BioData.pt elixir PORTUGAL

Ready for BioData Management?

Introduction To

Data Management Plans

Funding:

- ELIXIR Implementation Study "Impact evaluation at Node-level - getting it done"
- ELIXIR-CONVERGE Project “Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services”

BioData.pt elixir PORTUGAL

Ready for BioData Management?

Introduction To
Data Management Plans

Demystifying Data Management Plans

Jorge Oliveira
João Cardoso & Daniel Faria

What is a Data Management Plan (DMP)?

Learning Outcome 1:

Recognize the purpose of Data Management Plans

What is a DMP?

- A DMP is a formal document used to **plan and support** data management activities by anticipating **needs and requirements** in a (research) project, facility or institution
- It is the to **data management** what a **blueprint** is to construction

What does it include?

- A DMP should detail policies and methods pertaining to data:
  - Creation / collection
  - Documentation
  - Access
  - Preservation
  - Dissemination
- And ensure an adequate allocation of resources:
  - Human, Computational & Financial

Why Do We Need DMPs?

- The stick:
  - Many **funding** agencies now require that grant proposals be accompanied by a DMP
  - In particular, they require DMPs that demonstrate intent to comply with the **FAIR** data principles.
  - Monitoring of the quality and execution of these DMPs is still light, but expected to tighten

Why Do We Need DMPs?

- The carrot:
  - DMPs are valuable tools in the planning of research activities to **ensure the necessary resources** are devoted to data management
  - Adequate planning can facilitate the task of ensuring compliance of research outputs with the FAIR principles

What Should be in a DMP?

Learning Outcome 2:

List the main topics that should be covered by a DMP

What should be in a DMP?

Administrative & Legal Aspects

- Which **institutions** and **people**:
  - Are involved in the research activities, and what are their roles
  - Are responsible for the execution of the DMP, and what are their roles
- Who are the **contact persons or institutions** for the DMP

Responsibilities & Duties

- Which institutions (and people):
  - **Own** the data
  - Are responsible for **data collection**
  - Are responsible for **data protection** (if applicable)
  - Are responsible for **data security** (if applicable)

Costs & Resources

- What will it cost to:
  - **Analyse** data (hardware, software, man-hours)
  - **Prepare/clean/curate** data (man-hours, maybe software)
  - **Store** data (hardware, maybe man-hours)
  - **Publish** data (publication fees)
  - ...

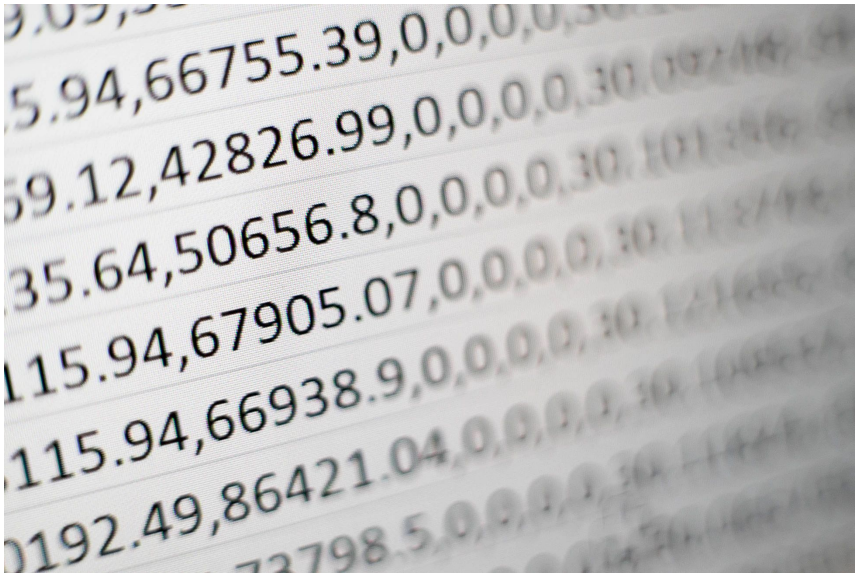
Project Description

- A **summary** of the project proposal:
  - Goals
  - Experiments
  - Methods

(Note: not needed if the DMP is part of a project proposal)

Data

- **Existing** data (if applicable):
  - Data source(s)?
  - Usage licence(s)?
  - Volume of data?
- Data to be **created** or **generated**:
  - How?
  - What types?
  - Volume of data?



Data

- Data **organisation**:
  - How data will be described (**metadata**)
  - How data will be structured (**data formats**)
  - How data will be interconnected (**data structures**)
  - Where data will be **stored** during the project

Data

- **Archiving, sharing** and **publishing** data:
  - Archiving:
    - Where?
    - For how long?
  - Sharing / publishing:
    - Where / how?
    - License?
    - Who can access?
    - Privacy & security?

DMPs: Present & Future

Learning Outcome 3:

Describe the current state and future directions of DMPs

DMPs: Present

- In current practice, DMPs are mainly seen as a bureaucratic hassle
- They are static documents, prepared for grant applications because they are mandatory, but never or rarely updated
- They are generally not validated during the research project, and are never published, which prevents external validation

DMPs: Present

- The fact that different funding bodies use different DMP templates makes it difficult for researchers to get familiar with them and to recognize the value
- Moreover, most templates are free text questionnaires that look more like surveys than planning documents, and are only human readable
- All this results in poor quality DMPs, of low practical value

DMPs: Present

Example: the H2020 DMP template

A questionnaire covering the following topics:

1. Data Summary
   - Describe the data to be acquired/produced
2. FAIR data
   - **Detail** how you'll comply with the FAIR principles
3. Allocation of resources
   - Who does what and what it costs
4. Data security
5. Ethical aspects
6. Other issues

DMPs: Future

- To be of practical use, **a DMP should be**:
  - A living document that is **updated as needed**
  - Both human and machine-readable
  - Comply with a common standard
  - Be shared

DMPs: Future

- The **Machine-Actionable DMP (maDMP)**:
  - **Machine** and **human** readable **descriptions**
  - **Automated** policy enforcement
  - **Shareable**
  - **Interoperable** DMP version
  - **Extensible**

Current DMPs

```
<admindata>
    <question>Who is the Principle Investigator?</question>
    <answer>The PI is John Doe from the JDU</answer>
</admindata>
```

maDMPs

Reuse of existing standards

Use of persistent identifiers

Use of controlled vocabularies

```
"dc:creator":[{
      "foaf:name":"John Doe",
      "@id":"orcid.org/000-1111-2222-3333",
      "foaf:mbox":"mailto:jdoe@jdu.edu",
       "madmp:institution":"JDU-John_Doe_University"
}],
```


DMPs: Future

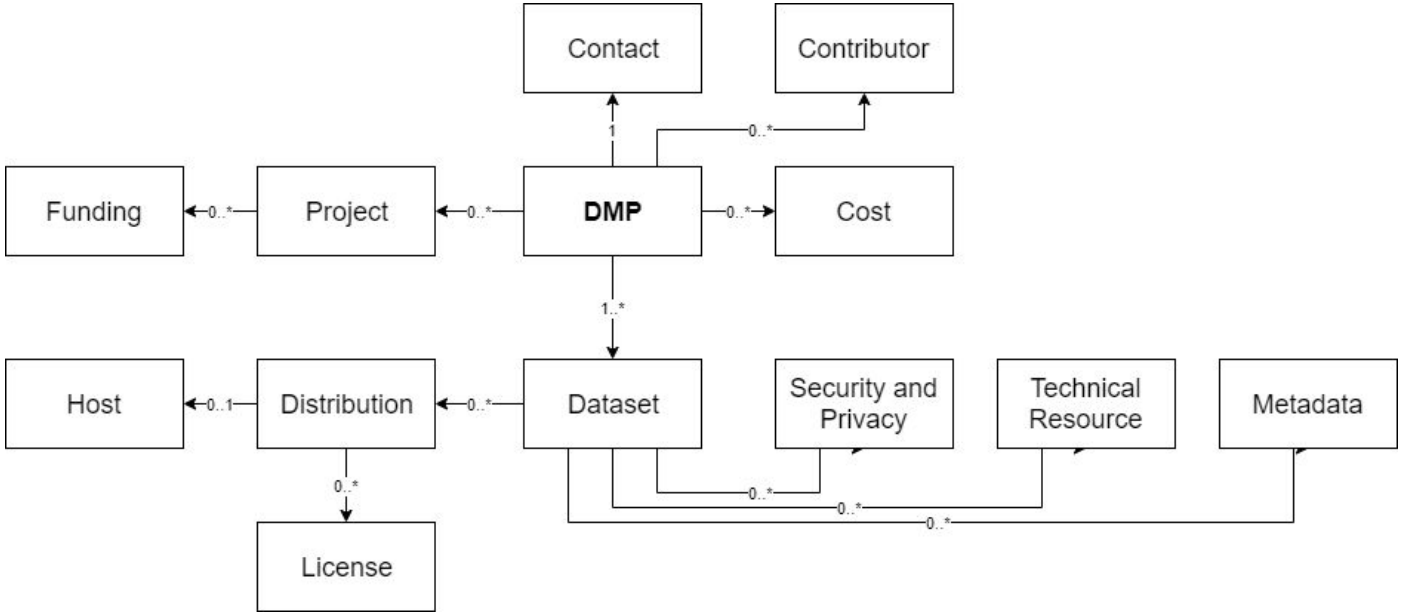
- The RDA DMP Common Standards Working Group was created to focus on the standardization of the knowledge contained in a DMP.
- Its objective was to establish a metadata application standard that defines a core set of elements for a DMP.
- The metadata application standard is modular in design, and allows for extensions.

RDA
RESEARCH DATA ALLIANCE

Scan for more!

DMPs: Future

- A **minimum set of universal terms** to ensure basic **interoperability** of systems using DMPs.



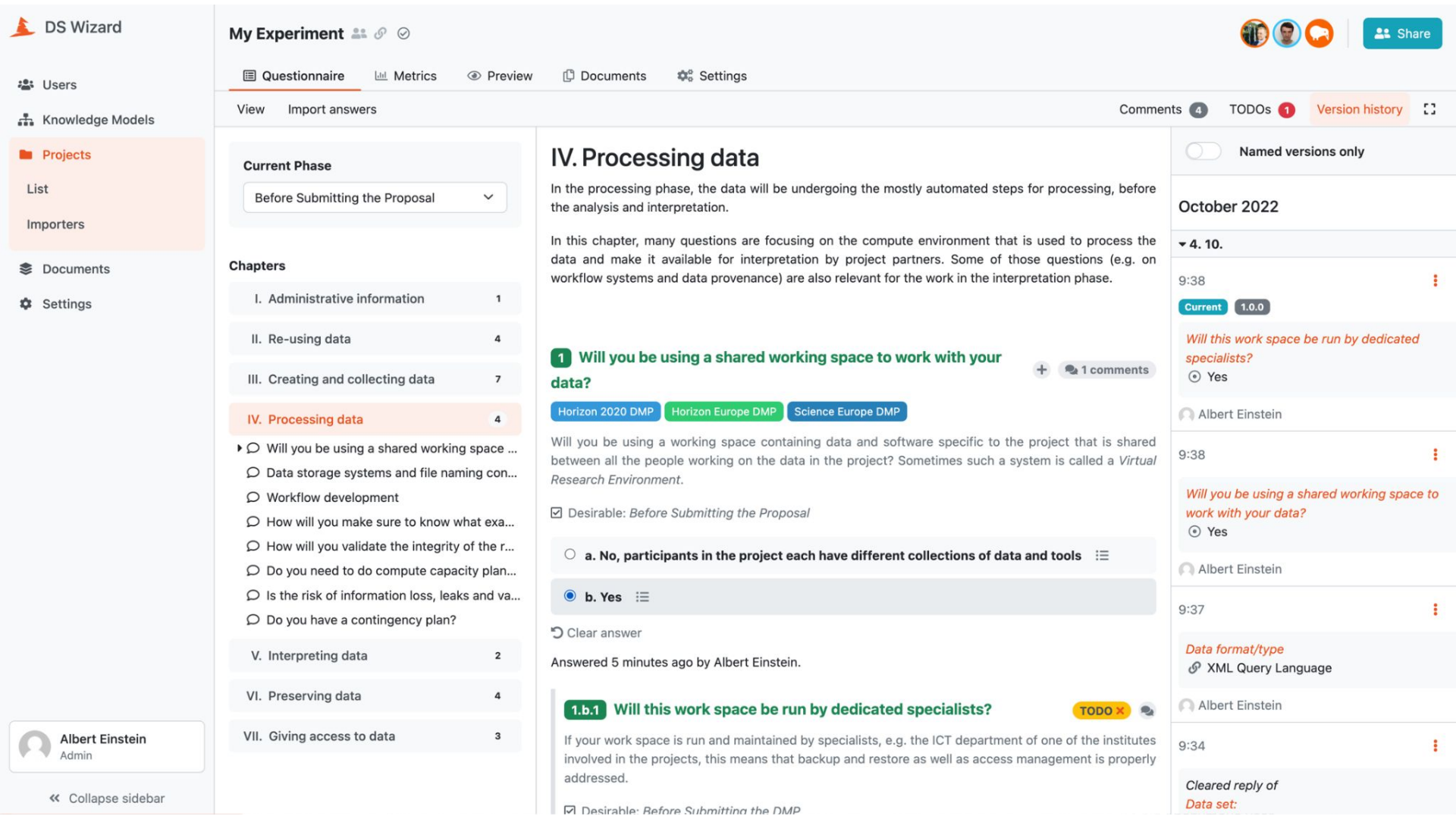
DMPs: Future

- Applications of a maDMP:
  - One DMP for all templates
  - DMP maturity model
  - Automation in both creation and monitoring during the project's life-cycle

Data Stewardship Wizard

(https://biodata-pt.ds-wizard.org/)

- Guided tool to create Data Management Plans
- Real-time Collaborative Support
- Templates (H2020, Science Europe, RDA-machine-actionable and others)
- Developed by ELIXIR and trusted by thousands of people worldwide
- Made with FAIR in mind


DS Wizard
Users
Knowledge Models
Projects
List
Importers
Documents
Settings
Albert Einstein
Admin
Collapse sidebar
My Experiment
Share
Questionnaire
Metrics
Preview
Documents
Settings
View
Import answers
Comments 4
TODOs 1
Version history
Current Phase
Before Submitting the Proposal
Chapters
I. Administrative information 1
II. Re-using data 4
III. Creating and collecting data 7
IV. Processing data 4
Will you be using a shared working space ...
Data storage systems and file naming con...
Workflow development
How will you make sure to know what exa...
How will you validate the integrity of the r...
Do you need to do compute capacity plan...
Is the risk of information loss, leaks and va...
Do you have a contingency plan?
V. Interpreting data 2
VI. Preserving data 4
VII. Giving access to data 3
IV. Processing data
In the processing phase, the data will be undergoing the mostly automated steps for processing, before the analysis and interpretation.
In this chapter, many questions are focusing on the compute environment that is used to process the data and make it available for interpretation by project partners. Some of those questions (e.g. on workflow systems and data provenance) are also relevant for the work in the interpretation phase.
1 Will you be using a shared working space to work with your data?
1 comments
Horizon 2020 DMP
Horizon Europe DMP
Science Europe DMP
Will you be using a working space containing data and software specific to the project that is shared between all the people working on the data in the project? Sometimes such a system is called a Virtual Research Environment.
Desirable: Before Submitting the Proposal
a. No, participants in the project each have different collections of data and tools
b. Yes
Clear answer
Answered 5 minutes ago by Albert Einstein.
1.b.1 Will this work space be run by dedicated specialists?
TODO
If your work space is run and maintained by specialists, e.g. the ICT department of one of the institutes involved in the projects, this means that backup and restore as well as access management is properly addressed.
Desirable: Before Submitting the DMP
Named versions only
October 2022
4. 10.
9:38
Current
1.0.0
Will this work space be run by dedicated specialists?
Yes
Albert Einstein
9:38
Will you be using a shared working space to work with your data?
Yes
Albert Einstein
9:37
Data format/type
XML Query Language
Albert Einstein
9:34
Cleared reply of
Data set:

FAIR Guidance

9 Will you need consent for any newly collected personal data?

- (●) a. No, We do not collect any new personal data
  Reusability
- ( ) b. No, our work on personal data can be done using another legal base
  Reusability
- ( ) c. Yes, we will collect consent for our specific use of the data
  Reusability
- ( ) d. Yes, we will collect consent for our use as well as reuse of the data
  Reusability

FAIR Metrics

Metric	Measure
Findability	0.70
Accessibility	0.22
Reusability	1.00
Good DMP Practice	0.67
Openness	0.69

Other DMP creation tools

DMPONLINE

https://dmponline.dcc.ac.uk/

argos OpenAIRE

https://argos.openaire.eu/

RDMkit (https://rdmkit.elixir-europe.org/)

- Created in ELIXIR-CONVERGE Project by European Experts
- The Research Data Management toolkit for Life Sciences
- **Best practices and guidelines** to help you make your data FAIR:
  - Findable
  - Accessible
  - Interoperable
  - Reusable

Browse all topics by

Data life cycle

Start here to get an overview of research data management based on stages in the data life cycle.

Your role

Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.

Your domain

Learn about data management tasks that affect your domain or research community, and the solutions adopted to address them.

Your tasks

Find guidelines and solutions for tackling common data management tasks.

Tool assembly

Find concrete combinations of tools and resources assembled into an ecosystem for research data management.

National resources

Find pointers to country specific information resources and national research data management practices.

All tools and resources

Browse the RDMkit's catalogue of tools and resources for research data management.

All training resources

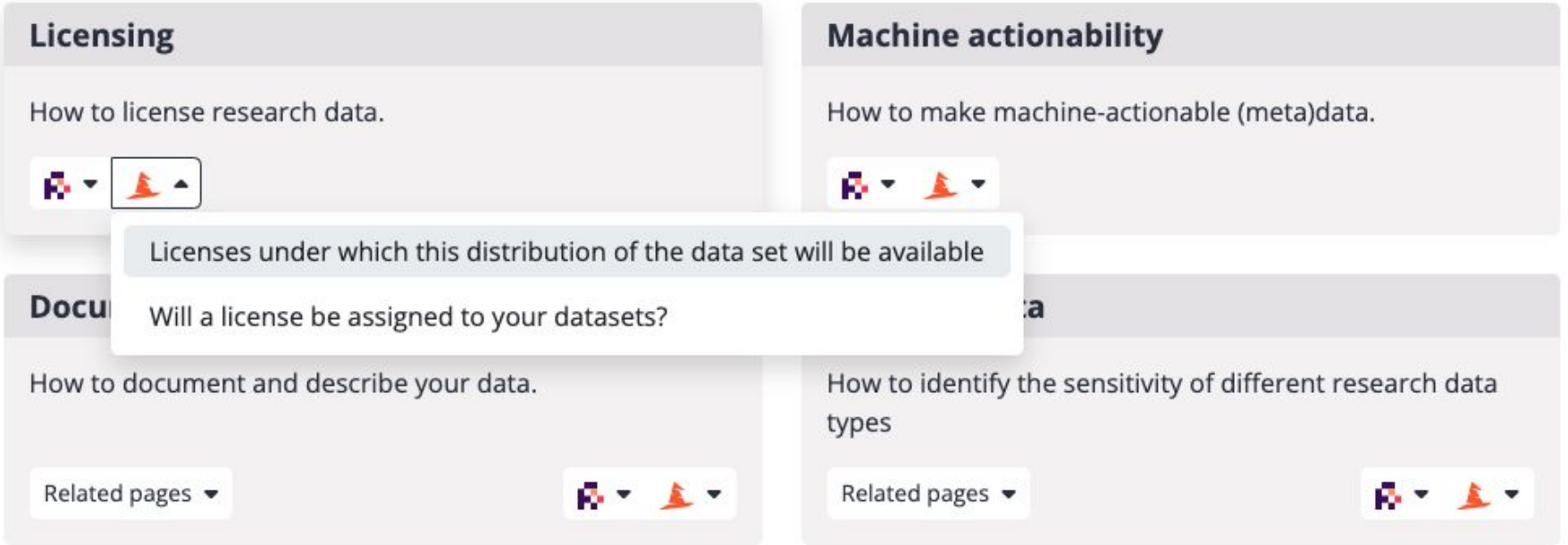
Browse all training resources mentioned in RDMkit pages.

RDMkit

Data Life Cycle

- Descriptions
- Specific considerations
- Links to tasks
- Links to training materials
- Links to DSW

Links to DSW & FAIR Cookbook



The Take Home Message

- The benefits of DMPs:
  - Promote good data management practices
  - Assist in compliance with FAIR data principles
  - Ensure adequate allocation of resources data management activities
  - Enable accountability
- The benefits of maDMPs:
  - Automation (creation, validation, policy enactment)
  - Increase usefulness

Introduction To
Data Management Plans

Questions?

BioData.pt elixir PORTUGAL

Ready for BioData Management?

Introduction To
Data Management Plans

Organization:
Ana Portugal Melo - INESC-ID / BioData.pt
Daniel Faria - INESC-ID / IST / BioData.pt
Jorge Oliveira - INESC-ID / IST / BioData.pt

@BioData_PT

@BioData.PT

@BioData-PT

www.BioData.pt

inesc id lisboa | GULBENKIAN CIÊNCIA | TÉCNICO LISBOA | BioData.pt

elixir CONVERGE

Introduction To
Data Management Plans

Acknowledgements

- ELIXIR Implementation Study "Impact evaluation at Node-level - getting it done"
- ELIXIR-CONVERGE Project “Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services”