DSW
DATA STEWARDSHIP WIZARD

DSW Workshop for Finnish Data Support Personnel

elixir CZECH REPUBLIC

FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Tereza Machačová
tereza.machacova@ds-wizard.org

Marek Suchánek
marek.suchanek@ds-wizard.org

13 April 2021

Today‘s Plan (Outline)

- Introduction and motivation
- How to use DSW as a researcher
- Short hands-on exercise
- How to customize DSW
- Wrap-up
- Discussion and questions

Introduction to DMP

- Data Management Planning (DMP) is important for (research) projects
- DMP = what, why and how will be done to avoid any inclarities during project
- Data Stewardship is complicated and covers very various topics
- DMP ≠ extensive static document with a long text to satisfy grant agency

Templates

... and many more

Preflight Checklist

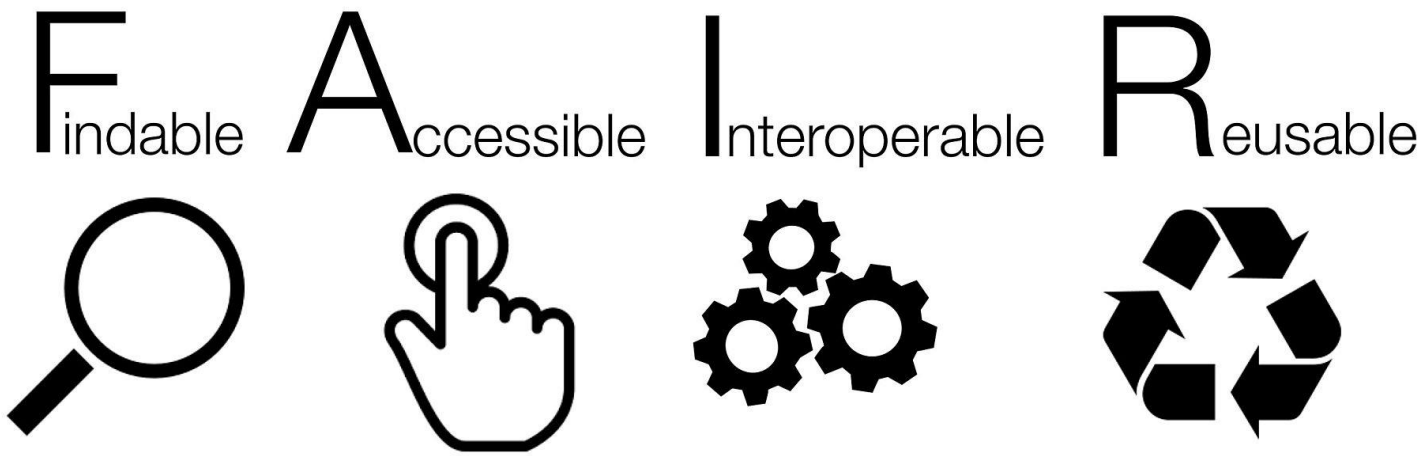
Adobe Stock | RioPatuca Images

What is Data Stewardship Wizard?

- Open-source tool developed as a part of ELIXIR CZ research infrastructure
- Tool suitable for everyone (from beginners to data stewards)
- Serves as a check-list before starting the project
- Supports data management planning with respect to (current) best practices

FAIR Data

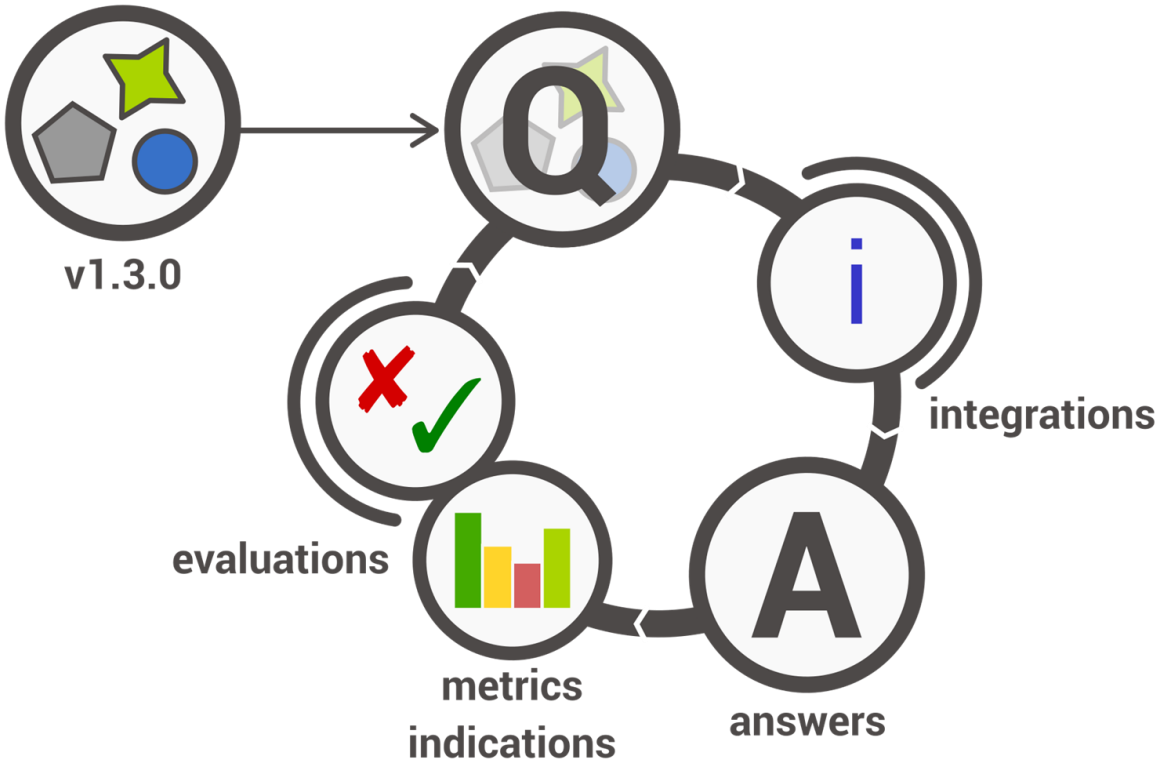
- What is FAIR?
- Principles applied on data to make them:



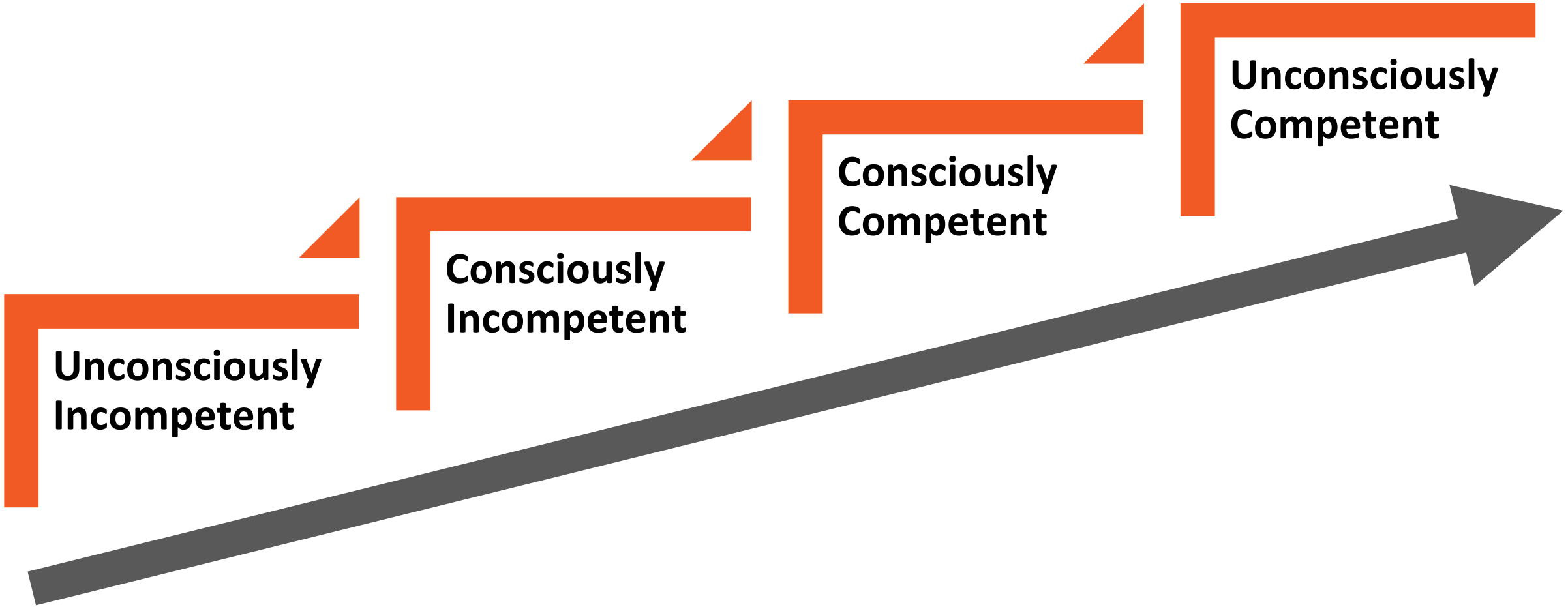
DSW Main Ideas

- Minimum of writing = DMP is not an essay, as little writing as possible
- Guidance = DSW guides users through the smart Questionnaire
- Flexibility = easy to edit the content and integrate with other services
- Openness = anybody can use it and create own content
- User-oriented = DSW development is strongly user feedback driven

Returning to DMP



Educational Aspect of DSW



Smart Questionnaire

DSW

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

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

Desirable: *Before Submitting the Proposal*
Data Stewardship for Open Science: *atq*
External links: *Google dataset search*, *Datacite Search*

- ○ **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?

Desirable: *Before Submitting the Proposal*
Data Stewardship for Open Science: *ezi*

- ○ **a. No**
- ◉ **b. Yes**

Clear answer

Data Stewardship for Open Science

DS Wizard Go to App

Data Stewardship for Open Science: Chapter 1.1

With kind permission of CRC Press Taylor & Francis Group

Is there pre-existing data?

What's up?

For many decades if not centuries, virtually every experiment started with the collection or creation of 'observations' and in fact data. In social sciences and humanities the tendency to 'reuse' data that had been created earlier, in all kinds of surveys and increasingly of course from sources such social media maybe already somewhat more established. However, in many of the hard experimental sciences, the generation of new data specifically generated to answer a hypothetical question is still so commonplace that careful thinking about the actual need to generate new data may just not be on the radar screen. Obviously, data creation will need to continue, but increasingly we have to ask the question whether such new data are absolutely necessary to answer the question we want to answer. With more and more data becoming available in reusable format, there may well be existing data collections 'Other People's' Data and associated Services (OPEDAS) that without or with some extra effort needed, can answer at least part of the question or least may be crucial for the interpretation of your own data.

Do

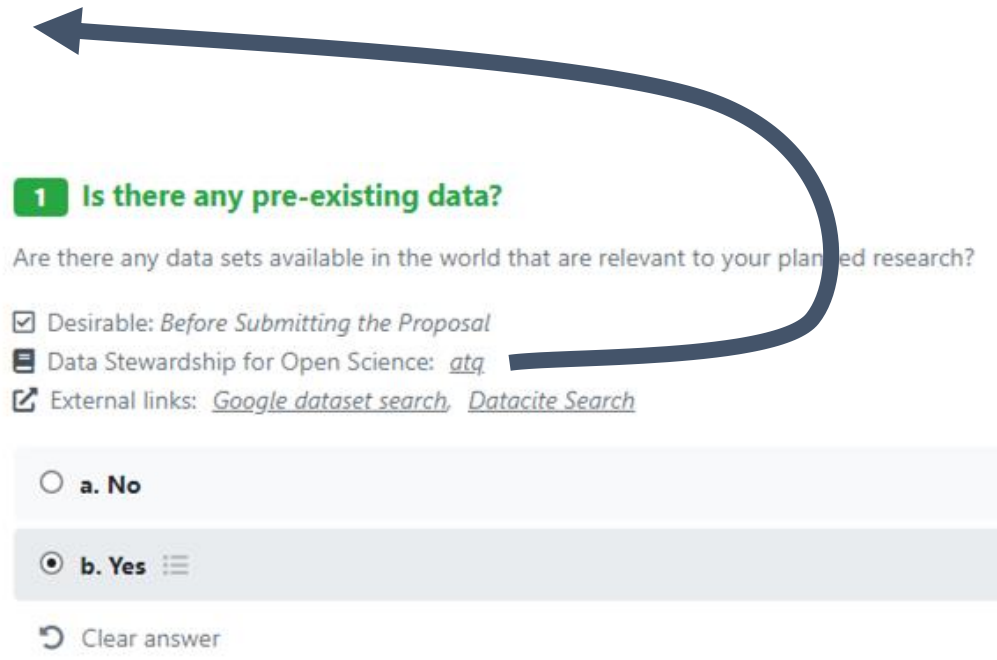
- Search for data sets (OPEDAS) that may be re-usable and can help you to reduce the number of new data sets you may have to generate (and steward later on).
- Include annotated collections of data and curated databases in your search.
- Check the accessibility and license situation attached to the relevant data sets you found.
- Check their interoperability. They may be relevant but not interoperable with your analysis pipelines. In that case you may have to extract, transform and load (ETL) them or decide that -although relevant- they are not reusable for your purpose.
- Ensure that using OPEDAS will not restrict in any way the use of your results later on, including copyright and freedom to operate on the request of IPR.
- Check how to cite and acknowledge OPEDAS.
- Consider to actively involved OPEDAS owners in your research in order to make optimal use of their data.
- Speak to colleagues who did similar experiments before to find out about potential OPEDAS you may consider to use.

Don't

- Assume no OPEDAS exist without thorough checking using all your possibilities.
- Start an experiment without properly checking with colleagues about the best approach and OPEDAS out there.
- budget for data generation in your study without justifying to the funder why the generation of the data is necessary.
- Move into actual experimentation without consulting a data expert.

Links

- DS Question GitHub resources repository: atq

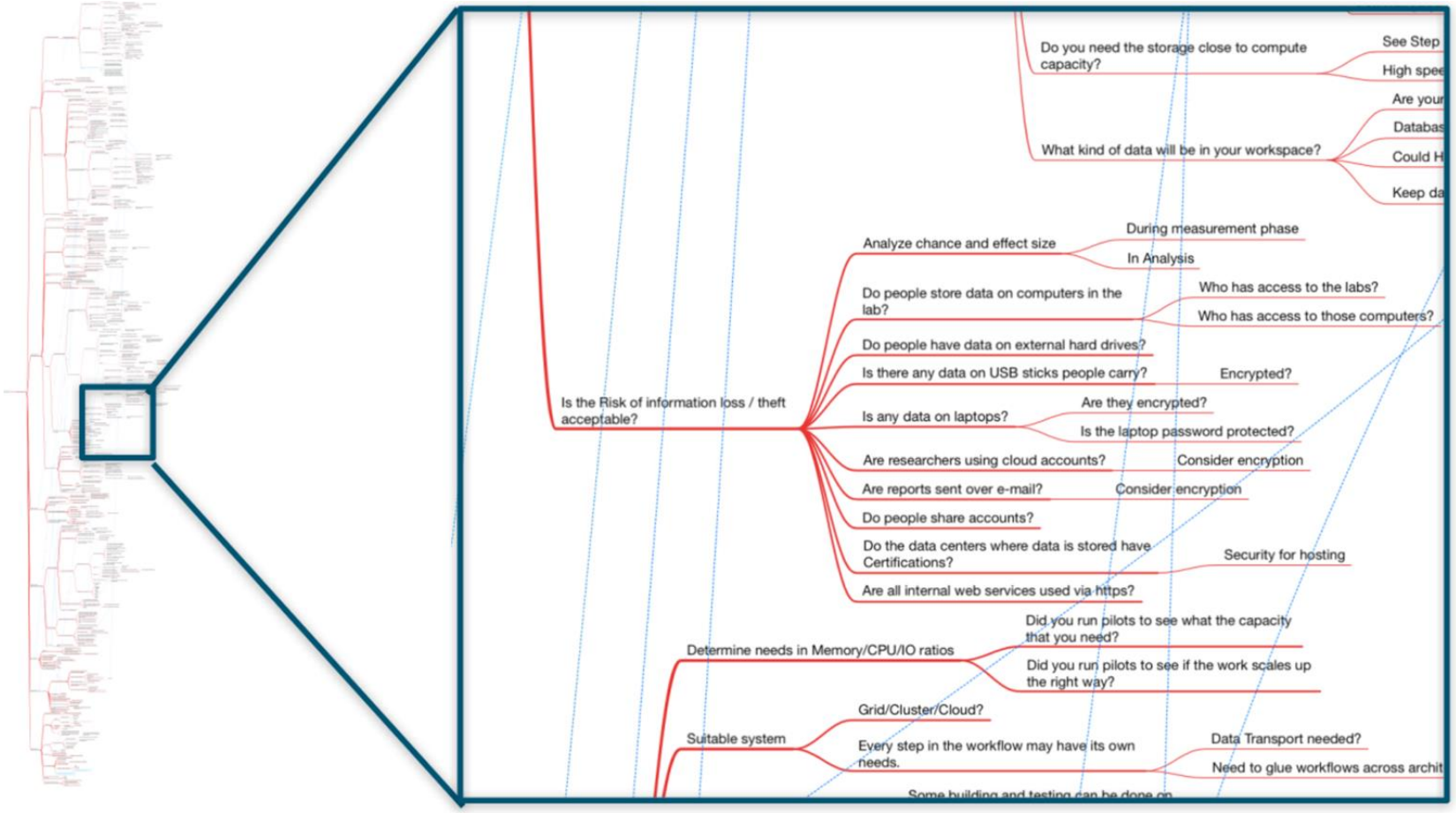


Data Stewardship for Open Science:

Implementing FAIR Principles

By Barend Mons

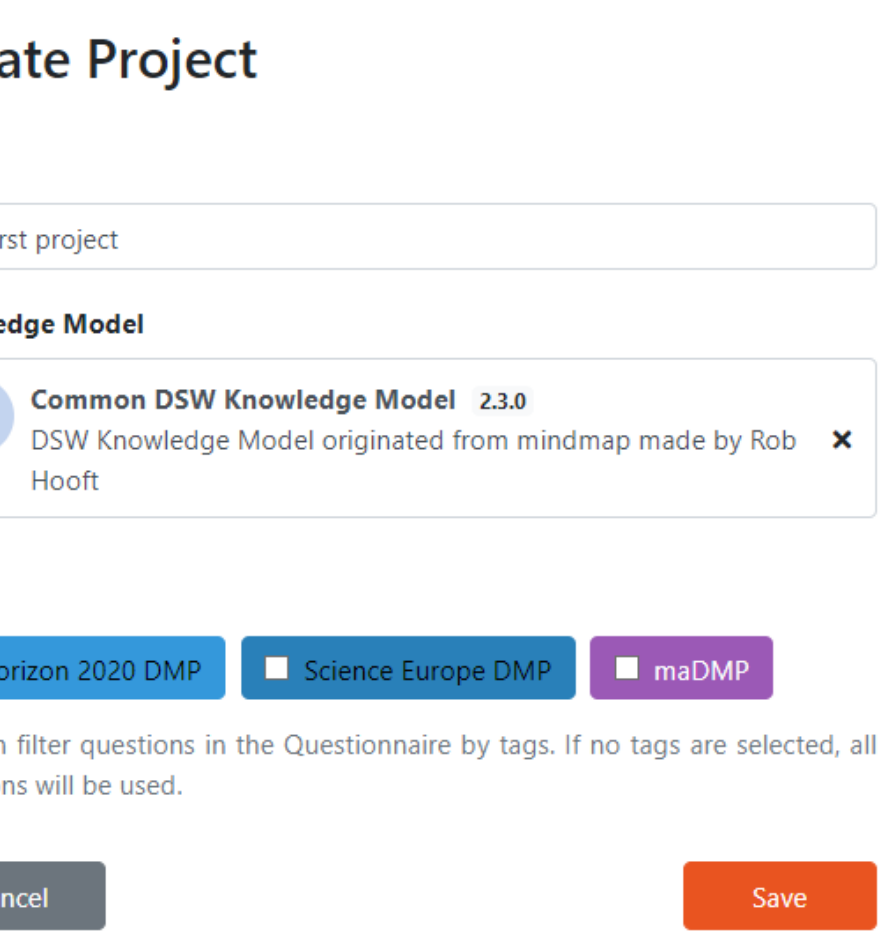
Common DSW Knowledge Model

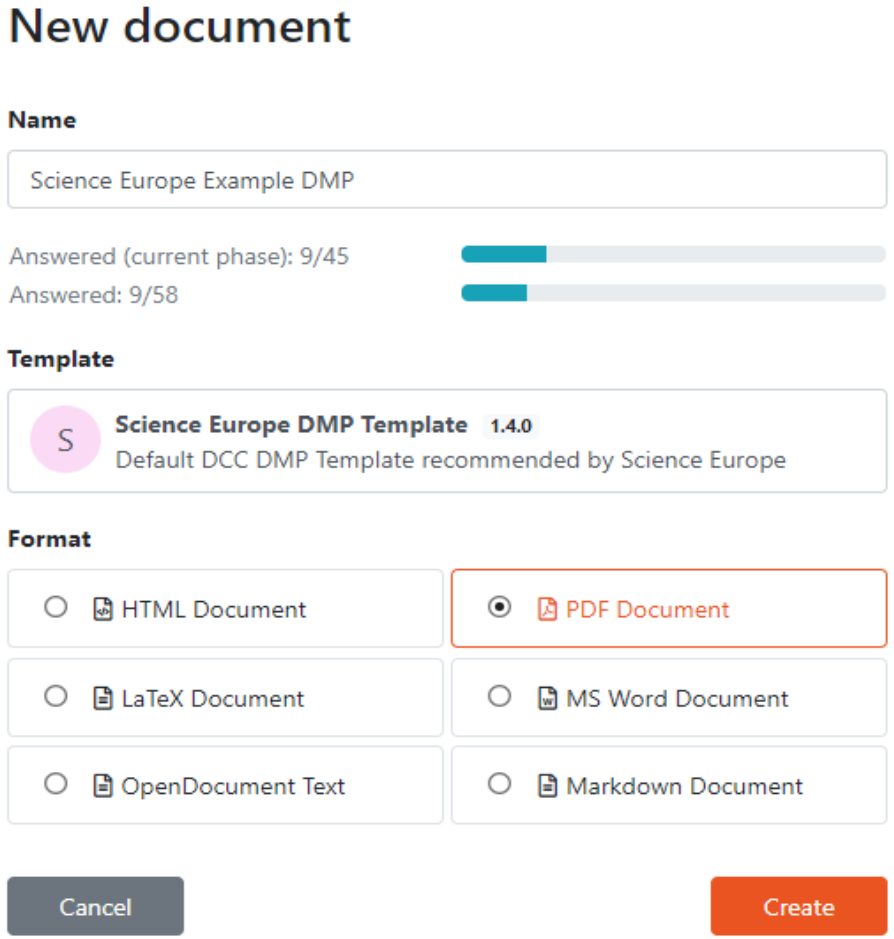


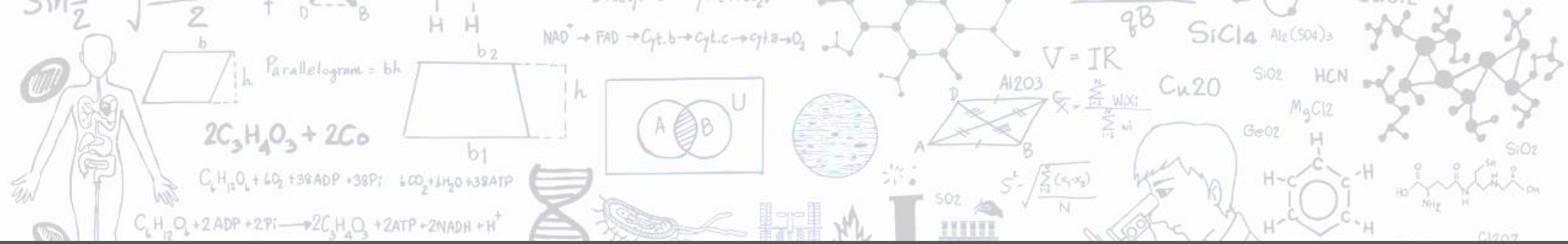
- Complex decision tree
- ~600 nodes
- Originated from mindmap made by Rob Hooft (DTL, ELIXIR NL)

https://zenodo.org/record/2614820

Knowledge Model vs Template



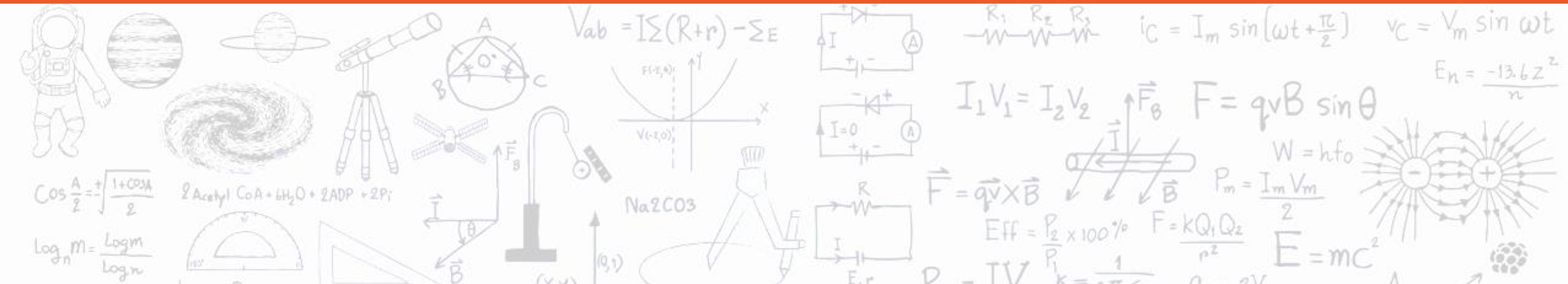




DSW

DATA STEWARDSHIP WIZARD

How to use (Researcher)



Goals of Researchers

- Create a DMP for own project
- Learn about Data Management
- Collaborate with colleagues
- Be productive

Create Project (DMP)

Create Project

Name

My first project

Knowledge Model

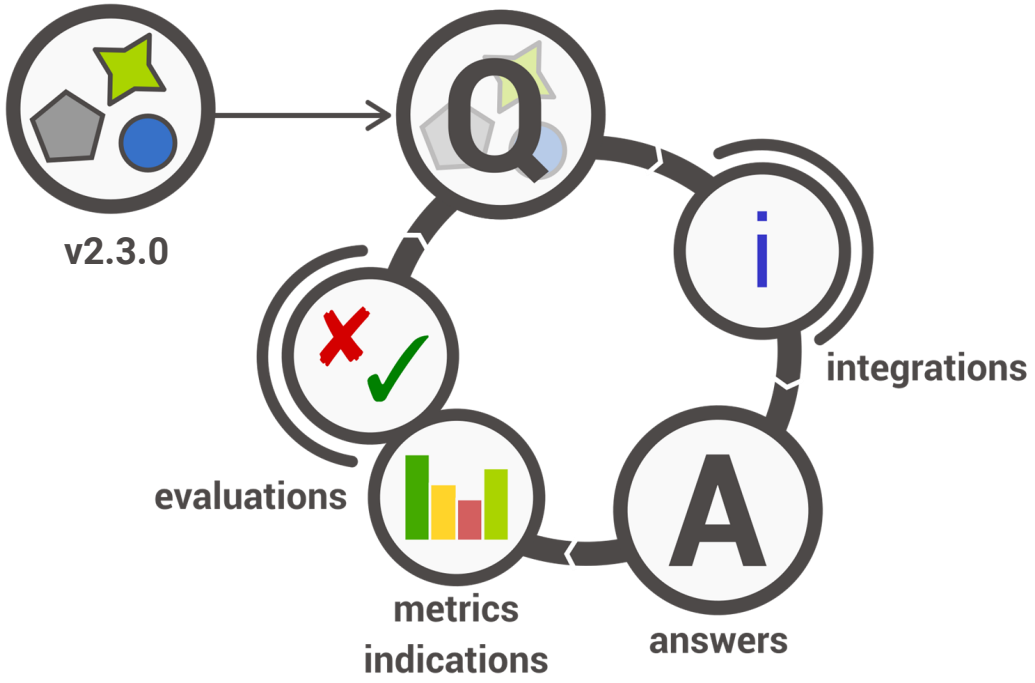
C **Common DSW Knowledge Model** 2.3.0
DSW Knowledge Model originated from mindmap made by Rob Hooft ×

Tags

☐ Horizon 2020 DMP ☐ Science Europe DMP ☐ maDMP

You can filter questions in the Questionnaire by tags. If no tags are selected, all questions will be used.

Cancel Save



Key Factor: Good Guidance

Explanation

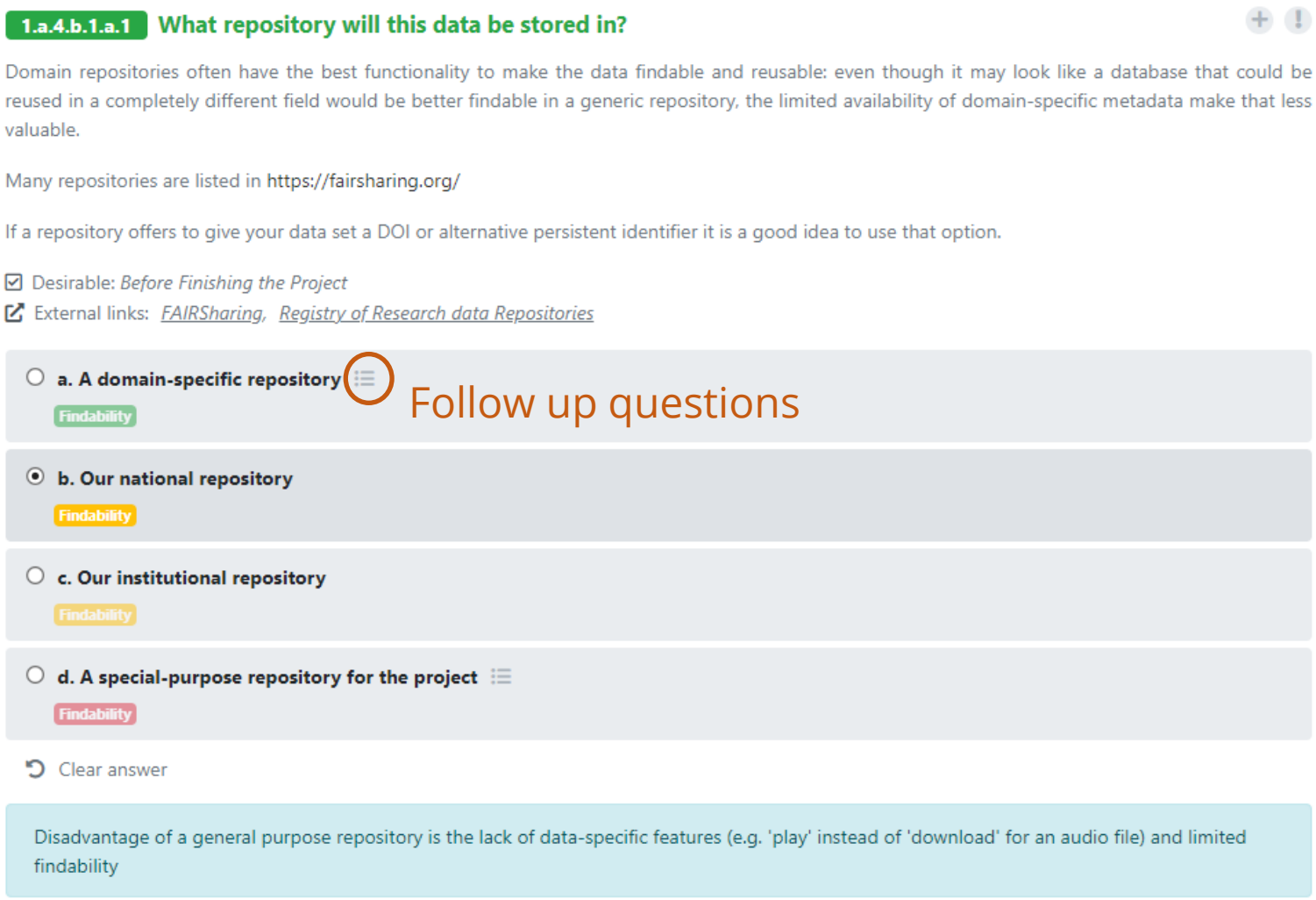
Choice of Option

FAIR metrics

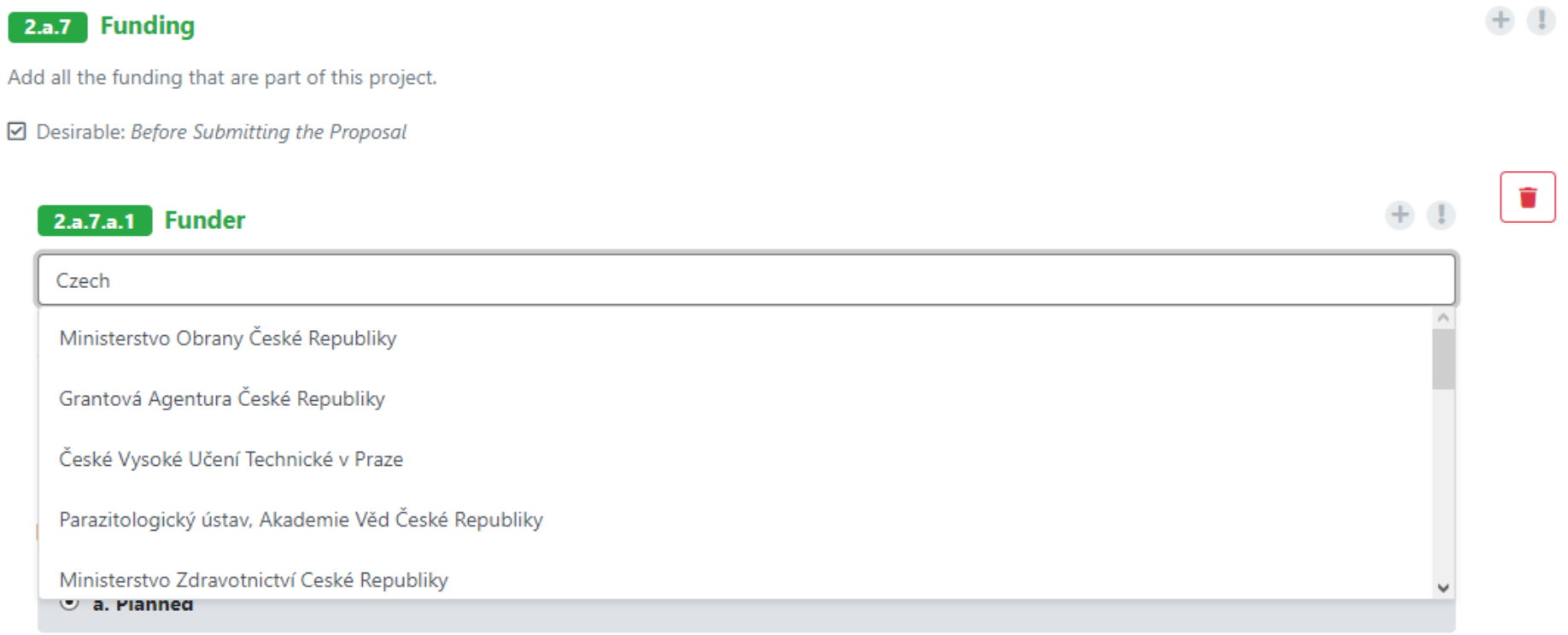
TODO, Feedback

References

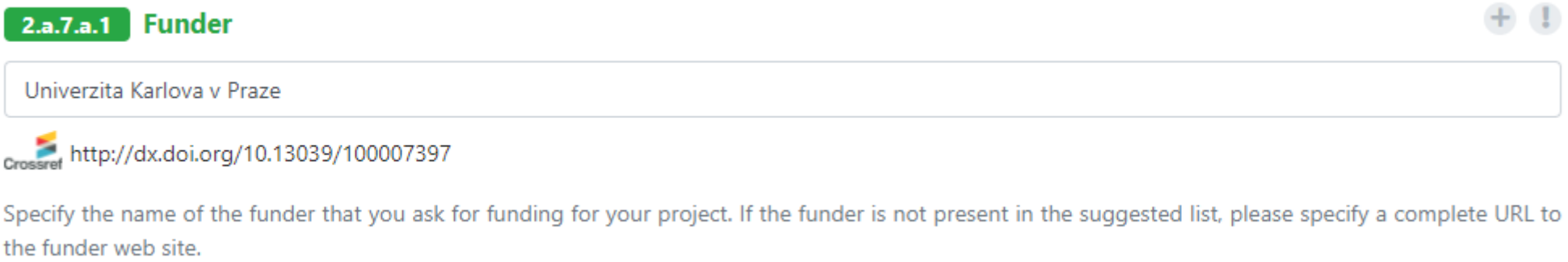
Recommendation



Answers through Integration



Linking the Answers to URI



FAIRsharing.org
standards, databases, policies

Crossref

ROR

WIKIDATA

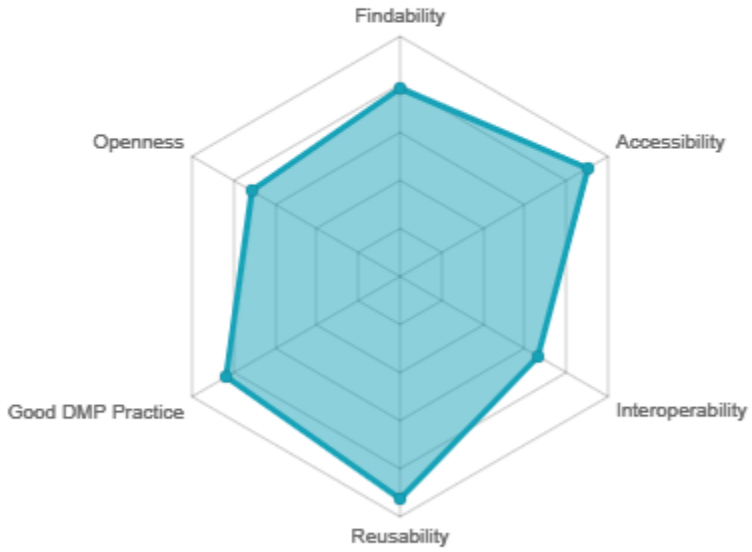
Indications and Metrics

Summary Report

Answered (current phase): 100/100

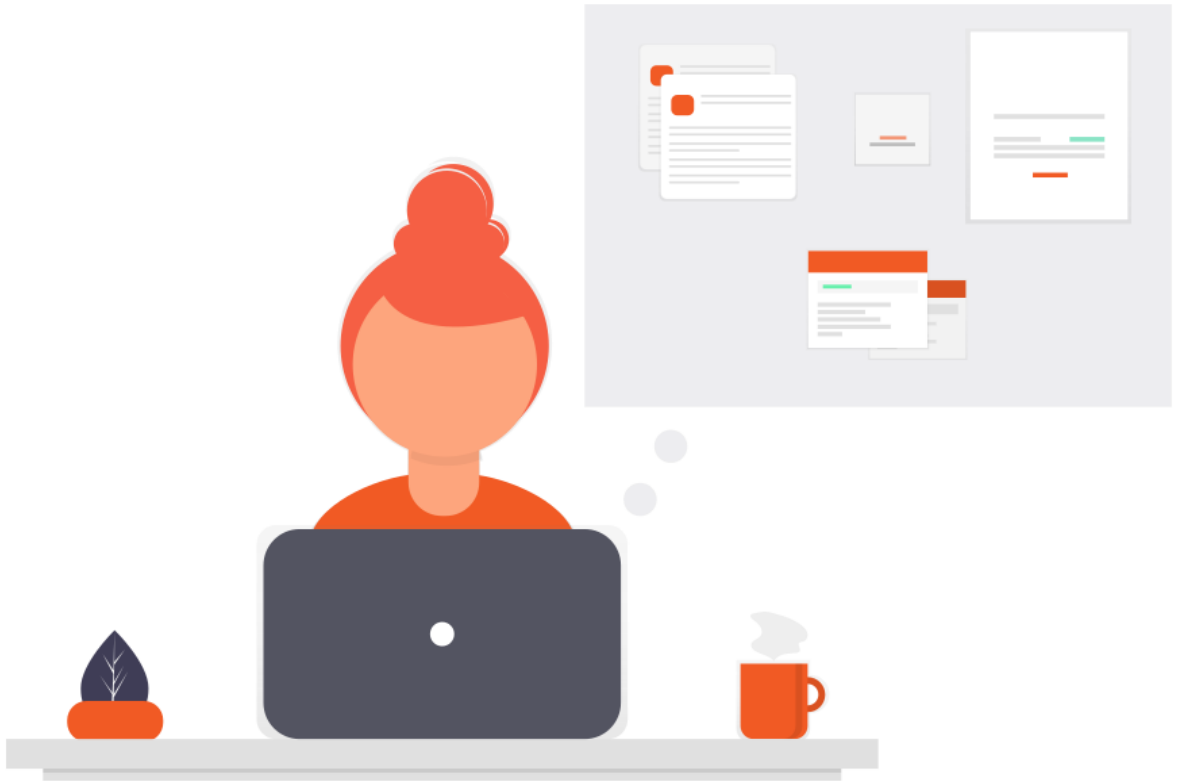
Answered: 293/305

Metric	Measure
Findability	0.78
Accessibility	0.90
Interoperability	0.67
Reusability	0.92
Good DMP Practice	0.84
Openness	0.72

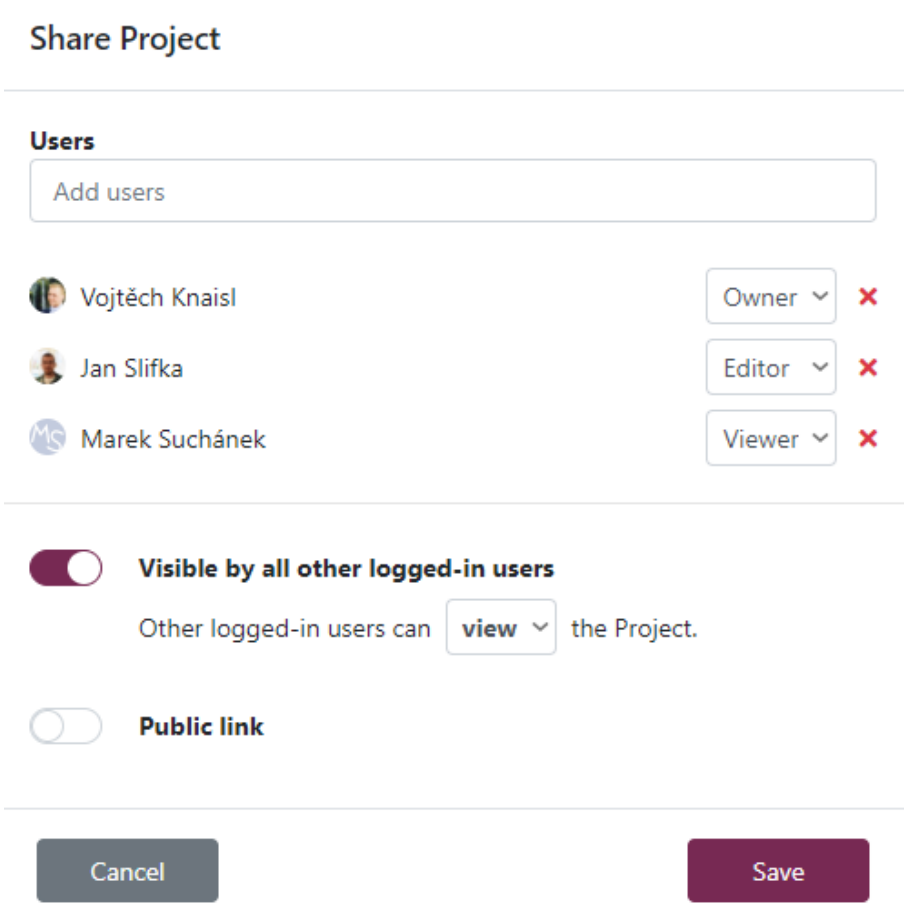


Demonstration

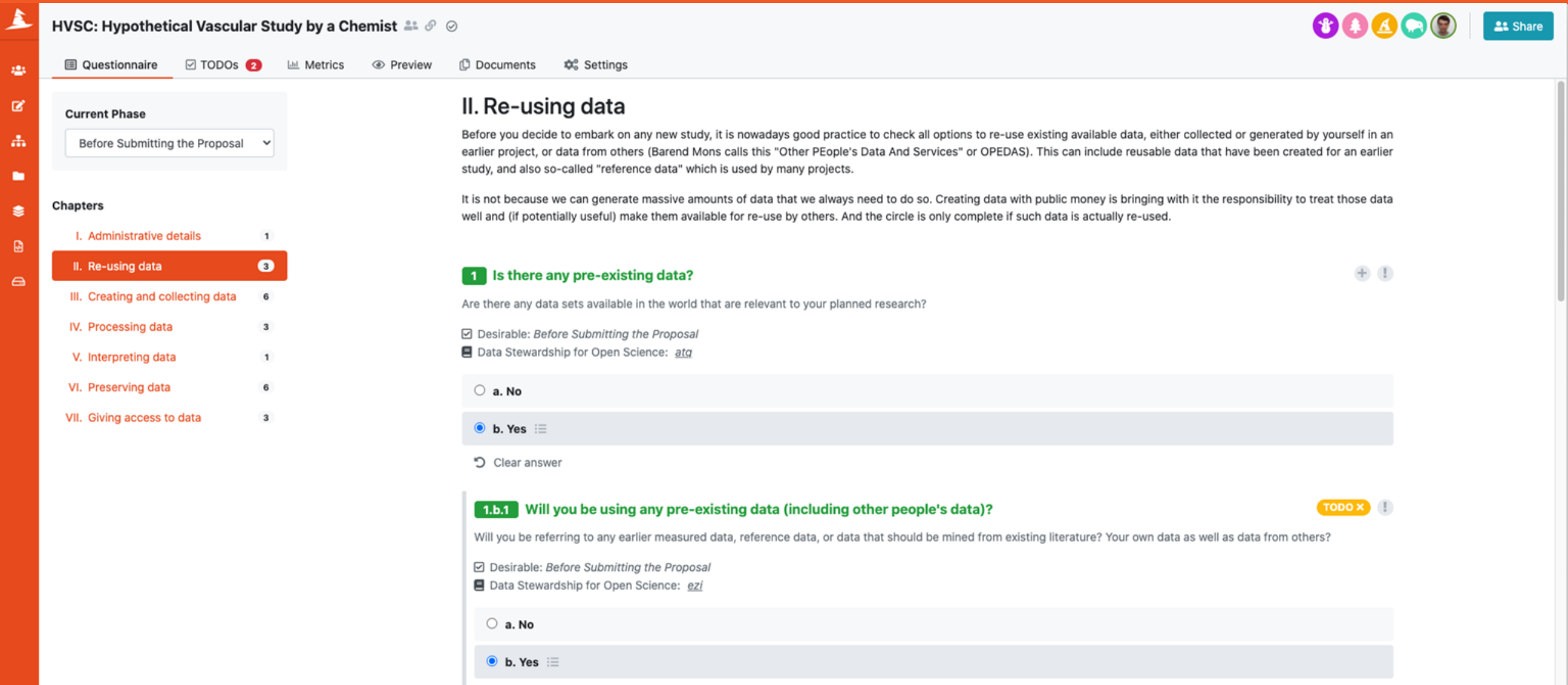
DSW



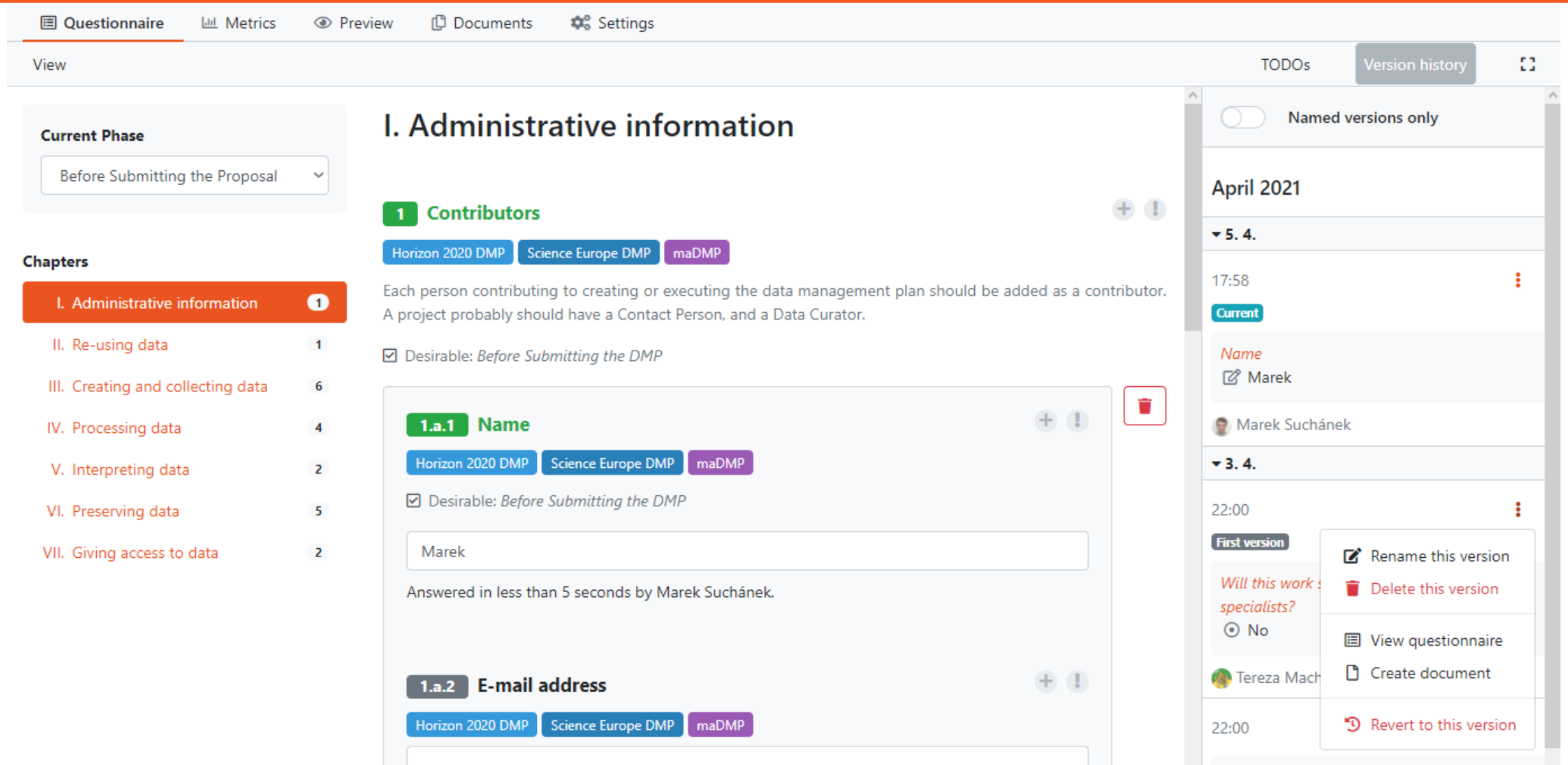
Working Together with your Colleagues



Working Together with your Colleagues



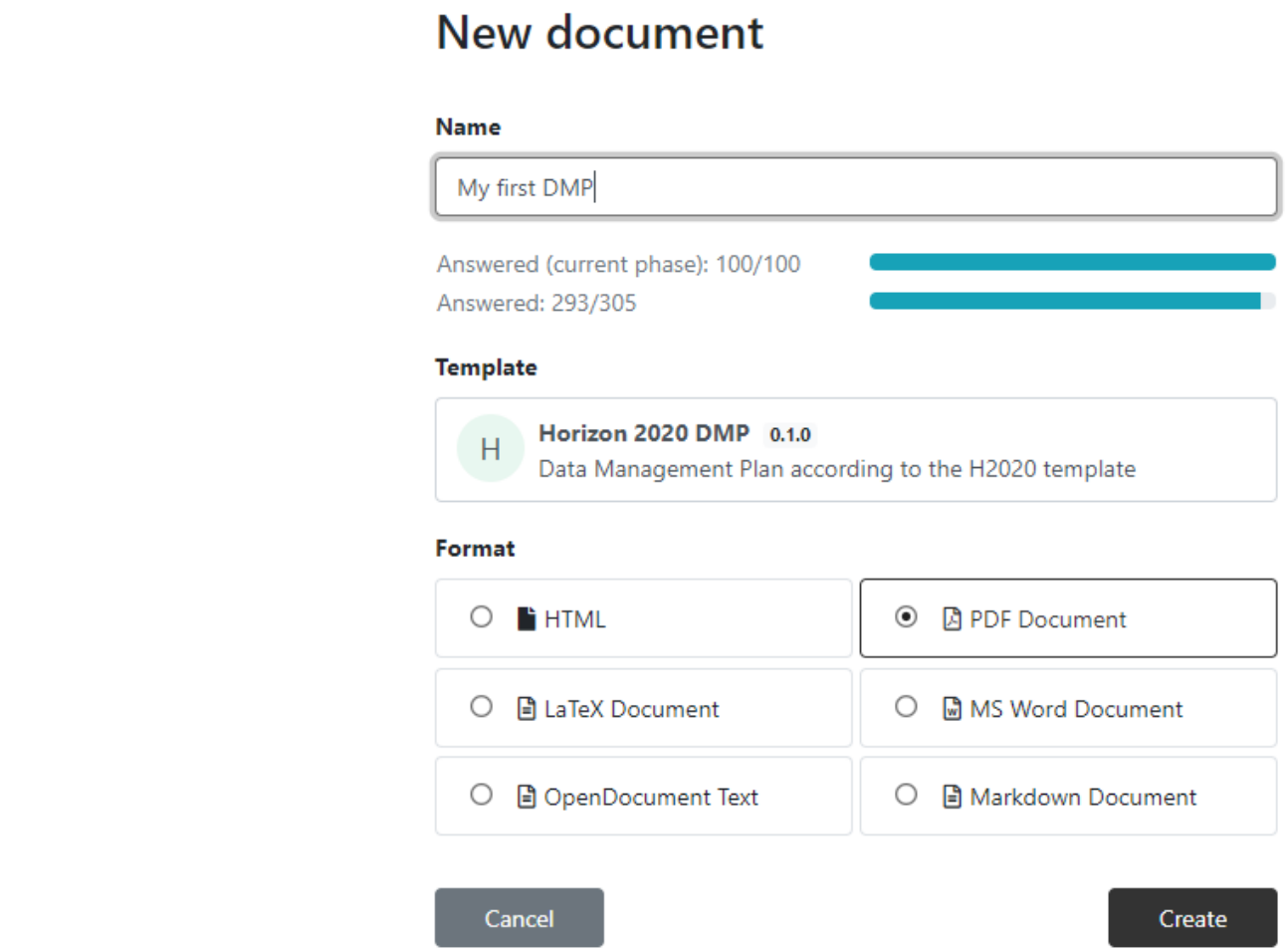
Version History



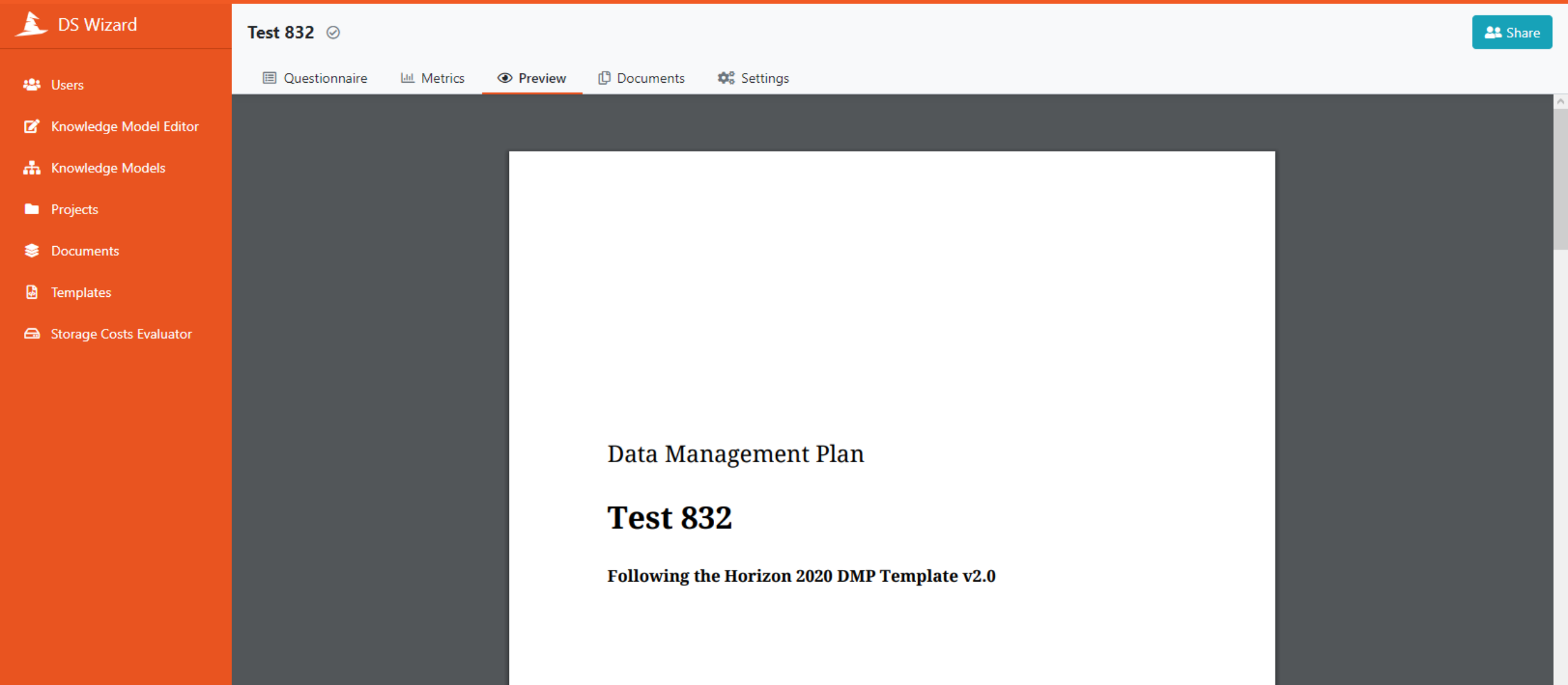
Demonstration

DSW

DMP Export in Various Templates



Document Preview and Download



Demonstration

Break & Hands-on

- Let's have a break ~15 minutes
- Grab a snack or coffee
- Try out what was shown
- Create a project (KM: ...)
  - Fill in basic information about the project
  - Generate H2020 document
  - Generate maDMP (JSON or Turtle RDF)

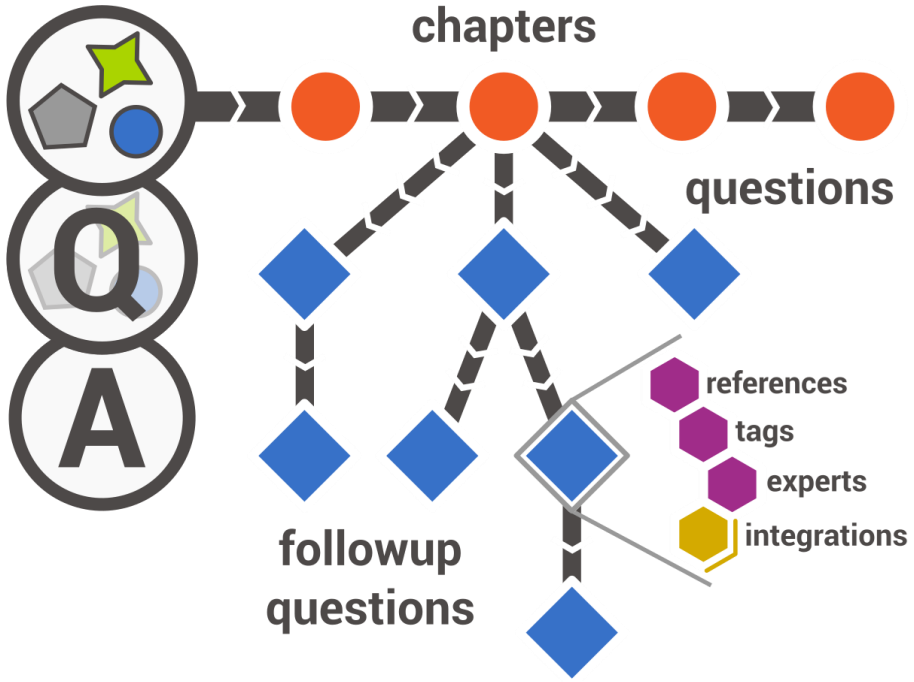
DSW

DATA STEWARDSHIP WIZARD

How to use (Data Steward)

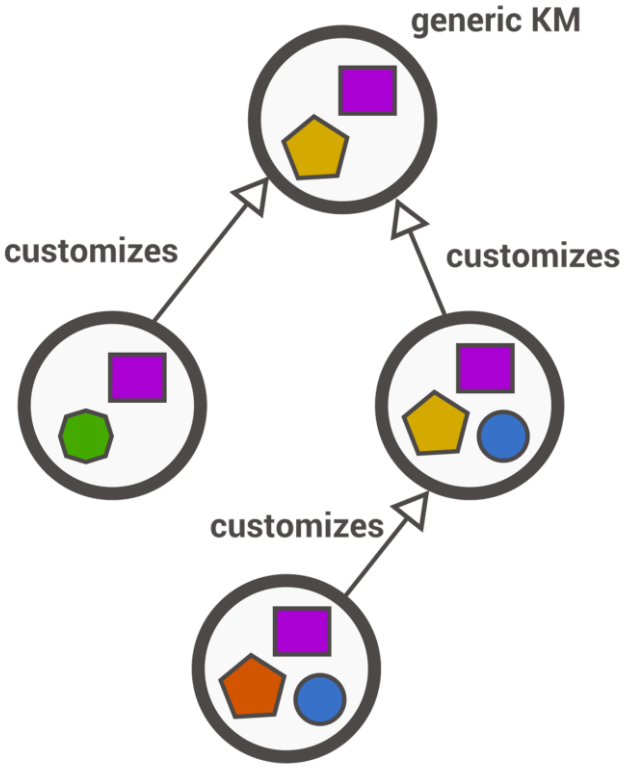
What is „Knowledge Model“

- Questionnaire structure is not fixed and strictly given
- Contains:
  - Chapters
  - Questions (different types)
  - Answers (choice of option)
  - References and Experts
  - Tags
  - Integration for answer suggesting



Questionnaire Customization

- Built-in Knowledge Model Editor
- Create „from scratch“
- Edit existing ones:
  - Delete non-relevant questions
  - Add relevant ones
  - Change existing questions
  - Structure change

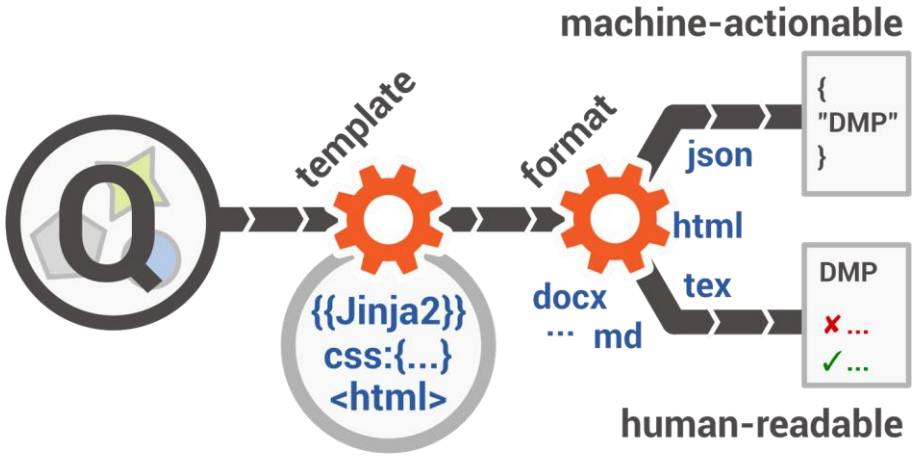


Demonstration

DSW

Development of Own Templates

- Not hard-coded, possible to change and create custom templates
- For easier development „DSW Template Development Kit“
- You can watch tutorial on YouTube



Integrations and Customizations

- It is possible to further customize and integrate DSW in your workflow:
  - User authentication using OpenID identity providers
  - Document submission to other services (e.g. institutional repository)
  - Branding (colors, logo, icons, email templates, etc.)
  - Custom integration questions (answer suggestions)
  - DSW exposes REST API for other integrations and scripting:
    https://api.demo.ds-wizard.org/swagger-ui/
    (Python SDK is being under development)

Wrap-up & Call to Action

- You now know how to use DSW to start planning
- For planning, register in researchers.ds-wizard.org
- For testing, register in demo.ds-wizard.org
- In case of any questions or feedback, let us know
  - info@ds-wizard.org
  - ideas.ds-wizard.org
  - github.com/ds-wizard

Acknowledgement

The development and operation of DSW is supported by ELIXIR CZ research infrastructure (MŠMT Grant No.: LM2018131).

Questions & Discussion

DSW

DATA STEWARDSHIP WIZARD

Thank you!

info@ds-wizard.org