



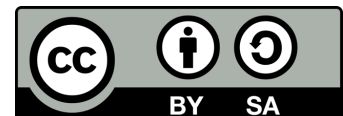
DMPs, maDMPs, and DSW: An Outlook

SND Network meeting in Gothenburg
Gothenburg, 24 April 2024

Marek Suchánek

marek.suchanek@ds-wizard.org

 0000-0001-7525-9218



Marek and his Hats

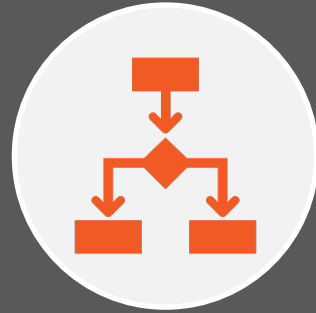


- Data Stewardship Wizard [2018]
 - Czech Technical University in Prague (ELIXIR LM, NRP, OSTrails)
 - Institute of Organic Chemistry and Biochemistry (ELIXIR-STEERS)
- FAIR Wizard [2023]
 - Codevance Solutions
- RDA DMP Common Standards WG [2022]
- Model-Driven Development and Normalized Systems [2017]





DMP
Challenges



maDMPs

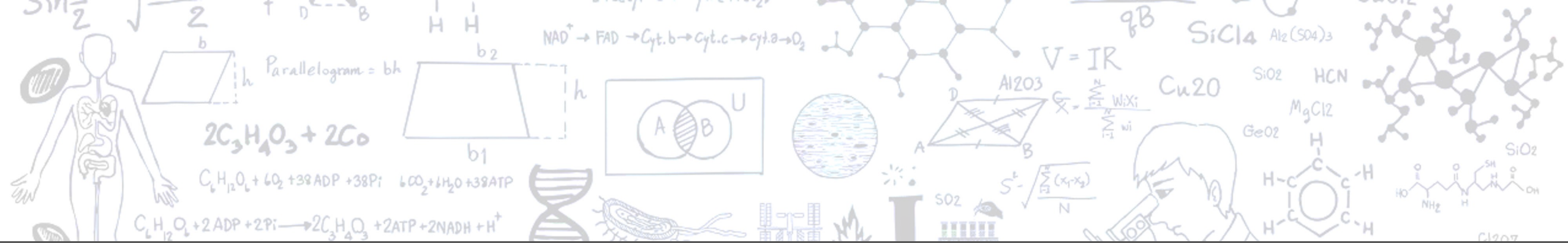


DSW
Approach

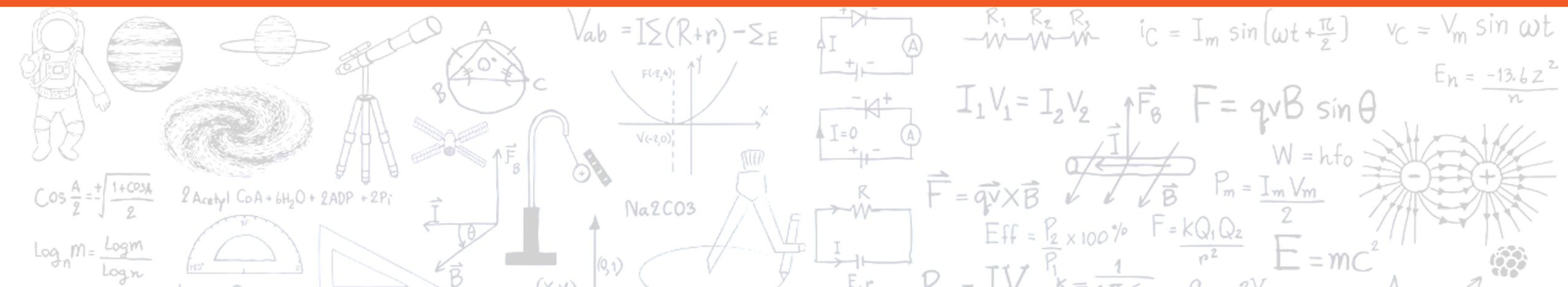


Trends and
Visions





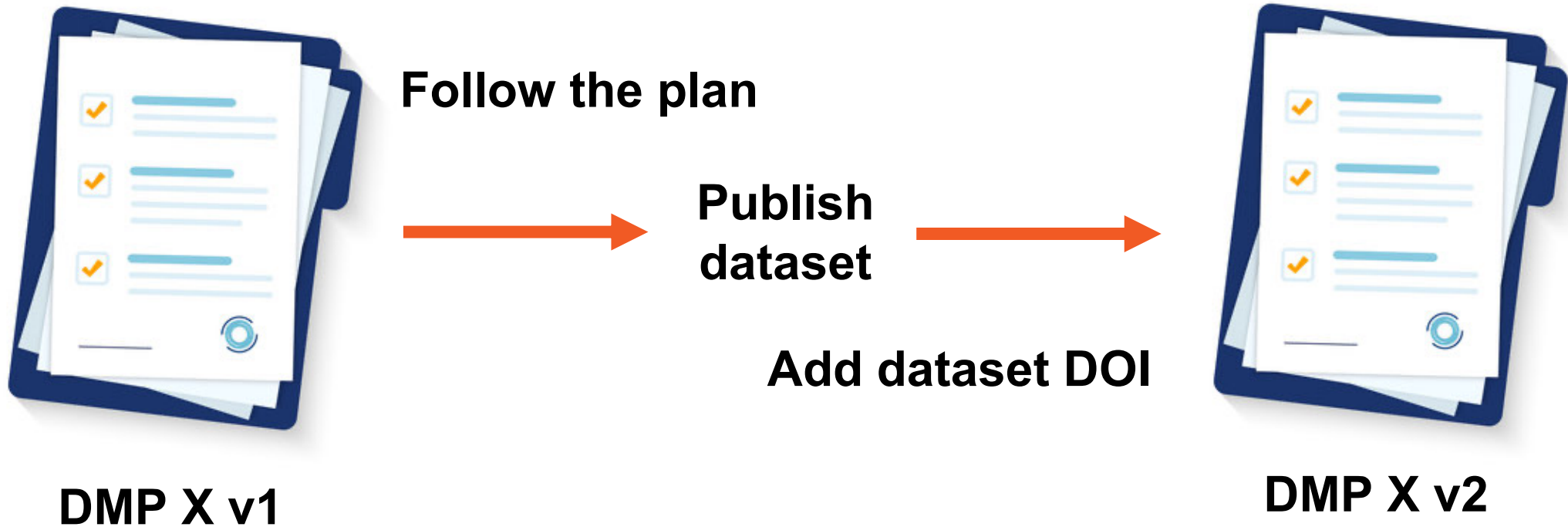
Part I: DMP Challenges



Challenge: Various Requirements



Challenge: Living Document, Versions

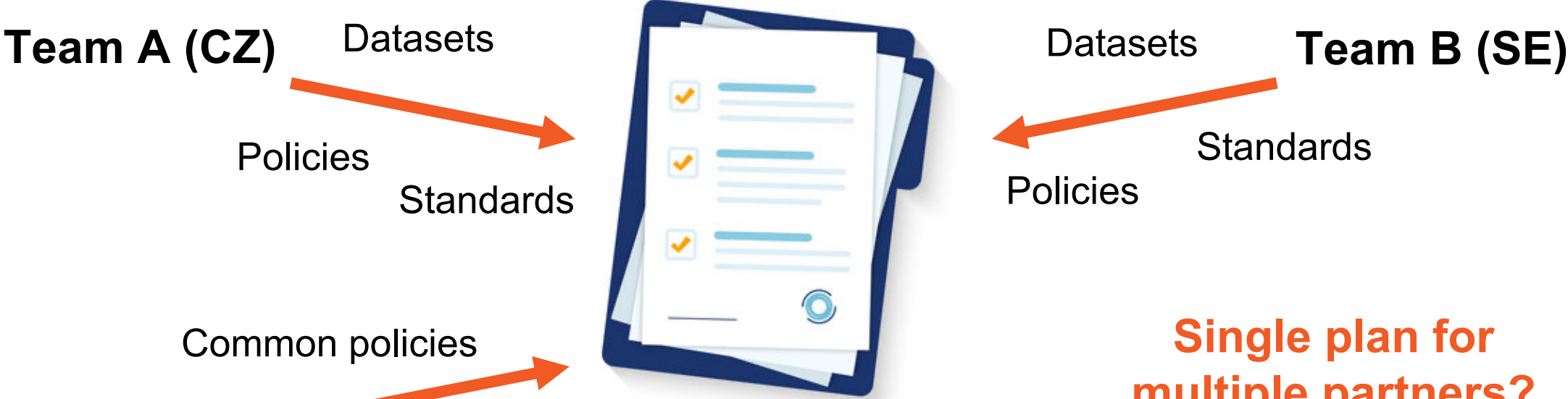


Tracking changes?

Automated updates?

Referring to DMP?

Challenge: Collaboration



Single plan for multiple partners?

Consistency?

Coordinated updates?

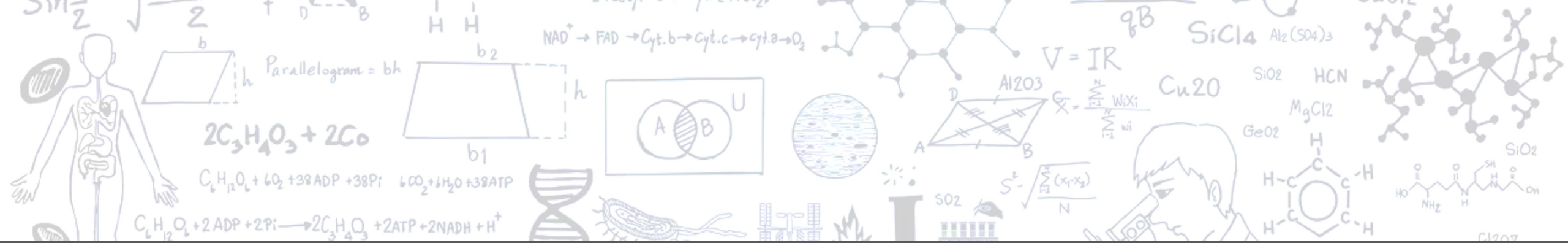
Challenge: Evaluating DMPs



Rapid feedback?

Unbiased response?

Full assessment?



DSW

DATA STEWARDSHIP WIZARD

Part II: Machine-Actionable



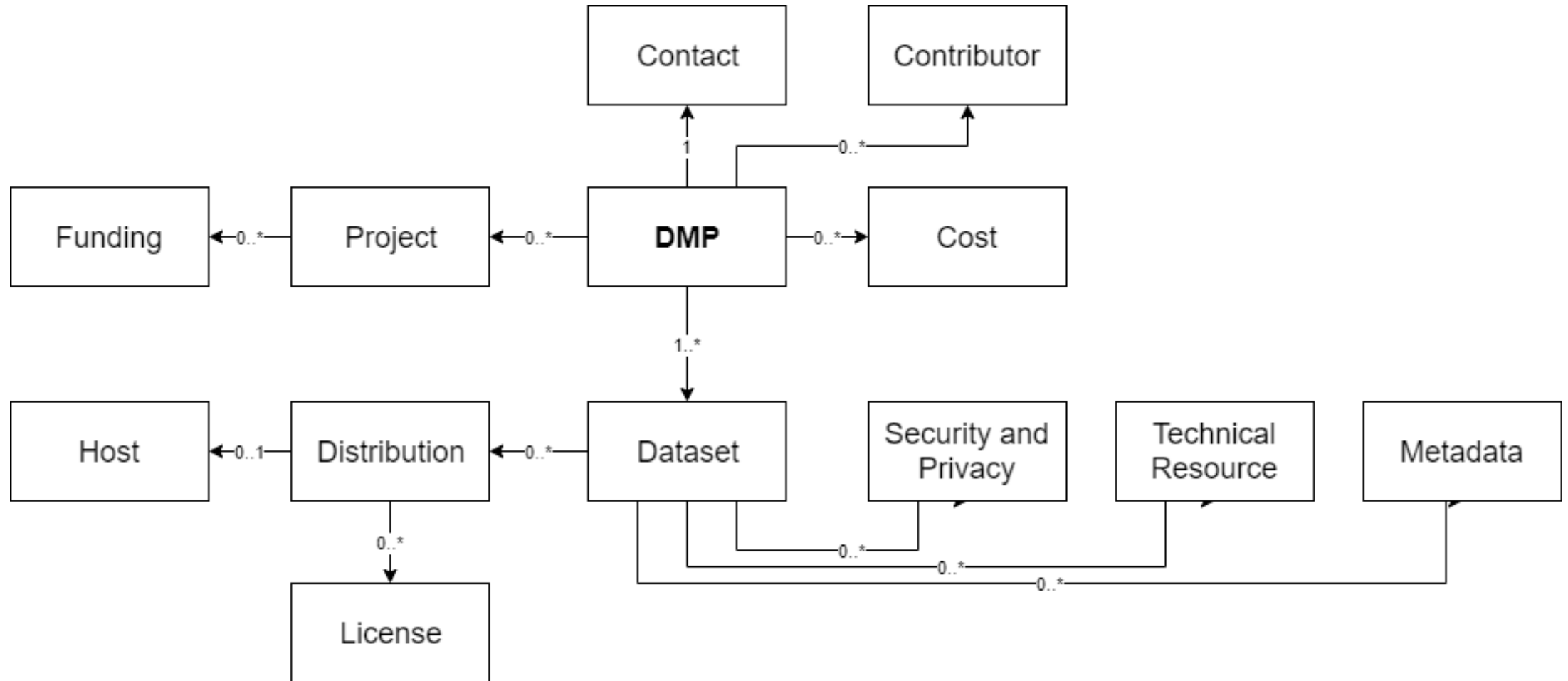
Machine-Actionable DMPs (maDMPs)



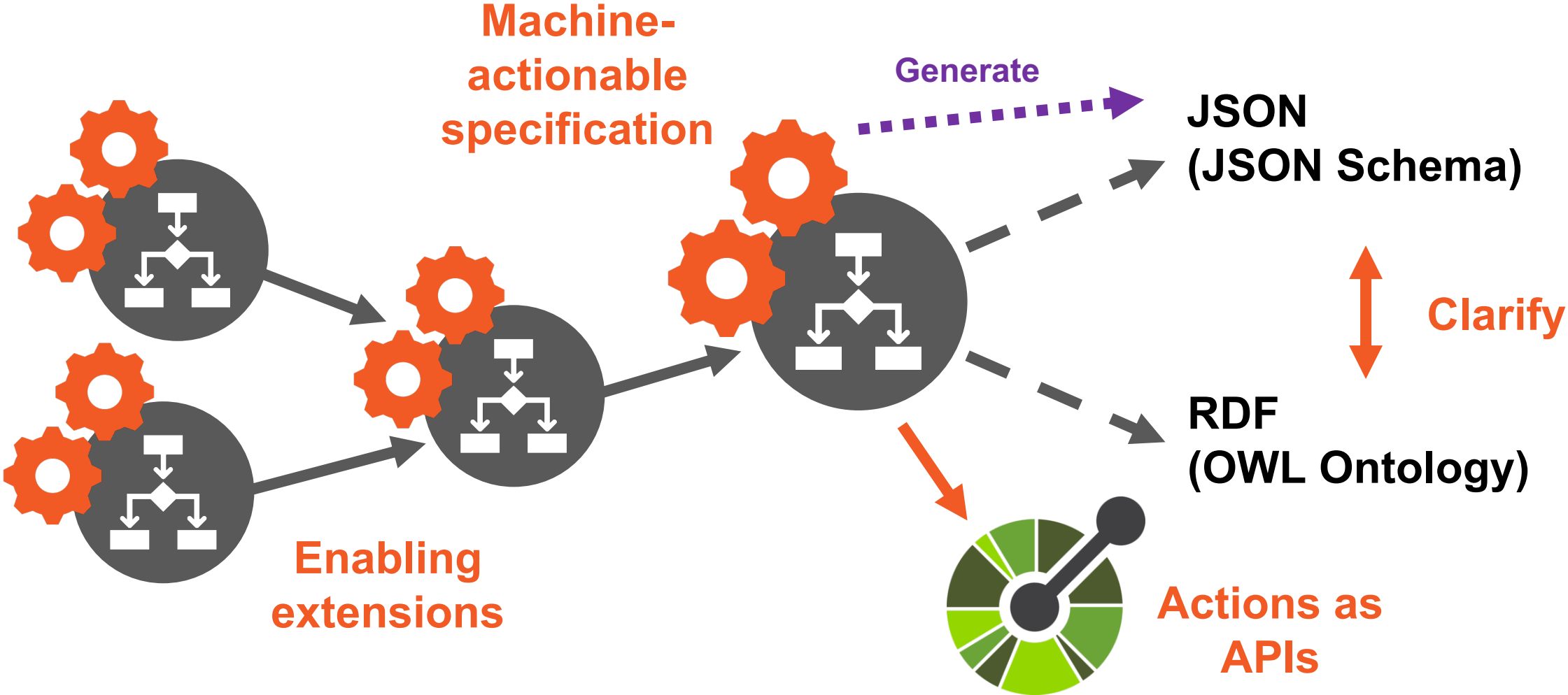
- Usually maDMP = DMP according to **RDA DMP Common Standard (DCS) for maDMPs**
- Started **>5 years ago (in 2019)**, pushed during hackathon 2020 (all main DMP tools)
- Covers **common properties** (the ultimate core) of DMPs
- Uses widely used ontologies/vocabularies such as DCAT, DCTerms, or DataCite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

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

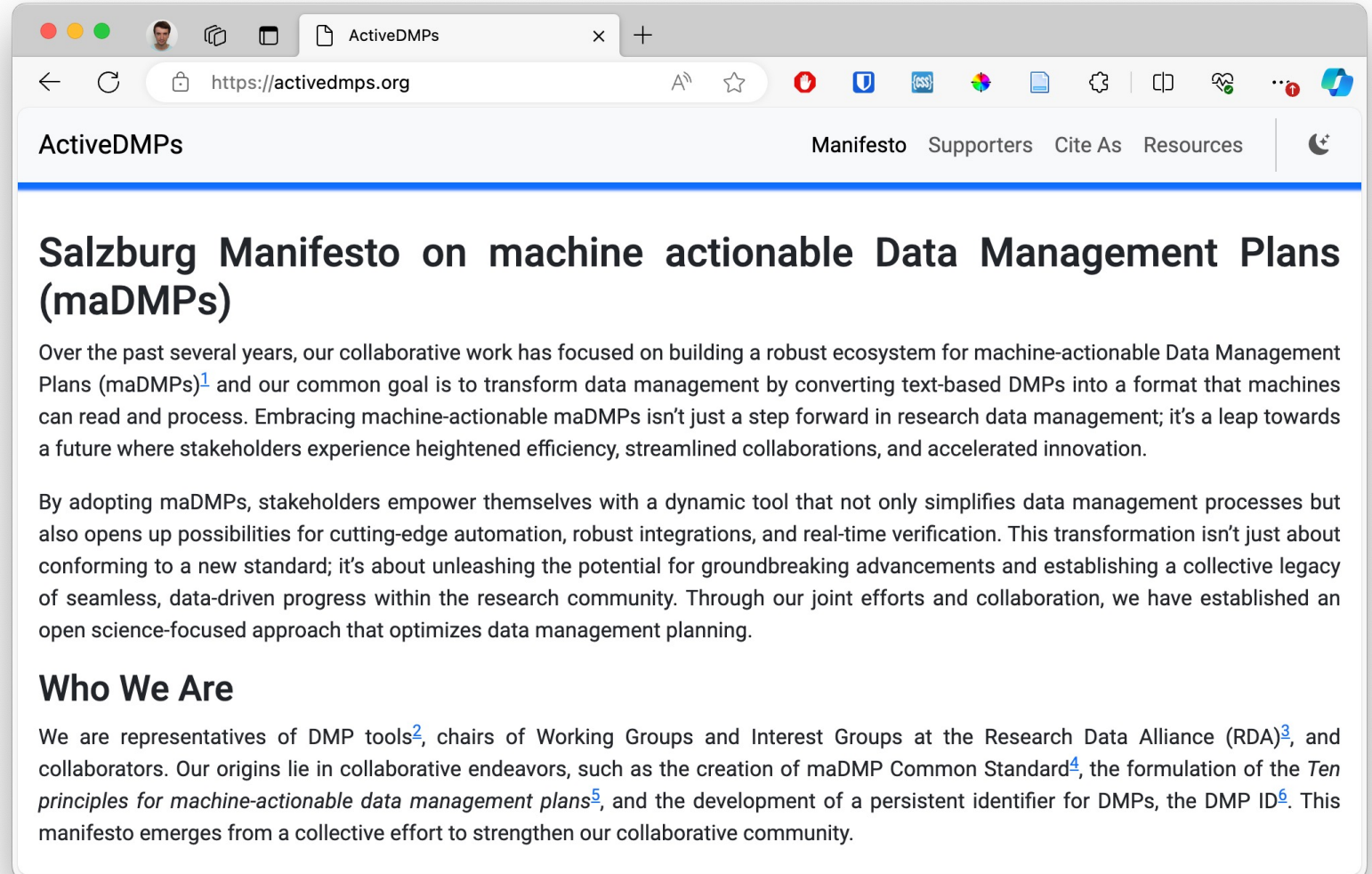
RDA DMP Common Standard for maDMPs



Plans for maDMPs



- During RDA Plenary 22
- Collaboration
- Community of practice
- Adopting and promoting
- All key DMP platforms
- You can also support it:
activeDMPs.org

A screenshot of a web browser displaying the homepage of activeDMPs.org. The browser's address bar shows the URL https://activedmps.org. The page title is 'ActiveDMPs' and the navigation menu includes 'Manifesto', 'Supporters', 'Cite As', and 'Resources'. The main heading is 'Salzburg Manifesto on machine actionable Data Management Plans (maDMPs)'. The text below the heading describes the collaborative work on building a robust ecosystem for machine-actionable Data Management Plans (maDMPs) and the goal of transforming text-based DMPs into a machine-readable format. It also mentions the adoption of maDMPs by stakeholders and the establishment of an open science-focused approach. The 'Who We Are' section identifies the representatives as DMP tools, chairs of Working Groups and Interest Groups at the Research Data Alliance (RDA), and collaborators, listing their origins in collaborative endeavors like the creation of the maDMP Common Standard, the formulation of the Ten principles for machine-actionable data management plans, and the development of the DMP ID.

ActiveDMPs

Manifesto Supporters Cite As Resources

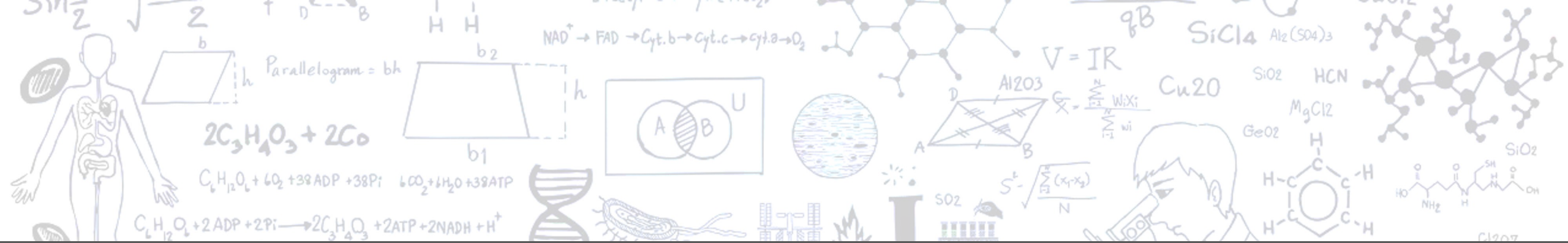
Salzburg Manifesto on machine actionable Data Management Plans (maDMPs)

Over the past several years, our collaborative work has focused on building a robust ecosystem for machine-actionable Data Management Plans (maDMPs)¹ and our common goal is to transform data management by converting text-based DMPs into a format that machines can read and process. Embracing machine-actionable maDMPs isn't just a step forward in research data management; it's a leap towards a future where stakeholders experience heightened efficiency, streamlined collaborations, and accelerated innovation.

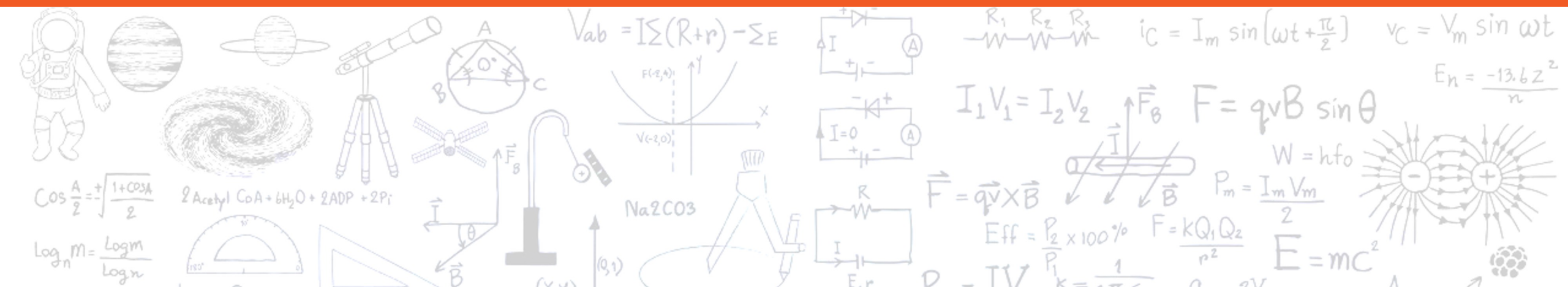
By adopting maDMPs, stakeholders empower themselves with a dynamic tool that not only simplifies data management processes but also opens up possibilities for cutting-edge automation, robust integrations, and real-time verification. This transformation isn't just about conforming to a new standard; it's about unleashing the potential for groundbreaking advancements and establishing a collective legacy of seamless, data-driven progress within the research community. Through our joint efforts and collaboration, we have established an open science-focused approach that optimizes data management planning.

Who We Are

We are representatives of DMP tools², chairs of Working Groups and Interest Groups at the Research Data Alliance (RDA)³, and collaborators. Our origins lie in collaborative endeavors, such as the creation of maDMP Common Standard⁴, the formulation of the *Ten principles for machine-actionable data management plans*⁵, and the development of a persistent identifier for DMPs, the DMP ID⁶. This manifesto emerges from a collective effort to strengthen our collaborative community.



Part III: Approach of DSW

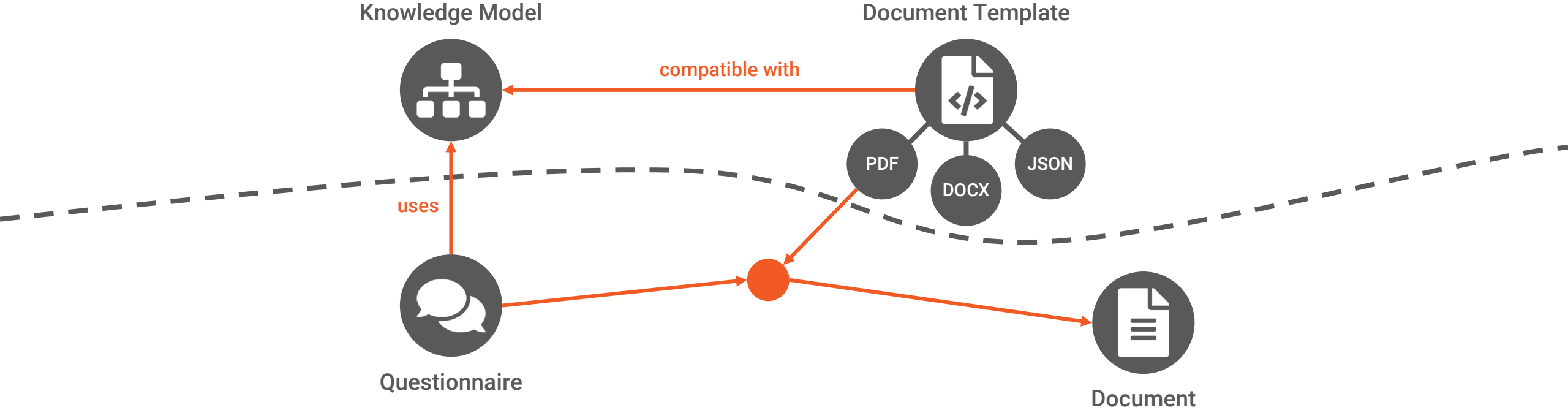


Data Stewardship Wizard (DSW)



- **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

Data Stewards



Researchers

DSW: Questionnaire



Questionnaire

Description

Answers

Who answered
and when

Advice for answer

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.

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

Answered in less than 5 seconds by Albert Einstein.

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

Title

References

(FAIR) metrics

DSW: Questionnaire



Questionnaire

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



Search

Get reply with URI

Open link with details

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.EtYkWo>

Clear answer

FAIRsharing.org search through all content

Banana Breeding Tracker Database (BBTbase)

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 extraction, culturing, and herding. This database also provides...

DSW: Questionnaire



Questionnaire

Before S

	Albert Einstein	Owner	✖
	Nikola Tesla	Editor	✖
	Isaac Newton	Viewer	✖

rch Pro

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**

the rese

ork are c

Before S

ew Documents Settings

Comments 2 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...

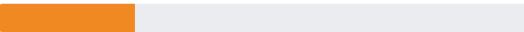







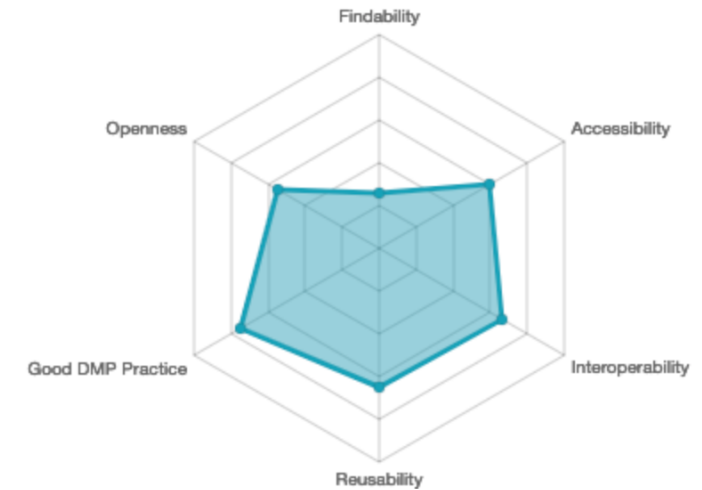
Questionnaire

Summary Report

Answered (current phase): 15/38 

Answered: 41/101 

Metric	Measure	Progress
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 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 Management Plan

Science Europe Example

Contact person: **Jana Freeman** (jana.freeman@ds-wiz.cz)
0000-0000-0000-0001
[Czech Technical University in Prague](http://www.ctu.cz)

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 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 ICP-MS.

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 and will be used for the analysis of As and Se. The equipment is very well described and known.
- **ICP-MS**
This dataset will be collected by experts in the project and will be used for the analysis of As and Se. 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 the analysis of As and Se. We will have only a small amount of data stored in the project.

DSW: Knowledge Models



Knowledge Model

Common DSW Knowledge Model

Knowledge Model Phases Question Tags Preview Settings

... > Will this data set be published? > Yes > Specify where you will distrib... > What repository will this data ...

- No
 - Why not?
 - It is raw data that never needs to be rep
 - This data set only led to unpublishable r
 - It is intermediate data that is easily repr
 - There is no value, nobody will ever reus
 - It is too expensive
 - It was lost
 - There are other reasons
 - What other reasons?
- Yes
 - Specify where you will distribute this data t
 - What repository will this data be stored
 - A domain-specific repository
 - Our national repository
 - Our institutional repository
 - A special-purpose repository for the
 - FAIRSharing
 - Registry of Research data Repositorie
 - FAIRCookbook on depositing to a ger
 - Who will the data in this place be share
 - Open: The data will be shared with a
 - Shared: The data will be shared with
 - Closed: This is an archive, it is not m
 - Licenses under which this distribution o
 - Under what license will the data set l
 - They will be freely available for ar

Question

80a682bd Move Delete

Type

Options

⚠ Changing a question type will remove all answers.

Title

What repository will this data be stored in?

Text

Editor Preview

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/>

You can use Markdown and see the result in the preview tab.

Versatility of DSW



Data Steward

(Hierarchical) guidance

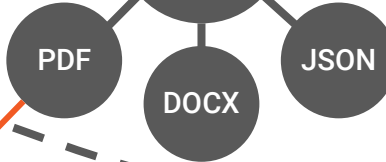
Desired outputs

Knowledge Model

Document Template



compatible with



uses Project templates

Project actions*



Questionnaire

Submission services

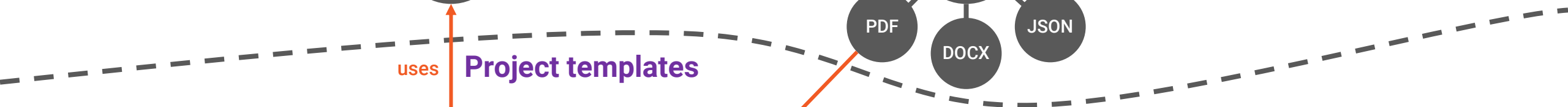


Document

Researcher

Project importers

+ OpenAPI

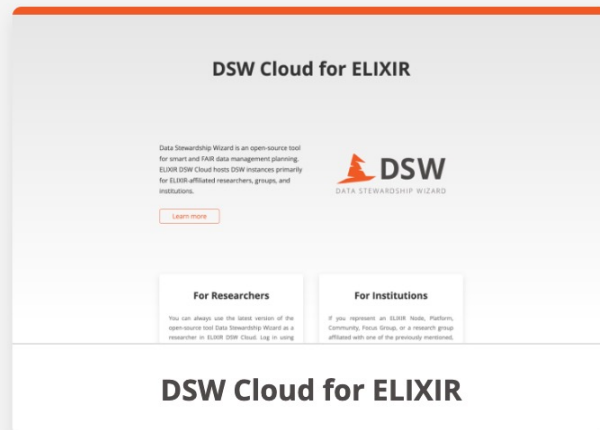


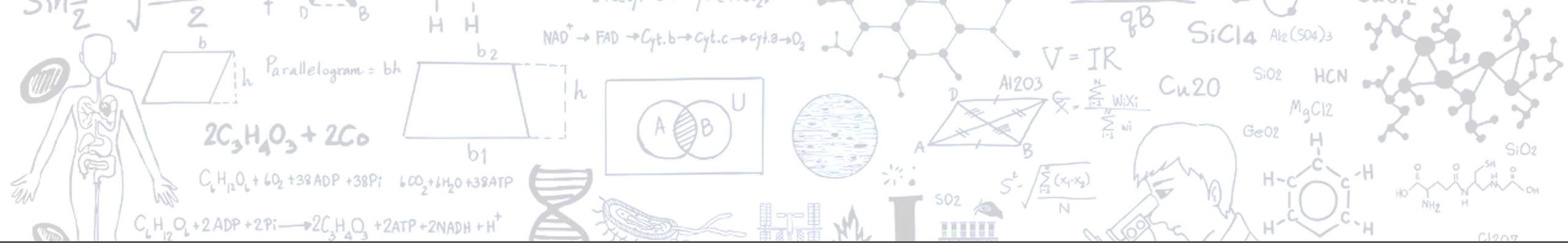
- **Projects support development** of features, standardization, interoperability, collaborations, piloting, etc.
- Projects do not provide **long-term sustainability and operations** for having services running and then institutions need to find a way how to run them
 - Some institutions run on their 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

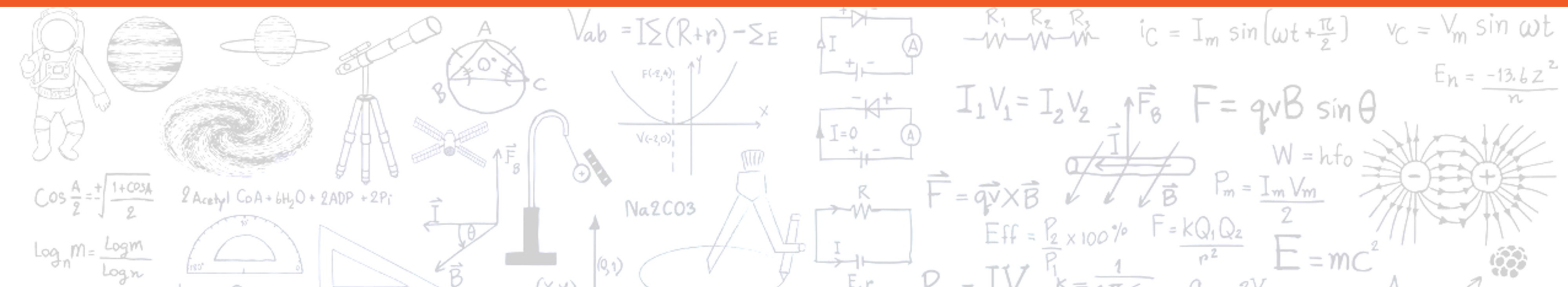


- DSW is open-source project, anyone can host it locally (using the documentation) and use it
- There are also providers / hostings of pure DSW or solutions that use DSW as a component





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 documents
- Sometimes providing trainings or workshops (esp. if domain-specific or national)
- No commitments 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



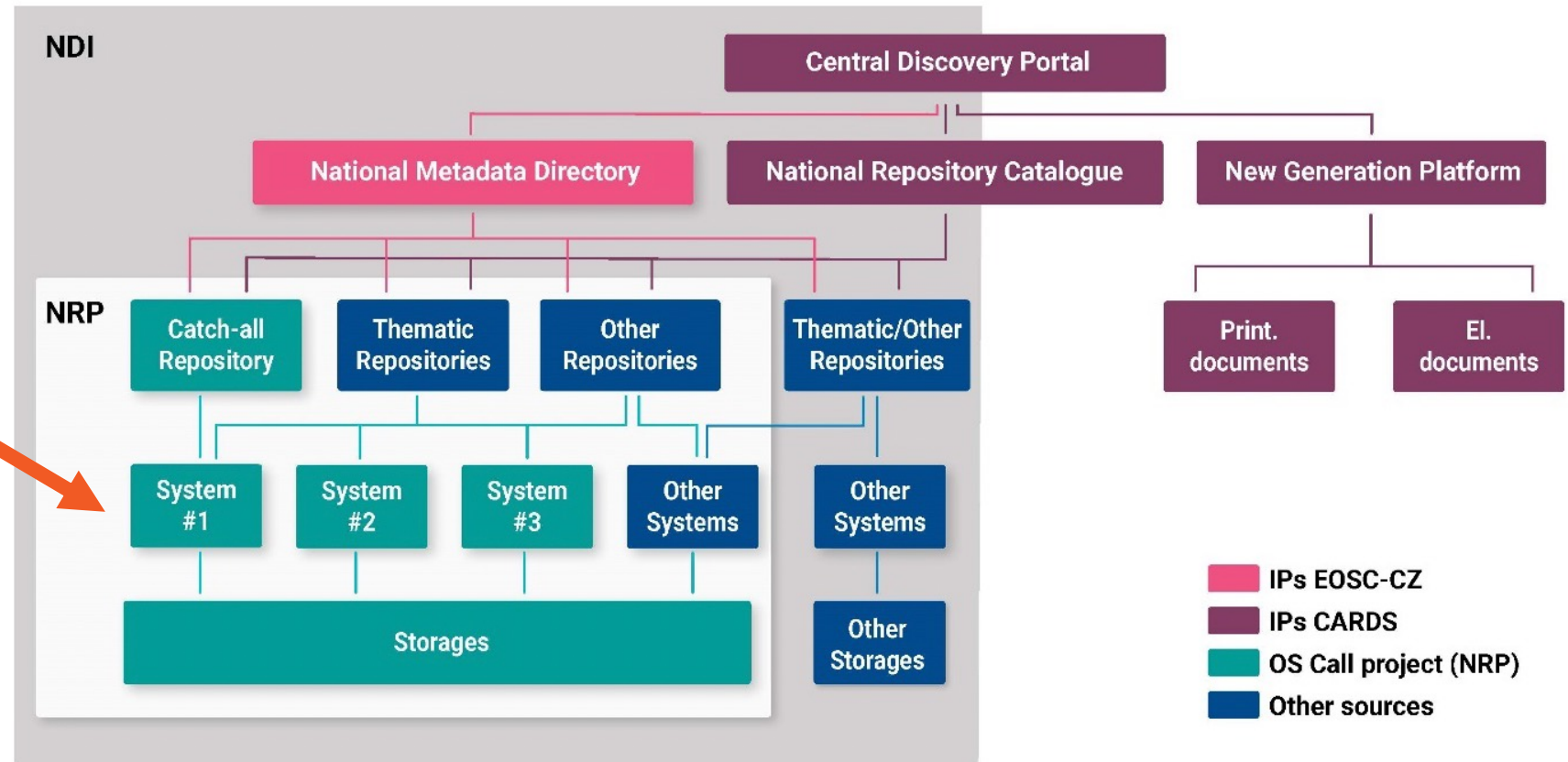
- No huge changes recently...
 - DSW has several UI/UX, bug-fixes, and preparation for “**Project Actions**”
 - DMP Tool brought **DMP ID** some time ago, with strong collaboration with DataCite
 - Argos has **Blueprints** to move away from funder-centric model (similar to DSW)
 - DMPonline is mainly 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



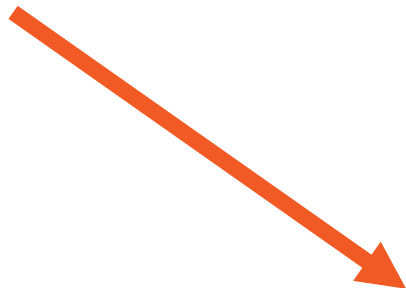
- Czech National Repository Platform (2024-28) project brings national services for RDM

Czech Horizon: National Repository Platform

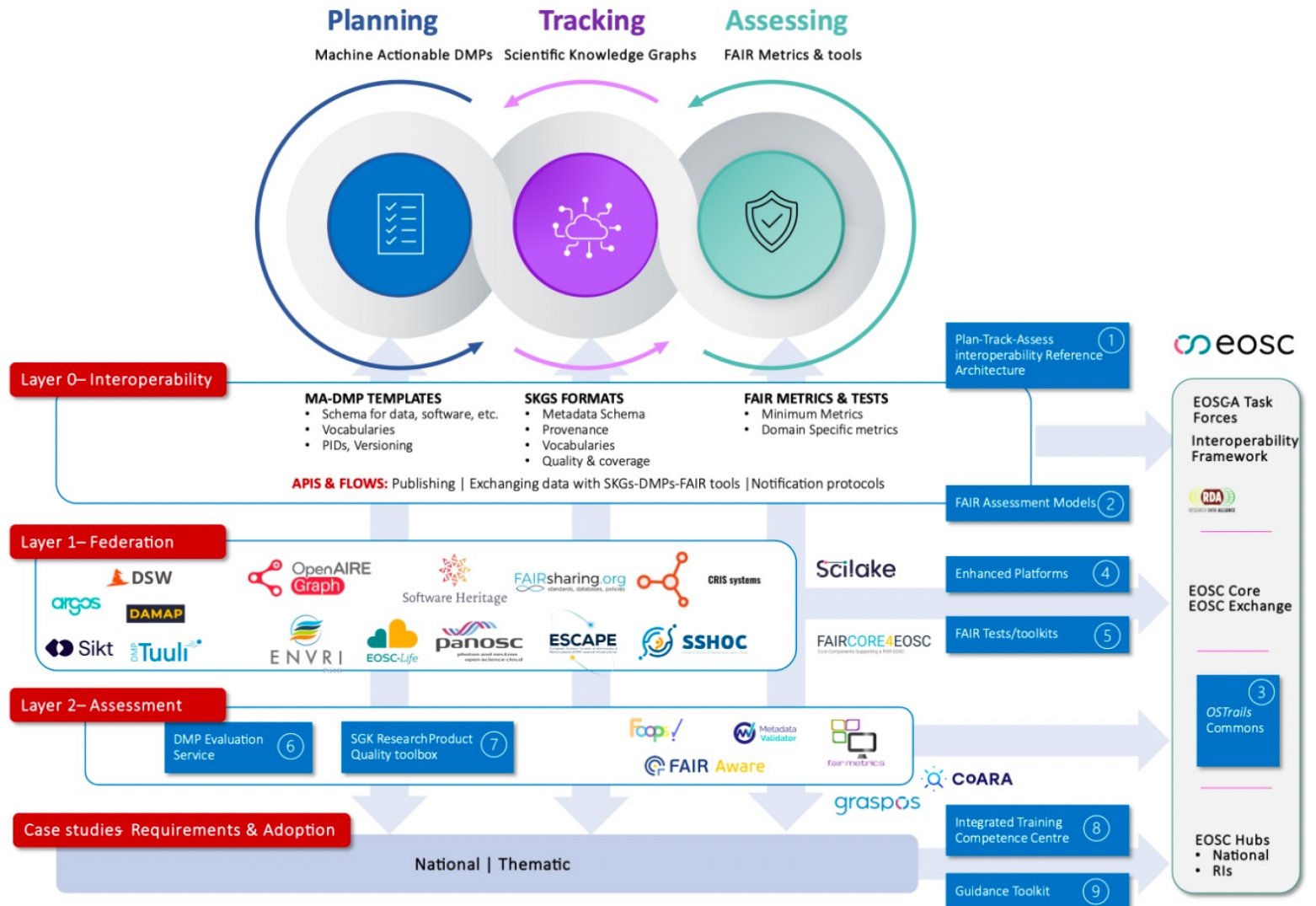


- **Localized content** for Czech national environment, national maDMP extension
- **New features, services, and enhancements** (planned):
 - Knowledge model and document template **translations** (i18n)
 - DSW instance bootstrapping/sandboxing (e.g. for trainings)
 - Selecting items as answers in a 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



<https://ostrails.eu>



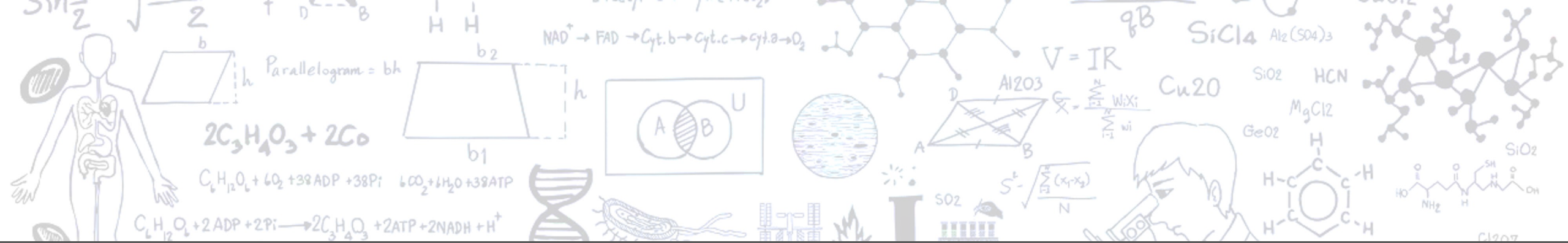
Horizon Europe (EOSC): OSTrails



- 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 quality 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

- 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)

- 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



DSW

DATA STEWARDSHIP WIZARD

Thank you!

