DSW

DATA STEWARDSHIP WIZARD

DMPs, maDMPs, and DSW: An Outlook

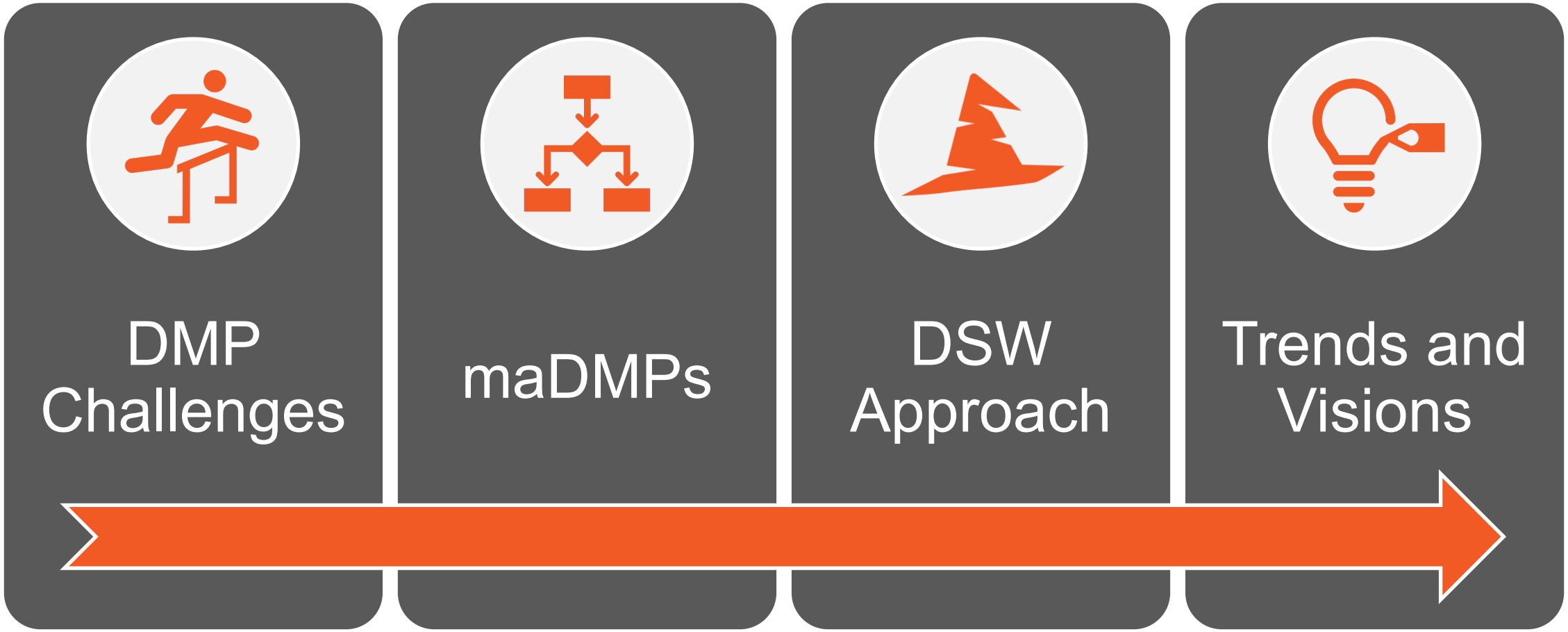
SND Network meeting in Gothenburg
Gothenburg, 24 April 2024

Marek Suchánek
marek.suchanek@ds-wizard.org
0000-0001-7525-9218

Marek and his Hats

- Data Stewardship Wizard [2018]
  - Czech Technical University in Prague (ELIXIR LM, NRP, OSTrails)
  - Institute of Organic Chemistry and Biochemistry (ELIXIR-STEERS)
- FAIR Wizard [2023]
  - Codevence Solutions
- RDA DMP Common Standards WG [2022]
- Model-Driven Development and Normalized Systems [2017]

Outline



DSW

DATA STEWARDSHIP WIZARD

Part I: DMP Challenges

Challenge: Various Requirements

Funder requirements

FAIR

Open

Reproducible

Recommended services

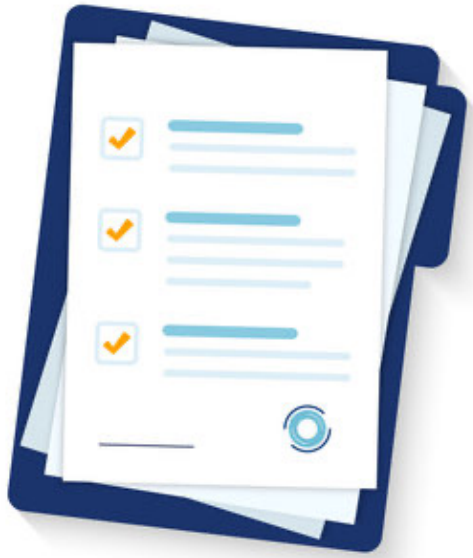
Institutional requirements

Internally compliant

Metadata catalogs

(Inter)national requirements

Law compliant



Domain best practices

FAIR

Reproducible

Domain-specific requirements

Effortless

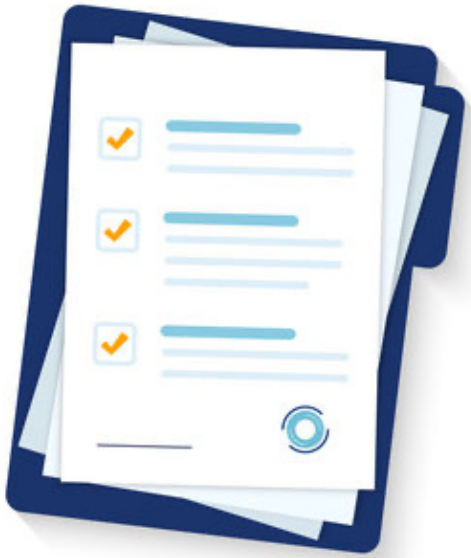
Easy to follow

Personal requirements

Consistency?

Feasibility?

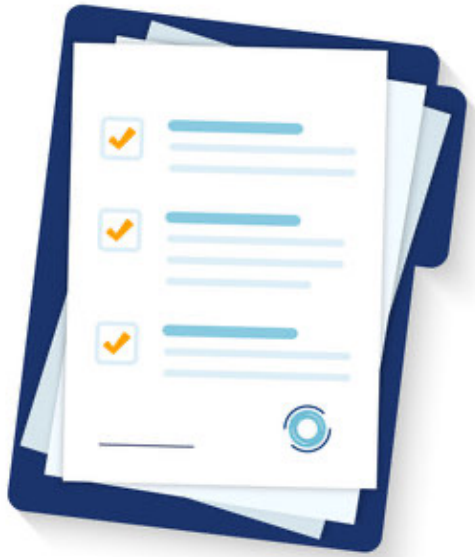
Challenge: Living Document, Versions



Follow the plan

Publish dataset

Add dataset DOI



DMP X v1

DMP X v2

Tracking changes?

Automated updates?

Referring to DMP?

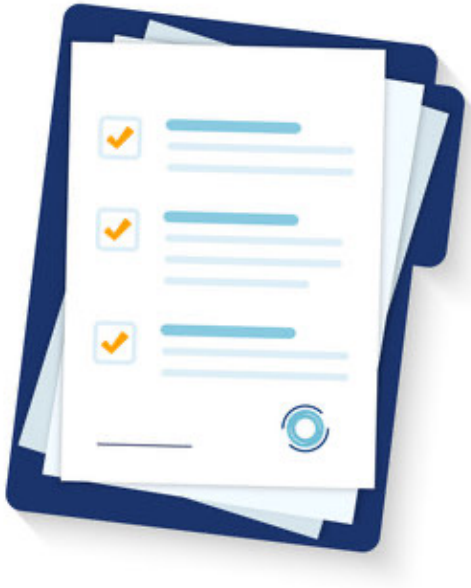
Challenge: Collaboration

Team A (CZ)

Datasets

Policies

Standards



Datasets

Team B (SE)

Standards

Policies

Common policies

Project coordinator

Common standards

Single plan for multiple partners?

Consistency?

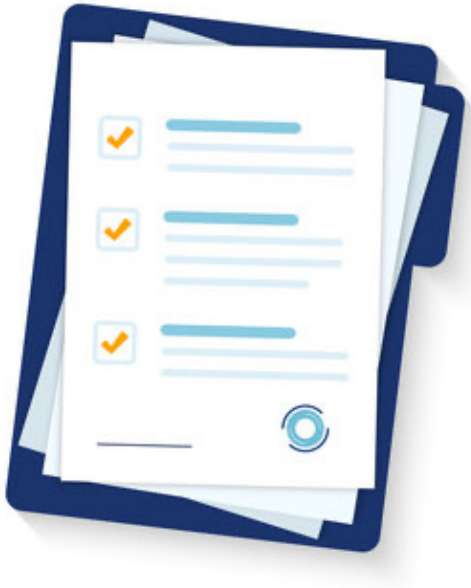
Coordinated updates?

Challenge: Evaluating DMPs

Funder

Compliance?

(Valid DMP?)



Complete?

Feasible?

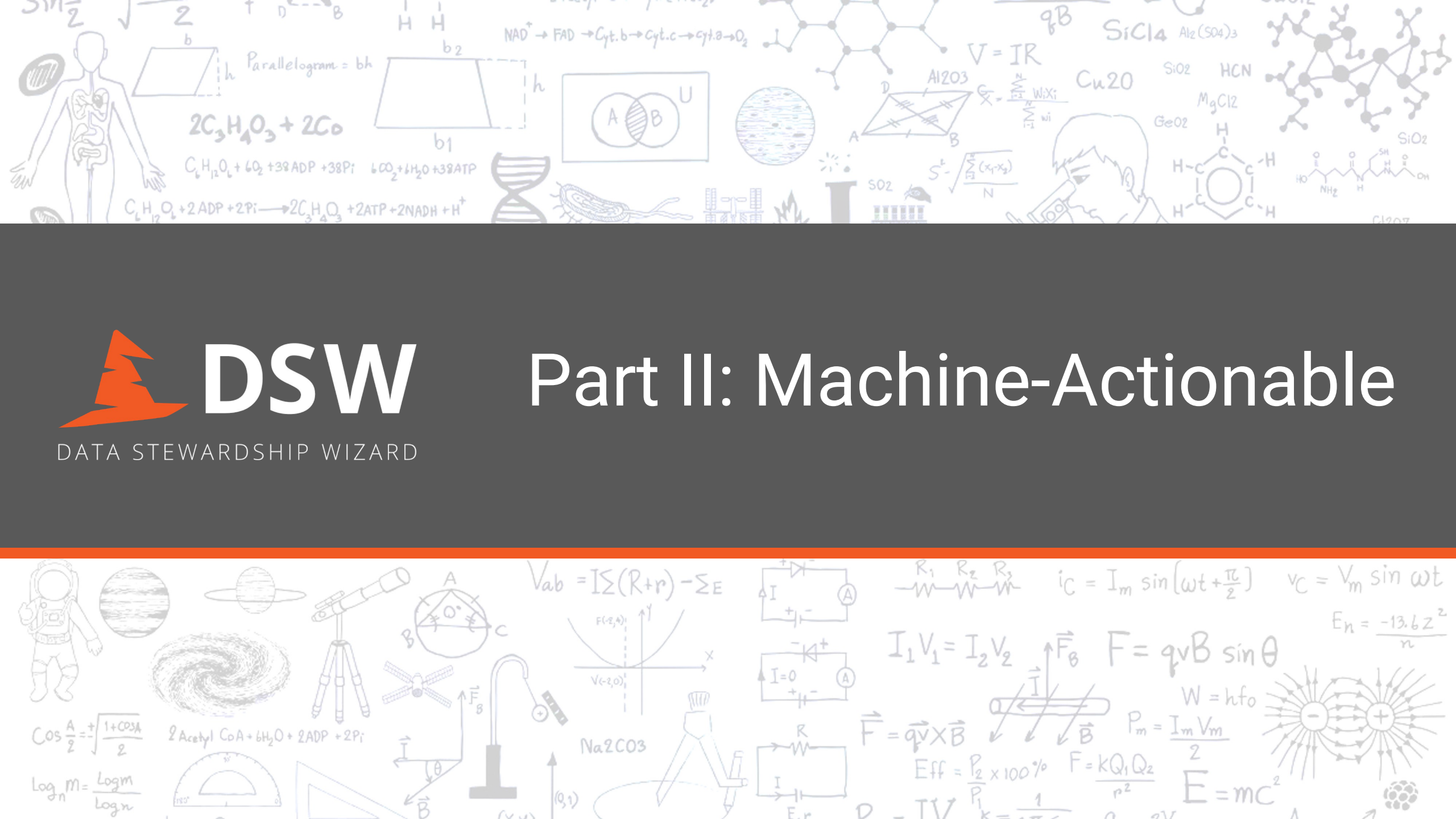
Space for improvement?

Researcher / Institution

Rapid feedback?

Unbiased response?

Full assessment?



DSW

DATA STEWARDSHIP WIZARD

Part II: Machine-Actionable

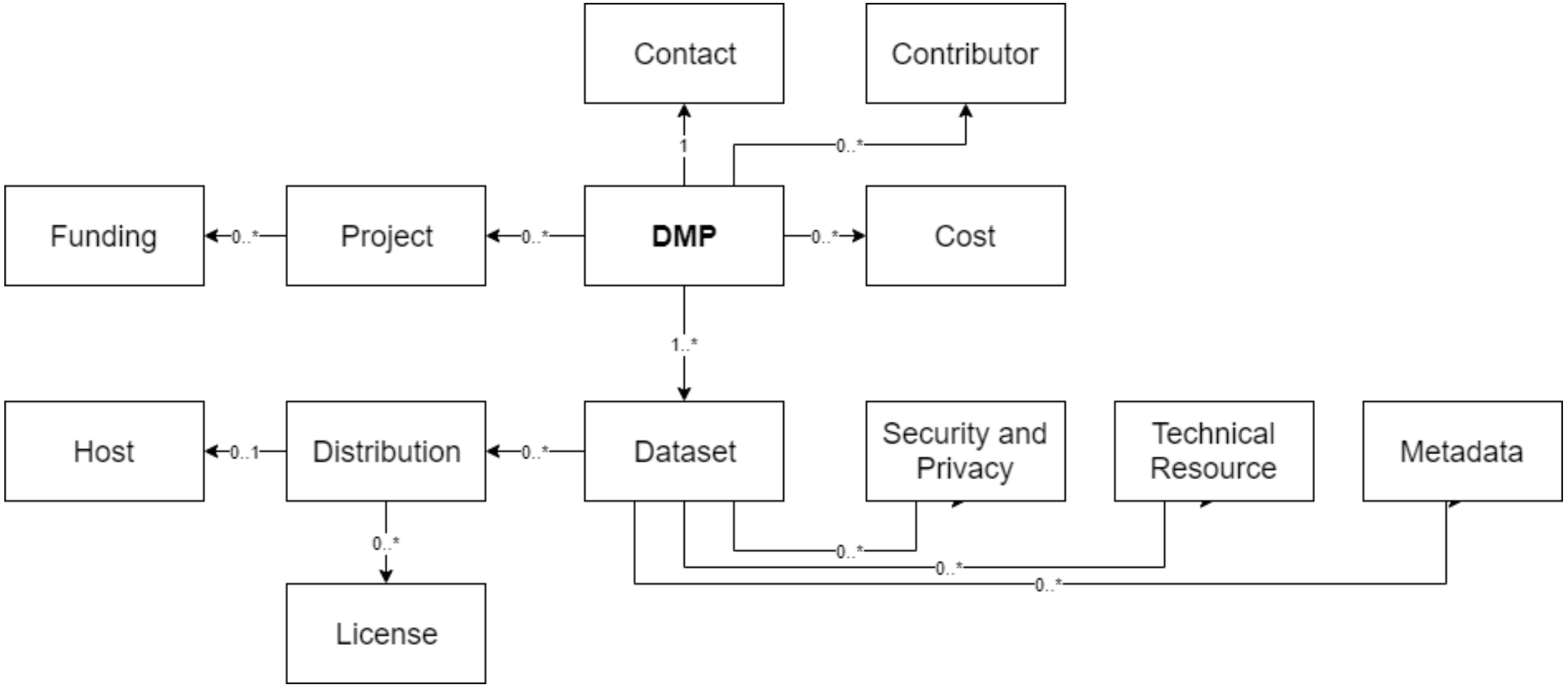
Machine-Actionable DMPs (maDMPs)

DSW

- Usually maDMP = DMP according to **RDA DMP Common Standard (DCS) for maDMPs**
- Started **>5 years ago (in 2019)**, pushed during hackathon 2020 (all main DMP tools)
- Covers **common properties** (the ultimate core) of DMPs
- Uses widely used ontologies/vocabularies such as DCAT, DCTerms, or DataCite
- Unfortunately, some parts are not so machine-actionable (e.g. ethical_issues_description)

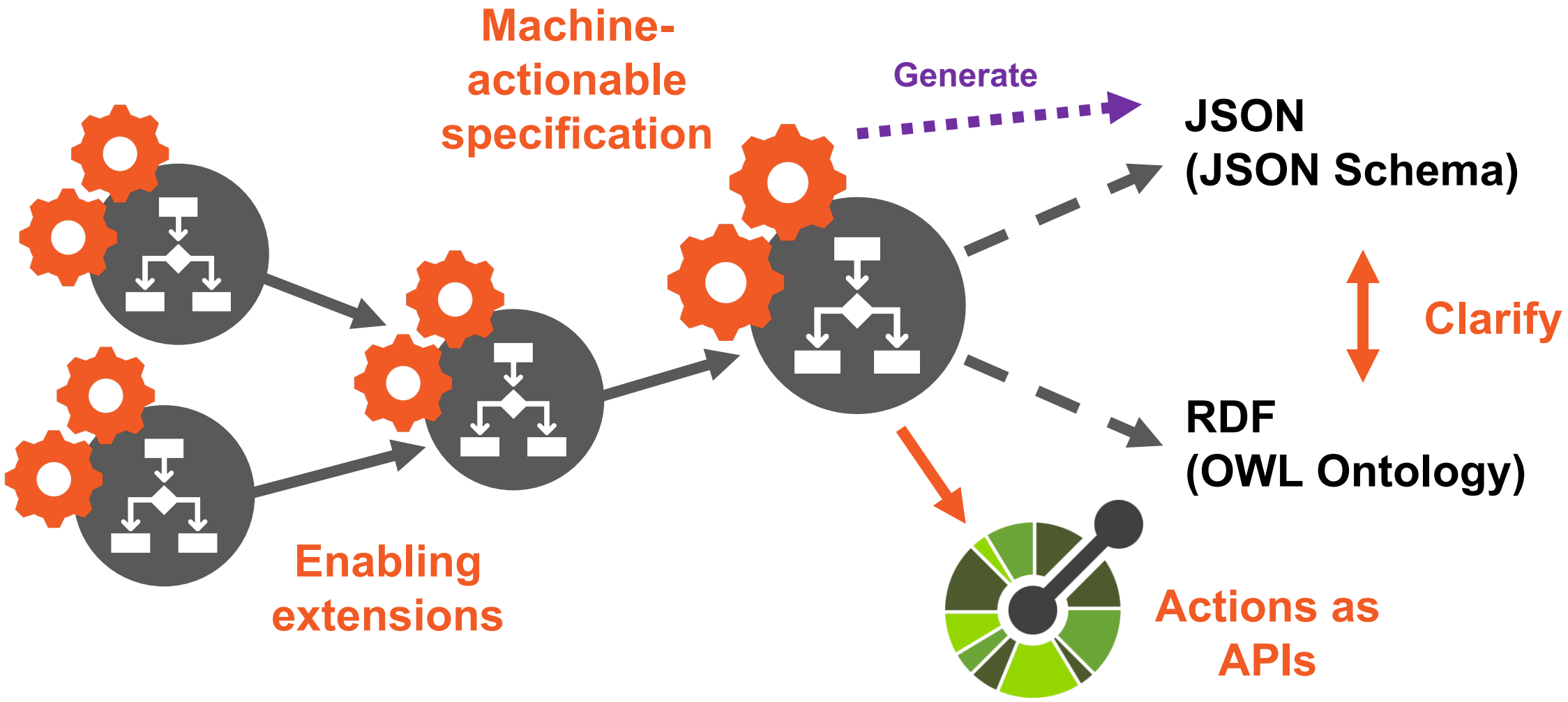
- **maDMPs can help with all the mentioned challenges**

RDA DMP Common Standard for maDMPs



https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard

Plans for maDMPs

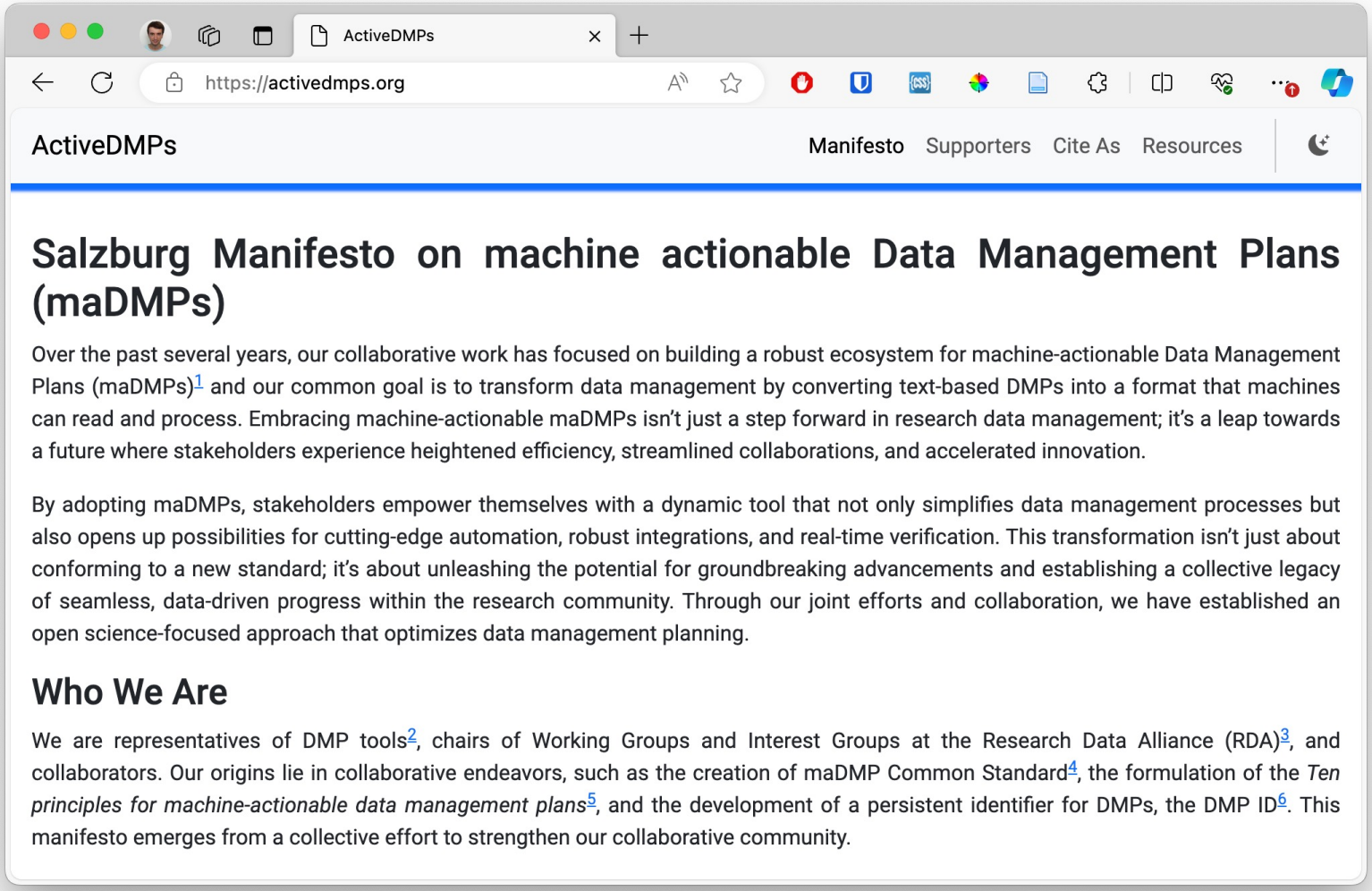


Salzburg Manifesto

- During RDA Plenary 22
- Collaboration
- Community of practice
- Adopting and promoting
- All key DMP platforms

- You can also support it:

  activeDMPs.org



DSW

DATA STEWARDSHIP WIZARD

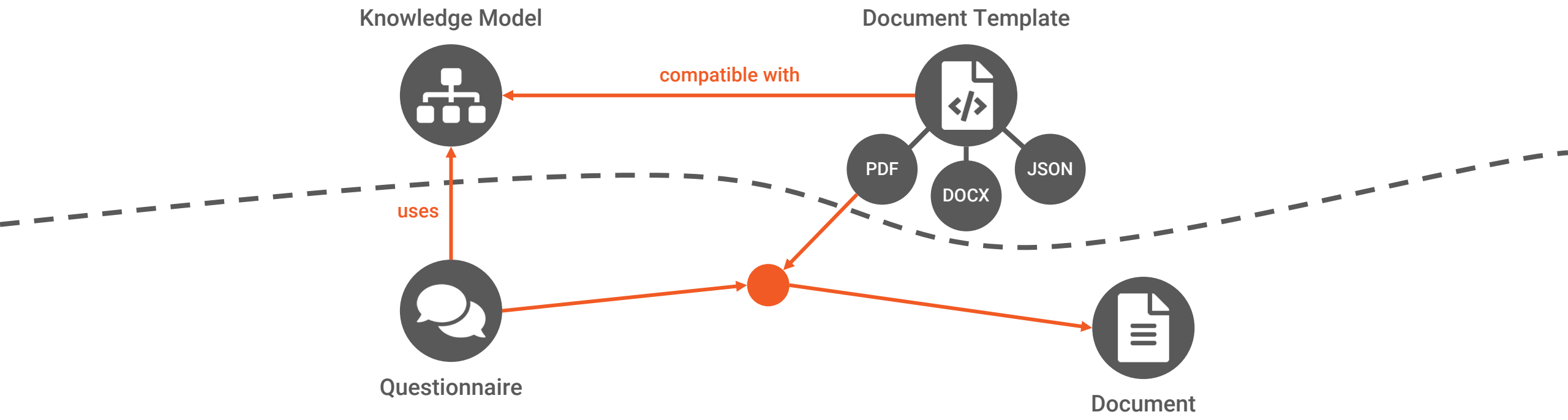
Part III: Approach of DSW

Data Stewardship Wizard (DSW)

- **Open-source platform** for data management planning
- Started within ELIXIR but now getting **contributions from various projects**
- **Avoiding writing** as much as possible, **asking only relevant questions**
- Plan within a **single questionnaire**, get **different documents** (outputs) when needed
- Customizable with **local guidance**, in as many layers as needed

DSW in a Nutshell

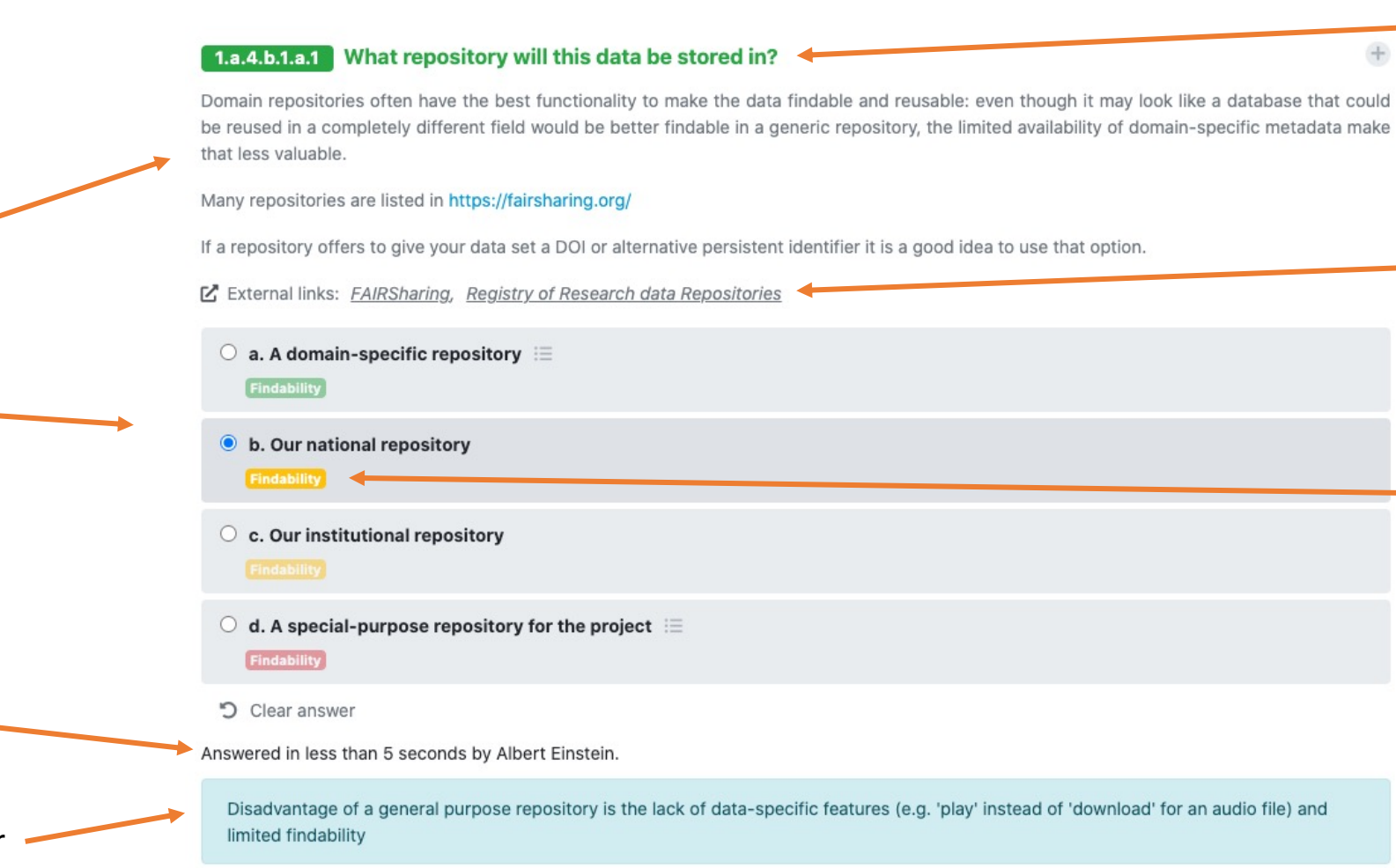
Data Stewards



Researchers

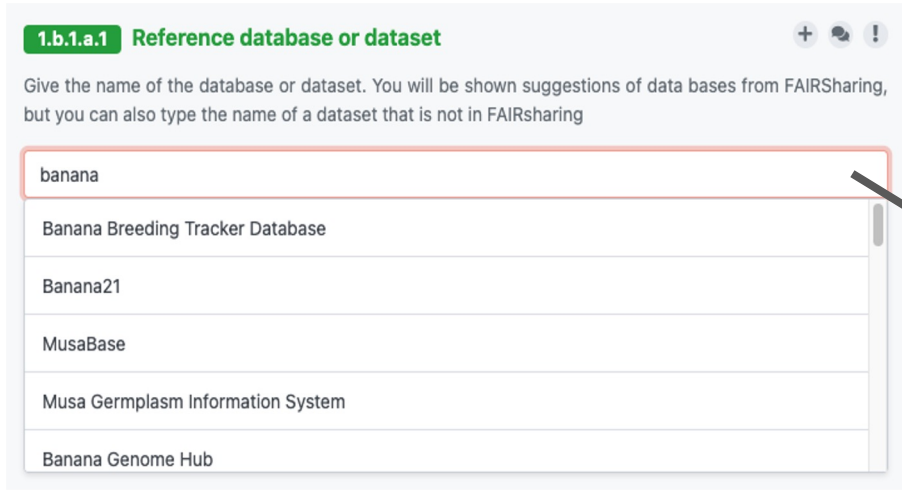
DSW: Questionnaire

Questionnaire



DSW: Questionnaire

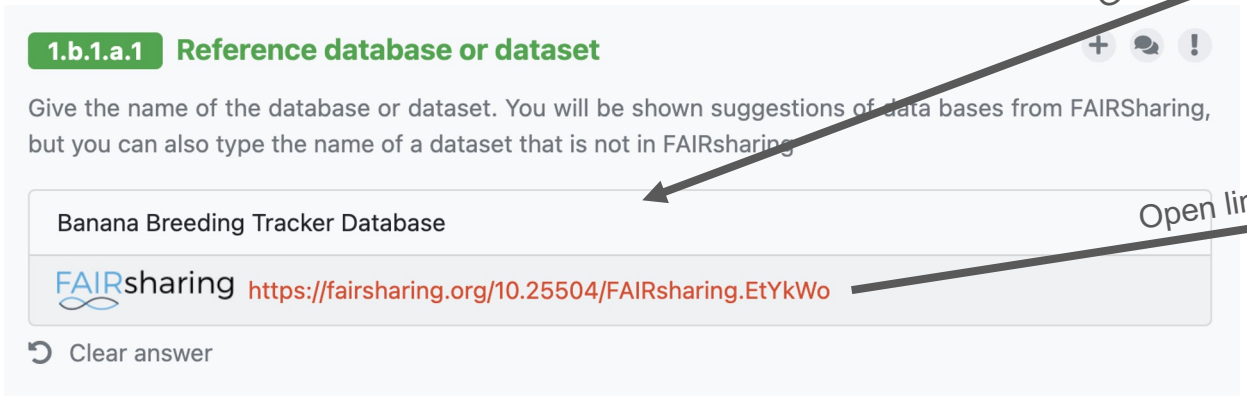
Questionnaire



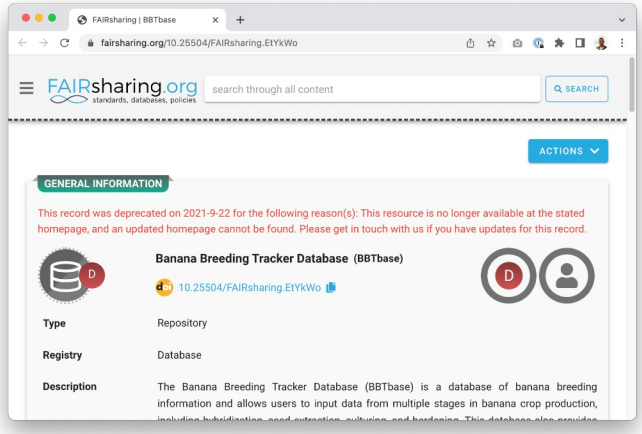
Search

FAIRsharing.org
standards, databases, policies

Get reply with URI

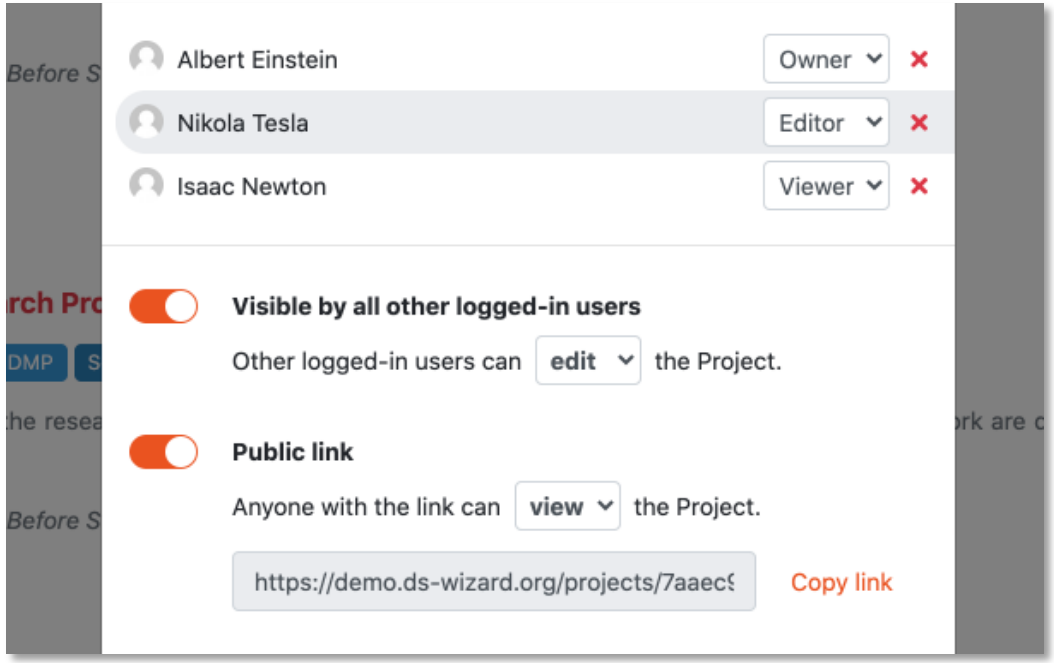


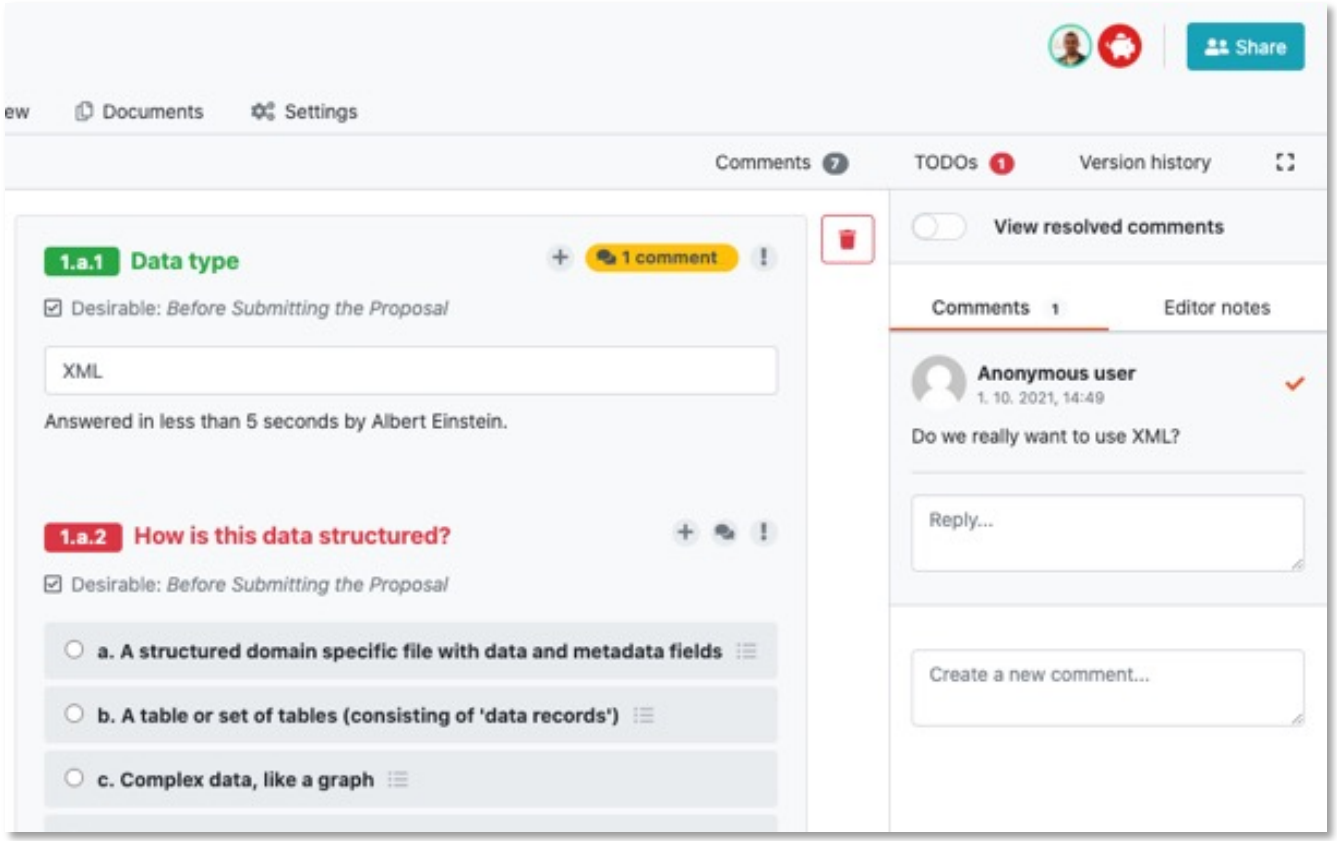
Open link with details



DSW: Questionnaire

Questionnaire





DSW: Questionnaire

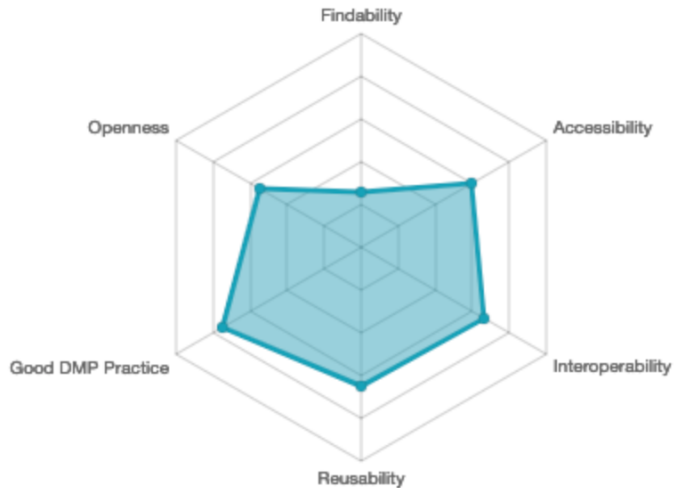
Questionnaire

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



DSW: Questionnaire

Document


```
{
    "dmp": {
        "title": "DMP in a planning phase",
        "description": "Example of a DMP describing a project in which source code will be crea
        "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 Management Plan

Science Europe Example

Contact person: ***Jana Freeman*** *(jana.freeman@ds-wiz*
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*
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 Hyphena

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 i produced from the meadows in the vicinity of old meta Příbram, Kutná Hora, and Nalžovské Hory (Czech Republi herbivorous herds. Total and speciation analysis of As and will be conducted using a hyphenated technique of HPLC a

Section A: Data Collection

1. What data will you collect or create?

Instrument datasets

The following instrument datasets will be acquired in the p

- **HPLC**

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

  The equipment is very well described and known.

- **ICP-MS**

  This dataset will be collected by experts in the project 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/0

We will use the following already existing non-reference da

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

  We already have a copy of this dataset.

Data formats and types

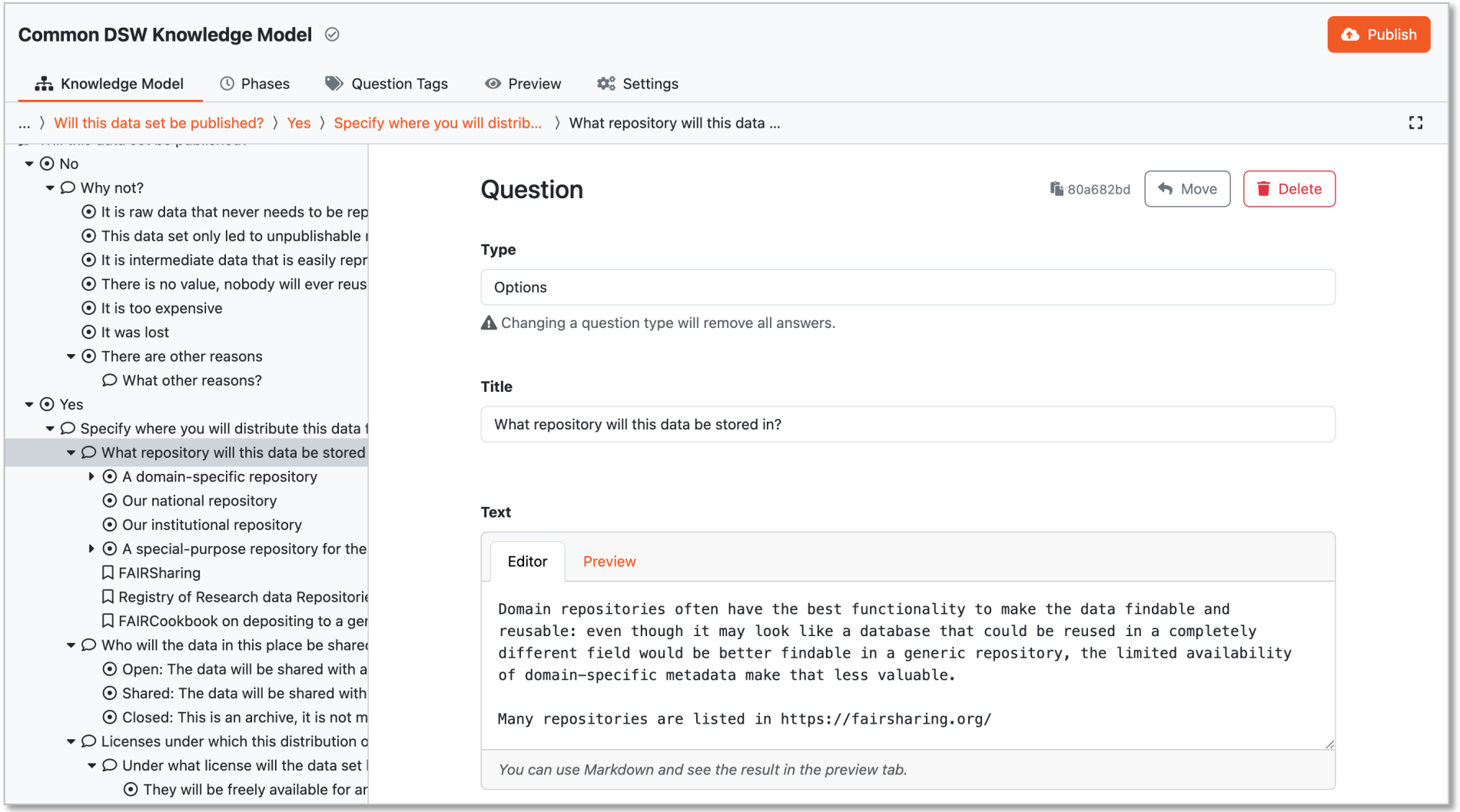
We will be using the following data formats and types:

- **Chemistry vocabulary**

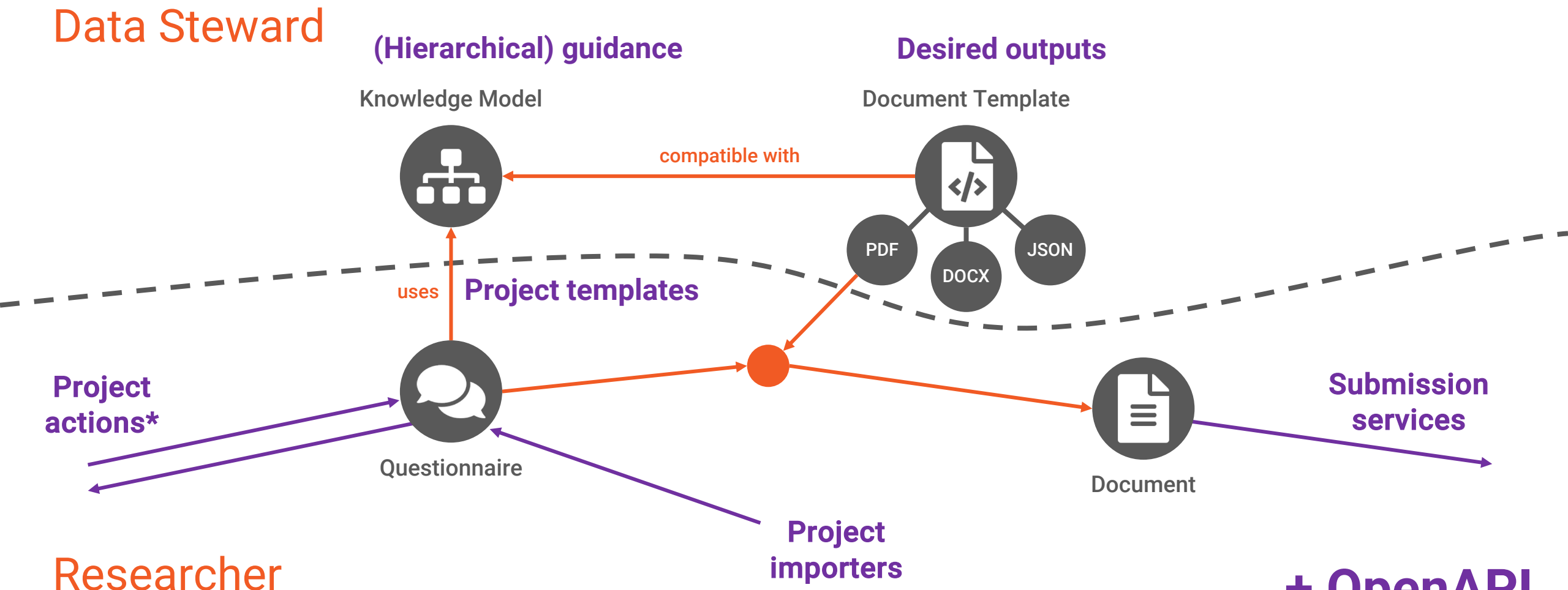
  It is a standardized format. This is a suitable format f
  We will have only a small amount of data stored in th

DSW: Knowledge Models

Knowledge Model



Versatility of DSW



+ OpenAPI

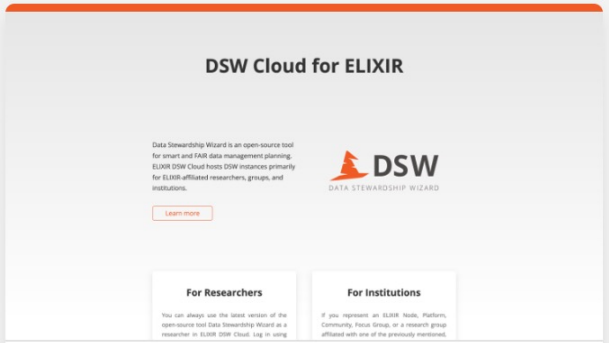
Sustainability and Commercial Side

DSW

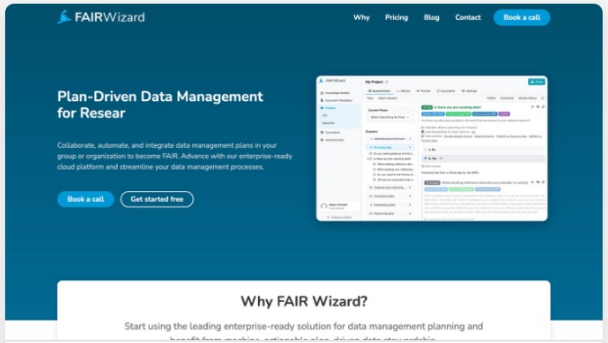
- **Projects support development** of features, standardization, interoperability, collaborations, pilotting, etc.
- Projects do not provide **long-term sustainability and operations** for having services running and then institutions need to find a way how to run them
  - Some institutions run on their own (on-premise, self-hosted) = need people and infrastructure, can adjust according to local needs and have data locally
  - Some institutions do not have own capacities / resources and rather want to get solution, support, SLAs, trainings, etc. (and pay for it)

From DSW to providers & FAIR Wizard

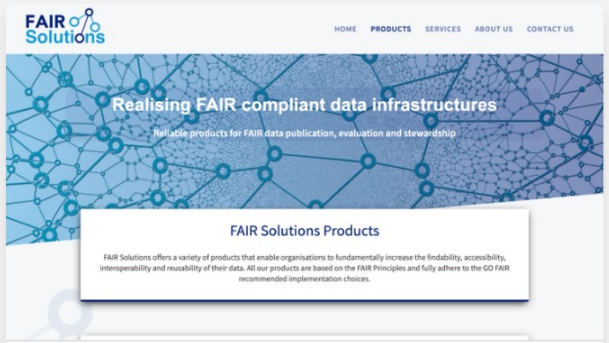
- DSW is open-source project, anyone can host it locally (using the documentation) and use it
- There are also providers / hostings of pure DSW or solutions that use DSW as a component



DSW Cloud for ELIXIR



FAIR Wizard



FAIR Solutions

DSW

DATA STEWARDSHIP WIZARD

Part IV: Trends & Future

Trends and Observations: Institutions

- First “extreme“: **RDM responsibilities assigned to librarians** who have no time / resources
  - They seek to provide some quick solution (existing tools, guides, etc.)
  - They do not have time to adjust content for local needs and help researchers
  - Researchers need to struggle and hope to get help elsewhere
- Second “extreme“: **dedicated department for RDM** with own data stewards and IT staff
  - Customize tools to provide seamless integrations within environment
  - Content tailored to local needs, sometimes even domain-specific parts
  - Providing reviews and also “CRM” for researchers to assist them with planning

Trends and Observations: Funders

- There are **no significant changes** visible, providing templates and guidance documnets
- Sometimes providing trainings or workshops (esp. if domain-specific or national)
- No committments to become also FAIR or machine-actionable (e.g. maDMP extensions)
- Some obviously **not reviewing (properly)** DMPs but just require them
- Some **limit number of pages** (e.g. NIH, 2 pages of DMPs)
- Review process would be currently very complicated (domain-specific, long text, very technical, legal stuff, etc.) but **essential for improving quality**

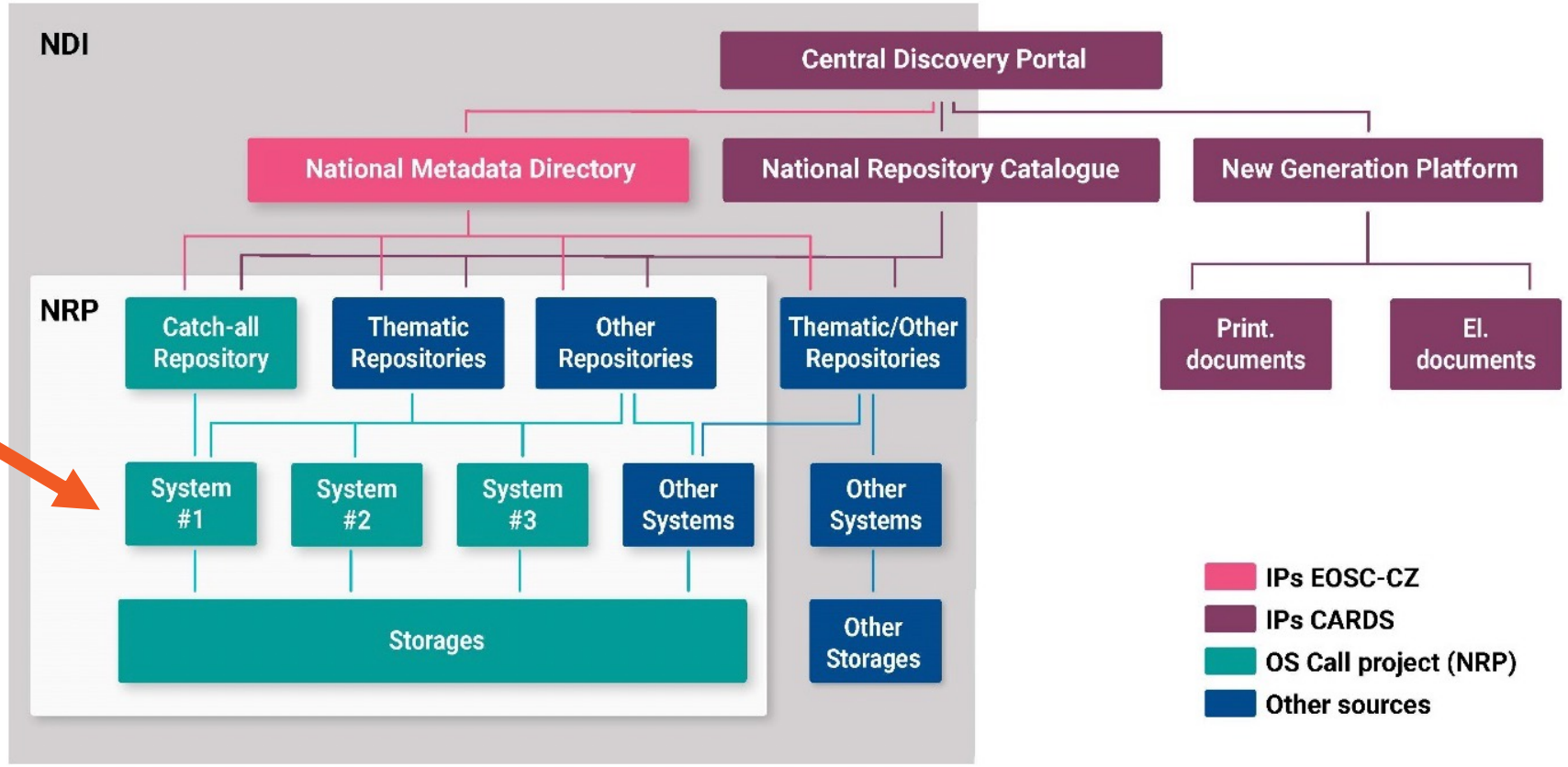
Trends and Observations: Tools

- No huge changes recently…
  - DSW has several UI/UX, bug-fixes, and preparation for “**Project Actions**”
  - DMP Tool brought **DMP ID** some time ago, with strong collaboration with DataCite
  - Argos has **Blueprints** to move away from funder-centric model (similar to DSW)
  - DMPonline is mainly maintaining the codebase (e.g. dependency updates)
  - RDMO is maintaining and improving but no significant features (focusing on content?)
  - DAMAP is relatively new, heavily based on **DCS and integrations** (TU Wien, T. Miksa)
- … but that may change soon

What is on the Horizon(s)?

- Many ongoing efforts on various levels (national, research infrastructures, European)
- DSW is part of several such efforts (Czech, ELIXIR, EOSC)

Czech Horizon: National Repository Platform



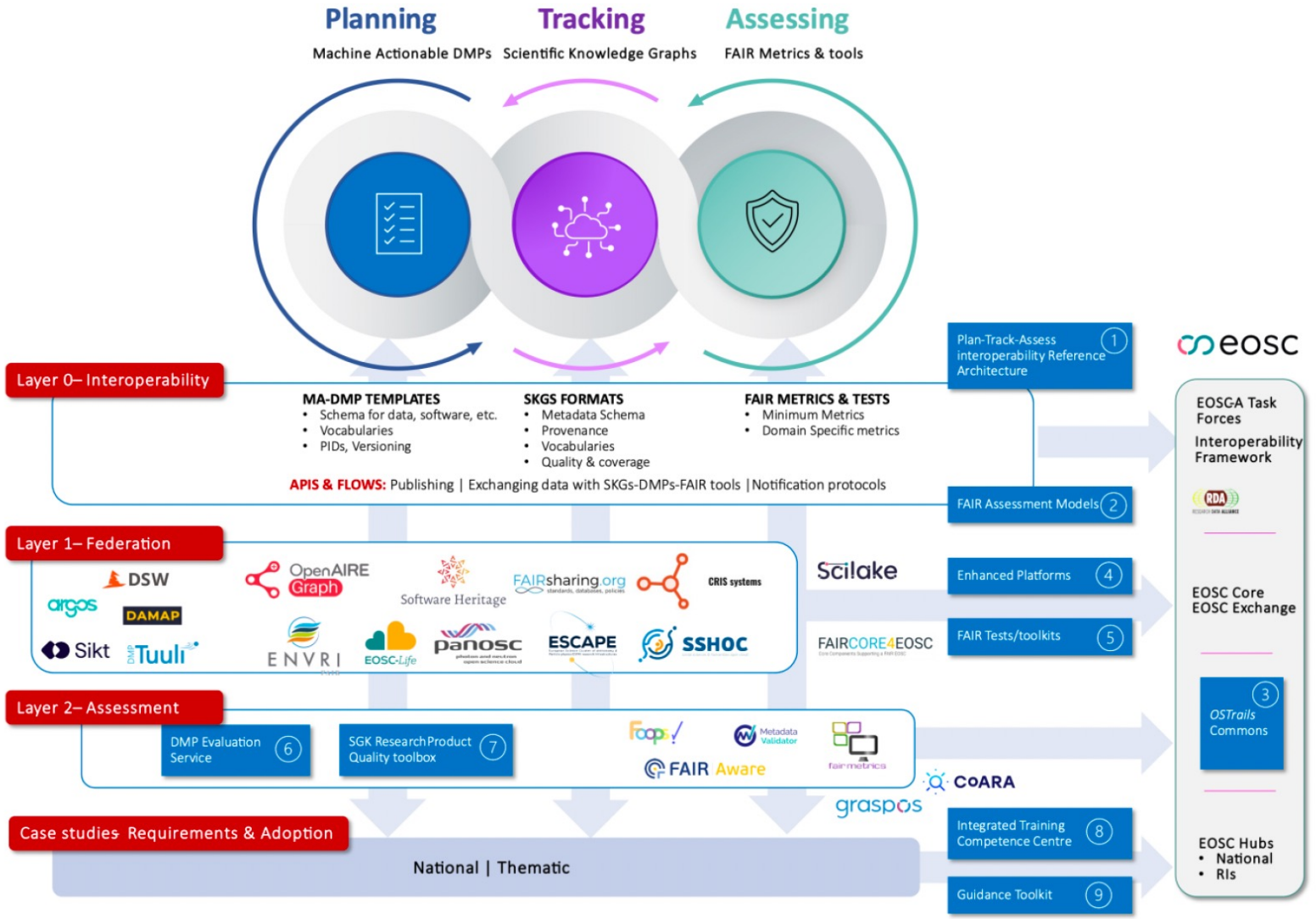
- Czech National Repository Platform (2024-28) project brings national services for RDM

Czech Horizon: National Repository Platform

- **Localized content** for Czech national environment, national maDMP extension
- **New features, services, and enhancements** (planned):
  - Knowledge model and document template **translations** (i18n)
  - DSW instance bootstrapping/sandboxing (e.g. for trainings)
  - Selecting items as answers in a questionnaire, references in chapters, cross-references
  - **DMP visualization** and improved progress indications
  - **Assigning PID to DMPs** (submissions)
  - Improving UI/UX and **accessibility (WCAG)**

Horizon Europe (EOSC): OSTrails

DSW



ostrails

https://ostrails.eu

Horizon Europe (EOSC): OSTrails

DSW

- Devise **interoperability framework between DMPs, SKGs**, and **FAIR assessment tools**
- **Specifying APIs and data structures** for exchanging information automatically
- Heavily **based on DCS / maDMPs** in terms of DMP part
- **DMP Evaluation Service** for assessing qualitied of DMPs and interoperable
- **OSTrails Commons for easier adoption** during project but also after / outside it
- Promoting (not only technical) **collaboration between tools** in project (~45) and beyond

Global Horizon: More _MPs

- Researchers need to manage **data** properly; they should have a **data management plan**!
- Researchers need to manage **research software** properly; …
- Researchers need to manage **ML models** properly; …
- Researchers need to manage **physical samples** properly; …
- … how many management plans should researchers prepare? or a single “**ROMP”**?
- There are emerging efforts on such things, SMPs are getting momentum (incl. maSMPs)
  - ELIXIR-STEERS (DSW), RDMO, and several templates (e.g. NWO)

Summary

DSW

- Data Management Plans are important; but **not necessarily as traditional documents**
- **Various challenges** around DMPs which can be addressed using **machine-actionability**
- Institutions and national efforts seems to go (slowly) in the right direction, funders not (yet)
- FAIR and other practices are **expanding to other areas** of research outputs/artifacts
- There are **many ongoing activities** and options, hopefully converging to ultimate goal of making (data) **management planning more efficient** for all stakeholders

DATA STEWARDSHIP WIZARD

Thank you!