

Standards, Interoperability, Quality and Reuse

Tools and Resources from the FAIRagro Workbench

Daniel Bauer, Constantin Breß, Sven Gedicke, Jascha Jung, Susanne Lachmuth, Antonia Leidel, Mahdi Hedayat Mahmoudi, Daniel Martini, Markus Möller, Gabriel Schneider, Claus Weiland for the FAIRagro Consortium

Senckenberg Society for Nature Research, FIZ Karlsruhe, University of Bonn, KTBL e.V., ZALF, IPK Gatersleben, Julius Kühn-Institute, ZB MED

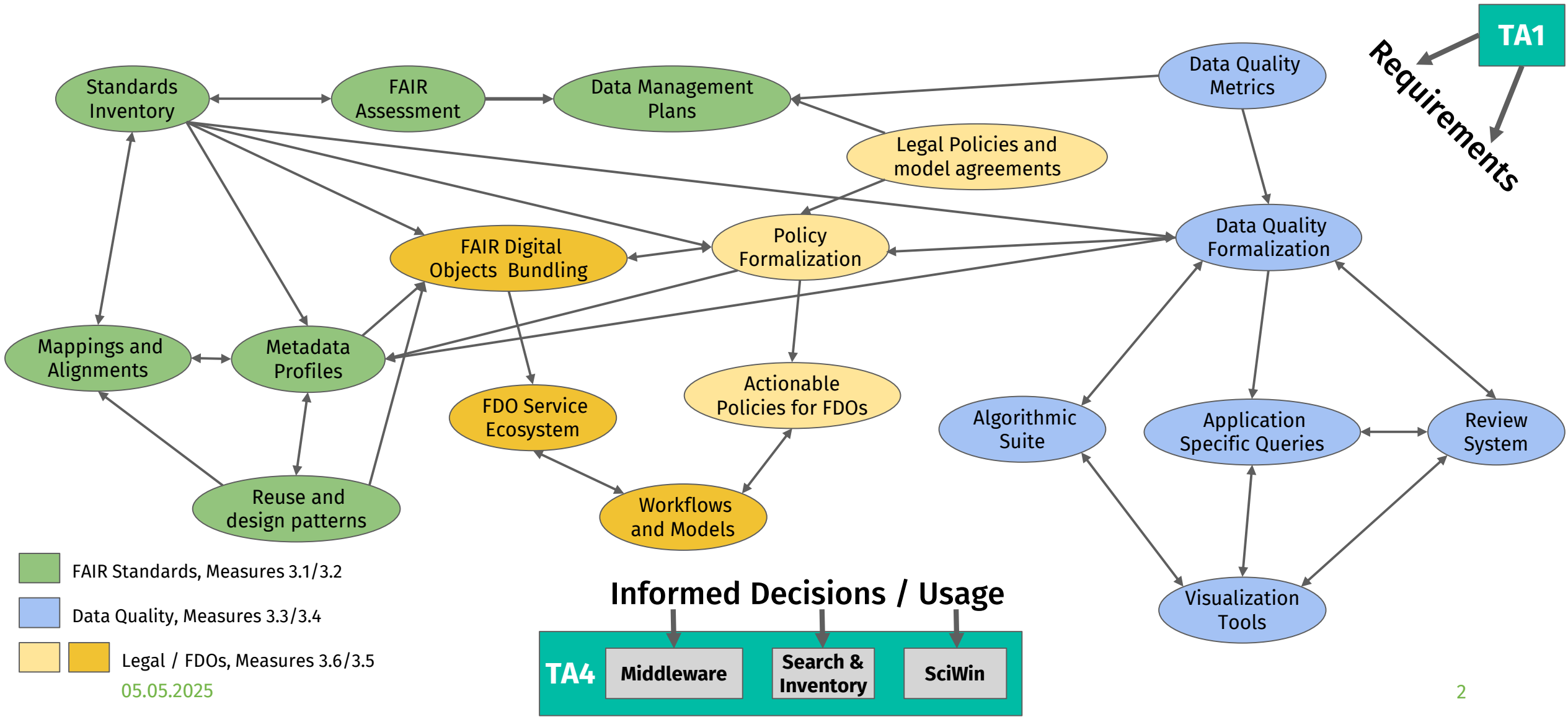
Funded by
DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

project number 501899475

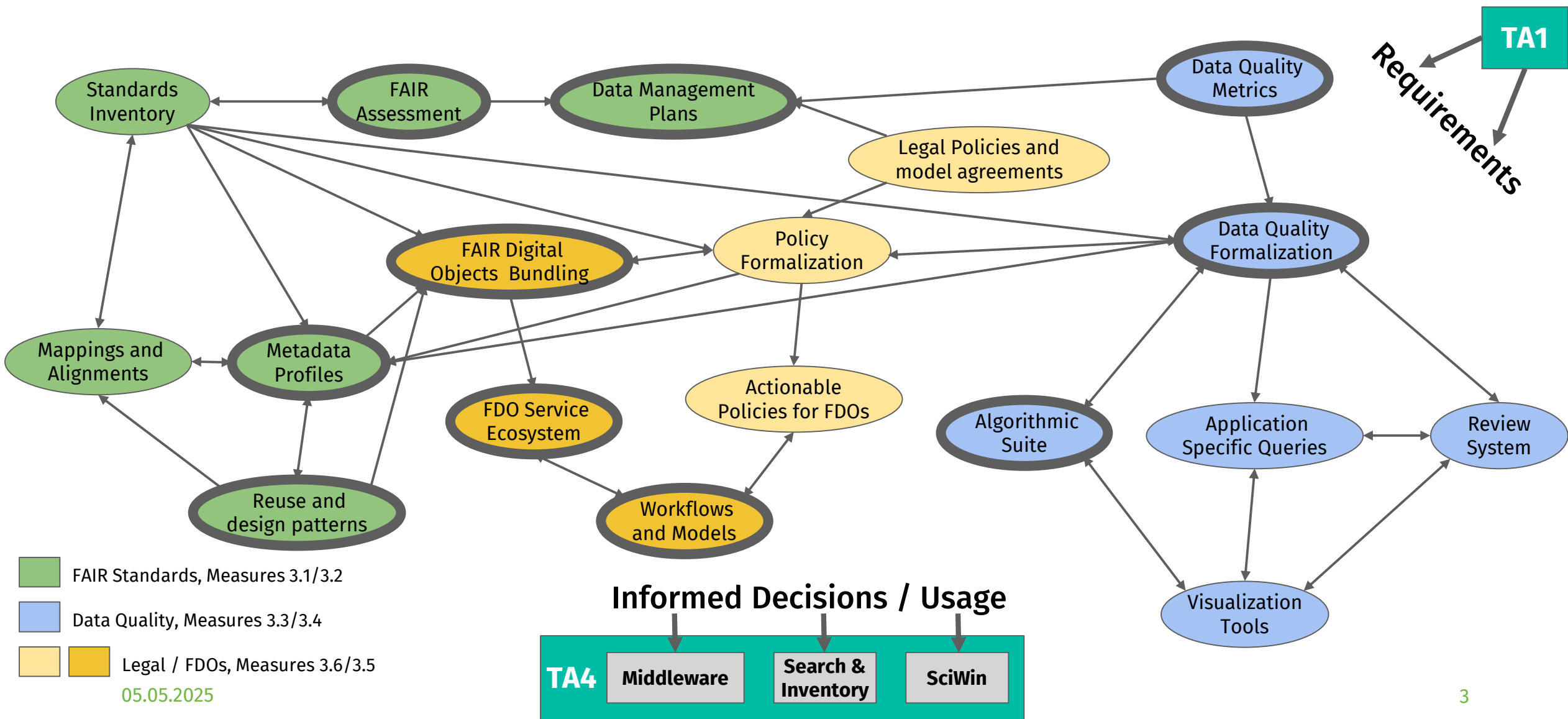
In cooperation with

nfdi

TA3 - Standardization, Interoperability and Quality



TA3 - Standardization, Interoperability and Quality



Measuring Rulers: FAIR Assessment


Recent Developments FAIR Assessment


- Functional:
 - Automated Assessment
 - Mapping into machine readable file formats using DQV
 - Storage Backend (RDF Knowledge Graph) and Dockerization
 - SPARQL Query Endpoint and Query Templates
- Current Development:
 - FAIRagro RDIs published data assessment
 - Unpublished Data Set Assessment
 - Github-Publication (https://github.com/fairagro/FAIR_evaluation_repository), CoRDI, password-protected web app for testing
- Planned:
 - Feedback using natural language


FAIR Assessment - Result Visualization and Download




FAIR Assessment - SPARQL Endpoint and Query Templates

 FAIR Evaluation

 SPARQL Explorer

 About

 [GitHub Repository](#)

FAIR_Metrics

Querying endpoint: http://fair_evaluation_repository-fuseki-1:3030/FAIR_Metrics/sparql

Choose a template query:

Findability Metrics

Enter your SPARQL query here:

PREFIX dqv: <http://www.w3.org/ns/dqv#>
 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
 PREFIX fairagro: <http://fairagro.net/ontology#>

Execute Query

	metric	label	definition
0	https://	Data Identifier Explicitly In Metadata	Metric to test if the metadata contains the unique identifier to the data. This is doi
1	https://	Data Identifier Persistence	Metric to test if the unique identifier of the data resource is likely to be persistent.
2	https://	Grounded Metadata	Tests whether a machine is able to find 'grounded' metadata, i.e., metadata terms
3	https://	Identifier Persistence	Metric to test if the unique identifier of the metadata resource is likely to be persis
4	https://	Metadata Identifier Explicitly In Metadata	Metric to test if the metadata contains the unique identifier to the metadata itself.
5	https://	Searchable in Major Search Engine	Tests whether a machine is able to discover the resource by search, using Microso

Inspection Report: Schema.org for harmonizing metadata between Research Data Infrastructures (RDIs)

- Goal: Improve the findability of datasets from FAIRagro RDIs and make them available in the FAIRagro Search Hub
- Schema.org use differs in implementation between RDIs

```
"author": [  
  {  
    "@type": "Organization",  
    "contactPoint": {  
      "@type": "ContactPoint",  
      "contactType": "technical support",  
      "email": "***@zalf.de",  
      "name": "***",  
      "url": ""  
    },  
    "name": "Leibniz Centre for Agricultural Landscape Research"  
  },  
]
```

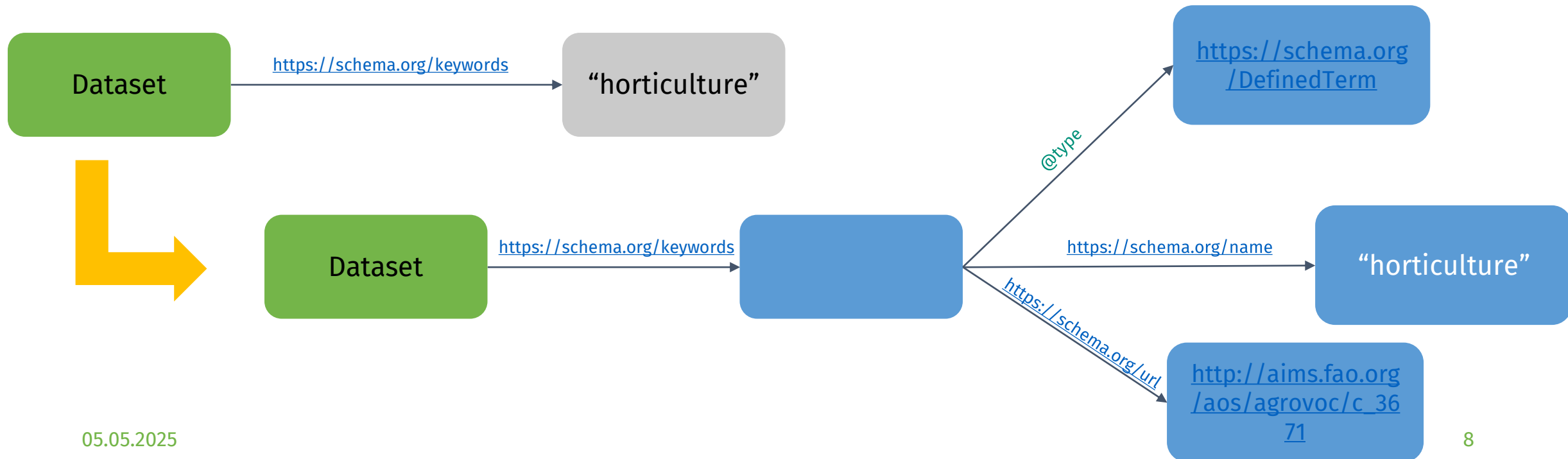
RDI 1

```
"creator": [  
  {  
    "@type": "Person",  
    "familyName": "***",  
    "givenName": "***",  
    "identifier": [  
      "https://orcid.org/****-****-****-****",  
      {  
        "@type": "PropertyValue",  
        "propertyID": "orcid",  
        "value": "****-****-****-****"  
      }  
    ]  
  }  
]
```

RDI 2

Schema.org: Optimizing what is already there

- A first set of optimization recommendations for first phase RDIs has been collected
 - Range between improved structuring of already available information and provision of additional, domain specific information



Agrischemas: Express what we need for searching

- Idea: Build upon [Schema.org](https://schema.org) and [Bioschemas](https://bioschemas.org) to express domain specific metadata required for searching
 - RDIs already offer “basic” Schema.org interface
 - Leveraging terminologies (e.g. [AGROVOC](https://agro.voc)) for semantic enrichment
- Describing core concepts such as crops, soils, diseases, pesticides etc.
- Recommendations for terminology resources for properties of these concepts
- First draft of the extension open for comments later this year

Diagnosis questions: Data Management Plans (DMPs)

- DMPs are created for planning and simplifying the **research data management** of a project
- Plans the **entire data life cycle** (e.g., from the origin of the data to publication and re-use scenarios)
- A **"living"** document → modify DMP during project

Why should you create DMPs?

- Consider important issues early on (e.g., storage, sensitive data, legal issues)
- DMPs are often required for funding

Our goal:

- Customize funder DMP templates for the FAIRagro use cases



„Data life cycle diagram“ by Elixir (https://rdmkit.elixir-europe.org/media_kit), licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>).

RDMO: FAIRagro templates

- Based on funding agency templates (DFG, Horizon Europe, etc.)
- Provide (English and German) descriptions with additional information and links to domain specific and FAIRagro resources
- Provide customized option sets for use cases
- Provide reusable text blocks for questions


[Back to project](#)
[Management](#)
[Test Instance](#)
[Feed](#)

Where will the data (including metadata, documentation and, if applicable, relevant code) be stored or archived after the end of the project?

Specify where and how your data will be stored. The repository should be as domain-specific as possible. If you are unsure where to find a suitable repository see [Where to find a fitting repository?](#) in the [FAIRagro Toolbox](#).

Domain-specific repositories could include:

- [OpenAgrar](#) for agricultural science documents and data
- [BonaRes](#) for agricultural, soil and landscape ecological datasets
- [PANGAEA](#) for earth and environmental sciences
- [e!DAL-PGP](#) for plant genomics and phenomics research data
- [Environmental Data Initiative \(EDI\)](#) for environmental data
- [European Soil Data Centre \(ESDAC\)](#) for soil-related data

Alternatively, [OpenDOAR](#) offers a list of open access repositories, e.g., [Zenodo](#)

- ☐ OpenAgrar
- ☐ BonaRes
- ☐ PANGAEA
- ☐ e!DAL-PGP
- ☐ Environmental Data Initiative (EDI)
- ☐ European Soil Data Centre (ESDAC)
- ☐ Zenodo
- ☐ Archive service of the institution
- ☐ Cross-disciplinary archive services such as RADAR:
- ☐ Other:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why.

A default answer has been provided for the text field. However, it can be edited by clicking into the text field and modifying the text as needed.

Yes, the metadata will be provided with a public domain license (CC0) in accordance with the funding agreement and made accessible. This enables unrestricted reuse of the metadata by the scientific community and the public. Default

Collaboration & future plans

- Using a RDMO instance hosted by the DMP4NFDI basic service
 - Involvement in NFDI cross consortia working groups and events
 - Alignment of template developments with DMP and NFDI communities
 - Involvement in training on DMPs
-
- Survey planned to gather more feedback from the use-cases on the current state of the FAIRagro templates



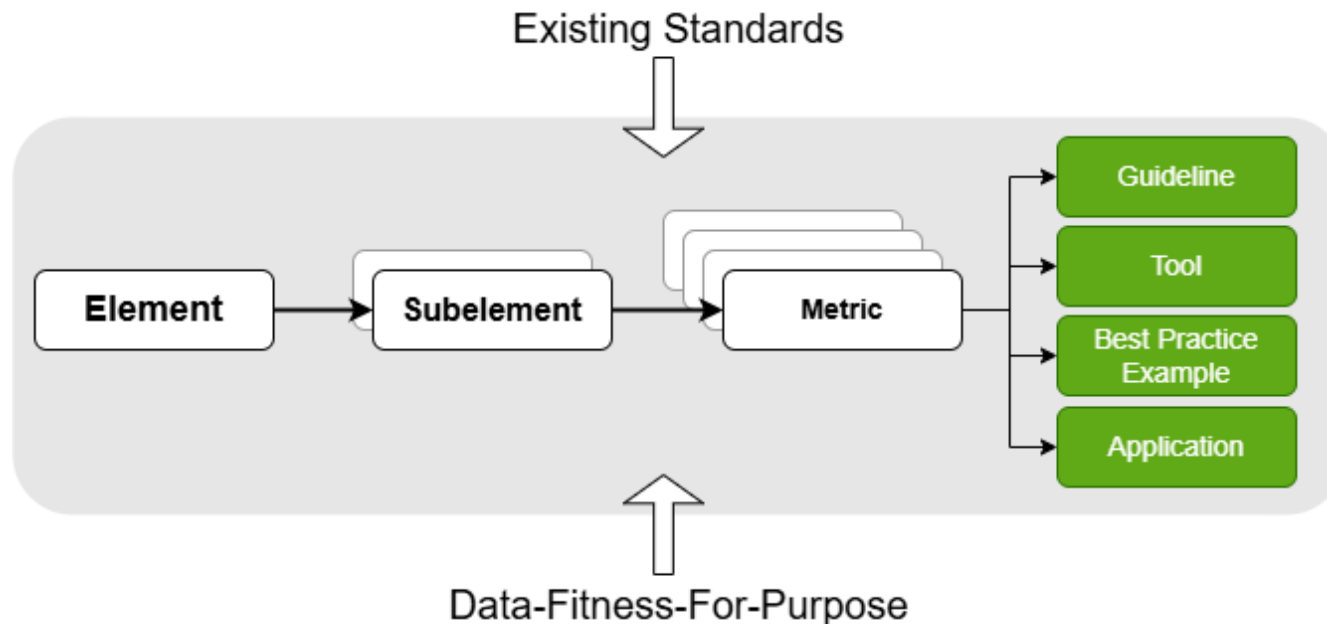
Image by [Mohamed_hassan](#), licensed under pixabay [Content License](#).

Data Quality (DQ) and Data-Fitness-For-Purpose (DFFP)

The accuracy of data and data sources is highly associated with trust and with having confidence that the **quality of data** is sufficient to serve as evidence base for critical decision making.

Lokers et al. (2016) <https://doi.org/10.1016/j.envsoft.2016.07.017>

Degree to which a set of inherent characteristics of data fulfils the requirements [of a producer and potential user]. ISO19157-1, 2023. Geographic information: Data quality. Part 1: General requirements. Geneva, Switzerland.

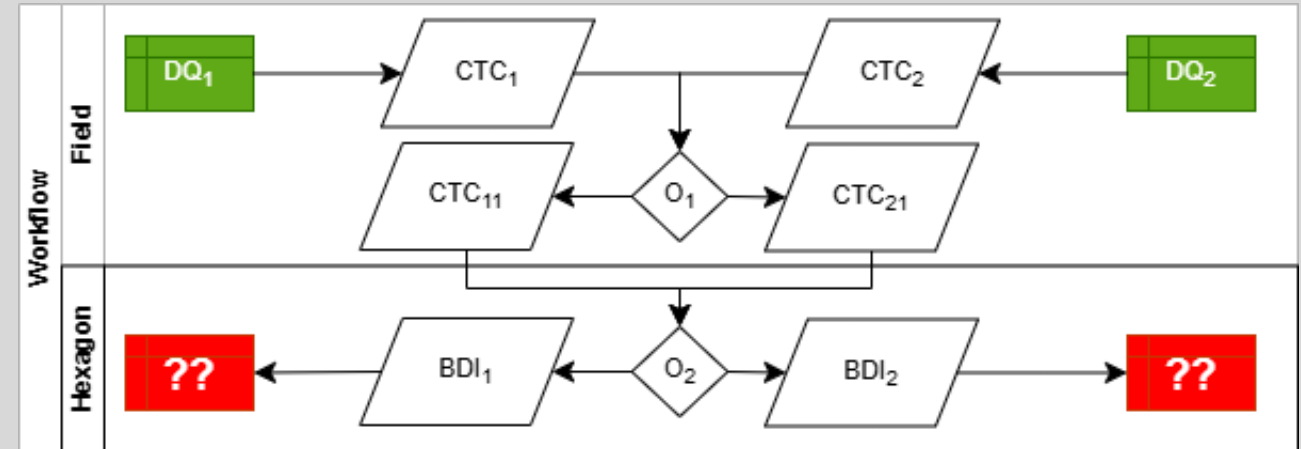
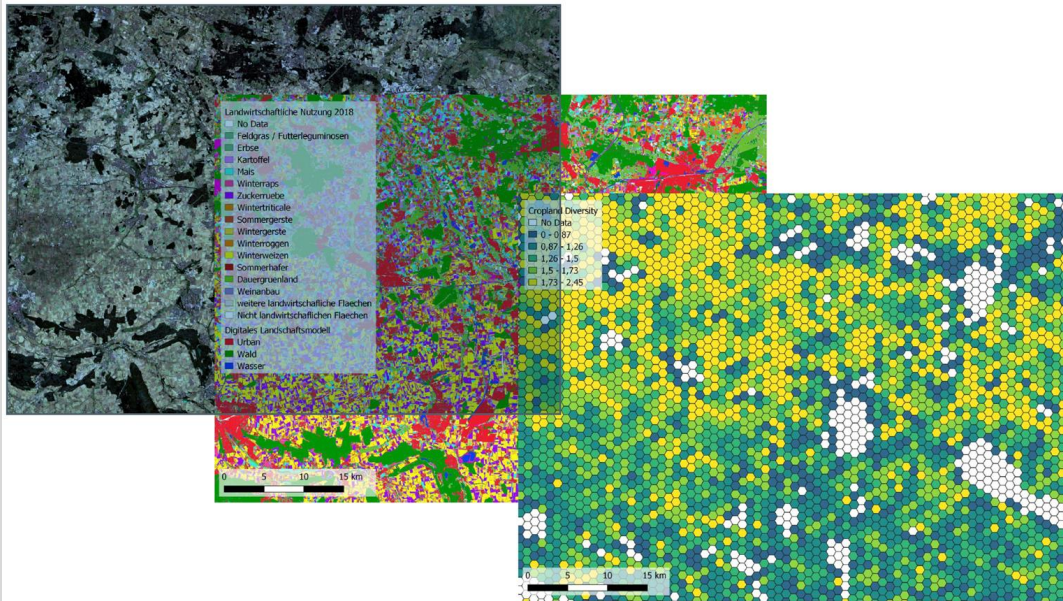


An use case-based evaluation procedure

Data Quality (DQ) and Data-Fitness-For-Purpose (DFFP)

A use case-based evaluation procedure

Biodiversity Indicator calculation workflow



Germany-wide Crop Type Classification (CTC)

Blickensdörfer et al. (2022) <https://doi.org/10.1016/j.rse.2021.112831>

Preidl et al. (2020) <https://doi.org/https://doi.org/10.1016/j.rse.2020.111673>

Data Quality (DQ) and Data-Fitness-For-Purpose (DFFP)

An use case-based evaluation procedure

Biodiversity Indicator calculation workflow

Established and standardised thematic accuracy metrics provided by the CTC producers were insufficient to detect local and regional inaccuracies as well as attributing Biodiversity Indicator calculation DQ information



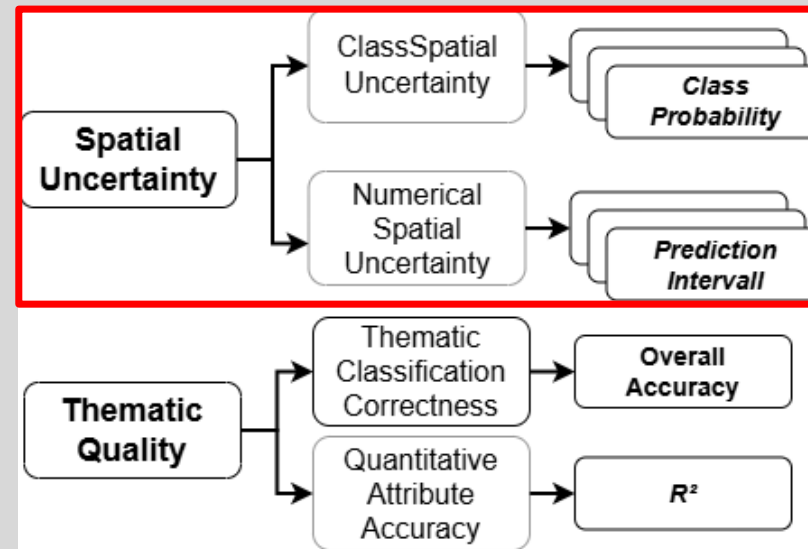
BUT: Local classification uncertainty metrics are by-products of ML-algorithms but are not typically provided by data producers.



Derivation of additional DQ metrics by data producers characterizing local uncertainty

Uhlott, J. et al. (2024)

<https://doi.org/10.5281/zenodo.14259254>



FAIRagro DFFP extension

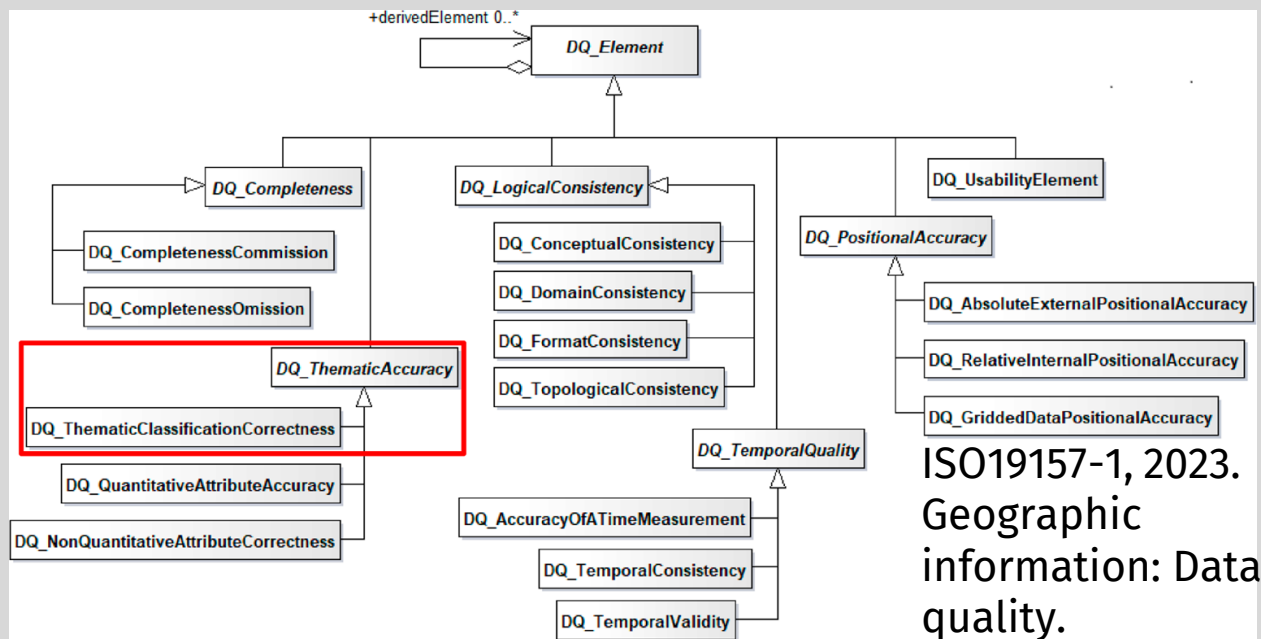
ISO19157-1, 2023.
Geographic information: Data quality.

Data Quality (DQ) and Data-Fitness-For-Purpose (DFFP)

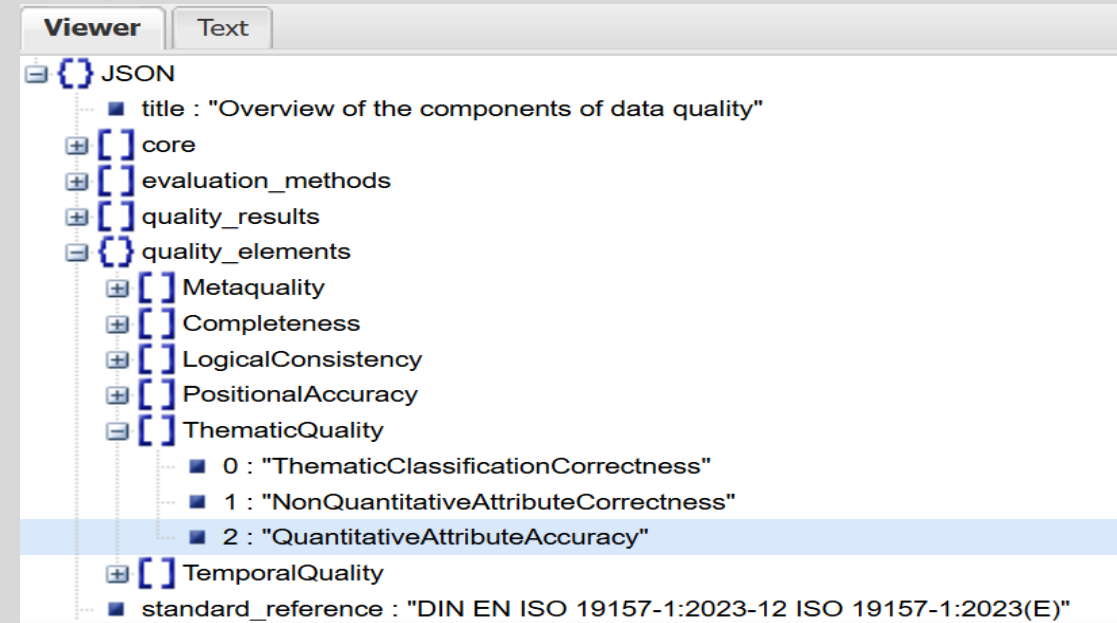
Next steps

Automation: Deriving machine-readable representations using LLM

Structured extraction of data quality information from unstructured sources (e.g., ISO standards)



ISO19157-1, 2023.
Geographic
information: Data
quality.



Data Quality tools in FAIRagro

DQ-Kit/DQ Toolbox



Data Quality tools in FAIRagro



Enhance your data quality assurance with DQ-Kit.

Simply upload your CSV table and let DQ-Kit provide comprehensive statistical insights into your scientific data. DQ-Kit performs automated statistical summaries, quality assessments, and alerts you to potential anomalies that could affect the accuracy and reliability of your data.

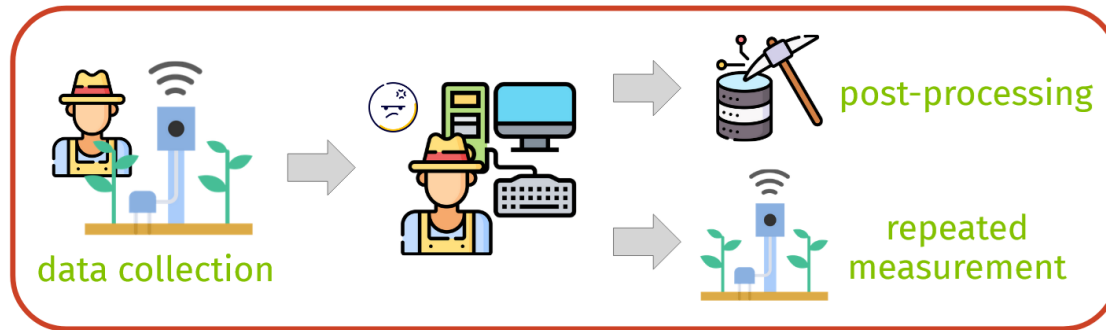
DQ-Kit helps data authors refine their datasets before publication, enhancing data quality and scientific impact. It also enables data reusers to efficiently evaluate and select datasets that meet their specific research needs.

<https://dqkit.bonares.de/>

Lachmuth, S., Dönmez, C., Hoffmann, C., Specka, X., Svoboda, N., Helming, K., 2025. Facilitating Effective Reuse of Soil Research Data: The BonaRes Repository. European J Soil Science 76, e70103.
<https://doi.org/10.1111/ejss.70103>

Data Quality tools in FAIRagro

DQ Toolbox



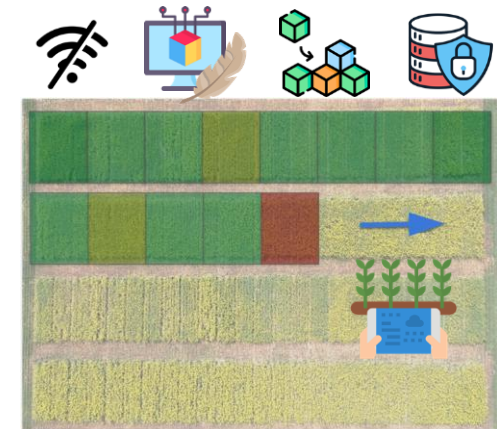
classical workflow



our approach

idea

- address data quality early in the data lifecycle
- integrate quality assessment into data collection process
- lower barrier to uploading quality information with the data



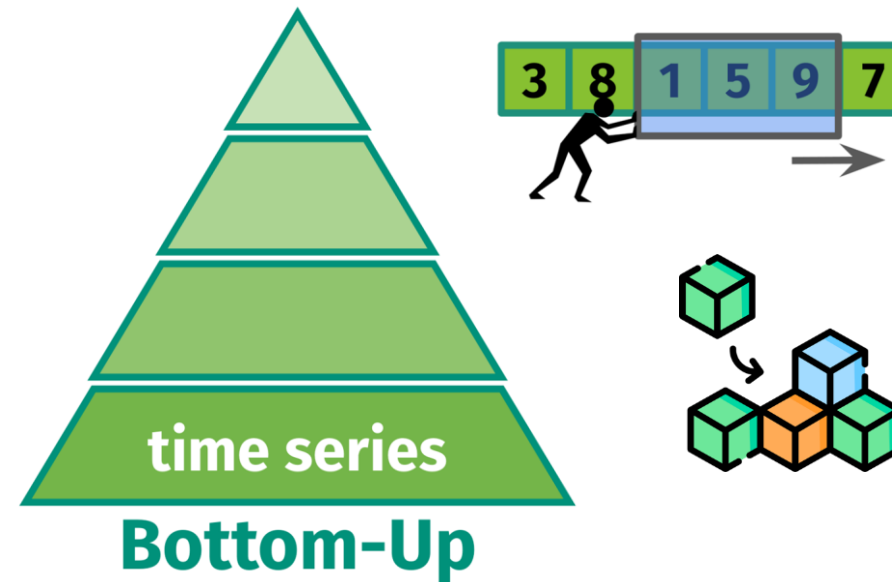
Icons designed by Freepik from Flaticon.

Data Quality tools in FAIRagro

DQ Toolbox

current status

- focus on **time series data**
- memory-efficient **sliding window algorithms**
 - summary statistics
 - outlier analysis
 - clustering
- **modular design**

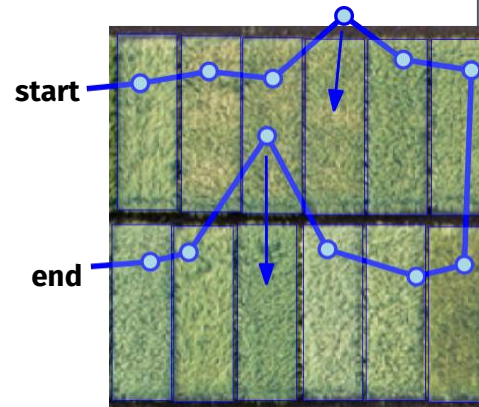


Data Quality tools in FAIRagro

DQ Toolbox

current status

- focus on **time series data**
- memory-efficient **sliding window algorithms**
 - summary statistics
 - outlier analysis
 - clustering
- **modular design**



development foci

1) plausibility

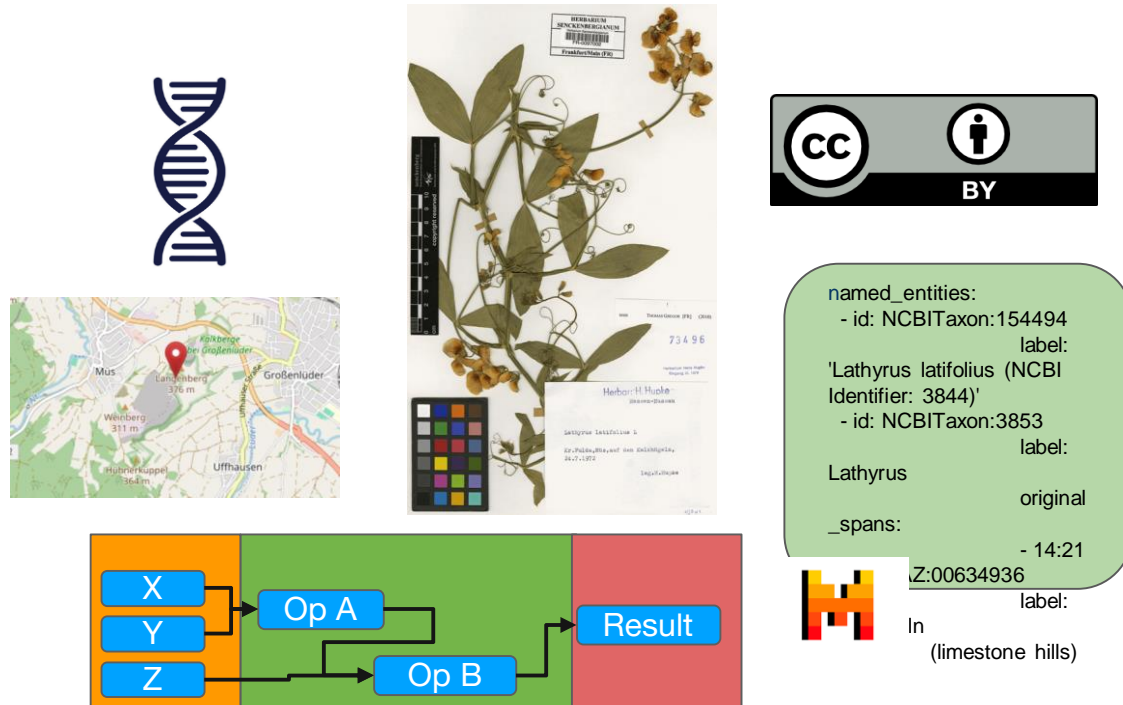
- use statistics to provide **plausibility probability**
- visual and computational **comparison with existing datasets**

2) measurement-to-plot assignments

- **matching algorithms** that integrate GPS data, plot layout, and user trajectory
- quantification of **assignment accuracy**
- intuitive **visual representations**

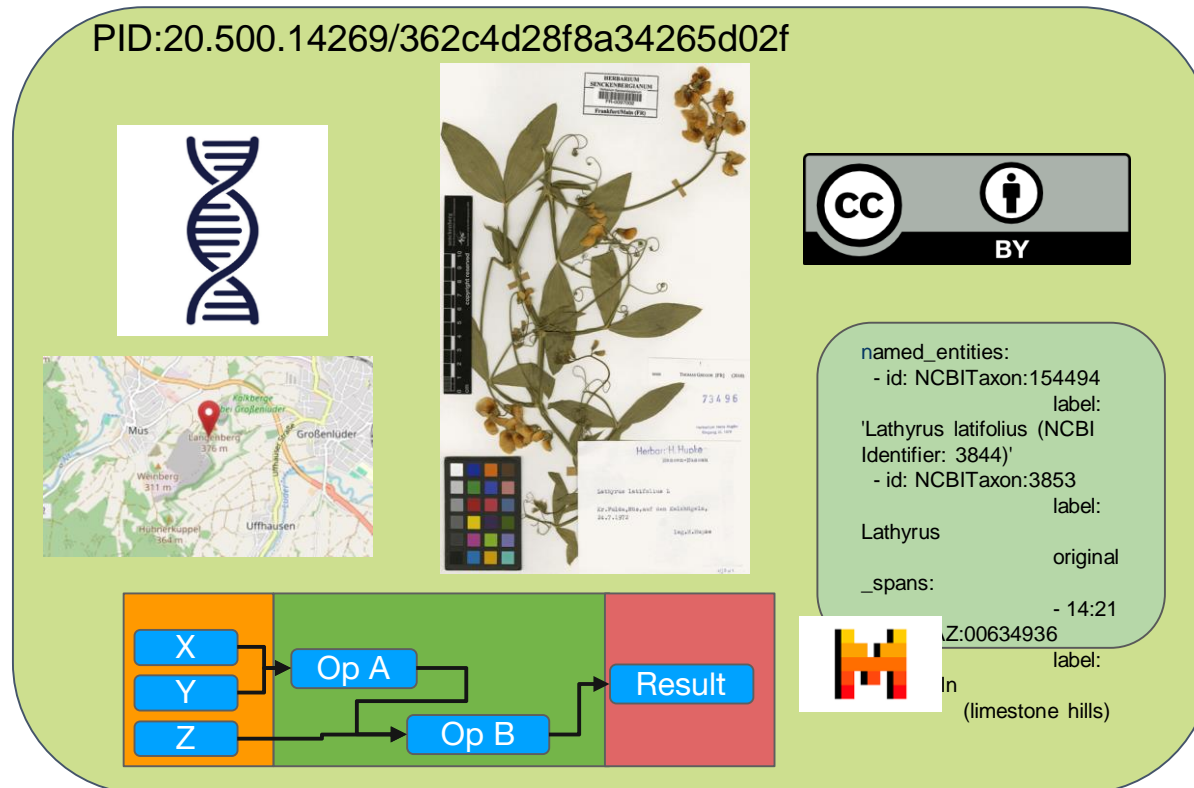
Icons designed by Freepik from Flaticon.

Packaging: FAIR Digital Objects



- **Downstream processing** of material and digital assets **delivers a variety of derived data products** (sequences, measurements, geolocations) **and semantic artefacts** (workflows, metadata, mappings)
- Requirement: **Facilitate access and reuse** of data products for researchers by **realizing stability, identifiability and persistence** of research artifacts
- Approach: **Use unified data model** to assemble all the core information and content in one container format → **FAIR Digital Objects (FDOs)**

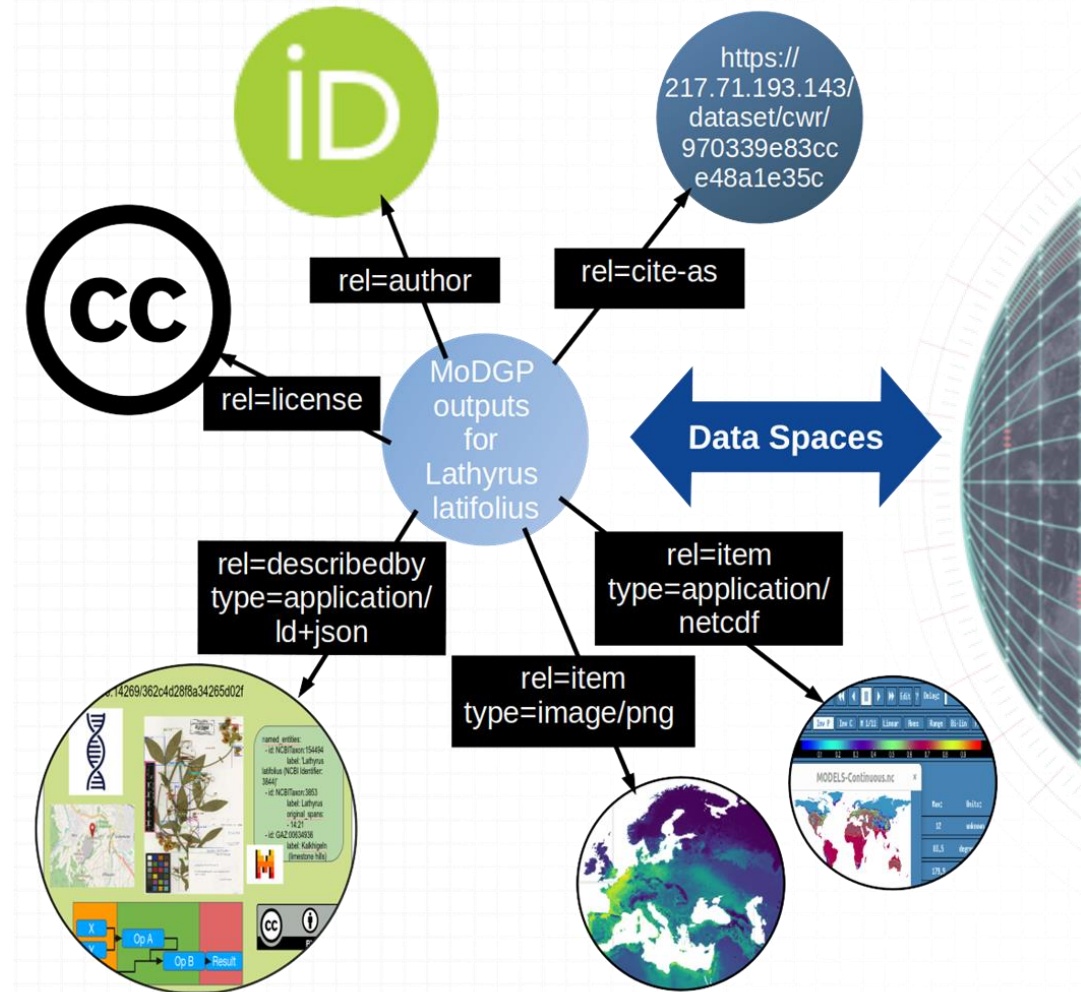
Packaging: FAIR Digital Objects



- **Downstream processing** of material and digital assets **delivers a variety of derived data products** (sequences, measurements, geolocations) **and semantic artefacts** (workflows, metadata, mappings)
- Requirement: **Facilitate access and reuse** of data products for researchers by **realizing stability, identifiability and persistence** of research artifacts
- Approach: **Use unified data model** to assemble all the core information and content in one container format → **FAIR Digital Objects (FDOs)**

Implementing webby FDOs (combing RO-Crate and FAIR Signposting)

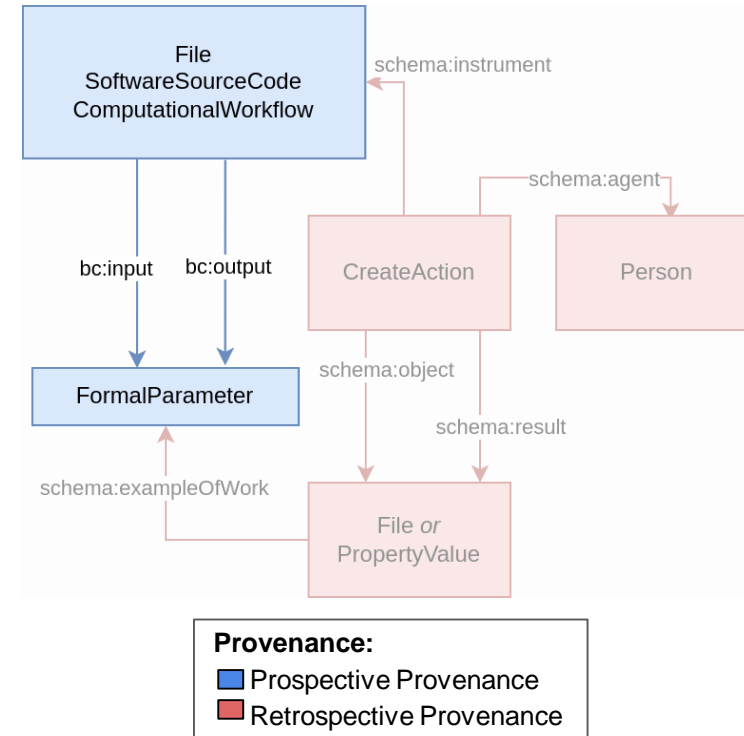
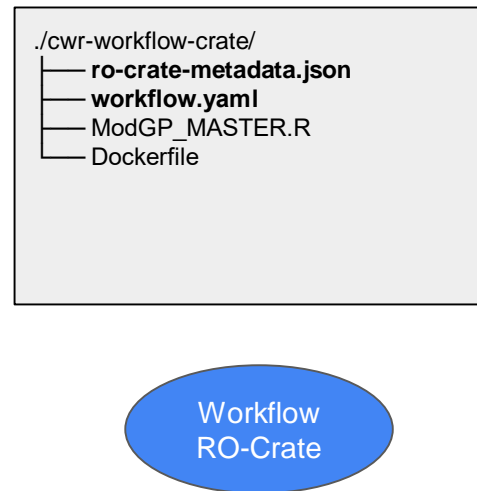
- We use **RO-Crate** + **FAIR Signposting** to set up an implementation path for FDOs based on semantic web technologies (→ **webby FDOs**).
- **RO-Crate**: Provides a **lightweight container for the FAIR-compliant representation of research data** with metadata, workflow descriptions (CWL) and links to schema mappings to make the data cross-domain interoperable .
- **FAIR Signposting**: Adds **machine-interpretable links to human-readable digital objects** using link relation types ([RFC8288](https://www.rfc-editor.org/rfc/rfc8288)) such as item, cite-as, describedBy to present the web topology of a resource



FAIR Workflows

Workflow Run RO-Crate:

- Extension of RO-Crate to capture the prospective/retrospective provenance of computational workflows
- Bundle together all WF associated products (inputs, outputs, code)
- Prototype: workflow.earth

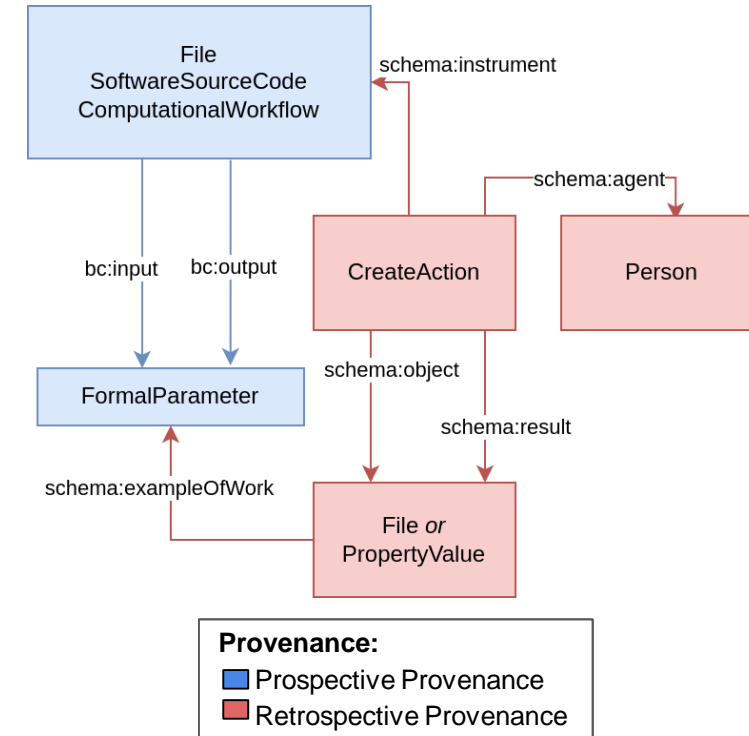
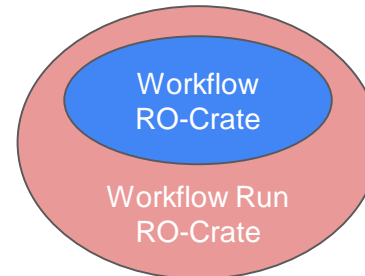
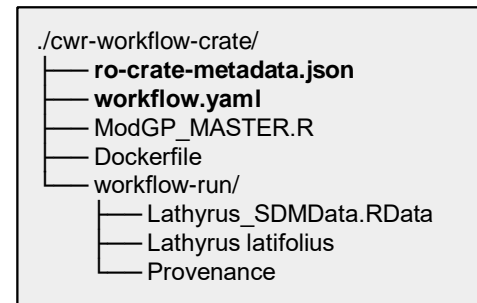


Recording provenance of workflow runs with RO-Crate
<https://arxiv.org/abs/2312.07852>

FAIR Workflows

Workflow Run RO-Crate:

- Extension of RO-Crate to capture the prospective/retrospective provenance of computational workflows
- Bundle together all WF associated products (inputs, outputs, code)
- Prototype: workflow.earth

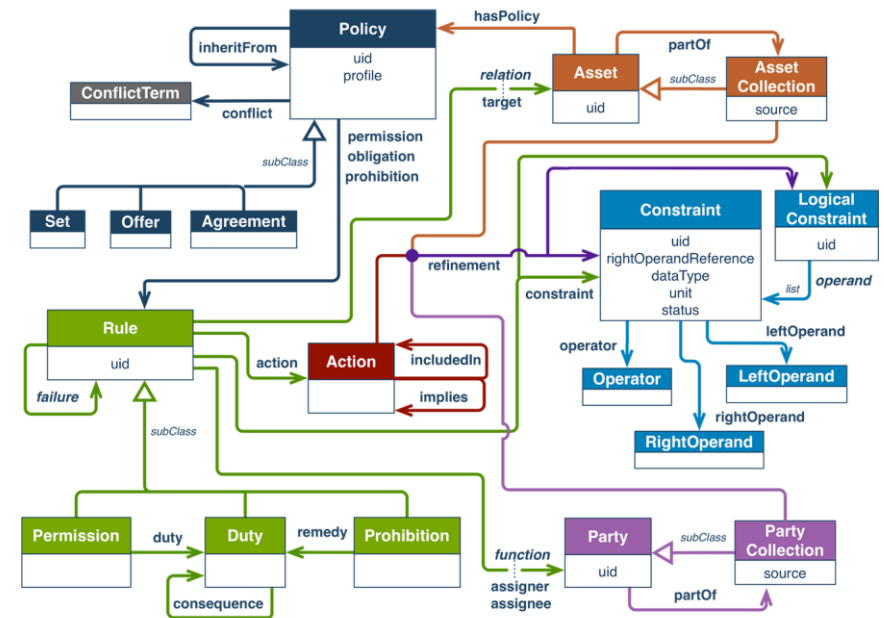


Recording provenance of workflow runs with RO-Crate
<https://arxiv.org/abs/2312.07852>

Guide Rails: machine-actionable policies

Machine-actionable licensing

- Problem
 - Licensing neither easily and universally comparable and customizable
 - CC Licenses do not work for every dataset
 - Other widely used standard licenses do not exist (unlike for software licenses, e.g. spdx)
- Solution
 - ODRL (v2.2): Standardized language used for expressing rights and permissions related to digital assets and data
 - Allows for granular expression of rights and limitations, e.g. only useable in the EU



Guide Rails: machine-actionable policies

Machine-actionable licensing

- Current work
 - [ODRL Builder](#)
 - Coordination of similar projects in NFDI WG LARA (Sect. ELSA)
 - Extension of FDOs by ODRL Metadata

```
{
  "@context": "http://www.w3.org/ns/odrl.jsonld",
  "@type": "Set",
  "uid": "http://example.com/policy:1010",
  "permission": [{
    "target":
      "http://example.com/asset:local_soil.json",
    "action": "cc:Reproduction",
    "constraint": {
      "leftOperand": "purpose",
      "operator": "eq",
      "rightOperand":
        "cc:NonCommercialUse"
    }
  }]
}
```

Workshop tour wrap-up

Services and Tools (Prototypes)

- FAIR Assessment Service
- FAIRagro RDMO instance for DMPs
- DQ-Kit, DQ Toolbox
- Workflow.Earth
- ODRL Builder

Guidance and Documentation

- schema.org & agrischemas extension
- DMP templates
- Data Quality Metrics
- FDO/RO-Crate Packaging
- Licensing Metadata

TA3 Upcoming Events

- Focus Topics TA3 Jour fixe: DeNBI Hackathon Results OntoGPT, 2025-05-22, 14:00
- Workshop on (Standards) Inventory and (FAIR) Assessment (D3.1.3, 2025-06-24/25)