

BioMedData Deliverable D3.3

Post graduate curriculum for Data Stewards in Norway within the natural sciences

[Federico Bianchini](#), [Nazeefa Fatima](#), [Korbinian Bösl](#)

Funded under project [295932](#) by the Research Council of Norway

There is a growing need for a post-graduate Data Steward curriculum in Norway due to the increasing demand for professionals with expertise in managing and curating data in both academia and industry. Many job advertisements for data steward positions require specialised knowledge in scientific fields, as well as a range of other skills related to policy, research support, and infrastructure. BioMedData is working to address this need by developing programs and initiatives to support the education and professional development of data stewards in Norway.

These efforts aim to equip individuals with the necessary skills and knowledge to effectively manage and utilise data in a variety of settings, thus contributing to the advancement of research and innovation.

This curriculum is the first draft for a post master or PhD certification to bring the additional necessary data stewardship skills to professionals with relevant domain knowledge in a targeted and flexible approach.

Preliminary modules

Course name: Data stewardship in natural sciences

Audience: post-graduates - master's graduates, PhD researchers and early-career researchers within the natural sciences

Language: English

Credit: 15, with 2.5 credits per module and a final 5 credit project (1 Credit = 25h learning time)

Duration: 1 year

Name (ECTS)	Content
1. Introduction to RDM (2.5)	FAIR principles, data life cycle, Data Steward roles, National landscape/resources, legal and ethical requirements, guidelines and policies
2. Fundamental programming skills (2.5)	Software Carpentry: POSIX, git, high-level language: R Python, databases & SQL, Machine Learning in R Python, reproducible computing
3. Data Stewardship of (quantitative) natural science data (2.5)	DMPs, Metadata management system(s), technical compliance with ELSI requirements, data reproducibility, data reuse
4. Role(s) of data steward in practice (2.5)	Service provision, Communication strategies, Stakeholder involvement, project management from a DS perspective, change management
5. Project work (5)	

Name of module: Introduction to RDM

ECTS Credits: 2.5

Language of Instruction: English

Objectives and content:

This module is a comprehensive introduction to the different tasks in data stewardship. It will offer a mixture of topics covering both technical, and legal and organizational aspects about working as a data steward in a national and international-oriented context.

The module will cover following topics:

- Introduction to the FAIR principles
- The steps of the data life-cycle and associated solutions
- The various roles a data steward typically assume (policy, research or infrastructure-oriented) and the ideal interactions between these roles
- The national landscape of available resources, relevant to the various steps of the data life-cycle and compliance to the FAIR data principles. Integration or more general connections to international services will also be discussed

The aim of the module is to:

- Provide a comprehensive introduction to data stewardship and its challenges.
- Explain the FAIR principles in practical terms.
- Provide an overview of the data life cycle with solutions at each stage.
- Describe the possible role of a data steward and related competencies and tasks.
- Inform about national tools of relevance.

Learning Outcomes:

1. Knowledge:
 - a. The candidate will be able to state the FAIR principles and describe the impact of their implementation to research data.
 - b. The candidate will develop a basic understanding of the usage of the technologies behind the FAIR principles (e.g. identifiers).
 - c. The candidate will understand the data life cycle and be able to relate it to one or more specific fields of their competence.
 - d. The candidate will be able to describe and classify the various data management related activities
 - e. The candidate will be able to identify key legal and ethical requirements in the context of research data management in Norway and the European system
2. Skills:
 - a. The candidate will be able to assess the degree of compliance to the FAIR data principles of datasets

- b. The candidate will be able to assign tools and services to a stages of the data life cycle
- 3. General competence:
 - a. The candidate will understand the role of a data steward in a research project
 - b. The candidate will understand how the usage of standardised tools and services can improve data management routines

Recommended previous knowledge:

None, although a general understanding of the research process in an academic environment is assumed.

Credit Reduction due to Course Overlap

–

Teaching and learning methods:

remote/in person:

Compulsory Assignments and Attendance

Form of Assessment

Digital Multiple choice

Examination Support Material

None

Grading scale

Pass / Fail

Reading List:

1. [The FAIR Guiding Principles for scientific data management and stewardship](#)
2. [Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship](#)
3. [Using Personas to Visualize the Need for Data Stewardship](#)
4. [A teaching and training handbook for higher education institutions](#)
5. <https://www.bbmri-eric.eu/elsi-knowledge-base/>
6. https://commission.europa.eu/law/law-topic/data-protection_en
7. <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/>
8. [Indigenous Data Sovereignty and Policy | Maggie Walter, Tahu Kukutai,](#)

Name of module: Fundamental programming skills

ECTS Credits: 2.5

Language of Instruction: English

Objectives and content:

This module provides the technical framework of the activity of a research data steward. It consists of several self-contained sub-modules adapted from workshops developed by the Carpentries (<https://carpentries.org/>). These modules are:

1. Unix shell
2. Version control with git
3. A high level language (R or python)
4. Databases and SQL
5. A hands-on introduction to Machine learning
6. Notions of reproducible computing (package management systems, workflow management system)

Due to the nature of this module, a considerable amount of hands-on activity is proposed to the learners, including guided exercises

Learning Outcomes:

1. Knowledge:
 - a. The candidate will understand the formal logic behind programming languages.
 - b. The candidate will recognise the importance of version control for collaborative working
 - c. The candidate will classify the challenges associated with reproducible computing
2. Skills:
 - a. The candidate will develop a working understanding of the six sub-topics covered in this module
 - b. The candidate will apply these basic skills to solve exemplary exercises
 - c. Candidates will be able to share code between each other and work collaboratively

Recommended previous knowledge:

None

Credit Reduction due to Course Overlap

This module can be entirely omitted if the candidate has already obtained the required competencies from other sources, including the Carpentries workshops. Due to the natural division of the module in sub-topics, a candidate should be able to pick only the lessons of interest. Each sub-module, in this course, not taken will reduce the credit amount by 0.5

2.5 ECTS Reduction due to Course Overlap:

[INF100](#)

[INFO132](#)

[DATA110](#)

INF109;

MIX101

[INF620](#)

[INF630](#)

[LING123](#)

Teaching and learning methods:

remote/in person:

Frontal lecture with demonstration and guided exercises. Collaborative coding between participants.

Compulsory Assignments and Attendance

Attendance and participation in the guided exercises is compulsory.

Form of Assessment

Practical (digital) examination (programming).

Examination Support Material

All written and printed examination aids are allowed.

Grading scale

Pass / Fail

Reading List:

1. <https://swcarpentry.github.io/shell-novice/>

2. <https://swcarpentry.github.io/git-novice/>
3. <https://swcarpentry.github.io/python-novice-inflammation/>
4. <http://swcarpentry.github.io/python-novice-gapminder/>
5. <http://swcarpentry.github.io/r-novice-inflammation/>
6. <https://swcarpentry.github.io/r-novice-gapminder/>
7. <https://swcarpentry.github.io/sql-novice-survey/>
8. <https://carpentries-incubator.github.io/deep-learning-intro/>
9. <https://carpentries-incubator.github.io/r-ml-tabular-data/>

Name of module: Data Stewardship of (quantitative) natural science data

ECTS Credits: 2.5

Language of Instruction: English

Objectives and content:

In this module, we will address more practically some of the points raised in module 1.

We will focus on one of the main ingredients for FAIR data, i.e. metadata. We will describe typical technological implementations of metadata standards, including JSON(-LD) schemas. We will, moreover, discuss metadata management systems. We will introduce semantic annotation of metadata via ontologies and discuss the integration of these in the scientific workflow and into datasets.

Finally, we will focus on integrating all the information from this and the previous modules into a data management plan. We will discuss the scope of DMP in a research context and discuss machine-actionability.

Learning Outcomes:

1. Knowledge:
 - a. The candidate will understand the pivotal role of metadata standards for data FAIRification.
 - b. The candidate will familiarise with the concept of metadata validation and, more in general, machine-actionability.
 - c. The candidate will understand the meaning of semantic annotation and its implication towards FAIR and AI.
 - d. The candidate will understand technical implications and representation of legal and ethical requirement of scientific projects
2. Skills:
 - a. The candidate will operate metadata schema in a semantic format such as JSON(-LD) or RDF.
 - b. The candidate will organise metadata using a management system.
 - c. The candidate will annotate metadata using ontologies
 - d. The candidate will search datasets using semantically enhanced search systems
 - e. The candidate will generate a data management plan
3. General competence:
 - a.

Recommended previous knowledge:

The previous two modules (or analogue training) are required. High-level knowledge of a scientific domain of relevance is assumed.

Credit Reduction due to Course Overlap

Teaching and learning methods:

remote/in person:

Compulsory Assignments and Attendance

Form of Assessment

Practical examination – DMP and organisation of semantically annotated metadata

Examination Support Material

Grading scale

Pass / Fail

Reading List:

1. Data Stewardship for Open Science, Barend Mons 2018, Chapman & Hall ISBN 9781032095707
2. [Semantic Web - W3C](#)
3. [Understanding the Data Management Plan as a Boundary Object through a Multi-stakeholder perspective | International Journal of Digital Curation](#)
4. [RDA DMP Common Standard for Machine-actionable Data Management Plans | Zenodo](#)
5. [GA4GH Data Security Toolkit](#)

Name of module: Role(s) of data steward in practice

ECTS Credits: 2.5

Language of Instruction: English

Objectives and content:

In this module, the focus is on the environment where a data steward is expected to work and the various interactions associated with it. We will discuss and provide examples about service provision. We will discuss communication strategies and how to involve stakeholders. In this context, data stewards often act in coordinating and moderating roles. We will, finally, discuss project management and how a data steward can support project managers and promote organisational change via the adoption of better practices and how these practices can be formalised into policies.

Learning Outcomes:

1. Knowledge:
 - a. The candidate will understand the basic interaction of a data steward with researchers and support personnel within a project.
 - b. The candidate will define communications strategies.
 - c. The candidate will compare ideal scenarios to its real-life experience
 - d. The candidate will appraise the condition of the real-life experience and start designing a strategy for improvement
2. Skills:
 - a. The candidate will map stakeholders
 - b. The candidate will demonstrate communications strategies
 - c. The candidate will create a gap-analysis based on material from this course and real-life experience form e.g. a certain research field.
 - d. The candidate will create new policies based on this gap-analysis.
3. General competence:
 - a.

Recommended previous knowledge:

All the previous modules are required

Credit Reduction due to Course Overlap

Teaching and learning methods:

remote/in person:

Compulsory Assignments and Attendance

Form of Assessment

Oral examination

Examination Support Material

none

Grading scale

Pass / Fail

Reading List:

1. [Engaging Researchers with Data Management: The Cookbook](#)

Name of module: Project work

ECTS Credits: 5

Full-time/Part-time:

Language of Instruction: English

Objectives and content:

In this final module the candidates are expected to demonstrate they have acquired the competence for working as a data steward either in individual projects or in groups. In this project, the candidates will pick a project (suggestions for projects might be provided from the instructors)

Possible projects include:

1. End-to-end analysis of support needs of a research project:
 - Analysis of the typical data, metadata and controlled vocabularies/ontologies
 - Tools used to process the data and metadata in the project
 - Data Management Plan in human and machine actionable format
 - Analysis of the dataflow
 - Roles and stakeholders in the project from a data perspective
 - Relevant ELSI
 - Data sharing and preservation strategies
2. Conceptualization and development of data management policies for organisations
 - Review of the organisational infrastructure
 - Review of existing legal and ethical framework and relevant existing organisational policies
 - Alignment of strategies and goals of organisation, funders and governmental bodies
 - Formulation of policy documents and examples for practical guidance for implementation
 - Registration of standards and repositories included in policies in relevant FAIR indices
3. Conceptualization of a training unit for end-users and creation of training material
 - Identification of target audience, learning objectives and outcomes
 - Review of existing open training material
 - Conceptualization of schedule, exercises, lectures and necessary related material
 - Creation and provision of the material under consideration of universal design
 - FAIR Deposition and indexing of the material

Recommended previous knowledge:**Credit Reduction due to Course Overlap****Teaching and learning methods:**

remote/in person:

Compulsory Assignments and Attendance**Form of Assessment**

The candidate will present the results of the project in a written format and discuss them as part of the final oral examination.

Examination Support Material**Grading scale**

Pass / Fail

Reading List:

1. [Practical Guide to the International Alignment of Research Data Management - Extended Edition | Zenodo](#)
2. [Guidance Document Presenting a Framework for Discipline-specific Research Data Management | Zenodo](#)