

# ELIXIR-CH Data Stewardship Wizard (DSW)



Faculté de biologie et de médecine



## Dr Vassilios Ioannidis

Lead Computational Biologist - SIB/Vital-IT group  
Data Steward, FAIR Specialist - UNIRIS - Direction UNIL

<https://orcid.org/0000-0002-4209-2578>

## Dr Cécile Lebrand

Responsable du service de soutien pour la gestion des publications et des données de recherche FBM UNIL/CHUV - BiUM

Data Steward – UNIRIS - Direction UNIL

<https://orcid.org/0000-0002-2750-3164>

## Dr Grégoire Rossier

Project Manager, SIB/Training/Vital-IT group

<https://orcid.org/0000-0002-0065-5053>



Vital-IT & CHUV Library  
FAIR RDM - 2023

# ACKNOWLEDGMENTS

---



# Outline

## 1. Data Management Plan tools

- a. Data Stewardship Wizard (DSW) Overview and Demo
- b. Data Stewardship Wizard – work with your data
- c. Presentation of user data & discussions

# REMINDER

---

## Research Data management (RDM)

**Data handling** steps during the lifetime of a research project (active phase), but also before (planning) and after (preservation, reuse).

## Data management plan (DMP)

Written **document** describing the process of controlling the information generated during a research project.

# IN A NUTSHELL

---

- Data Management Plan (DMP) tool, but also for Research Data Management (RDM)
- Mind map imagined in 2012 in NL, implemented by ELIXIR-CZ in 2016
- Bring together researchers AND data stewards
- Machine-actionable DMP
- FAIR and other metrics
- Customizable by data stewards (institution-specific)

Knowledge model (KM)

Projects (template)

Document (template)

- **ADMIN**

- Generic knowledge model(s) (KMs)
- Document templates
- Users' management

- **DATA STEWARDS**

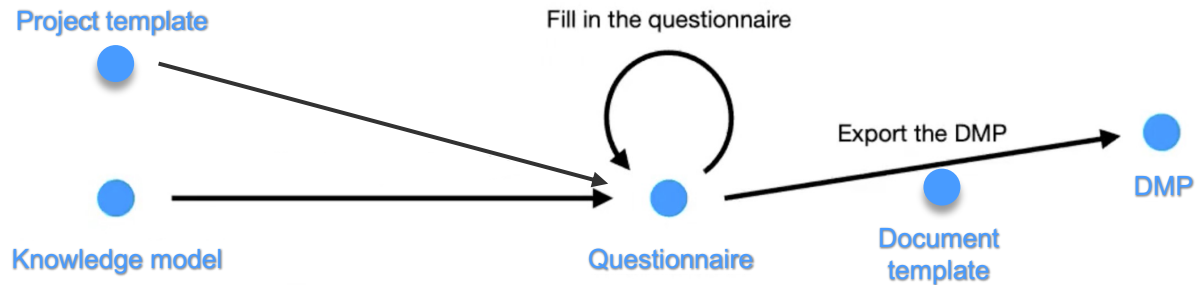
- Edition / customization of KMs
- Project Templates (PTs)
- Help researchers with DMPs

- **RESEARCHERS**

- Create projects (using or not PTs)
- Responsible of DMPs

## Researchers Workflow

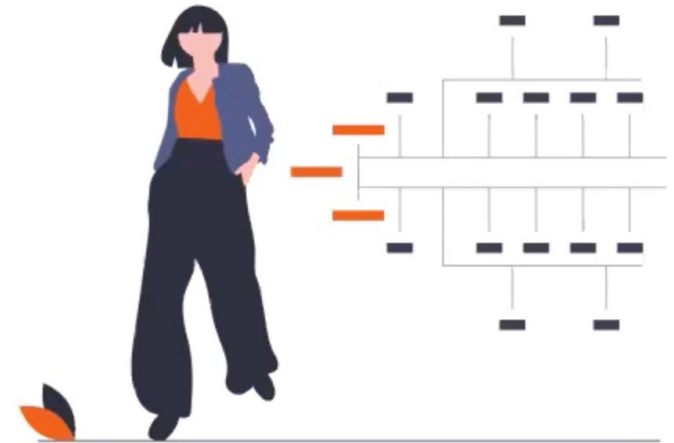
### New project



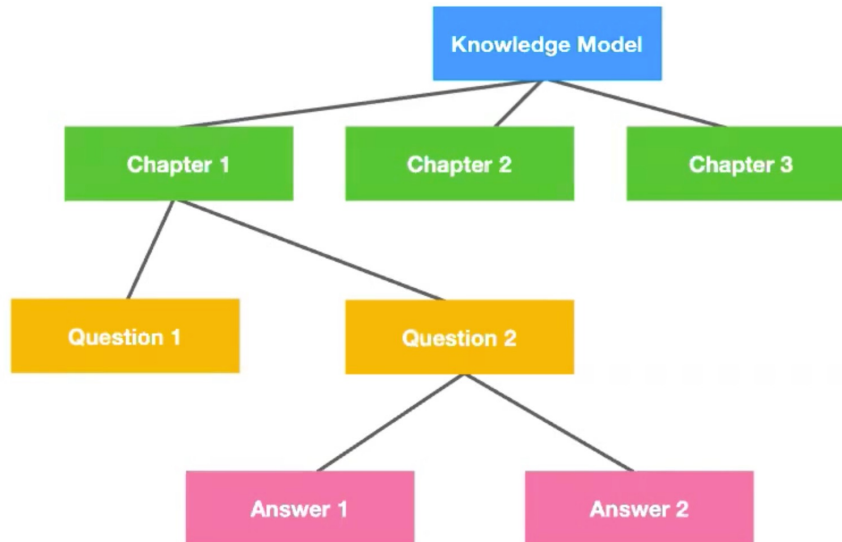


## Knowledge Model

- Contains the **knowledge** about what should be asked and how
- **Tree-like structure** of Chapters, Questions, Answer and References or Experts
- “**Template**” for the Questionnaire
- Provided by DSW or **institution data stewards**



## Knowledge Model



- Core DS Knowledge Model
  - Design of experiment
    - Is there any pre-existing data?
      - No
      - Yes
      - atq
    - Will reference data be created?
    - Will you be storing samples?
    - Will you be collecting experimental data?
  - Data design and planning
    - What data formats/types will you be using?
      - Data format/type:
        - Is this a standard data format used by others too?
        - Does this data format enable sharing and long term archiving?
          - No
          - Yes
        - njy
      - Will you be using new types of data?
      - How will you be storing metadata?
      - During the project, will you be archiving data (using so-called 'cold storage')?
      - Will you need a shared working space to work with your data?
      - Is the risk of information loss, leaks and vandalism acceptably low?
      - Do you need to do compute capacity planning?
    - Data Capture/Measurement
    - Data processing and curation

# KNOWLEDGE MODEL (KM)

## Swiss Biomedical Sciences KM ✓

Publish

Knowledge Model Phases Question Tags Preview Settings

Swiss Biomedical Sciences... Warnings 4

Expand all Collapse all

- Swiss Biomedical Sciences KM
  - Administrative information
  - Re-using data
  - Creating and collecting data
  - Describing data
  - Processing data
  - Preserving data
  - Sharing data
  - Findability
  - Accessibility
  - Interoperability
  - Reusability
  - Good DMP Practice
  - Openness
  - DMP and Research Planning
  - RDM guidance
  - Before Submitting the Proposal
  - Before Submitting the DMP
  - Before Finishing the Project
  - After Finishing the Project
  - Science Europe DMP
  - maDMP
  - Horizon 2020 DMP
  - Horizon Europe DMP

### Knowledge Model

6f6241d9

#### Chapters

- [Administrative information](#)
- [Re-using data](#)
- [Creating and collecting data](#)
- [Describing data](#)
- [Processing data](#)
- [Preserving data](#)
- [Sharing data](#)

[+ Add chapter](#)

#### Metrics

- [Findability](#)

# KNOWLEDGE MODEL (KM) – PARENT / CHILD

---



DSW “LIFE SCIENCES” KM (PARENT)

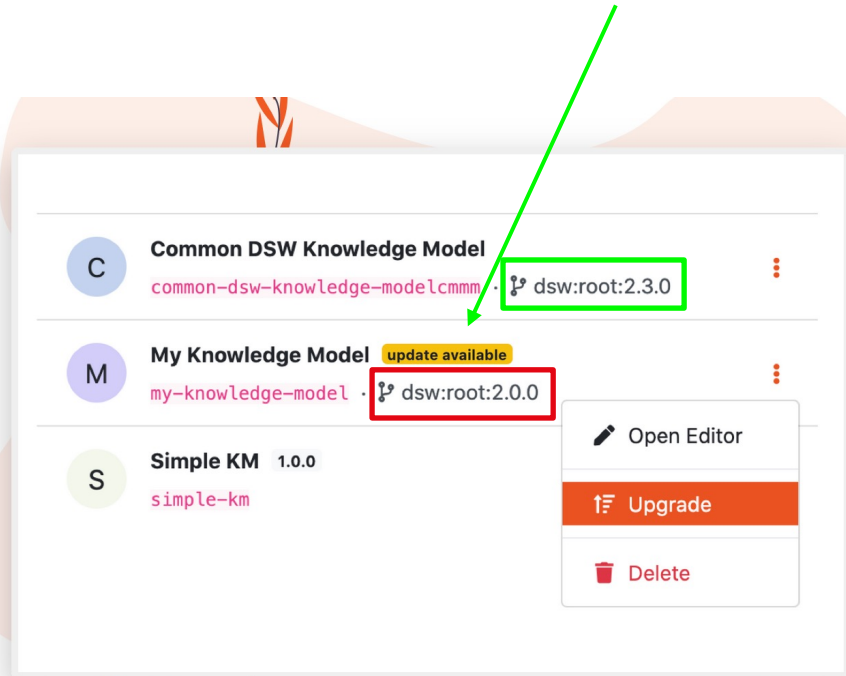


SWISS BIOMEDICAL KM (CHILD)

# KNOWLEDGE MODEL (KM) – SYNCHRO UPDATES

## Evolvability

It is a common practice that Knowledge Models have to change sometimes. When a **parent Knowledge Model** has a new version, its customizations have the option to upgrade so all the Knowledge Models can still be up to date. There is also a very similar process for Questionnaires when their Knowledge Model changes.



## Create Project

 **From Project Template** Custom 

**Name**

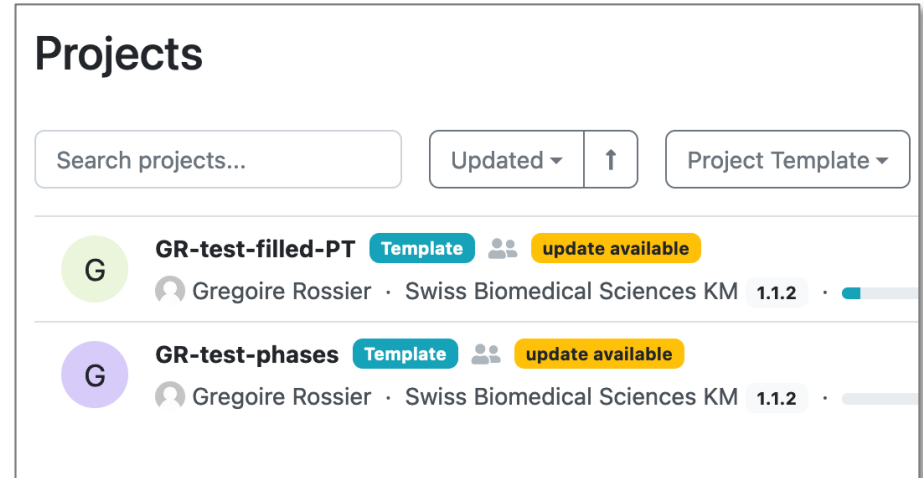
**Knowledge Model**

Cancel



Create



- Selected KM
- Pre-filled answers
- By Data Stewards


A screenshot of a web interface titled 'Projects'. At the top, there is a search bar with the placeholder text 'Search projects...'. To its right are two buttons: 'Updated' with a downward arrow and an upward arrow, and 'Project Template' with a downward arrow. Below these are two project entries. Each entry starts with a circular profile picture containing the letter 'G'. The first entry is 'GR-test-filled-PT', followed by a blue 'Template' badge, a user icon, and a yellow 'update available' badge. Below this is the text 'Gregoire Rossier · Swiss Biomedical Sciences KM 1.1.2' and a progress bar. The second entry is 'GR-test-phases', followed by a blue 'Template' badge, a user icon, and a yellow 'update available' badge. Below this is the text 'Gregoire Rossier · Swiss Biomedical Sciences KM 1.1.2' and a progress bar.

**Projects**


Search projects... Updated ↑ Project Template ▾


**G** **GR-test-filled-PT** **Template**  **update available**  
Gregoire Rossier · Swiss Biomedical Sciences KM 1.1.2 · 

**G** **GR-test-phases** **Template**  **update available**  
Gregoire Rossier · Swiss Biomedical Sciences KM 1.1.2 · 

EPD 

[Questionnaire](#) [Metrics](#) [Preview](#) [Documents](#) [Settings](#)

View [Import answers](#) [TODOs](#) [Comments](#) [Version history](#) 

**Current Phase**  
DMP and Research PI 

**Chapters**

- I. Administrative inf...** 2
  - Contributors
  - Research Project(s)
- II. Re-using data 1
- III. Creating and colle... 7
- IV. Processing data 3
- V. Preserving data 2
- VI. Sharing data 6


## I. Administrative information

### I.1 Contributors

**FBM RDM** **FNS DMP**

Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

Desirable: *DMP and Research Planning*


 Add

### I.2 Research Project(s)

**FBM RDM** **FNS DMP**

Add each of the research project(s) that you are (or will be) working on and for which the data and work are described in this DMP. Give each project a small identifying name for yourself.

Desirable: *DMP and Research Planning*

 Add



## Question Type

### Type

✓ Options

List of Items

Value

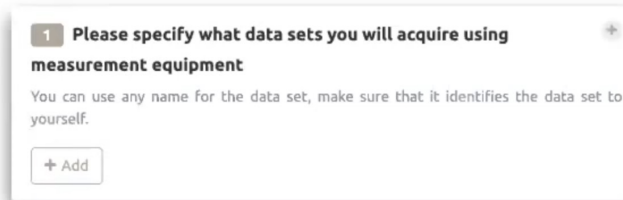
Integration

Multi-Choice



## Question Type: List of Items

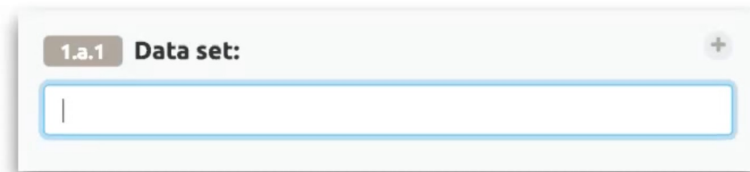
- Multiple **items** of the same type
- Each item has the same set of questions

A screenshot of a single question item from a list. The item has a grey header bar with a small square icon containing the number '1' on the left and a plus sign on the right. The text in the header reads 'Please specify what data sets you will acquire using measurement equipment'. Below the header, there is a paragraph of text: 'You can use any name for the data set, make sure that it identifies the data set to yourself.' At the bottom of the item is a button with a plus sign and the text '+ Add'.

= BLOCK OF QUESTIONS

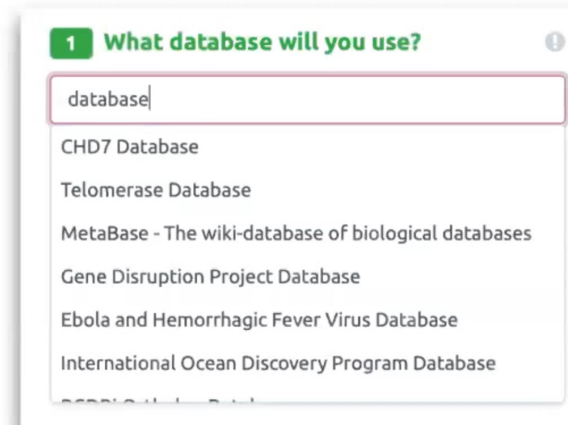
## Question Type: Value

- Simple value that users type
- Value types
  - String
  - Number
  - Date
  - Date time
  - Time
  - Text
  - Email
  - URL
  - Color

A screenshot of a user interface element. It shows a light gray rectangular box with a thin blue border. At the top left of the box, the text '1.a.1 Data set:' is displayed in a small, dark font. To the right of this text is a small gray square containing a white plus sign. Below the text and icon is a wide, empty text input field with a light blue border and a vertical cursor on the left side.

## Question Type: Integration

- Similar to Value type
- The answers can be taken **from external resource**
- More complex to set up

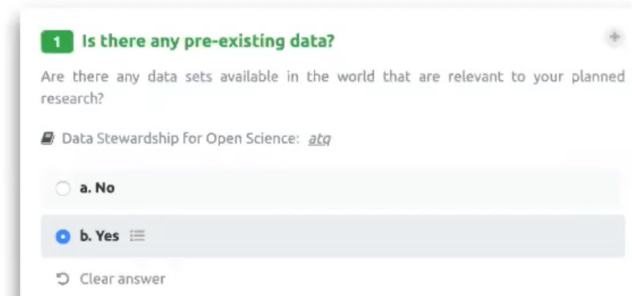
A screenshot of a question interface. At the top, a green box contains the number '1' and the text 'What database will you use?'. Below this is a search input field containing the text 'database'. A dropdown menu is open, showing a list of database names: 'CHD7 Database', 'Telomerase Database', 'MetaBase - The wiki-database of biological databases', 'Gene Disruption Project Database', 'Ebola and Hemorrhagic Fever Virus Database', and 'International Ocean Discovery Program Database'.

1 What database will you use?

- CHD7 Database
- Telomerase Database
- MetaBase - The wiki-database of biological databases
- Gene Disruption Project Database
- Ebola and Hemorrhagic Fever Virus Database
- International Ocean Discovery Program Database

## Question Type: Options

- Closed list of **answers**
- Users can pick one
- Answers can have **follow-up questions**

A screenshot of a question interface from the DSW tool. The question is titled '1 Is there any pre-existing data?' and asks 'Are there any data sets available in the world that are relevant to your planned research?'. The question is attributed to 'Data Stewardship For Open Science: atg'. Two radio button options are shown: 'a. No' and 'b. Yes', with 'b. Yes' selected. A 'Clear answer' button is located at the bottom of the question card.

1 Is there any pre-existing data?

Are there any data sets available in the world that are relevant to your planned research?

Data Stewardship For Open Science: *atg*

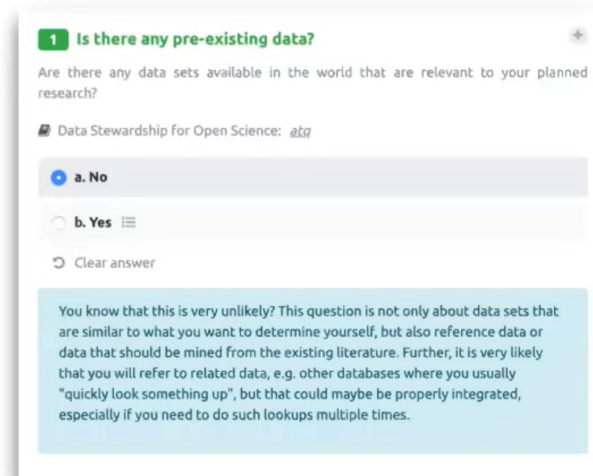
a. No

b. Yes

Clear answer

## Answer Advice

- When an answer is selected, advice can be provided
- Guide users to achieve better results

A screenshot of a survey question from the Data Stewardship Wizard. The question is "1 Is there any pre-existing data?" and asks if there are any data sets available in the world relevant to the user's planned research. The user has selected "a. No". A light blue advice box is displayed below the options, providing guidance on how to interpret the question and where to look for data.

**1 Is there any pre-existing data?**

Are there any data sets available in the world that are relevant to your planned research?

Data Stewardship for Open Science: *atq*

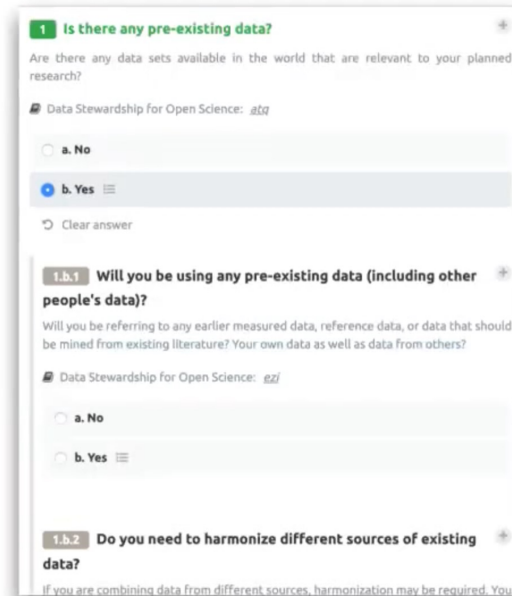
a. No

b. Yes

You know that this is very unlikely? This question is not only about data sets that are similar to what you want to determine yourself, but also reference data or data that should be mined from the existing literature. Further, it is very likely that you will refer to related data, e.g. other databases where you usually "quickly look something up", but that could maybe be properly integrated, especially if you need to do such lookups multiple times.

## Follow-up Questions

- When an answer is selected, more questions can be unfolded
- Used for questions that are relevant only until certain conditions

A screenshot of a web form interface for Data Stewardship Wizard. The form contains three questions, each with a title, a description, a source attribution, and radio button options. The first question, '1 Is there any pre-existing data?', has 'b. Yes' selected. The second question, '1.b.1 Will you be using any pre-existing data (including other people's data)?', has 'b. Yes' selected. The third question, '1.b.2 Do you need to harmonize different sources of existing data?', is partially visible. A 'Clear answer' link is located between the first and second questions. Each question has a plus sign in the top right corner, indicating it can be expanded.

**1 Is there any pre-existing data?**

Are there any data sets available in the world that are relevant to your planned research?

Data Stewardship for Open Science: *atq*

a. No

b. Yes

[Clear answer](#)

**1.b.1 Will you be using any pre-existing data (including other people's data)?**

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

Data Stewardship for Open Science: *gzj*

a. No

b. Yes

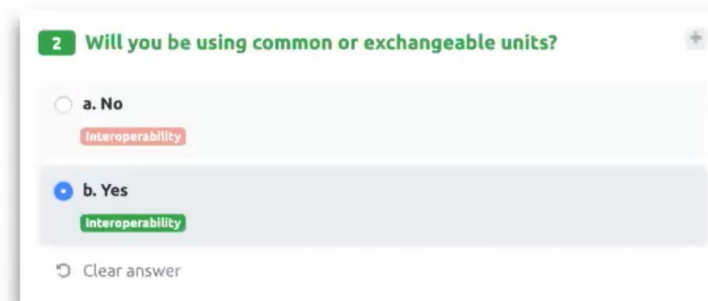
**1.b.2 Do you need to harmonize different sources of existing data?**

If you are combining data from different sources, harmonization may be required. You



## FAIR Metrics

- Each answer can affect different **FAIR metrics**
- Result is calculated as a weighted average of selected answers

A screenshot of a digital form interface. At the top, a green box contains the number '2' and the question text 'Will you be using common or exchangeable units?'. Below the question are two radio button options: 'a. No' and 'b. Yes'. The 'b. Yes' option is selected, indicated by a blue dot. Below each option is a red box labeled 'Interoperability' for 'a. No' and a green box labeled 'Interoperability' for 'b. Yes'. At the bottom of the form is a 'Clear answer' button with a circular arrow icon.



## FAIR Metrics

- **Weight** — how important the answer is

Not important 0  1 Very Important

The weight scale is a horizontal bar with a color gradient from orange on the left to green on the right. The number "0" is positioned at the left end and "1" at the right end.

- **Measure** — how does it affect the result

Bad 0  1 Good





The measure scale is a horizontal bar with a color gradient from orange on the left to green on the right. The word "Bad" is positioned at the left end and "Good" at the right end.

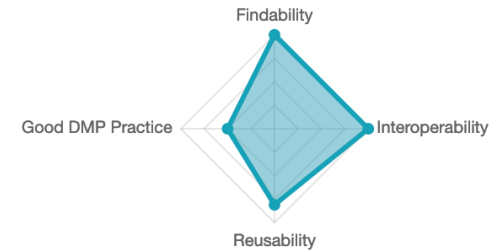
Questionnaire **Metrics** Preview Documents Settings

## Summary Report

Answered (current phase): 53/92 

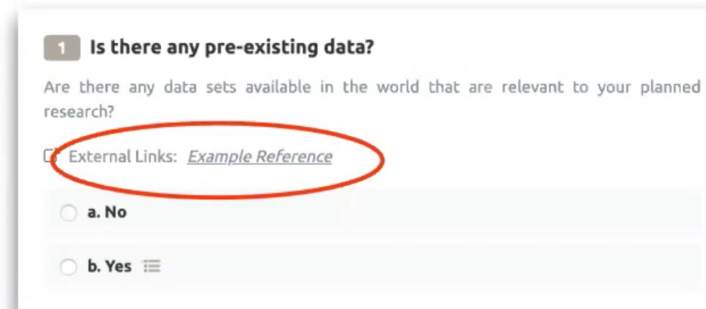
Answered: 53/95 

Metrics	Measure	
Findability	1.00	
Interoperability	1.00	
Reusability	0.81	
Good DMP Practice	0.50	



## Reference

- Can be used to provide links to **external resources**
- User can open links when filling the Questionnaire

A screenshot of a questionnaire question. The question is numbered '1' and asks 'Is there any pre-existing data?'. Below the question is a text input field containing 'External Links: [Example Reference](#)', where the link text is underlined and circled in red. Below the input field are two radio button options: 'a. No' and 'b. Yes'.

**1 Is there any pre-existing data?**

Are there any data sets available in the world that are relevant to your planned research?

External Links: [Example Reference](#)

a. No


b. Yes ☰

## Expert


- Can be used to provide contacts to experts
- Users can see the contact by the question

**1 Is there any pre-existing data?**

Are there any data sets available in the world that are relevant to your planned research?

 Experts: *John Brown (john.brown@example.com)*

a. No

b. Yes 

# PROJECT PHASES

Red = answer during this Phase

Green = fully answered

Grey = no need to answer during this Phase

Questionnaire Metrics Preview Documents Settings

View Import answers

SHOW ALL  
HIDE ALL

- ✓ Answered by
- ✓ Phases
- ✓ Question tags
- ✓ Non-desirable questions
- Metric values

II. Re-using data

III. Creating and collec... 10

IV. Describing data 1

**V. Processing data 2**

- ▶  Will you be using a shared ...
- ▶  Data storage systems and fi...
  - Workflow development
- ▶  How will you make sure to k...
  - Do you need to do compute...
  - Is the risk of information los...
  - Do you have a contingency ...

VI. Preserving data 2

**V.2.a.1 How much storage space will the project require for all data and software, including temporary storage?**

FBM RDM FNS DMP

Desirable: *DMP and Research Planning*

a. So little that it is not a problem

b. I can specify the total amount ☰

**V.2.a.2 Are you using a filesystem with files and folders?**

FBM RDM

Are some of the data in the project stored in a filesystem with files and folders?

Desirable: *RDM guidance*

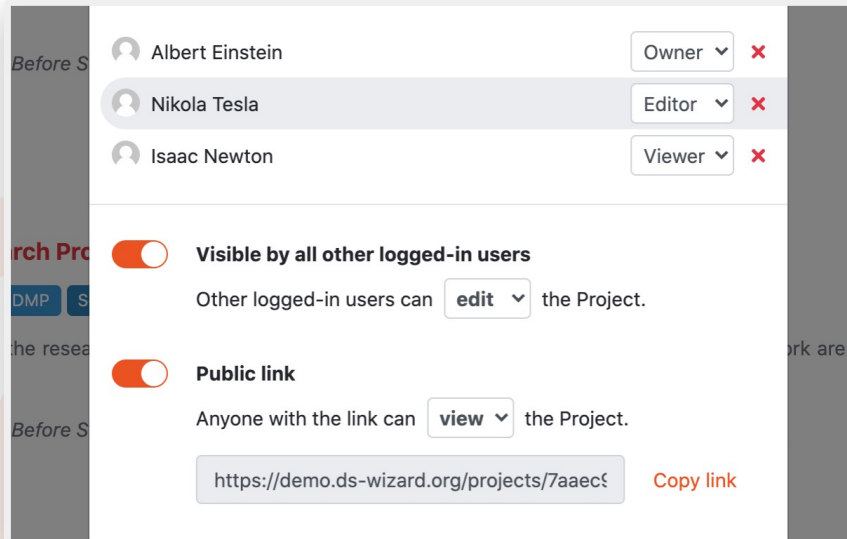
External links: [RDMkit on Data Organisation](#)

a. No

b. Yes

**✓ V.2.a.3 Will you be storing data in an "object store" or a "document store" system?**

FBM RDM

A screenshot of a web interface for project sharing and user access. It shows a list of users with their roles and a section for visibility and public link settings. The user list includes Albert Einstein (Owner), Nikola Tesla (Editor), and Isaac Newton (Viewer). The visibility settings include a toggle for 'Visible by all other logged-in users' (set to 'edit') and a toggle for 'Public link' (set to 'view'). A public link is displayed as 'https://demo.ds-wizard.org/projects/7aac9' with a 'Copy link' button.

Before S

Albert Einstein Owner ✕

Nikola Tesla Editor ✕

Isaac Newton Viewer ✕

**Visible by all other logged-in users**

Other logged-in users can **edit** the Project.

**Public link**

Anyone with the link can **view** the Project.

<https://demo.ds-wizard.org/projects/7aac9> Copy link

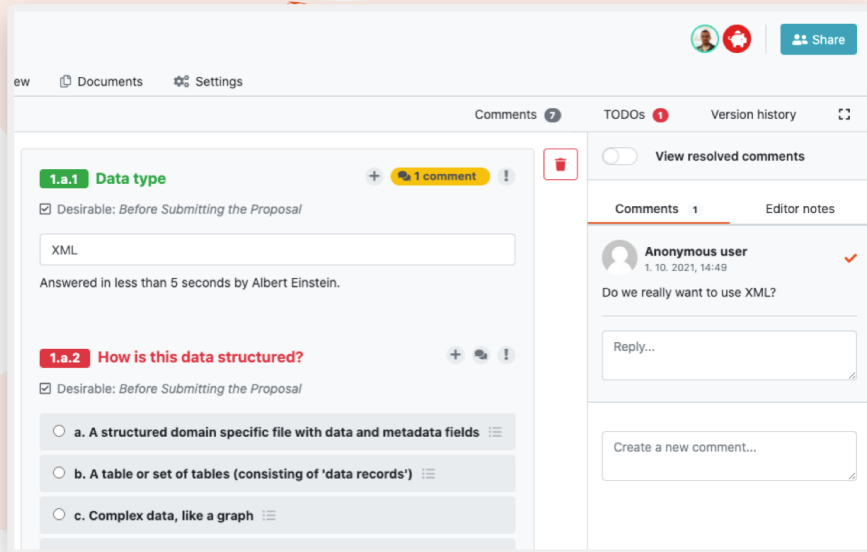
## Project Sharing & User Access

**Share** your Projects with your colleagues whether they have a DSW user account or not. Furthermore, you can set their **access rights** as well and let them just view, edit or even own the Project.

## Comments & Editor Notes

Get feedback and discuss your project. You can share it in **comment mode** for a review or write **editor notes** visible only to the project editors.





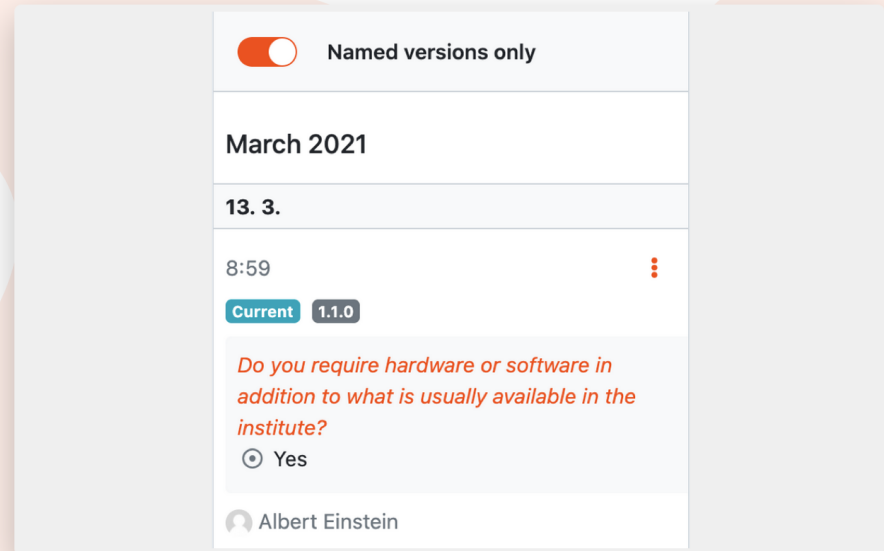
## Online Collaboration

You can **collaborate online in real-time** on your data management plans with other users or anyone who you send the link to. Just like you would expect it in a cutting-edge online platform.



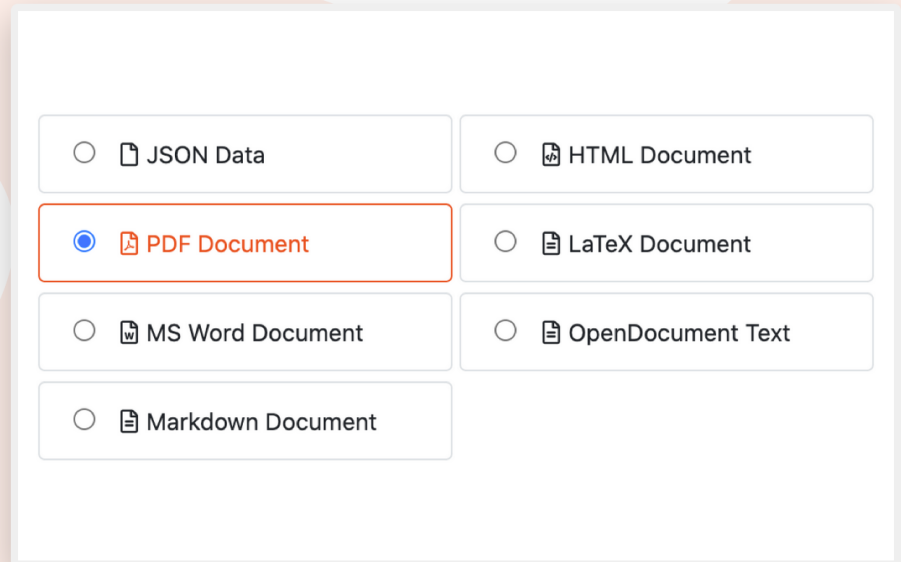
## Version History








Every change you make in your project is saved. You can browse the history, watch who makes changes and when, and create a version for your project's specific state.



## File Formats

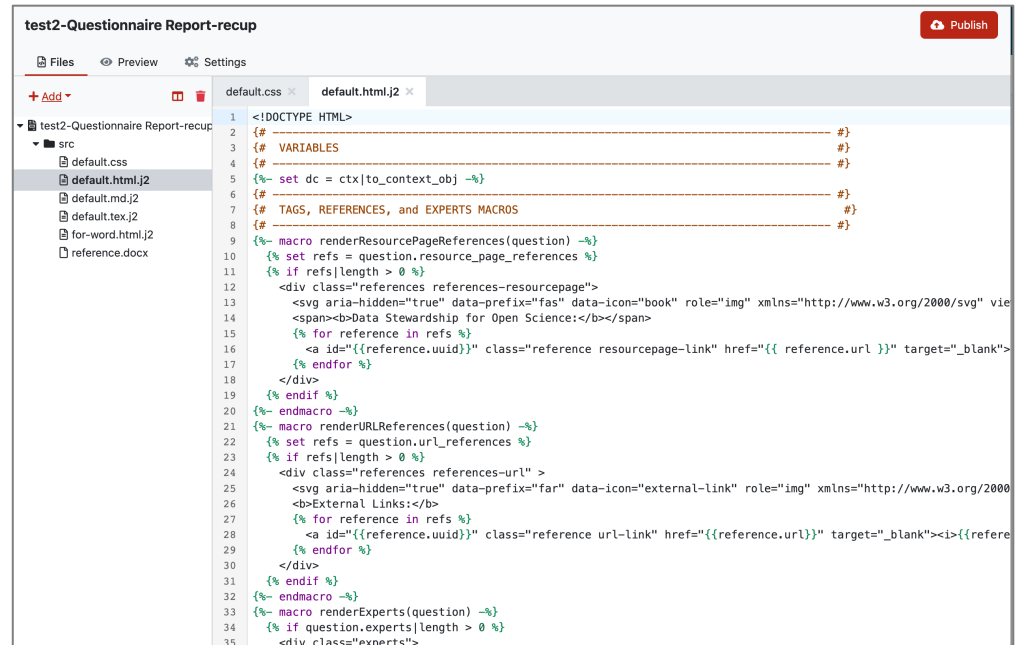
Each template can specify the number of formats that can be used for export. For example, the Questionnaire Report, as well as Horizon 2020 and Science Europe templates, can be exported as HTML, a Word document, LaTeX, Markdown, an OpenDocument, or a PDF. Similarly, the Machine-Actionable DMP can be exported into JSON but also into various RDF formats such as RDF/XML, Turtle, TRIG, or JSON-LD.

A screenshot of a web interface for selecting file formats. It consists of a white rectangular box with rounded corners and a subtle drop shadow, set against a light orange background. Inside the box, there are seven radio button options arranged in a grid. The 'PDF Document' option is selected, indicated by a blue dot and a red border around its label. Each option includes a small icon representing the file format: a document icon for most, and a data icon for JSON Data.

<input type="radio"/>  JSON Data	<input type="radio"/>  HTML Document
<input checked="" type="radio"/>  PDF Document	<input type="radio"/>  LaTeX Document
<input type="radio"/>  MS Word Document	<input type="radio"/>  OpenDocument Text
<input type="radio"/>  Markdown Document	

# DOCUMENT TEMPLATES

- Questionnaire -> document
- Several output formats



The screenshot shows a document editor interface for a questionnaire report. The title bar reads "test2-Questionnaire Report-recup" and includes a "Publish" button. The interface is divided into three main sections: a file explorer on the left, a preview area in the middle, and a code editor on the right.

The file explorer on the left shows a tree structure for the project "test2-Questionnaire Report-recup". It includes a "src" folder containing several files: "default.css", "default.html.j2" (which is selected), "default.md.j2", "default.tex.j2", "for-word.html.j2", and "reference.docx".

The code editor on the right displays the content of "default.html.j2". The code is a J2 template for an HTML document, starting with a DOCTYPE declaration and followed by several macros for rendering references and experts. The code is as follows:

```
1 <!DOCTYPE HTML>
2 {# ----- #}
3 {# VARIABLES ----- #}
4 {# ----- #}
5 {%- set dc = ctx|to_context_obj -%}
6 {# ----- #}
7 {# TAGS, REFERENCES, and EXPERTS MACROS ----- #}
8 {# ----- #}
9 {%- macro renderResourcePageReferences(question) -%}
10 {% set refs = question.resource_page_references %}
11 {% if refs|length > 0 %}
12 <div class="references references-resourcepage">
13 <svg aria-hidden="true" data-prefix="fas" data-icon="book" role="img" xmlns="http://www.w3.org/2000/svg" vie
14 <span><b>Data Stewardship for Open Science:</b></span>
15 {% for reference in refs %}
16 <a id="{{reference.uuid}}" class="reference resourcepage-link" href="{{ reference.url }}" target="_blank">
17 {% endfor %}
18 </div>
19 {%- endmacro -%}
20 {%- macro renderURLReferences(question) -%}
21 {% set refs = question.url_references %}
22 {% if refs|length > 0 %}
23 <div class="references references-url" >
24 <svg aria-hidden="true" data-prefix="far" data-icon="external-link" role="img" xmlns="http://www.w3.org/2000
25 <b>External Links:</b>
26 {% for reference in refs %}
27 <a id="{{reference.uuid}}" class="reference url-link" href="{{reference.url}}" target="_blank">{{refere
28 {% endfor %}
29 </div>
30 {%- endmacro -%}
31 {%- macro renderExperts(question) -%}
32 {% if question.experts|length > 0 %}
33 <div class="experts">
```

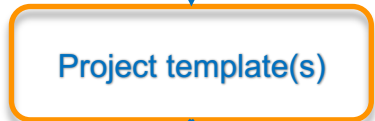
# PLANNED WORKFLOW FOR LIFE SCIENCES



ELIXIR-CH admin



Data Stewards

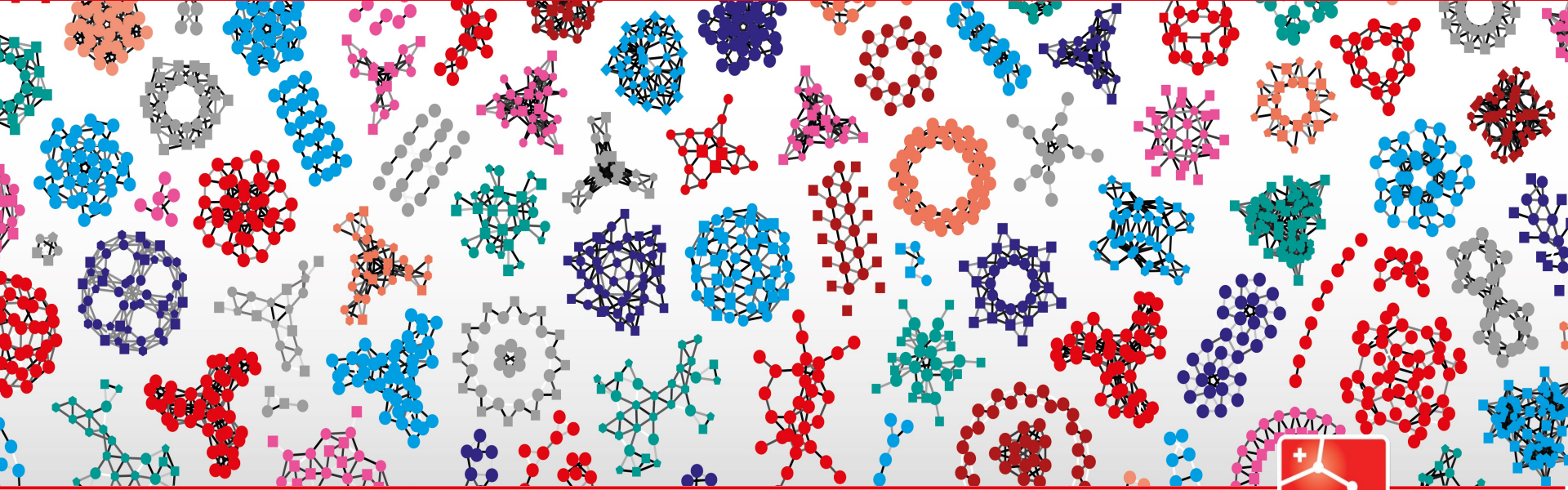


Researchers



DEMO





Swiss Institute of  
Bioinformatics

# Thank you!