – adapt the structure of the interview with the researchers to your own needs.

Suggested structure of the interview with researchers

Conducting the interview:

Ask the questions in broader circles (green), let the scientist talk and ask for details (*in italics*). Below are a few points that can help to have a successful conversation:

- It is important to find **common interest** with the researcher.
- It is possible that the researcher **will not be able to answer** some, perhaps even most, of the questions because he or she will not be familiar with the concept of OS and will generally find it difficult to identify with the need to share his or her data.
- The interview should be conducted **with interest in the work and research of** the researcher being interviewed and not that you've come to lecture them on what to do.
- OS issues **should not be presented as dogma** but should help to show the potential path that science is taking and how to incorporate its elements into a given project.
- It is a good idea **to actively listen to** the researcher, **lighten the** conversation **with humor** and calm non-verbal communication.

Introduction: Tell us about your research, what you do. Using an example of specific research you have conducted, describe (briefly) the entire research process of you as an individual/team.

a. *Explain to the respondent what we mean by scientific data.*
b. *Do you work alone or in a team?*
c. *Do you conduct research with grant support? Which grants?*
d. *Is the process described above in any way formalized and generally applicable within the domain? Or is it just a set of "unwritten rules and procedures" that are always adapted to the specific research/project at hand?*

PART I: Research data life cycle

1. Project planning: In the planning phase of a research project, do you also address data handling?

a. *At the time of planning, do you think about whether you have all the necessary tools and expertise (not only) for working with data?*
b. *Do you know what a DMP is? Do you use DMP?*

2. Data collection: What data do you use/collect in your research?

a. *Are you using existing data (e.g., from an existing database, data from other researchers, your own data from previous research)? Where and how do you collect it?*
    i. *Is the data you use subject to any license/will it be possible to share your resulting data that is based on licensed third-party data?*
b. *What volume and type of data do you use/generate (e.g., text, tables, images, videos...)?*
c. *Do you address any ethical issues in the collection and subsequent processing of data (e.g., ethics committee approval for data collection on human participants, informed consent from participants...)?*
    i. *Does the informed consent include consent to disclose/share data?*

3. Data processing: How do you process the data? Specifically, we are interested in:

a. *Who do you share data with during the research and how?*
b. *Where do you store the data during the research? Do you have backups?*

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

National Repository Platfrom for Research Data (NRP)
CESNET, association of legal entities
Correspondence address: Generála Píky 430/26, 160 00, Prague
info@eosc.cz; www.eosc.cz

7

    c. *Do you have a system within the group/for yourself for organizing data (e.g., file naming, directory organization, data organization tools)?*

    d. *Do you document your data? In what way? Are there standardized formats in your domain? What information do you consider important for good, complete and understandable data documentation?*

    e. *Do you have assigned responsibilities within the team as to(i.e., who is responsible for what)*

    f. *Do you deal with sensitive / personal data in your work? How?*

4. Data analysis: How do you analyze your data?

    a. *Do you use any special tools/services for data analysis?*

5. Data preservation: How do you handle the data after the project ends? Do you keep it?

    a. *How do you select data for long-term preservation?*
    b. *How long do you plan to keep the data after the end of the project?*
    c. *How/where do you store the data?*
    d. *If there was an institutional repository for long-term preservation of data, would you use it or would you prefer to use other options?*
        i. *Why yes / no*

6. Data sharing: Do you share your data with anyone? How do you share your data?

    a. *Who do you share your research data with and under what conditions? How are these conditions communicated? Why yes/no?*
    b. *When do you share your data (e.g., during the research, after the research is over...)?*
    c. *How do you share your data (e.g., repository, upon request...)?*
    d. *Do you share documentation along with the data? Do you store/share the documentation with the data or separately?*
        i. *Why yes / no?*
    e. *Do you consider any legal issues (e.g., licensing) when sharing data?*
    f. *Have you ever encountered any external requests for data sharing (e.g., from funders, publishers...)?*

7. Cost: Does working with data cost you any money? How do you cover these costs?

PART II: Support

8. Have you ever encountered a problem during your work with data that you needed help with?

    a. *Who did you go to for advice/help?*
    b. *Who gave you the advice/help in the end?*
    c. *Are you missing a tool or service for working with data?*
    d. *Would you like to see the institution become more involved in this issue? In what way?*

Areas that could be affected: DMP; Data collection, search for existing datasets; Use of internal service or core facilities; Use of open access core facilities (e.g., Large Research Infrastructures, or ERIC); Data documentation; Data cleaning; Analysis (qualitative or quantitative), visualization; Finding a suitable repository for long-term data preservation; Advice on which data to preserve at all and how to handle common operational documents (e.g., Presentations, informed consents, etc.), if the project also includes teaching materials or conference papers in the form of presentations, how to handle these; Meeting the requirements of facilities , funders or publishers; Ethical and legal issues (collection, processing, sharing and storage of data), Other Open Science practices (citizen science, pre-registration, open peer review, etc.).

Appendix 3 – guideline for training design; resource and detailed guide: EOSC Synergy

Worksheet A. Initial training analysis

Goals and audience

Your title and broad topics *Include a short descriptive title. Why do learners need to know about your topic/service? What does it offer?*	
Training goal *Why are you doing this training? What would success look like? This can include developing skills, gaining knowledge and changing attitude, improving service awareness/use etc.*	
Audience *Who is your target audience? This could be roles (e.g., researcher, students), knowledge level (e.g., aimed at beginners) or task based - e.g., those who want to be able to do a specific task.*	
Benefits/outcomes for learners *What will learners gain from your training? What will they be able to do? What will they know? It is helpful to use the phrase 'By the end of this training you will: Be able to…. Be familiar with… Have practiced…*	
Pre-requisites *Is there any knowledge that is useful / required? Any other courses that it would be useful to complete beforehand?*	

Content and resources

Content *This should be a rough outline at this stage - you will think about this in more detail in the Design stage. What topics will you cover? How many sessions and what length?*	
Delivery method *Live sessions or self-learning? A mix? You may not have a choice in some of these areas depending on the constraints you have.*	
Timescale *When do you plan to deliver the training? Allow enough time to develop and test content.*	
Who will develop/deliver? *Do you have the right skills, or will you need support? Where will you get support?*	
Software, tools, infrastructure required *This might be to deliver your course (e.g., a webinar tool) or for learner activities (e.g., collaboration, live coding). Note that you may decide to use additional tools once you've designed your activities.*	
Materials to re-use *What are the materials you will re-use? What is their format? videos, slides, documentation etc. You can save time by reusing other materials. How will you keep track of this to give credit later?*	
Re-use of your material *Do you want others to re-use your material? How? Material can be shared in different stages/shape and may need to be in particular formats depending on where it will be reused.*	

Worksheet B. Topics and timelines

<table>
<tr><td colspan="3">Topic 1:</td></tr>
<tr><td>Learning activities</td><td>Duration</td><td>Assessment (if applicable)</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td colspan="3">Topic 2:</td></tr>
<tr><td>Learning activities</td><td>Duration</td><td>Assessment (if applicable)</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

Add more topics if needed….

Worksheet C. Reminder and analysis

Write your learning outcomes:

Write review of Topics and Learning Activities:

10