FAIR characterization data of genetic resources: making it real?

Anne-Françoise Adam-Blondon

FAIR characterization data of genetic resources: making it real?

A-F Adam-Blondon (INRAE, FR)

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme
of the European Union

COST Action
CA17111
INTEGRAPE

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

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action
CA17111
INTEGRAPE

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

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

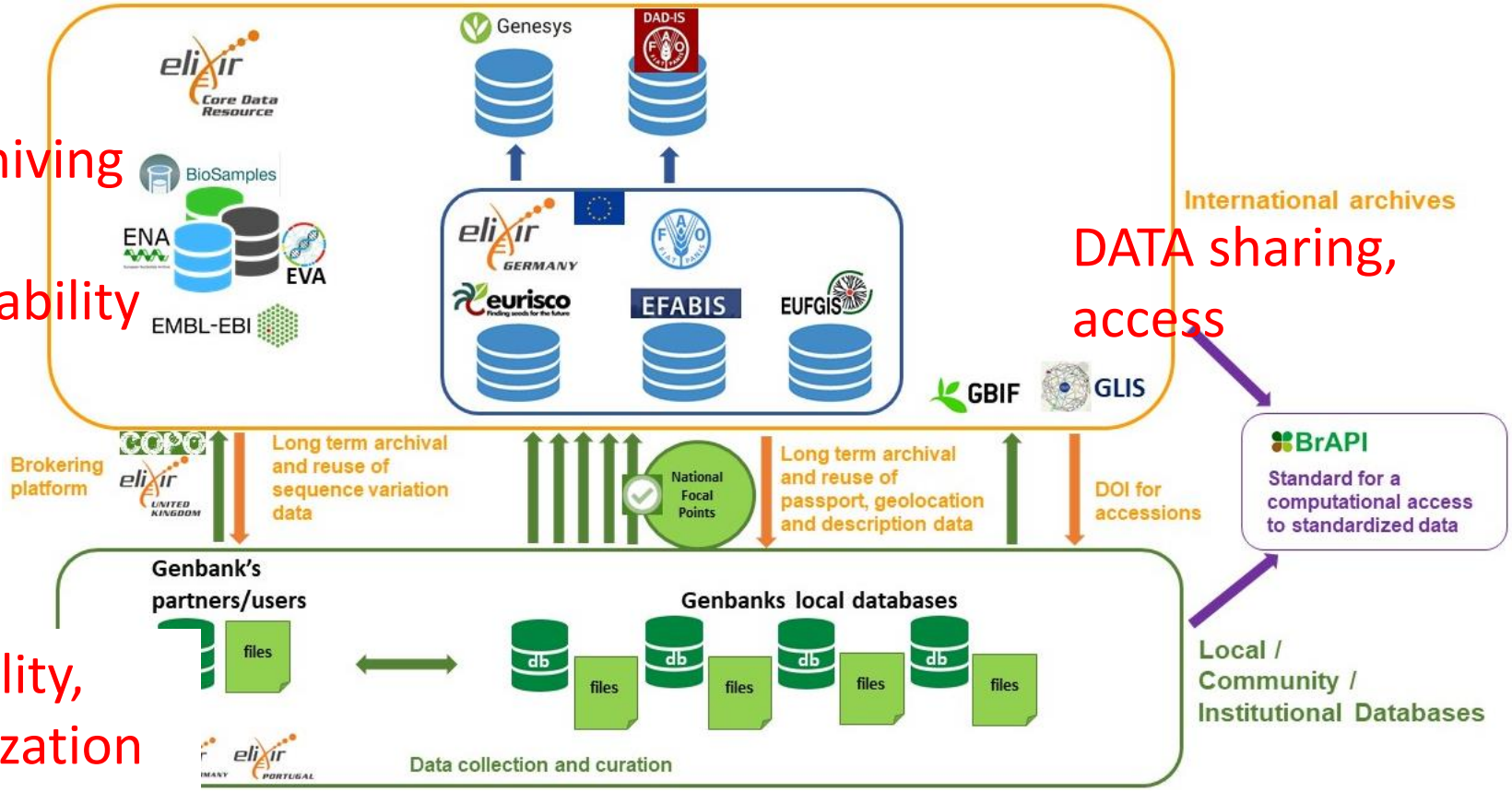
Building « federation » of interoperable information systems about RG

- 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

DATA archiving
Ensuring interoperability

DATA sharing, access

DATA quality, standardization



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

elixir CONVERGE

https://elixir-europe.org/

AGENT

Activated Genbank Network

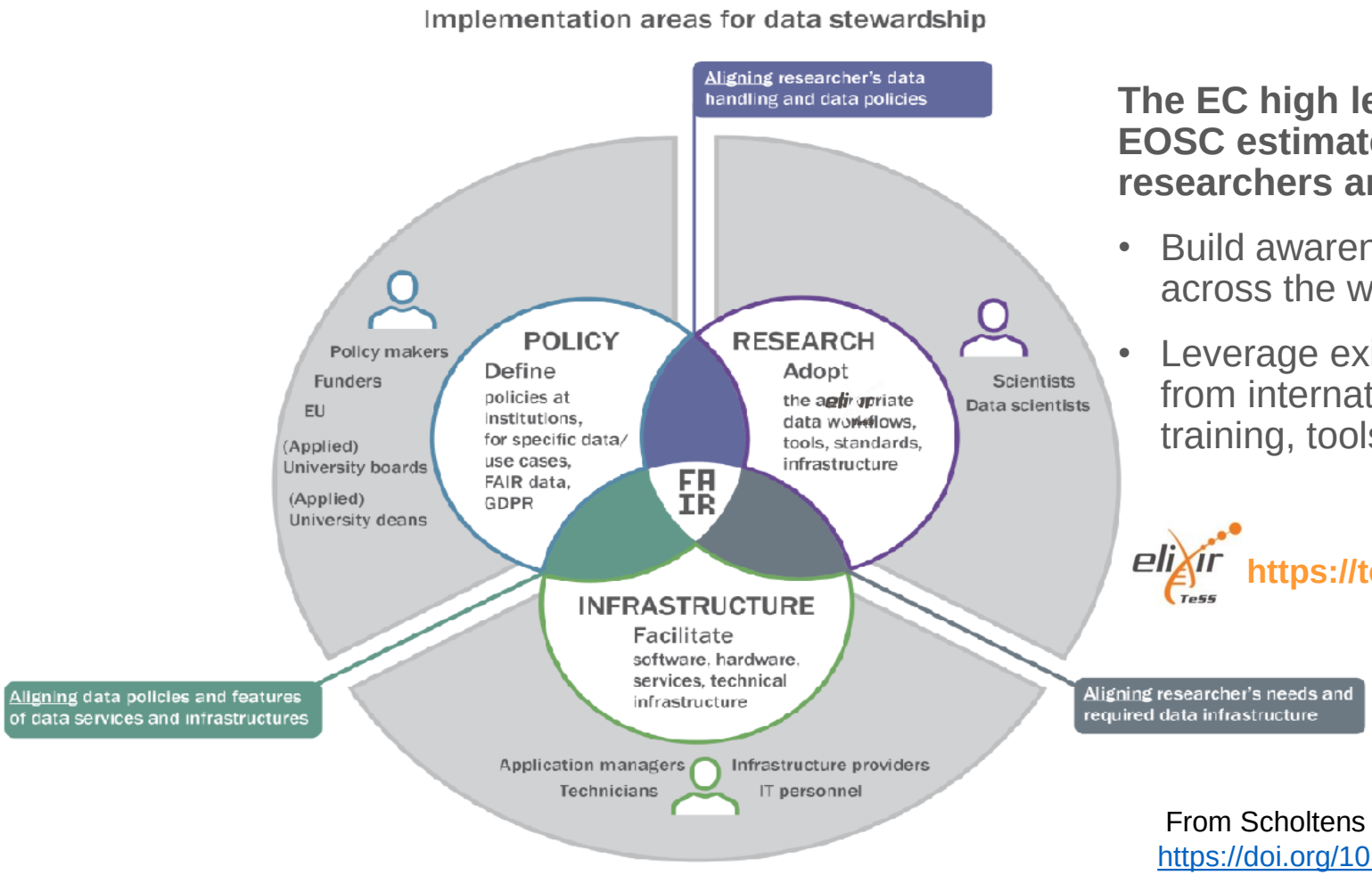
https://agent-project.eu/

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action CA17111 INTEGRAPE

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



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, ressources

elixir TeSS **https://tess.elixir-europe.org/**

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

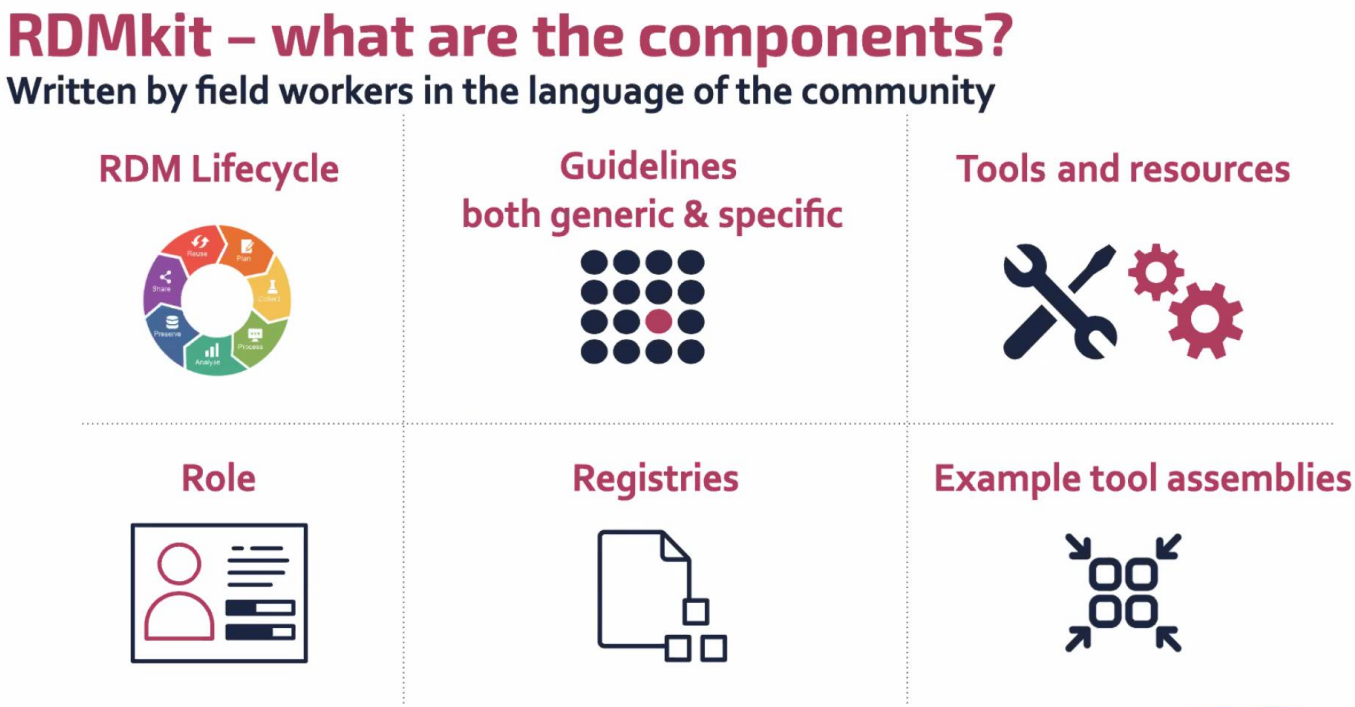
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action CA17111 INTEGRAPE

Guidelines: capitalizing across Europe

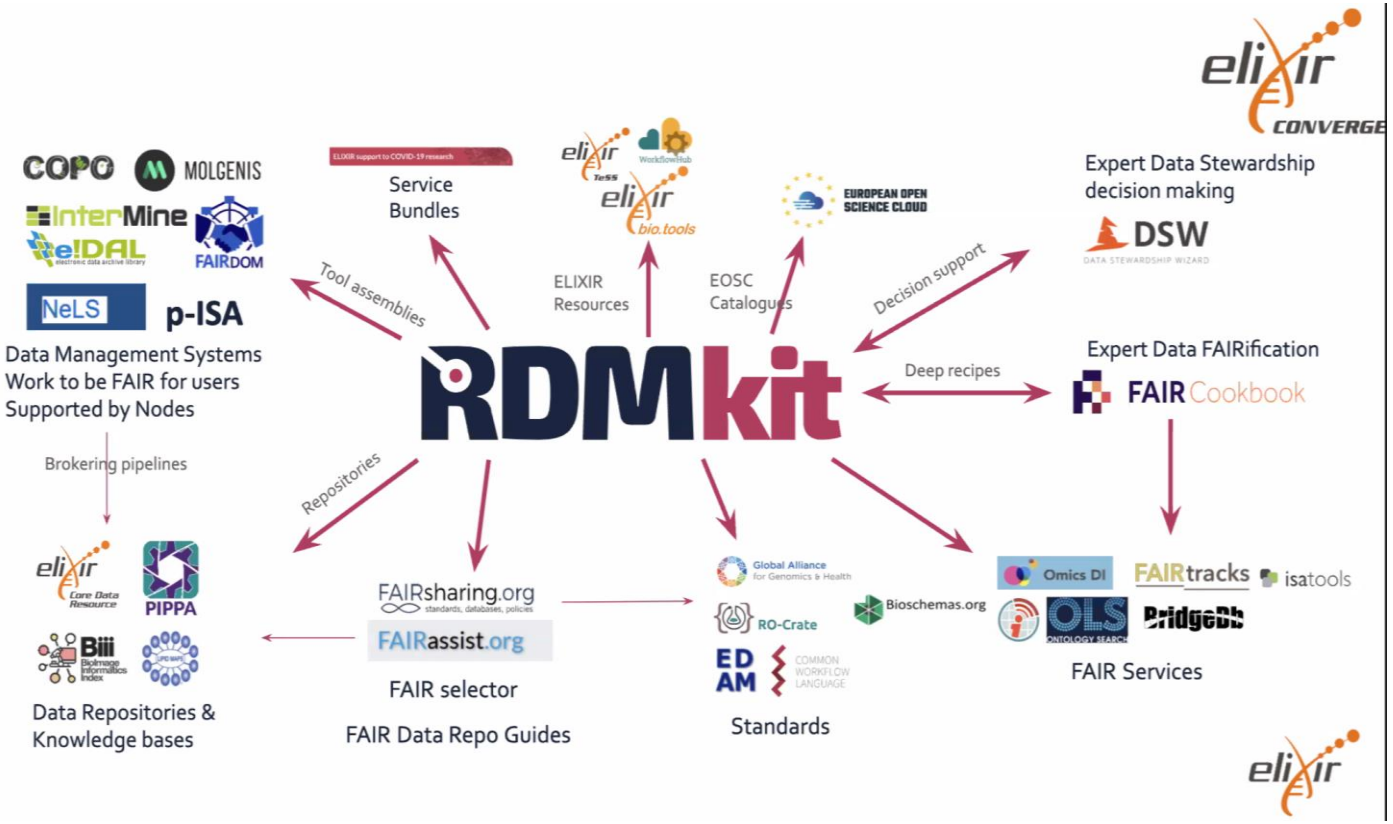
ELIXIR-CONVERGE project: development of a comprehensive repository of guidelines for FAIR data management in the Life Sciences with a participatory approach.
https://rdmkit.elixir-europe.org/ **first version launched in April 2021**



Funded by the Horizon 2020 Framework Programme of the European Union

Guidelines: capitalizing across Europe

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



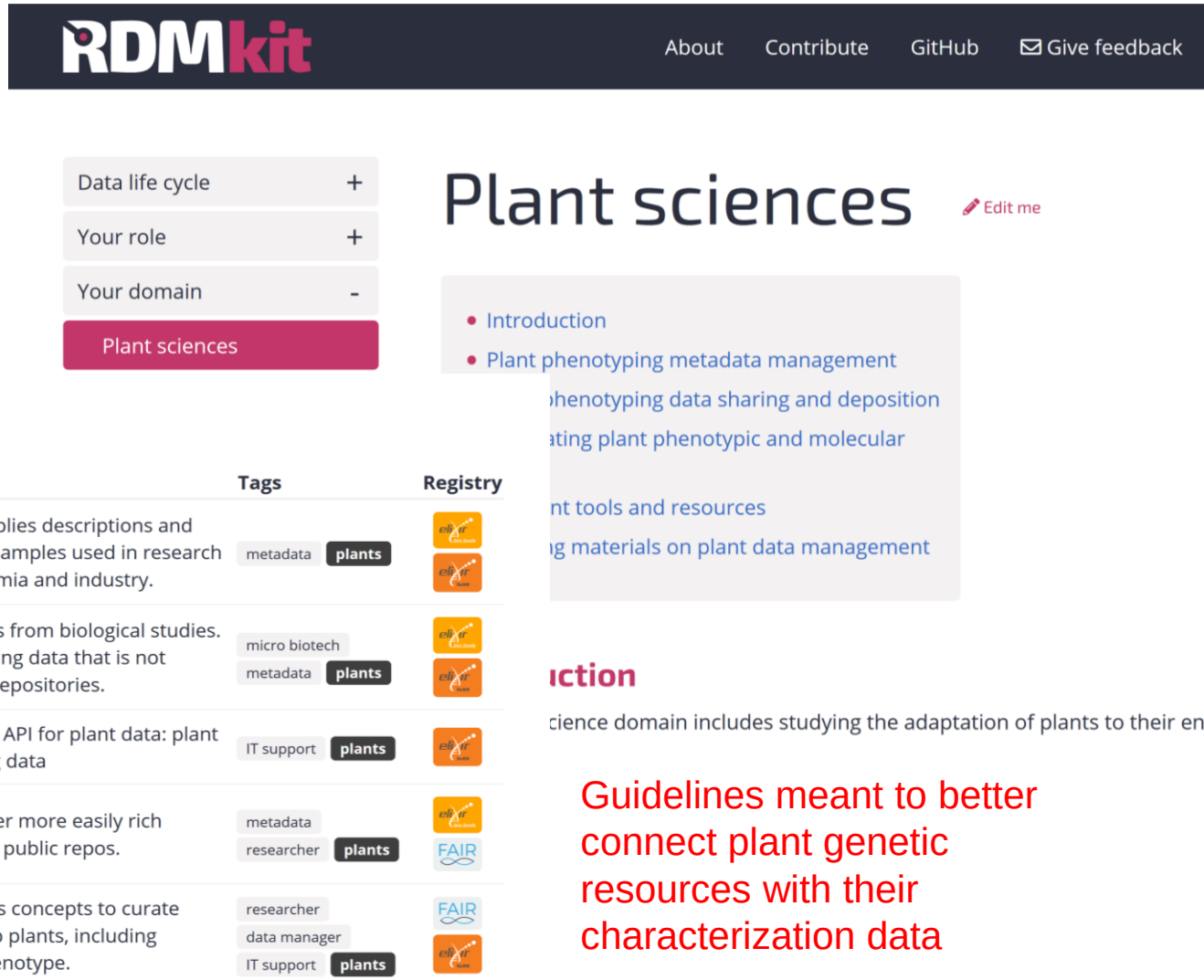
cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action CA17111 INTEGRAPE

Guidelines: domain specificities

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



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

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)**

Use them and give feedback!

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

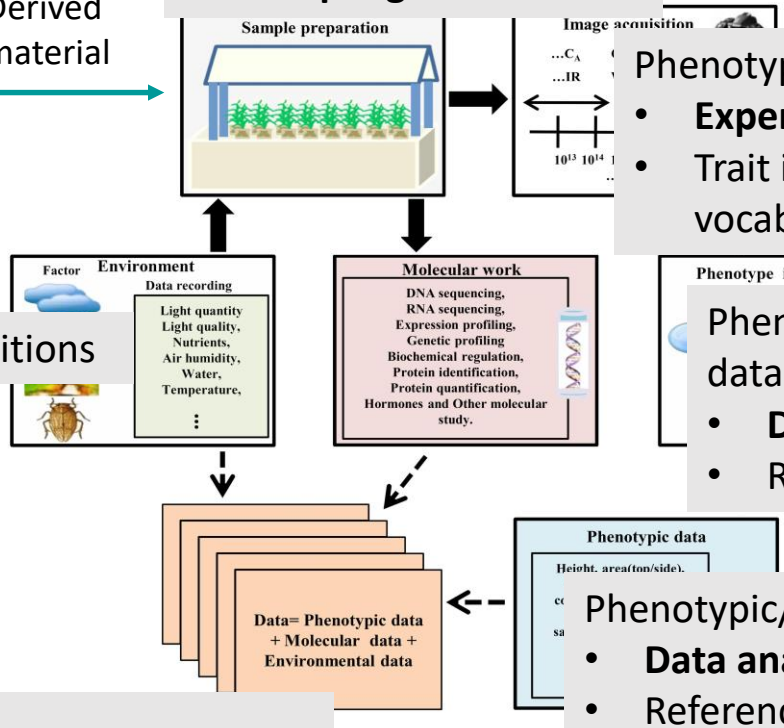
COST Action
CA17111
INTEGRAPE

From guidelines to implementation : planning the data and metadata capture

Lots, Cuttings, Samples
- Unique identifiers
- Developmental stages
- **Sampling method**

Derived material

Set of Genebank Accessions
- Unique identifiers



Phenotypic/genomic raw data
- **Experimental protocole**
- Trait identifiers/standards vocabularies

Environmental conditions

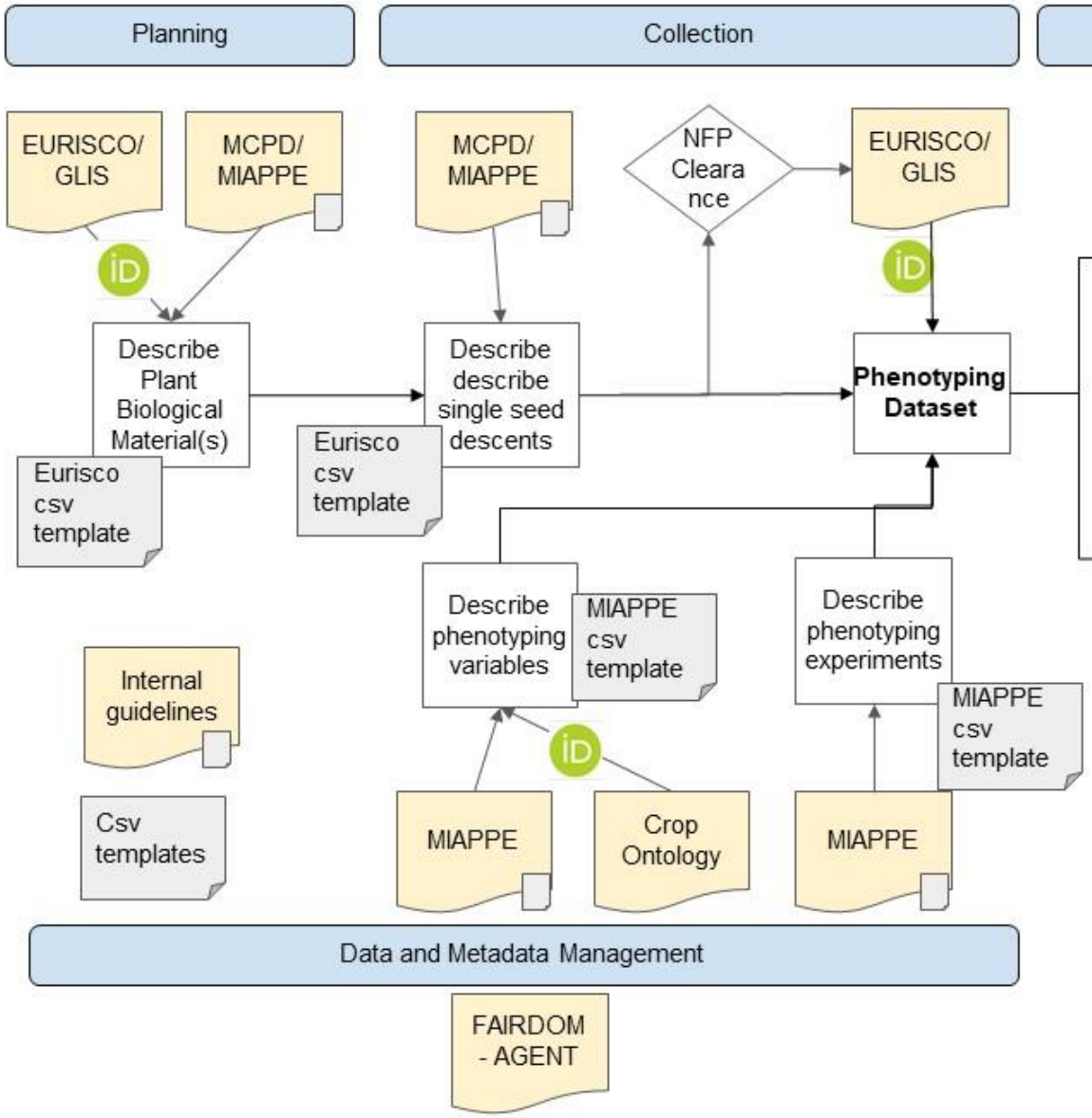
Phenotypic/genomic processed data (e.g. quality control)
- **Data analysis process**
- Reference data sets

Phenotypic/genomic curated data
- **Data analysis process**
- Reference data sets

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

izon 2020 Framework Programme
nion

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/

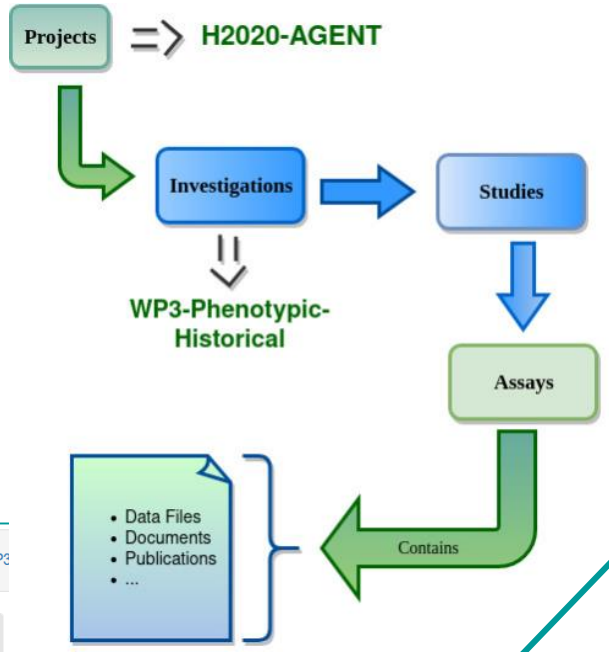
- Develop projects containers within FAIRDOMhub
- Develop a private instance of Seek/FAIRDOM
  - CESGO (https://www.cesgo.org/fr/): a service for data management proposed by the GenOuest facility of ELIXIR-FR
  - pISA-Tree: Standard Directory Tree as a support for Reproducible Research (https://github.com/ablejec/pISA-tree) 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**

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action CA17111 INTEGRAPE

FAIRDOM supports the ISA data model



WP3 H2020-AGENT

Snapshots: No Snapshots

Studies: No Studies
Assays: No Assays

Created: 9th Feb 2021 at 07:42

WP3-Phenotypic-Experiment-Precision H2020-AGENT

No description specified

Snapshots: *No snapshots*

Studies: *No Studies*
Assays: *No Assays*

Created: 9th Feb 2021 at 07:39

WP3-Phenotypic-Historical H2020-AGENT

No description specified

Snapshots: *No snapshots*

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/fairdom/investigations/2

- 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

mme

AGENT

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?

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action CA17111 INTEGRAPE

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.

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action
CA17111
INTEGRAPE

Aknowledgements

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**

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

COST Action
CA17111
INTEGRAPE

Thank you!

cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme
of the European Union

COST Action
CA17111
INTEGRAPE