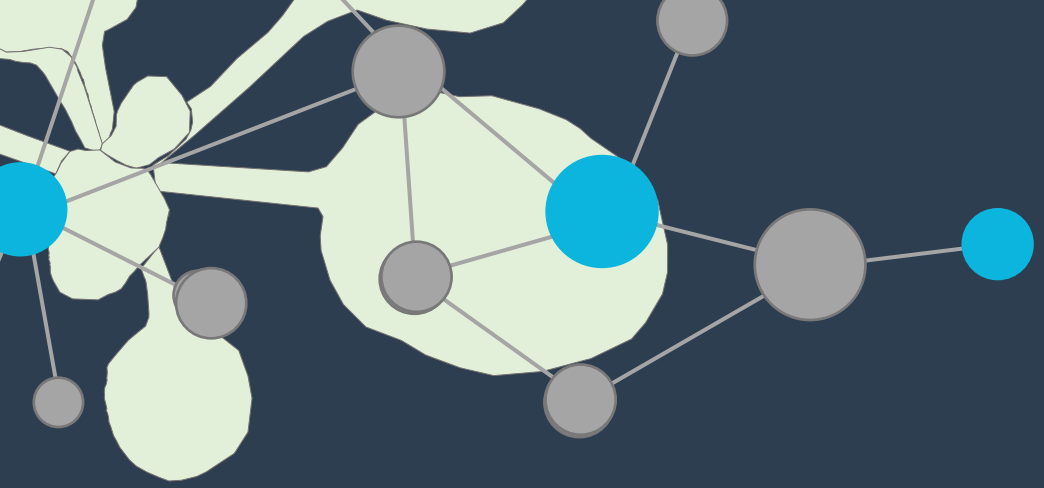


RPTU Summer School on RDM

Cristina Martins Rodrigues

17th July 2023

Agenda

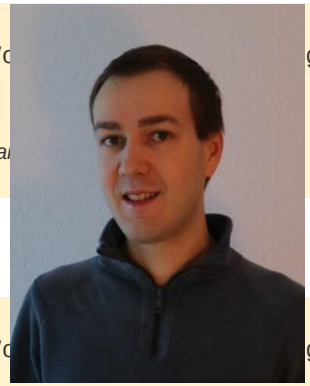
covering the aspects of Research Data Management

	Monday (17.7.)	Tuesday (18.7.)	Wednesday (19.7.)	Thursday (20.7.)	Friday (21.7.)
09:00 - 10:30	Opening / Keynote *Rodrigues*	Versioning & Collaboration *Garth*	Organization and Archives *Wetzels*	Workflows and Data Processing *Garth / Kappe*	RDM Planning *Josch / Niederprüm*
	coffee break	coffee break	coffee break	coffee break	coffee break
11:00 - 12:30	Process Models *Mühlhaus*	Versioning & Collaboration *Garth*	Databases *Doniparthi*	Workflows and Data Processing *Garth / Kappe*	Closing + Open Session *all*
	lunch break	*lunch break*	*lunch break*	*lunch break*	
14:00 - 15:30	Process Models *Mühlhaus*	Galaxy *Gallardo Alba*	group hike & Bremerhof dinner	RDM in practice *Brilhaus*	

Data PLANT

Agenda

covering the aspects of Research Data Management

	Monday (17.7.)	Tuesday (18.7.)	Wednesday (19.7.)	Thursday (20.7.)	Friday (21.7.)
09:00 - 10:30	Ope *Rod* 	Ver *Gar* 	Orga *Wetz* 	Wo g *Ga* 	RDM *Josch* 
11:00 - 12:30	Proc *Mühlhaus*	Ver *Garth*	Data *Doniparthi*	Wo g *Garth / Kappe*	Closi *all*
14:00 - 15:30	Proc *Mühlh* 	Ga *Gal* 		RDM in pract *Brill* 	

Data PLANT

Why Research Data Management?

Relevance of research data management

https://youtu.be/N2zK3sAtr-4

Data PLANT

"Data Sharing and Management Snafu in 3 Short Acts" YouTube, uploaded by NYU Health Sciences Library, 19 Dec 2012, www.youtube.com/watch?v=N2zK3sAtr-4

Benefits of research data management

- Fast Findability of Data
- Knowledge Preservation
- Avoid Data Loss
- Referenceability
- …

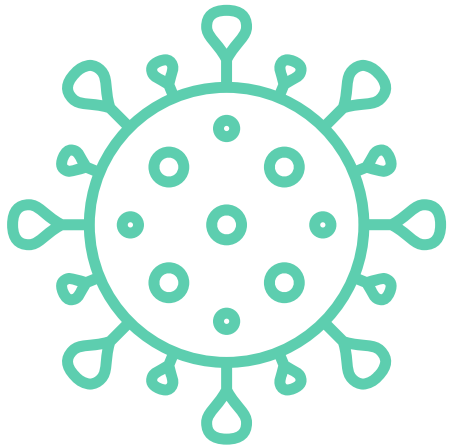
Data PLANT

Research Data Life Cycle

Data PLANT

Benefits of research data management

based on the example of Covid-19



Data PLANT

Your data as part of the big picture

Data PLANT

Your data as part of the big picture

Data PLANT

Research Data Management

Is part of Good Research Practice

Guidelines for Safeguarding Good Research Practice

Code of Conduct

https://zenodo.org/record/6472827

Data PLANT

What is Research Data Management?

What is Research Data Management?

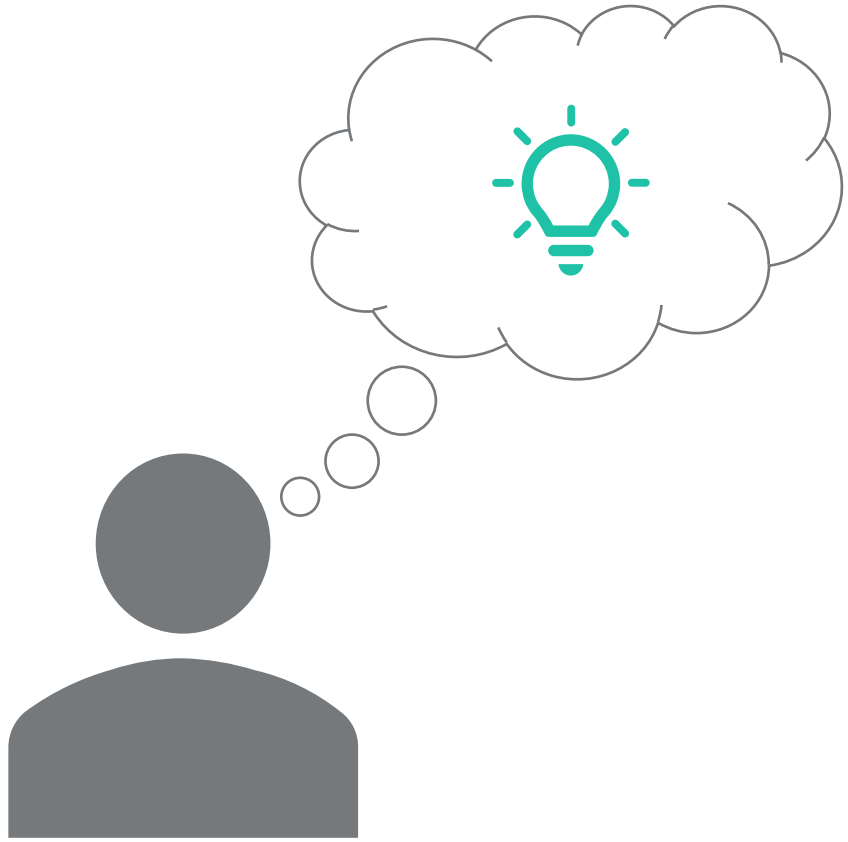
for each of us?

State something about the following:

- Who am I
- What area do I come from
- What data do I work with
- What does research data management mean for me
- What do I expect from this Summer School

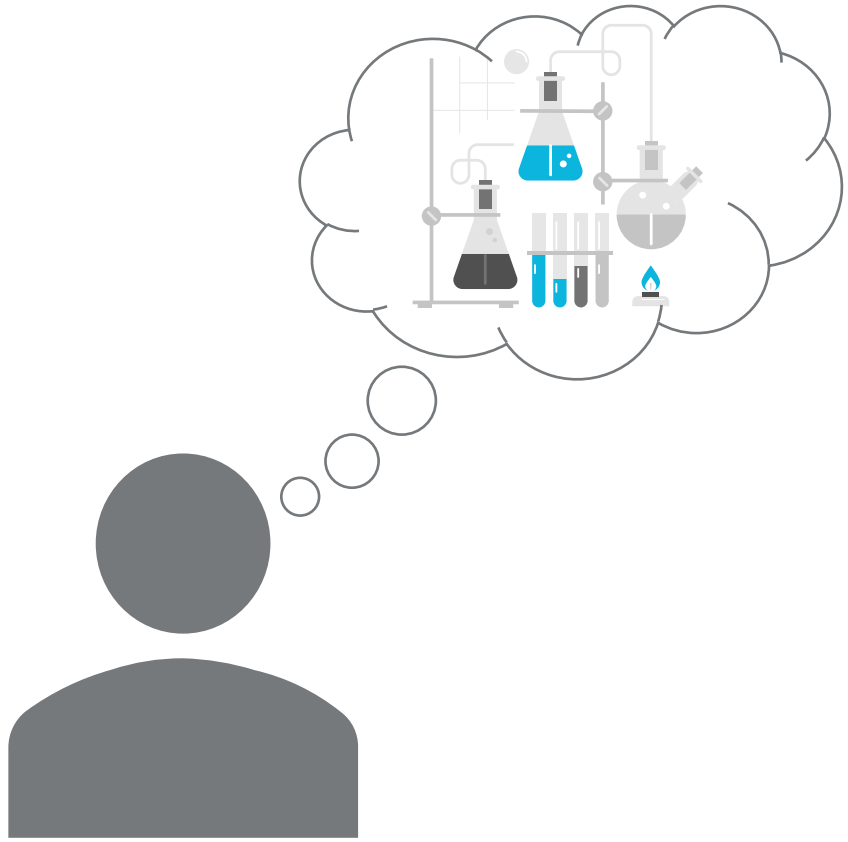
Reproducibility of results

is a fundamental aspect of scientific integrity



Reproducibility of results

is a fundamental aspect of scientific integrity



Data PLANT

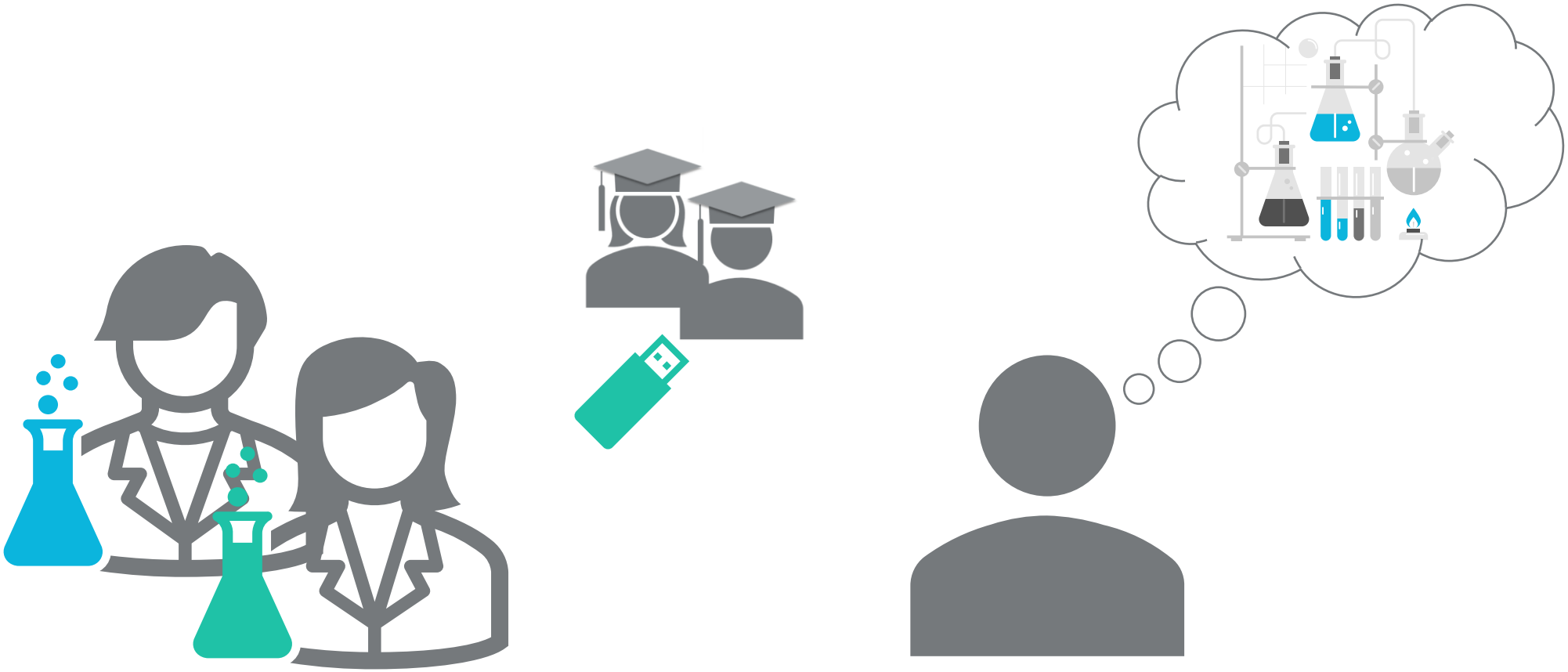
Reproducibility of results

is a fundamental aspect of scientific integrity

Data PLANT

Reproducibility of results

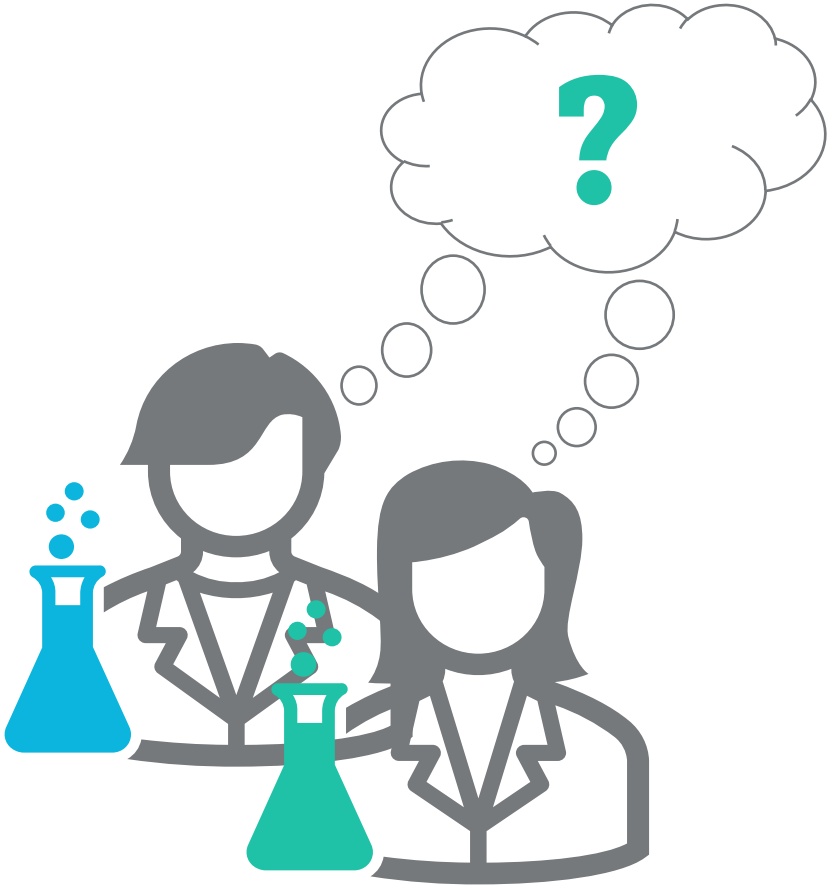
is a fundamental aspect of scientific integrity

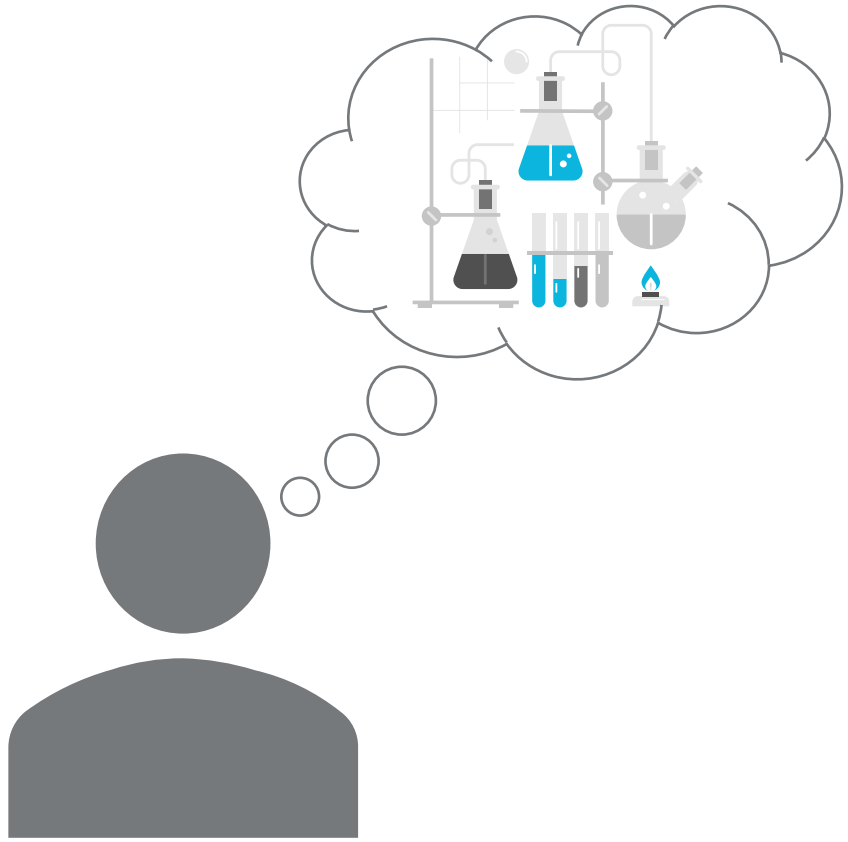


Data PLANT

Reproducibility of results

is a fundamental aspect of scientific integrity

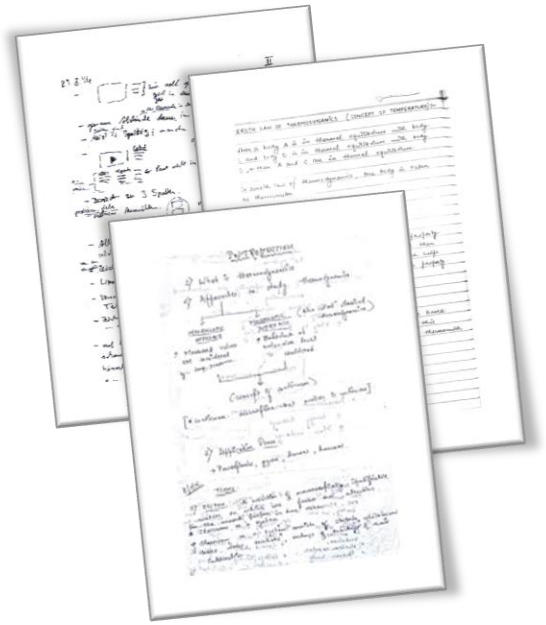




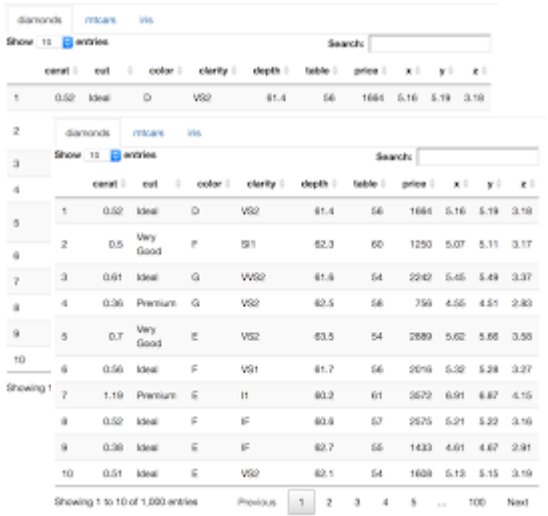
Reproducibility of results

is a fundamental aspect of scientific integrity

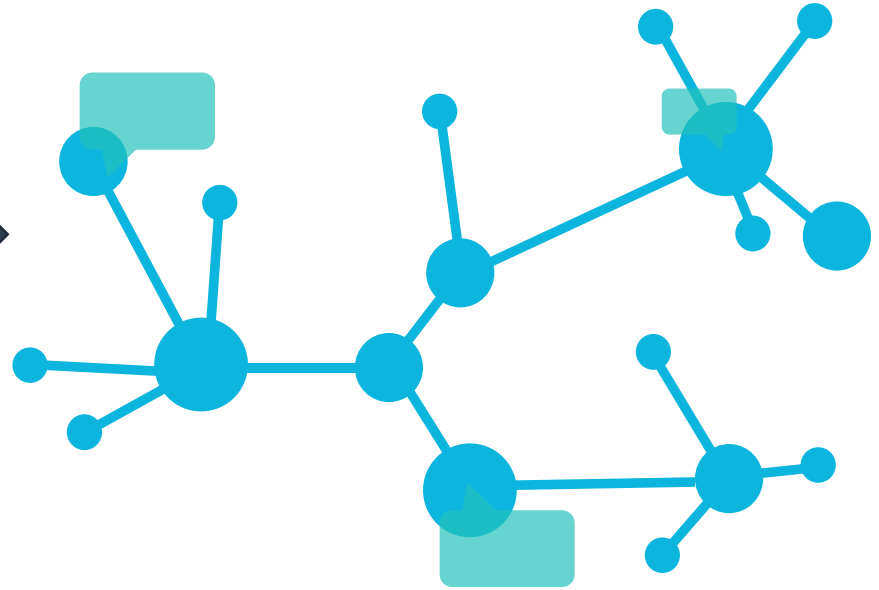
notes



digital information



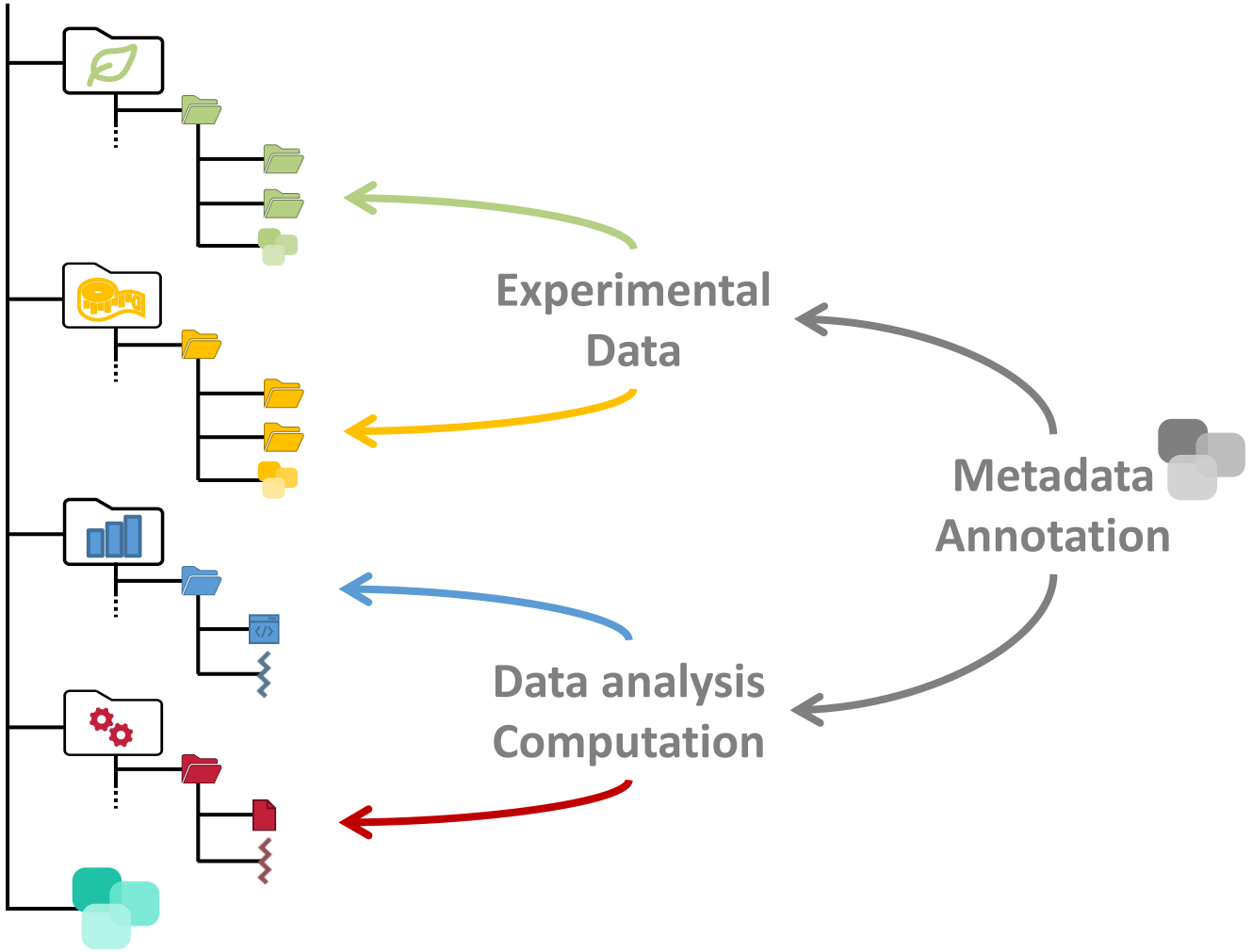
knowledge



Data PLANT

Improvement of Research Resources

by using standards and data organization



Data PLANT

Collaboration and Integration

on and of research data

Data PLANT

Long-term Archiving and Accessibility

of research data

storage

compute

authentication

elixir

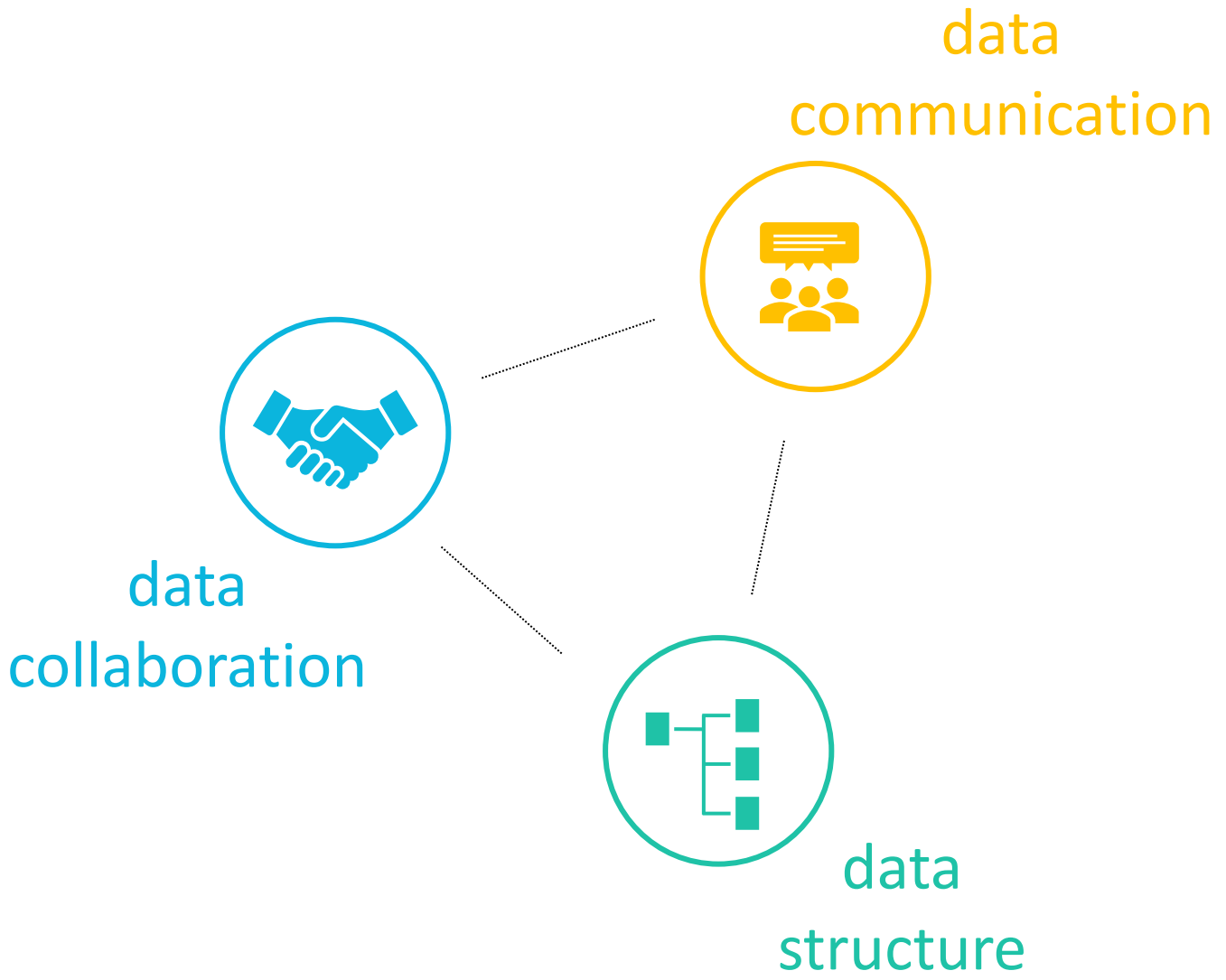
de.NBI
GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE

BinAC
Bioinformatik Astrophysik Cluster

Galaxy

Data PLANT

Research Data Management



Data PLANT

Research Data Management

FINDABLE

ACESSIBLE

INTEROPERABLE

REUSABLE

https://www.go-fair.org/fair-principles/

Data PLANT

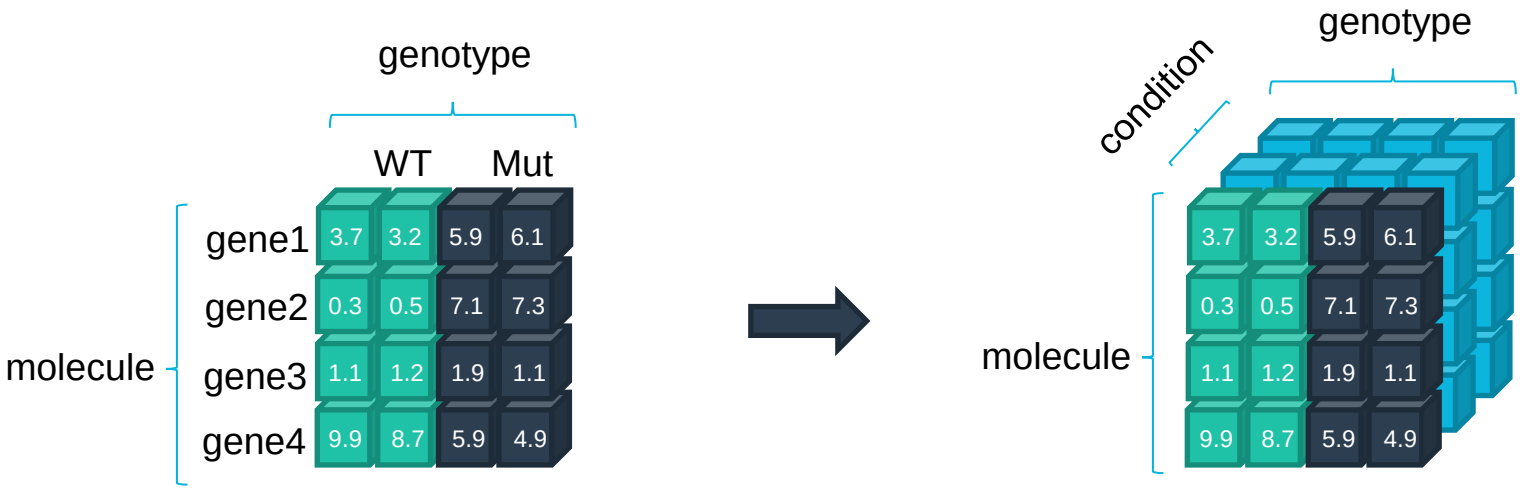
Essential elements for being FAIR

Metadata

Data PLANT

Essential elements for being FAIR

Metadata



Additional measurements

=

Knowledge gain

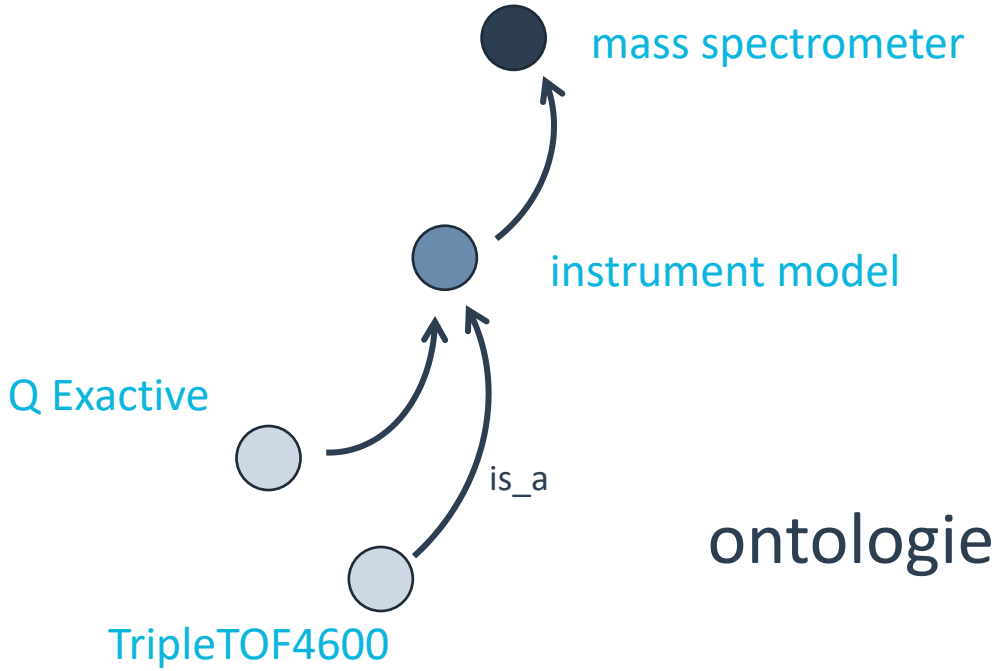
Data PLANT

Essential elements for being FAIR

Ontologies

dictionaries

=



ontologies

Meta data need to “speak“ the same language

Data PLANT

Findable means…

(Meta)data should be easy to locate for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier
  - Stability to locate and cite the data
  - Persistent identifier (PID) e.g. a Digital Object Identifier (DOI)

- F2. Data are described with rich metadata (see also R1.)
  - Necessary context information for the correct interpretation
  - Administrative metadata (author, license, etc.)

Data PLANT

ORCID
Connecting Research and Researchers

Findable means…

(Meta)data should be easy to locate for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

- F3. (Meta)data are registered or indexed in a searchable resource
  - Metadata are used to create easily searchable indexes of records
  - Basis for an information platform
- F4. Metadata specify the data identifier
  - The association between the metadata and the dataset is established by identifying the PID in the metadata

Accessible means…

Data and metadata should be archived long-term and made available by standard communication protocols for easy retrieval by machines and humans.

- A1. (Meta)data are retrievable by their identifier using a standardized communications protocol
  - Data retrieval should be possible without specialized tools
    A1.1 The protocol is open, free, and universally implementable
    A1.2 The protocol allows for an authentication and authorization procedure
- A2. Metadata are accessible, even when the data are no longer available
  - Dataset need a landing page

Interoperable means…

Data should be available in a format that allows exchange, interpretation and (semi-)automatic combination with other data.

- I1. (Meta)data use a formal, shared, and broadly applicable language for knowledge representation
  - Using controlled vocabularies/ontologies/thesauri and a clearly defined framework e.g., in the sense of the Semantic Web
  - Data need to interoperate with applications or workflows for analysis, storage, and processing
- I2. (Meta)data use vocabularies that follow FAIR principles
  - Controlled vocabularies must be properly identified (referenceable) and documented

Interoperable means…

Data should be available in a format that allows exchange, interpretation and (semi-)automatic combination with other data.

- I3. (Meta)data include qualified references to other (meta)data
  - Relationships between datasets must be specified in the (meta)data
  - Datasets must be referenced and linked

Reusable means…

Metadata and data should be well-described so that they can be replicated and/or combined in different settings.

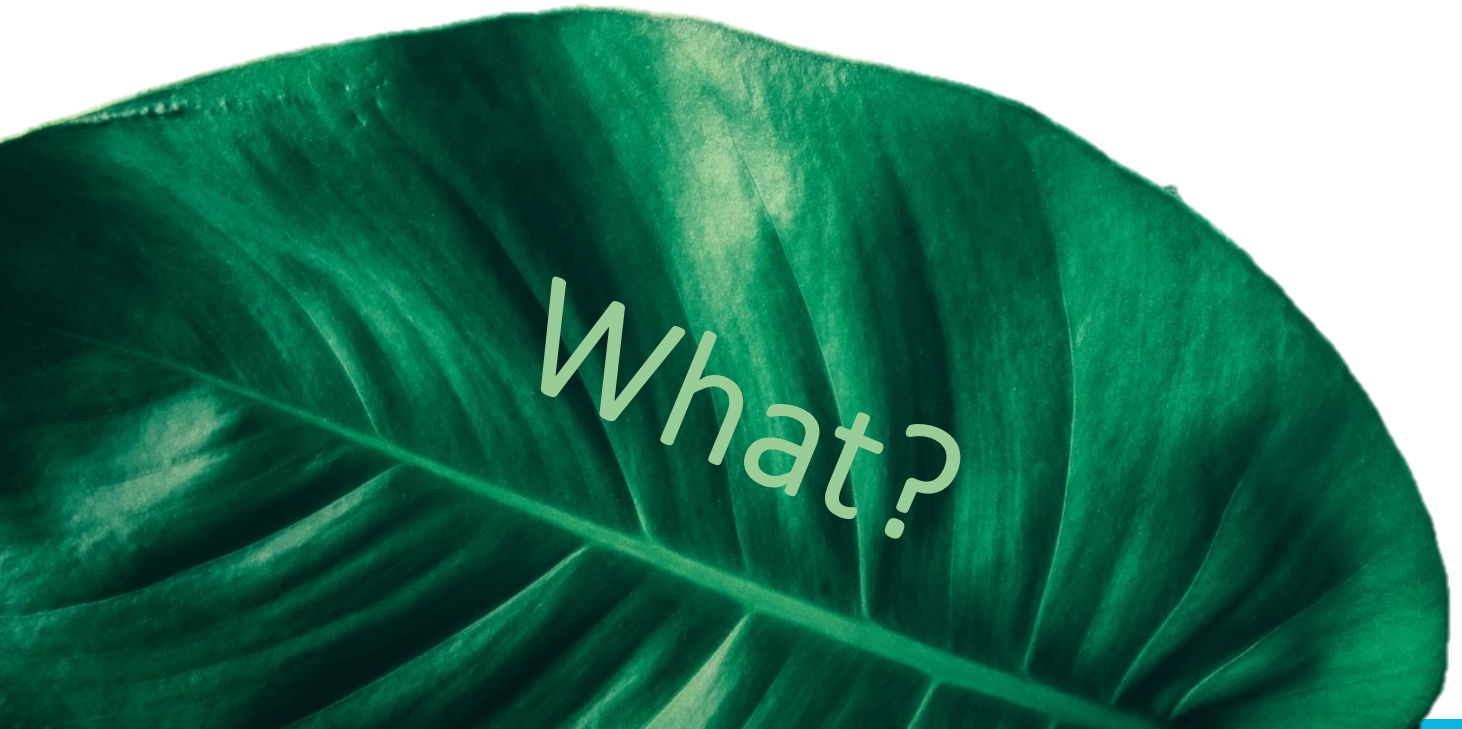
- R1. Meta(Data) have a plurality of accurate and relevant attributes
  - Extensive description of the data generation process. This may include e.g. experimental protocols, the manufacturer of the machine or sensor used for data creation, the software used for analyses, etc.

  - R1.1: (Meta)data are released with a clear and accessible data usage license
  - R1.2: (Meta)data are associated with detailed provenance
  - R1.3. (Meta)data meet **domain-relevant community standards**


Metadata
GEO
Gene Expression Omnibus
ENA
Repository
European Nucleotide Archive
MetaboLights
PRIDE Archive
PRoteomics IDEntifications Database
nextflow
Workflow
Galaxy
EUROPE
Ontology

Data PLANT


Metadata
GEO
Gene Expression Omnibus
ENA
Repository
European Nucleotide Archive
MetaboLights
PRIDE Archive
PRoteomics IDEntifications Database
nextflow
Workflow
Galaxy
EUROPE
Ontology
What?


What?


Where?


How?
Where?
What?

How?

Where?

What?


How?
Where?
What?

Nationale Forschungsdaten Infrastructure (NFDI)

National Research Data Infrastructure (NFDI)

National Research Data Infrastructure

2016

RfII proposal for the foundation of a NFDI

2018

Agreement of the federal and state governments (BLV) on the establishment of a NFDI

2020

Foundation of National Research Data Infrastructure (NFDI) Association

Data PLANT

National Research Data Infrastructure

https://www.youtube.com/watch?v=uJ01g9m8uE4

Data PLANT

"What is the National Research Data Infrastructure (NFDI)?" YouTube, uploaded by DFG bewegt, 7 Oct 2021, www.youtube.com/watch?v=uJ01g9m8uE4

The Consortia

Humanities and Social Sciences

- BERD@NFDI: NFDI for Business, Economic and Related Data
- KonsortSWD: Consortium for the Social, Educational, Behavioural and Economic Sciences
- NFDI4Culture: Consortium for Research Data on Material and Immaterial Cultural Heritage
- NFDI4Memory: The Consortium for the Historically Oriented Humanities
- NFDI4Objects – Research Data Infrastructure for the Material Remains of Human History
- Text+: Language and text-based research data infrastructure

Engineering Sciences

- NFDI4DataScience: NFDI für Datenwissenschaften und Künstliche Intelligenz
- NFDI4Energy: Nationale Forschungsdateninfrastruktur für die interdisziplinäre Energiesystemforschung
- NFDI4Ing: NFDI für die Ingenieurwissenschaften
- NFDI-MatWerk: NFDI für Materialwissenschaft & Werkstofftechnik
- NFDIxCS – Nationale Forschungsdateninfrastruktur für und mit Computer Science

The Consortia

Life Sciences

- DataPLANT: Plant research data
- FAIRagro: FAIR Data Infrastructure for Agrosystems
- NFDI4Immuno – National Research Data Infrastructure for Immunology
- GHGA: National Research Data Infrastructure for Immunologyv
- NFDI4Biodiversity: Biodiversity, Ecology and Environmental Data
- NFDI4BIOIMAGE: National research data infrastructure for microscopy and bioimage analysis
- NFDI4Health: NFDI personal health data
- NFDI4Microbiota: NFDI for Microbiota Research

Natural Sciences

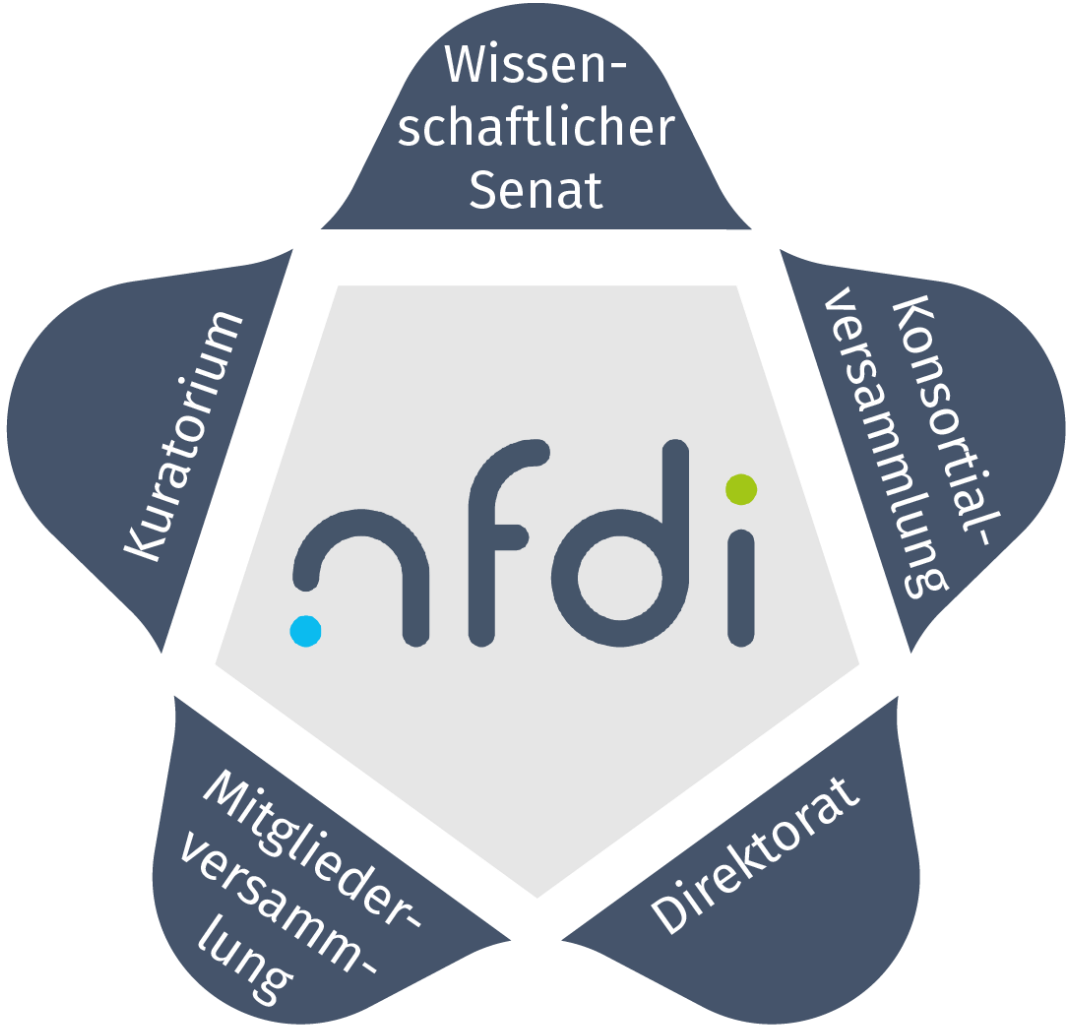
- DAPHNE4NFDI: Data from PHoton and Neutron Experiments for NFDI
- FAIRmat: FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids
- NFDI4Cat: NFDI for sciences related to catalysis
- MaRDI: Mathematical Research Data Initiative
- NFDI4Chem: Chemistry consortium for the NFDI
- NFDI4Earth: NFDI Consortium Earth System Sciences
- PUNCH4NFDI: Particles, Universe, NuClei and Hadrons for the NFDI

Basic Services

- Base4NFDI: Basic services for NFDI

The Association

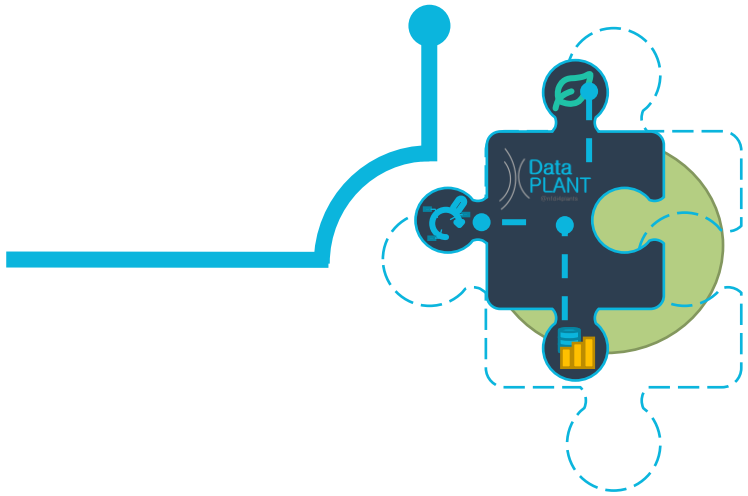
NFDI e.V.

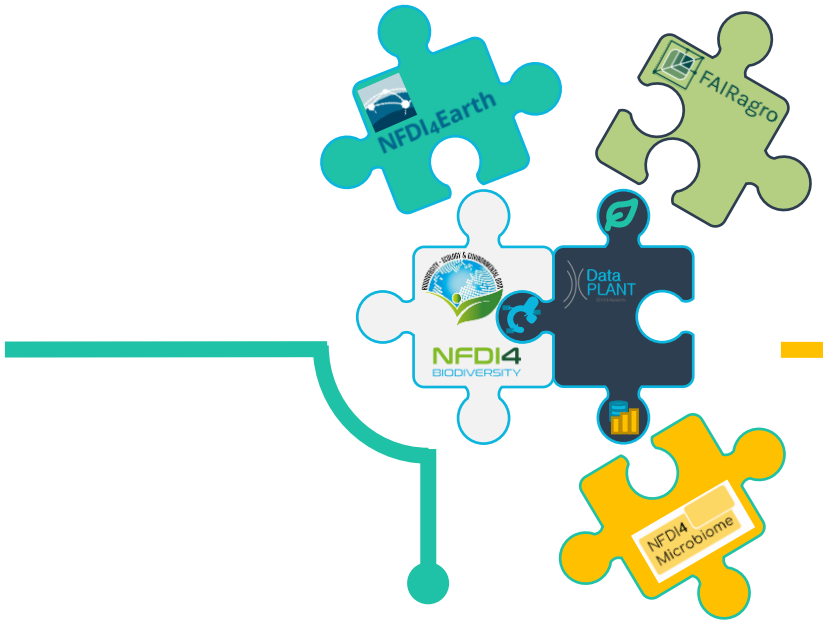


The Future

of the NFDI

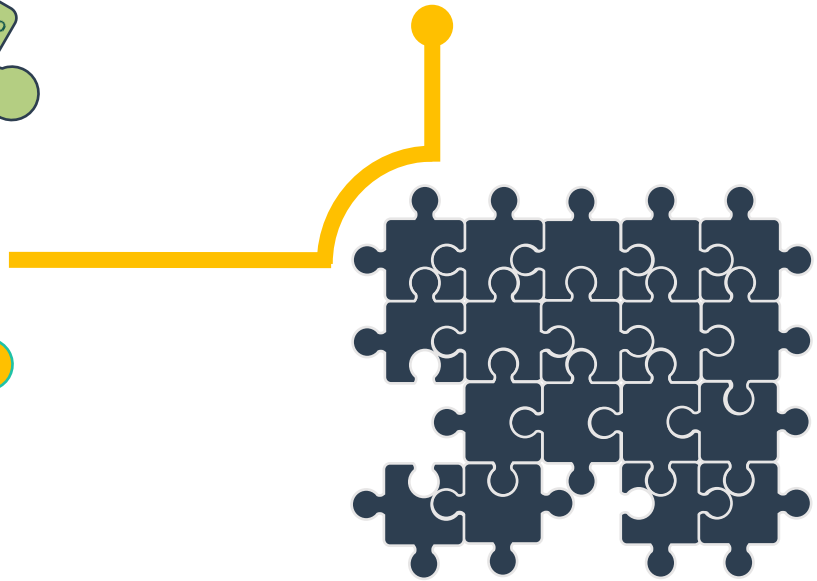
Capture the
own community





Docking "close"
communities

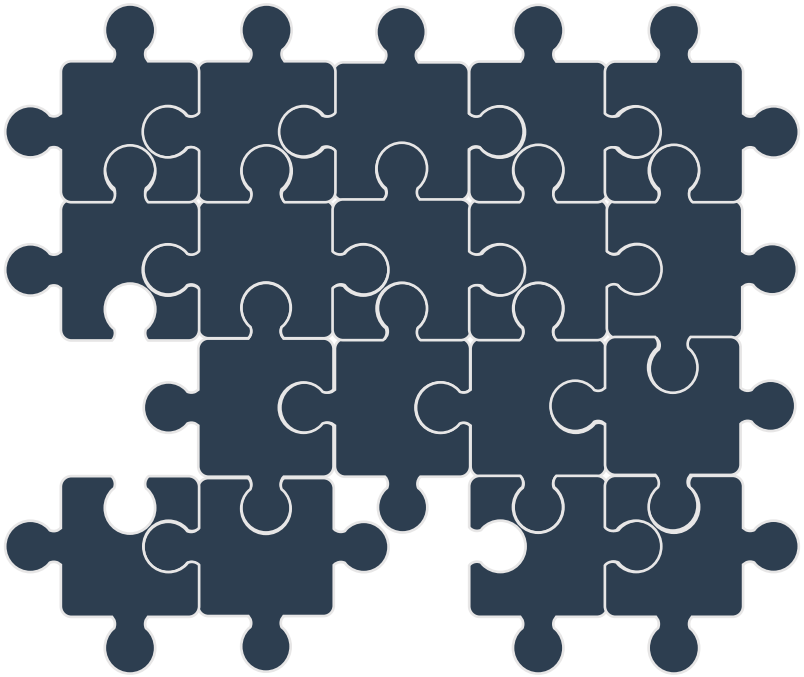
Broad coverage
by the NFDI



Data
PLANT

From formation to sustainability

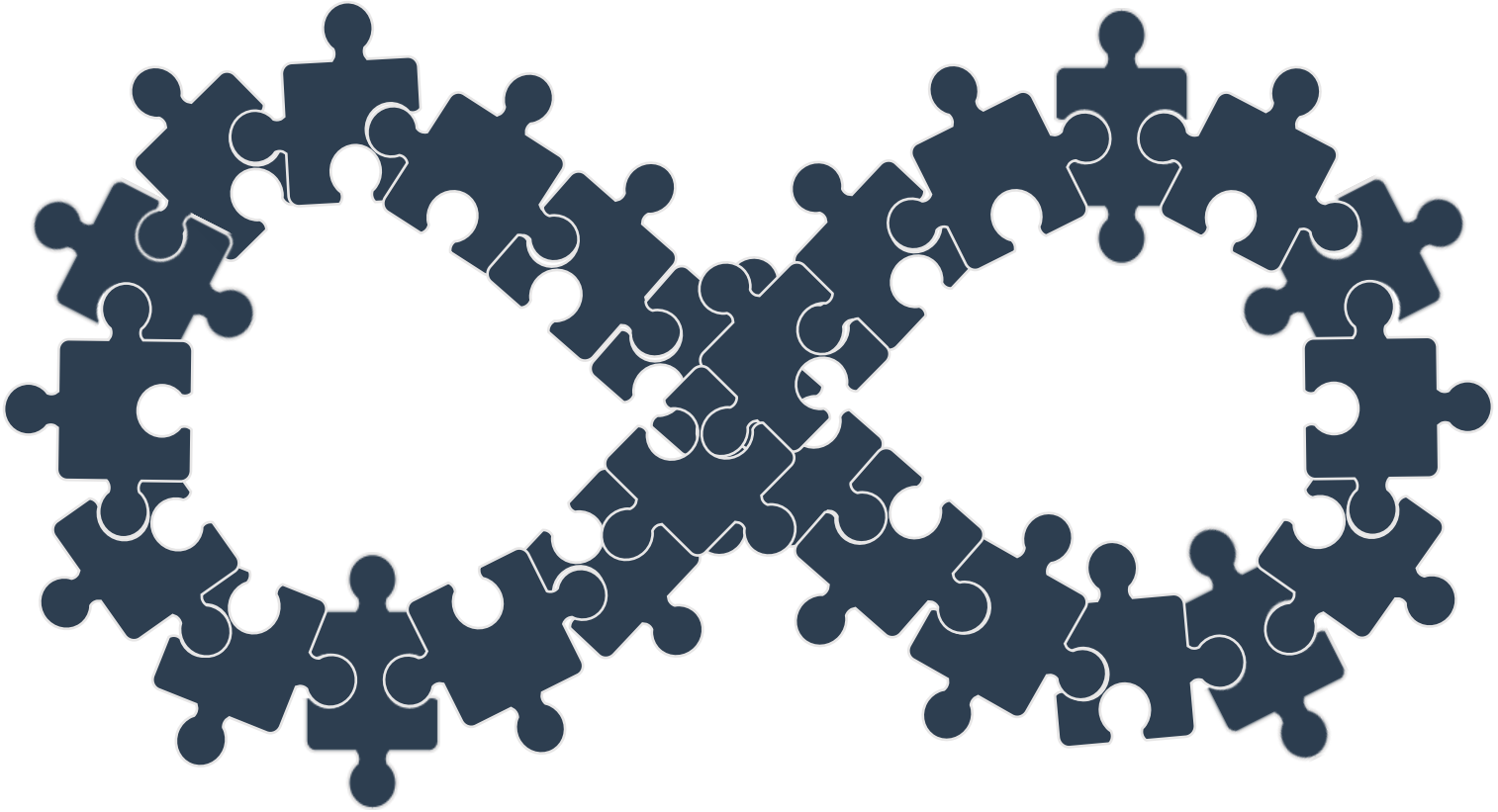
of the OneNFDI



Data PLANT

From formation to sustainability

of the OneNFDI



Data PLANT

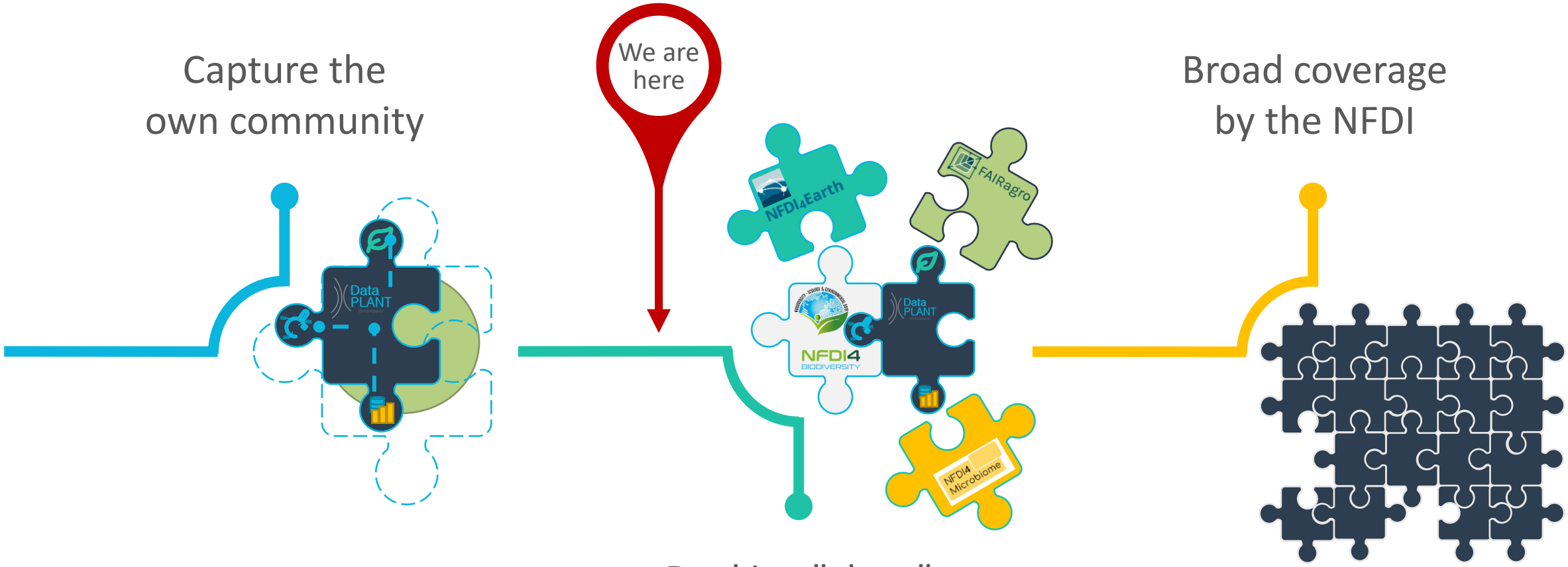
The Future

of the NFDI

Capture the own community

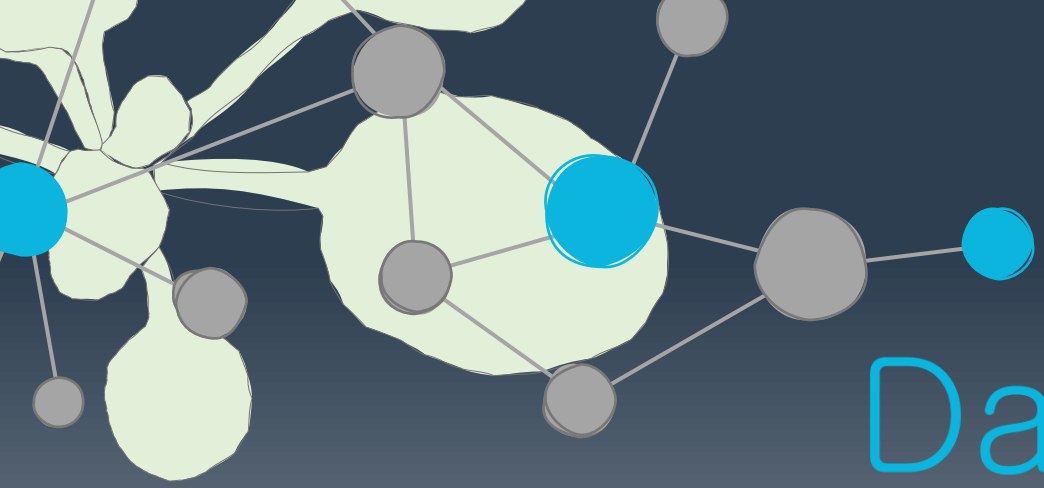
We are here

Broad coverage by the NFDI



Docking "close" communities

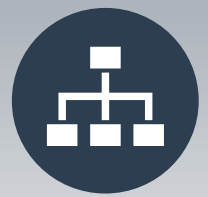
Data PLANT



DataPLANT

FAIR RESEARCH IN YOUR HANDS

Plan



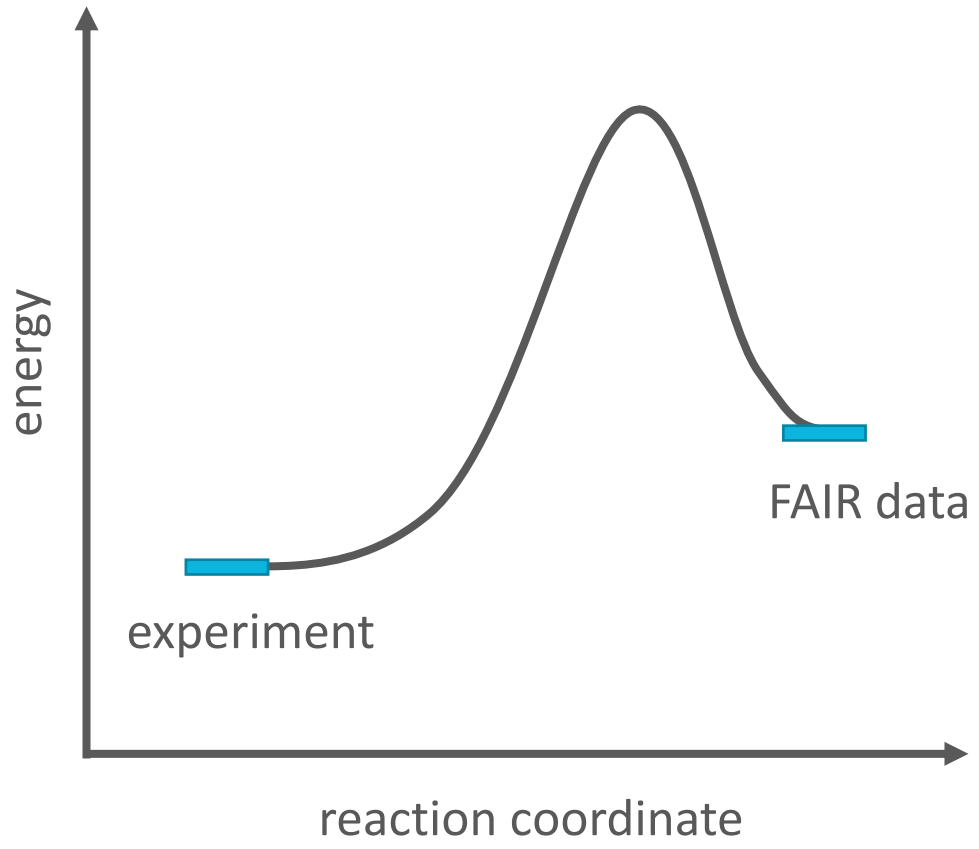
Structure

Manage

Share

Store

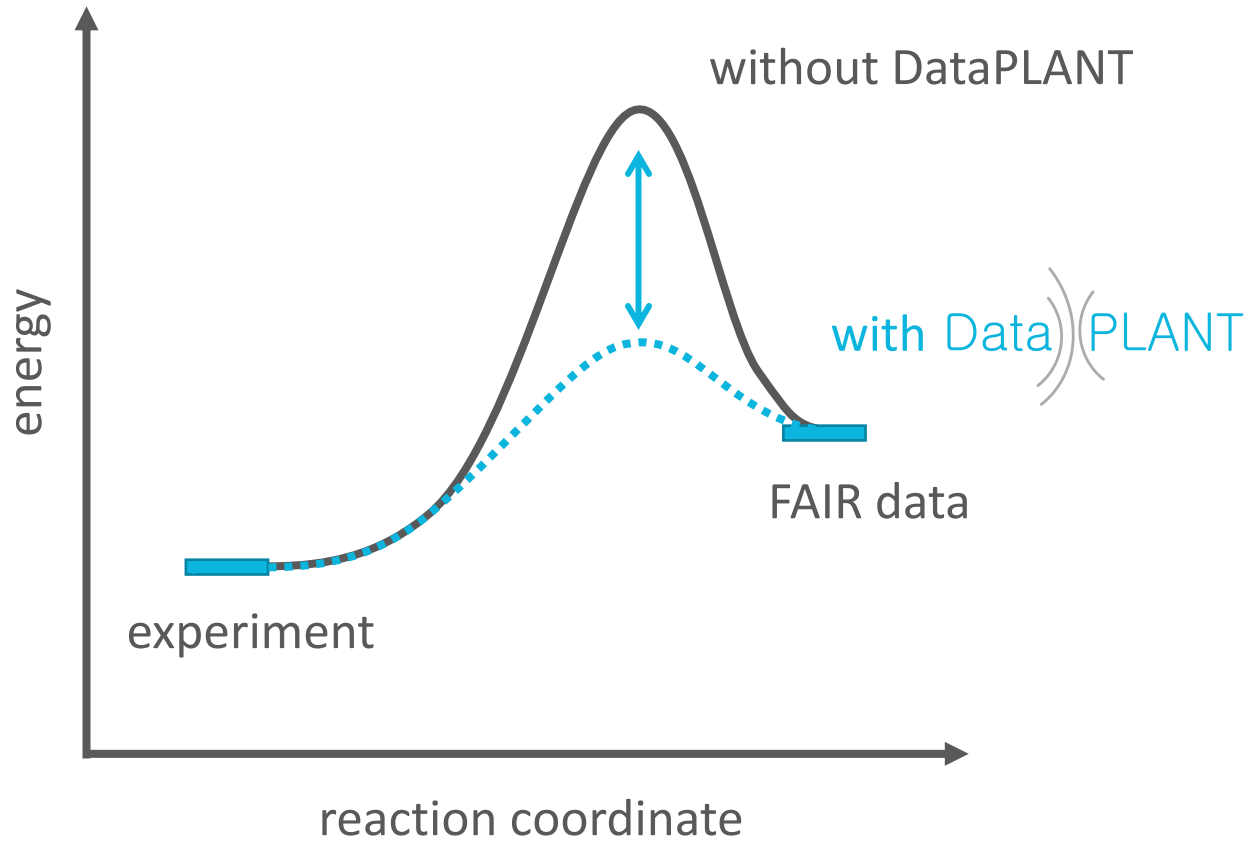
Going FAIR requires energy investment



Data PLANT

DataPLANT as an enzyme

lowering the activation energy.



Annotated Research Context (ARC)

a FAIR digital object implementation of DataPLANT

FINDABLE

ACCESSIBLE

REUSABLE

INTEROPERABLE

Metadata Templates

Controlled Vocabularies

Standards

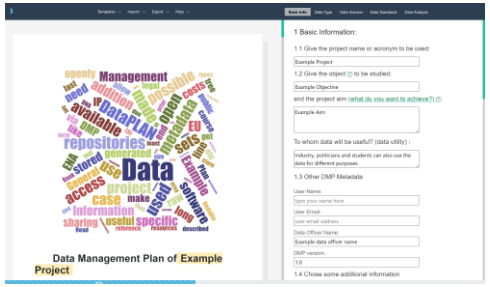
Data PLANT

Tools and Services

provide support in being FAIR

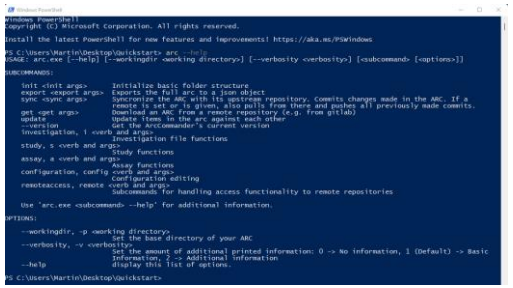
Data
Management
Planning

DATAPLAN



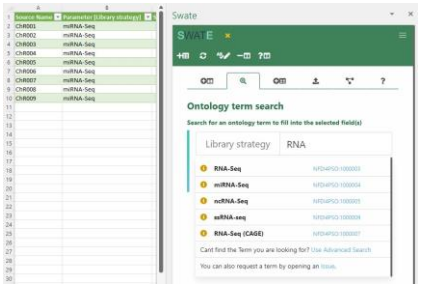
ARC Initialization
ARC Management

ARC COMMANDER



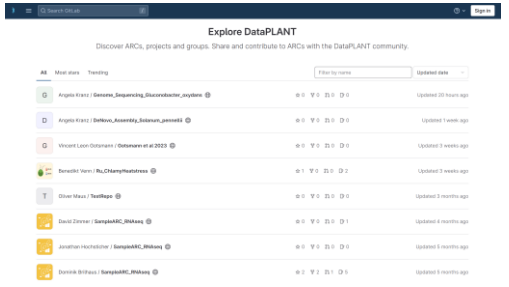
Metadata Annotation
Term Finding
Template Finding

SWATE



Storage
Sharing
Version Control

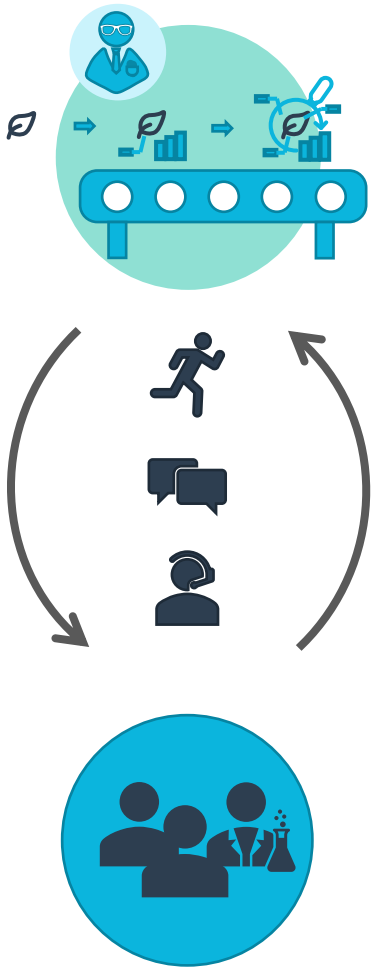
DATAHUB



Data PLANT

Let´s be FAIR together

https://linktr.ee/DataPLANT



Helpdesk
https://helpdesk.nfdi4plants.org/

E-Mail
info@nfdi4plants.org

Website
https://www.nfdi4plants.org/

Organizational

Agenda

Covering the aspects of Research Data Management

	Monday (17.7.)	Tuesday (18.7.)	Wednesday (19.7.)	Thursday (20.7.)	Friday (21.7.)
09:00 - 10:30	Opening / Keynote *Rodrigues*	Versioning & Collaboration *Garth*	Organization and Archives *Wetzels*	Workflows and Data Processing *Garth / Kappe*	RDM Planning *Josch / Niederprüm*
	coffee break	coffee break	coffee break	coffee break	coffee break
11:00 - 12:30	Process Models *Mühlhaus*	Versioning & Collaboration *Garth*	Databases *Doniparthi*	Workflows and Data Processing *Garth / Kappe*	Closing + Open Session *all*
	lunch break	*lunch break*	*lunch break*	*lunch break*	
14:00 - 15:30	Process Models *Mühlhaus*	Galaxy *Gallardo Alba*	group hike & Bremerhof dinner	RDM in practice *Brilhaus*	

Data PLANT