



Praktické využití Data Stewardship Wizard pro tvorbu DMP

14. 3. 2023

Marek Suchánek

marek.suchanek@ds-wizard.org

 0000-0001-7525-9218

- Data Stewardship Wizard
- Horizon Europe DMP šablona
- Praktická ukázka tvorby DMP
- Příklady vyplněných DMP

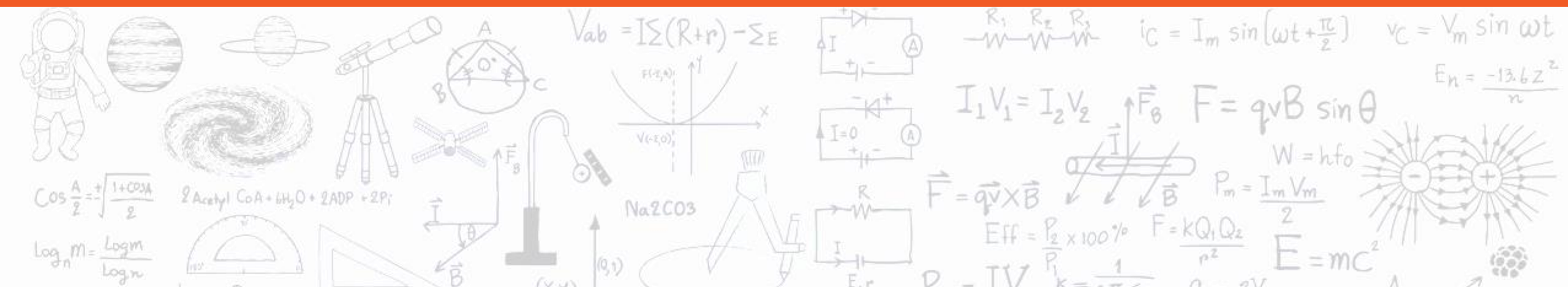




DSW

DATA STEWARDSHIP WIZARD

Data Stewardship Wizard



- Open-source vyvíjený ve spolupráci **ELIXIR CZ** a **ELIXIR NL**
- Expertní systém pro plánování správy dat a tvorbu data management plánů (DMP)
 - *“From burden to benefit”*
 - *“Plans are worthless, but planning is everything”*
- Doporučený nástroj v **Horizon Europe Program Guide**

- Minimum psaní = plán není eseje, psaní jen tam, kde je to nezbytné
- Vedení = DSW vede uživatele skrz tzv. *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

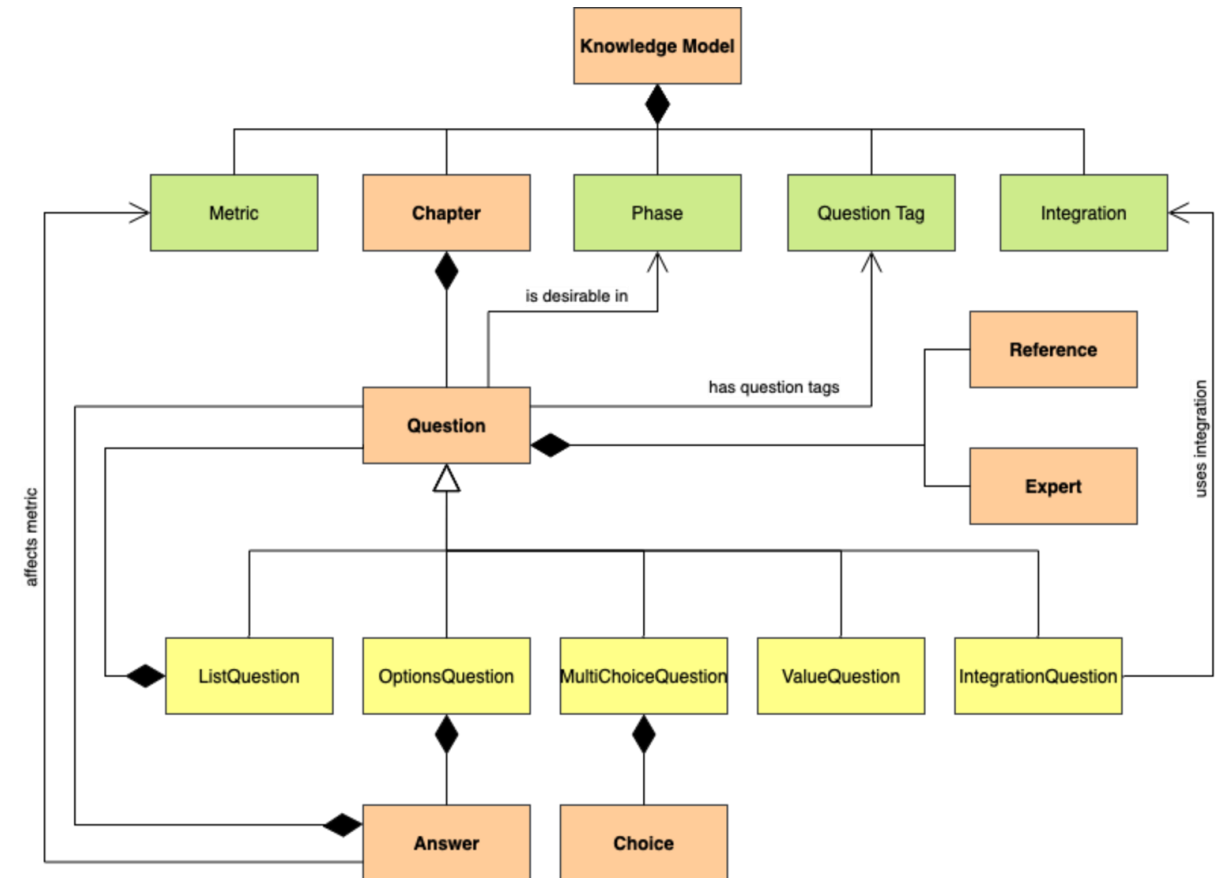


Knowledge Model



Knowledge Model (KM)

- Obsahuje **znalosti** o tom, na co a jak je potřeba se ptát
- **Šablona** pro strukturovaný dotazník (questionnaire)
- Stromová struktura sestávající z **kapitol, otázek, odpovědí**, navazujících otázek a dalších zdrojů



Data Stewardship Wizard



Knowledge Model



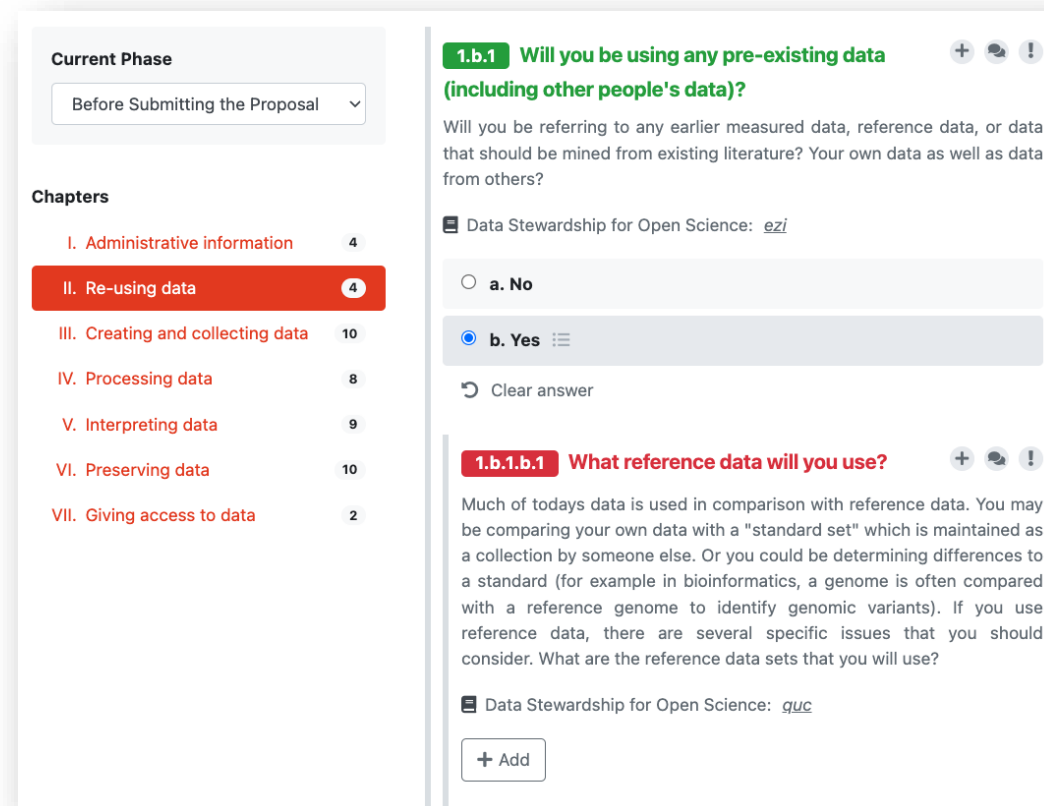
uses



Questionnaire

Dotazník (Questionnaire)

- Interaktivní formulář pro získávání informací o projektu založený na Knowledge Modelu
- Odpovídat je možné v libovolném pořadí
- Jsou zobrazovány pouze relevantní otázky na základě předchozích odpovědí



The screenshot shows a questionnaire interface with a sidebar and a main content area.

Current Phase
Before Submitting the Proposal

Chapters

I. Administrative information	4
II. Re-using data	4
III. Creating and collecting data	10
IV. Processing data	8
V. Interpreting data	9
VI. Preserving data	10
VII. Giving access to data	2

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?

Data Stewardship for Open Science: [ezi](#)

a. No

b. Yes

Clear answer

1.b.1.b.1 What reference data will you use?

Much of today's data is used in comparison with reference data. You may be comparing your own data with a "standard set" which is maintained as a collection by someone else. Or you could be determining differences to a standard (for example in bioinformatics, a genome is often compared with a reference genome to identify genomic variants). If you use reference data, there are several specific issues that you should consider. What are the reference data sets that you will use?

Data Stewardship for Open Science: [quc](#)

+ Add

- Hodnota (Value) – text, email, datum, číslo,...
- Výběr odpovědi (Options) – umožňuje větvení dotazníku
- Výběr z více možností (Multi-Choice)
- Seznam položek (List of items) – každá položka má stejné otázky (pod-dotazník)
- Integrace (Integration) – propojení s externí službou

Otázka s výběrem jedné možnosti

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

Popis → 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.

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

Odpovědi →

- a. A domain-specific repository **Findability**
- b. Our national repository **Findability** → **FAIR metriky**
- c. Our institutional repository **Findability**
- d. A special-purpose repository for the project **Findability**

Kdo odpověděl → Answered in less than 5 seconds by Albert Einstein.

Informace k odpovědi → 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

- Propojení s **externí databází** nebo službou na úrovni Knowledge Modelu
- Odpověď na otázku není pouze text ale i **odkaz na vybranou položku**



Integrační otázka

1.b.1.a.1 Reference database or dataset

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

Banana Breeding Tracker Database

Banana21

MusaBase

Musa Germplasm Information System

Banana Genome Hub

Integrační otázka

1.b.1.a.1 Reference database or dataset

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

- Banana Breeding Tracker Database
- Banana21
- MusaBase
- Musa Germplasm Information System
- Banana Genome Hub

Vyhledávání v externí
databázi



Integrační otázka

1.b.1.a.1 Reference database or dataset

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

- Banana Breeding Tracker Database
- Banana21
- MusaBase
- Musa Germplasm Information System
- Banana Genome Hub

Vyhledávání v externí databázi




FAIRsharing.org
standards, databases, policies

Získání odpovědi s odkazem

1.b.1.a.1 Reference database or dataset

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

Banana Breeding Tracker Database

 <https://fairsharing.org/10.25504/FAIRsharing.EYkWo>

Clear answer

Integrační otázka

1.b.1.a.1 Reference database or dataset

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

- Banana Breeding Tracker Database
- Banana21
- MusaBase
- Musa Germplasm Information System
- Banana Genome Hub

Vyhledávání v externí databázi



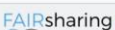
FAIRsharing.org
standards, databases, policies

Získání odpovědi s odkazem

1.b.1.a.1 Reference database or dataset

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

Banana Breeding Tracker Database

 <https://fairsharing.org/10.25504/FAIRsharing.EYkWo>




Clear answer


Otevření odkazu a zobrazení detailů

FAIRsharing.org search through all content

GENERAL INFORMATION

This record was deprecated on 2021-9-22 for the following reason(s): This resource is no longer available at the stated homepage, and an updated homepage cannot be found. Please get in touch with us if you have updates for this record.

 **Banana Breeding Tracker Database (BBTbase)**  

 10.25504/FAIRsharing.EYkWo

Type: Repository

Registry: Database

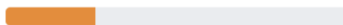





Description: The Banana Breeding Tracker Database (BBTbase) is a database of banana breeding information and allows users to input data from multiple stages in banana crop production, including hybridization, seed selection, planting and breeding. This database also provides...

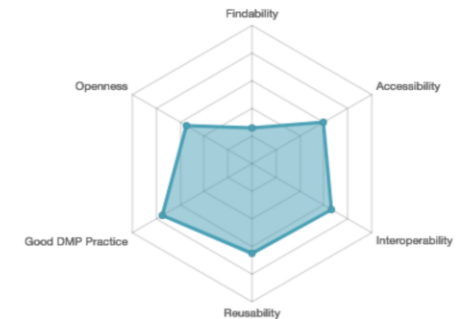
FAIR metriky

- Každá odpověď může ovlivňovat **FAIR metriky**
- Výsledná hodnota se počítá jako vážený průměr všech odpovědí, které danou metriku ovlivňují

Summary Report

Answered (current phase): 15/38
Answered: 41/101

Metric	Measure	
Findability	0.26	
Accessibility	0.60	
Interoperability	0.67	
Reusability	0.65	
Good DMP Practice	0.75	
Openness	0.55	



Online spolupráce



Before S

Albert Einstein Owner ✕

Nikola Tesla Editor ✕

Isaac Newton Viewer ✕

Visible by all other logged-in users
Other logged-in users can **edit** the Project.

Public link
Anyone with the link can **view** the Project.

<https://demo.ds-wizard.org/projects/7aaec?> **Copy link**

rch Pro

DMP S

he rese

ork are c

Before S

ew Documents Settings

Comments 7 TODOS 1 Version history

1.a.1 Data type + 1 comment !

Desirable: Before Submitting the Proposal

XML

Answered in less than 5 seconds by Albert Einstein.

1.a.2 How is this data structured? + !

Desirable: Before Submitting the Proposal

- a. A structured domain specific file with data and metadata fields
- b. A table or set of tables (consisting of 'data records')
- c. Complex data, like a graph

View resolved comments

Comments 1 Editor notes

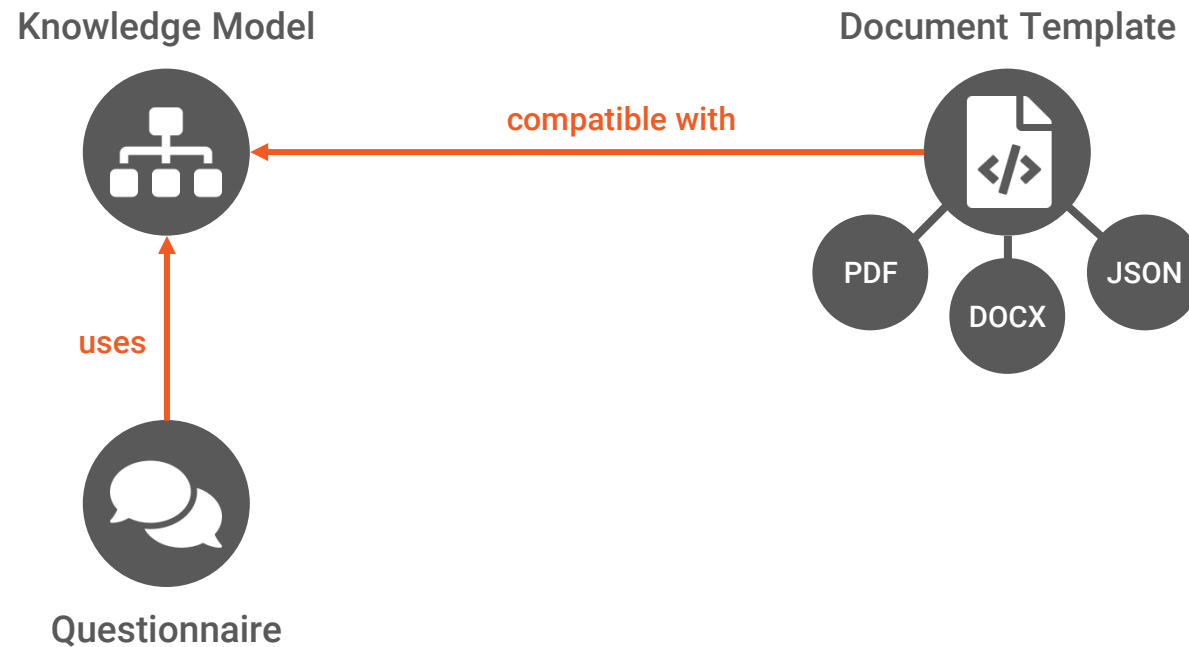
Anonymous user
1. 10. 2021, 14:49 ✓

Do we really want to use XML?

Reply...

Create a new comment...

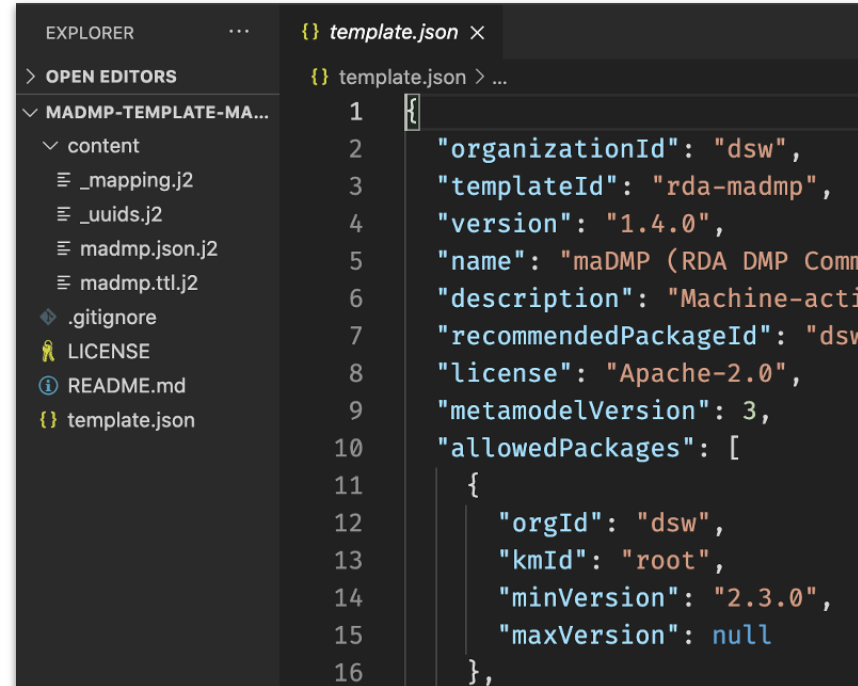
Data Stewardship Wizard



Šablona pro dokumenty (Document Template)



- Tvoření šablon vyžaduje více **technických dovedností**
- Šablony jsou složeny z **JSON** metadat, **Jinja2** šablon a dalších souborů
- Tvoření usnadňuje **DSW Template Development Kit (TDK)**

A screenshot of a code editor showing a JSON file named 'template.json'. The left sidebar shows a file explorer with a project structure including folders like 'content' and files like '_mapping.j2', '_uuids.j2', 'madmp.json.j2', 'madmp.ttl.j2', '.gitignore', 'LICENSE', 'README.md', and 'template.json'. The main editor area shows the JSON content of 'template.json' with line numbers 1 through 16. The JSON object contains metadata for a document template, including organization ID, template ID, version, name, description, recommended package ID, license, metamodel version, and a list of allowed packages with their own metadata.

```
1 {  
2   "organizationId": "dsw",  
3   "templateId": "rda-madmp",  
4   "version": "1.4.0",  
5   "name": "maDMP (RDA DMP Comm",  
6   "description": "Machine-acti",  
7   "recommendedPackageId": "dsw",  
8   "license": "Apache-2.0",  
9   "metamodelVersion": 3,  
10  "allowedPackages": [  
11    {  
12      "orgId": "dsw",  
13      "kmId": "root",  
14      "minVersion": "2.3.0",  
15      "maxVersion": null  
16    },  
17  ]  
18 }
```



- Horizon 2020 DMP
- **Horizon Europe DMP**
- Science Europe DMP
- Machine-Actionable DMP (RDA DMP Common Standard)
- Questionnaire Report



Data Management Plan

Science Europe Example

Contact person: **Jana Freeman** (jana.freeman@ds-wizc.cz)
0000-0000-0000-0001

[Czech Technical University in Prague](#)

Based on: *Common DSW Knowledge Model, 2.3.0*

Created by: **Jana Freeman** (jana.freeman@ds-wizc.cz)
DSW

Generated on: 24 Jun 2021

Projects

We will be working on the following projects and for those described in this DMP.

Arsenic and Selenium Speciation Using Hyphenated

Start date: 1.1.2021

End date: 31.12.2021

Funding: [Grantová Agentura České Republiky: grant \(planned\)](#)

The main goal of this study is to determine whether arsenic is produced from the meadows in the vicinity of old metal mines in Příbram, Kutná Hora, and Nalžovské Hory (Czech Republic) herbivorous herds. Total and speciation analysis of As and Se will be conducted using a hyphenated technique of HPLC and

Section A: Data Collection

1. What data will you collect or create?

Instrument datasets

The following instrument datasets will be acquired in the project:

- **HPLC**

This dataset will be collected by experts in the project using the following equipment.

The equipment is very well described and known.

- **ICP-MS**

This dataset will be collected by experts in the project using the following equipment.

The equipment is very well described and known.

Re-used datasets

We will use the following reference datasets:

- **Chemical Component Dictionary** (<http://dx.doi.org/10.26434/chem-cd>)

We will use the following already existing non-reference datasets:

- **Previous in-house Arsenic and Selenium analysis**

We already have a copy of this dataset.

Data formats and types

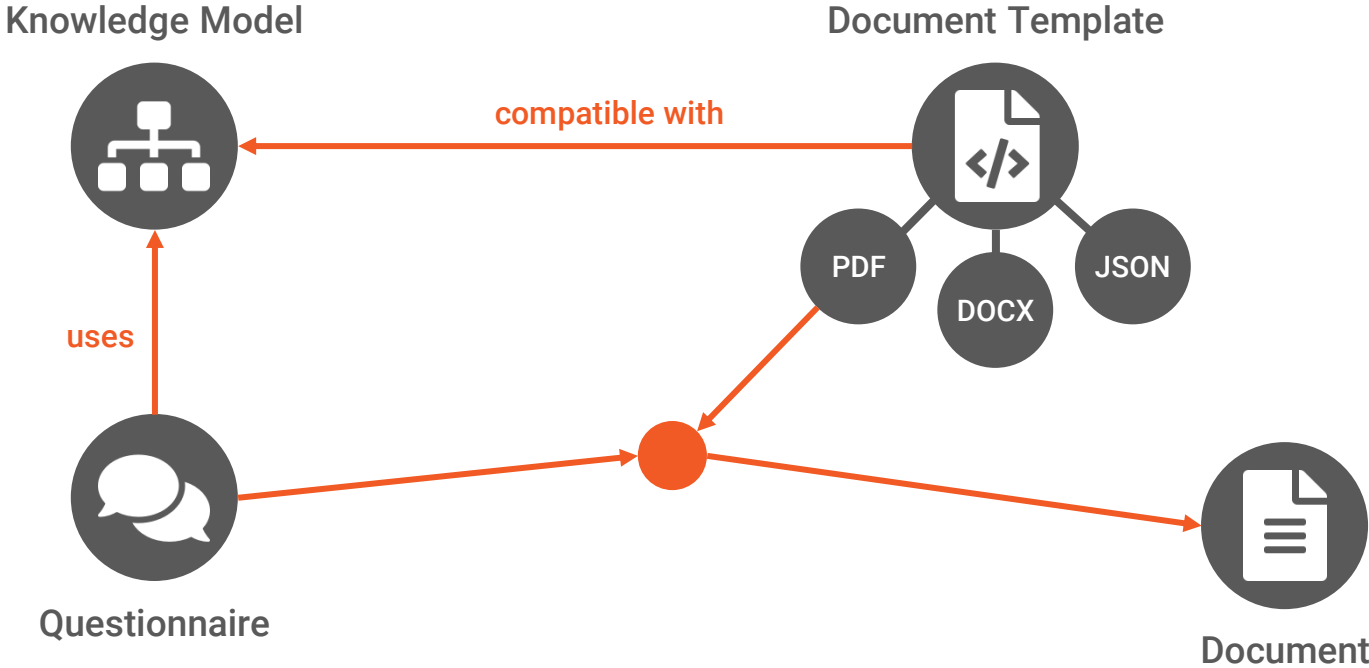
We will be using the following data formats and types:

- **Chemistry vocabulary**

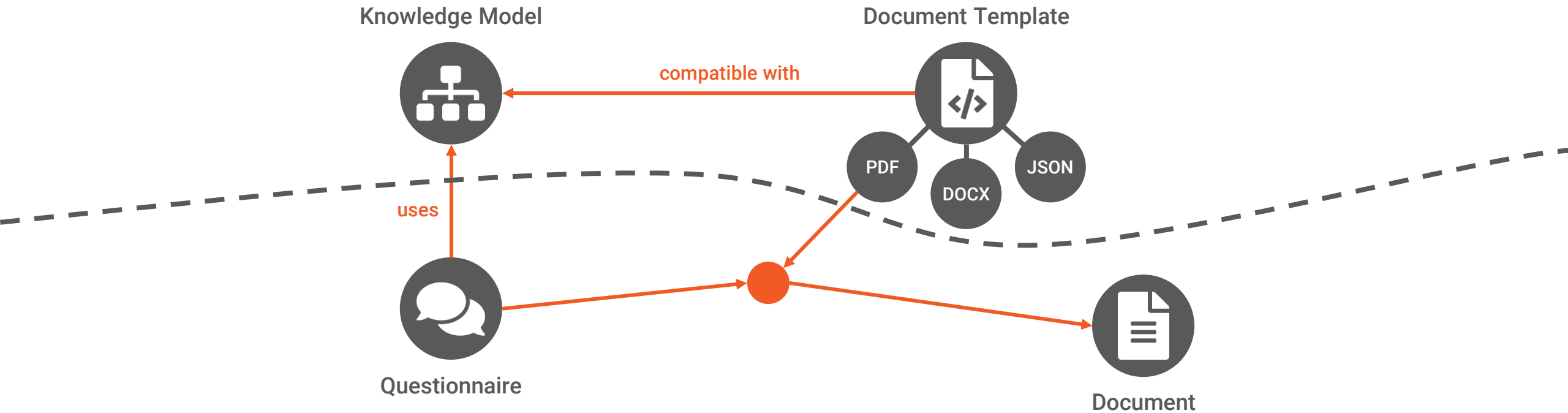
It is a standardized format. This is a suitable format for data storage. We will have only a small amount of data stored in the project.

```
{
  "dmp": {
    "title": "DMP in a planning phase",
    "description": "Example of a DMP describing a project in which source code will be created",
    "created": "2019-02-22T13:20:15.5",
    "modified": "2019-02-22T13:20:15.5",
    "project": [],
    "contact": {
      "mbox": "TMiksa@sba-research.org",
      "name": "Tomasz Miksa",
      "contact_id": {
        "identifier": "https://orcid.org/0000-0000-0000-0000",
        "type": "orcid"
      }
    }
  },
  "language": "eng",
  "ethical_issues_exist": "no",
  "dmp_id": {
    "identifier": "https://doi.org/10.0000/00.0.1234",
    "type": "doi"
  },
  "dataset": [
    {
      "title": "Source Code",
      "description": "Proof of concept implementation"
```

Data Stewardship Wizard



Data Steward



Výzkumník (Researcher)

A umí toho mnohem více...

- Projektové fáze
- Projektové šablony
- Komentáře a poznámky autorů
- Historie změn a verzování
- Migrace na novější verze KM
- Integrace pomocí Widgetu
- Odeslání dokumentu
- Jednotné přihlášení přes OpenID
- ... a další



ideas.ds-wizard.org

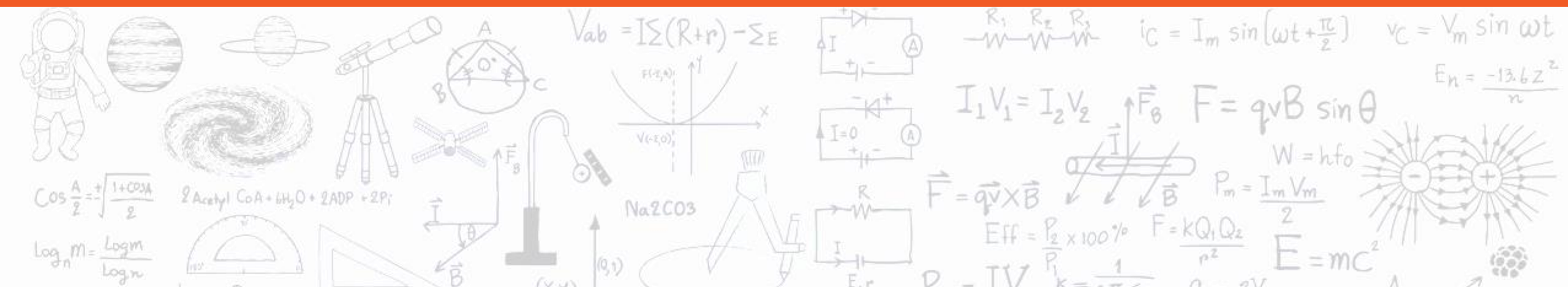
GitHub.com/ds-wizard



DSW

DATA STEWARDSHIP WIZARD

Horizon Europe DMP



- Šablona DMP pro „Horizon“ projekty v rámci období 2021–2027 (Evropská komise, [Programme Guide](#))
- Využívá se i pro jiné projekty, např. OP JAK
- Vychází z Horizon 2020 DMP šablony
- Důraz na FAIR data, ale i další výstupy projektu
- Překlad šablony od NTK ([10.48813/sstg-4g21](#))



- **Shrnutí dat** = jaká data (znovupoužitá, nasbíraná, odvozená, generovaná, ...), objem dat, formáty dat, komu a jak budou užitečná (tzv. data utility)
- **FAIR Data** = persistentní identifikátory, repozitáře, metadata, otevřenost dat (licence) a zabezpečení, čitelnost dat, použití standardů a ontologií, formáty, provázání na jiné datové daty, zajištění kvality dat, dokumentace
- **Další výstupy výzkumu/projektu** = popis jiných výstupů než dat, např. software, workflows, modely, protokoly, ale i fyzické vzorky, materiály, atd. a jejich FAIRness
- **Alokace zdrojů** = jak a z čeho alokujeme zdroje na FAIRness a další práci s daty (např. dlouhodobé uchování v repozitářích) a kdo je zodpovědný za data management

- **Bezpečnost dat** = opatření pro zabezpečení dat: přístup k datům, přenos citlivých dat, zajištění důvěryhodnosti i při dlouhodobém uložení
- **Etická stránka dat** = etické a právní překážky pro sdílení dat, souhlas se zpracováním a ukládáním dat, případně informace o anonymizaci/pseudonymizaci dat
- **Ostatní** = další relevantní informace k DMP, především popis postupů pro práci s daty vycházející z národních, grantových, sektorových nebo institucionálních požadavků, předpisů nebo doporučení (např. odkázání se na interní metodiku práce s daty)

DMP popisuje celý *Data Life Cycle*



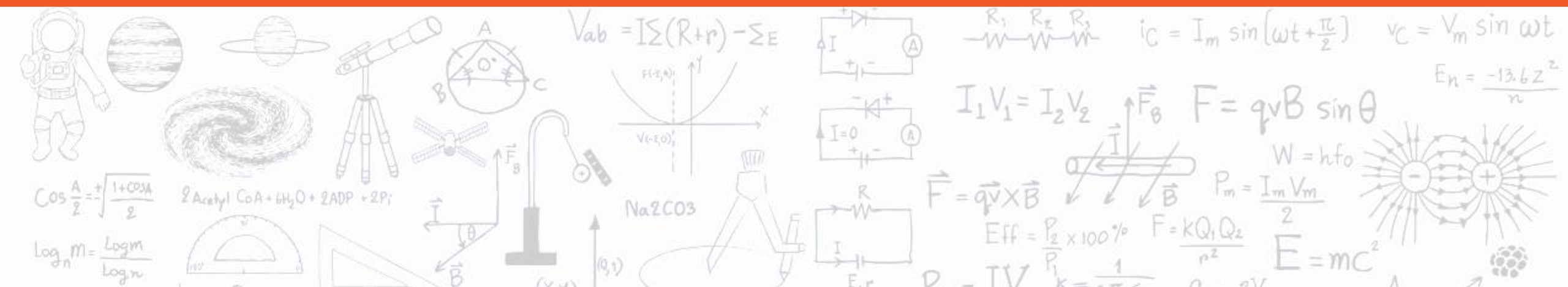
DMP je živý dokument



- Plán se logicky může v průběhu projektu měnit
 - Nové technologie a možnosti pro práci s daty
 - Nové výzkumné okolnosti během projektu mající vliv na práci s daty
 - Nové právní a jiné normy regulující práci s daty
- Jedná o tzv. *živý dokument*
- Některé grantové žádosti vyžadují aktualizaci DMP v pozdějších fázích projektu
- Na začátku projektu nemusí být přesný, dokumentuje naši připravenost (odhady, plán)
- Na konci projektu by měl DMP reflektovat / dokumentovat výsledný stav



Praktické použití DSW



Možnosti použití DSW



- Veřejná ELIXIR instance DSW: <https://researchers.ds-wizard.org>
 - Základní obsah (KM a šablony)
- Institucionální instance v rámci ELIXIR
 - Závislost na infrastruktuře ELIXIR / CESNET
 - Úprava uživatelů, obsahu a nastavení
 - Možnosti SSO a integrace s interními systémy
- Vlastní institucionální instance (on-premise / cloud, např. <http://fair-wizard.com>)
 - Plná kontrola nad integracemi, konfigurací a daty

Ukázka použití DSW pro plánování



Poděkování



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



Webinář byl podpořen v rámci projektu Rozvoj kapacit pro oblast výzkumu a vývoje na Mendelově univerzitě v Brně II , CZ.02.2.69/0.0/0.0/18_054/0014577.

Děkuji za pozornost



- Data Stewardship Wizard

<https://ds-wizard.org>

 [@dswizard_org](https://twitter.com/dswizard_org)

info@ds-wizard.org



- [RDMkit](#)
- [FAIR Principles](#)
- [Data Stewardship Wizard](#)
- [DSW Guide \(guide.ds-wizard.org\)](#)