

Jak na tvorbu Data management plánu



Tereza Macháčová
tereza.machacova@ds-wizard.org

Jan Slifka
jan.slifka@ds-wizard.org

- Data Management Planning (DMP)
- Formát dat, jejich množství, účel vzniku, doba a místo uložení...
- Použité licence, sdílení dat, platnost přístupu...
- Plány jsou důležité pro (výzkumné) projekty
- Vyžadován funderem
- Pomáhá proti nejasnostem
- Pomáhá proti ztrátě dat

- Data Stewardship je komplikovaná problematika pokrývající různá témata
 - úložiště, licence, sdílení, FAIR
- Plán nemá být jednorázovým statickým dokumentem
- Plán má být aktuální a pravdivý



... a mnoho dalších

Cíl výzkumníků

- Vytvořit DMP pro projekt
- Získat potřebné znalosti pro DMP
- Spolupracovat s kolegy
- Být produktivní



Předletová příprava



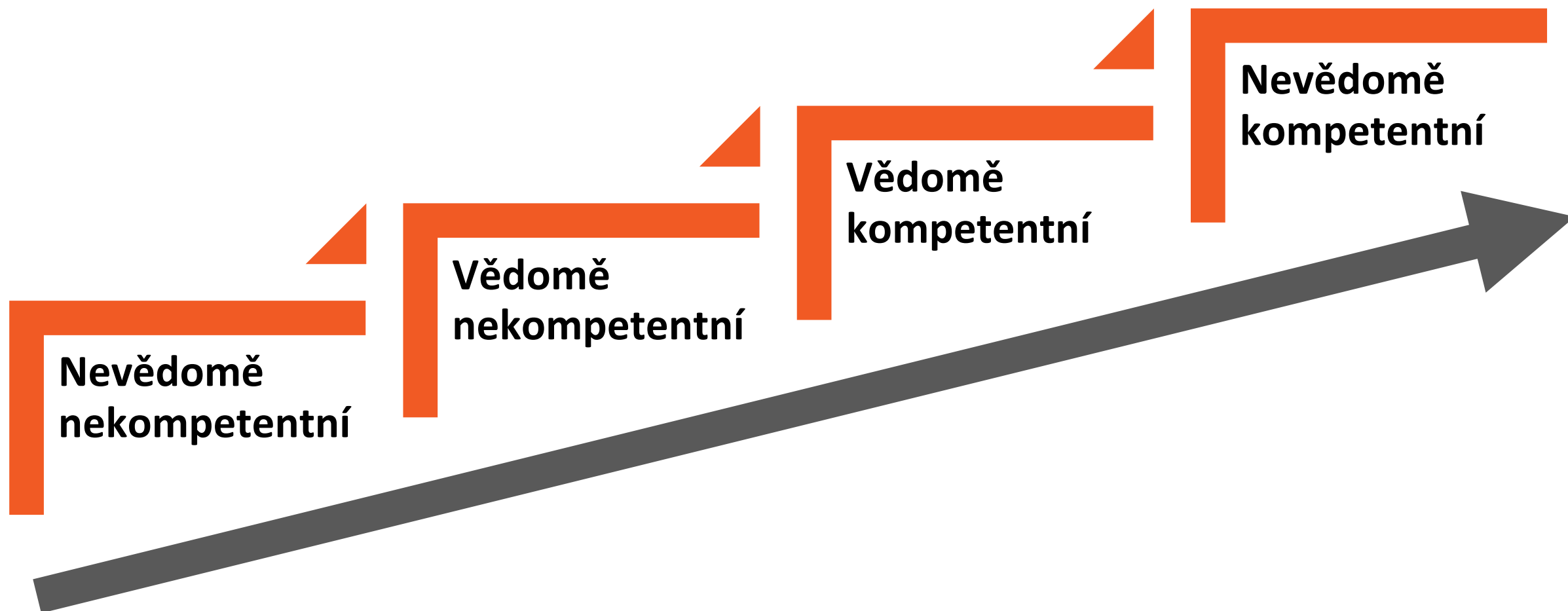
Operations Checklist	
Parking Brake	Set
Fuel Flow	Cutoff
Battery Switch	On
Hydraulic Pump	ON
Landing Gear	On
Flaps	Check
Spoiler	Up
Fuel Amount	Retracted
De-Ice	Check
Passenger Sign	Off
Check Weather	Off
Flight Services	
Transponder	Standby
Anti Collision Ligths	On
Engine Start Switches	Check
Thrust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

Co je Data Stewardship Wizard?



- Open-source nástroj vyvíjený v rámci ELIXIR
- Nástroj vhodný pro každého (od začátečníků po data stewardy)
- Slouží jako check-list před startem projektu
- Podporuje správu dat s ohledem na současné best practice (FAIR, machine-actionability)

- Minimum psaní = plán není eseje, psaní jen tam, kde je to nezbytné
- Vedení = DSW vede uživatele skrz smart Questionnaire
- Flexibilita = lze upravovat obsah a integrovat s jinými službami
- Otevřenost = kdokoli je může využít a vytvářet vlastní obsah
- Orientace na uživatele = vývoj se výrazně řídí na základě zpětné vazby



Vytváření projektu (dotazníku)

Create Project

Name

Knowledge Model

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

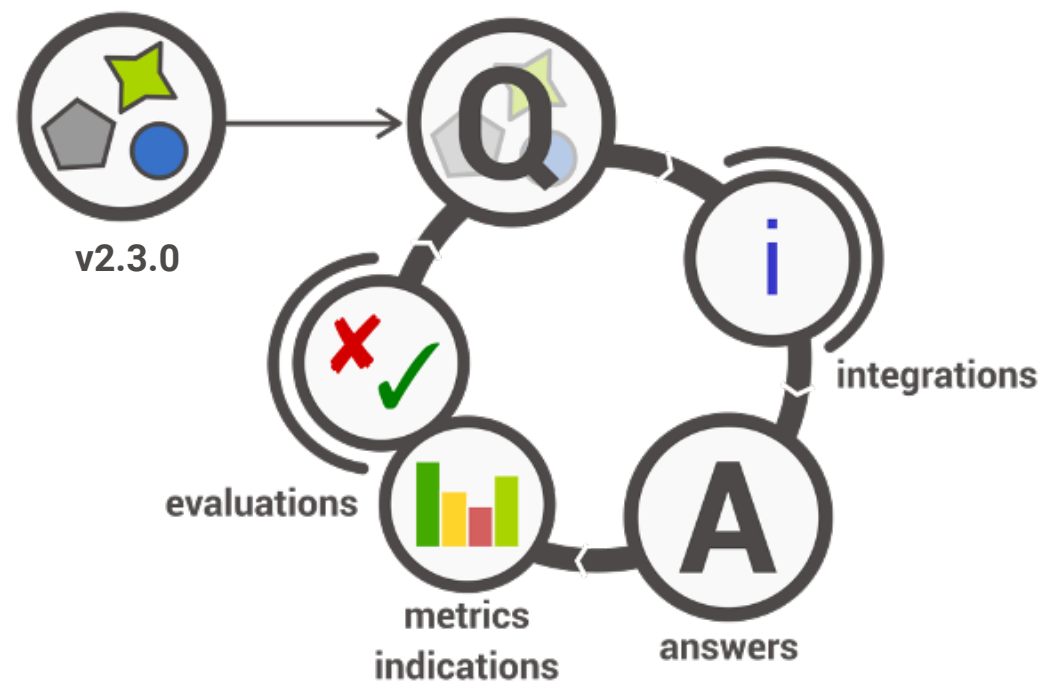
Tags

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



DMP jako chytrý dotazník

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



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*

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](#)


Typy otázek

- Výběr
- Multichoice
- List
- Hodnota
- Integrace

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

List a hodnota



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*

+ Add

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*

1.a.2 E-mail address



Horizon 2020 DMP Science Europe DMP maDMP

- Součástí odkaz - kvalitní metadata a usnadnění práce

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

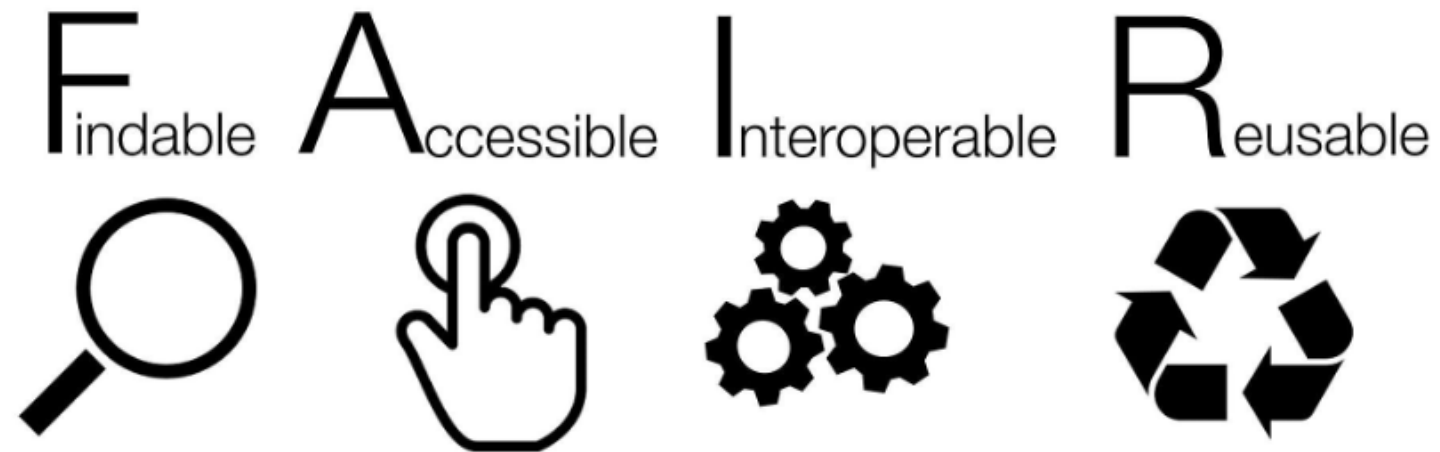
Parazitologický ústav, Akademie Věd České Republiky

Ministerstvo Zdravotnictví České Republiky

a. Planned



- Co je FAIR?
- Principy aplikované na data



- „Informace strukturovaná konzistentním způsobem tak, že stroje a počítače mohou být naprogramované pro snadné zpracování a vyhodnocování této struktury a jejího obsahu“
- Usnadnění práce
- Kvalitnější data a metadata

Základem je kvalitní „guidance“

Krátké vysvětlení

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

Výběr odpovědi

FAIR metriky



Reference

Doporučení







Summary Report

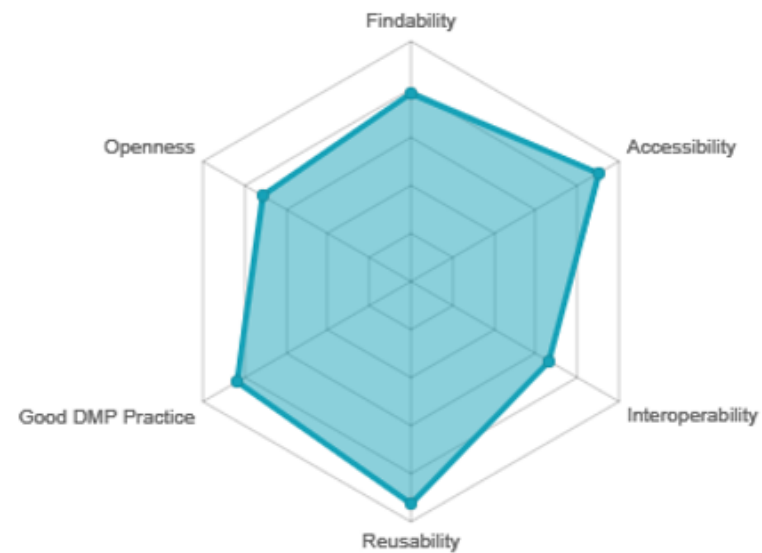
Answered (current phase): 100/100



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	



Spolupráce při plánování



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 the Project.

Public link

Cancel

Save

Spolupráce při plánování



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)?

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

V. Interpreting data

The interpretation of the data consists of the last steps of processing (often with manual interventions), visualisation, and data integration. In this chapter many questions about data interoperability will come up.

1 List the data formats you will be using and their structure TODO X !

Give each type of data a name that you recognise.

If you have data in many different structures, integrating the data may be more challenging.

Desirable: *Before Submitting the Proposal*

+ Add

2 Will you be doing integration of different data types? + !

If you are getting different types of data from different sources and want to use them together it is likely that you will need to match items and glue everything together. This can be done with traditional table database technology, but it is also possible to use Linked Data and RDF.

This is an advanced subject that you may want to skip if this is not an issue for you. On the other hand, if this is your expertise we would like your help in improving the questions in this section.

Processing data

- Will you be using a shared working space to work with your data?
- How will you validate the integrity of the results?

Interpreting data

- List the data formats you will be using and their structure

Historie verzí



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

Komentáře a editorské poznámky

My example

Questionnaire Metrics Preview Documents Settings

View Comments 2 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 3
- V. Interpreting data 2
- VI. Preserving data 5
- VII. Giving access to data 2




I. Administrative information

1 Contributors  2 comments 

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*




2 Research Project(s)   

Horizon 2020 DMP Science Europe DMP maDMP


Add each of the research project(s) that are you 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.

View resolved comments

Comments 2 Editor notes

 **Marek Suchánek** 14. 10. 2021, 16:43  

Who should we add here as contributor?

 **Anonymous user** 14. 10. 2021, 16:44

Add me as well, please!!

Knowledge Model vs Document template



Create Project

Name

My first project

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

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

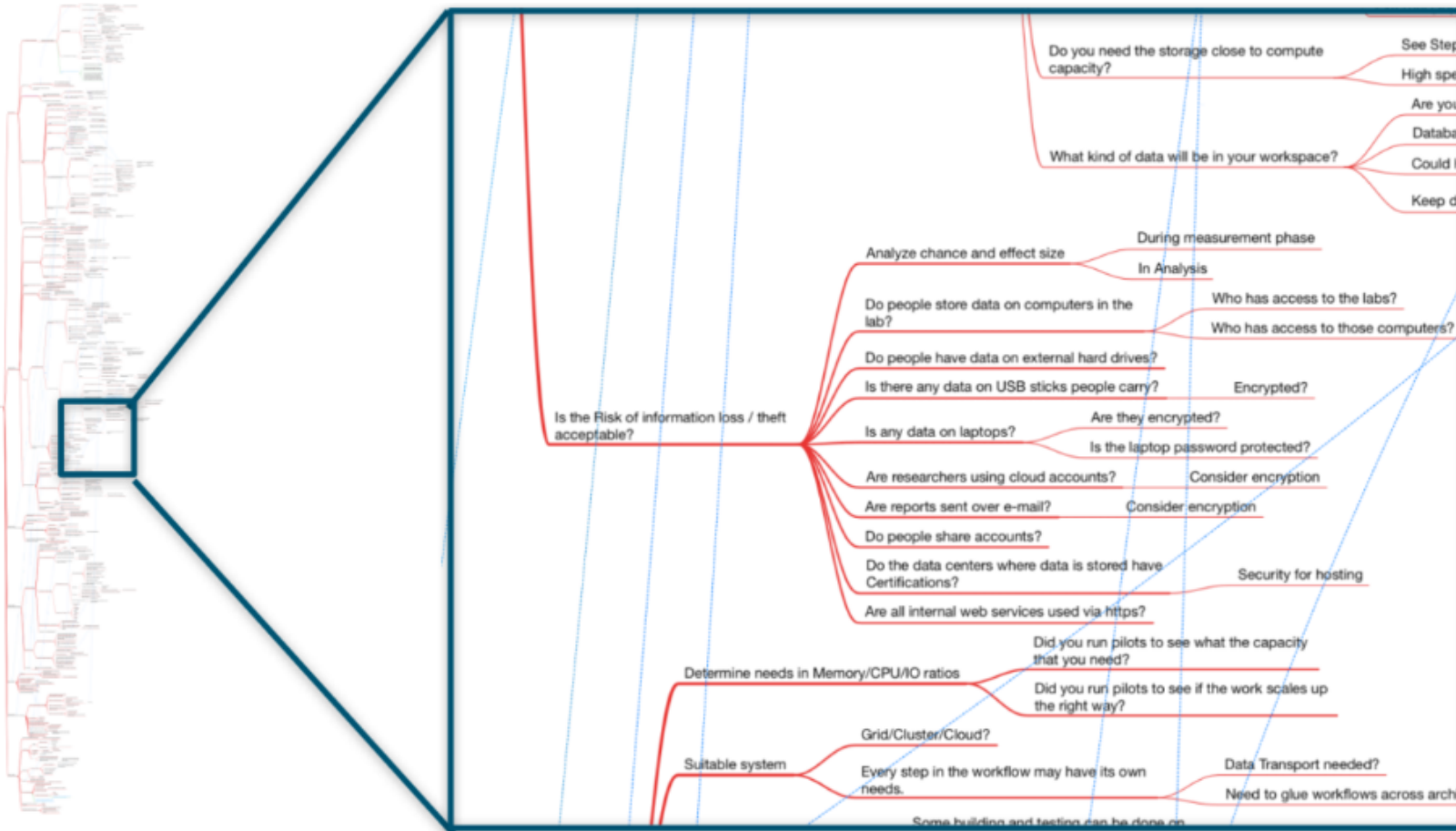
OpenDocument Text

Markdown Document

Cancel

Create

Common DSW Knowledge Model



- Komplexní rozhodovací strom
- Podle mindmapy kterou vytvořil Rob Hooft (DTL, ELIXIR NL)
- ~600 uzlů

Document template

- Horizon 2020
- Science Europe
- maDMP
- Questionnaire Report
- Horizon Europe již brzy!

Default document template

S **Science Europe DMP Template** 1.6.0 ×
Default DCC DMP Template recommended by Science Europe

H **Horizon 2020 DMP** 1.3.0
Data Management Plan according to the H2020 template

Q **Questionnaire Report** 2.1.0
Exported questions and answers from a questionnaire

m **maDMP (RDA DMP Common Standard)** 1.7.0
Machine-actionable DMP according to RDA Common Standard

S **Science Europe DMP Template** 1.6.0
Default DCC DMP Template recommended by Science Europe


Ukázkový DMP ✓

 Share

 Questionnaire

 Metrics

 Preview

 Documents

 Settings

Filter by name...


Created ▾



New document

D

DMP v LaTeX formátu

 LaTeX Document · Science Europe DMP Template

Updated less than 5 seconds ago



V

Varianta pro H2020


 MS Word Document · Horizon 2020 DMP

Updated less than a minute ago



U

Ukázkový DMP

 PDF Document · Science Europe DMP Template

Updated 9 minutes ago



Náhled dokumentu



- 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

Create Project

Name

Knowledge Model

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

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

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*

1.a.2 E-mail address + !

Horizon 2020 DMP

Science Europe DMP

maDMP



Zjistěte víc

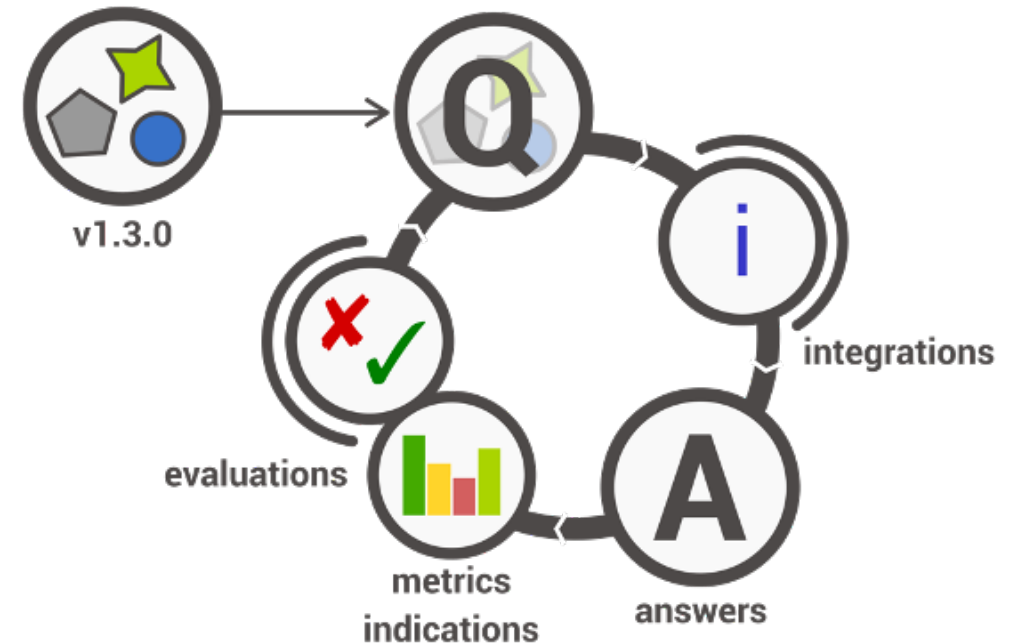


- Webová stránka (ds-wizard.org)
- GitHub (github.com/ds-wizard)
- User Guide (guide.ds-wizard.org)
- Twitter ([@dswizard_org](https://twitter.com/dswizard_org))
- Ideas (ideas.ds-wizard.org)
- Slack
- DSW Coffee



Zopakování hlavních myšlenek DSW

- Minimum psaní = typy otázek, follow-up otázky
- Vedení = odkazy, kniha, FAIR, rady
- Flexibilita = vlastní knowledge modely, šablony, integrace
- Otevřenost = open-source, zdarma
- Orientace na uživatele = ideas, DSW Coffee



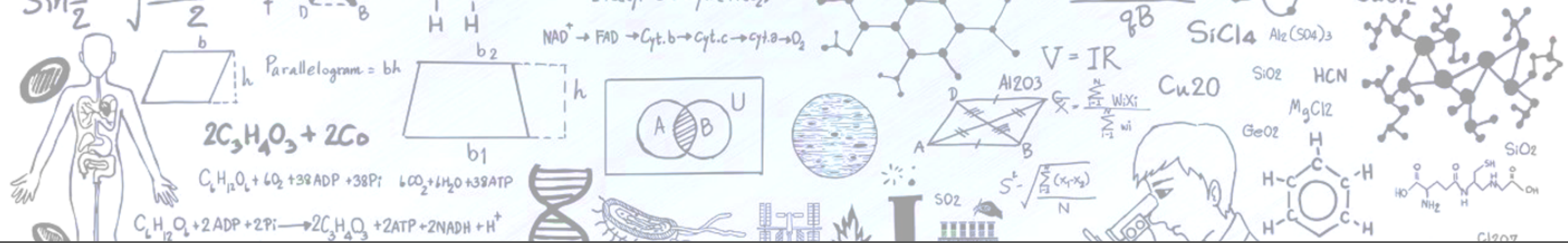
Poděkování



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

Otázky a diskuze





Děkujeme!

info@ds-wizard.org

