

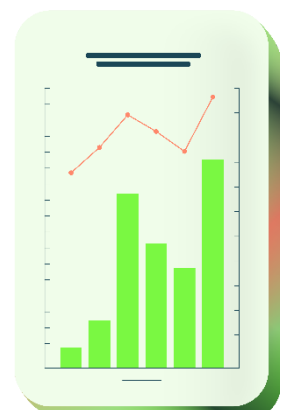
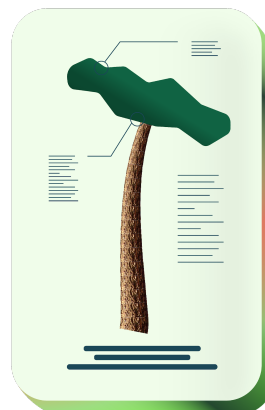
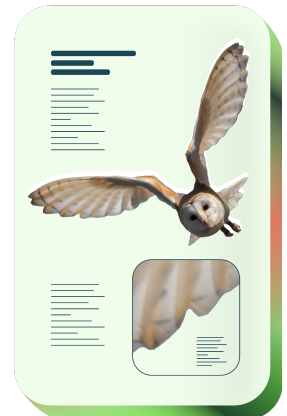


NFDI 4  
**BIODIVERSITY**  
BIODIVERSITY, ECOLOGY & ENVIRONMENTAL DATA

# **Template for the NFDI4Biodiversity & GfÖ Winter School**

Data Management in Ecology and  
Environmental Science

5th - 9th of December 2022



## Authors

Juliane Röder<sup>2</sup>, Marlen Fischer<sup>2</sup>, Daniel Tschink<sup>1</sup>, Ortrun Brand<sup>2</sup>

<sup>1</sup>German Federation for Biological Data (GFBio e.V.), Mary-Somerville-Str. 2-4, 28359 Bremen, Germany

<sup>2</sup>Philipps-University of Marburg, Biegenstraße 36, 35032 Marburg, Germany

## Published by

Philipps-University of Marburg, Biegenstraße 36, 35032 Marburg, Germany

We would like to thank each and every one for making this event truly exceptional and inspiring.

To the esteemed lecturers - Carsten Dormann, Birgit Gemeinholzer, Matthias Grenié, Dennis-Kenji Kipker, Ivaylo Kostadinov, Jimena Linares, Wolfgang Müller, Michael Oellermann, Hanieh Saeedi, and Sophie Wolf - your captivating lectures and insightful guidance have expanded our understanding and opened our minds to new possibilities in handling research data.

To the organisational support provided by Jens Nieschulze and the Ecological Society of Germany, Austria and Switzerland (GfÖ) - our collaboration helped us to develop a program close to the researchers' needs.

A special appreciation goes to Marthe Irene Eisner for her commitment and enthusiasm in bringing this course to life.

Last but not least, we want to thank Claas Thido Pfaff, Ivaylo Kostadinov, and Jörg Brünecke for their technical support. Your dedication to ensuring smooth operations and addressing any challenges allowed us to focus on our learning without any disruptions.

The knowledge and insights we've gained from this course have been invaluable, and we are grateful for the opportunity to be part of such a remarkable learning experience.

This document is published under the [Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).



This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

The work was supported by the German Research Foundation DFG under the grant agreement number [442032008](https://www.dfg.de/grants/442032008) (NFDI4Biodiversity). NFDI4Biodiversity is part of NFDI, the National Research Data Infrastructure in Germany ([www.nfdi.de](https://www.nfdi.de)).

# Content

Abstract.....	1
1. Introduction.....	1
2. Course objectives and outcomes.....	3
3. Required resources and materials.....	3
4. Course program.....	5
Day 1 – Research Data Management & Data Management Plan (DMP).....	5
Day 2 – Data standards and legal aspects.....	6
Day 3 – Handling spatial data.....	7
Day 4 – Data integration, data annotation & taxonomic harmonisation.....	8
Day 5 – Data publication.....	9
5. Lessons learned and outlook.....	10
6. References.....	11
7. Supplementary material.....	12
A. Course material.....	12
B. Tools and software used during the course.....	13
C. Course logistics and schedule.....	14
D Application forms.....	16
E. Anonymous survey for course evaluation.....	17

## Abstract

This document describes a one-week course program for a Winter School on research data management (RDM) which was organised in 2022 by partners from the NFDI4Biodiversity consortium. The course aimed to provide young scientists in ecology and environmental science with state-of-the-art skills and knowledge in handling scientific data throughout the data life cycle. Lectures and practical sessions were given by experts from renowned universities, data centers, and research institutions in the biodiversity community, covering general aspects of RDM and aspects specifically important in handling ecological data. The participants were introduced to Jupyter notebooks, and learned how to combine software tools like Jupyter, R, Python, RightField and OpenRefine. The course was conducted remotely and targeted PhD students (R1) and Early-Career-Researchers (R2). This publication summarises detailed information about the structure, organisation and lessons learned and is supposed to be used as a guide for similar events.

## 1. Introduction

The amount of data produced in ecological and biodiversity research is rapidly growing, and increasingly large and heterogeneous datasets are used to answer complex systemic questions. Much of the value and reliability of these studies depend on proper data (quality) management and workflows. Consequently, data competence has to develop in pace with modern technologies used in the domain.

Improved research data management allows for efficient spending of public funding by enabling data sharing and reuse, promoting collaborative efforts, and reducing redundancy in data collection. It also contributes to building knowledge better by facilitating robust and reproducible research, enabling evidence-based decision-making, and enhancing the transparency and credibility of scientific findings. Consequently, the most important third-party funder in Germany, the German Research Foundation (DFG) (DFG et al., 2021), incorporated RDM in their mandatory code of conduct “Safeguarding Good Research Practice” (DFG, 2022).

In the context of the biodiversity crisis, high-quality biodiversity data are critical for modelling current trajectories of biodiversity and ecosystem decline, informing policy decisions, and developing effective conservation strategies. By actively participating in research data management and contributing high-quality biodiversity data, ecologists can make a meaningful contribution to the knowledge base of institutions at local to global levels, like local environmental agencies and the expert syntheses of the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES).

As more data become available, managing these large and diverse datasets becomes crucial to ensure their usability for future research. This extends to making old data sets available, with the challenge of data integration. For biological data, matching taxonomic information is a specific and essential form of data integration: matching old and alternative names to valid species names and species concepts (Grenié et al., 2022). Handling and analysing large and complex

data sets can be challenging, and usually involves many steps of data cleaning and reformatting. So called “executable papers” or reproducible scientific articles, i.e. data, code, analyses and the manuscript text combined in a Jupyter Notebook (Jupyter.org, 2023) or a [R Markdown](#) file, can improve transparency and reproducibility of the scientific process (Frank & Hartgerink, 2017; Perkel, 2022) and can be a valuable educational tool (Davies et al., 2020).

The German National Research Data Infrastructure (NFDI) was funded by the DFG to provide access to research data, to integrate this data and make it usable in a sustainable and qualitative manner for the entire German science system. NFDI4Biodiversity is one of a total of 26 NFDI consortia, consisting of 50 scientific institutions, museums, natural history societies, state offices, and other institutes and expert groups, providing a wide range of services for handling biodiversity and environmental data. NFDI4Biodiversity offers access to standards, protocols, and formats for data exchange, as well as technical documentation, a data repository search, and the ability to submit data sets and create data management plans according to community standards. The consortium also organises and participates in community events, including training for students and researchers.

The NFDI4Biodiversity & GfÖ Winter School on Data Management in Ecology and Environmental Science was a one-week program for First Stage Researchers (R1) (European Commission, 2011, but see European Commission, 2022a, 2022b) with a background in ecology and environmental science. The course, held on 5-9 December 2022, focused on theoretical and practical aspects of research data management, including metadata standards, data integration, and taxonomic harmonisation. It covered central aspects of the data life cycle and emphasised handling data according to the FAIR principles (Findable, Accessible, Interoperable, Reusable, (Wilkinson et al., 2016)). The course used data and tools provided by partners in the NFDI4Biodiversity network, including plant trait data from the iDiv PlantHub project (iDiv 2023) and cloud services from the German Network for Bioinformatics Infrastructure (de.NBI). Lecturers from renowned universities, data centres, and research institutions in the biodiversity community shared their knowledge in lectures and practical sessions, providing state-of-the-art skills and knowledge in research data management. The course aimed to develop skills in handling scientific data along the data life cycle and promote the use of open teaching formats aligned with the FAIR principles.

The following institutions were involved in the preparation and in the implementation of the Winter School:

- Coordination:
  - [German Federation for Biological Data \(GFBio e.V.\)](#)
  - [Philipps-Universität Marburg](#)
  - [Ecological Society of Germany, Austria and Switzerland \(GfÖ\)](#)
  - [Göttingen State and University Library \(SUB\)](#)
- Technical Infrastructure:
  - [German Network for Bioinformatics Infrastructure \(de.NBI\)](#)
  - [Helmholtz Centre for Environmental Research \(UFZ\)](#)
  - [German Federation for Biological Data \(GFBio e.V.\)](#)

- Lectures (alphabetical order):
  - [Albrecht-Ludwigs-University of Freiburg](#)
  - [German Centre for Integrative Biodiversity Research \(iDiv\)](#)
  - [German Federation for Biological Data \(GFBio e.V.\)](#)
  - [Heidelberg Institute for Theoretical Studies \(HITS gGmbH\)](#)
  - [University of Leipzig](#)
  - [MARUM - Center for Marine Environmental Sciences](#)
  - [National Research Data Infrastructure for Personal Health Data \(NFDI4Health\)](#)
  - [PANGAEA - Data publisher for Earth & Environmental Science](#)
  - [Senckenberg - Leibniz Institution for Biodiversity and Earth System Research \(SGN\)](#)
  - [University of Bremen](#)
  - [University of Kassel](#)

## 2. Course objectives and outcomes

Our aim was to develop a course program specifically for the needs of ecologists, based on open software and free of charge. In this course, we combined tools ecologists are already using (e.g. R, Python, Excel) with innovative and interdisciplinary tools (e.g. Jupyter, OpenRefine, RightField), and we showed how those tools could be used to integrate research data management in the participants' daily research routine. We further introduced openly available resources for research data management like global biodiversity databases and repositories, and state-of-the-art knowledge about scientific best practice and the legal foundations of handling research data. We structured the course program to cover all parts of the data life cycle in only five days.

In detail, the participants learned how to create data management plans (DMPs) (e.g. GFBio 2023), about metadata standards, and about tools that could help to create metadata. The participants learned how to extract data from open access databases, how to assure data quality, and how to integrate different data sets. Data integration also covered the challenging task of taxonomic harmonisation. The participants learned about legal aspects of handling ecological data, and finally, they learned how to publish data sets.

## 3. Required resources and materials

To make sure that practical sessions were effective and every participant was able to ask questions, the number of places was restricted to 20. We reserved half of the places for GfÖ members, and the other half for non-members. The course was conducted remotely through Zoom, with English as the primary language of instruction. Both features made the course more accessible, but everyone involved needed access to a computer with a stable internet connection, a camera and a microphone. Each day of the course consisted of two lectures with theoretical aspects and a practical session (see [course program](#)).

We scheduled two hosts per day, and four hosts for the duration of the five-day course. Additionally, one person was available for tech support during each of the practical sessions. A total of nine lecturers were involved, i.e. two to three

lecturers per day. The organising team consisted of seven people, including three experts who set up the virtual teaching infrastructure in the de.NBI cloud. We used BinderHub with Jupyter environments based on Git(Hub) repositories – one of them featuring RStudio.

The lecturers prepared slides and scripts in R, Python, or Git repositories including code, data and documentation. Most of the lecturers agreed on publishing their slides, and most also agreed to have a video recording of their lecture published after the course. We did not record any of the practical sessions, to facilitate exchange between participants and lecturers.

Participants used a spreadsheet software on their computer (e.g. Excel or LibreOffice) and were supposed to install the free tools RightField (Wolstencroft et al., 2011) and OpenRefine to be able to follow the hands-on session about data annotation (Day 4).

We compiled a detailed list of tasks for the organising team, the lecturers and the participants (see Supplement C). We also listed all tools and databases that were used in this course (Supplement B), as well as all published teaching materials and recorded lectures (Supplement A).

## 4. Course program

### Day 1 – Research Data Management & Data Management Plan (DMP)

<b>08:30 – 09:00 Warm-Up</b> by Daniel Tschink (GFBio e.V.) <ul style="list-style-type: none"><li>• Networking/Introduction</li><li>• Welcome &amp; organisational aspects</li></ul>
<b>09:00 – 10:20 Lecture 1</b> by Carsten Dormann (Albrecht-Ludwigs-University of Freiburg)  <b><i>Keynote - Pitfalls in ecological data sets</i></b> <ul style="list-style-type: none"><li>• Naming conventions</li><li>• Standards</li><li>• Reproducibility, etc.</li></ul>
<b>10:20 – 10:40 Break</b>
<b>10:40 – 12:00 Lecture 2</b> by Ivaylo Kostadinov (GFBio e.V.)  <b><i>Research Data Management &amp; Data Literacy</i></b> <ul style="list-style-type: none"><li>• Local organisation</li><li>• Folder structures</li><li>• Backup strategies</li><li>• Versioning, etc.</li></ul> <p>➤ Recorded talk on <a href="#">YouTube</a></p>
<b>12:00 – 13:00 Break</b>
<b>13:00 – 14:55 Practical session 1</b> by Jimena Linares (GFBio e.V.)  <b><i>Create a Data Management Plan (DMP)</i></b> <ul style="list-style-type: none"><li>• DMPs as part of local organisation, for FAIRness and policy compliance</li><li>• Parts of a DMP</li><li>• Create a DMP based on your research</li></ul> <p>➤ Slides of the talk are published on <a href="#">Zenodo</a></p> <p>➤ Tool introduced and used during the lecture: <a href="#">GFBio Data Management Plan (DMP)</a></p>
<b>14:55 – 15:00 Wrap-up day 1 and outlook day 2</b>



## Day 2 – Data standards and legal aspects

### 08:30 – 09:00 Warm-Up

- Quick introduction of participants (60 sec per person):  
Who are you?
  - Your name, institution, position?
  - What is your project about?
  - What are your expectations for the Winter School?

### 09:00 - 11:00 Practical session 2 by Sophie Wolf (University of Leipzig)

#### *Introduction to Jupyter*

- What is Jupyter
  - How to use it
- Tutorial as a [Jupyter Book on GitHub](#)

### 11:00 - 11:20 Break

### 11:20 - 12:40 Lecture 3 by Dennis-Kenji Kipker (University of Bremen)

#### *Legal aspects*

- Sensitive data
  - Rights of use
  - Licences
  - Copyright
- Slides of the talk are published on [Zenodo](#)
- Recorded talk on [YouTube](#)

### 12:40 - 13:40 Break

### 13:40 - 14:55 Lecture 4 by Birgit Gemeinholzer (University of Kassel)

#### *Data standards*

- Metadata Standards for (Meta)Data structures
- Examples e.g. [EML](#), [Darwin-Core](#), [ABCD](#)

### 14:55 - 15:00 Wrap-up day 2 & outlook day 3

## Day 3 – Handling spatial data

### 08:30 – 09:00 Warm-Up

- What are your experiences in handling data?
  - Best vs. worst experiences?
  - Do you have any data management routines at your working group/institution or for your project? Why not?
  - What kind of support would you, as an individual researcher, most urgently need?

### 09:00 – 10:30 Lecture 5 by Hanieh Saeedi (pre-recorded) (SGN)

#### ***Investigating marine biodiversity using digitized natural history collections & open-access databases: following FAIR data principles (Topic 1 and 2)***

- Status of Marine Biodiversity
- Challenges and opportunities of studying biodiversity
- The importance of natural collections in biodiversity studies
- Technology and digitization advances

### 10:30 – 10:40 Break

### 10:40 – 12:10 Lecture 6 by Hanieh Saeedi (pre-recorded) (SGN)

#### ***Investigating marine biodiversity using digitized natural history collections & open-access databases: following FAIR data principles (Topic 3)***

- Open-access databases for marine biodiversity studies
- [World Register of Marine Species \(WoRMS\)](#)
- [Ocean Biodiversity Information System \(OBIS\)](#)

### 12:10 – 13:00 Break

### 13:00 – 14:30 Practical session 3 by Hanieh Saeedi (SGN)

#### ***Handling spatial marine data***

- Extracting data from open-access databases such as GBIF, OBIS via online portals and R
- [Darwin Core](#)/data standards used in OBIS and GBIF

### 14:55 – 15:00 Wrap-up day 3 & outlook day 4

## Day 4 – Data integration, data annotation & taxonomic harmonisation

### 08:30 – 09:00 Warm-Up

- Why don't we share data?
  - Inspiration: Figure 1 from Gomes et al. (2022) Why don't we share data and code? Perceived barriers and benefits to public archiving practices *Proc R Soc B* 289: 20221113 <https://doi.org/10.1098/rspb.2022.1113>

### 09:00 - 10:20 Lecture 7 by Wolfgang Müller (HITS gGmbH)

#### **Data integration & annotation**

- Data structuring/labeling
  - How/when can terminology services be employed
  - How to structure/annotate data
  - FAIRification of data
  - Useful techniques in Excel, VBA, Python, Google Refine, RightField
  - Tools introduced and used during the lecture:
    - [OpenRefine](#)
    - [RightField](#)
    - [ReStoRunT](#)
- Recorded talk on [YouTube](#)

### 10:20 - 10:40 Break

### 10:40 - 12:00 Lecture 8 by Matthias Grenié (iDiv Leipzig)

#### **Naming conventions & taxonomic harmonization**

- The lecture was based on Grenié et al. (2022)
  - The shiny app “taxonomicexplorer” visualizes relationships between taxonomic databases and R packages (Grenié et al. 2021)
- Slides of the talk are published on [Zenodo](#)
- Recorded talk on [YouTube](#)

### 12:00 - 13:00 Break

### 13:00 - 14:55 Practical session 4 by Sophie Wolf (University of Leipzig)

#### **Jupyter Hub**

- Data analysis/visualization using Jupyter
  - Combining Jupyter, R, Python
- Tutorial as a [Jupyter Book on GitHub](#)

### 14:55 - 15:00 Wrap-up day 4 & outlook day 5

## Day 5 – Data publication

### 08:30 – 09:00 Warm-Up

- Open questions and feedback
  - Link to Indico [survey for anonymous feedback](#)

### 09:00 - 10:20 Lecture 9 & Practical session 5 by Michael Oellermann (PANGAEA)

#### **Data publication**

- Why share and publish data?
  - Where can data be archived and published?
  - Persistent Identifiers (PIDs)
  - Practical session in groups:
    - Finding suitable data repository using [re3data.org](#) or [fairsharing.org](#)
    - Comparing data publications among different data repositories
- Slides of the talk and results of the practical sessions are published on [Zenodo](#)
- Recorded talk on [YouTube](#)

### 10:20 - 10:40 Break

### 10:40 - 12:00 Lecture 10 & Practical session 6 by Michael Oellermann (PANGAEA)

#### **Data archiving**

- How to publish research data
  - How to curate your data
  - How to submit and cite your data
  - Practical session:
    - Group work: Peer review your data using practice data table on [Google Jamboard](#)
    - Live submission of data to [GFBio](#) or [PANGAEA](#)
    - Data curation checklist using a [Jupyter Notebook](#) in [Binder/GoogleColab](#)
- Slides of the talk and results of the practical sessions are published on [Zenodo](#)
- Recorded talk on [YouTube](#)
- Jupyter Notebook on GitHub on [how to check data before submission to PANGAEA](#)

### 12:00 - 13:00 Break

### 13:00 - 14:00 Wrap-up day 5 & Wrap-up overall & Goodbye

## **5. Lessons learned and outlook**

We invited participants to give feedback on the structure and the contents of the course in two different ways: 1) during the warm-up session of the last day of the course, and 2) via an anonymous survey. We received valuable feedback from both approaches. The participants e.g. suggested adjustments to the course content, like more input on handling spatial data, more programming with R than with Python, and more focus on terrestrial ecology and corresponding databases. For the next course, we are also planning to facilitate the exchange between lecturers during the planning phase, in order to improve coherence of lecture contents. In addition, we are planning to improve our backup plans in case one of the lecturers cannot support the course on short notice. For the organizing team, it was crucial to agree on the scope of work for each partner early in the planning process. We have to improve some technical details, though, e.g. using only one form during the application phase. For the next course, we will prepare and circulate a metadata template including keywords for lecture slides that are supposed to be uploaded to Zenodo. Finally, we are planning to select a slightly more homogenous group of participants for the next course. To accomplish this, we are planning to add questions about career status and programming skills in the application form.

We are looking forward to repeating and improving this course!

## 6. References

- Davies, A., Hooley, F., Causey-Freeman, P., Eleftheriou, I., & Moulton, G. (2020). Using interactive digital notebooks for bioscience and informatics education. *PLOS Computational Biology*, 16(11), e1008326. <https://doi.org/10.1371/journal.pcbi.1008326>
- Deutsche Forschungsgemeinschaft. (2022). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. <https://doi.org/10.5281/zenodo.6472827>
- DFG, Fischer, C., Heidler, R., Klöble, K., Porschke, A., Stellmach, F., & Weigelt, M. (2021). *DFG Förderatlas 2021—Kennzahlen zur öffentlich finanzierten Forschung in Deutschland*. <https://www.dfg.de/sites/foerderatlas2021>
- EURAXESS. (2023). *Research profiles descriptors*. <https://euraxess.ec.europa.eu/europe/career-development/training-researchers/research-profiles-descriptors> Accessed: 2023-05-15
- European Commission, Directorate General for Research & Innovation, Directorate B - European Research Area, Skills. (2011). *Towards a European framework for research careers*. European Commission, Brussels. [https://euraxess.ec.europa.eu/sites/default/files/policy\\_library/towards\\_a\\_european\\_framework\\_for\\_research\\_careers\\_final.pdf](https://euraxess.ec.europa.eu/sites/default/files/policy_library/towards_a_european_framework_for_research_careers_final.pdf)
- European Commission. Directorate General for Research and Innovation. (2022a). *The European competence framework for researchers*. [https://research-and-innovation.ec.europa.eu/system/files/2023-04/ec\\_rtd\\_research-competence-presentation.pdf](https://research-and-innovation.ec.europa.eu/system/files/2023-04/ec_rtd_research-competence-presentation.pdf)
- European Commission. Directorate General for Research and Innovation., IDEA – Inspire, Debate, Engage and Accelerate Action Consult., Technopolis Group, & CWTS. (2022b). *Knowledge ecosystems in the new ERA: Using a competence based approach for career development in academia and beyond*. Publications Office. <https://data.europa.eu/doi/10.2777/150763>
- Frank, M., & Hartgerink, C. (2017). *RMarkdown for writing reproducible scientific papers*. <https://libscie.github.io/rmarkdown-workshop/handout.html> Accessed: 2023-05-15
- GFBio (2023). GFBio Data Management Plan. <https://dmp.gfbio.org/>, accessed 2023-07-18.
- Grenié, M., Berti, E., Carvajal-Quintero, J., Dädlow, G. M. L., Sagouis, A., & Winter, M. (2022). Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. *Methods in Ecology and Evolution*, 2041-210X.13802. <https://doi.org/10.1111/2041-210X.13802>
- Grenié, M., Berti, E., Carvajal-Quintero, J., Winter, M., & Sagouis, A. (2021). Matching Species Names Across Biodiversity Databases: Sources, tools, pitfalls and best practices for taxonomic harmonization. *Biodiversity Information Science and Standards*, 5, e75359. <https://doi.org/10.3897/biss.5.75359>
- iDiv.de (2023). PlantHub (SU). <https://www.idiv.de/de/forschung/plattformen.html#29>, Accessed 2023-07-18.
- Jupyter.org (2023). Introduction to the JupyterLab and Jupyter Notebooks. <https://jupyter.org/try-jupyter/retro/notebooks/?path=notebooks/Intro.ipynb>, Accessed: 2023-07-18.

- Perkel, J. M. (2022). Cut the tyranny of copy-and-paste with these coding tools. *Nature*, 603(7899), 191–192. <https://doi.org/10.1038/d41586-022-00563-z>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wolstencroft, K., Owen, S., Horridge, M., Krebs, O., Mueller, W., Snoep, J. L., du Preez, F., & Goble, C. (2011). RightField: Embedding ontology annotation in spreadsheets. *Bioinformatics*, 27(14), 2021–2022. <https://doi.org/10.1093/bioinformatics/btr312>

## 7. Supplementary material

### A. Course material

#### A.1 Lecture slides

- List of all four Zenodo records: <https://t1p.de/4BioDiv-WinterSchool>
- Single records (alphabetical order):
  - Grenié, M. (2022). *Taxonomic Harmonization: A problem you didn't know you had!* <https://doi.org/10.5281/ZENODO.7417951>
  - Kipker, D.-K. (2022). *Legal framework of research data management with special focus on copyright law and data protection law.* <https://doi.org/10.5281/ZENODO.7441388>
  - Linares, J., & Tschink, D. (2023). *DMPs as part of local organisation, for FAIRness and policy compliance.* <https://doi.org/10.5281/ZENODO.7638635>
  - Oellermann, M. (2022). *Publishing Research Data in Open Access Repositories.* <https://doi.org/10.5281/ZENODO.7441751>

#### A.2 Lecture recordings

- Playlist of all six videos: <https://www.youtube.com/playlist?list=PL06Unzn1hDrhZ4mfsdfeypJajuGsuzdfm>
- Single videos (alphabetical order):
  - Grenié, M. (2022). *Taxonomic Harmonization.* <https://www.youtube.com/watch?v=vOgadsdBGeo>
  - Kipker, D.-K. (2022). *Legal Aspects.* <https://www.youtube.com/watch?v=YSI2IUyWEiU>
  - Kostadinov, I., & Linares, J. (2022). *Data Management and Data Literacy.* <https://www.youtube.com/watch?v=CUCgHrEdt-M>
  - Müller, W. (2022). *Data Integration and Annotation.* <https://www.youtube.com/watch?v=F350FyqCkWA>
  - Oellermann, M. (2022a). *Data Search: Publish your Research Data pt. 1.* <https://www.youtube.com/watch?v=agCfSBWdRXM>
  - Oellermann, M. (2022b). *Data Search: Publish your Research Data pt. 2.* <https://www.youtube.com/watch?v=U6kr3DfS9HA>

### A.3 GitHub repositories

- Oellermann, M., & Riemann-Campe, K. (2022). *PANGAEA Data Curation Checklist* (0.9.0).  
[https://github.com/pangaea-data-publisher/community-workshop-material/tree/master/Python/Data\\_curation\\_checklist](https://github.com/pangaea-data-publisher/community-workshop-material/tree/master/Python/Data_curation_checklist)
- Wolf, S. (2022). *Jupyter Workshop Winterschool 2022* [Python, R].  
[https://github.com/sojwolf/Jupyter\\_Workshop\\_Winterschool\\_2022](https://github.com/sojwolf/Jupyter_Workshop_Winterschool_2022)

## B. Tools and software used during the course

### B.1 Organizing

- BinderHub (technical platform) <https://binderhub.readthedocs.io>
- Indico (forms and surveys) <https://getindico.io>
- Zoom <https://zoom.us>

### B.2 Teaching

- Virtual boards
  - Google Jamboard [https://edu.google.com/intl/ALL\\_de/jamboard](https://edu.google.com/intl/ALL_de/jamboard)
  - Miro <https://miro.com>
- Engagement tools
  - Mentimeter <https://www.mentimeter.com>
  - Slido <https://www.slido.com>

### B.3 Coding

- Git <https://git-scm.com>
- GitHub <https://github.com>
- Google Colab <https://colab.research.google.com>
- Jupyter <https://jupyter.org>
- Jupyter Notebook <https://jupyter-notebook.readthedocs.io>
- Python <https://www.python.org>
- R <https://www.r-project.org>
- RStudio <https://posit.co/download/rstudio-desktop>

### B.4 Data management and annotation

- DMPs
  - GFBio Data Management Plan <https://dmp.gfbio.org>
- Metadata standards
  - Morpho (EML editor offline) <https://github.com/NCEAS/morpho>
  - ezEML (EML editor online) <https://ezeml.edirepository.org/eml>
  - DarwinCore <https://dwc.tdwg.org>
  - DarwinCore templates
    - Occurrence data  
<https://ipt.gbif.org/manual/en/ipt/latest/occurrence-data#templates>
    - Checklist data  
<https://ipt.gbif.org/manual/en/ipt/latest/checklist-data#templates>
    - Sampling event data  
<https://ipt.gbif.org/manual/en/ipt/latest/sampling-event-data#templates>



- Resource metadata  
<https://ipt.gbif.org/manual/en/ipt/latest/resource-metadata#template>
- Data annotation (in spreadsheets)
  - Excel <https://www.microsoft.com/de-de/microsoft-365/excel>
  - LibreOffice <https://www.libreoffice.org>
  - RightField <https://rightfield.org.uk>
  - Open Refine <https://openrefine.org>
  - ReStoRunT <https://github.com/mertova/ReStoRunT>
  - Ontologies for biodiversity data:
    - <https://bioportal.bioontology.org>
    - <https://fairsharing.org>

## B.5 Databases and data repositories (alphabetical order)

- DRYAD <https://datadryad.org>
- Fairsharing.org <https://fairsharing.org>
- Figshare <https://figshare.com>
- GBIF – Global Biodiversity Information Facility <https://www.gbif.org>
- OBIS – Ocean Biodiversity Information System <https://portal.obis.org>
- PANGAEA – Data publisher for Earth & Environmental Science  
<https://pangaea.de>
- re3data <https://www.re3data.org>
- TRY – Plant Trait Database <https://www.try-db.org>
- WoRMS – World Register of Marine Species <https://www.marinespecies.org>
- Zenodo <https://zenodo.org>

## C. Course logistics and schedule

### C.1 ... for the organizing team

#### **Preparation:**

- Develop the format and decide on the topics
- Invite lecturers
- Set-up a programming environment for the course (tools: [Jupyter](#), [BinderHub](#), [Python](#), [RStudio](#))
- Organize a pre-course meeting of all lecturers to discuss technical requirements and lecture contents
- Set-up a website and [application form](#) (tool: [Indico](#))
- Advertise the course on the project website, via mailing lists and Twitter
- Select participants from applications and create a waiting list as long as the course's capacity
- Inform participants, applicants on the waiting list, and unsuccessful applicants
- Set-up a virtual room on a video conference platform (our tool: Zoom)
- Circulate information on technical requirements for participants
- Inform lecturers about the level of programming skills of the participants
- Organize an extra meeting to discuss details of the technical platform
- Prepare a [survey for anonymous feedback for participants](#) (tool: [Indico](#))

**During the course:**

- Open the virtual room 30 min before the lecture
- Warm-Up session
- Hosting the lectures, incl. time-keeping
- Record lectures (but not practical sessions to facilitate exchange between lecturers and participants)
- Keep notes during lectures and practical session about questions that need a follow-up, or suggestions for improvements
- Wrap-up and outlook for the next day

**After the course:**

- Circulate the link to a [survey for anonymous feedback](#) to the participants
- Circulate course material and [links to course material](#) to the participants
- Prepare and circulate certificates of attendance to participants
- Analyze results of the survey to improve this course
- Support lecturers uploading their slides in the [NFDI4Biodiversity Zenodo Community](#)
- Process recorded lectures and upload the recordings in the [NFDI4Biodiversity YouTube Channel](#)

## C.2 ... for the lecturers

**Preparation:**

- Agree to be a lecturer for the course
- Prepare your lecture slides, engagement tools and material for practical sessions
- Meet with the organizing team and the other lecturers before the course to discuss technical requirements and lecture contents
- If necessary, join an extra meeting on details about the technical platform
- Organize to have a stable internet connection for the course
- Organize to have a computer, camera and microphone for the course
- Decide if you want to share your slides in the [NFDI4Biodiversity Zenodo Community](#)
- Decide if you agree to have your lecture recorded and published in the [NFDI4Biodiversity YouTube Channel](#)

**During the course:**

- Join the virtual room in time for a short technical check
- Present your lecture
- Answer questions of the participants

**After the course:**

- Upload your slides in the [NFDI4Biodiversity Zenodo Community](#), if you agreed to do this

### C.3 ... for the participants

#### **Preparation:**

- Apply for the course
- Organize to have time for the course
- Organize to have a stable internet connection for the course
- Organize to have a computer, camera and microphone for the course
- [Download and install required software](#)

### D Application forms

We asked applicants about their scientific background, their motivation to join the course, their programming skills, and about their previous experiences with RDM.

#### D.1 Registration form (form 1)

- Name
- Email Address
- Affiliation
- GfÖ Membership (Yes - No)

#### D.2 Application form (form 2)

- Name
- Email Address
- Research topic (max. 20 words)
- Letter of motivation (max. 300 words)
- Basic information
  - Where are your data from?  
Answers:
    - I collect data myself
    - I use data sets from other researchers (e.g. open data)
  - What kind of data do you use? (max. 20 words)
- Software skills
  - What experience level do you have using R software?  
Answers:
    - None (never used before)
    - Basic (I know how to import & handle data)
    - Advanced (I know how to use packages & visualize data)
    - Professional (I am writing & publishing my own scripts)
  - Software skills: Are you familiar with R Studio? (Yes - No)
  - What experience level do you have using Python software?  
Answers:
    - None (never used before)
    - Basic (I know how to import & handle data)
    - Advanced (I know how to use packages & visualize data)
    - Professional (I am writing & publishing my own scripts)
  - Are you familiar with Jupyter Notebook? (Yes - No)

- Data skills
  - Did you already publish data? (Yes - No)
  - Which databases do you use for publishing data? (max. 20 words)
  - Which databases do you use for searching data? (max. 20 words)
  - Did you ever download open biodiversity data? (Yes - No)
  - Which portal did you use for downloading data? (max. 20 words)
  - Are you familiar with the Darwin Core data standard? (Yes - No)

## E. Anonymous survey for course evaluation

The first eight questions were single-choice questions, and the last three questions required text answers. We circulated the link to the survey during the warm-up session of the last day of the course, and again in the follow-up email two weeks after the course.

### Preparation:

- Would you have liked more information in advance?  
Answers: Yes - No

### During the Winter School:

- ... the aims and structure were clear  
Answers: I do not agree - I rather disagree - I rather agree - I fully agree
- ... the lecturer did use their time pretty well  
Answers: I do not agree - I rather disagree - I rather agree - I fully agree
- ... the lecturers were always open for questions and further assistance  
Answers: I do not agree - I rather disagree - I rather agree - I fully agree
- ... the lecturer showed expertise and were thematically well prepared  
Answers: I do not agree - I rather disagree - I rather agree - I fully agree

### In general:

- The time frame of the Winter School was appropriate  
Answers: I do not agree - I rather disagree - I rather agree - I fully agree
- How satisfied are you in terms of the Winter School?  
Answers: I do not agree - I rather disagree - I rather agree - I fully agree
- Would you recommend the Winter School to others?  
Answers: Yes - Perhaps - No
- Do you have any suggestions on how we could improve the Winter School?  
Answers: Text
- Can you name one aspect that you particularly liked?  
Answers: Text
- What were your expectations when you attended the workshop? Were they fulfilled?  
Answers: Text