



**HAL**  
open science

# FAIR characterization data of genetic resources: making it real?

Anne-Françoise Adam-Blondon

► **To cite this version:**

Anne-Françoise Adam-Blondon. FAIR characterization data of genetic resources: making it real?. INTEGRAPE 2021 - Achievements, industry needs and next steps, Anne-Françoise Adam-Blondon; Mario Pezzotti, Jun 2021, Virtuel, France. hal-03316847

**HAL Id: hal-03316847**

**<https://hal.inrae.fr/hal-03316847v1>**

Submitted on 6 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Data integration to maximise  
the power of omics  
for grapevine  
improvement

# FAIR characterization data of genetic resources: making it real?

A-F Adam-Blondon (INRAE, FR)

# Genetic Resources are key for changing our model of agriculture and meet current challenges

- Adapt to climate change
- No use of chemicals
- Increase biodiversity



Develop agro-ecological practices

# Enhancing access to data on genetic resources to better monitor their conservation and enhance their use

Development of a common « European Genetic Resources Strategy » with strong recommendations on:

- Enhancing the characterization of genetic resources as a lever for their better use
- Developing FAIR compliant data management practices as a lever for enhancing access to data and knowledge on genetic resources



An instrument developed by ECPGR: the facilitation of public-private European Evaluation (EVA) Networks of genetic resources

<https://www.ecpgr.cgiar.org/european-evaluation-network-eva>



The ECPGR European Evaluation Network (EVA) for Plant Genetic Resources for Food and Agriculture (PGRFA) is an international project aimed at increasing the use of crop genetic diversity and the diversity of stakeholders in plant breeding. EVA is of strategic importance for Europe and provides an opportunity to promote sustainable use of PGRFA to facilitate adaptation of European agriculture to climate change and to contribute towards achieving related Sustainable Development Goals (SDGs).

In collaborative projects involving public and private sector partners and through participatory plant breeding actions EVA is generating standardized evaluation data (both phenotypic and genotypic data) for numerous crop accessions and landraces available in European gene banks. EVA is implemented through crop-specific networks, covering both cereal and vegetable crops.

Right now there are networks for six different crops. Follow the links to find out more about each crop network!

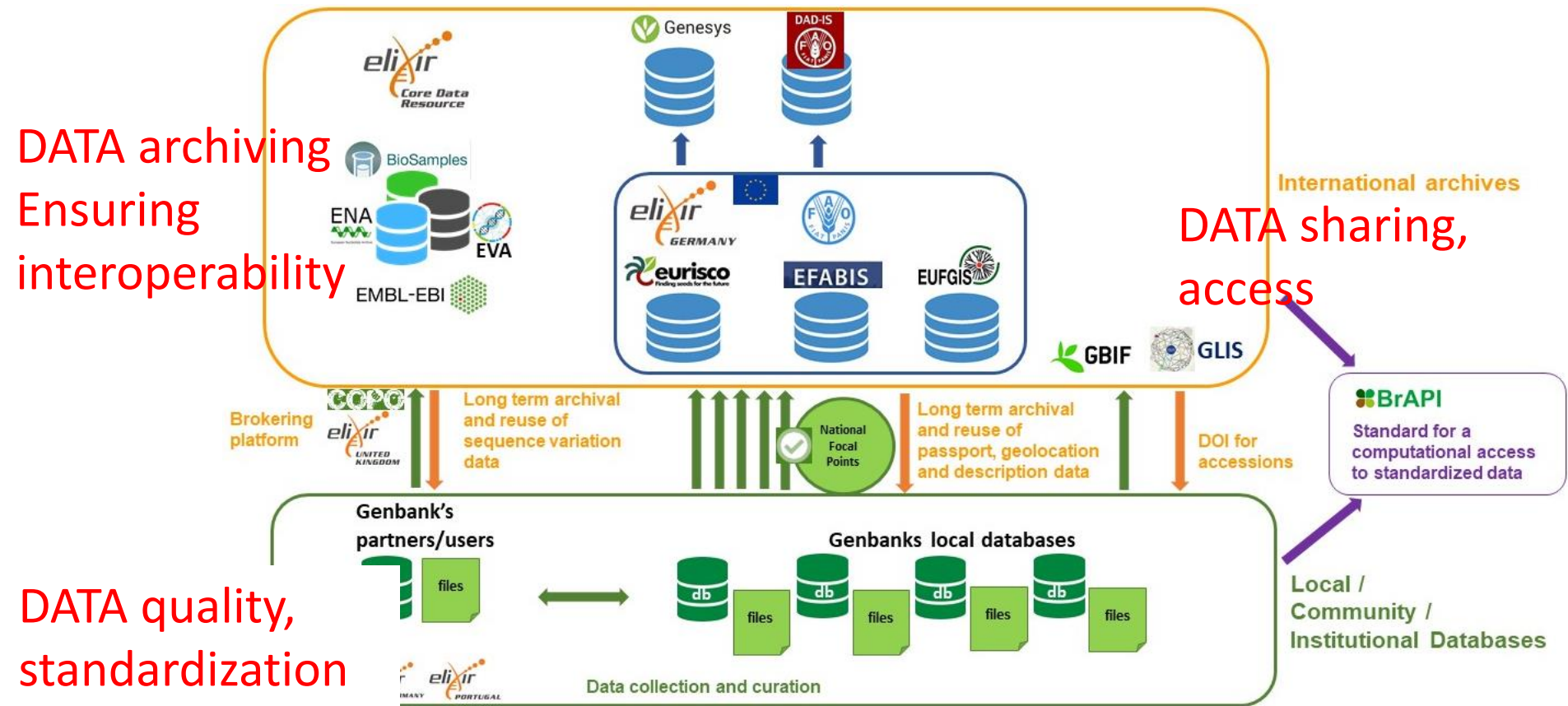
For more information or if you are interested in joining the EVA networks, please contact EVA coordinator [maria.dimitrakou@ecpgr.cgiar.org](mailto:maria.dimitrakou@ecpgr.cgiar.org)



The screenshot shows a web interface with a map of Europe on the left and a list of EVA networks on the right. The networks listed are: EVA Carrot network, EVA Lettuce network, EVA Maize network, EVA Pepper network, and EVA Wheat and Barley network.

# Building « federation » of interoperable information systems about RG

- Networking with many stakeholders at the international level : facilitators such as international infrastructures (GBIF, ELIXIR), consortia (DivSeek), governed networks (ECPGR, ERFP, EUFGIS)
- Networking with many stakeholders at the national level: facilitators National Focal Points/Inventories, national infrastructures/institutes



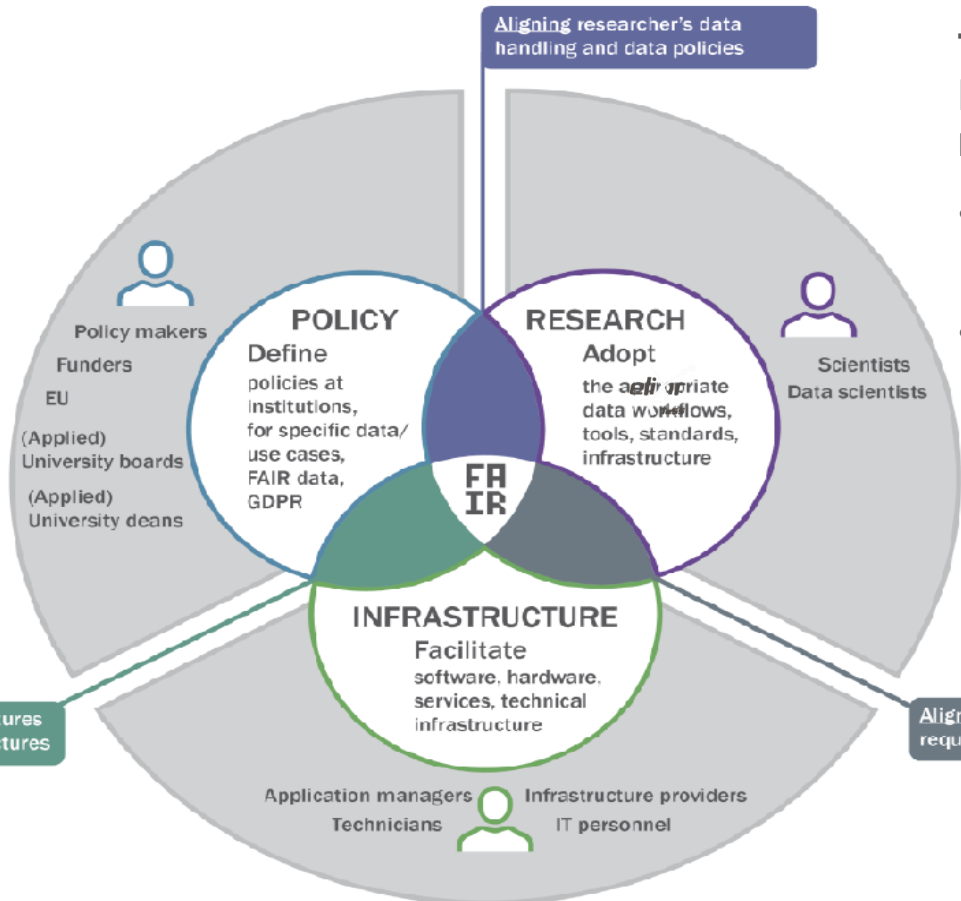
# Needs for sustaining the building of such federation

- **Skills:** data stewards
- **Guidelines, recommendations** for FAIR compliant data management: general and domain related specifications; associated to catalogs
- **Operational implementations** of these guidelines by:
  - Partners from the ELIXIR plant Science community
  - Genbank managers
  - Wheat and barley european public-private evaluation network
  - Academic researchers



# SKILLS: bringing the different dimensions of data stewardship in the GR documentation activities

Implementation areas for data stewardship



The EC high level expert group EOSC estimated 3 FTE per 100 researchers are necessary

- Build awareness and capacity across the whole network
- Leverage existing resources from international initiatives: training, tools, resources



<https://tess.elixir-europe.org/>

From Scholtens et al. (2019)  
<https://doi.org/10.5281/zenodo.3471707>

# Guidelines: capitalizing across Europe

ELIXIR-CONVERGE project: development of a comprehensive repository of guidelines for FAIR data management in the Life Sciences with a participatory approach.

<https://rdmkit.elixir-europe.org/> first version launched in April 2021

## RDMkit – what are the components?

Written by field workers in the language of the community

### RDM Lifecycle



### Guidelines both generic & specific



### Tools and resources



### Role



### Registries



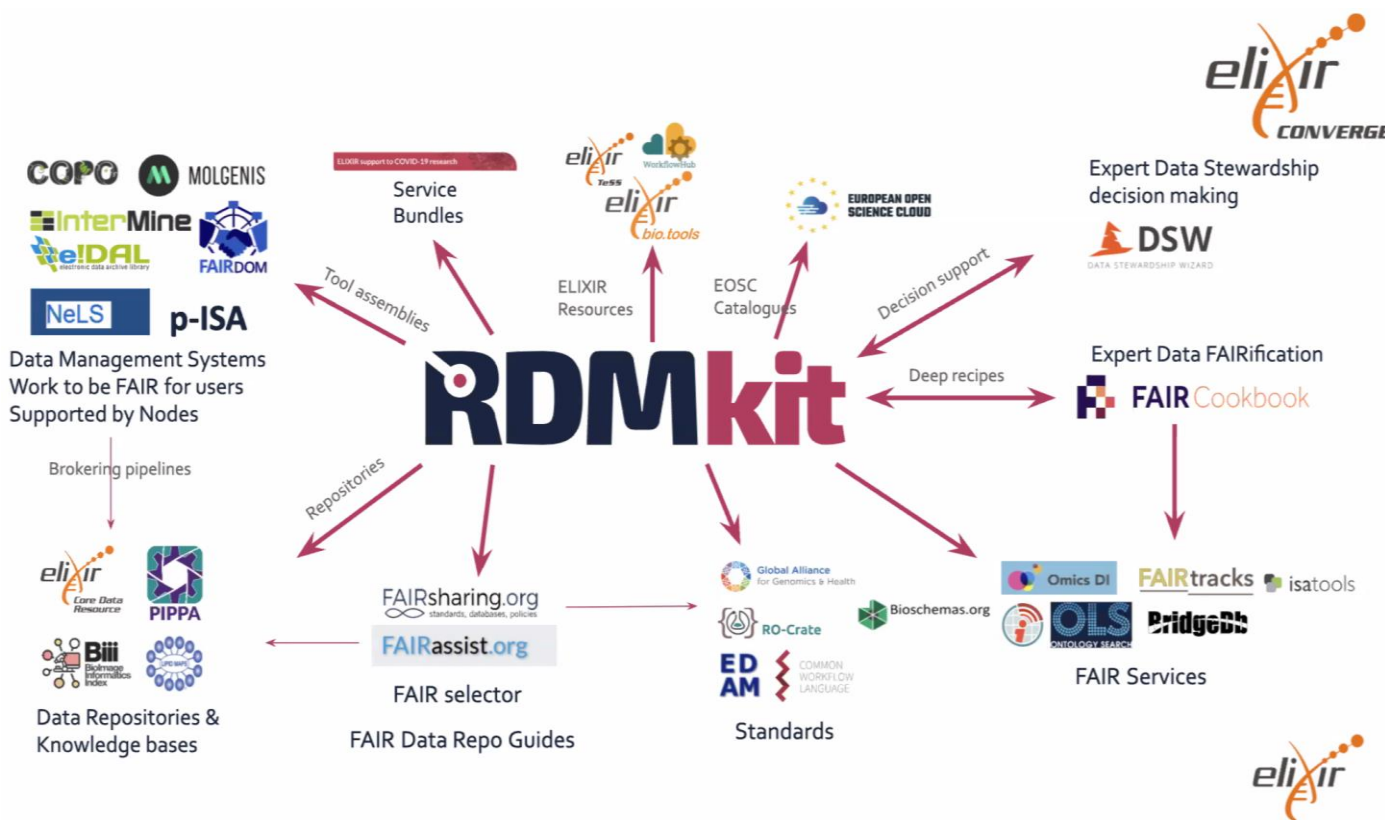
### Example tool assemblies





# Guidelines: capitalizing across Europe

ELIXIR-CONVERGE project: A gateway to resources supporting FAIR data management for research in the Life Sciences.



# Guidelines: domain specificities

ELIXIR-CONVERGE project: Domain pages -> Plant Sciences pages of the ELIXIR Plant Science community.



## Plant sciences [Edit me](#)

- Introduction
- Plant phenotyping metadata management

phenotyping data sharing and deposition  
 ating plant phenotypic and molecular  
 nt tools and resources  
 g materials on plant data management

### Relevant tools and resources

Tool or resource ⓘ	Description	Tags	Registry
BioSamples	BioSamples stores and supplies descriptions and metadata about biological samples used in research and development by academia and industry.	metadata <b>plants</b>	
BioStudies	A database hosting datasets from biological studies. Useful for storing or accessing data that is not compliant for mainstream repositories.	micro biotech metadata <b>plants</b>	
BrAPI	Specification for a standard API for plant data: plant material, plant phenotyping data	IT support <b>plants</b>	
COPO	Portal for scientists to broker more easily rich metadata alongside data to public repos.	metadata researcher <b>plants</b>	 
Crop Ontology	The Crop Ontology compiles concepts to curate phenotyping assays on crop plants, including anatomy, structure and phenotype.	researcher data manager IT support <b>plants</b>	 

### ction

science domain includes studying the adaptation of plants to their en

Guidelines meant to better connect plant genetic resources with their characterization data

# COST Integrape is declining for grapevine the plant domain guidelines

<http://www.integrape.eu/index.php/resources/data-management>



Use them and give feedback!

You are here: [Home](#) ▶ [Resources](#) ▶ [Data management](#)

## Data management

### Guidelines for Data Management

The scope of these guidelines is to give recommendations to provide meaningful information on experiments, starting with the plant material used. Additionally, we set up an ontology for the organs, some of them being not present in general plant ontologies, as well as some recommendations to describe the phenological stages. This will allow a more accurate and standard description of grapevine biological samples. This will support the grapevine research community in opening its data according to the FAIR principles.

- [How to describe an experiment](#)
- [How to submit sequence data to ENA](#)
- [How to submit metabolomic data to MetaboLights](#)
- [How to standardize JBrowse's tracks](#)
- [Apollo Manual Curation Guide for the PN40024.v4 assembly \(under construction\)](#)

# From guidelines to implementation : planning the data and metadata capture

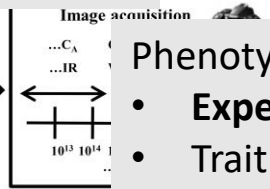
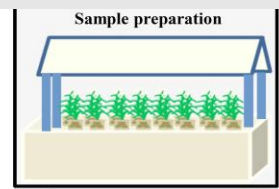
**Set of Genebank Accessions**

- Unique identifiers

Derived material

Lots, Cuttings, Samples

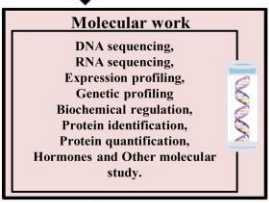
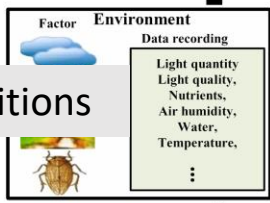
- Unique identifiers
- Developmental stages
- **Sampling method**



Phenotypic/genomic raw data

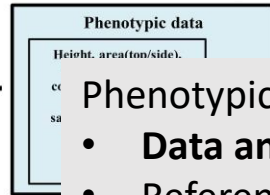
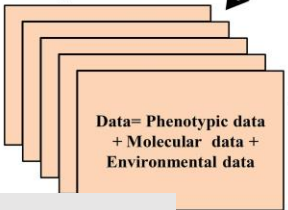
- **Experimental protocols**
- Trait identifiers/standards vocabularies

Environmental conditions



Phenotypic/genomic processed data (e.g. quality control)

- **Data analysis process**
- Reference data sets



Phenotypic/genomic curated data

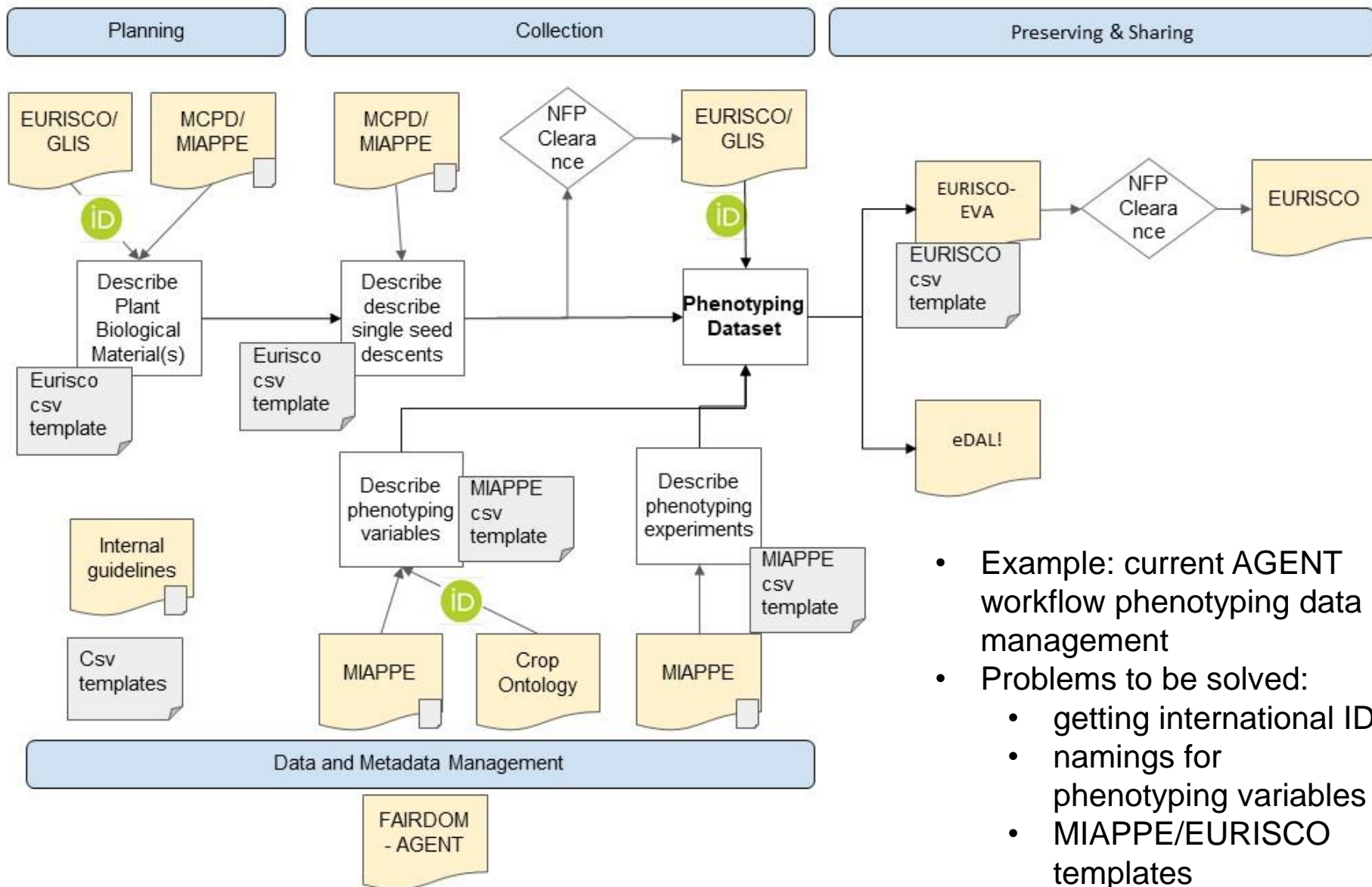
- **Data analysis process**
- Reference data sets

**Final published dataset**

- Project description
- Authors
- **Data integration & analysis process**
- DATASET WITH ALL METADATA



# From guidelines to implementation : planning the data and metadata capture



- Example: current AGENT workflow phenotyping data management
- Problems to be solved:
  - getting international ID
  - namings for phenotyping variables
  - MIAPPE/EURISCO templates

# Platforms facilitating the capture of metadata

Seek/FAIRDOM: <https://fairdomhub.org/>

- Develop projects containers within FAIRDOMhub
- Develop a private instance of Seek/FAIRDOM
  - CESCO (<https://www.cesgo.org/fr/>): a service for data management proposed by the GenOuest facility of ELIXIR-FR
  - pISA-Tree: Standard Directory Tree as a support for Reproducible Research (<https://github.com/ablejec/pISA-tree>) used for data management in the H2020 adapt project (Developing heat- and drought-stress tolerant potatoes)
  - **FAIRDOM instance for the data management of the H2020 AGENT project**



# FAIRDOM supports the ISA data model

Projects => H2020-AGENT



Investigations



Studies



Assays

WP3-Phenotypic-Historical

- Data Files
- Documents
- Publications
- ...

Contains

WP3  Created: 9th Feb 2021 at 07:42	H2020-AGENT
WP3-Phenotypic-Experiment-Precision No description specified Snapshots: No snapshots	H2020-AGENT
WP3-Phenotypic-Historical No description specified Snapshots: No snapshots	H2020-AGENT
Studies: CREA-CI-Phenotypic-Historical-wheat, CRI-Phenotypic-Historical-Barley, CRI-Phenotypic-Historical-Wheat, ICARDA-Phenotypic-Historical-Barley, ICARDA-Phenotypic-Historical-Wheat, IHAR-Phenotypic-Historical-Barley, IHAR-Phenotypic-Historical-Wheat, INIA-Phenotypic-Historical-Wheat, INRAe-Phenotypic-Historical-Wheat, IPGR-SADOVO_Barley_Historical_Phenotypic_Data, IPK-Phenotypic-Historical-Wheat, NPPC-Phenotypic-Historical-Barley, NPPC-Phenotypic-Historical-Wheat, WBF-Agroscope-Phenotypic-Historical-Wheat, WUR Test study upload phenotypic data, WUR-Phenotypic-Historical-Wheat, WURCGN-Phenotypic-Historical-Barley, WURCGN-Phenotypic-Historical-Wheat, [IPGR-SADOVO]-Phenotypic-Historical-Wheat Assays: Assay: [IPGR-Sadovo]-Phenotypic-Historical-Wheat, CREA-CI Phenotypic historical study wheat assay, CRI-Phenotypic-Historical-Wheat, CRI-Phenotypic-historical-barley	

Selected: WP3-Phenotypic-Historical (Investigation)  
 Description: No description  
 SEEK ID: <https://urgi.versailles.inrae.fr/airdom/investigations/2>

- CRI-Phenotypic-Historical-Wheat
    - CRI-Phenotypic-Historical-Wheat
  - CRI-Phenotypic-Historical-Barley
    - CRI-Phenotypic-historical-barley
      - CRI-Spring-Barley-Phenotypic-Historical-Data
  - ICARDA-Phenotypic-Historical-Wheat
    - ICARDA-Phenotypic-Historical-Wheat-Assay-2020
      - ICARDA-Phenotypic-Historical-Wheat-2020-MARCHOUCH-Eurisco
    - ICARDA-Phenotypic-Historical-Wheat-Assay-2019
      - ICARDA-Phenotypic-Historical-Wheat-2019-AREC-eurisco
    - ICARDA-Phenotypic-Historical-Wheat-Assay-2018
      - ICARDA-Phenotypic-Historical-Wheat-2018-MARCHOUCH-eurisco
    - ICARDA-Phenotypic-Historical-Wheat-Assay-2017
      - ICARDA-Phenotypic-Historical-Wheat-2017-TERBOL-eurisco
      - ICARDA-Phenotypic-Historical-Wheat-2017-MARCHOUCH-eurisco
      - ICARDA-Phenotypic-Historical-Wheat-2017-AREC-eurisco
  - IHAR-Phenotypic-Historical-Barley
    - IHAR-Phenotypic-Historical-Barley
      - IHAR-Phenotypic-Historical-winter barley (v2)
      - IHAR-Phenotypic-Historical-spring barley (v1)
  - WBF-Agroscope-Phenotypic-Historical-Wheat
    - WBF-Agroscope-historical-assay-test
      - WBF-Agroscope-template-wheat-test
    - WBF-Agroscope-historical-assay-wheat
  - WUR Test study upload phenotypic data
  - INRAe-Phenotypic-Historical-Wheat
    - INRAe-Phenotypic-Historical-Wheat-Assay
      - URGI\_phenotypic\_Historical\_Wheat\_eurisco\_example



# Trainings/datathons, coordination and governance

- Are necessary to understand and agree on details (this is where the devil lies usually)
- To agree on who is responsible of what when decision and actions are necessary (the devil is here also)



## Examples:

Who is responsible for issuing DOI to accessions (Laboratories, Genbanks, national focal points or EURISCO)?

Who is responsible for issuing unique Sample identifiers and do we use also project internal ID in addition of BioSample IDs?



# Still a long way to go but we can help each others

- Sharing experience on guidelines implementations
- Contributing to raise awareness
  - In the scientific community about guidelines and best practices
  - To our institutes and funders: on the importance of skills and means for data stewardship.
    - See for instance OECD (2020), "Building digital workforce capacity and skills for data-intensive science", *OECD Science, Technology and Industry Policy Papers*, n° 90, Éditions OCDE, Paris, <https://doi.org/10.1787/e08aa3bb-en>.



# Acknowledgements

## GenRes Bridge partners (WP4)

- M. Bozzano (EFI, coord), M. Westergren (Gozdis)
- S. Weise (IPK), T. Van Hintum (WUR)
- E. Charvolin-Lemaire (INRAE), D. Laloe (INRAE), M. Boichard (INRAE)
- S. Sharrock (BGCI)
- E. Sturaro (Unipd)
- S. Goritschnig (Cgiar/ECPGR)

## ELIXIR CONVERGE partners (too many...)

- Hub: N. Blomberg, J. Arenas (coord. and management)
- C. Goble (UK) and F. Coppens (BE; WP3 and RDMkit)
- S. Cappella and L. Portell (SP; WP5 and use cases)
- D. Faria: Plant use case (PT)

## AGENT partners (too many...)

- N. Stein (IPK, coord)
- P. Kersey (RBGK), M. Alaux (INRAE), S. Weise (IPK), C. Pommier (INRAE), M. Lange (IPK), R. Finkers (WUR), *J. Destin (INRAE)*
- V. Holubec (CRI; link with phenotyping WP)

- The GenRes Bridge project has received fundings from the European Union's Horizon 2020 research and innovation programme under grant agreement No 817580
- The ELIXIR-CONVERGE project has received fundings from the European Union's Horizon 2020 research and innovation programme under grant agreement No
- The AGENT project has received fundings from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862613

Thank you!

