

Data Management with Data Stewardship Wizard



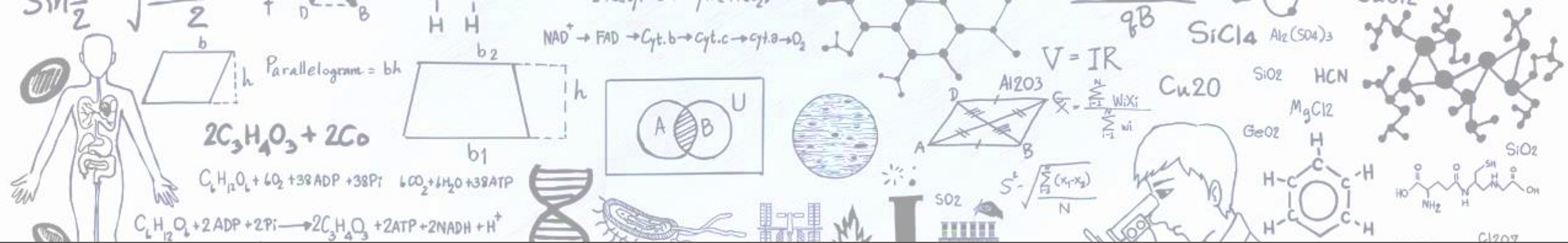
**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**



CODEVENCE

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

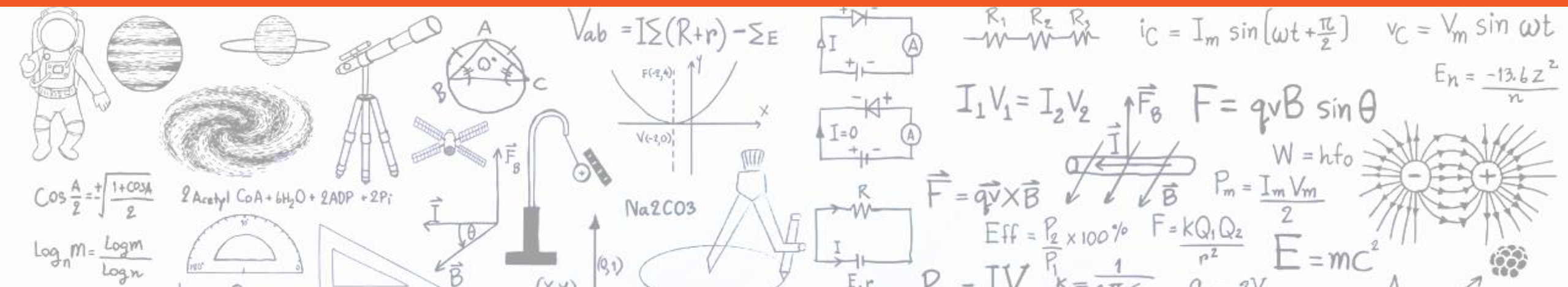
- Motivation
- Features
- Embedding in the FAIR ecosystem
- How to start
- Adoption
- Governance
- Learn more



DSW

DATA STEWARDSHIP WIZARD

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 \neq an essay exercise to satisfy a funder
- From beginners to data stewards

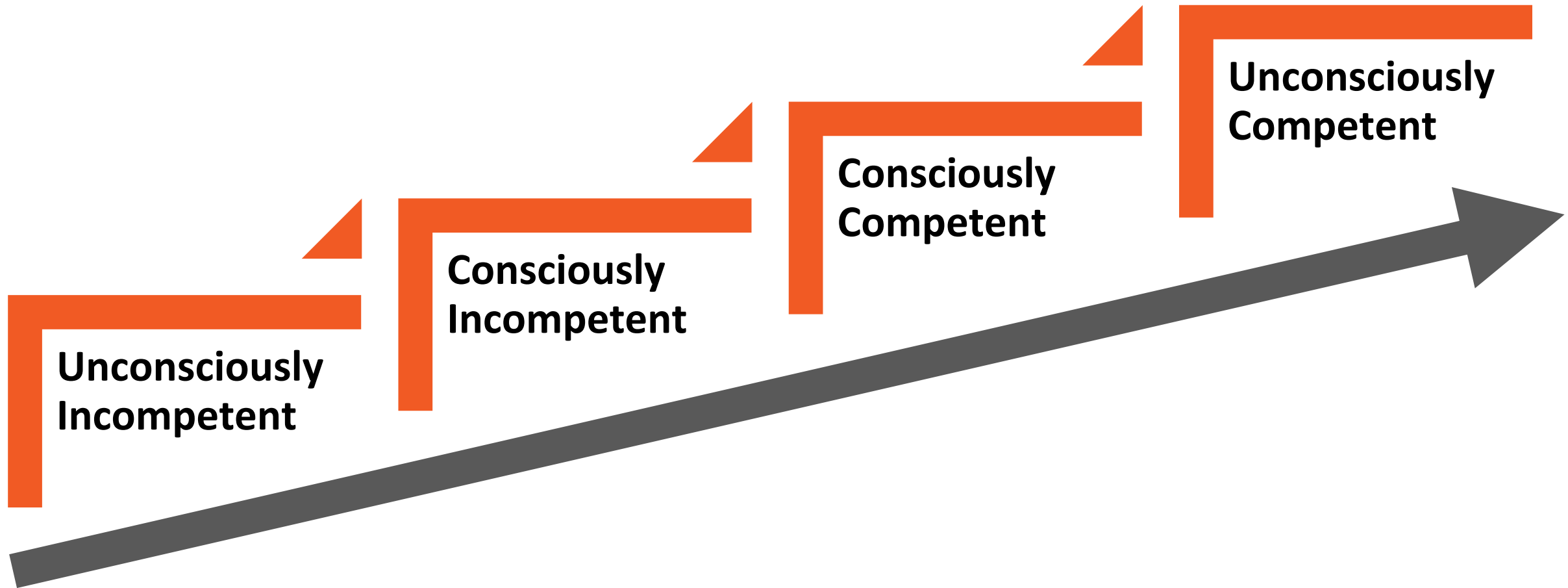
One size does not fit all

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

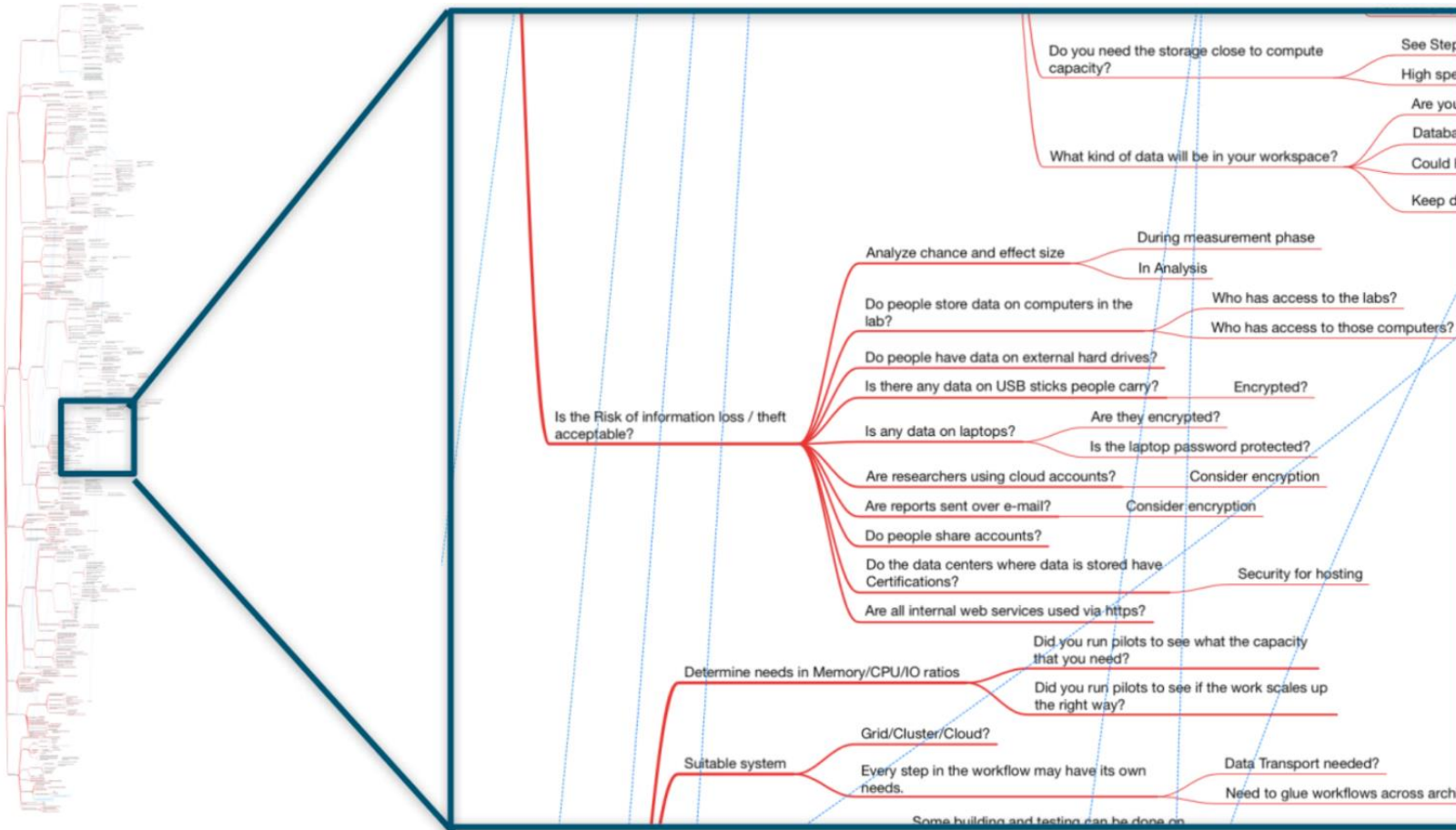


Challenge #1

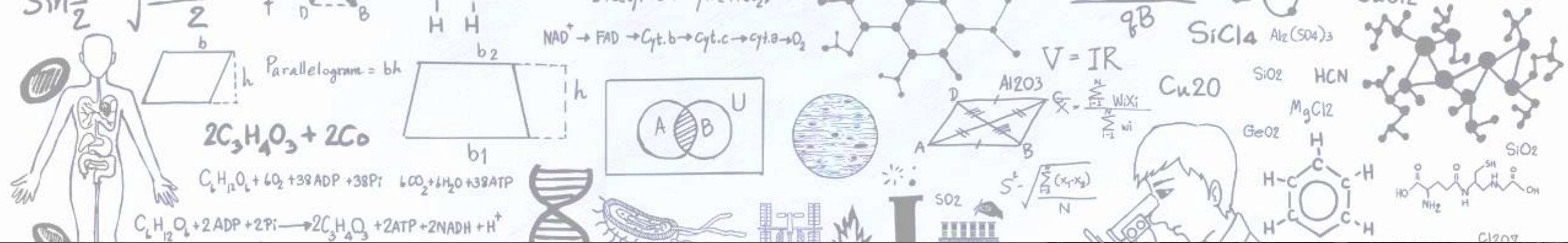
How do you know your DMP is good?



Challenge #2 Making a good DMP encompasses a vast knowledge



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

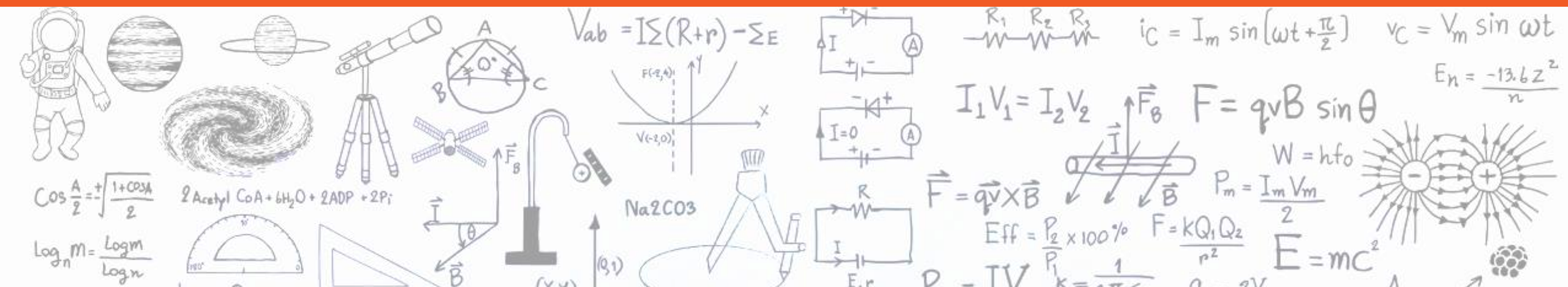


Features

- Guidance
- Learning
- Effectiveness



<https://www.wikiwand.com>



Guidance: Pilot's Check List



Operations Checklist	
Parking Brake	Set
Fuel Flow	Cutoff
Battery Switch	On
Hydraulic Pump	ON
Landing Gear	On
Flaps	Check
Spoiler	Up
Fuel Amount	Retracted
De-Ice	Check
Passenger Sign	Off
Check Weather	Off
Flight Services	
Transponder	Standby
Anti Collision Lgths	On
Engine Start Switches	Check
Thrust Reverser Switches	On

Guidance: Smart Questionnaire



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

Explanation

Project phases
References

Choice of Option

FAIR metrics

1.a.4.b.1.a.1 What repository will this data be stored in?



Domain repositories often have the best functionality to make the data findable and reusable: even though it may look like a database that could be reused in a completely different field would be better findable in a generic repository, the limited availability of domain-specific metadata make that less valuable.

Many repositories are listed in <https://fairsharing.org/>

If a repository offers to give your data set a DOI or alternative persistent identifier it is a good idea to use that option.

☒ Desirable: *Before Finishing the Project*

 External links: [FAIRSharing](#), [Registry of Research data Repositories](#)

☐ a. A domain-specific repository 


Findability

☒ b. Our national repository

Findability

☐ c. Our institutional repository

Findability

☐ d. A special-purpose repository for the project 

Findability


 Clear answer

Disadvantage of a general purpose repository is the lack of data-specific features (e.g. 'play' instead of 'download' for an audio file) and limited findability

Recommendation



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](#)

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

Data Stewardship for Open Science:

Implementing FAIR Principles

By *Barend Mons*







Guidance: Indications and Metrics

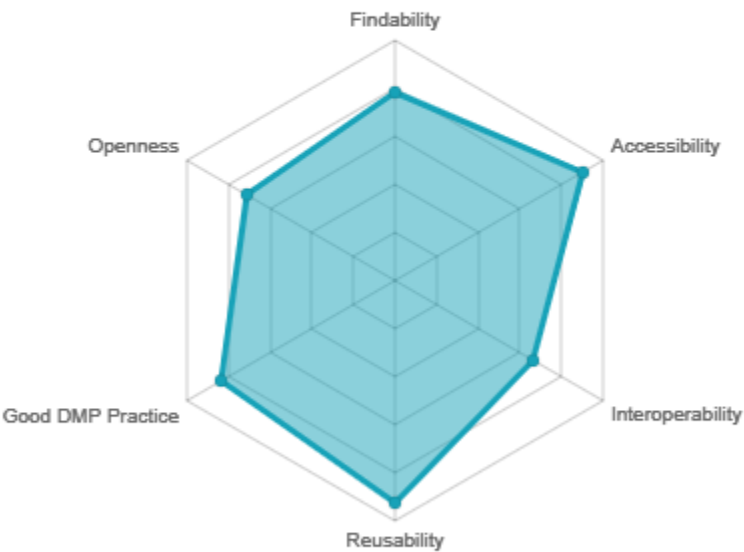


Summary Report

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

2.a.7 Funding

Add all the funding that are part of this project.

☒ Desirable: *Before Submitting the Proposal*

2.a.7.a.1 Funder

Czech

Ministerstvo Obrany České Republiky

Grantová Agentura České Republiky

České Vysoké Učení Technické v Praze

Parazitologický ústav, Akademie Věd České Republiky

Ministerstvo Zdravotnictví České Republiky

☒ a. Planned

Effectiveness: Collaboration - Sharing



Share Project

Users

Add users



Vojtěch Knaisl

Owner ▾



Jan Slifka

Editor ▾



Marek Suchánek

Viewer ▾



Visible by all other logged-in users

Other logged-in users can **view** ▾ the Project.



Public link

Cancel

Save

Effectiveness: Collaboration - Online



HVSC: Hypothetical Vascular Study by a Chemist

Questionnaire

TODOs 2

Metrics

Preview

Documents

Settings

Current Phase

Before Submitting the Proposal

Chapters

I. Administrative details1

II. Re-using data3

III. Creating and collecting data6

IV. Processing data3

V. Interpreting data1

VI. Preserving data6

VII. Giving access to data3

II. Re-using data

Before you decide to embark on any new study, it is nowadays good practice to check all options to re-use existing available data, either collected or generated by yourself in an earlier project, or data from others (Barend Mons calls this "Other PEOple's Data And Services" or OPEDAS). This can include reusable data that have been created for an earlier study, and also so-called "reference data" which is used by many projects.

It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others. And the circle is only complete if such data is actually re-used.

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](#)

☐ 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

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 x




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



3 Will you be using new types of data?

Horizon 2020 DMP

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

Answered 4 minutes ago by Marek Suchánek.

4 How will you be collecting and keeping your metadata?

Horizon 2020 DMP

Science Europe DMP

For the re-usability of your data by yourself or others at a later stage, a lot of information about the data, how it was collected and how it can be used should be stored with the data. Such data about the data is called metadata, and this set of questions are about this metadata

TODOs 3

Version history

☐ Named versions only

March 2021

▼ 21. 3.

15:51

Current

Set phase to *Before Submitting the DMP*

Marek Suchánek

15:51

First draft

Is there any pre

☒ No

Marek Suchánek

15:51

Will you be using a shared working space to work with your data?

- Rename this version
- Delete this version
- View questionnaire
- Create document
- Revert to this version

Effectiveness: DMP Export in Various Templates



New document

Name

Science Europe Example DMP

Answered (current phase): 9/45

Answered: 9/58

Template

--



Questionnaire Report 1.3.0

Exported questions and answers from a questionnaire



Horizon 2020 DMP 1.1.0

Data Management Plan according to the H2020 template



Science Europe DMP Template 1.4.0

Default DCC DMP Template recommended by Science Europe



maDMP (RDA DMP Common Standard) 1.4.0

Machine-actionable DMP according to RDA Common Standard

New document

Name

Science Europe Example DMP

Answered (current phase): 9/45

Answered: 9/58

Template



Science Europe DMP Template 1.4.0

Default DCC DMP Template recommended by Science Europe

Format



HTML Document



PDF Document



LaTeX Document



MS Word Document



OpenDocument Text



Markdown Document

Cancel

Create

Data Management Plan

Science Europe Example DMP

Contact person: **Jana Freeman** (jana.freeman@ds-wizard.org,

[0000-0000-0000-0001](https://orcid.org/0000-0000-0000-0001))

[Czech Technical University in Prague](https://www.cvut.cz/en/)

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

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

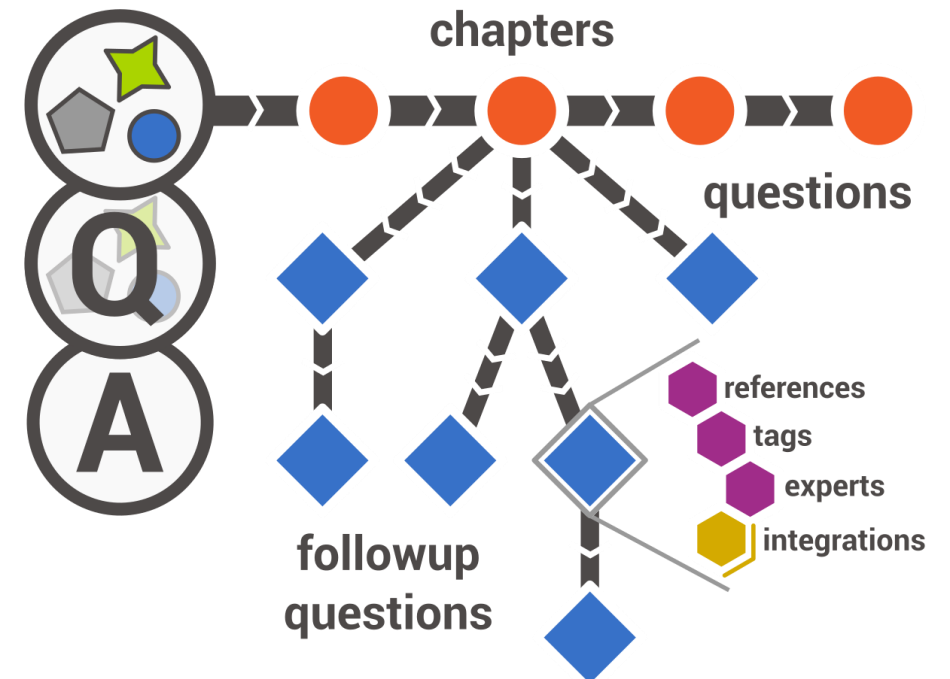
Generated on: 18 Mar 2021

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

One Size Does not Fit All

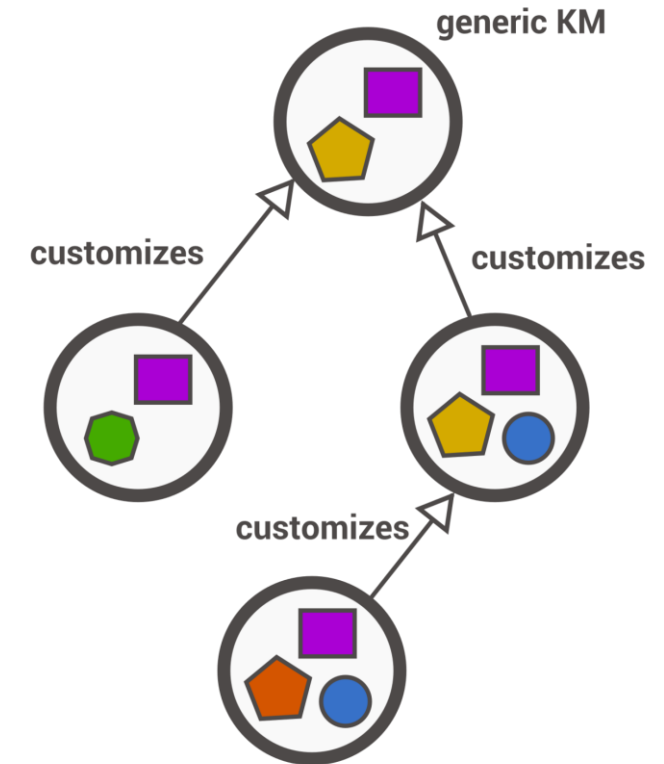
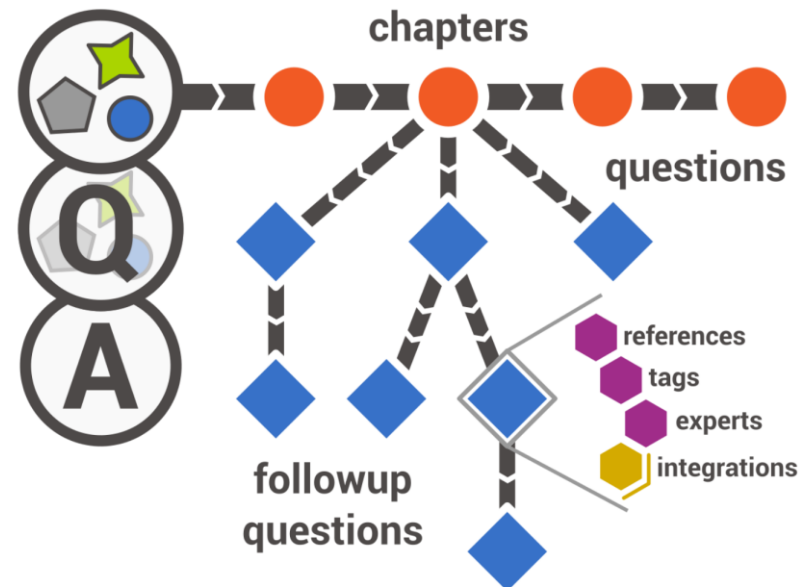
The Questionnaire is fully customisable

- 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



Integrations through REST API

2.a.7.a.1 Funder



Univerzita Karlova v Praze

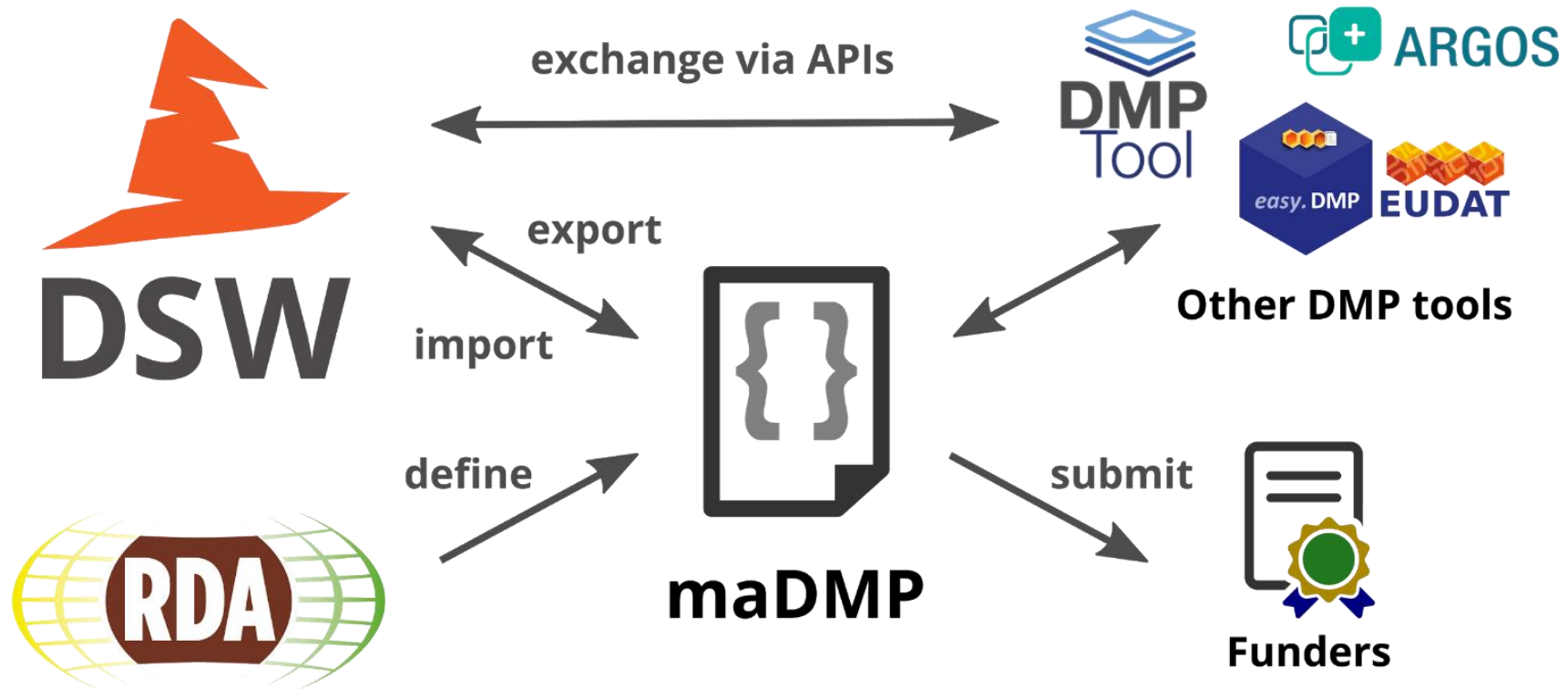


<http://dx.doi.org/10.13039/100007397>

Specify the name of the funder that you ask for funding for your project. If the funder is not present in the suggested list, please specify a complete URL to the funder web site.



Machine-Actionable DMPs



Machine-Actionable DMPs



New document

Name

My first maDMP

Answered (current phase): 9/45

Answered: 9/58

Template



maDMP (RDA DMP Common Standard) 1.4.0

Machine-actionable DMP according to RDA Common Standard

Format

☐ JSON

☒ Turtle

☐ N3

☐ RDF/XML

☐ JSON-LD

☐ N-Triples

☐ Trig

Cancel

Create

```
@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 dcs: <https://w3id.org/dcs/ns/core#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcs-lang: <https://w3id.org/dcs/id/iso639-3/> .

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

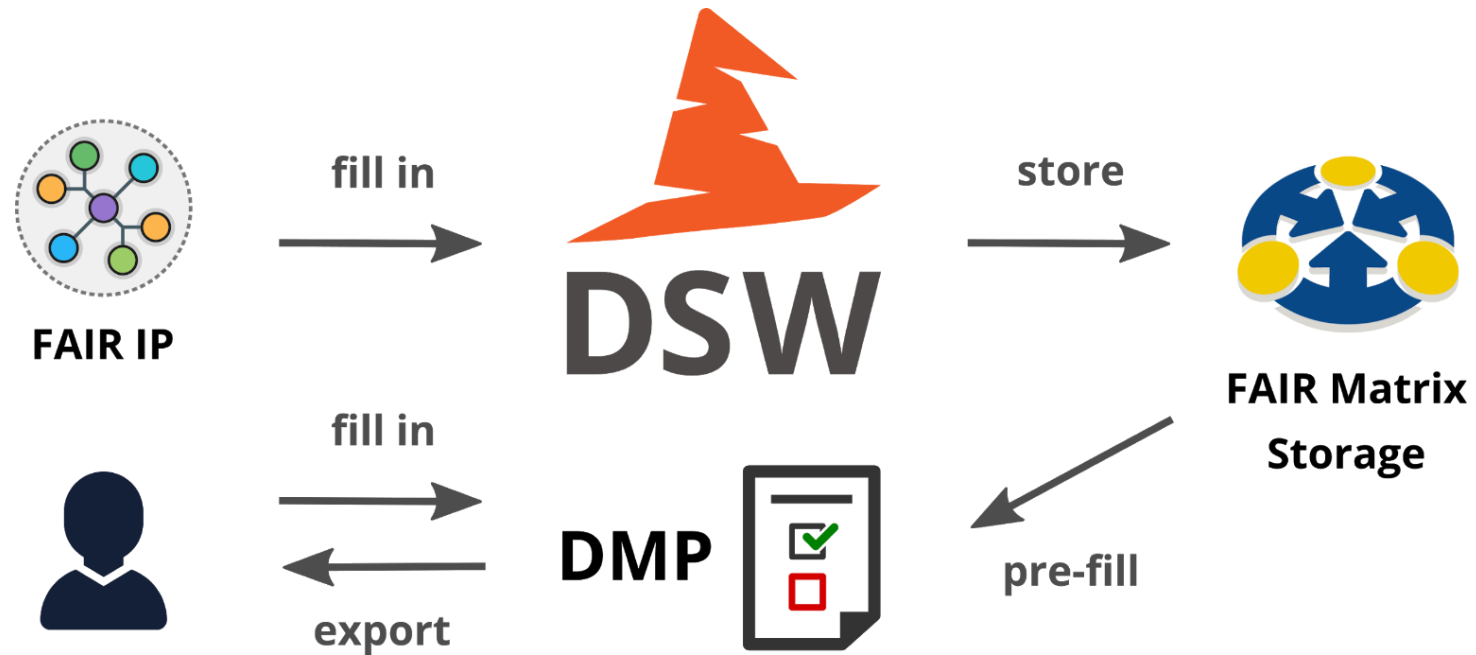
# DMP contact

# 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)
```

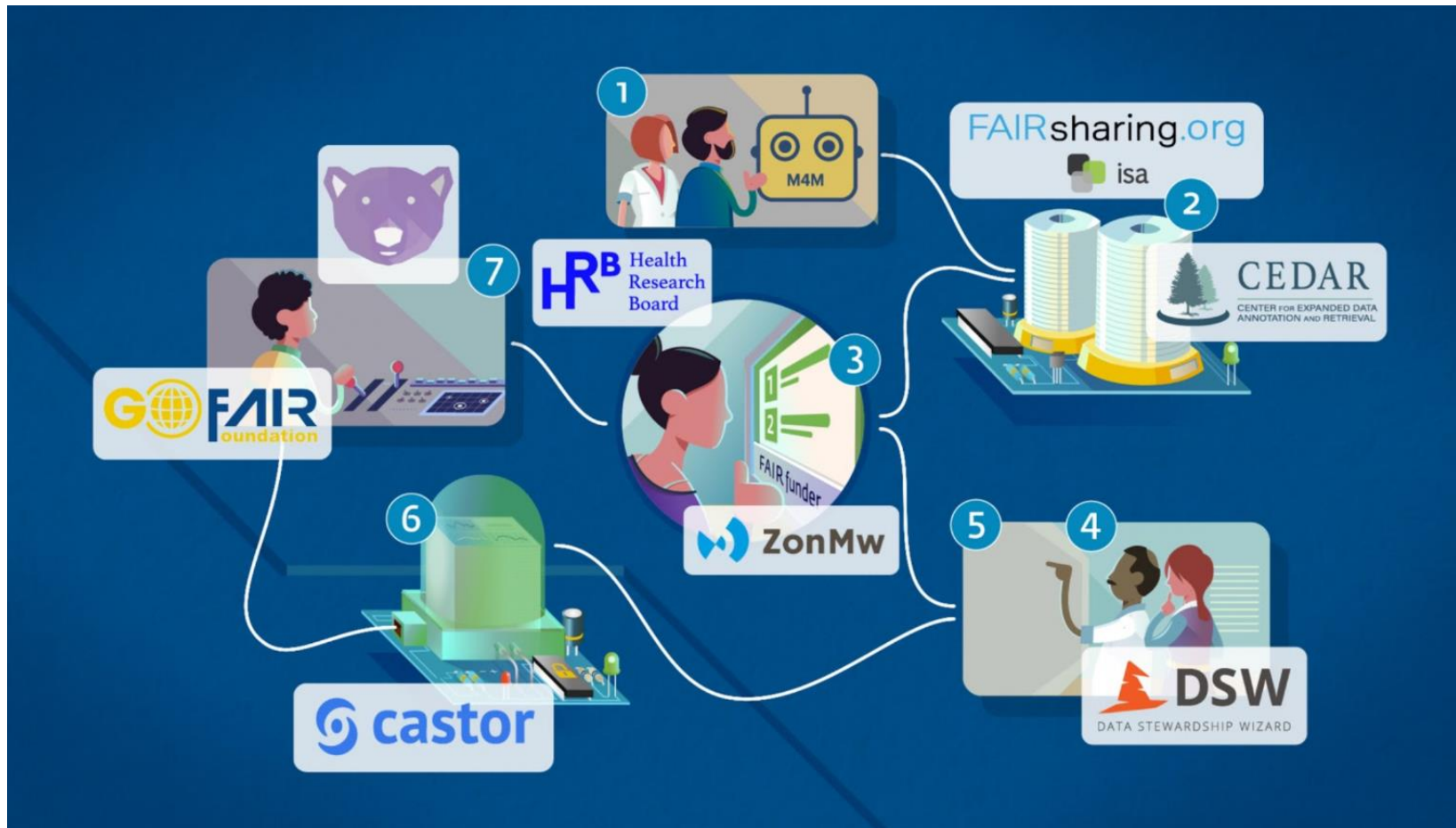

- **FAIR Implementation Profile** represents the implementation strategies of a FAIR community
- Such implementation strategies entail decisions about which FAIR-enabling resources are used, planned-to-be-used, or not-used

"Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence"

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



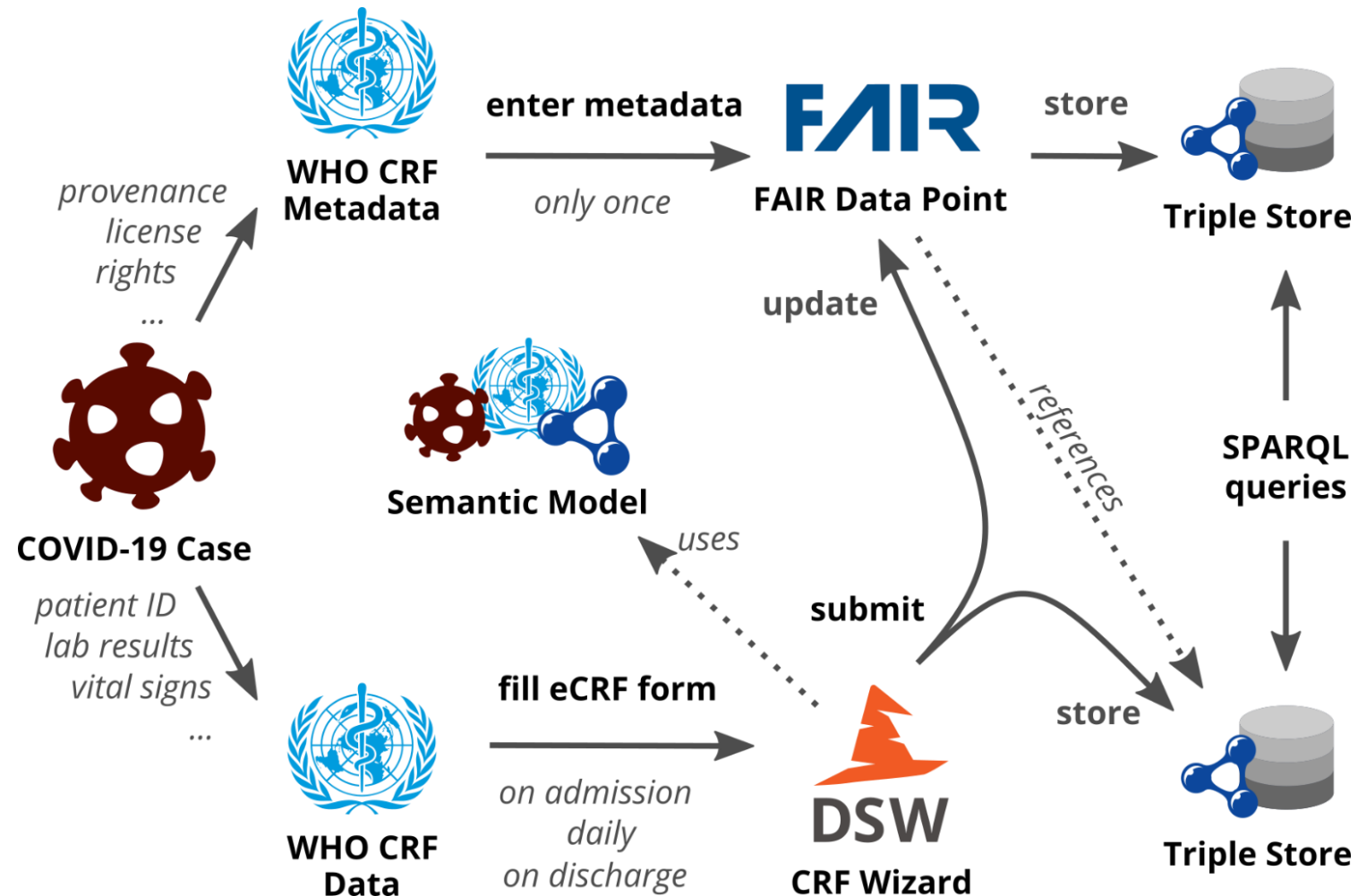
FAIR Funder Study




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

<https://arxiv.org/abs/1902.11162>

CRF Wizard



AboutGitHubContributeGive feedbackSearch...

Data life cycle +

Your role +

Your domain +

Your problem +


All tools and resources

Tools assembly +

Are you working in the Life Sciences with data? Do you feel overwhelmed when you think about Research Data Management?

The ELIXIR Research Data Management Kit (RDMKit) is an online guide containing good data management practices applicable to research projects from the beginning to the end. Developed and managed by people who work every day with life science data, RDMKit has guidelines, information and pointers, organised in many different ways to help you with problems throughout the data's life cycle.

Find resources by clicking on the RDM data lifecycle




Or find resources by:

Your role +

Your domain +

Your problem +

Tools assembly +



<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.



How to Start



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

Demo	Researchers	Self-Managed	DSW Cloud
For exploring the DSW features	For individual researchers	For organizations	For organizations
<ul style="list-style-type: none">• Easy to sign up and use• A shared instance with other users• Not for serious usage	<ul style="list-style-type: none">• Easy to sign up and use• Ready to use Knowledge Models• Privacy and stability	<ul style="list-style-type: none">• All the DSW features available• Your own instance• You need to host and run the instance by yourself	<p>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. *</p>
<ul style="list-style-type: none">✓ Questionnaires & DMPs✓ Knowledge Models Management	<ul style="list-style-type: none">✓ Questionnaires & DMPs	<ul style="list-style-type: none">✓ Questionnaires & DMPs✓ Knowledge Models Management✓ User Management & Organization Settings✓ Features configuration & visual style customizations	<ul style="list-style-type: none">✓ Questionnaires & DMPs✓ Knowledge Models Management✓ User Management & Organization Settings✓ Features configuration & visual style customizations
Try the demo	Start planning	Contact us	Contact us

Free for research institutions (ELIXIR funded)

- 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



Pinar Alper
ELIXIR Luxembourg



Korbinian Bösl
ELIXIR Norway, Centre for Digital Life
Norway, Computational Biological Unit -
University of Bergen



Flora D'Anna
ELIXIR Belgium, VIB-UGent Center for Plant
Systems Biology



Kristina Hettne
Centre for Digital Scholarship, Leiden
University Libraries



Erik Hjerde
ELIXIR Norway, The Arctic University of
North Norway



Niclas Jareborg
ELIXIR Sweden, National Bioinformatics
Infrastructure Sweden



Yvonne Kallberg
ELIXIR Sweden, National Bioinformatics
Infrastructure Sweden



Brane Leskošek
ELIXIR Slovenia, University of Ljubljana,
Faculty of Medicine, IBMI



Barbara Magagna
Environment Agency Austria



Tomasz Miksa
TU Wien & SBA Research



Barend Mons
GO FAIR



João Moreira
University of Twente



Valentina Pasquale
Istituto Italiano di Tecnologia



Erik Schultes
GO FAIR

- 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



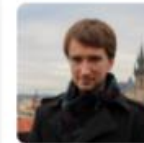
Rob Hooft
DS Knowledge Expert



Vojtěch Knaisl
Backend Developer



Jan Slifka
Frontend Developer



Marek Suchánek
Leonardo da Vinci



Jana Freeman
Community Manager



Tereza Macháčová
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).



Thank you!

info@ds-wizard.org

