

FAIR data supply in federated data ecosystems


-

challenges, concepts and opportunities

2022-06-20

[Dr. Matthias Lange](#) (Leibniz Institute of Plant Genetics and Crop Plant Research - IPK)  [0000-0002-4316-078X](#)  [@langemger](#)

Prof. Dr. Juliane Fluck (Deutsche Zentralbibliothek Medizin - ZBMED)  [0000-0003-1379-7023](#)  [@FluckJuliane](#)

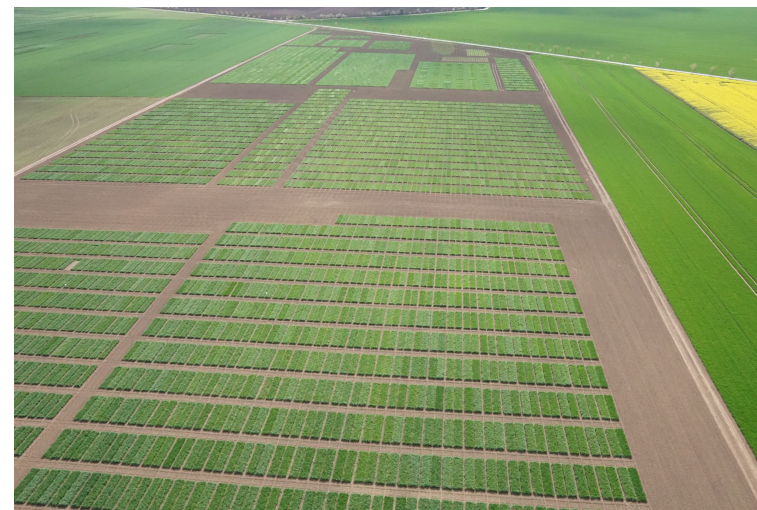
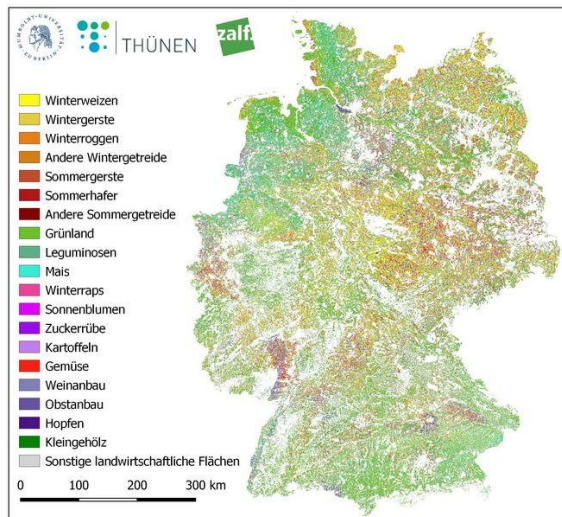
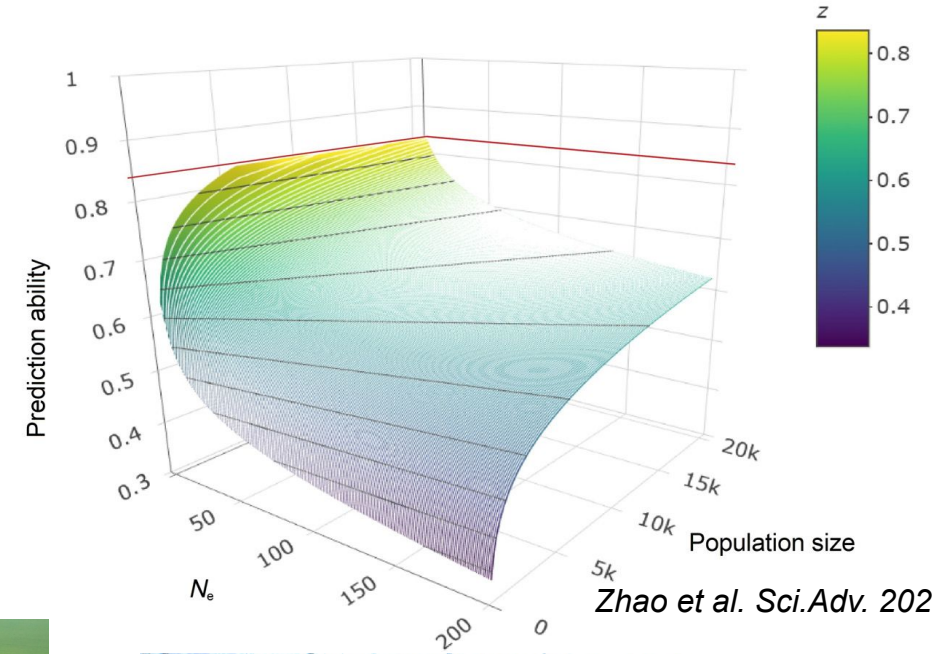
Prof. Dr. Björn Usadel (Forschungszentrum Jülich - FZJ)  [0000-0003-0921-8041](#)  [@usadellab](#)

FAIR as concept to Integrate Data Services towards a Data Science



FAIRagro

- precise breeding: find best genetic combination for new varieties
 - analysis and evaluation of plants in field and greenhouse
 - molecular analysis in the lab
 - environment => localisation, time (clima, soil, pathogens, fertilisation ...)
- selection based on broad integrated high quality data



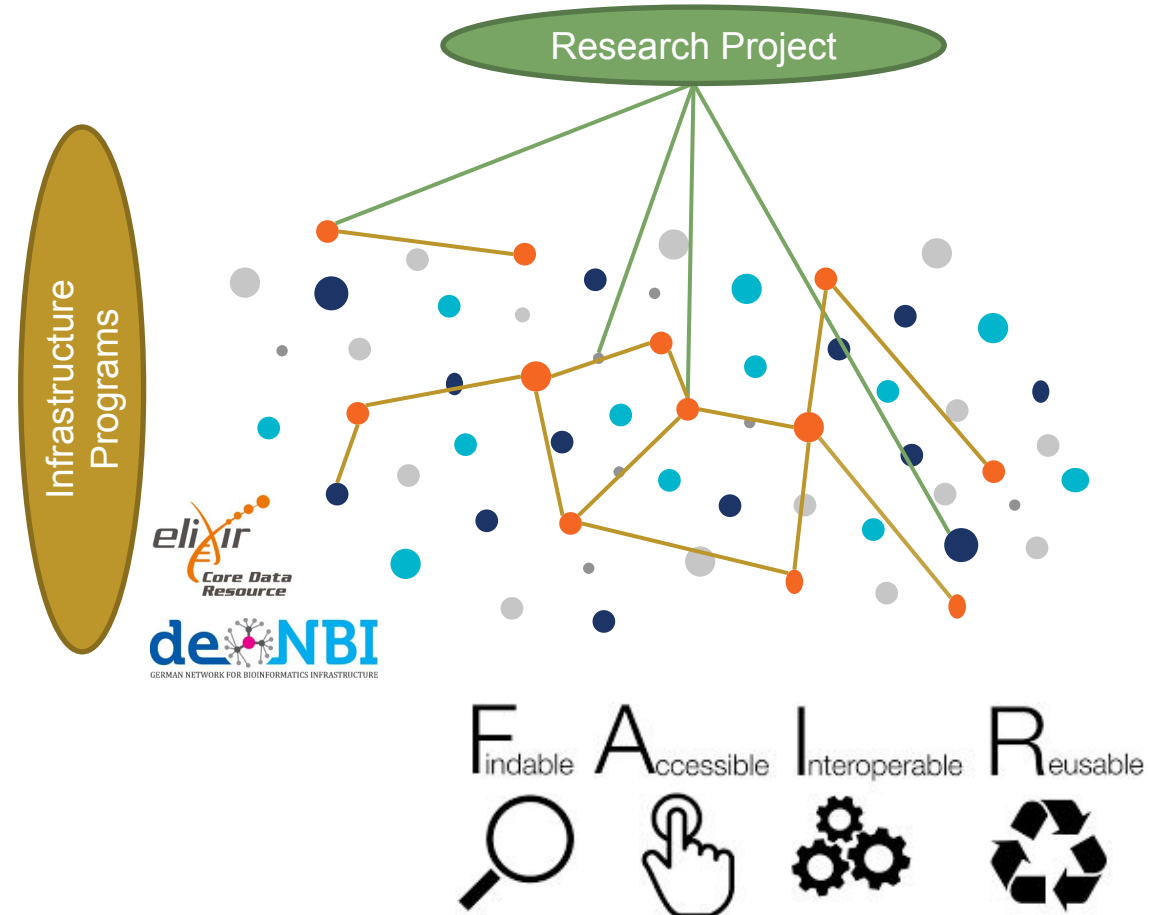
(map agricult. use in Germany © Thünen-Institut)

Federated Research Data Infrastructure (RDI)



FAIRagro

- Distributed research data
- Heterogeneous data services
- Autonomous data service
- Integrated access
- On demand data integration



FAIR - Findability

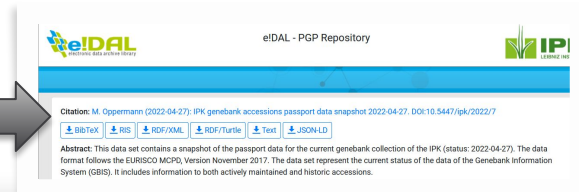


FAIRagro

- persistent globally unique identifier (PID)
 - connect things, places, people
 - coordinated by NFDI - section infra - WG PIDs
- resolving services map ID to resource (+ meta data)
- harmonized meta data schema
 - data about digital entities = basic summary
 - administrative, structural, descriptive
- data and meta data indexed in global search engines



<https://dx.doi.org>
10.5447/ipk/2022/7



The screenshot shows the FAIRsharing.org profile for the 'Plant Genomics and Phenomics Research Database'. The profile includes the following information:

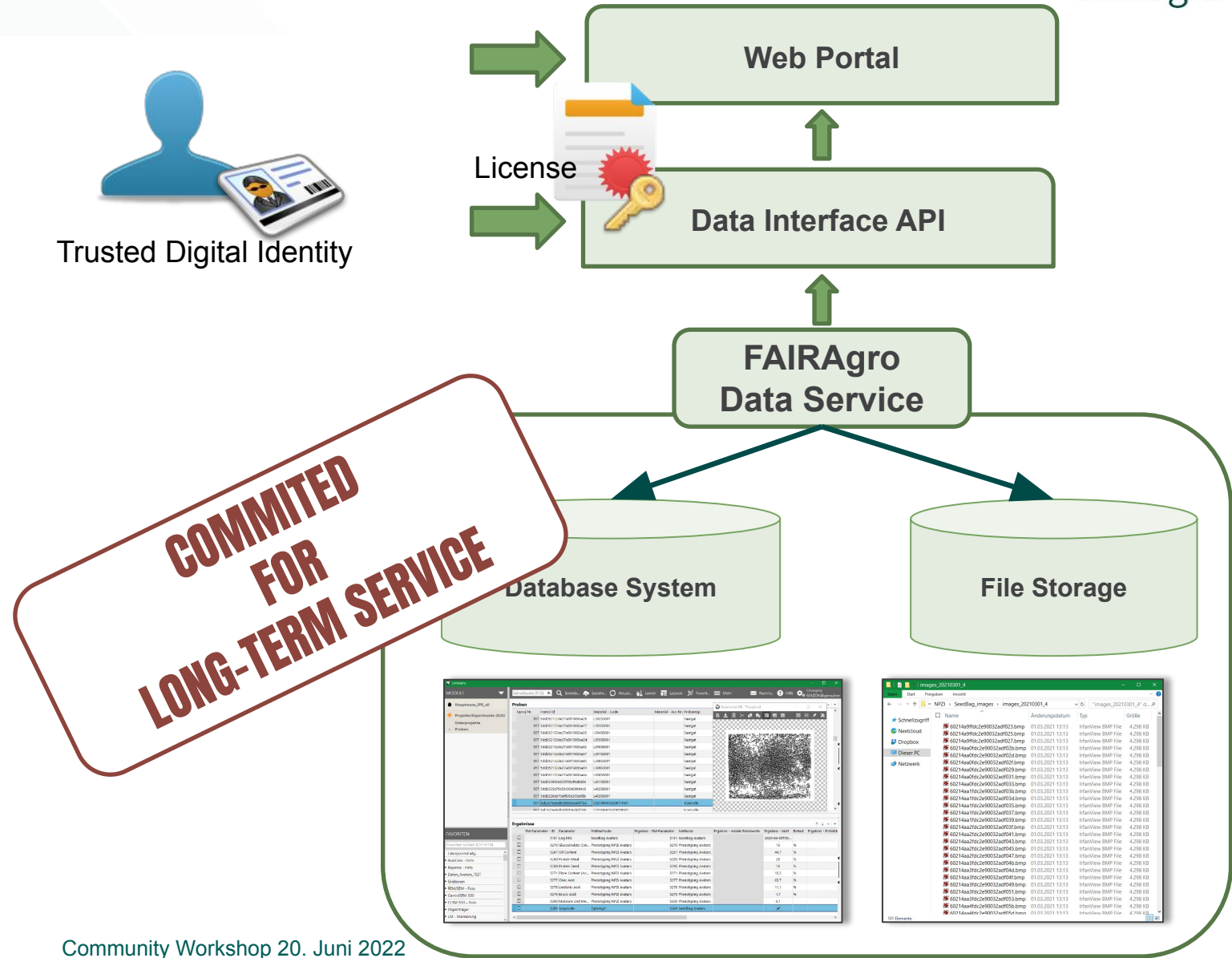
- GENERAL INFORMATION:** Repository, Registry: Database, Description: This repository provides several plant genomic and phenotypic datasets resulting from IPK and German Plant Phenotyping Network (DPPN) research activities. It was established in January 2015. The background of the study is in plant genetic resources, in particular the association of genotypes and phenotypes. This archive aims to provide data from the 'system plant' - from the root to bloom and seed, as well from sequence analysis to systems biology.Ge, Homepage: <http://edal-gpp.ipk-gatersleben.de>, Year of Creation: 2011, Maintainers: arendd (IP), Countries developing this resource: Germany, Subjects: Genomics, Life Science, Domains: Phenotype, Genotype, Taxonomic Range: Plantae.
- License:** R (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International)
- Standard(s):** R (Research Data Alliance), D (DataCite), U (Universal Data Catalog), D (Data DOI)
- API:** R (Research Data Alliance), D (DataCite), U (Universal Data Catalog), D (Data DOI)
- Record Status:** ready, deprecated, uncertain, In Development

FAIR - Accessibility

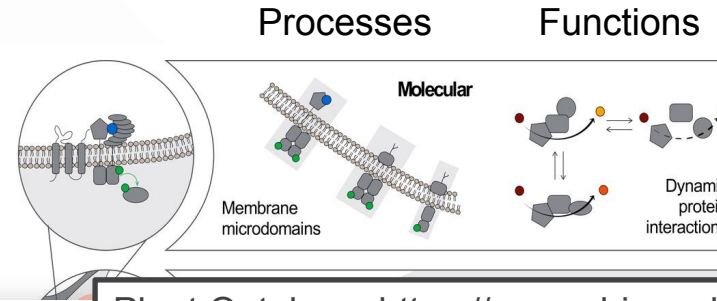


FAIRagro

- Standardised Protocol
- Authorisation and Authentication
- Meta Data and Data Longevity



- Knowledge Representation Languages



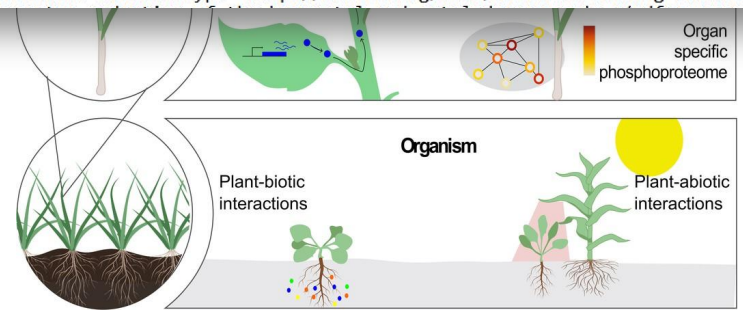
Plant Ontology: <https://www.ebi.ac.uk/ols/ontologies/po>

```

<!-- http://purl.obolibrary.org/obo/PO_0004708 -->
<owl:Class rdf:about="http://purl.obolibrary.org/obo/PO_0004708">
  <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/PO_0020141"/>
  <rdfs:subClassOf rdf:resource="http://purl.obolibrary.org/obo/PO_0025099"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://purl.obolibrary.org/obo/BFO_0000050"/>
      <owl:someValuesFrom rdf:resource="http://purl.obolibrary.org/obo/PO_0019018"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <obo:IAO_0000115 rdf:datatype="http://www.w3.org/2001/XMLSchema#string">A stem node (PO:0020141) that is part of a plant (PO:0019018) where the scutellum (PO:0020110) is attached.</obo:IAO_0000115>
  <oboInOwl:hasDbXref rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PO_GIT:145</oboInOwl:hasDbXref>
  <oboInOwl:hasExactSynonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">nodo escutelar (Spanish, exact)
</oboInOwl:hasExactSynonym>
  <oboInOwl:hasExactSynonym rdf:datatype="http://www.w3.org/2001/XMLSchema#string">胚盤節 (Japanese, exact)</oboInOwl:hasExactSynonym>
  <oboInOwl:hasOBONamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#string">plant_anatomy</oboInOwl:hasOBONamespace>
  <oboInOwl:id rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PO:0004708</oboInOwl:id>
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Found in grasses where no hypocotyl is present. The
  
```

- Structured Vocabularies

- Models/Formats for Data and Meta Data

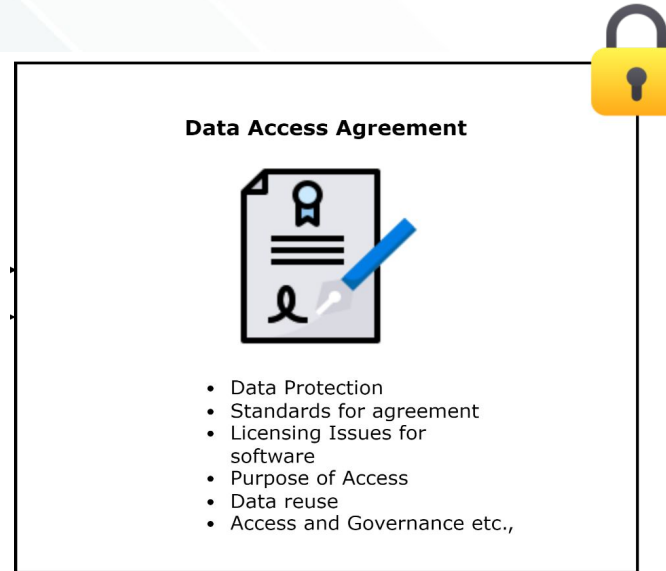


FAIR - Reusability



FAIRagro

- Proper License



Federal Ministry of Education and Research

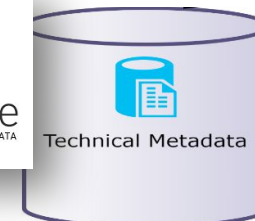
Förderlinie Datentreuhändermodelle



nfdi Nationale Forschungsdaten Infrastruktur

Section ELSA

- Standard Meta Data Schemas



- Data Access Protocols
- Authentication Properties
- Data Structures
- Data Formats
- Contract Metadata
- Abstraction Levels
- Masking Rules etc.,

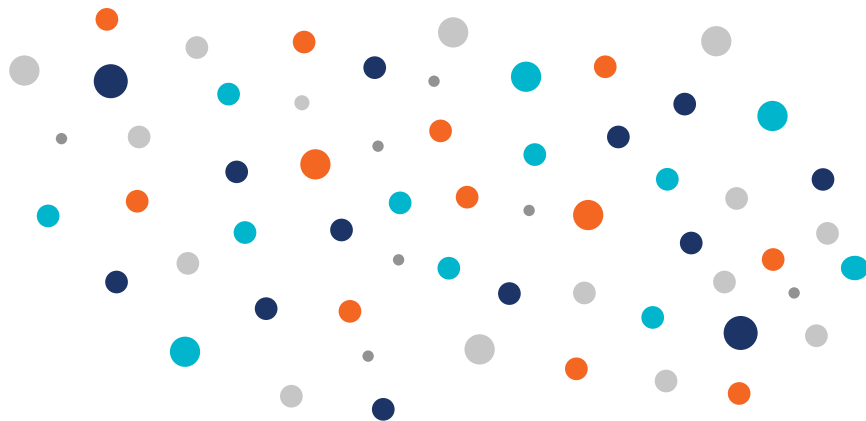
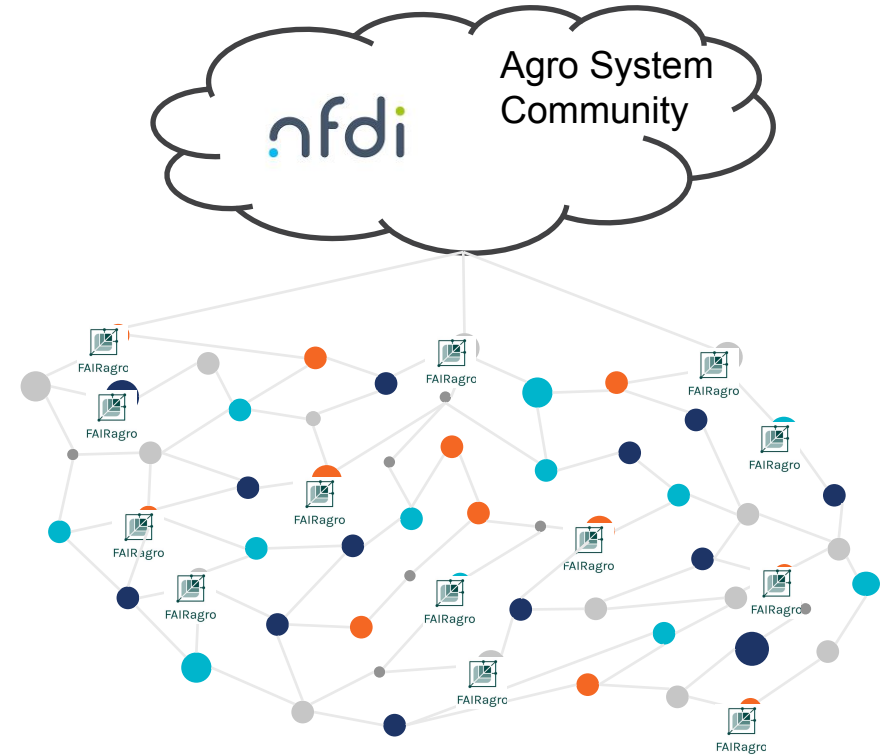
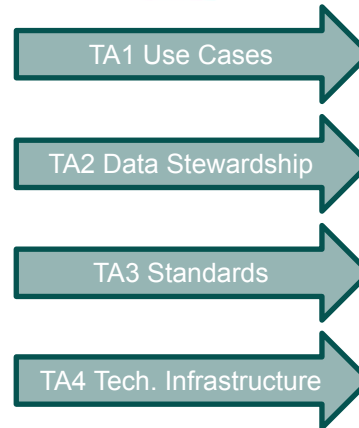


- Investigation
- Study
- Person
- Biological Material
- Environmental Factors
- Experimental Factors
- Sample etc.,

FAIR Agro Escorts RDM Life Cycle to build a Federated Infrastructure



FAIRagro



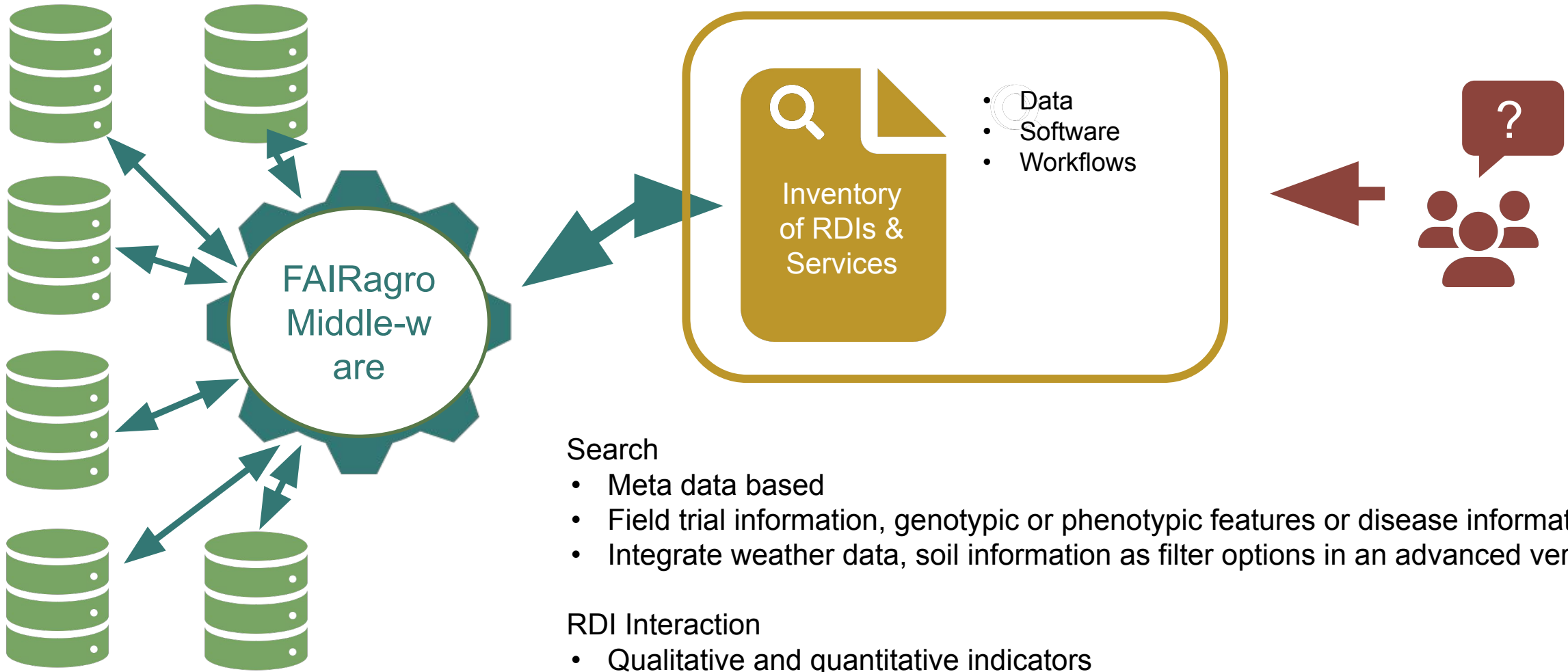
Phase 1	Phase 2	Phase 3
<ul style="list-style-type: none"> e!DAL-PGP (IPK) BonaRes (ZALF) OpenAgrar (JKI) National Soil & Forest Inventories (Thünen) PUBLISSO Repository for Life Sciences (ZB MED) 	<ul style="list-style-type: none"> PhenoRob DB (UBN) PlabiPD (FZJ) Edaphobase (SGN) GBIS/I, LIMS (IPK) JKI-DataCube (JKI) SRADI (TUM) BonaRes Knowledge Library (UFZ) Open Data Server (DWD) 	<ul style="list-style-type: none"> Cross-NFDI infrastructures <p><i>Additional infrastructures from agrosystem community</i></p>

[FAIRagro Research Data Infrastructures in detail](#)

Search Portal over Federated RDI



FAIRagro



Research Data Infrastructures (RDIs)

Search

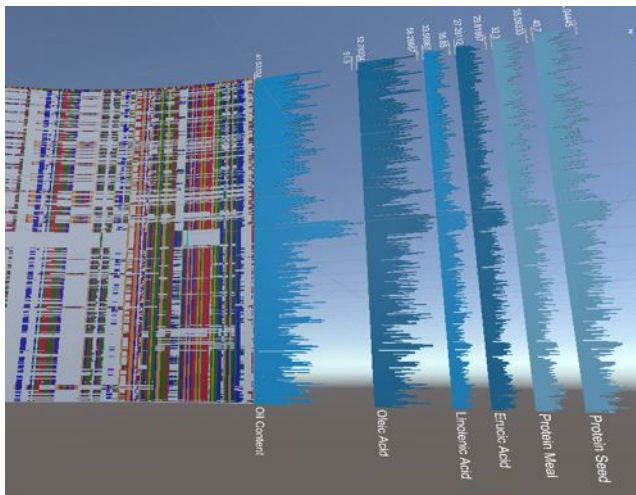
- Meta data based
- Field trial information, genotypic or phenotypic features or disease information
- Integrate weather data, soil information as filter options in an advanced version

RDI Interaction

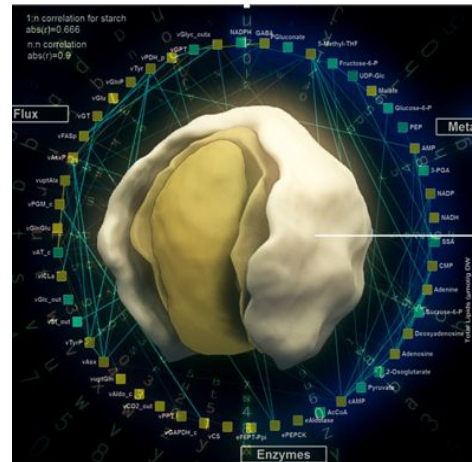
- Qualitative and quantitative indicators
- User feedback on data quality and missing data
- Features data curation by data curators (Community crowd sourcing)

Examples from running initiatives in plants: AVA

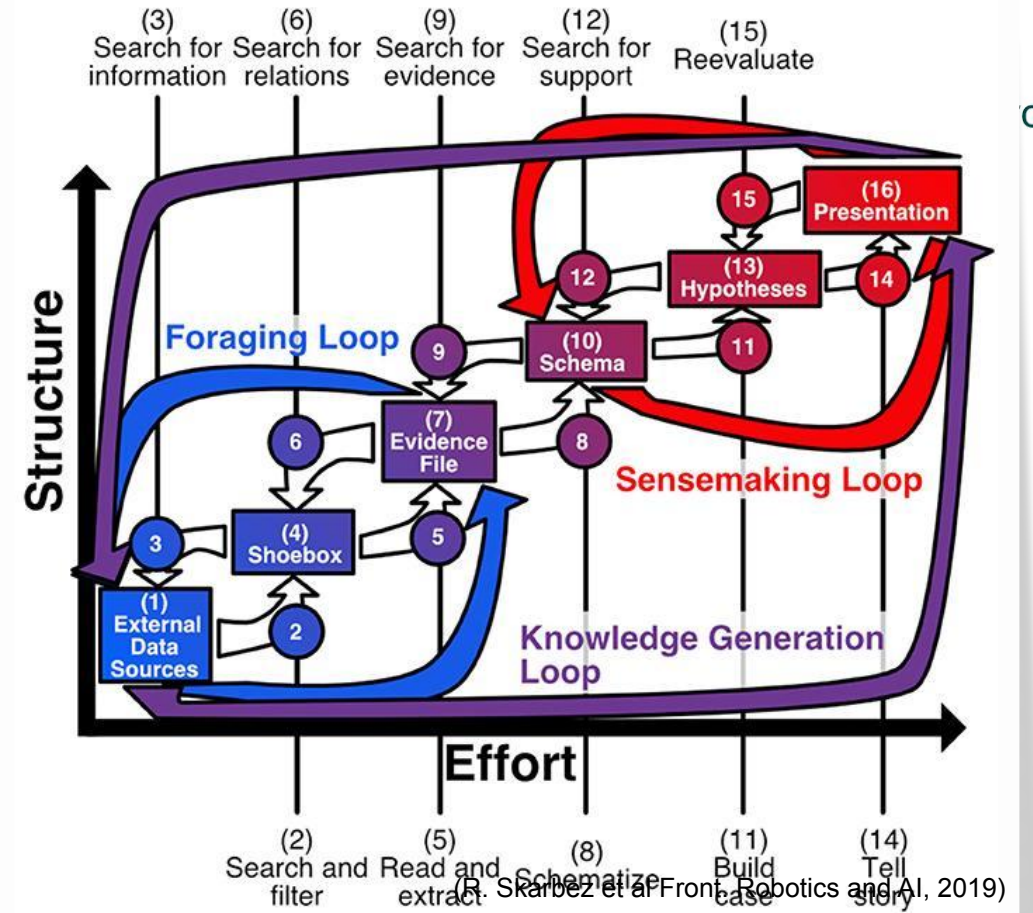
- immersive analytics of broad biological datasets
 - support analytical reasoning, decision making
 - enhanced data interaction and display technologies
 - training and education
- FAIR by federated data flows and standards
- support human knowledge generation loop



enhanced 3D visualisation of genotype phenotype correlations



interactive 3D seed model:



AVATARS –
Advanced Virtuality and Augmented Reality
AppRoaches in Seeds to Seeds

<https://www.avatars-project.de/>

- Integrated Access to Services is Benfitial
- FAIR is conceptual frame
- FAIRAgro
- Portal as Central Entry Point

Thanks to

YOU

FAIRAgro Consortium

IPK groups BIT, QG, HET

