End-User Course on FAIR Data Management for Life Sciences Research in Norway

Organised by:

BioMedData Deliverable D4.1 - End-User Course on FAIR Data Management for Life Sciences Research in Norway

About BioMedData

The **BioMedData** project, initiated by ELIXIR Norway, aims to connect the largest research infrastructures (RIs) within the Life Sciences in Norway with the scope of improving their data management practises, aiming at the production of findable, accessible, interoperable and reusable (FAIR) data. In this context, ELIXIR Norway has a coordinating role and each of the other RIs participates in the BioMedData activities through a dedicated Project Area Liaison (PAL). The infrastructures contributing to this course are:

- **ELIXIR Norway** The national node of the infrastructure for biological information
- **NorSeq** The national consortium for sequencing and personalised medicine
- **NOR-Openscreen** The consortium for chemical biology and early drug discovery.
- **NORMOLIM** The Norwegian Molecular Imaging Infrastructure
- **NALMIN** The Norwegian Advanced Light Microscopy Imaging Network
- **NORCRYST** The Norwegian Macromolecular Crystallography Consortium
- **BBMRI-NO** The national RI for clinical and population-based biobanks.
- **MoBa Genetics** The Norwegian Mother, Father and Child Cohort Study
- **GBIF Norway** The Norwegian node of the Global Biodiversity Information Facility

This deliverable aims at describing and defining the contents of the first BioMedData end-user course, to take place in spring 2022. This activity builds upon two pillars: the existing training activity on Data Management Planning currently offered by ELIXIR Norway and Digital Life Norway and the report on data management gaps published by the BioMedData participants in March 2021. The latter provides a set of field-specific data management recommendations that enrich the content of the former, thus providing the end-users with concrete knowledge on FAIR data based on their domain.

The course aims to cover topics covering the whole data life cycle (as shown in the Research Data Management kit by ELIXIR Europe) with a specific focus on the implementation of the FAIR data principles.

Course Information

The course is expected to be delivered online over a period of three weekdays. The target audience includes master's students, PhD candidates, postdoctoral fellows, early-career researchers, and technical staff involved in scientific research in the life sciences in Norway. The course is particularly relevant for users and/or collaborators of the research infrastructures listed on the previous page.

Learning Objectives:

A. To walk through data life cycle steps with an introduction to tools relevant to life sciences research, including in a context of computational analysis for life sciences.
B. To understand FAIR principles in the context of each data cycle step
C. To understand the data curation processes that enable data reuse and repurposing
D. To know the different RDM guidelines and practices used by the various RIs and institutions in Norway.

Learning Outcomes:

A. To be able to define and connect RDM tasks within a domain of interest.
B. To be able to determine different tools that are available, globally and in Norway, to carry out RDM tasks.
C. To be able to identify metadata standards and to associate relevant deposition repositories
D. To be able to analyse and evaluate FAIRness of Data (and Management Plan)

Learning Activities:

The course will begin with an interactive session to get to know the audience. Each session will include time for a short discussion including questions and answers. In addition, we aim to include a quiz and/or an exercise involving group activity relevant to each session. Learners are recommended to take a quiz during or after the course to help revise and understand the content learned throughout the course.

Course Curriculum

The course is structured based on data life cycle steps that serve as sessions, and each session includes topics that serve as modules.

Introduction

All the research infrastructures represented in BioMedData will provide a short introduction about their activity. This will help the attendants to identify the contact/reference person for their scientific domain.

Data Planning

The session will cover an introduction to the FAIR principles and how the FAIRification process works, what to consider for data management planning, and relevant national and global resources (such as Data Stewardship Wizard) that could be used to create a FAIR data management plan (DMP). The learning activities for this session will include reviewing

sample DMPs as a team, and a hands-on session dedicated to getting familiar with the Norwegian instance of DSW.

Data Collection

The session will provide practical information about best practises for data collection through some examples from sequencing, imaging and field observations. We will touch upon data safety and security and present the national sources addressing these requirements with particular focus on the Norwegian e-Infrastructure for Life Sciences (NeLS) and the National e-Infrastructure for Research Data (NIRD) provided by ELIXIR Norway and Sigma2 respectively.

There will be an additional part of this session dedicated to metadata management, that will include topics such as identification of metadata, types and examples of metadata in different life sciences datasets, and metadata standards.

Data Processing

The topics covered in this session will be data wrangling (including an introduction to computational tools that allow data conversion to human-readable format), national services that can help with processing of sensitive data, ontologies, and controlled vocabularies.

Data Analysis

The session will provide insights on local (high-performance) computing resources available at institutions across Norway, and an introduction to computational work-bench called Galaxy for which

Data Archiving

The course user will learn the core requirements for data archiving and get to know the most frequently used ELIXIR data deposition databases, as well as national and institutional archiving solutions. The session will also cover the general concepts of the central legal obligations.

Data Sharing

The course user will learn about dataset registrations in Current Research Information Systems (CRIS) databases, the concepts and general requirements of controlled access, standard licences for software and datasets, as well as versioning of code and its deposition.

Data Reusability

The session will cover information on searching datasets with different tools and strategies and understand the usage of standardised metadata for data discovery. The course users will learn to apply the existing standards to provide data availability statements in their publications.

Data Ethics

The user will know about the existing ethical guidelines, the research ethics act and the health research act, as well as the general implications of the GDPR in research.

Funding Source

The course is funded by the Norwegian Research Council through BioMedData as Forskningsinfrastruktur 295932