

Training Workshop on Research Data Management in (Bio-)Medicine

Justine Vandendorpe and Birte Lindstädt,
ZB MED - Information Centre for Life Sciences, Gleueler Str. 60,
50931 Cologne, Germany

DOI: 10.4126/FRL01-006399944

Release date: 2023-02-10

This is a report about the Training Workshop on Research Data Management in (Bio-)Medicine that was designed by the ZB MED - Information Centre for Life Sciences. This report contains information on the training concept such as the content and how it is organised. It also contains information on the collaborations that were established and the training workshop editions that were already given.

Concept

This training workshop is intended for an audience of all-qualification-level (bio-)medical researchers who are interested in being introduced to best practices in Research Data Management (RDM). It was designed by the ZB MED - Information Centre for Life Sciences, acting as a NFDI4Health and NFDI4Microbiota partner.

The workshop is intended to be the result of a collaboration between ZB MED and a local RDM centre from a university with a faculty of medicine, public health or similar. Such centres - referred to as partners in this report - usually support the researchers of their institutions with general advice and with local infrastructure and services. In the context of this workshop, partners thus cover general aspects of RDM and introduce their local infrastructure and services. ZB MED complements this with (bio-)medicine-specific aspects of RDM (e.g. privacy issues), introduce ZB MED services (e.g. RDMO4Life, the ELN guide, LIVIVO), and present results of NFDI4Health and NFDI4Microbiota. In practice, for instance for the topic of metadata, this would look as follows: someone from the partner institution would introduce the concept of metadata, and then someone from ZB MED would present (bio-)medicine-specific examples of metadata.

This workshop runs offline or online over 7 hours and includes interactive activities. We usually recommend splitting the workshop over two to four days, and to include breaks every 1 or 1.5 hours.

Outline

The workshop starts with welcoming the participants and explaining the practical details. A pad with information that might be relevant throughout the workshop (e.g. contact details of the trainers, schedule) can be shared with participants at the beginning of the workshop. The main workshop sections - outlined below - cover the fundamental concepts of RDM, the different steps of the research data life cycle and legal issues. The workshop ends with closing remarks and a Q&A session.

Fundamental concepts

- Research data
- Research Data Management (RDM)
- Metadata & metadata standards
- FAIR data principles
- Good scientific practice
- Policies & guidelines on managing research data (e.g. the publication guidelines of NFDI4Health)

Research data life cycle

- Planning (e.g. Data Management Plans (DMPs))
- Data collection (e.g. Electronic Lab Notebooks (ELNs))
- Data processing & analysis (e.g. reproducible data analysis)
- Data sharing & publishing (e.g. trusted data repositories)
- Data preservation (e.g. format recommendations)
- Data discovery & reuse (e.g. search engines such as the Study Hub NFDI4Health)

Legal issues

- Privacy issues
- Licences
- Legal issues when publishing research data

Practical information

Number of participants

This workshop best works for 10 to 30 participants. If there are less than 10 registrants, we postpone the workshop as it requires a lot of preparation. If there are more than 30 registrants, we convert the workshop into a two-hour introductory lecture on RDM, ending with an offer of subsequent workshops on specific topics (e.g. DMPs, reproducible data analysis). The reason being that the more participants there are, the less they seem to engage in interactive activities. Moreover, our support, for instance during exercises, is limited.

Registrations

We leave the responsibility of registration and promotion to the partner, but we provide them with a registration template. This template includes questions whose answers allow us to adapt the workshop to the audience:

- **What is your field of research?** [Free-form answer]
- **What is your career stage?**
 - Ph.D. candidate
 - Postdoctoral researcher
 - Junior professor
 - Professor
 - Director of an institute
 - Other (please specify)
- **Do you work in a wet lab?** [Yes / No]
- **Do you handle big data (i.e. data that are too large or too complex for ordinary computing devices to process)?** [Yes / No]
- **Do you handle sensitive data (e.g. personal health data)?** [Yes / No]
- **Do you write script, code and/or software?** [Yes / No]
- **What are your expectations about the workshop?** [Free-form answer]
- **Please write any questions you might have about research data management.**

At the beginning of the workshop, we reflect on participants' expectations and, for topics we are not covering, we refer them to relevant resources. We also adjust our teaching materials in the long run based on expectations and questions that often arise.

Platform for online training

We also leave the responsibility of setting up the meeting(s) to our partner. However, if they cannot do it, we do it ourselves with our department's Zoom account.

If the platform allows, we create instant polls before the workshop, download the poll report(s) afterwards, anonymize them and upload them to the shared folder (see below for more information on the shared folder).

Moderation

The moderator and time-keeper are designated during the first meeting with the partner. During the workshop, after each section, the moderator, or the trainer who has just presented, should ask the audience whether there are any questions. If the workshop takes place over several days, the moderator can announce the topics that will be covered in the next day(s) at the end of each day.

Slides

We provide our partner with an online neutral slide template, and encourage every trainer to directly work in the template in order to maintain a certain homogeneity. In the template, slides' header is in one colour for our partner (i.e. for general aspects of RDM, local infrastructure and services), and in another colour for ZB MED (i.e. (bio-)medicine-specific

aspects of RDM, ZB MED services and results of NFDI4Health and NFDI4Microbiota). We also provide slides from past workshops as examples.

To avoid disruptions in the layout, we download the slides in PDF before presenting them and, as a result, animations are not displayed.

Ideally, the slides should be ready sufficiently in advance in order to share them with registrants before the workshop. In reality, though, slides are often edited until the very last minute. Consequently, the slides are shared with the participants with a CC BY licence (with a disclaimer: “unless stated otherwise within the sources cited in this work”) after the workshop.

Interactive activities

Almost every section of this workshop includes an interactive activity such as a Q&A, an instant poll, a discussion, a video, an exercise, an assignment or a word cloud.

For Q&A, we give participants the opportunity to ask questions before and during the workshop, and we also have a dedicated 15-min Q&A session at the end of each day.

Instant polls can serve several purposes. They can act as a support for the content of the workshop, for instance by asking participants to share their opinion or experience (e.g. “What data formats do you use the most?”). Polls can also be used as comprehensive checks, i.e. to test the audience understanding (e.g. “FAIR data have to be open” [Yes / No]). Last but not least, polls can help in getting feedback (e.g. “Which section was the most understandable?”). For our full list of our polls, see the appendices below.

Discussions can take place with the whole group or in breakout rooms, as well as verbally or on online whiteboards. In this workshop, we usually include two discussions initiated by raising questions:

1. Discussion on Good Scientific Practice (GSP):

- What are examples of GSP?
- What are examples of scientific misconduct?
- How can we secure research integrity?

2. Discussion on Privacy Issues:

- Why is health data privacy important?
- What are the risks of not sharing clinical trial data?
- Is it possible to anonymize personal health data?
- What are alternatives to data anonymisation to ensure data privacy?

Playing a video during a workshop can also serve several purposes: to help hit some points, to drive a discussion, to tell stories, to re-engage or test the audience. When playing a video, it is good to express the objective right before, and discuss it after. Videos we usually include are:

- [The what, why and how of data management planning](#) (in English)

- [Electronic Lab Notebook ELN: Labfolder & eLabFTW | Tutorial Teil 1 – Einführung](#) (in German with English subtitles)
- [Electronic Lab Notebook ELN: Labfolder & eLabFTW | Tutorial Teil 2 – Benutzeroberfläche](#) (in German with English subtitles)
- [Electronic Lab Notebook ELN: Labfolder & eLabFTW | Tutorial Teil 3 - Workflow](#) (in German with English subtitles)
- [What is the National Research Data Infrastructure \(NFDI\)?](#) (in English)

Feedback

The way we collect feedback is agreed on during the first meeting with the partner. As the audience is composed of researchers from the partner's institution, we encourage them to use their own feedback system. If they do not have one, we can use an instant poll with the following questions:

- **Which section was the most understandable?**
- **Which section was the least understandable?**
- **On a scale from 1 to 10, how likely are you to recommend this workshop to your colleagues?**
- **How helpful was the workshop for you?**

Shared folder

To best prepare for the workshop, we share the following documents with our partner in a shared folder of a cloud service (e.g. Google Drive, Sciebo):

- Outline
- Teaching script
- Registration template
- Slide template
- Past presentation examples
- List of interactive activities and of instant polls for inspiration

Timeline

To best prepare for the workshop, we suggest following this timeline:

- **Month 6 to 12 before the workshop:** first meeting with the partner
 - Find (a) date(s) for the workshop
 - Agree on:
 - the outline
 - the distribution of the sections and time allocation
 - the moderation
 - when the slides should be ready
 - interactive activities
 - how to collect feedback
 - when to have the next meeting
- **Week 2 before the workshop**
 - Closing registration
 - Second meeting with the partner

- **Day 7 to 10 before the workshop**
 - Email to registrants:
 - Pre-workshop questionnaire (see “Registrations”)
 - Reminder about the practical details
 - If needed (e.g. for an exercise), inform them that they need to have their dataset close at hand during the workshop
 - If the platform allows, create instant polls
- **Post workshop**
 - Poll reports and feedback results are anonymised and uploaded to the shared folder
 - Poll results, online whiteboard figures, etc. are added to the slides, slides are merged into a single PDF, a CC BY licence is added and the resulting slides are uploaded to the shared folder and shared with the participants by the partner

Report on past workshops

In this section, you will find information on how we found potential partners, how we established and then kept contact, as well as information on the successful collaborations and workshops that we gave.

Selection of potential partners

We selected potential partners by identifying RDM centres from German universities with a faculty of medicine, health sciences, public health or similar.

First contact

To establish a first contact, we sent an email whose template can be found in the appendices. To run the first meeting with a potential partner, we have a protocol template to make sure we cover all the main topics, namely concept, number of participants, registrations, platform for online training, moderation, slides, interactive activities and feedback.

Existing collaborations

In addition to already-existing collaborations with three institutions, this first contact led to collaborations with 16 additional institutions, including four alliances: the Berlin University Alliance, “*Forschungsdatenmanagement und digitale Langzeitarchivierung in Kooperation der drei Universitäten Düsseldorf-Siegen-Wuppertal*” (FoDaKo), the Hessian Research Data Infrastructures (HeFDI) and “*Forschungsdatenmanagement in Brandenburg*” (FDM-BB).

Second contact

If we gave a workshop with a partner and, if they did not contact us again themselves, we contacted them again to continue our collaboration (see email template in the appendices). So far, that led to second joint training workshops with seven partners.

Past workshops

Since 2019, ZB MED has given more than 20 workshops and trained more than 400 researchers on the topic of RDM. A summary of the workshops given can be found in Table 1.

Alliance / institution	Types of event	Title	Place	Year
University of Cologne	Workshop	Introduction to Research Data Management	Cologne	2019
RWTH Aachen University	Workshop	Research Data Management	Aachen	2019
Max Delbrück Center for Molecular Medicine	Workshop	Research Data Management	Online	2020
CRC1382 Gut-Liver Axis	Workshop	Research Data Management in Medicine	Online	2020
Berlin University Alliance	Workshop	Workshop on Research Data	Online	2021
<i>“Forschungsdatenmanagement und digitale Langzeitarchivierung in Kooperation der drei Universitäten Düsseldorf-Siegen-Wuppertal” (FoDaKo)</i>	Workshop	Workshop on Research Data Management	Online	2021
Martin-Luther-University Halle-Wittenberg	Workshop	Workshop on Research Data Management	Online	2021
Johannes Gutenberg University of Mainz	Lecture	Workshop on Research Data	Online	2022
Hessian Research Data Infrastructures (HeFDI)	Workshop	Workshop on Research Data	Online	2022
Rhenish Friedrich Wilhelm University of Bonn	Workshop	Research Data Management in Medical and Biomedical Sciences	Online	2022
Friedrich Schiller University Jena	Workshop	Workshop on Research Data	Online	2022
Leipzig University	Workshop	Workshop on Research Data	Online	2022
Berlin University Alliance	Workshop	Workshop on Research Data Management	Online	2022
<i>“Forschungsdaten in</i>	Workshop	Workshop on Research	Online	2022

Brandenburg” (FDM-BB)		Data		
University of Oldenburg	Workshop	Workshop on Research Data Management	Online	2022
CRC296 LocoTact	Workshop	Workshop on Research Data Management in Medical and Biomedical Sciences	Online	2022
European Virus Bioinformatics Center	Workshop	Research Data Management	Online	2022
Digital Infrastructure for the Life Sciences	Workshop	Workshop on Research Data Management	Dortmund	2022
Friedrich Schiller University Jena	Workshop	Workshop on Research Data Management	Online	2022

Table 1. Past joint training workshops.

Prospects

To further develop our training offer, we could contact potential partners who have not responded after our first contact round. We have also thought about developing Bring Your Own Data workshops or a type of flipped classroom: we would record a video on a specific topic (e.g. DMPs), let people watch it on their own, and organise a follow-up discussion or Q&A session on the topic.

Appendices

Polls

Icebreaker poll

- **What is your subject area?** [Free-form answer]
- **What data format do you use the most?** [Multiple selection]
 - Artefacts, specimens, samples
 - Atomic coordinates (e.g. PDB)
 - Audiotapes or videotapes
 - Genome annotations (e.g. BAM/SAM, GFF/GTF, Pileup)
 - Geospatial data (ESRI Shapefiles, GML, KML, MapInfo, MID, MIF)
 - Images (DICOM, EPS, GIF, Illustrator, JPEG, JPG, PDF, PNG, STL, SVG, TIFF)
 - Lab notebooks, field notebooks, diaries
 - Nucleotide sequences or amino acid sequences (e.g. FASTA, FASTQ)
 - Scripts, code or software
 - Slides (e.g. PowerPoint)

- Spreadsheet (e.g. XLS)
- Text files (e.g. CSV, DOC, ODT, PDF/A, RTF, TXT, XML)
- Other (please specify)
- **Do you add metadata to your data?** [Yes / No]
- **Have you ever heard of the FAIR data principles?** [Yes / No]
- **Have you ever created a Data Management Plan (DMP)?** [Yes / No]
- **Have you ever published your data?** [Yes / No]

Poll on metadata

- **In which format(s) are metadata documented in your research group?** [Multiple selection]
 - In the file itself
 - In text files (e.g. README, XML file)
 - In input fields
 - In your own schemes
 - In standardised schemes
 - In a metadata repository
 - Other (please specify)
 - Metadata are not documented in my research group

Polls on the FAIR data principles

Credit for the two following polls: Dr. Roberto Cozatl and Dr. Gerrit Grenzebach from Martin-Luther-University Halle-Wittenberg.

Poll 1

- **FAIR data have to be open** [Yes / No / I don't know]
- **Open data are more useful than FAIR data** [Yes / No / I don't know]
- **The quality of FAIR and open data is extremely good** [Yes / No / I don't know]

Poll 2

What would be the appropriate access rights for each of the datasets described below?

- **A funded research project where access to the data will not be available until the project is complete** [Open / Closed / Embargoed]
- **Data from a seminar during which speakers have to read a story aloud and authorization to publish the data are available from all speakers** [Open / Closed / Embargoed]
- **Data from a seminar during which speakers have to read a story and authorization to publish the data are not available** [Open / Closed / Embargoed]
- **An open dataset which has been used in the context of a research final paper that has a licence allowing its usage, further reprocessing and subsequent publishing** [Open / Closed / Embargoed]
- **Interview materials about sensitive topics where anonymisation is not possible** [Open / Closed / Embargoed]

Poll on Electronic Lab Notebooks (ELNs)

- **Do you work in a wet lab?** [Yes / No]
- **Do you use a paper lab notebook?** [Yes / No]
- **Have you ever heard about Electronic Lab Notebooks (ELNs)?** [Yes / No]
- **Have you ever used an ELN?** [Yes / No]
- **Do you currently use an ELN?** [Yes / No]
- **If you currently use (an) ELN(s), which one(s)?** [Multiple selection]
 - Benchling
 - Chemotion
 - eLabFTW
 - eLabJournal
 - Labfolder
 - OpenBIS
 - RSpace
 - SciNote
 - Other (please specify)
 - None

Poll on data processing and analysis

- **Do you write scripts, code and/or software?** [Yes / No]
- **Which of these programming languages do you use?** [Multiple selection]
 - Bash
 - C, C++
 - Java
 - Julia
 - MATLAB
 - Python
 - R
 - Rust
 - Other (please specify)
 - None
- **Do you use version control?** [Yes / No]

Poll on data sharing

- **How is data shared within your research group or with external collaborators?**
 - Common line tool (e.g. DataLad)
 - Email
 - Mobile drive (e.g. USB stick)
 - Non-commercial cloud (e.g. B2DROP, Sciebo)
 - Proprietary cloud (e.g. Dropbox, Google Drive)
 - Web application (e.g. FAIRDOME-SEEK Hubs, Open Science Framework (OSF), SciFlow, WeTransfer)
 - Other (please specify)

Poll on digital preservation

- **Which of these two types of file formats do you use most in your research group?** [Open / Proprietary]

Feedback poll

- **Which section was the most understandable?**
 - Section 1
 - Section 2
 - ...
 - Everything was understandable
 - Nothing was understandable
- **Which section was the least understandable?**
 - Section 1
 - Section 2
 - ...
 - Everything was understandable
 - Nothing was understandable
- **On a scale from 1 to 10, how likely are you to recommend this workshop to your colleagues?** [1 to 10]
- **How helpful was the workshop for you?**
 - Extremely helpful
 - Very helpful
 - Somewhat helpful
 - Not so helpful
 - Not at all helpful

Email templates

First contact

Subject: Training Workshop on Research Data Management in (Bio-)Medicine

Dear Madam or Sir,

We would like to collaborate with the research data centre of the University of [...] in the organisation of a workshop on research data management for medical and biomedical sciences. The target group is all-qualification-level researchers.

We have already given this workshop multiple times, and we see synergies by combining the generic workshops of research data centres with our discipline-specific approach.

This workshop can be given offline over one day, or online (e.g. in four 2-hour sessions).

The specific structure of the workshop is attached to this email. Since, in our experience, the workshop participants speak English, we offer it in English. Of course, we could also offer it in German.

This offer from us is free of charge for universities.

We would be pleased to discuss this opportunity with you, and will be happy to answer any questions you may have.

Yours sincerely

[Signature]

Second contact

Subject: 2nd joint training workshop on RDM

Dear [partner's name],

We really enjoyed working with you and think our joint Training Workshop on Research Data Management in (Bio-) Medicine was a success. We would like to continue our cooperation and to further refine the concept we have already developed.

Would you like to offer this workshop again? If so, we could also start to give this workshop on a regular basis, for instance once a year.

If you are interested, we would suggest offering the workshop [approximate timeframe], over two to four days. Could you please let us know which month would suit you best?

We are looking forward to hearing from you.

Best wishes

[Signature]

Licence

This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) Licence. To view a copy of the licence, visit

<https://creativecommons.org/licenses/by/4.0/legalcode>.

