



DSW Workshop for Finnish Data Support Personnel



**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**

Tereza Macháčová
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



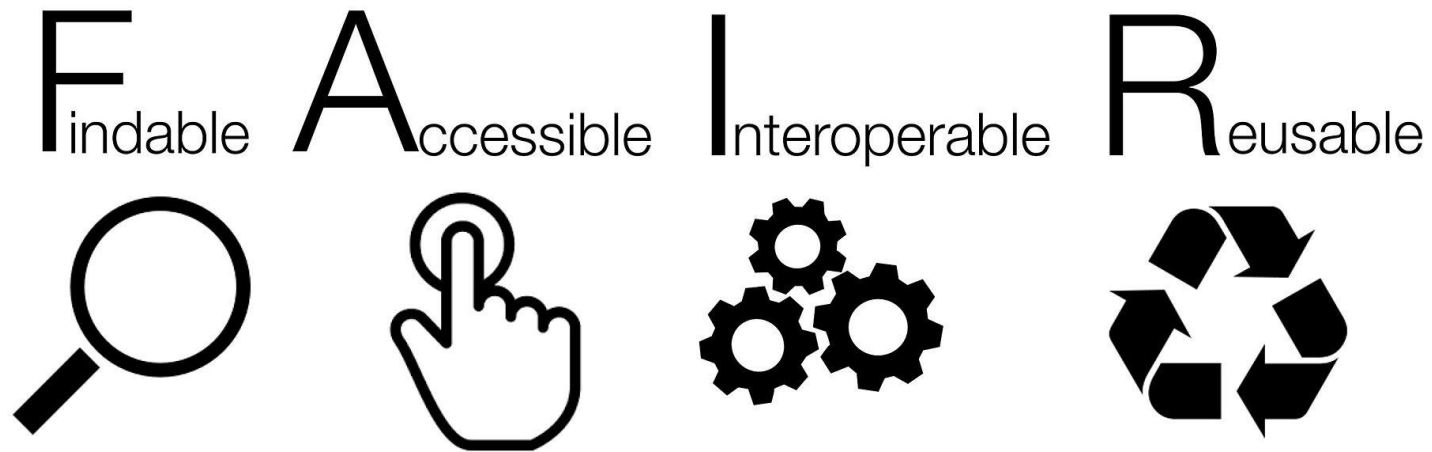
Operations Checklist	
Parking Brake	Set
Fuel Flow	Cutoff
Battery Switch	On
Hydraulic Pump ON	On
Landing Gear	On
Flaps	Check
Spoiler	Up
Fuel Amount	Retracted
De-Ice	Check
Passenger Sign	Off
Check Weather	Off
	Flight Services
Transponder	Standby
Anti Collision Lights	On
Engine Start Switches	Check
Throttle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

What is Data Stewardship Wizard?



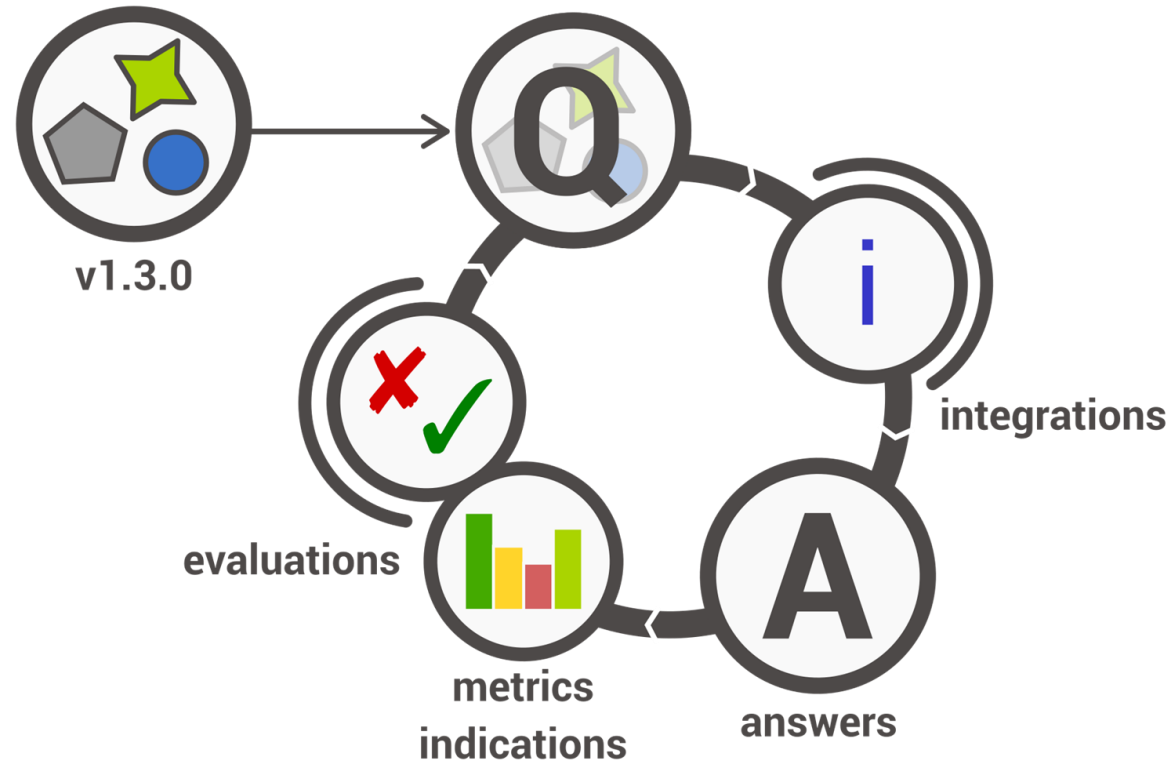
- Open-source tool developed as a part of ELIXIR CZ research infrastructure
- Tool suitable for everyone (from beginners to data stewards)
- Serves as a check-list before starting the project
- Supports data management planning with respect to (current) best practices

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

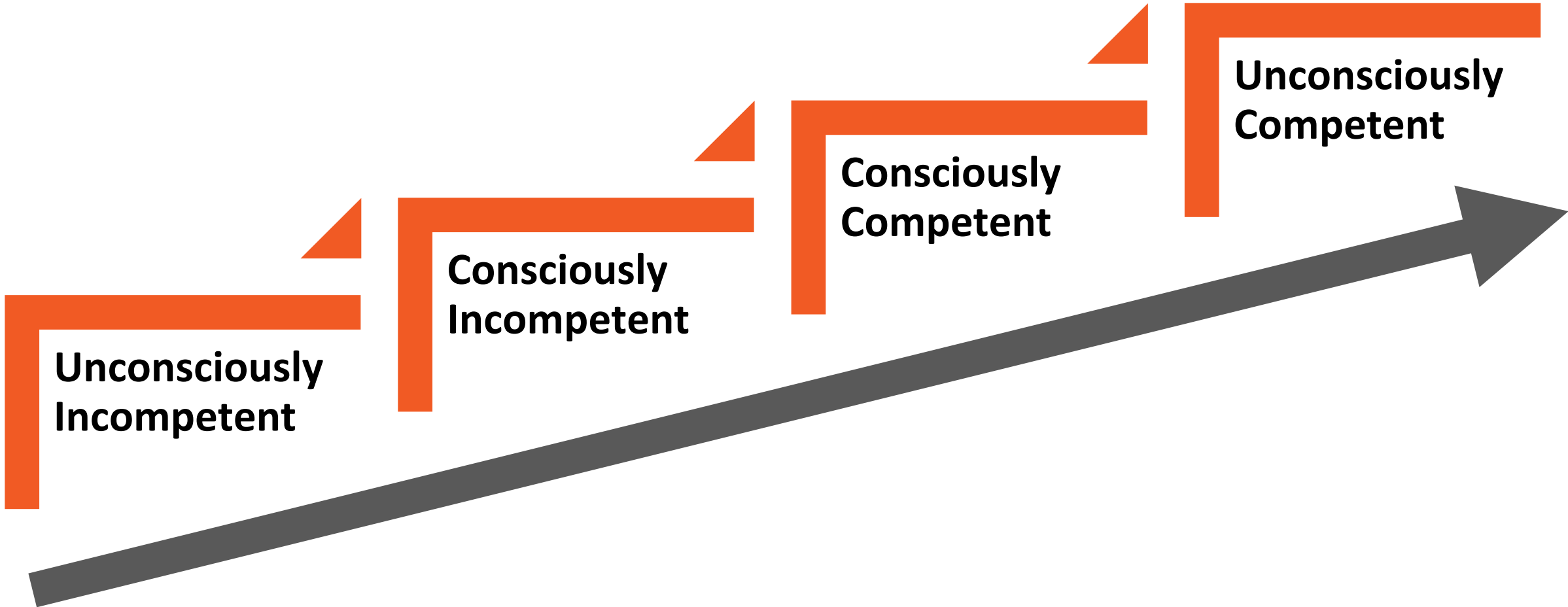


- Minimum of writing = DMP is not an essay, as little writing as possible
- Guidance = DSW guides users through the smart Questionnaire
- Flexibility = easy to edit the content and integrate with other services
- Openness = anybody can use it and create own content
- User-oriented = DSW development is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

Returning to DMP



Educational Aspect of DSW



**Unconsciously
Incompetent**

**Consciously
Incompetent**

**Consciously
Competent**

**Unconsciously
Competent**

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 ⋮

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 ⋮

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](#)

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

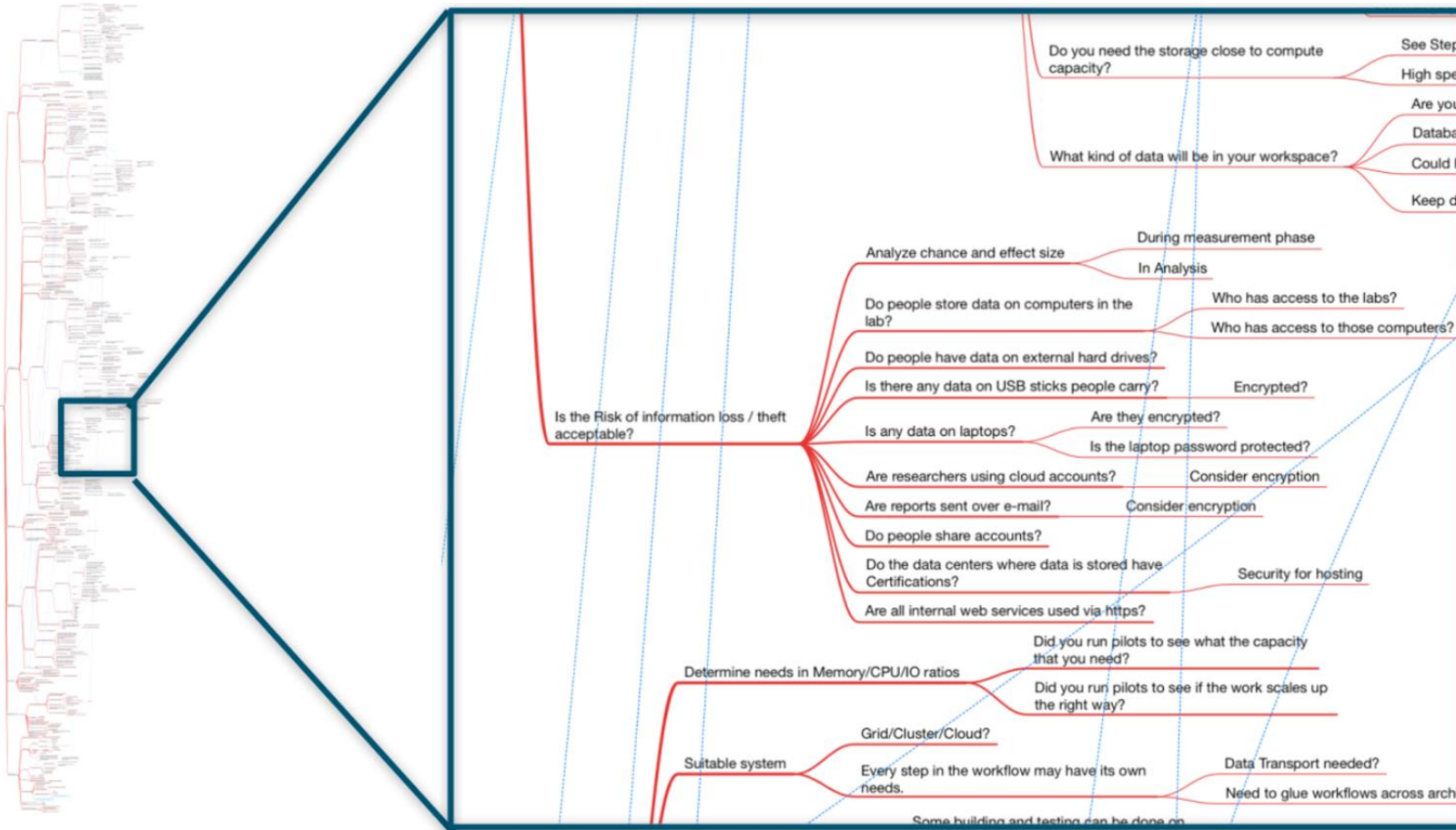
b. Yes

[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)

Knowledge Model vs Template



Create Project

Name

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.

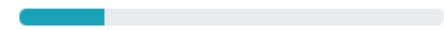
Cancel

Save

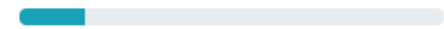
New document

Name

Answered (current phase): 9/45



Answered: 9/58



Template

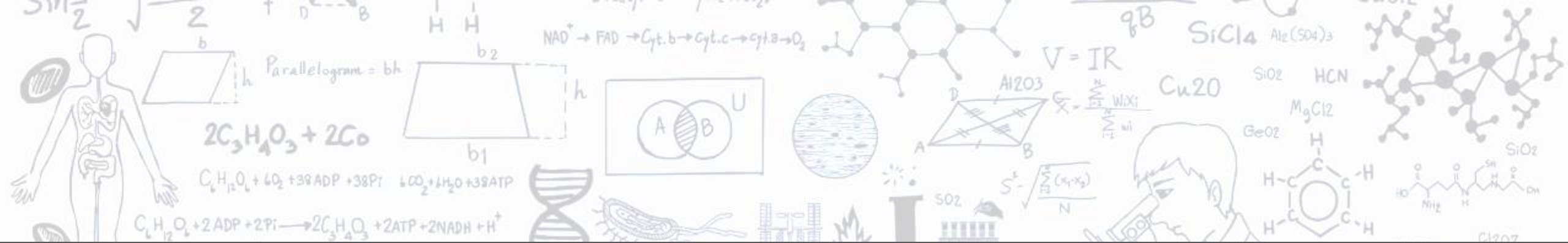
S **Science Europe DMP Template** 1.4.0
Default DCC DMP Template recommended by Science Europe

Format

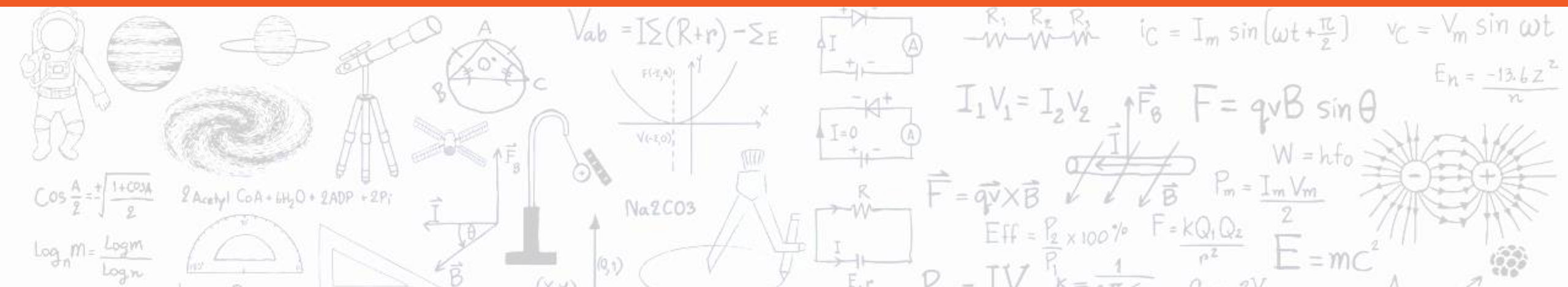
- HTML Document
- PDF Document
- LaTeX Document
- MS Word Document
- OpenDocument Text
- Markdown Document

Cancel

Create



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

Knowledge Model

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

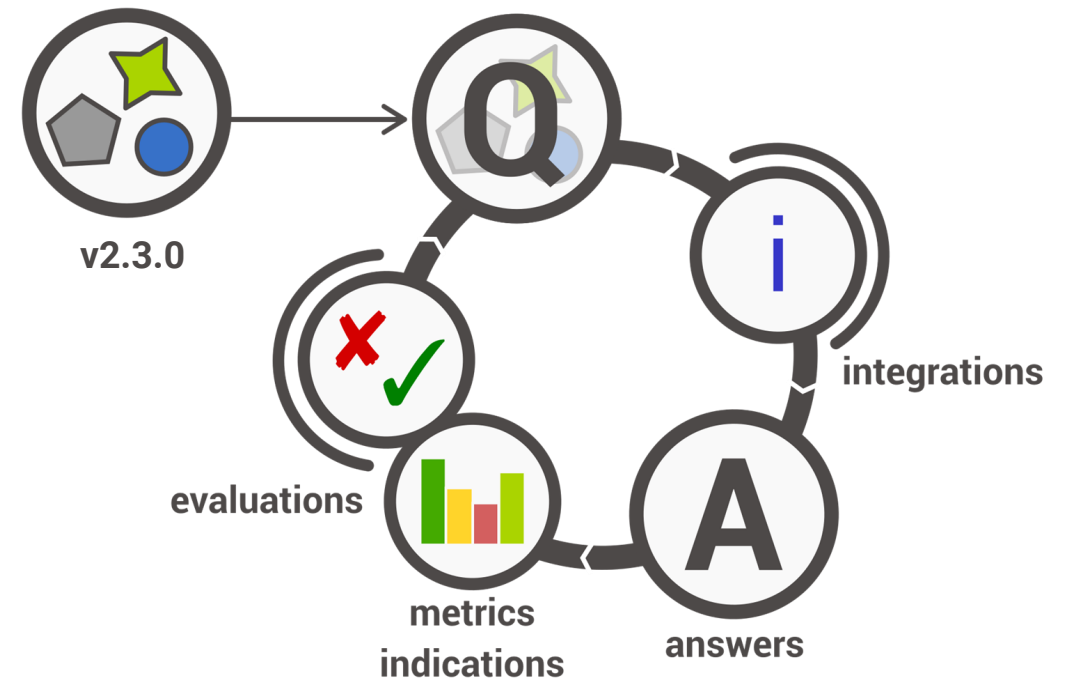
Tags

- 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

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



TODO, Feedback

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](#)

Choice of Option

a. A domain-specific repository

Findability

Follow up questions

b. Our national repository

Findability

c. Our institutional repository

Findability

d. A special-purpose repository for the project

Findability

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

FAIR metrics



References

Recommendation

Answers through Integration

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

Linking the Answers to URI

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.









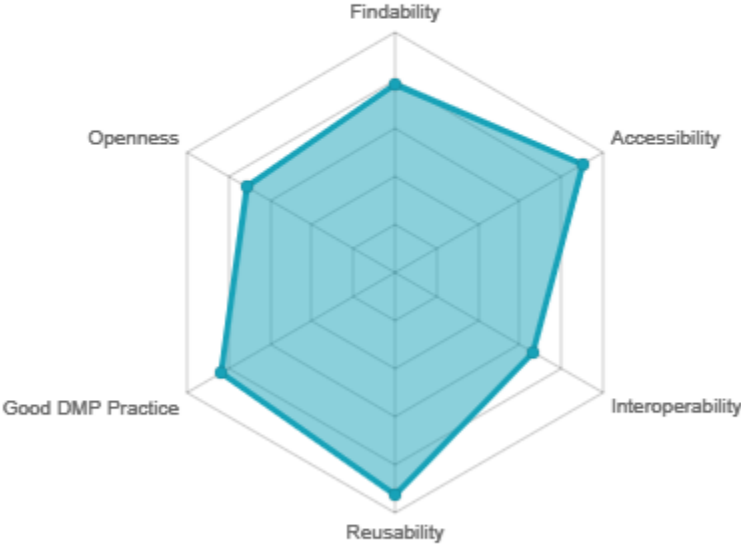
Indications and Metrics



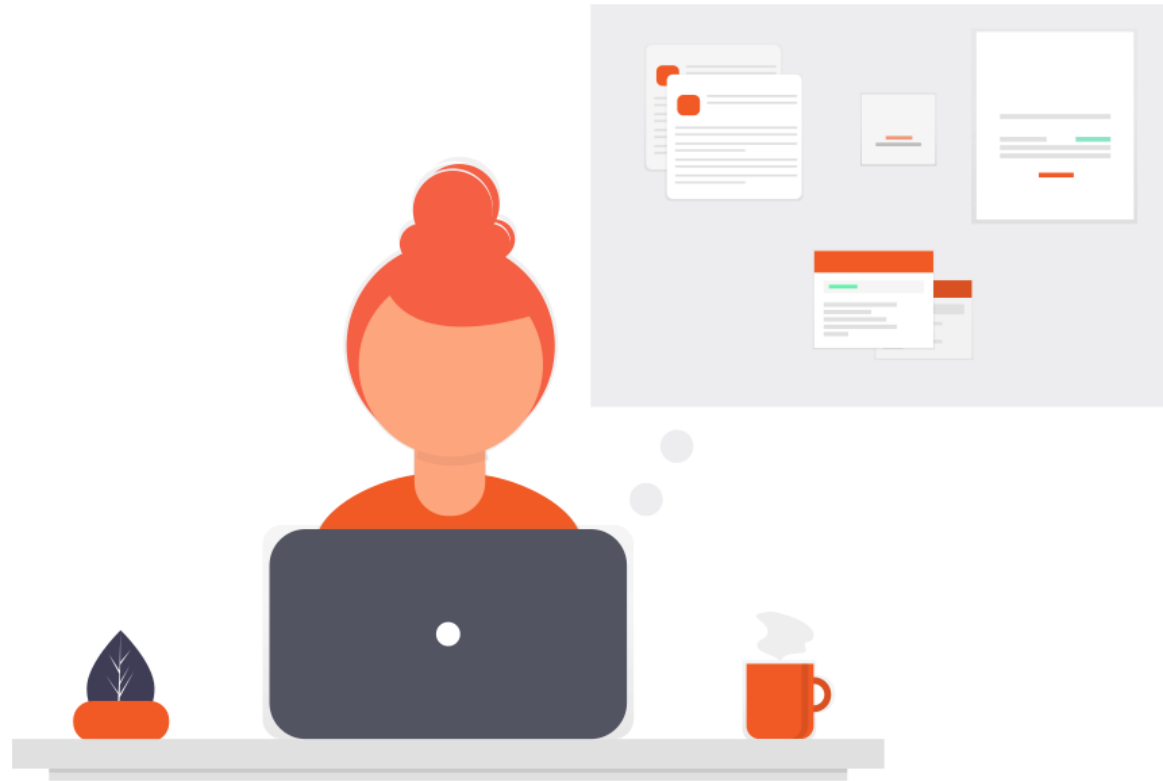
Summary Report

Answered (current phase): 100/100 
Answered: 293/305 

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



Demonstration



Working Together with your Colleagues



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

Working Together with your Colleagues



HVSC: Hypothetical Vascular Study by a Chemist



Share

Questionnaire **TODOs 2** Metrics Preview Documents Settings

Current Phase

Before Submitting the Proposal

Chapters

- I. Administrative details 1
- II. Re-using data 3**
- III. Creating and collecting data 6
- IV. Processing data 3
- V. Interpreting data 1
- VI. Preserving data 6
- VII. Giving access to data 3

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: [atg](#)

a. No

b. Yes

Clear answer

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

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: [ezf](#)

a. No

b. Yes

Version History



Questionnaire Metrics Preview Documents Settings

View TODOs **Version history**

Current Phase

Before Submitting the Proposal

Chapters

- I. Administrative information** 1
- II. Re-using data 1
- III. Creating and collecting data 6
- IV. Processing data 4
- V. Interpreting data 2
- VI. Preserving data 5
- VII. Giving access to data 2

I. Administrative information

1 Contributors + !

Horizon 2020 DMP Science Europe DMP maDMP

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: Before Submitting the DMP

1.a.1 Name + !

Horizon 2020 DMP Science Europe DMP maDMP

Desirable: Before Submitting the DMP

Marek

Answered in less than 5 seconds by Marek Suchánek.

1.a.2 E-mail address + !

Horizon 2020 DMP Science Europe DMP maDMP

Named versions only

April 2021

5. 4.

17:58 ⋮

Current

Name

Marek

Marek Suchánek

3. 4.

22:00 ⋮

First version

Will this work for specialists?

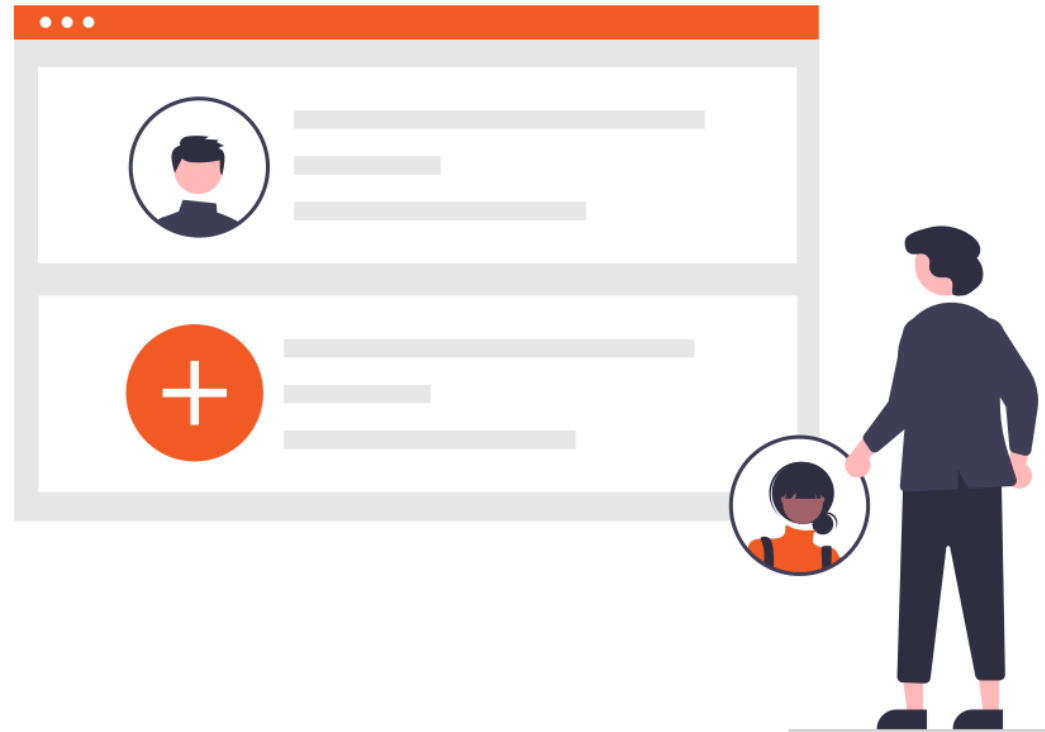
No

Tereza Mach

22:00

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

Demonstration



DMP Export in Various Templates



New document

Name

My first DMP|

Answered (current phase): 100/100

Answered: 293/305

Template

H **Horizon 2020 DMP** 0.1.0
Data Management Plan according to the H2020 template

Format

HTML

PDF Document

LaTeX Document

MS Word Document

OpenDocument Text

Markdown Document

Cancel

Create

Document Preview and Download



- DS Wizard
- Users
- Knowledge Model Editor
- Knowledge Models
- Projects
- Documents
- Templates
- Storage Costs Evaluator

Test 832

Share

- Questionnaire
- Metrics
- Preview
- Documents
- Settings

Data Management Plan

Test 832

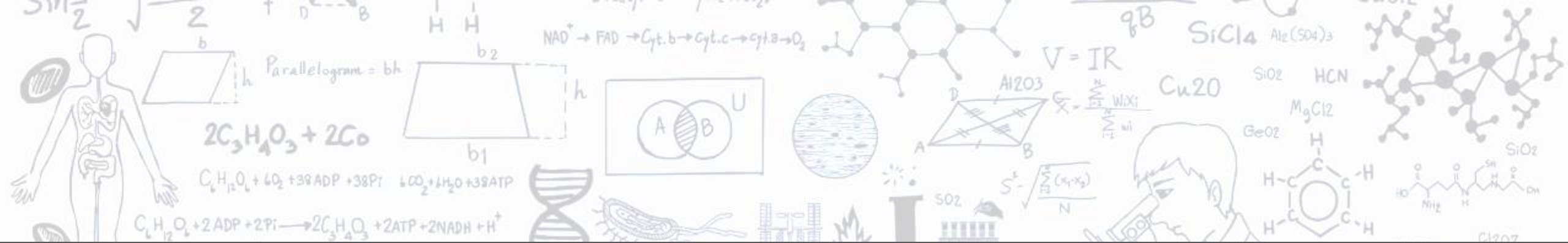
Following the Horizon 2020 DMP Template v2.0

Demonstration

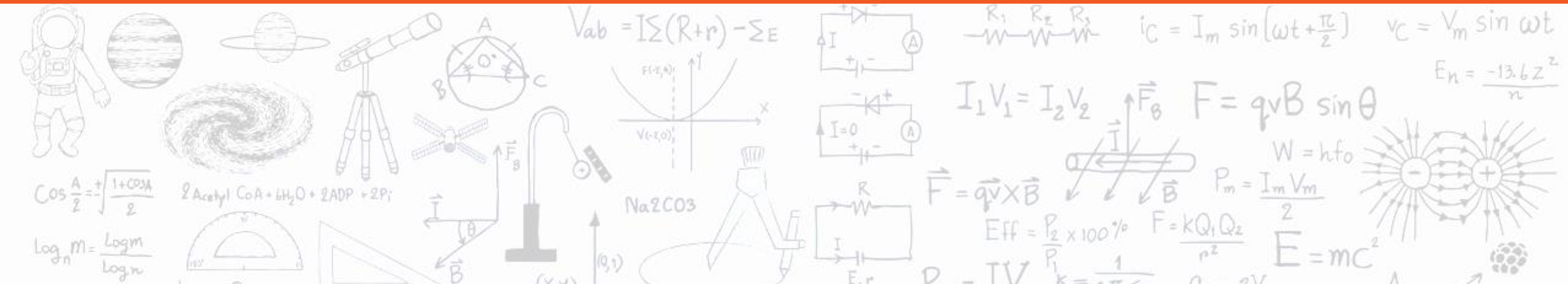


- Let's have a break ~15 minutes
- Grab a snack or coffee
- Try out what was shown
- Create a project (KM: ...)
 - Fill in basic information about the project
 - Generate H2020 document
 - Generate maDMP (JSON or Turtle RDF)



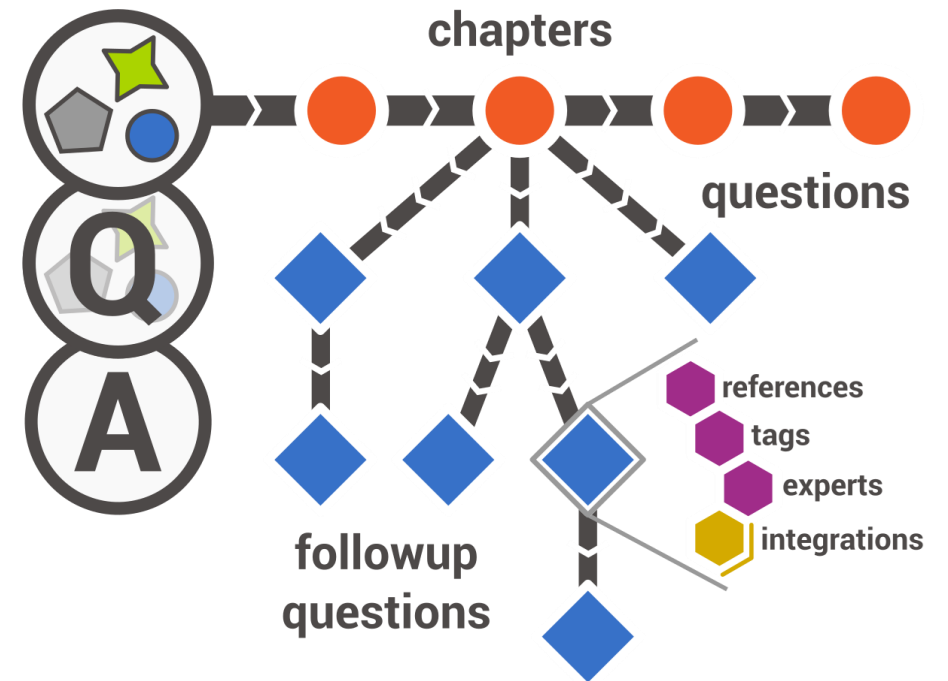


How to use (Data Steward)



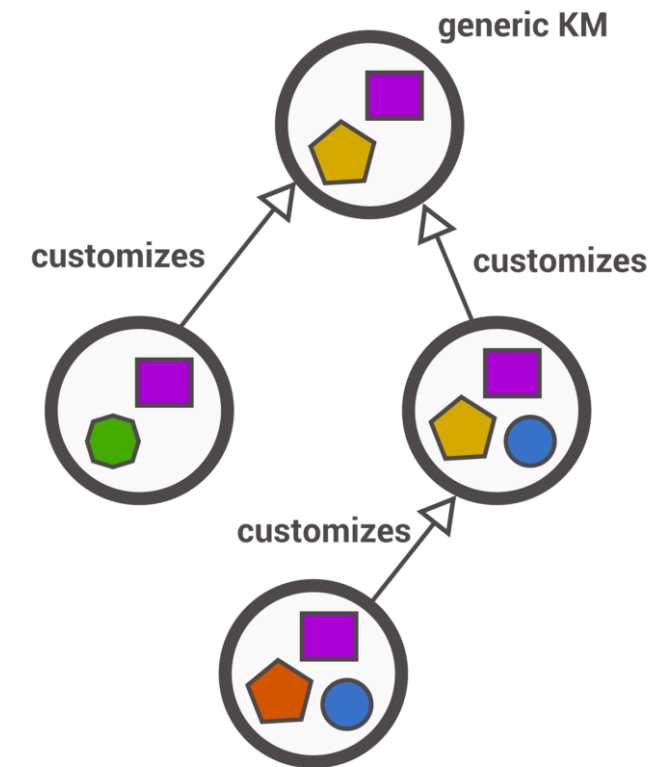
What is „Knowledge Model“

- 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

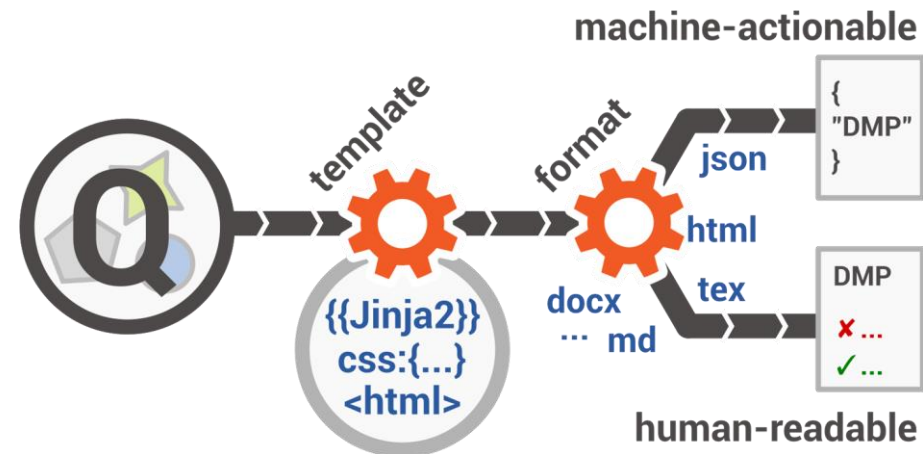


Demonstration



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





Thank you!

info@ds-wizard.org

