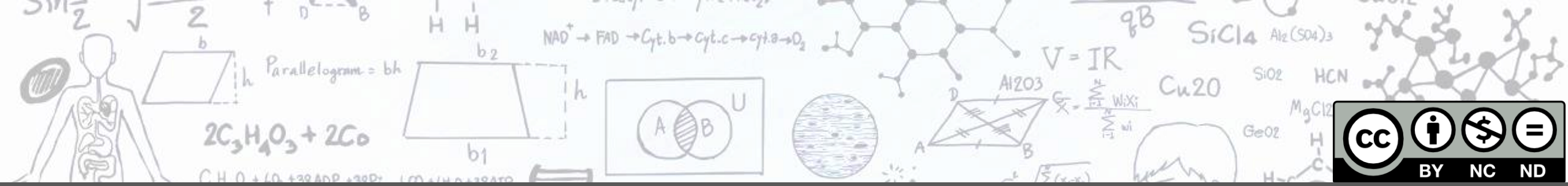


DSW
DATA STEWARDSHIP WIZARD

Data Management with Data Stewardship Wizard

elixir CZECH REPUBLIC

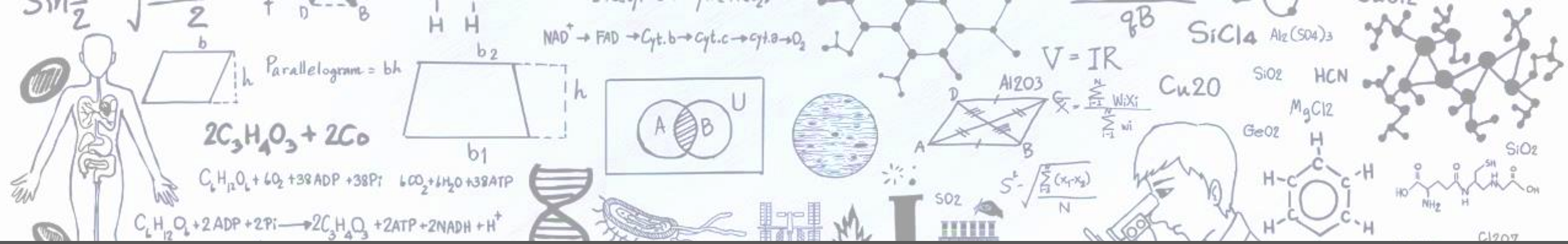
FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

CODEVENCE

Robert Pergl
robert.pergl@fit.cvut.cz
robert.pergl@codevence.com
Project Coordinator

Contents

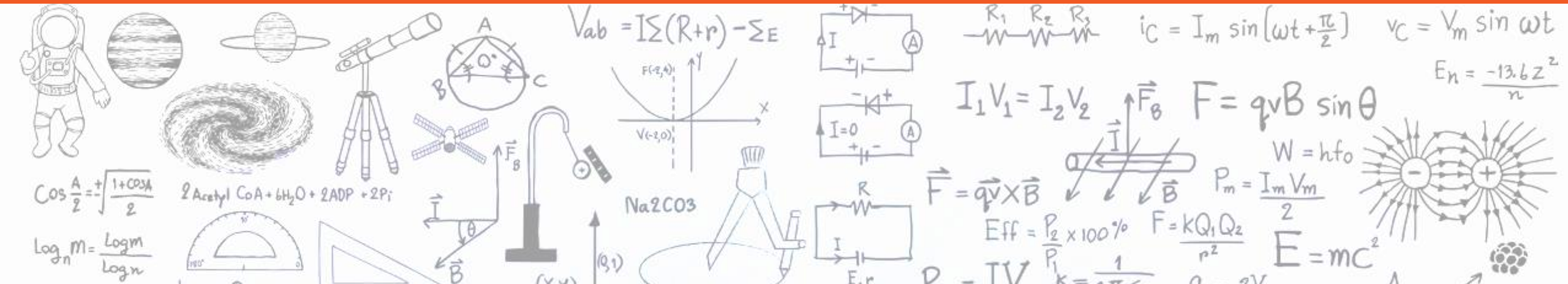
- Motivation
- Features
- Embedding in the FAIR ecosystem
- How to start
- Adoption
- Governance
- Learn more



DSW

DATA STEWARDSHIP WIZARD

Motivation



Motivation

Data Management "from burden to benefit"

- A good "Fahrplan" - knowing the steps and needed resources
- Avoiding forgetting something important
- DMP = what, why and how to do
- DMP ≠ an essay exercise to satisfy a funder
- From beginners to data stewards

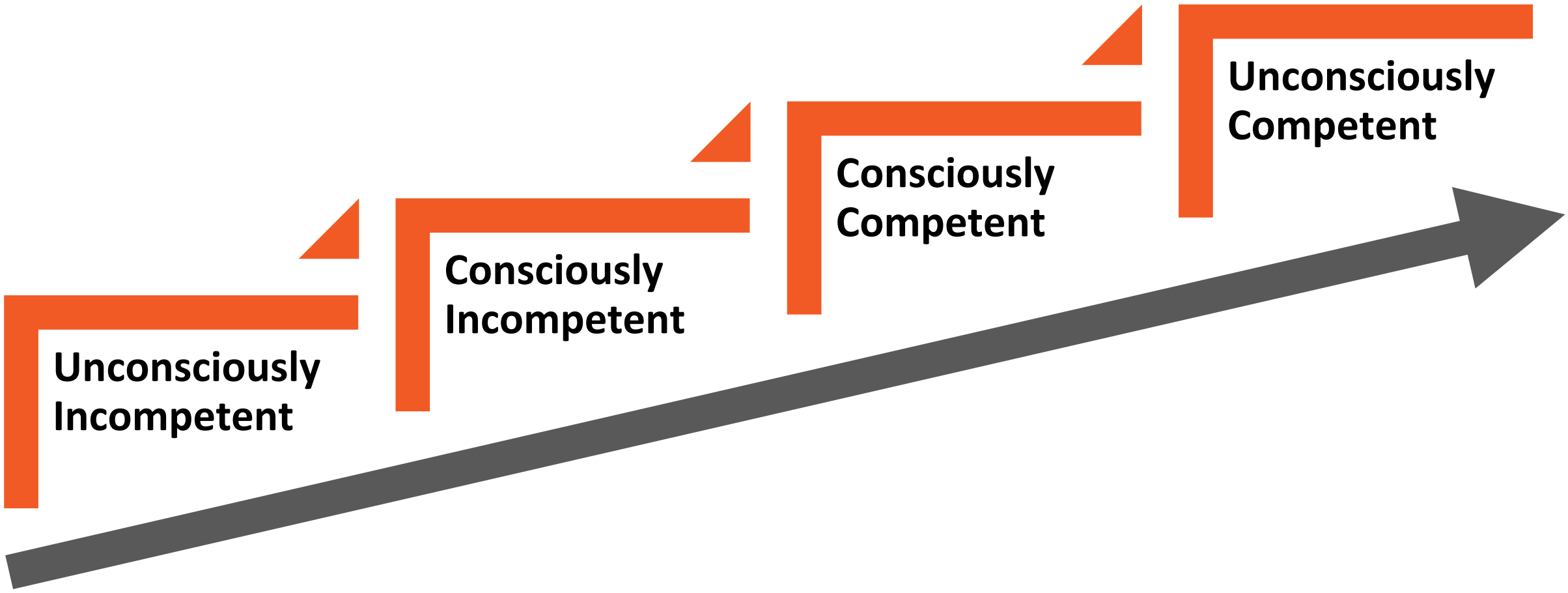
One size does not fit all

- Extensive customisation possibilities
- Openness (open-source code)

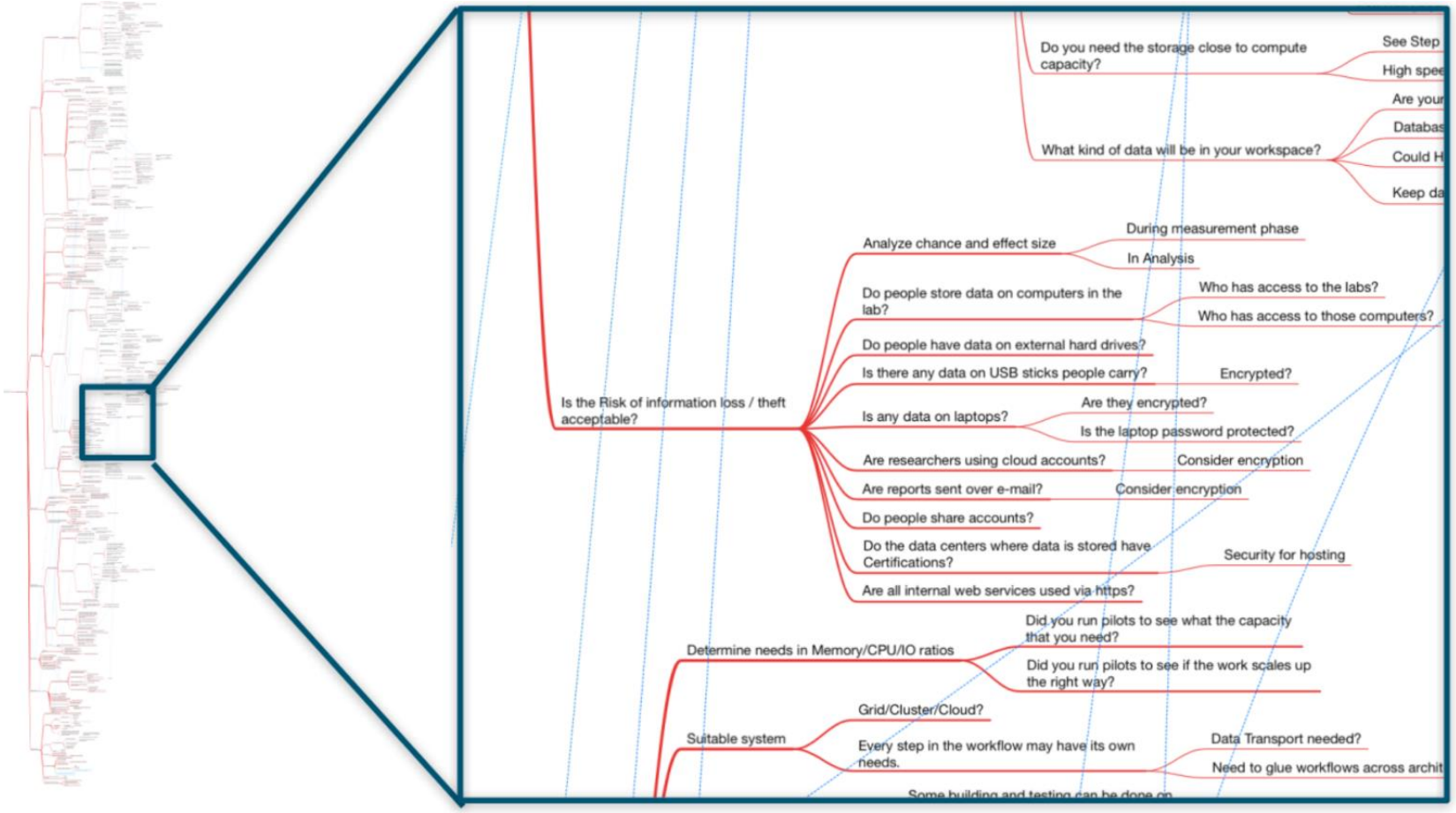
Challenge #1

How do you know your DMP is good?

DSW



Challenge #2 Making a good DMP encompasses a vast knowledge



- Rob Hooft (DTL, ELIXIR NL)
- Complex decision tree
- ~600 nodes

https://zenodo.org/record/2614820

DSW

DATA STEWARDSHIP WIZARD

Features

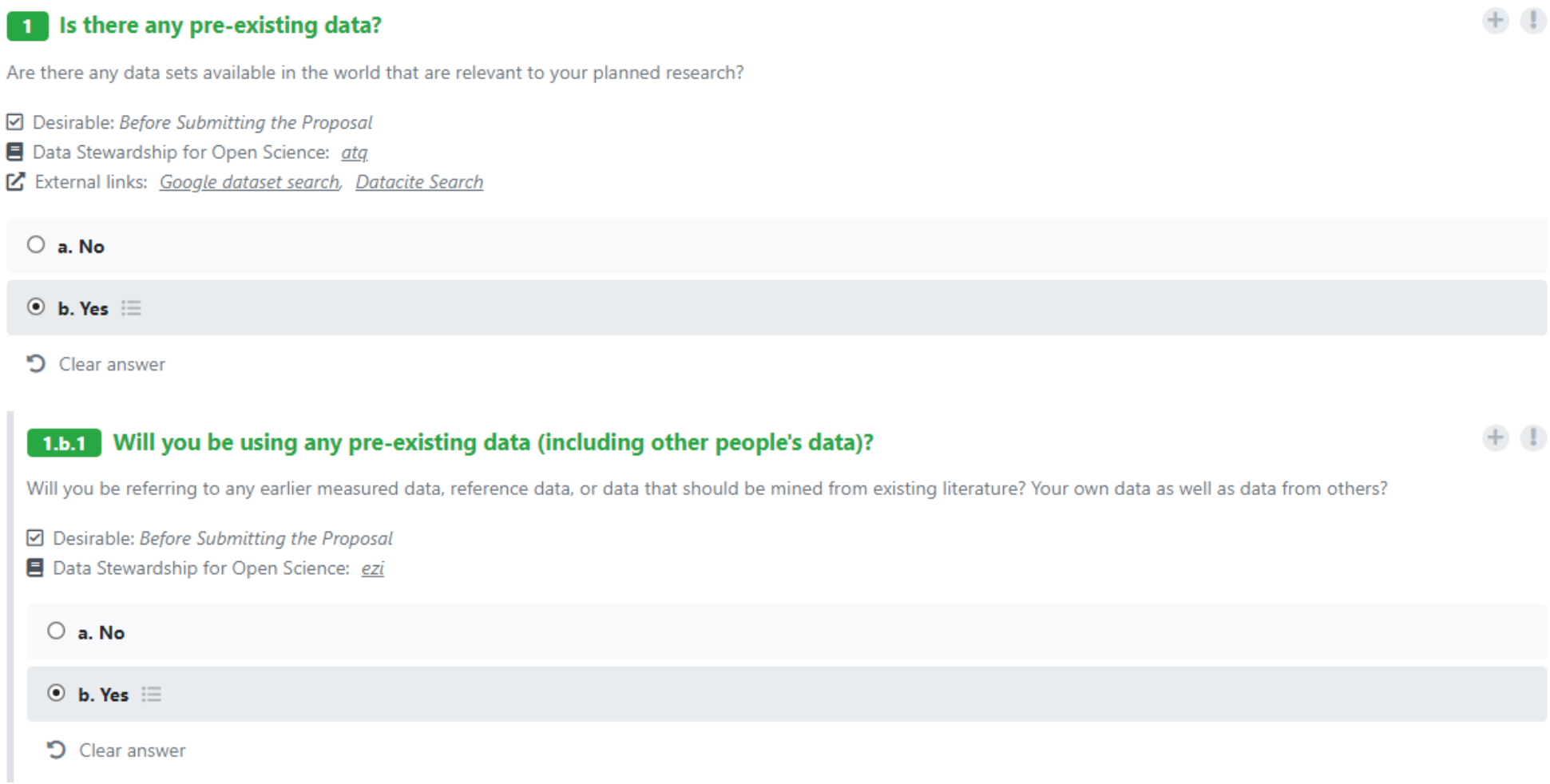
- **G**uidance
- **L**earning
- **E**ffectiveness

https://www.wikiwand.com

Guidance: Pilot's Check List

Adobe Stock | RioPatuca Images

Guidance: Smart Questionnaire



Guidance & Learning

Explanation

Project phases

References

Choice of Option

FAIR metrics

Recommendation

Learning

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

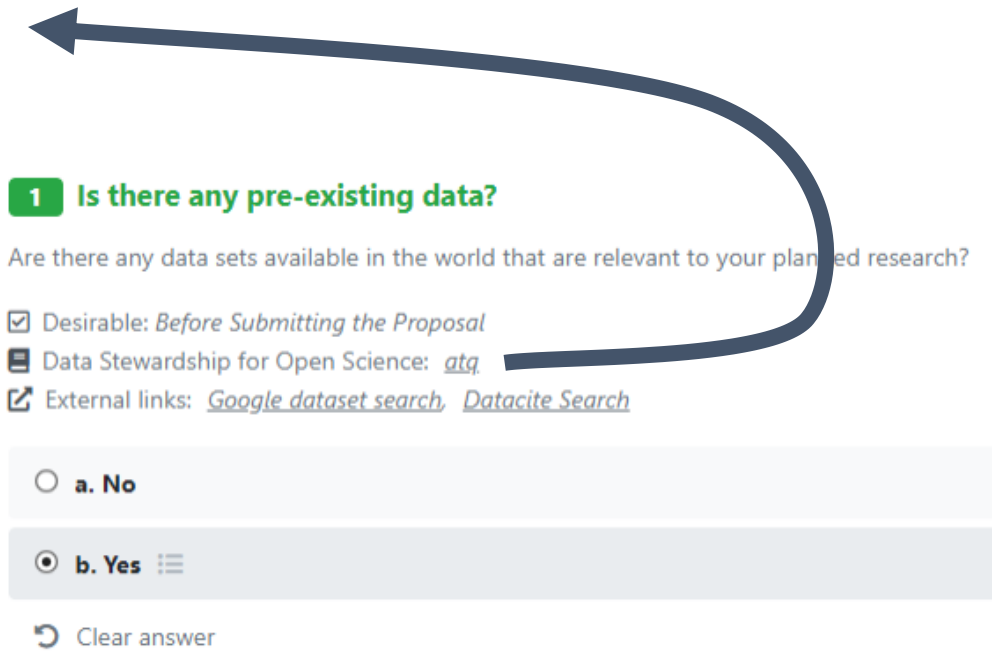
- Search for data sets (OPEDAS) that may be re-usable and can help you to reduce the number of new data sets you may have to generate (and steward later on).
- Include annotated collections of data and curated databases in your search.
- Check the accessibility and license situation attached to the relevant 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

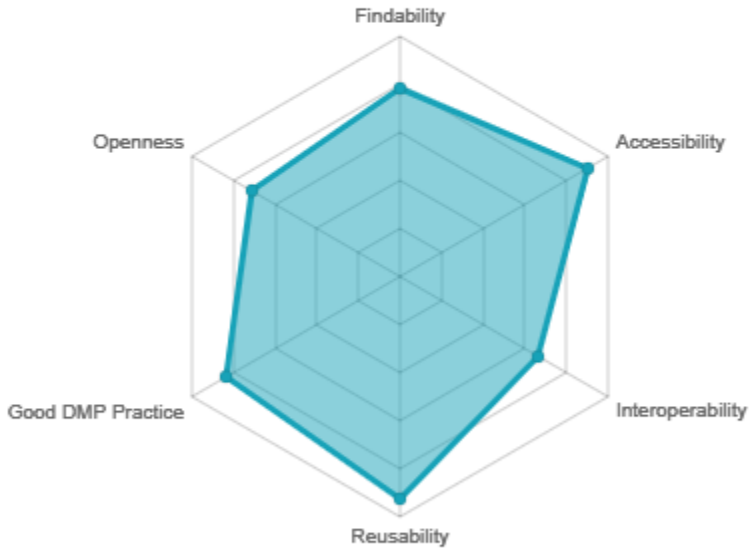
Guidance: Indications and Metrics

Summary Report

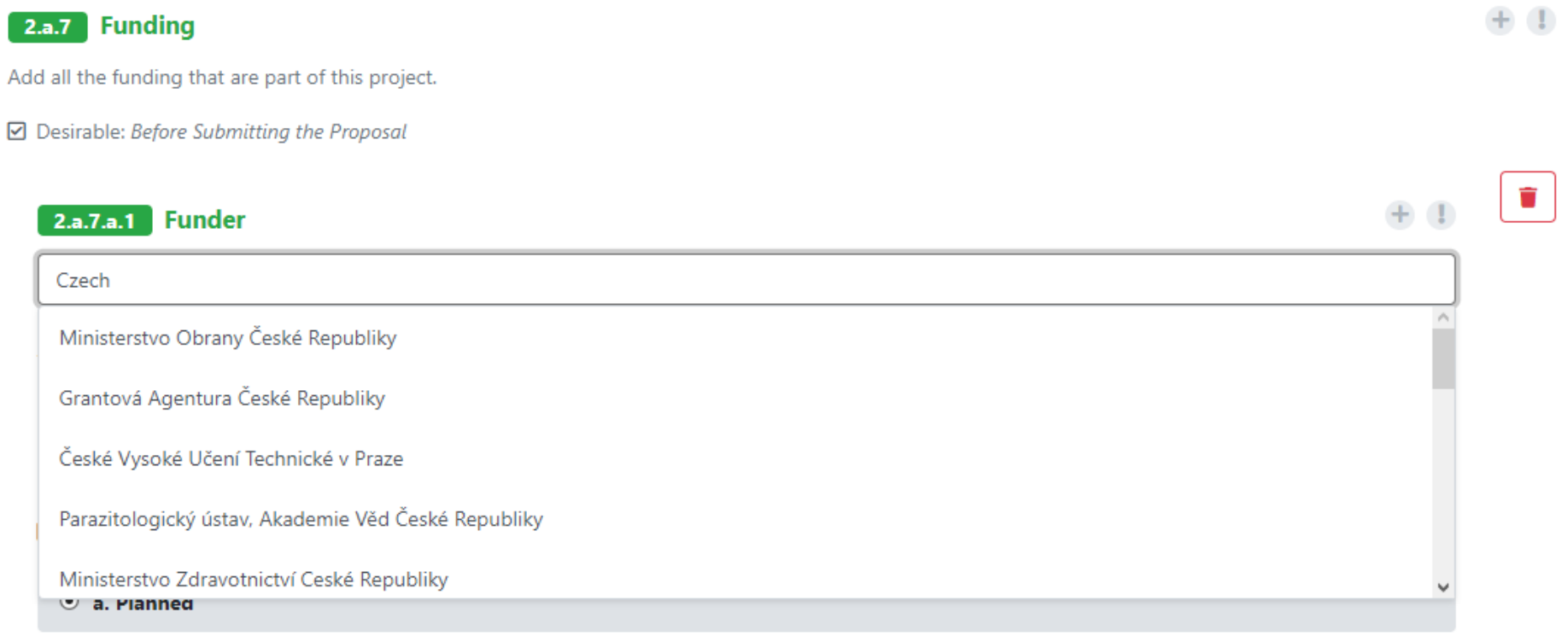
Answered (current phase): 100/100

Answered: 293/305

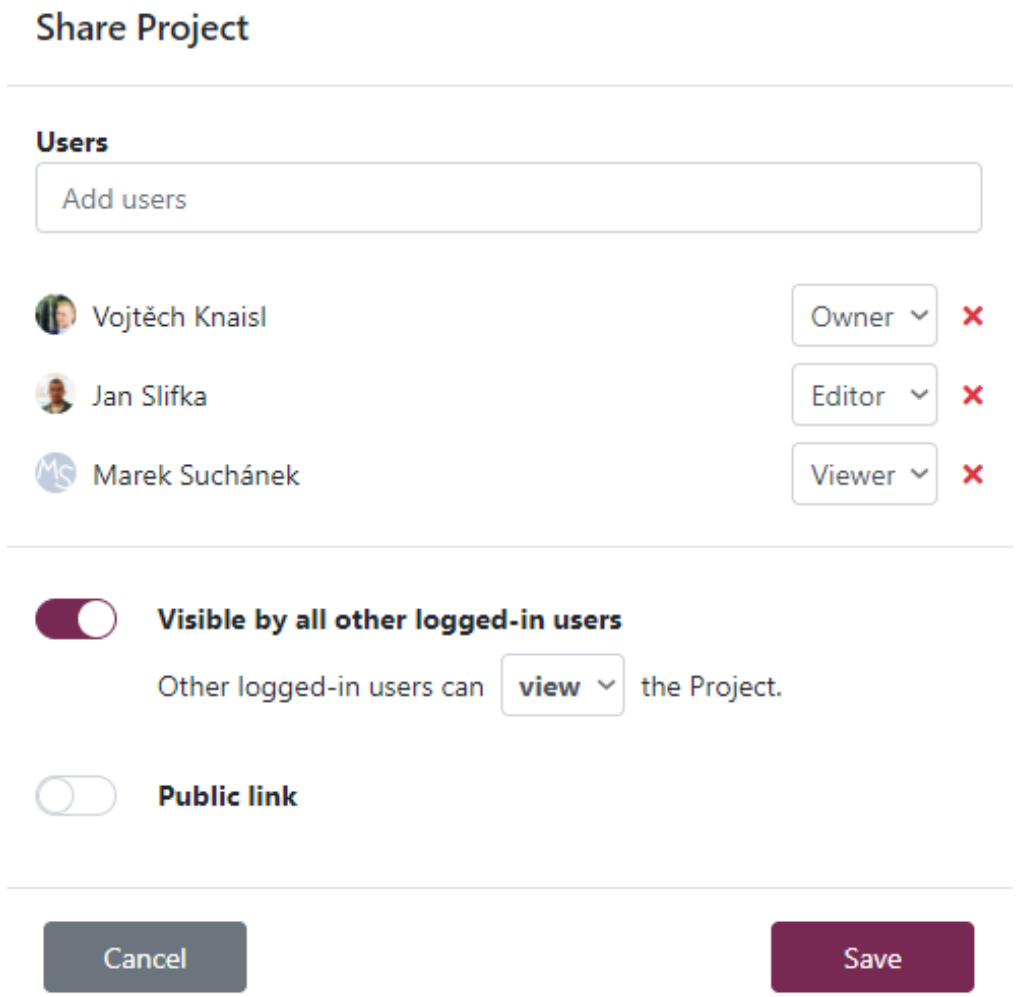
Metric	Measure
Findability	0.78
Accessibility	0.90
Interoperability	0.67
Reusability	0.92
Good DMP Practice	0.84
Openness	0.72



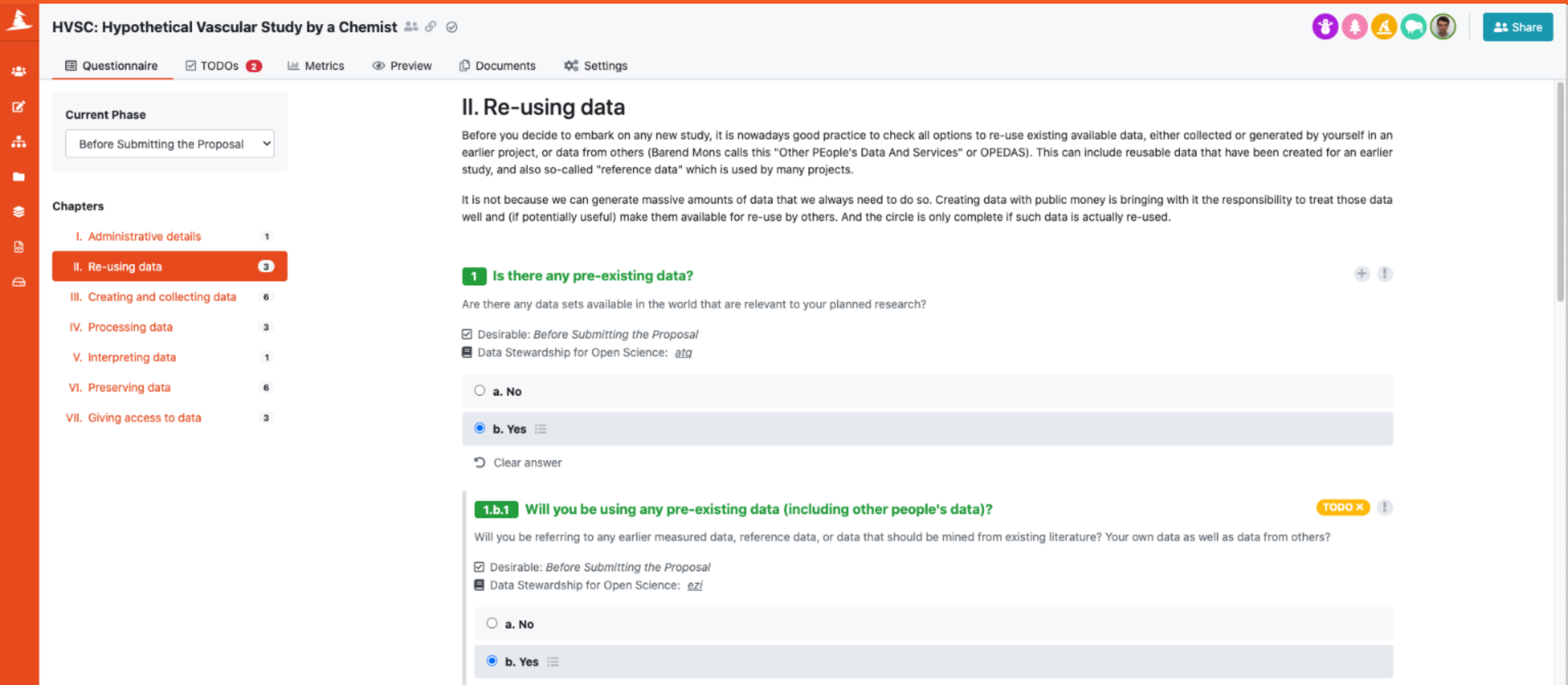
Effectiveness: Autocomplete



Effectiveness: Collaboration - Sharing



Effectiveness: Collaboration - Online



Effectiveness: TODOs

TODOs 4 | Version history

2 What existing encodings/terminologies/vocabularies/ontologies will you be using?

+ Add

3 Will you be using new types of data? TODO ×

Sometimes the type of data you collect can not be stored in a commonly used data format. In such cases you may need to make your own, keeping interoperability as high as possible.

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

- **a. No, all of my data will fit in common formats**
  Interoperability
- **b. Yes, I will need to use custom formats for some of my data**

Clear answer

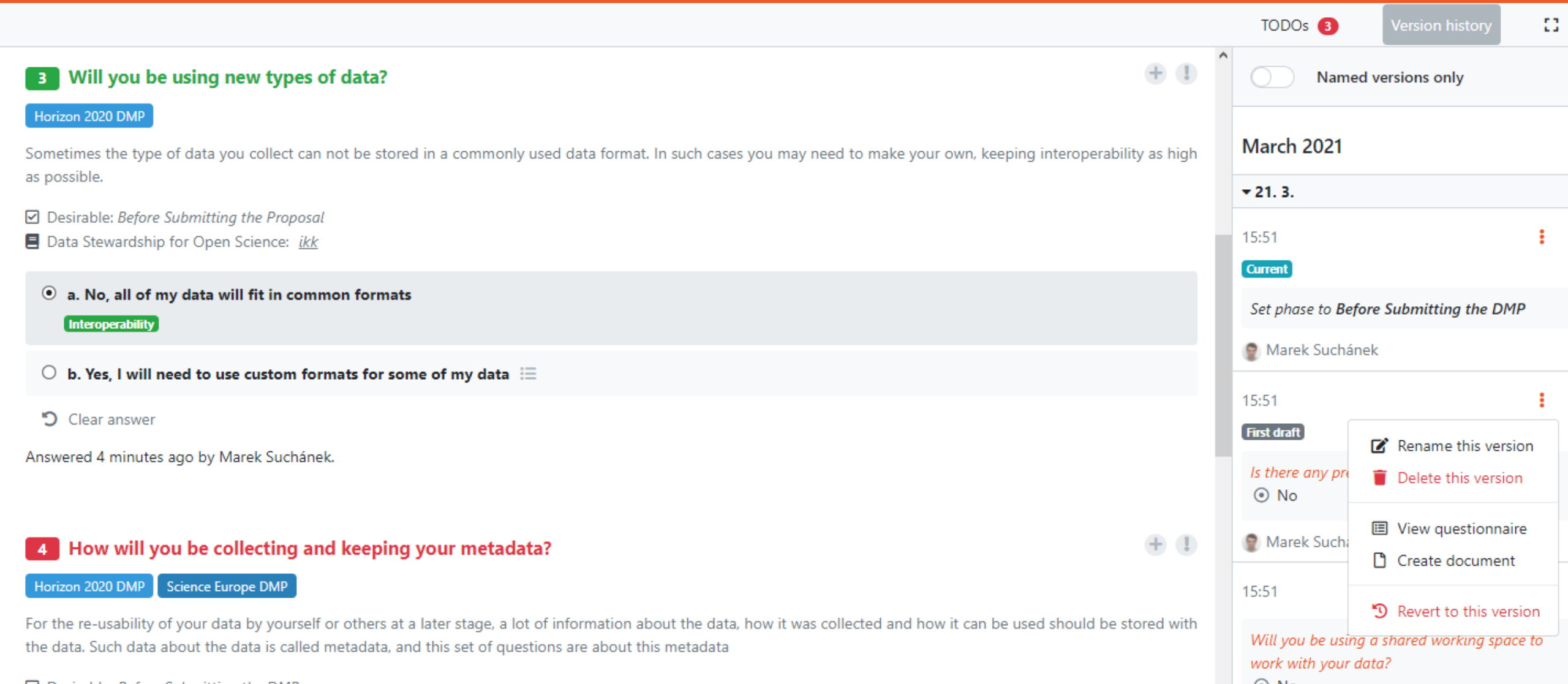
Re-using data
- Is there any pre-existing data?

Creating and collecting data
- Will you be using new types of data?

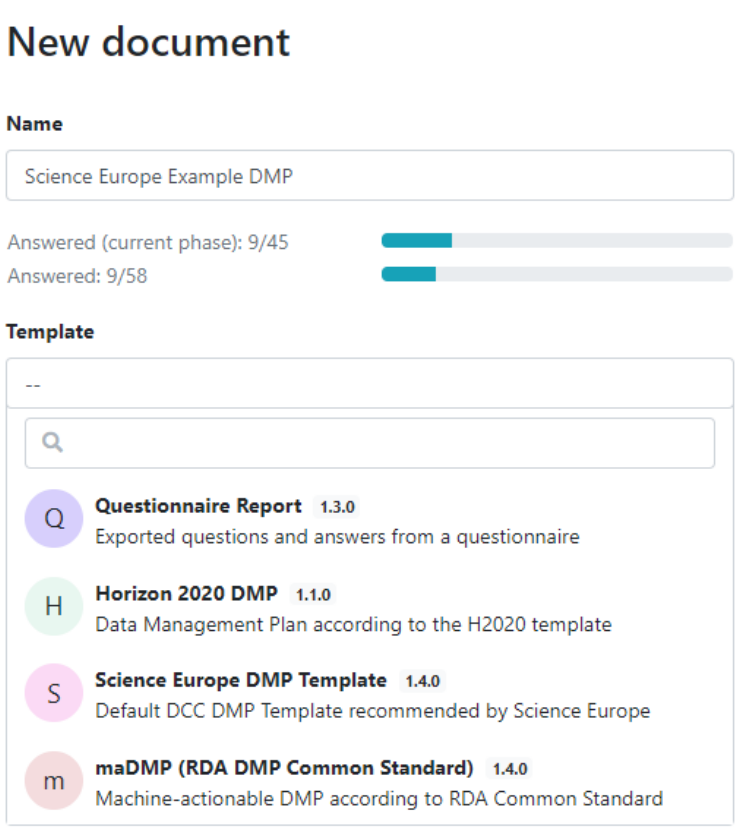
Processing data
- Data storage systems and file naming conventions

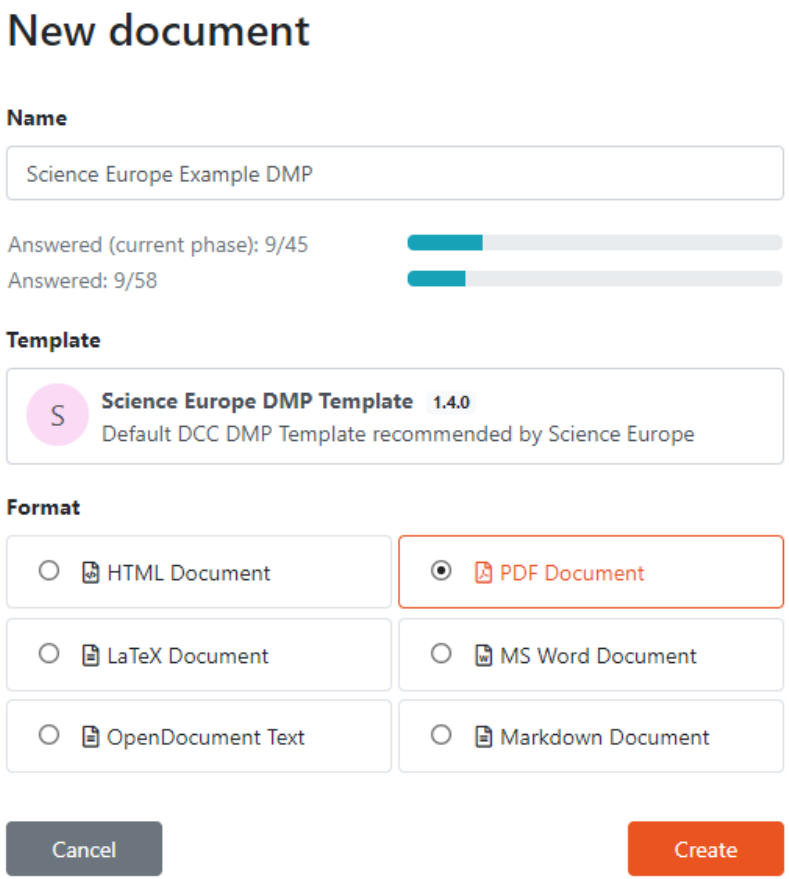
Giving access to data
- Can all of your data become completely open immediately?

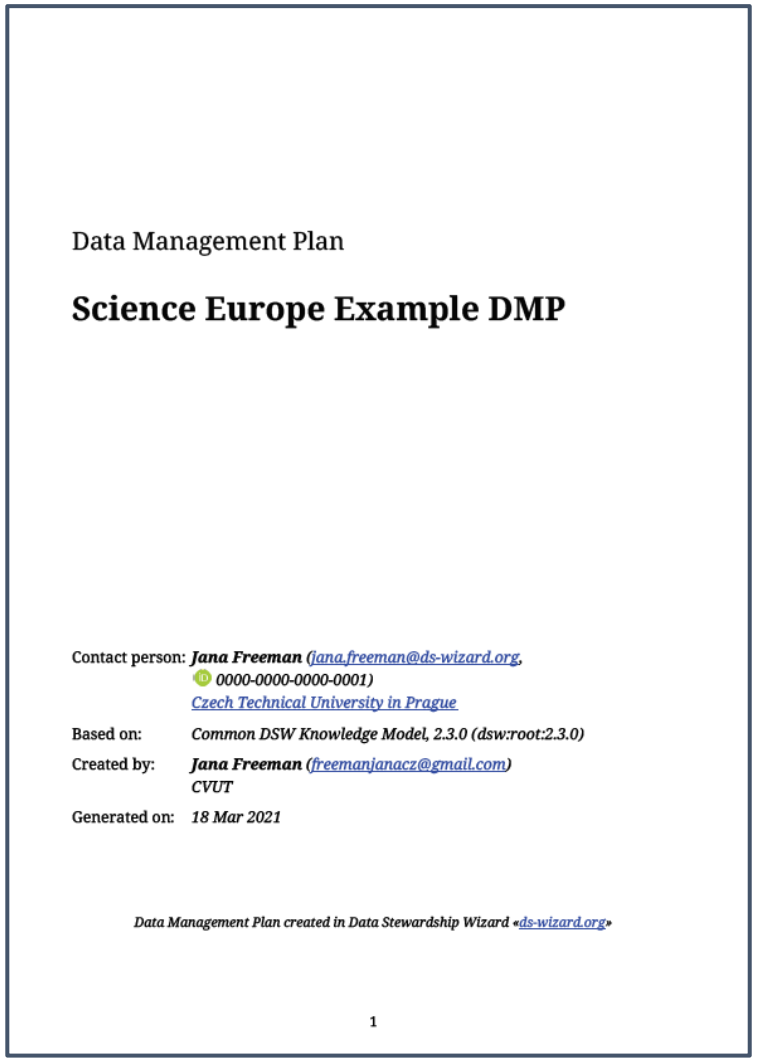
Effectiveness: Version history



Effectiveness: DMP Export in Various Templates







Data Management Plan

Science Europe Example DMP

Contact person: ***Jana Freeman*** *(jana.freeman@ds-wizard.org,*
0000-0000-0000-0001)
Czech Technical University in Prague

Based on: *Common DSW Knowledge Model, 2.3.0 (dsw:root:2.3.0)*

Created by: ***Jana Freeman*** *(freemanjanacz@gmail.com)*
CVUT

Generated on: *18 Mar 2021*

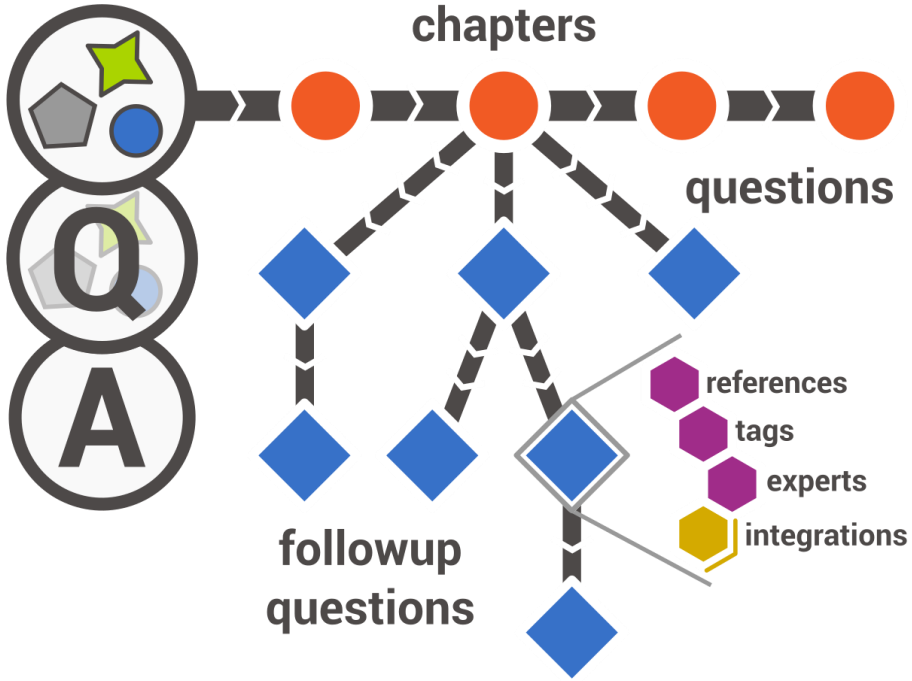
Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»

1

One Size Does not Fit All

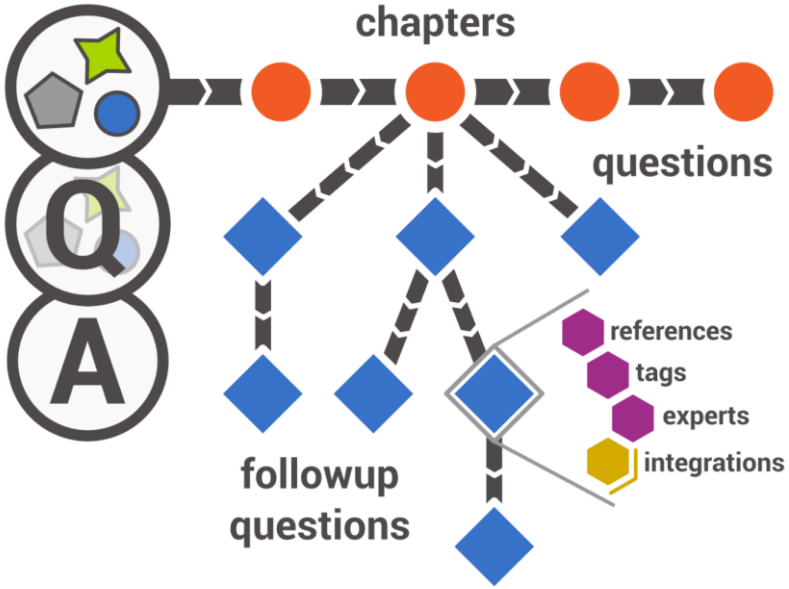
The Questionnaire is fully customisable

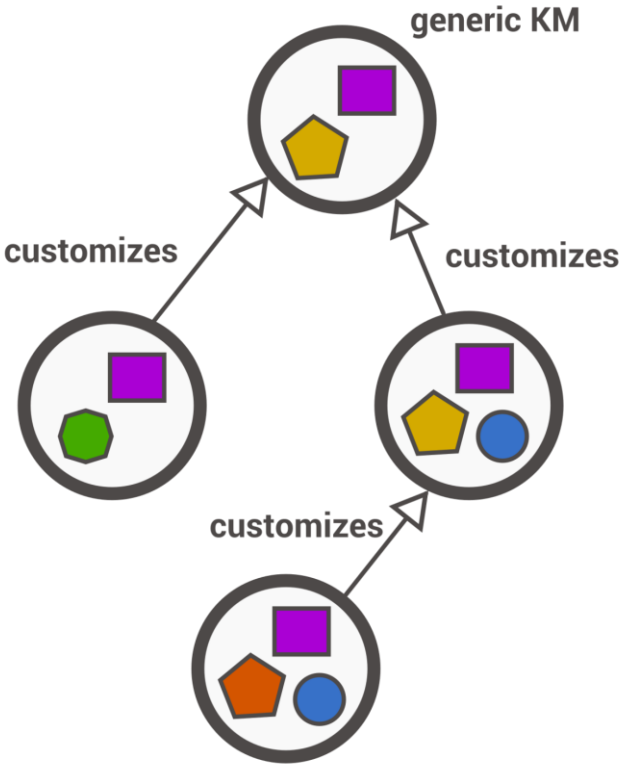
- Chapters
- Questions (different types)
- Answers (choice of option)
- References and Experts
- Tags
- Integration for answer suggesting



One Size Does not Fit All

- The questionnaire is fully customisable
- Built-in Knowledge Model Editor
- Create „from the scratch“
- Edit existing ones:
  - Delete non-relevant questions
  - Add relevant ones
  - Change existing questions
  - Structure change

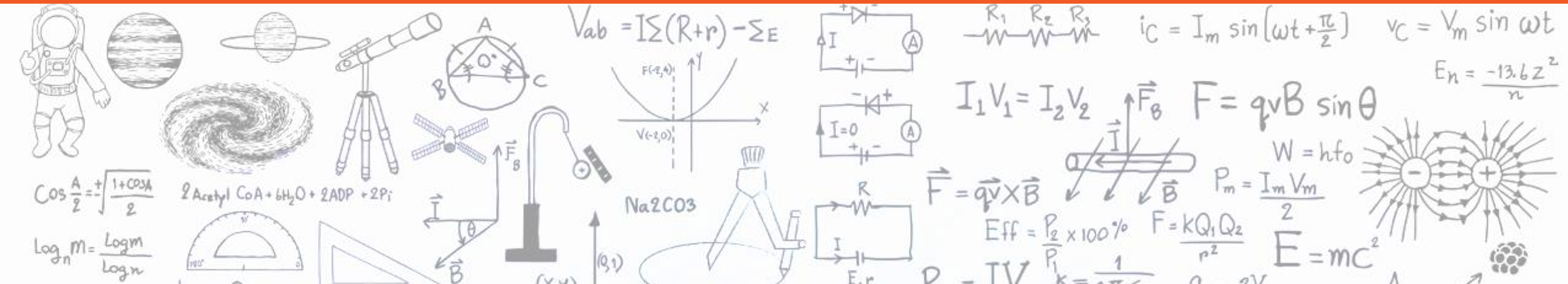




DSW

DATA STEWARDSHIP WIZARD

Embedding in the FAIR Ecosystem



Integrations through REST API

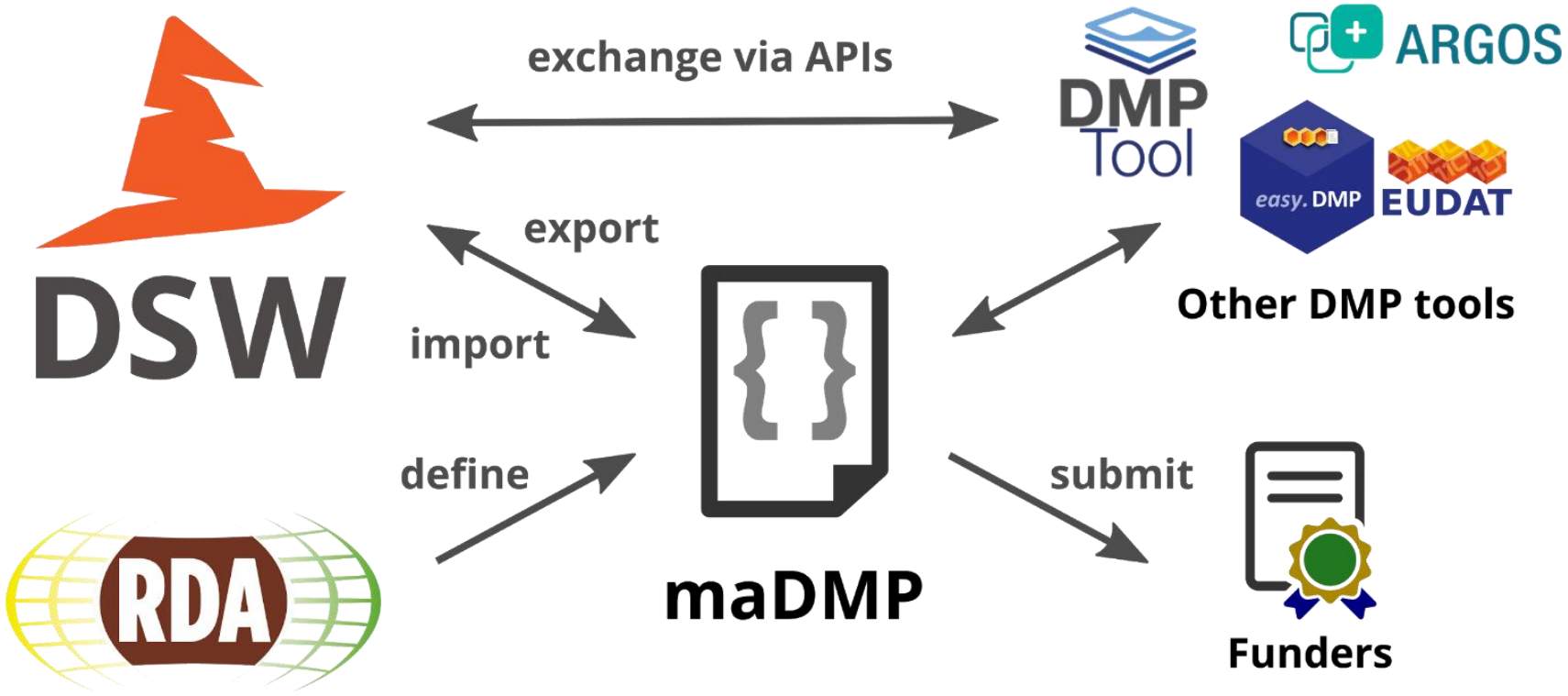
FAIRsharing.org
standards, databases, policies

Crossref

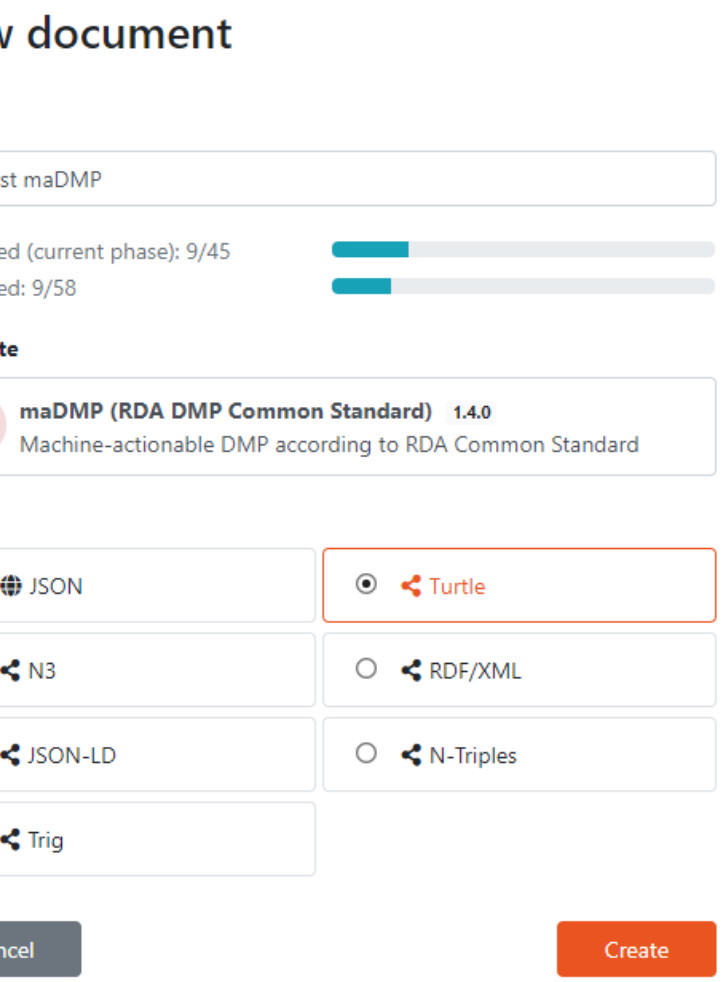
ROR

WIKIDATA

Machine-Actionable DMPs



Machine-Actionable DMPs




```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix dcso: <https://w3id.org/dcso/ns/core#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcs-lang: <https://w3id.org/dcso/id/iso639-3/> .

# DMP
<https://demo.ds-wizard.org/questionnaires/b3d89775-64aa-4342-916e-c010629036f7>
  a dcso:DMP ;
  dct:title "Marek" ;
  dct:description "This maDMP has been created using Data Stewardship Wizard (DSW, ds-wizard.
  org) and is based on knowledge model Common DSW Knowledge Model (dsw:root:2.3.0). The
  questionnaire used for this DMP is identified by UUID
  \"b3d89775-64aa-4342-916e-c010629036f7\" within https://demo.ds-wizard.org DSW instance." ;
  dct:language dcs-lang:eng ;
  dcso:ethicalIssuesExist "unknown" ;
  dcso:hasDMPId <https://demo.ds-wizard.org/questionnaires/
  b3d89775-64aa-4342-916e-c010629036f7/identifier> ;
  dct:created "2021-03-22T07:11:40Z"^^xsd:dateTime ;
  dct:modified "2021-03-22T07:11:40Z"^^xsd:dateTime .

# DMP identifier
<https://demo.ds-wizard.org/questionnaires/b3d89775-64aa-4342-916e-c010629036f7/identifier> a
dcso:DMPId ;
  dcso:identifier_type "url" ;
  dct:identifier "https://demo.ds-wizard.org/questionnaires/
  b3d89775-64aa-4342-916e-c010629036f7" .

# DMP contact


# This file has been exported from Data Stewardship Wizard (ds-wizard.org)
# It is using DCSO 3.0.2 (see https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard)
```

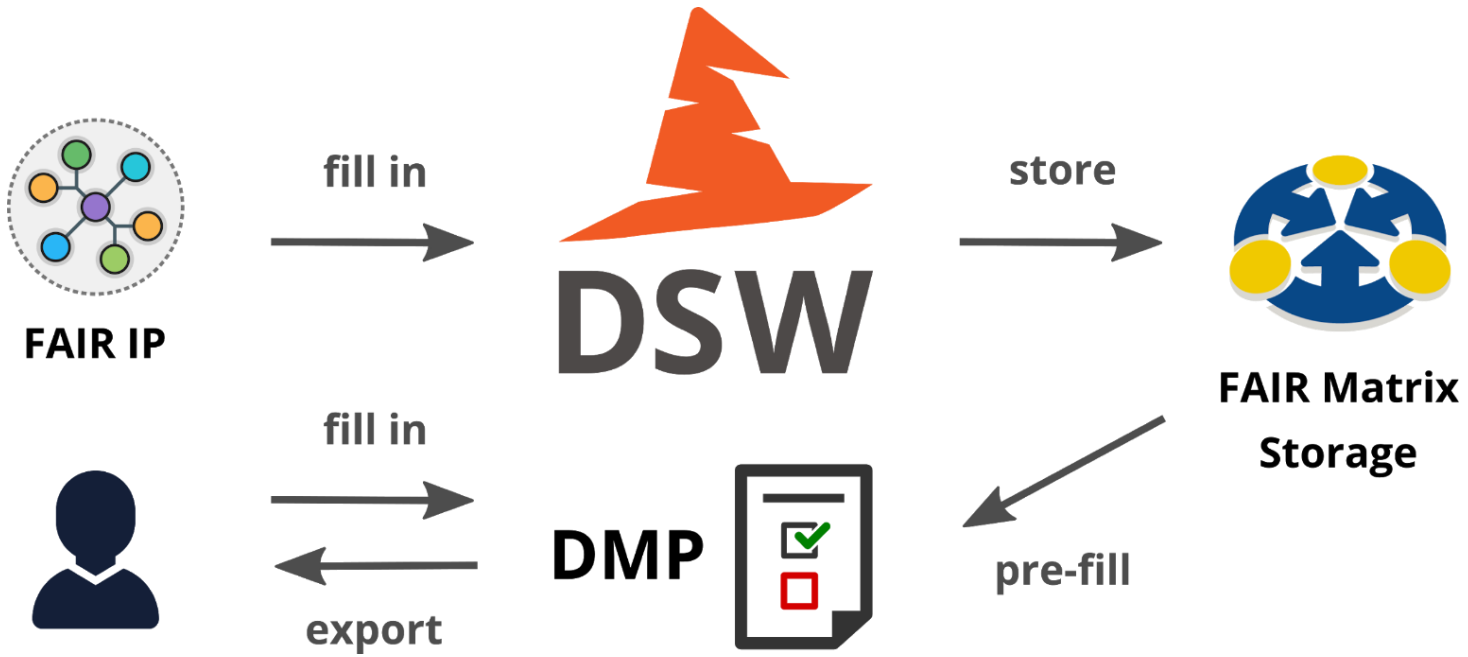
FIP Wizard

- **FAIR Implementation Profile** represents the implementation strategies of a FAIR community
- Such implementation strategies entail decisions about which FAIR-enabling resources are used, planned-to-be-used, or not-used

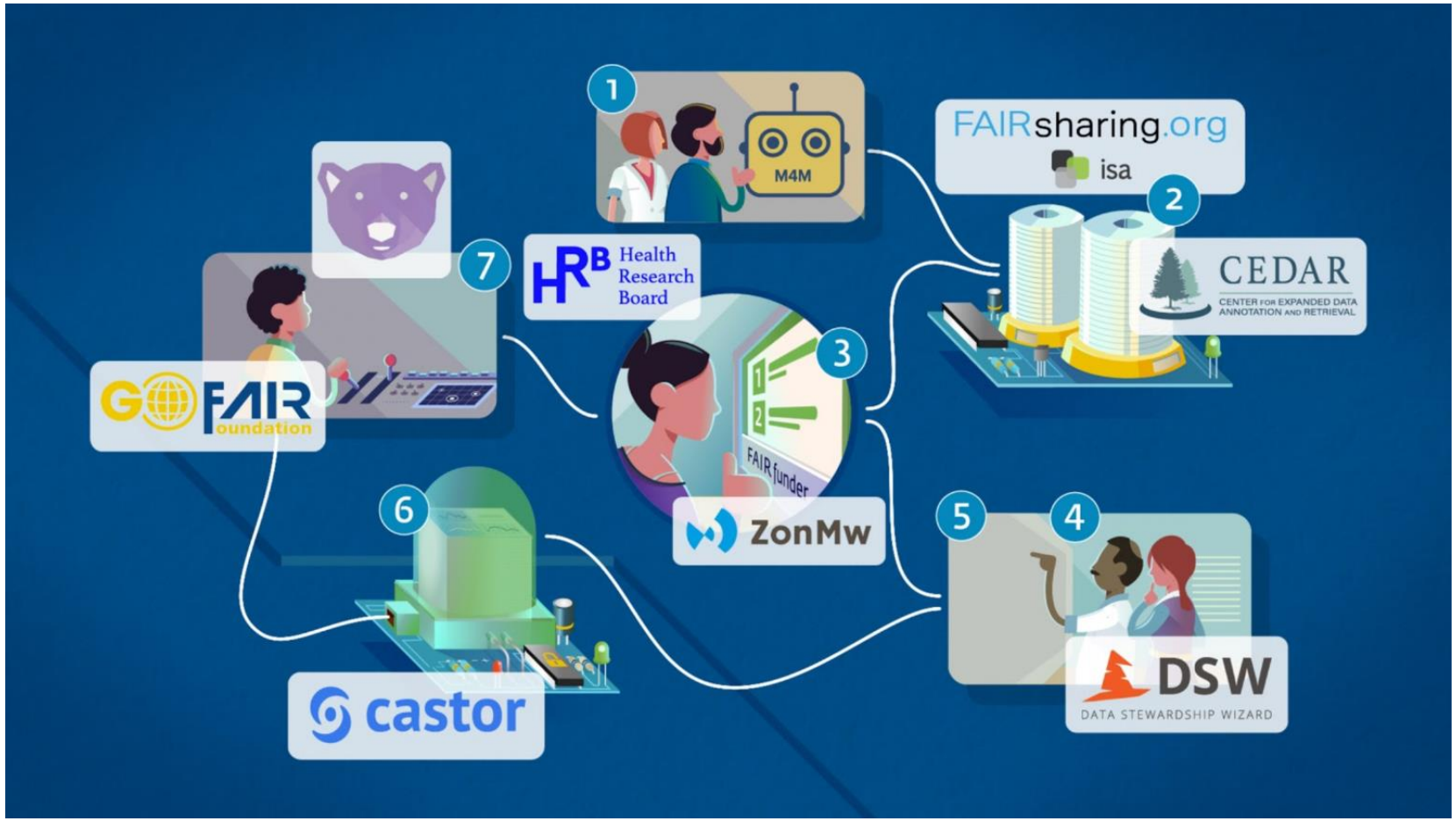
"Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence"

https://link.springer.com/chapter/10.1007/978-3-030-65847-2_13

FIP Wizard


FAIR IP
fill in
DSW
store
FAIR Matrix Storage
fill in
DMP
export
pre-fill

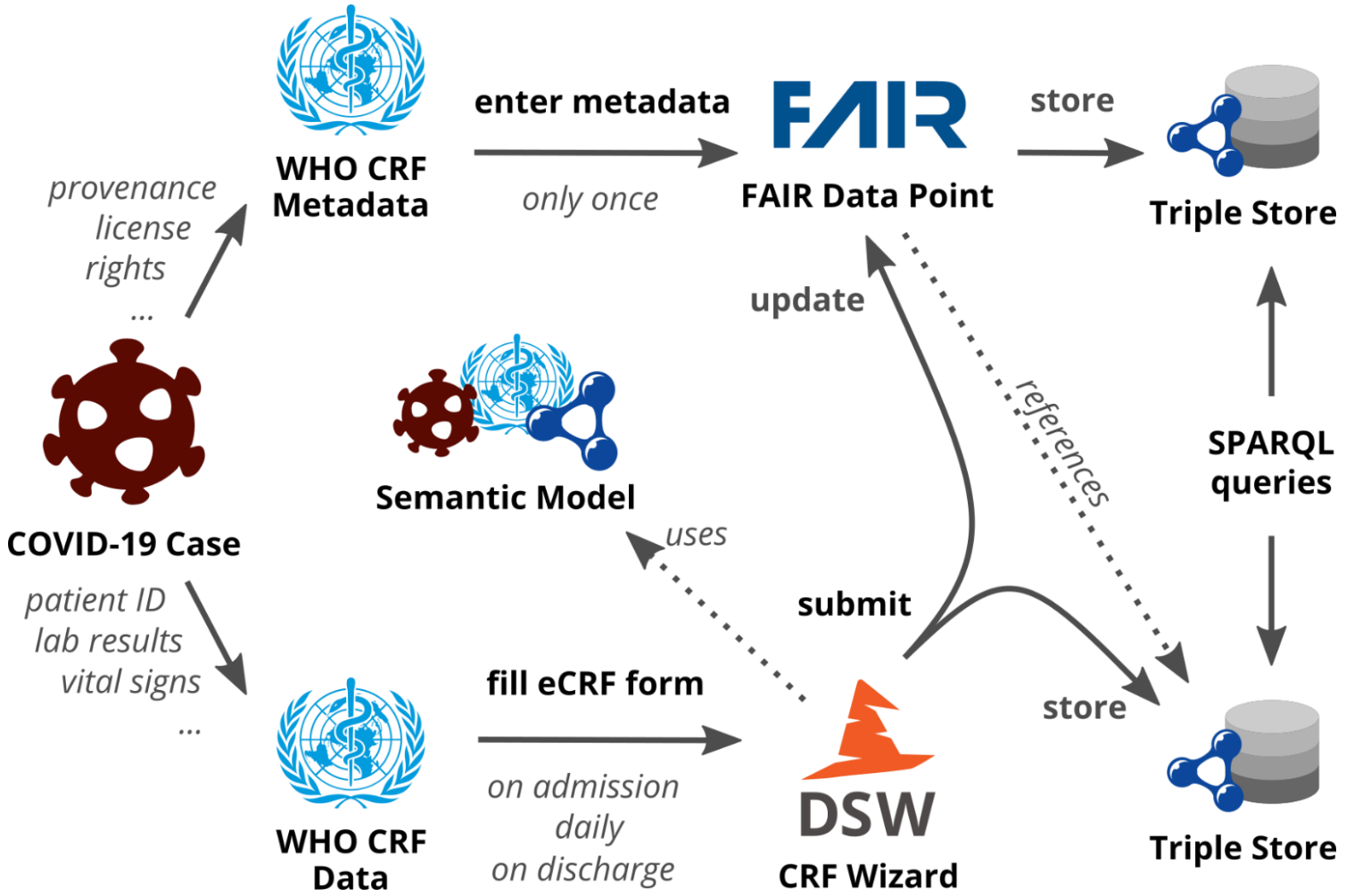
FAIR Funder Study



https://vimeo.com/toldt/download/356442595/b217c82a74

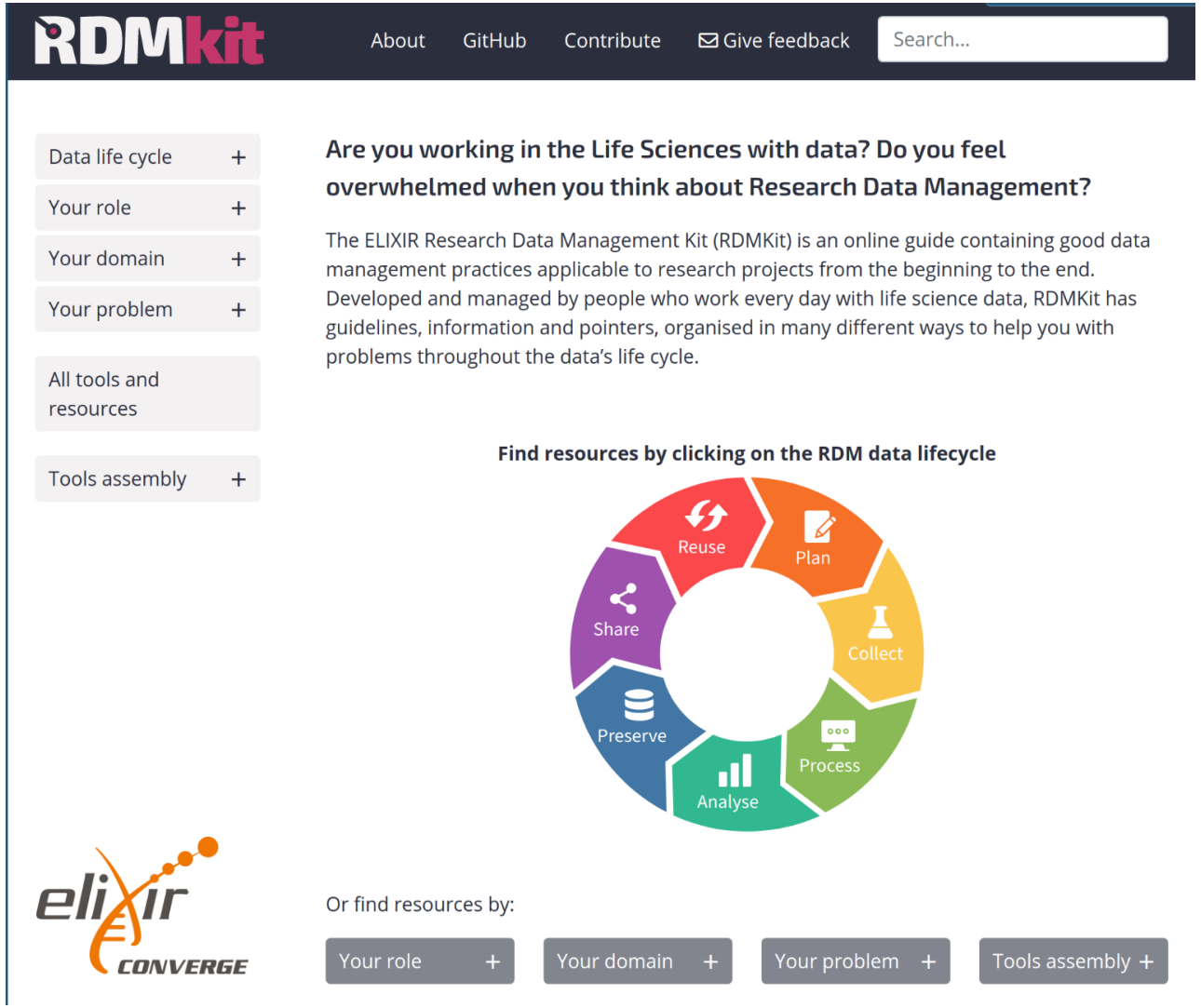
https://arxiv.org/abs/1902.11162

CRF Wizard



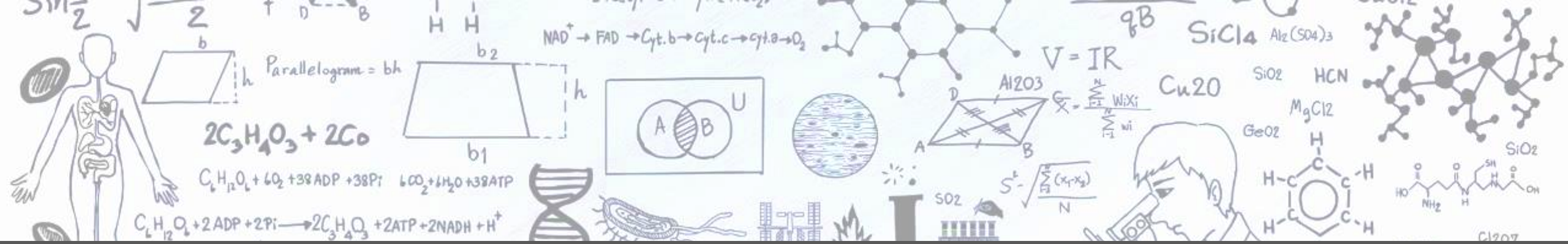
https://www.gofairfoundation.org/vodan-in-a-box-the-all-in-one-solution-for-easy-instalment-of-vodan-fair-data-points

ELIXIR CONVERGE



https://rdmkit.elixir-europe.org

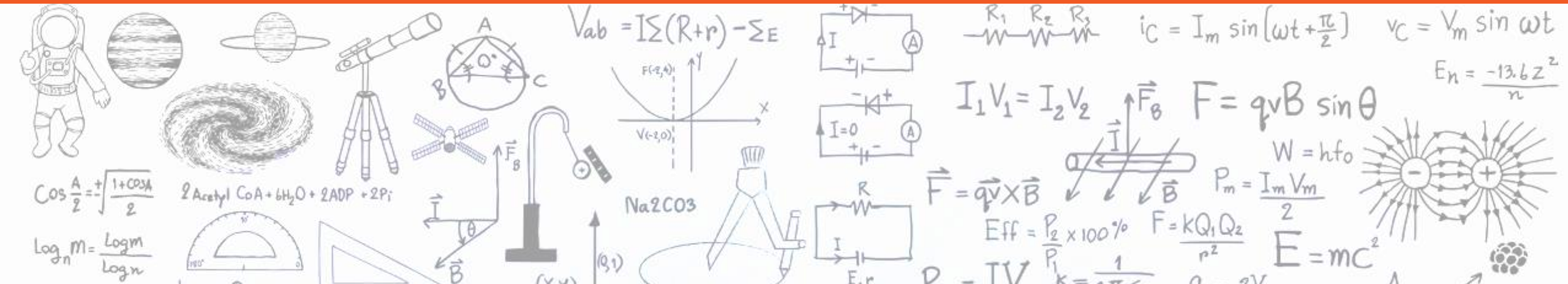
- We will use guided choices through the DSW to help people find the ELIXIR Tools and Toolkits most appropriate for their case.
- The Wizard is supposed to ask the same questions that an expert would ask if a researcher comes to ask for help.



DSW

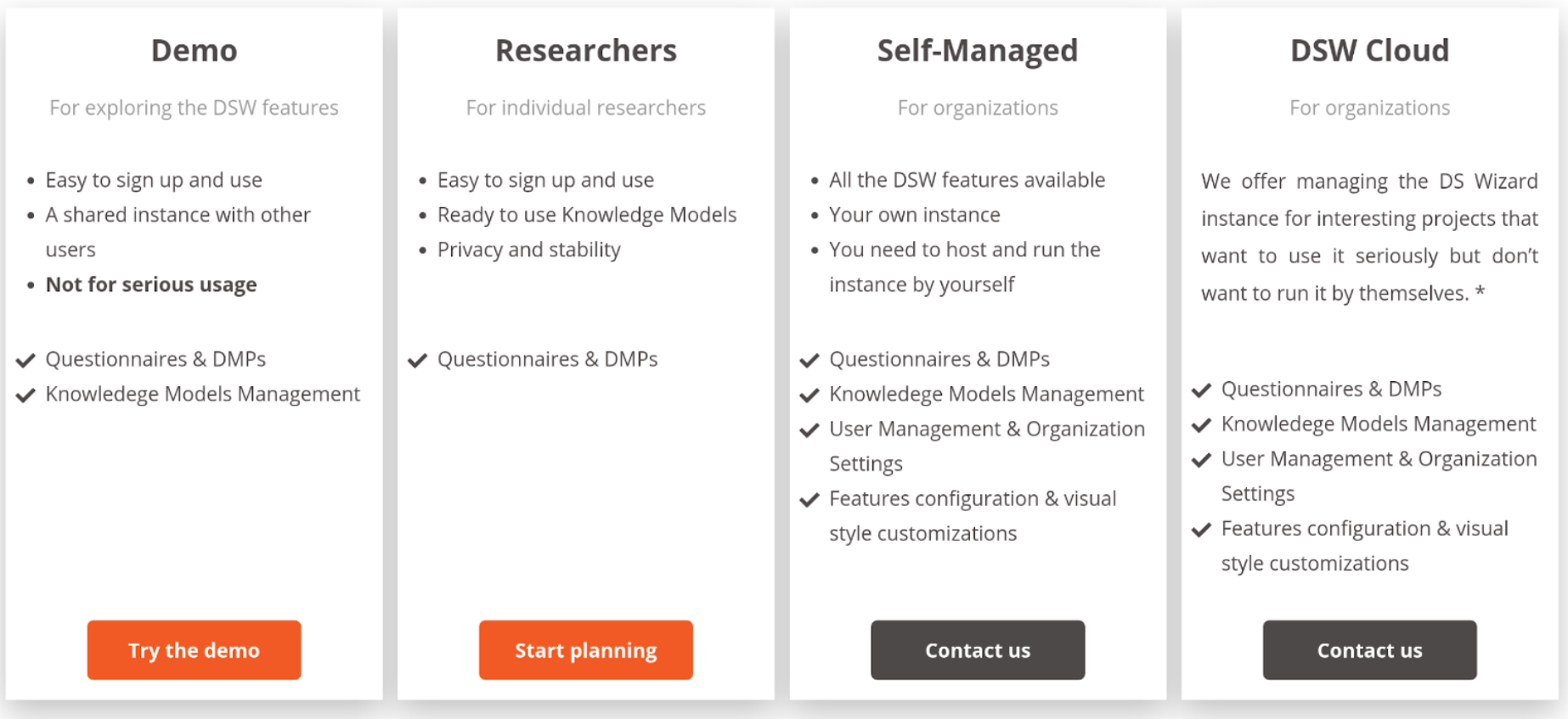
DATA STEWARDSHIP WIZARD

How to Start



How to Start

https://ds-wizard.org/get-started.html



Demo

For exploring the DSW features

- Easy to sign up and use
- A shared instance with other users
- **Not for serious usage**

- ✔ Questionnaires & DMPs
- ✔ Knowledege Models Management

Try the demo

Researchers

For individual researchers

- Easy to sign up and use
- Ready to use Knowledge Models
- Privacy and stability

- ✔ Questionnaires & DMPs

Start planning

Self-Managed

For organizations

- All the DSW features available
- Your own instance
- You need to host and run the instance by yourself

- ✔ Questionnaires & DMPs
- ✔ Knowledege Models Management
- ✔ User Management & Organization Settings
- ✔ Features configuration & visual style customizations

Contact us

DSW Cloud

For organizations

We offer managing the DS Wizard instance for interesting projects that want to use it seriously but don't want to run it by themselves. *

- ✔ Questionnaires & DMPs
- ✔ Knowledege Models Management
- ✔ User Management & Organization Settings
- ✔ Features configuration & visual style customizations

Contact us

Free for research institutions (ELIXIR funded)

Going professional

DSW

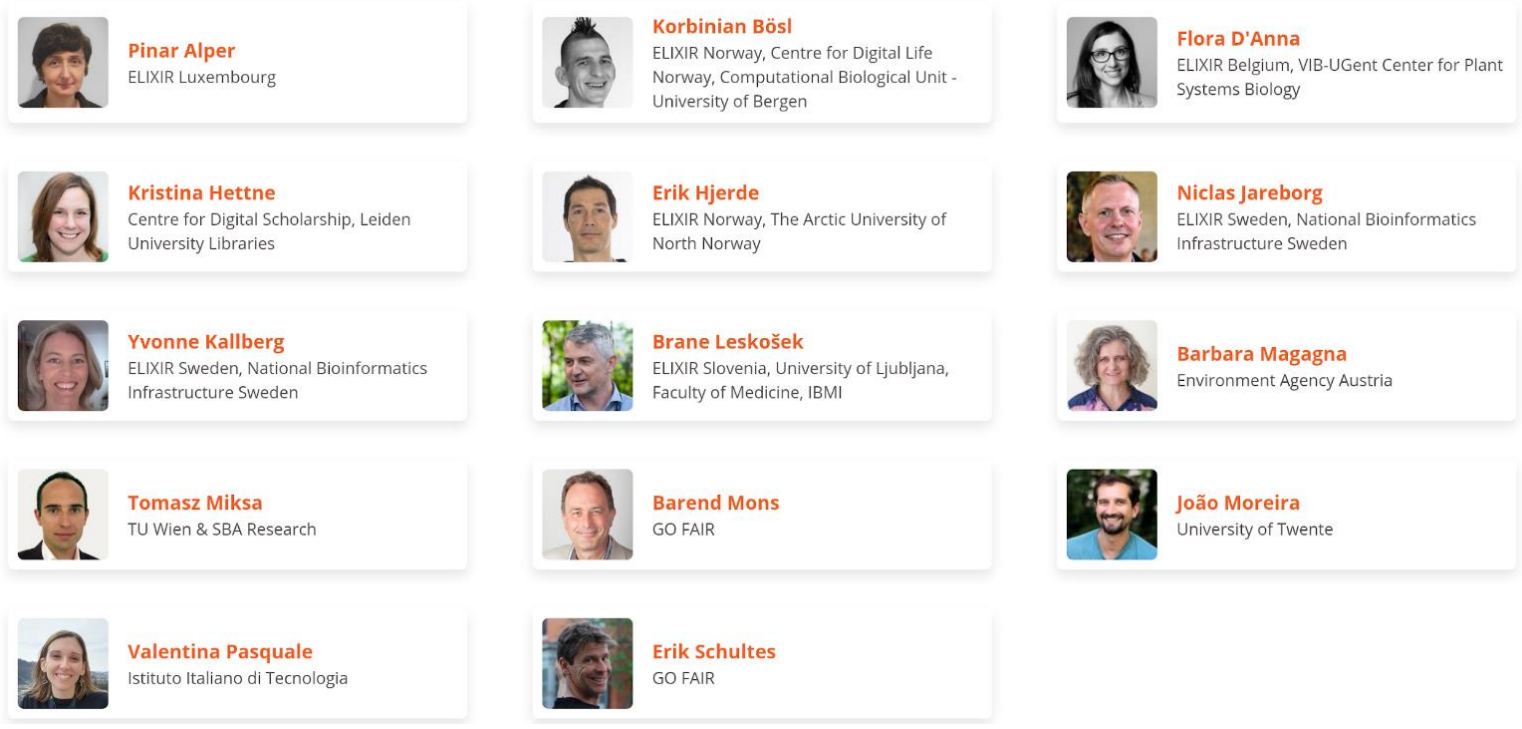
- Codevence professional services
  - Professional cloud instances with custom branding & styling
  - Deployment and configuration support
  - Consulting
  - Guaranteed user support
  - Tailored training plans
  - Custom document templates and exports
  - Integration into company infrastructure
  - Custom features development

Adoption

- ~1500 registered users in the ELIXIR DEMO/Researchers cloud instances
- 14 institutions having their own cloud version
- ~15 known on-premises installations

Governance

- Strongly user-driven:
  - User feedback & bug report – from the UI, goes into GitHub
  - User ideas https://ideas.ds-wizard.org
- Advisory Board



Learn More

- Features overview: https://ds-wizard.org/features.html
- Resources (webinars, workshops, presentations, publications, posters): https://ds-wizard.org/resources.html
- Comparison with other tools: https://ds-wizard.org/comparison.html

DSW Core Team

Robert Pergl
Project Coordinator

Rob Hooft
DS Knowledge Expert

Vojtěch Knaisl
Backend Developer

Jan Slifka
Frontend Developer

Marek Suchánek
Leonardo da Vinci

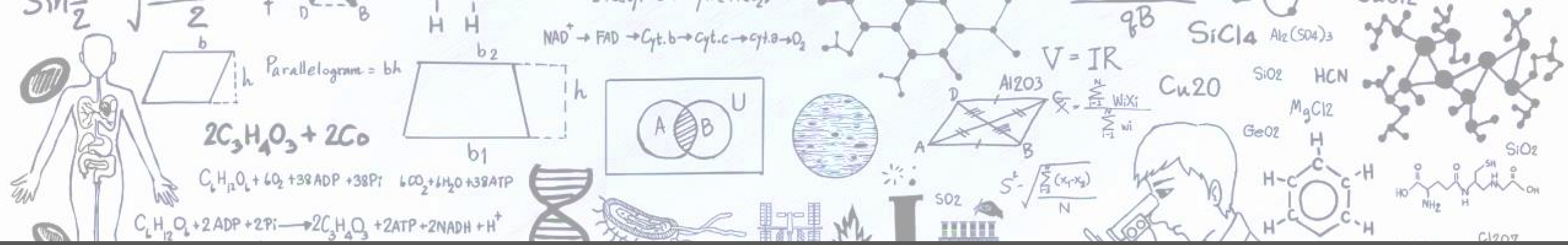
Jana Freeman
Community Manager

Tereza Machačová
Intern

Jiří Vondrášek
Business and Strategic Advisory

Acknowledgement

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



DSW

DATA STEWARDSHIP WIZARD

Thank you!

info@ds-wizard.org