Fakulta rybářství
a ochrany vod
Faculty of Fisheries
and Protection
of Waters

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice
Czech Republic

Ref. No. JU/09/02022/24 Vodňany,17. 6. 2024

Dean´s Measure No. 15/2024
Experimental Data and Metadata Processing by Doctoral Students

The Measure regulates the handling of experimental data and metadata that are created as part of the solution of a doctoral thesis in order to comply as much as possible with the principles of open science and the FAIR standard.

1. Definition of data and metadata, open science and the FAIR standard

Experimental data is understood as all data that resulted from measurements, data processing, or data analysis carried out by a Ph.D. student as part of his dissertation work. Metadata is then descriptive data for these experimental data, which store all important information for the reproducibility of the performed experiments (processing, analyses).

Open science is a term given to a set of concepts and practices that ensure the openness of the scientific process. It represents a new, modern way of conducting research and providing access to its results and research data. This is done through new digital technologies and tools, leading to enhanced scientific collaboration. Open science strives for the availability of scientific results that are reusable as well as transparency, which is a huge asset for scientific integrity at all stages of the scientific process. It emphasizes collaboration, the principle of free sharing, and the reuse of all scientific outputs. At the same time, scientific information should be available throughout all stages of the data life cycle, which helps to disseminate results more effectively.

The FAIR standard is composed of four basic data and metadata requirements

- Findable – the data and associated materials must be stored in a reliable and appropriate location, have sufficiently detailed descriptive metadata, and have a unique and persistent identifier, such as a Digital Object Identifier (DOI).
- Accessible – all metadata and data should be understandable to humans and computers under clearly defined conditions. Data should be stored in a trusted repository and freely downloadable. If data cannot be made available, metadata should be openly available.
- Interoperable - Metadata should use standardized formats and languages, standards for data description, and controlled vocabularies. It is also important to link between the dataset metadata and the metadata of other related publications and outputs, as well as to link to authors, institutions, other projects and outputs.

- Reusable data is provided with the least restrictive license that allows reuse; they have rich metadata descriptions, clear instructions and information on accessibility, provenance, and how to obtain the data. Standards common in the field are respected.

More detailed information about Open science and the FAIR standard can be found here: https://www.lib.jcu.cz/cz/otevrena-veda

More information about data, metadata, data management plan, and data management can be found in the course record on data management for Ph.D. students: the link to the video and presentation from the course.

2. Duties of a Ph.D. student

The student is required to manage all experimental data and metadata so that the person who will be authorized to work with the data is able to track the data, display it using standardized software tools, and be able to reproduce the retrieval of the data using metadata.
The student is obliged to create metadata for the measured or processed data so that the experiment or data processing can be reproduced. The metadata must conform to the conventions of the laboratory and the standards used in the student's field of research.
The student is obliged to store data and metadata in a secure repository and ensure their backup (storing data in at least two physically separate repositories), for example using the tools listed in point 4. If the student does not back up the data and it is lost, then the student will be financially involved in the restoration of this data, according to the decision of the supervisor and the head of the laboratory.
The student is required to ensure that data and metadata are shared with his/her supervisor and any persons with authorized access designated by the student's supervisor or lab manager. The student's supervisor has the right not to require data sharing.
Data and metadata may only be deleted with the approval of the student's supervisor and the lab manager.
If the student uses a service for storing data and metadata that might be unavailable after the completion of his/her doctoral studies (university OneDrive storage, web storage requiring login credentials for access), the student is obliged to store the data and metadata on storage accessible to the supervisor or provide login credentials for accessing the data.

3. Duties of a supervisor

The supervisor, through the laboratory and the faculty, is required to provide an adequate repository for storing and archiving the Ph.D. student's data.
The supervisor will provide training for the student in open science and the FAIR standard.
The supervisor shall inform the student of the storage practices, data backup, and metadata creation in the respective lab.

The supervisor is required to regularly check the student's compliance with the data management policy defined in this Measure. A record of the check will be made by the supervisor in the student's annual report. The supervisor shall use the services of a data steward to consult on the setup of the tools, to check the setup and compliance with the principles resulting from this Measure.

4. Recommended milestones

- Student training in data management – 1st-2nd year of study
- Creation of a data management plan - 1st-2nd year of study
- Ensuring sufficient internal storage capacity for data - 1st-2nd year of study
- Selecting an appropriate repository for open data – 2nd-3rd year of study
- Opening selected data – 3rd-4th year of study
- Handing over access to data to the supervisor - two months before graduation

5. Preferred tools and procedures

5.1. Data storage and backup

- In the case of internal data storage, the recommended tool is to store the data on the student's personal computer disk or the laboratory´s or institute´s server. For backup, it is recommended the following: the university's OneDrive (cloud-based data storage service), which is available to all students (contact the Faculty IT department for information about the service), or the ownCloud service operated by CESNET (attachment No. 1 - cloud-based data storage service with a 100GB limit - https://owncloud.cesnet.cz/). Another recommended archiving tool is the bioWES system (attachment No. 2 - http://www.biowes.org), which allows both storage and management of data and metadata. If the student stores the data on a server that provides data backup, then no further backup is necessary.
- for the case of opening datasets, it is recommended to use field or general repositories compliant with FAIR principles (Zenodo - https://zenodo.org/ Mendeley Data - https://data.mendeley.com/). To find a suitable repository for storing and publishing scientific data, the international data repository registry www.re3data.org can be used. For each repository entry, a series of pictograms can be found that provide information about the repository's facilities. For example, whether they are open access, whether they use a DOI identifier, etc.
- For more information on how to open datasets, contact the faculty data steward.

5.2. Data and metadata management

For the management of experimental data and the creation and management of metadata, it is recommended to use a system of so-called e-notebooks, which allow the

description of experimental work and the management of records. BioWES is recommended as an internal management system. The following external systems are recommended:

- https://www.elabftw.net/
- https://openbis.ch/
- https://www.bikalims.org/

For more information about the tools, contact the faculty data steward.

5.3. Data management plan

As a supporting tool for data and metadata management, it is recommended to create a data management plan that describes what data will be generated within the doctoral thesis, who can handle the data and how the FAIR principle will be ensured in handling the data. To create a data management plan, contact the faculty data steward or use the Data Stewardship Wizard tool (Appendix no. 3 - https://ds-wizard.org)

5.4. Training

Basic information on open science, data management, FAIR principles, and the development of a data management plan can be obtained at the annual (May) open science course in the Bioinformatics course. Video recording of the course: link to the course video and presentation.

Exceptions are decided by the Dean.
This Measure takes effect on 17th June 2024.

Prof. Dip.-Ing. Tomáš Policar, Ph.D.
Dean of FFPW USB

Attachment No. 1 CESNET ownCloud - instructions
Attachment No. 2 Data Stewardship Wizard - instructions
Attachment No. 3 bioWES - instructions

Attachment No. 1

CESNET ownCloud

OwnCloud is a cloud storage service suitable for smaller amounts of data that is also available via a web browser (an academic alternative to Dropbox, Onedrive, etc.).

Web interface: https://owncloud.cesnet.cz.

- It allows you to automatically synchronize data between your computers, mobile devices or share it with others.
- 100 GB of data space.
- Free synchronizing client for Linux/Windows/Mac OS + paid app (approx. 20 CZK) for mobile platforms Android, iOS, WindowsMobile (unofficial release)
- Possibility to connect as a disk via WebDAV e.g. in Windows with the command "net use r: https://owncloud.cesnet.cz/remote.php/webdav/ /user:cislo@cuni.cz

Login to ownCloud

1. Since 07/2018 the method of logging into the CESNET e-infrastructure has changed (until now it directly used the identities of users from the federation).
2. The login procedure is described in the instructions on https://du.cesnet.cz/cs/navody/owncloud/start#prvni_prihlaseni_k_e-infrastrukture_cesnet

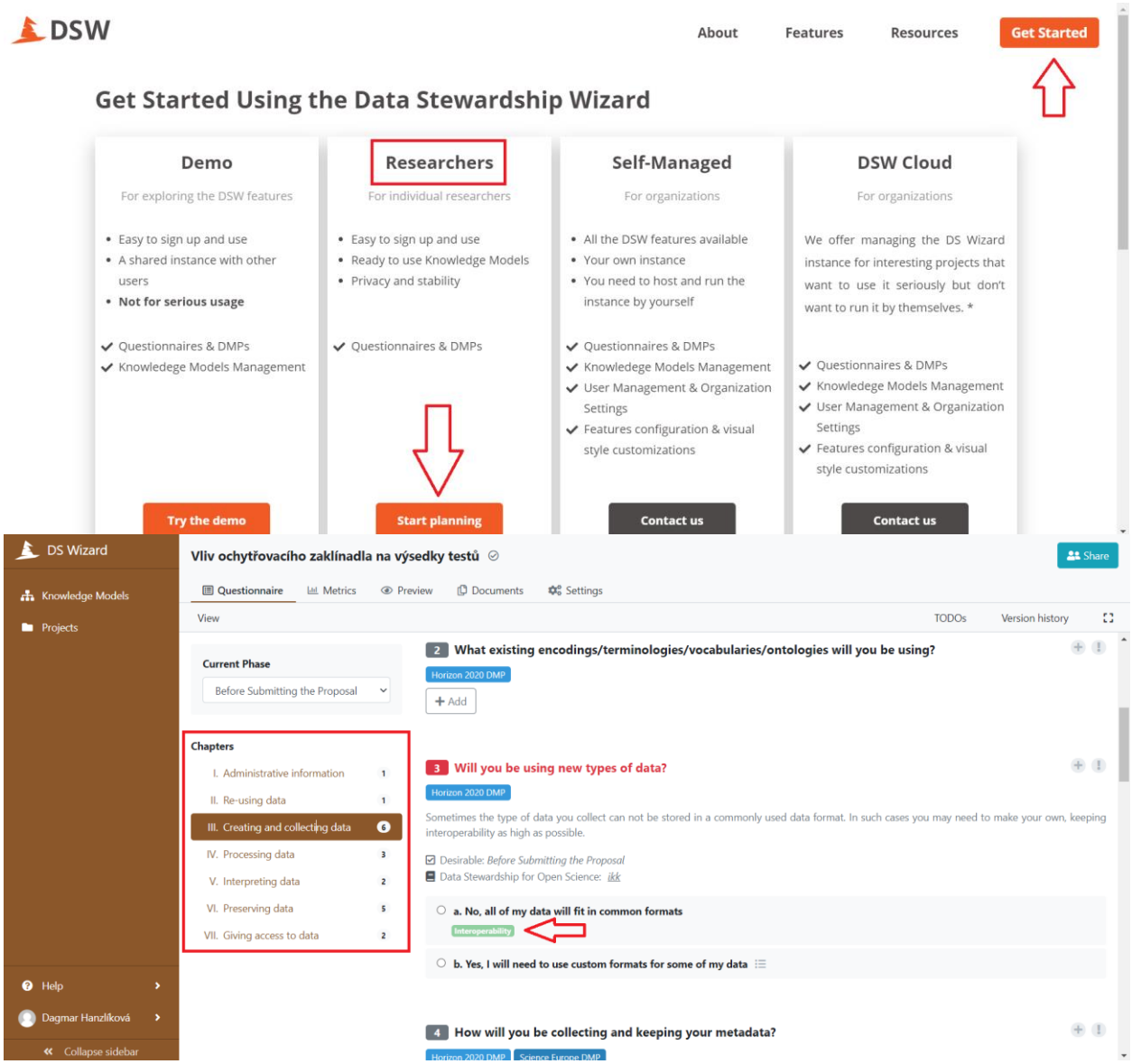
Setting up the ownCloud desktop storage client

1. Next, the client's registration is changed. He/she will no longer use a special name and password, but log in via the web interface. Clients will therefore need to re-register. The WebDAV protocol has specific passwords. Again, we refer to the instructions:
   https://du.cesnet.cz/cs/navody/owncloud/start#desktopovy_klient
   https://du.cesnet.cz/cs/navody/owncloud/start#pristup_pres_protokol_webdav
2. After the launch of the system, some mobile clients experienced a problem (Android, iOS), where registration to the user account failed. CESNET is working intensively on a solution and will keep us informed about the developments.

For more info and detailed help on ownCloud, see the instructions on the ownCloud website. Please read the CESNET data storage service usage policy before using it.

Attachment No. 2

Data Stewardship Wizard User Manual



Introduction

Data Stewardship Wizard is a comprehensive tool designed to streamline your organization's data governance, data quality assurance, management, and compliance processes. This guide will walk you through the key features of the Data Stewardship Wizard and enable you to effectively use it to its full potential.

Access to the guide

1. Login: Enter your login credentials on the login page to access the Data Management Wizard.
2. Dashboard: After logging in, you will be redirected to a dashboard that provides an overview of your data management tasks, alerts, and recent activity.

Key Features

1. Data quality control
   a) Data profiling: Analyze the quality of your data by generating detailed reports on data patterns, inconsistencies, and anomalies.
      - Steps:
        - Go to "Data Quality".
        - Select the dataset you want to profile.
        - Click on "Profile Data" to start the analysis.
        - Review the generated report to get statistics.

   b) Data cleaning: use cleaning tools to fix errors and standardize data formats
      - Steps:
        - Under "Data Quality", select "Cleaning".
        - Select the dataset and specify the cleaning rules.
        - Apply the rules and check the changes.

2. Data management
   a) Metadata management: maintain a comprehensive metadata repository to document the structure, relationships, and use of your data.
      - Steps:
        - Go to the "Metadata" section.
        - Add or update metadata items for your datasets.
        - Ensure that all items are properly documented with descriptions, data types, and appropriate tags.

   b) Policy enforcement: define and enforce data governance policies to ensure compliance with regulatory requirements.
      - Steps:
        - Enter the "Administration" section.
        - Create or update a management policy.
        - Assign policies to the appropriate datasets and monitor compliance.
3. Cooperation and workflow
   a) Task management: assign, track and manage data management tasks within your team.
      - Steps:

- In the "Tasks" section, create new tasks.
- Assign tasks to team members and set deadlines.
- Monitor the progress of each task through the dashboard.

b) Communication tools: Use the built-in communication tools to discuss data-related issues and collaborate effectively.

- Steps:
  - Use the "Messages" feature to send direct messages to team members.
  - Create discussion threads for specific datasets or tasks.
  - Attach relevant documents and links to your messages.

Advanced features

4. Data tracking
   a) Data Flow Tracking: Visualize the flow of data from its beginning to its final destination to understand the transformations and dependencies of the data.
   - Steps:
     - Go to the "Data Lineage" section.
     - Select a dataset to view its lineage.
     - Explore the visual representation of the data flow and identify any issues.
5. Reporting and analytics
   a) Custom reports: create custom reports to monitor data quality, governance and management activities.
   - Steps:
     - In the "Reports" section, create a new report.
     - Select the metrics and datasets you want to include in the report.
     - Customize the report format and schedule automatic generation if needed.

   b) Analytical panel: Use the analytics dashboard to gain insight into data trends and management performance.
   - Steps:
     - Go to the "Analytics" section.
     - Customize your dashboard to display the appropriate tables and charts.
     - Analyze the data to make informed decisions.

Video with guide: https://youtu.be/gcSPG_dyVUQ?si=EN7D7AIN0cnpJFoT

Attachment No. 3

Data management system bioWES

Installation

1. Go to 160.217.215.250 in your browser
2. Click on Create a new user account
a. Enter the user's email, first name, last name, and create a password
b. Your email and password are your login credentials to the systém

3. Install the log manager (for pre-installed computers, skip this)
a. Download the Protocol Manager installation - http://download.datapartner.cz/data/BioWes/releases/v1.12.213.0/BioWes.msi
b. Run the file
c. Install Protocol Manager in a place where you have full control over the files (create, delete, edit) - on your desktop
Server: 160.217.215.250
Database: dp_biowes_nas
User: ics
Password: Complex0750

Using Log Manager
1. Run the Log Manager - BioWes.exe
2. Log in with your login credentials (obtained from creating a new user)
3. Design a log template
4. Implement the actual experiment
5. Download the data from the server back to your computer

Using the web interface
1. Go to 160.217.215.250 in your browser
2. Log in to the server using your login credentials (obtained from creating a new user)
3. See the details of the experiment
4. Share the experiment
5. Search for a specific experiment

Video instructions:
https://youtu.be/M-H0tWuzkZA?si=ZHHDsB9_U5QVzKzB
https://youtu.be/T6r0Qbdk--0?si=RkEc6iXGWBnrd9D1
https://youtu.be/6Iell5EAQCs?si=tlRThgV4LrqIHI5Z