



Úvod do DMP s nástrojem Data Stewardship Wizard

25. 6. 2025



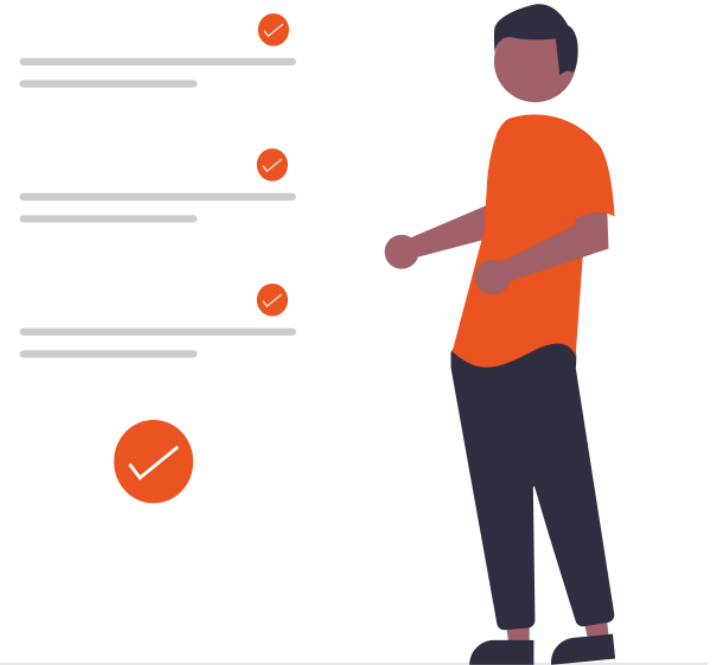
**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**

Kryštof Komanec

krystof.komanec@cvut.cz

 0000-0003-3856-1682

- Research Data Management
- Data Management Plan
- Data Stewardship Wizard



Data Life Cycle



Data Life Cycle: Plan



- Plánování jakým způsobem se bude během projektu pracovat s daty
- Výstupem by měl být **Data Management Plan**

Data Life Cycle: Collect



- Sbírání nových dat (metody se liší podle výzkumné oblasti)
- Použití existujících dat (např. z předchozích projektů)
- Zaznamenání původu dat (vzorky, výzkumníci, nástroje, atd.)
- Důraz na kvalitu zaznamenaných dat

Data Life Cycle: Process



- Převod dat ze zaznamenaného formátu do formátu vhodného pro analýzu
- Vyřazení špatných dat nebo dat z nízkou kvalitou
- Pseudonimizace/anonymizace citlivých dat

Data Life Cycle: Analyse



- Zkoumání nasbíraných dat
- Hlavní část výzkumu – získávání nových znalostí
- Workflow použité v analýze by mělo být reprodukovatelné
- Analýza velkých dat může vyžadovat velký výpočetní výkon

Data Life Cycle: Preserve



- Zajištění dlouhodobého uchování dat po ukončení projektu
- Možnost ověření výsledků projektu i po letech
- Využití dat v budoucnu pro jiné účely (výuka, jiný výzkum)

Data Life Cycle: Share



- Sdílení dat s dalšími lidmi (např. na dalším výzkumném projektu)
- Sdílení neznamená, že data musí být veřejně dostupná, mohou být sdílená pouze s omezeným přístupem
- Zvážení všech etických, právních, licenčních a jiných omezení

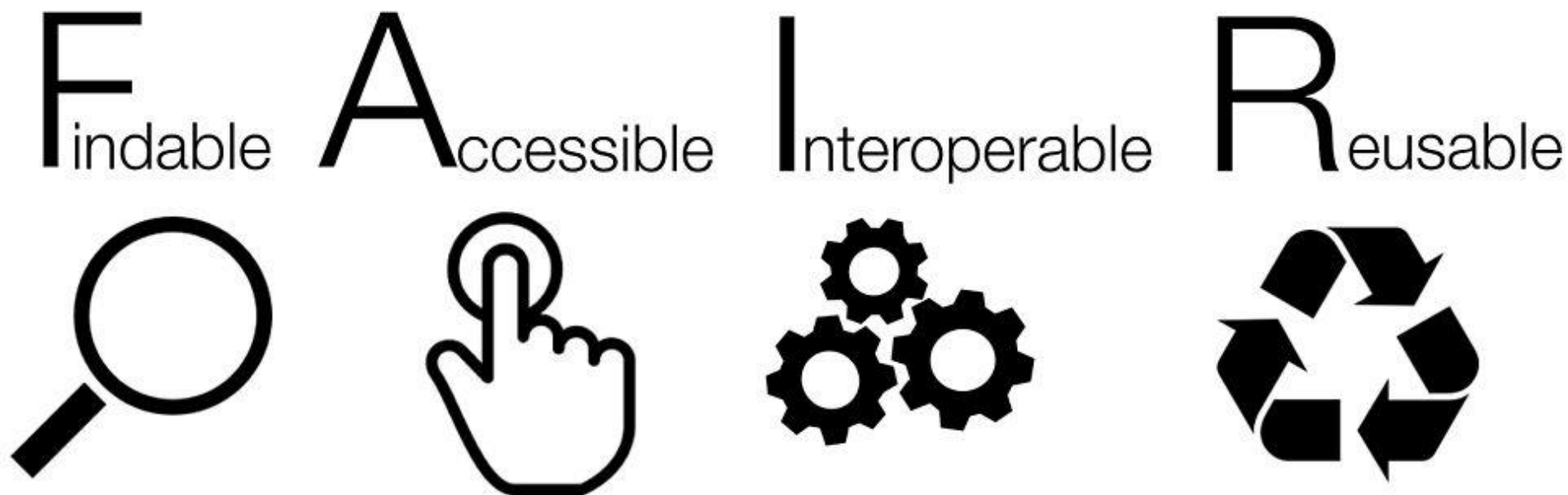
Data Life Cycle: Reuse

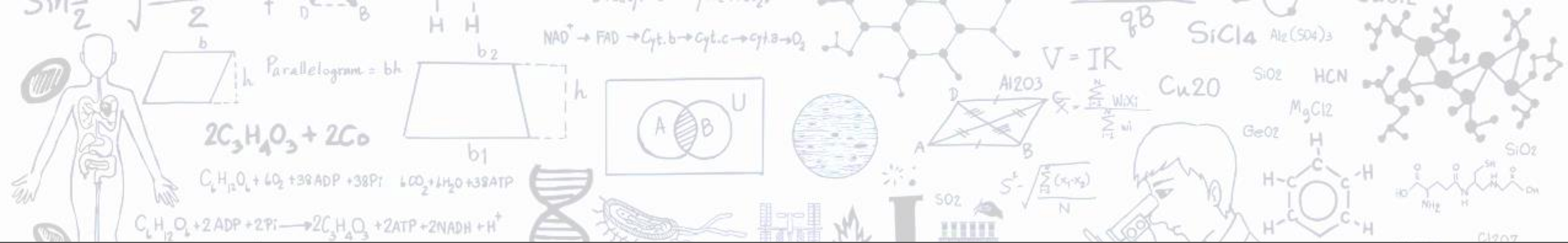


- Použití dat pro jiný účel, než pro který byla nasbírána, např.:
 - Jako referenční data pro jiný výzkum
 - Ověření výsledků původního výzkumu
 - Propojování výsledků více studií do meta-studií

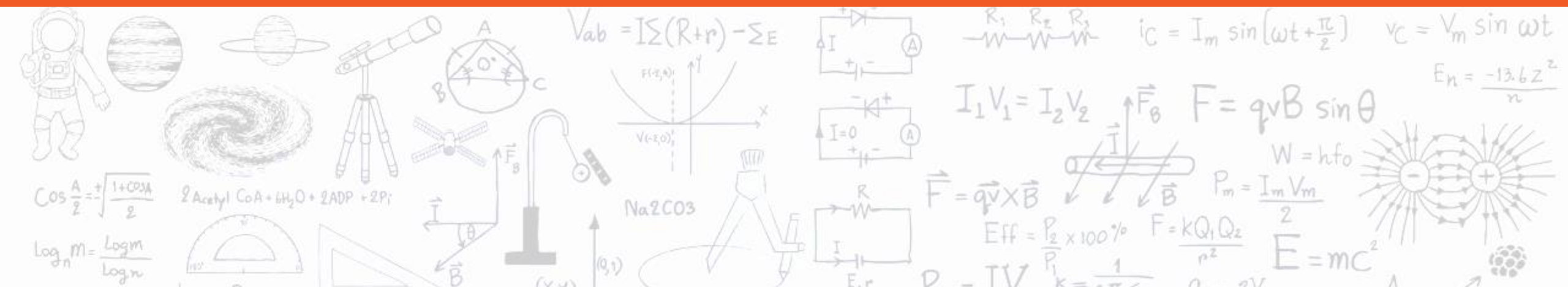
- **Researcher** = pracuje s daty během výzkumu, píše DMP pro svůj projekt
 - PhD, žadatelé o granty, projektoví manažeři, atd.
- **Data Steward** = stará se o práci s daty na různých úrovních
 - **Policy** = spolupráce s vedením a grantovými agenturami; dohlížení na správný postup, etické a právní záležitosti
 - **Research** = spolupráce s vědci; kontrola a pomoc s tvorbou DMP
 - **Infrastructure** = spolupráce s IT; IT řešení pro RDM, infrastruktura, atd.

- Hlavním cílem je vylepšit **znovupoužitelnost vědeckých dat**
- Zdůrazňují tzv. *machine-actionability* a kvalitní metadata





Data Management Plan



Data Management Plan (DMP)



- Obecné informace o projektu
- Popis dat, která budou během projektu použita, nasbírána a vytvořena
- Použití metadat a ontologií, dokumentace
- Ukládání dat, bezpečnost a strategie pro zachování dat po ukončení projektu
- Sdílení dat
- Náklady a lidské zdroje potřebné pro správu dat
- Etické a právní otázky, licence

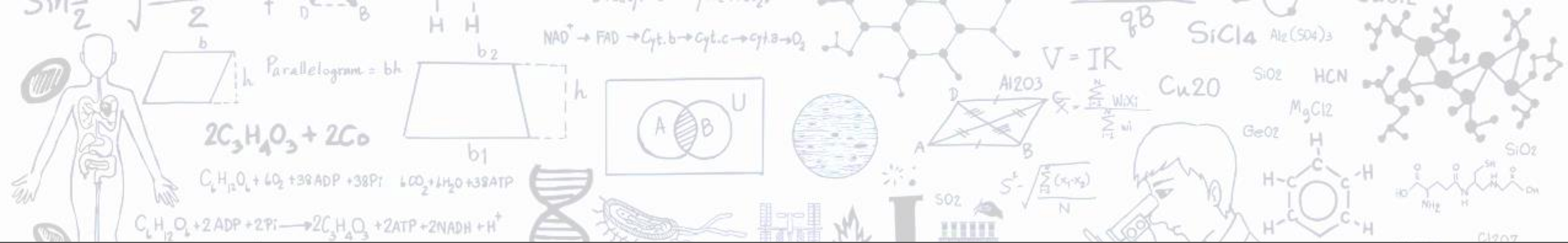
K čemu je dobré mít DMP



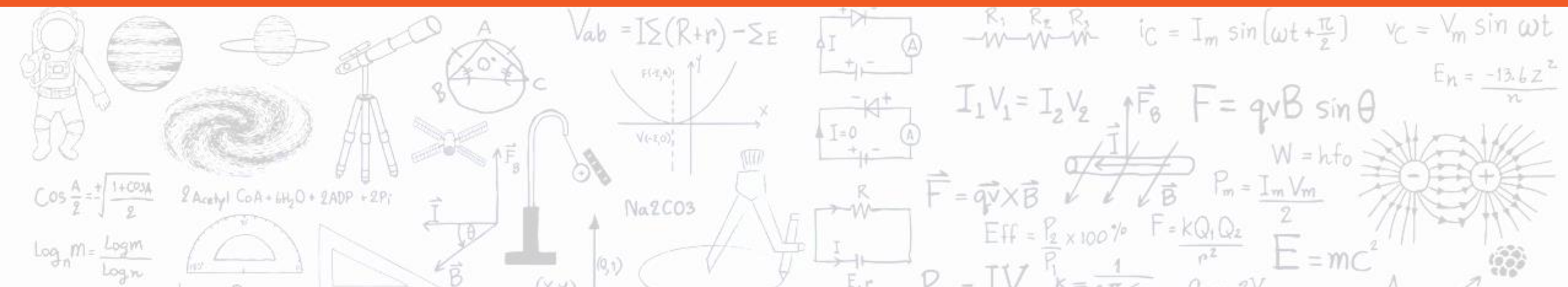
- Často je to vyžadované organizací nebo grantovou agenturou
- Pomůže při plánování prostředků a vybavení
- Definuje role a zodpovědnosti pro správu dat v týmu
- Pomůže identifikovat rizika a vybrat vhodná řešení na začátku projektu
- Usnadňuje sdílení, znovupoužitelnost a zachování dat
- Prevence vs. hašení požárů



... a mnoho dalších



Data Stewardship Wizard



- Open-source vyvíjený ve spolupráci **ELIXIR CZ** a **ELIXIR NL**
- Expertní systém pro plánování správy dat a tvorbu data management plánů
 - *“From burden to benefit”*
 - *“Plans are worthless, but planning is everything”*
- Doporučený nástroj v **Horizon Europe Program Guide**
- Vývoj nástroje i obsahu pokračuje také v rámci NRP

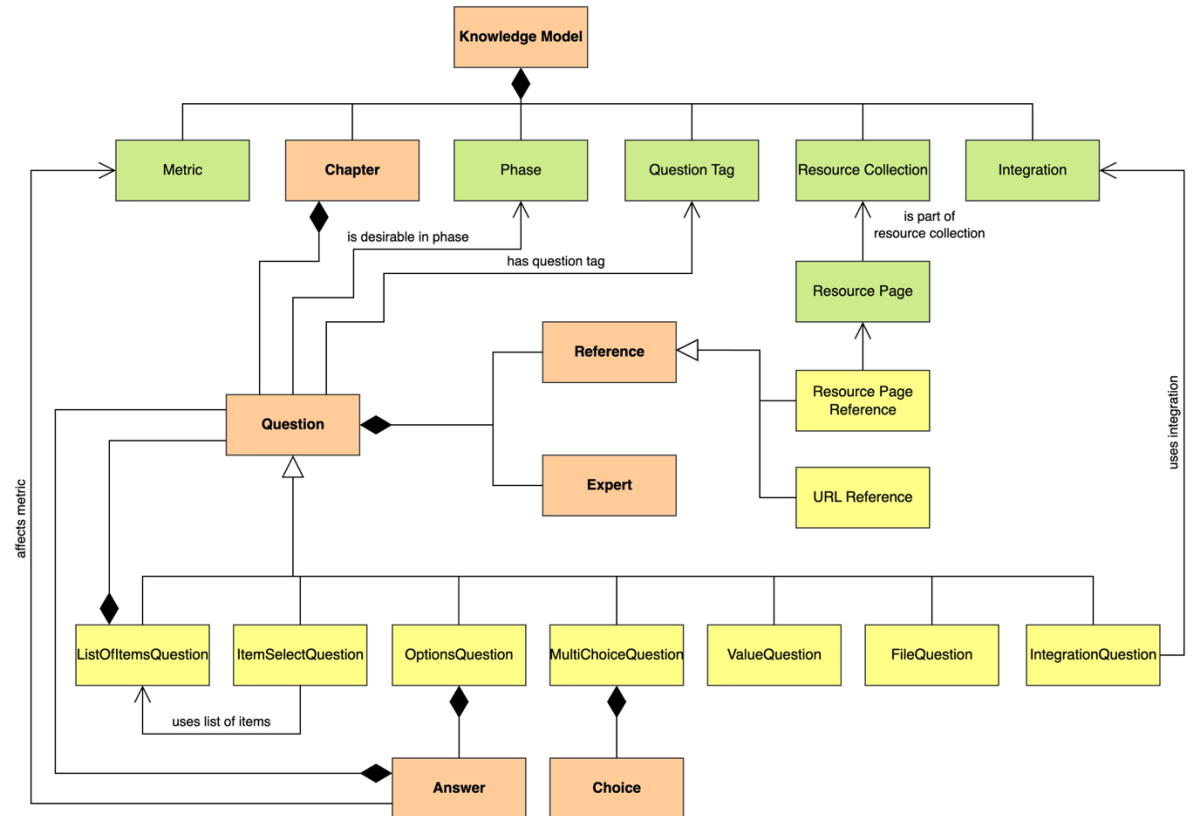
- Minimum psaní = plán není eseje, psaní jen tam, kde je to nezbytné
- Vedení = DSW vede uživatele skrz tzv. *smart questionnaire*
- Flexibilita = lze upravovat obsah a integrovat s jinými službami
- Otevřenost = kdokoli je může využít a vytvářet vlastní obsah
- Orientace na uživatele = vývoj se výrazně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



Knowledge Model Editor



Common DSW Knowledge Model

Publish

... > What existing reference data did ... > Will you be using this reference d... > Yes > What are the conditions of use fo...

Expand all Collapse all

Common DSW Knowledge Model

- Administrative information
- Re-using data
 - Do you need guidance to find existing data?
 - Is there any pre-existing data?
 - ☐ No
 - ☒ Yes
 - What existing reference data did you consider re-using?
 - Reference database or dataset
 - Where is this reference data available
 - Will you be using this reference data set?
 - ☒ No
 - ☒ Yes
 - Name and contact details of the owner of this data
 - What are the conditions of use for this database or dataset?
 - Do you know in what format the reference data is available?
 - Is the reference database or dataset versioned?
 - How will you make sure the same reference data will be available
 - What will you use this reference data set for?

Question Move Delete

Type

 Changing a question type will remove all answers.

Title

Text

Editor **Preview**

Although there is no world-wide rule for the application of copyright on data sets (copyright only applies to things that require a so-called "creative step"), it is wise to check for an explicit permission to use a data set and not to assume that data can be used freely just because you can access it.
Note that copyright laws explicitly **forbid** the use of a copyrighted work,

- Rozšíření nebo úpravy existujících KM
- Vhodné pro rozšíření pro jiné domény nebo instituce
- Příklad
 - **Common DSW Knowledge Model** – obecný KM pro data stewardship
 - **Life Sciences DSW Knowledge Model** – rozšiřuje obecný KM o otázky z Life Sciences domény
 - **Czech National Repository Platform KM** – WIP

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 displays the DSW questionnaire interface. On the left, a sidebar titled 'Current phase' shows 'Before Submitting the Proposal' as the active phase. Below it, a 'Chapters' list includes: I. Administrative Information (4), II. Re-using data (1), III. Creating and collecting d... (7), IV. Processing data (3), V. Interpreting data (2), VI. Preserving data (4), and VII. Giving access to data (3). The main content area is titled 'II. Re-using data' and contains introductory text about re-using data. It features two questions: 'II.1 Do you need guidance to find existing data?' with radio button options 'a. No' (selected) and 'b. Yes', and 'II.2 Is there any pre-existing data?' with radio button options 'a. No' and 'b. Yes'. The interface includes a 'Clear answer' button and a timestamp 'Answered less than a minute ago by Albert Einstein.'.

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 → Clear answer

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

- 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
- Výběr ze seznamu položek (Item select) – výběr položky vytvořené v otázce Seznam položek
- Soubor (File) – umožňuje nahrát jako odpověď soubor

- 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

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



FAIRsharing.org
standards, databases, policies

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

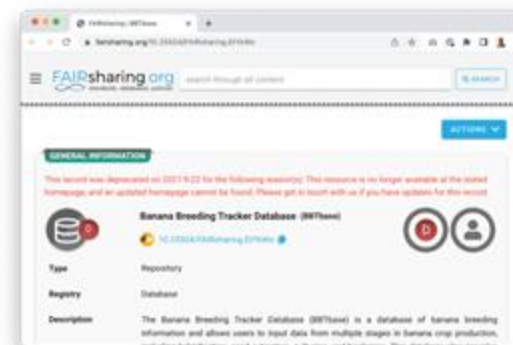
Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

Banana Breeding Tracker Database

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

Clear answer

Otevření odkazu a zobrazení detailů



FAIR metriky

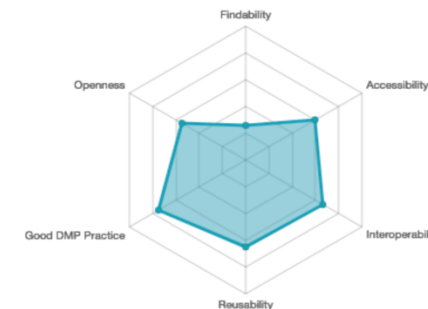
- Každá odpověď může ovlivňovat **FAIR metriky**
- Výsledná hodnota se počítá jako vážený průměr všech odpovědí, které danou metriku ovlivňují

Summary Report

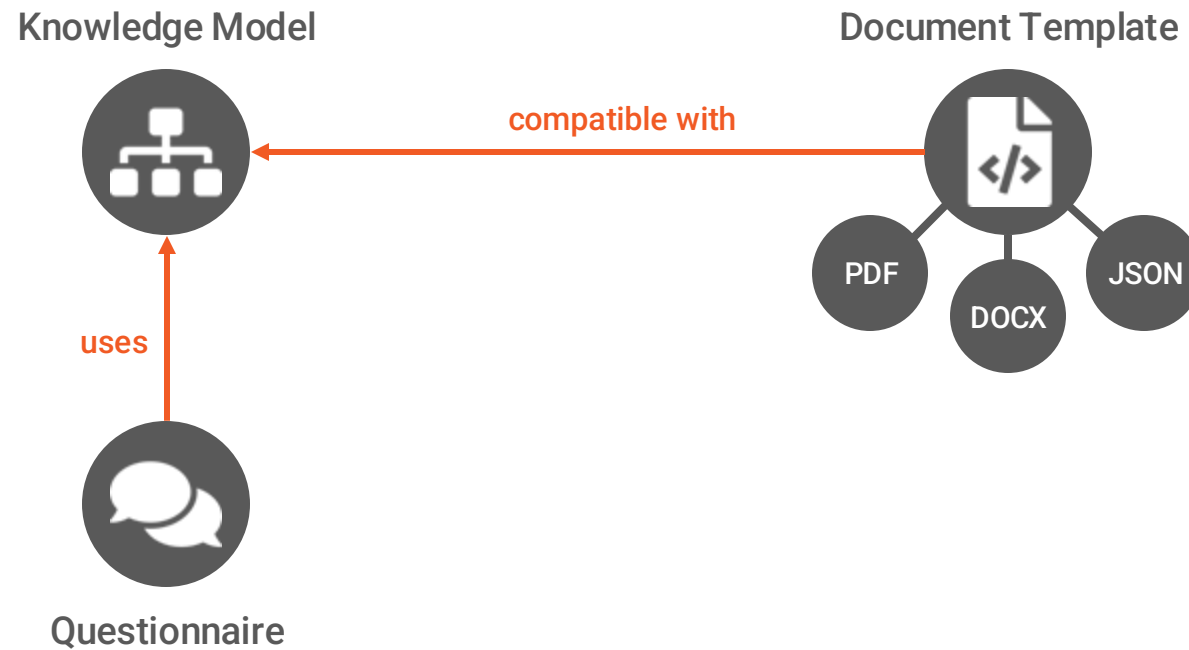
Answered (current phase): 15/38

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



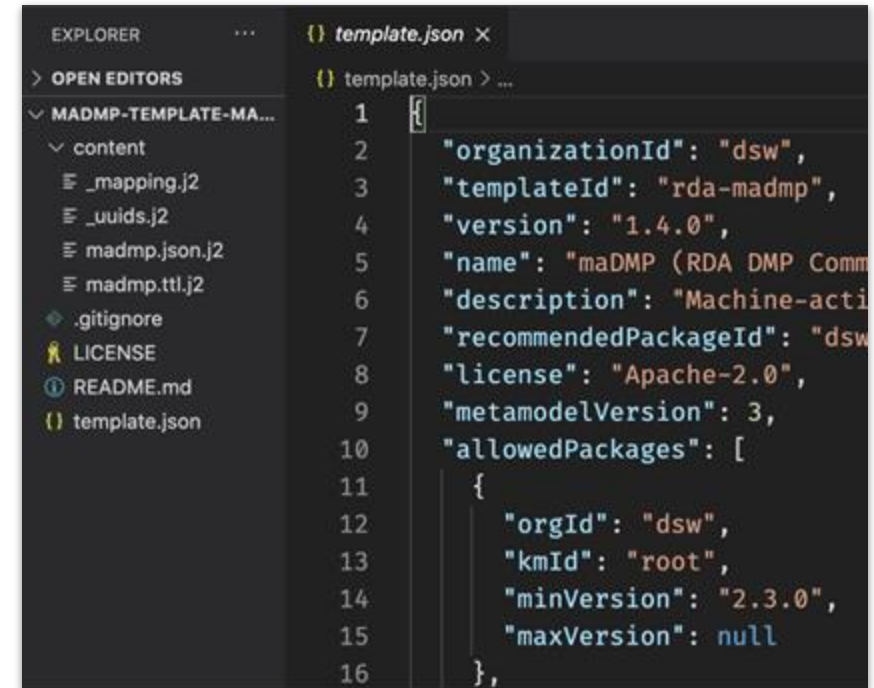
Data Stewardship Wizard



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)**

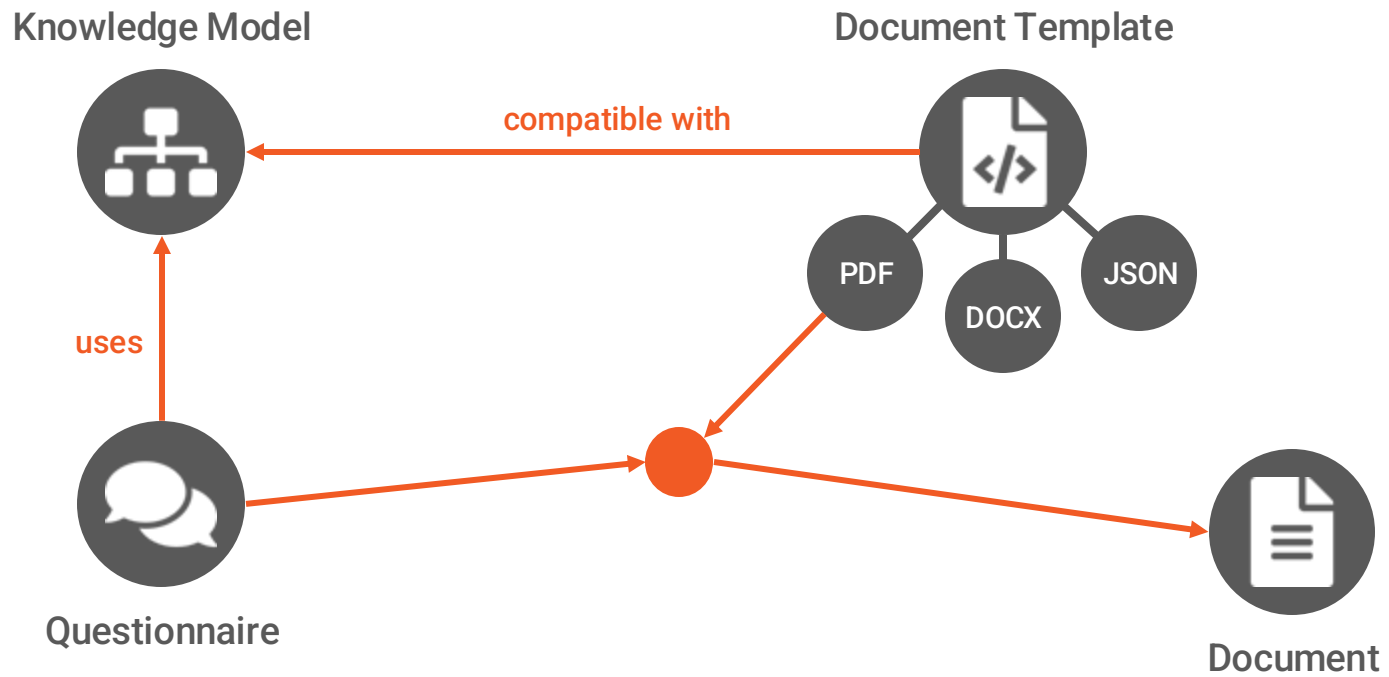
A screenshot of a code editor interface. On the left, the 'EXPLORER' pane shows a project structure with a folder 'MADMP-TEMPLATE-MA...' containing files like '_mapping.j2', '_uuids.j2', 'madmp.json.j2', 'madmp.ttl.j2', '.gitignore', 'LICENSE', 'README.md', and 'template.json'. The main editor area shows the content of 'template.json', which is a JSON object with the following fields: 'organizationId' (dsw), 'templateId' (rda-madmp), 'version' (1.4.0), 'name' (maDMP (RDA DMP Comm...), 'description' (Machine-acti...), 'recommendedPackageId' (dsw), 'license' (Apache-2.0), 'metamodelVersion' (3), and 'allowedPackages' (an array of objects). The 'allowedPackages' array contains one object with 'orgId' (dsw), 'kmId' (root), 'minVersion' (2.3.0), and 'maxVersion' (null).

```
{
  "organizationId": "dsw",
  "templateId": "rda-madmp",
  "version": "1.4.0",
  "name": "maDMP (RDA DMP Comm...",
  "description": "Machine-acti...",
  "recommendedPackageId": "dsw",
  "license": "Apache-2.0",
  "metamodelVersion": 3,
  "allowedPackages": [
    {
      "orgId": "dsw",
      "kmId": "root",
      "minVersion": "2.3.0",
      "maxVersion": null
    }
  ],
}
```



- Horizon 2020 DMP
- Horizon Europe DMP
- Science Europe DMP
- GAČR DMP
- MŠMT DMP
- AZV ČR DMPa
- Machine-actionable DMP (RDA Common Standard)
- Questionnaire Report

Data Stewardship Wizard



Vygenerované dokumenty



Data Management Plan

Science Europe Example

Contact person: **Jana Freeman** (jana.freeman@ds-wiz.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-wiz.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/chemrxiv-2019-08-01>)

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

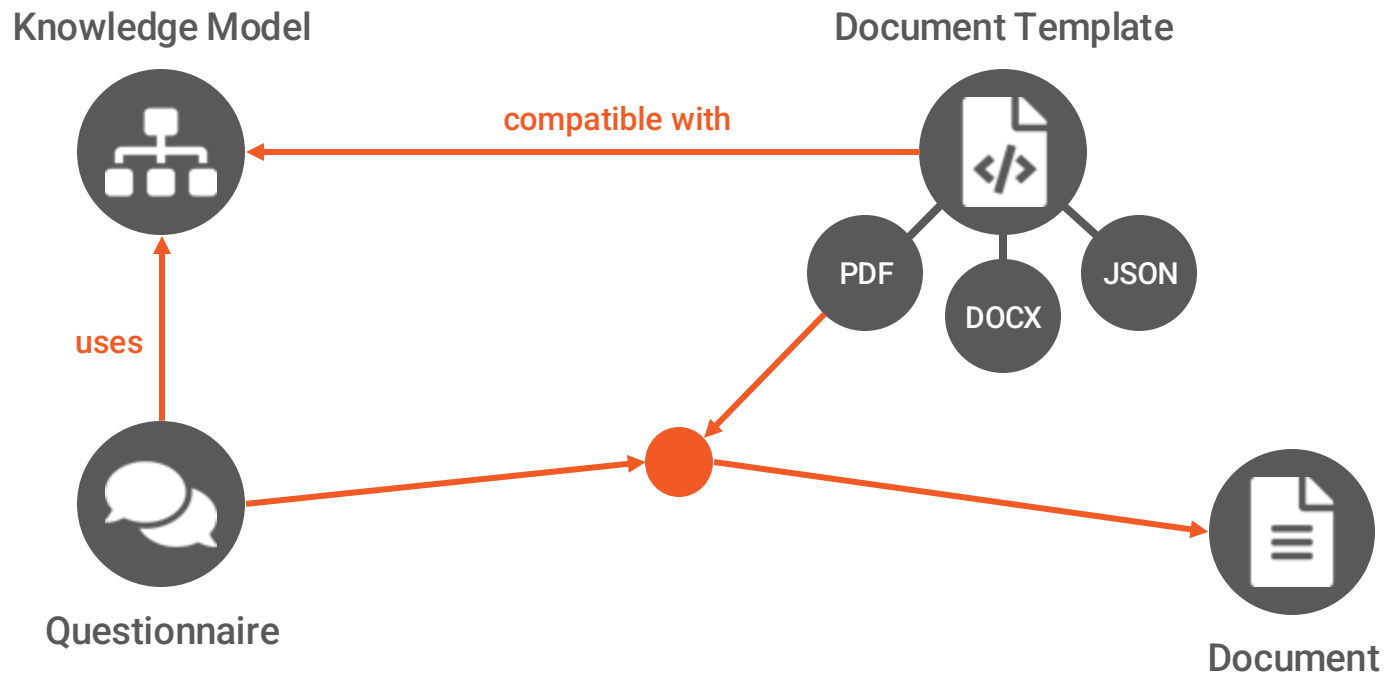
We will be using the following data formats and types:

• [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"
```

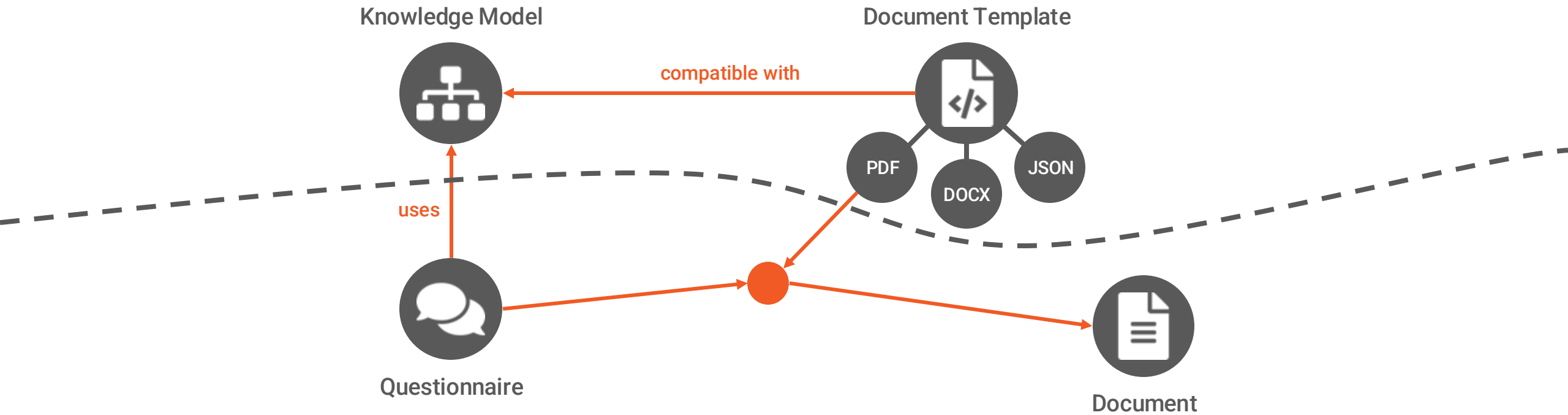
Data Stewardship Wizard



Data Stewardship Wizard



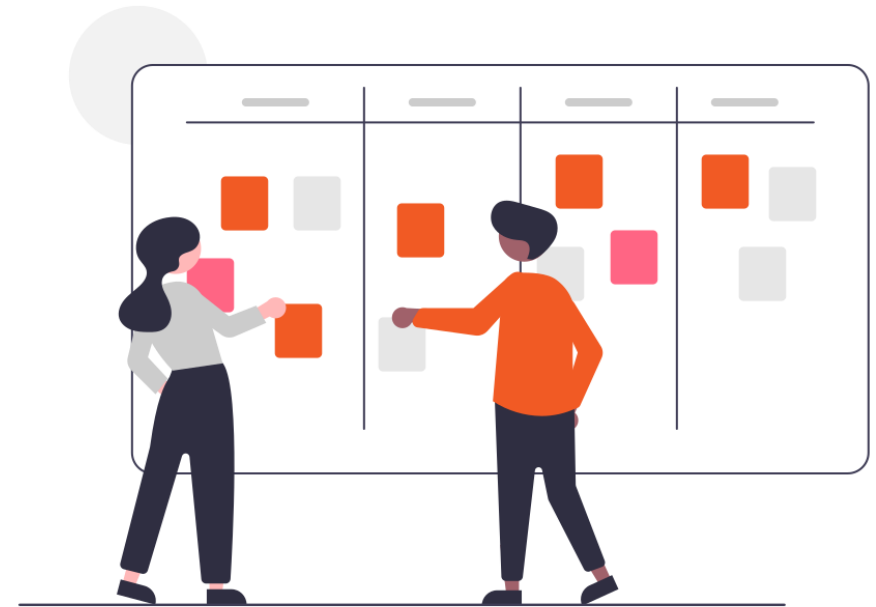
Data Steward



Researcher

A mnohem víc

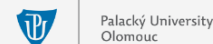
- Project phases
- Project templates
- Editor notes
- Version history
- Integration Widget
- Submission service
- Single-Sign On
- ...



ideas.ds-wizard.org

github.com/ds-wizard

Kde se DSW používá



Možnosti použití DSW



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

Poděkování (vývoj a provoz DSW)



Vývoj a provoz DSW je podpořen výzkumnou infrastrukturou ELIXIR CZ
(MŠMT Grant No.: LM2023055)

Děkuji za pozornost



- Data Stewardship Wizard
<https://ds-wizard.org>
 [@dswizard_org](https://twitter.com/dswizard_org)
info@ds-wizard.org
- <https://discord.gg/MW3H9tdMcT>



- [RDMkit](#)
- [FAIR Principles](#)
- [Data Stewardship Wizard](#)