

Fundamentals of ARCs

de.NBI RDM training – *Maximizing the Potential of Agricultural Research Data*

Dominik Brilhaus, [CEPLAS](#)

October 23rd, 2025



Introduce yourself

Please use zoom annotations

Used code / a programming language before? Experience with Git / GitLab / GitHub?

Operating system

- Windows
- Mac
- Linux

Experience with ARCs?

FAIR data stewardship

- Findable
- Accessible
- Interoperable
- Reusable

[nature](#) > [scientific data](#) > [comment](#) > [article](#)

[Open Access](#) | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

[Mark D. Wilkinson](#), [Michel Dumontier](#), [IJsbrand Jan Aalbersberg](#), [Gabrielle Appleton](#), [Myles Axton](#), [Arie Baak](#), [Niklas Blomberg](#), [Jan-Willem Boiten](#), [Luiz Bonino da Silva Santos](#), [Philip E. Bourne](#), [Jildau Bouwman](#), [Anthony J. Brookes](#), [Tim Clark](#), [Mercè Crosas](#), [Ingrid Dillo](#), [Olivier Dumon](#), [Scott Edmunds](#), [Chris T. Evelo](#), [Richard Finkers](#), [Alejandra Gonzalez-Beltran](#), [Alasdair J.G. Gray](#), [Paul Groth](#), [Carole Goble](#), [Jeffrey S. Grethe](#), [Jaap Heringa](#), [Peter A.C. 't Hoen](#), [Rob Hooft](#), [Tobias Kuhn](#), [Ruben Kok](#), [Joost Kok](#), [Scott J. Lusher](#), [Maryann E. Martone](#), [Albert Mons](#), [Abel L. Packer](#), [Bengt Persson](#), [Philippe Rocca-Serra](#), [Marco Roos](#), [Rene van Schaik](#), [Susanna-Assunta Sansone](#), [Erik Schultes](#), [Thierry Sengstag](#), [Ted Slater](#), [George Strawn](#), [Morris A. Swertz](#), [Mark Thompson](#), [Johan van der Lei](#), [Erik van Mulligen](#), [Jan Velterop](#), [Andra Waagmeester](#), [Peter Wittenburg](#), [Katherine Wolstencroft](#), [Jun Zhao](#) & [Barend Mons](#)✉

— Show fewer authors

[Scientific Data](#) 3, Article number: 160018 (2016) | [Cite this article](#)

<https://doi.org/10.1038/sdata.2016.18>

The FAIR principles



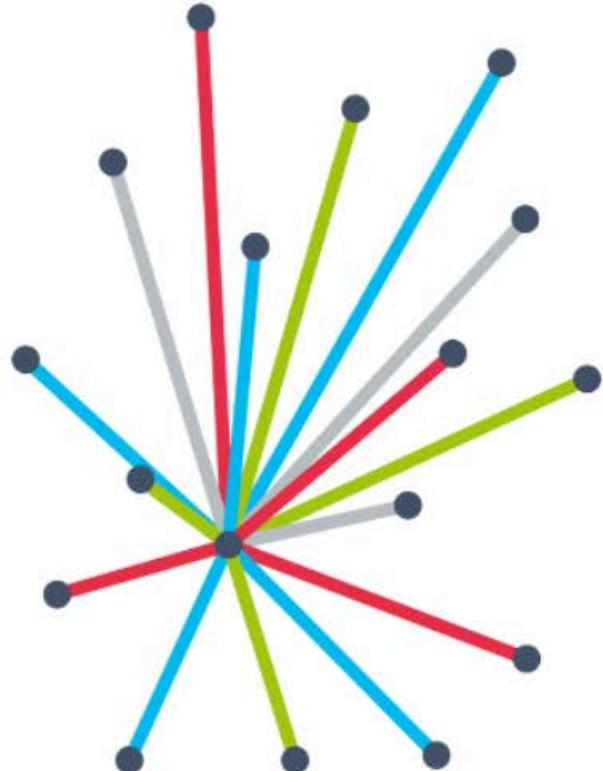
Is your data FAIR?

Findable | Accessible | Interoperable | Reusable

- Where do you store your data?
- How do you annotate your data?
- How do you share your data?
- What tools do you use to analyse your data?
- How do you reuse other people's data?



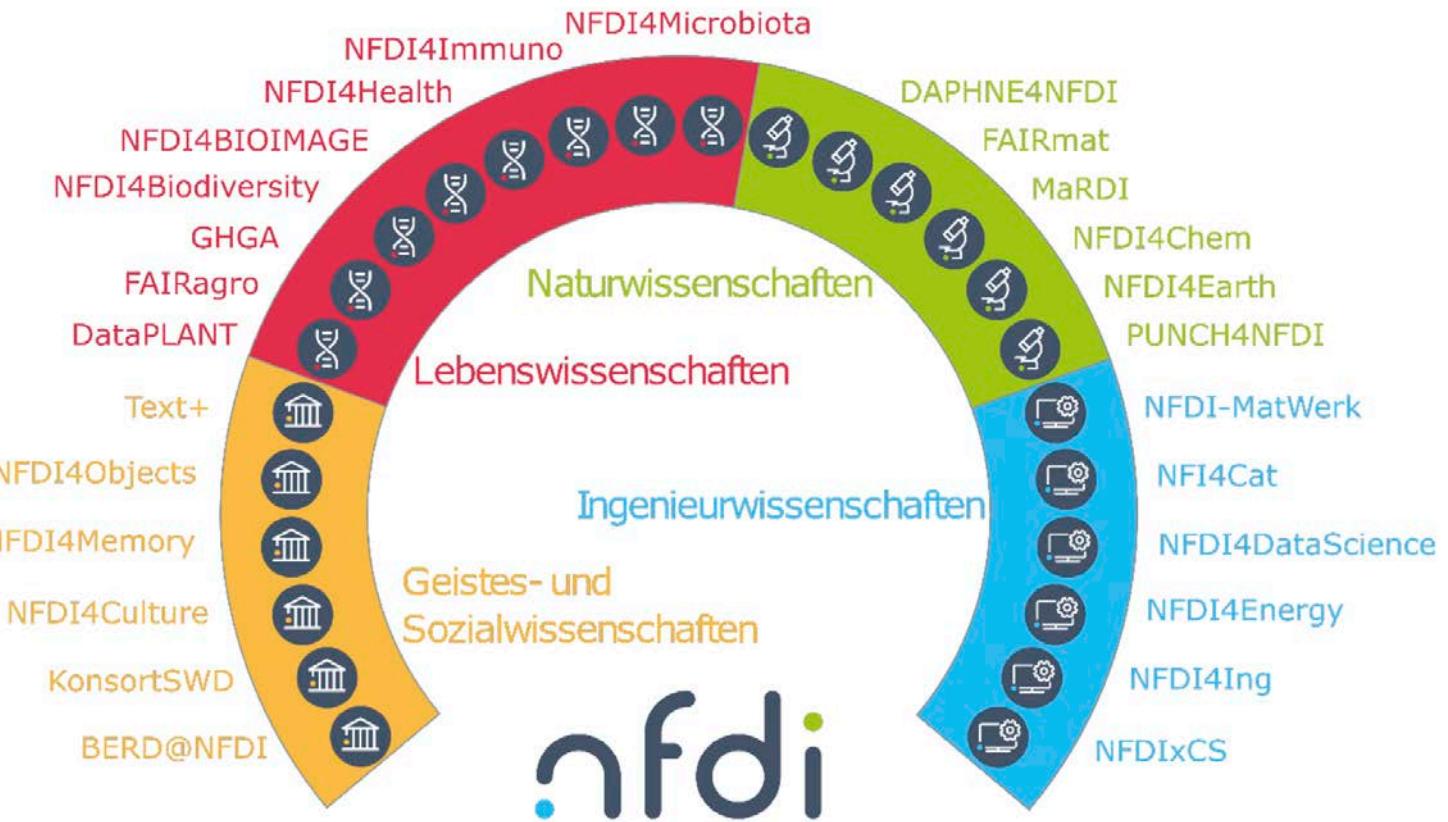
NFDI – Nationale Forschungsdaten Infrastruktur



The aim of the national research data infrastructure (NFDI) is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally.

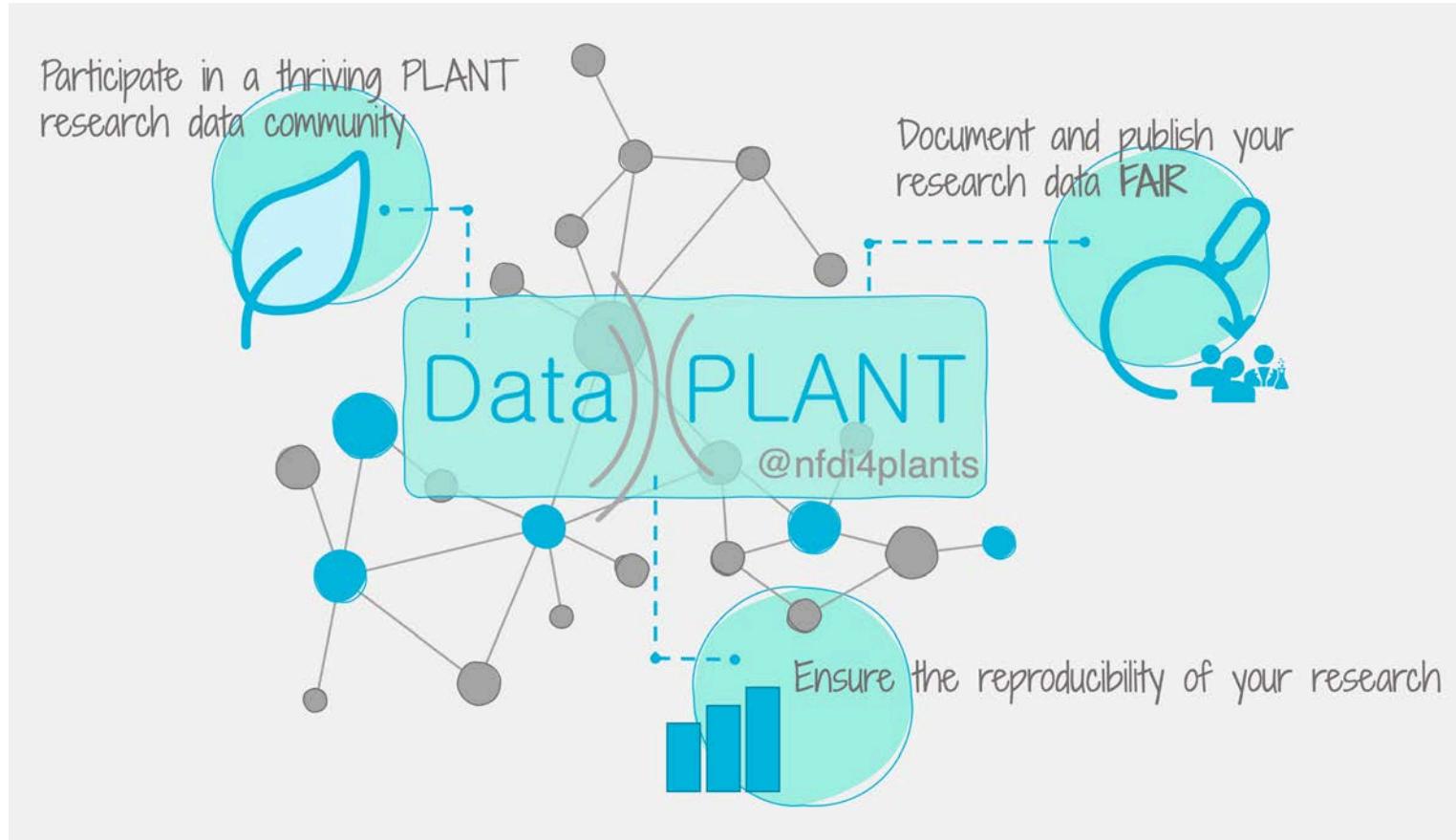
nfdi.de

NFDI – A community-targeted approach for RDM



nfdi.de

DataPLANT – NFDI4plants



<https://nfdi4plants.org>

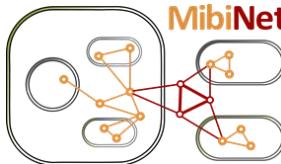
Data Stewardship between DataPLANT and the community

Community



CEPLAS

trr_341 plant
ecological
genetics



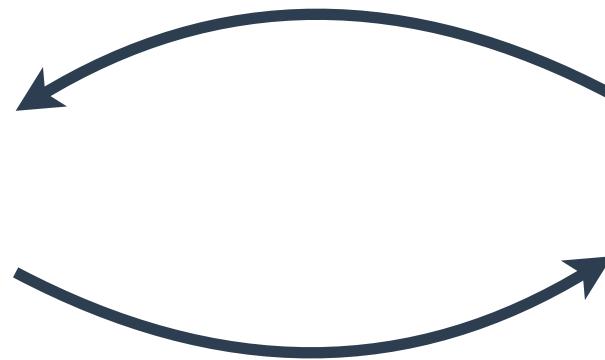
MibiNet

Domain experts
User experience
Training

nfdi4plants



Service provider
Developers
Tech experts



Resources



Info & materials

- DataPLANT Website: <https://nfdi4plants.org/>
- ARC website: <https://arc-rdm.org>
- Knowledge Base:
<https://nfdi4plants.org/nfdi4plants.knowledgebase/>

Tools and Services

- ARCitect: <https://github.com/nfdi4plants/arcitect>
- DataHUB: <https://git.nfdi4plants.org>

Continuous support

- HelpDesk: <https://helpdesk.nfdi4plants.org>
- Matrix for ad hoc support: <https://matrix.to/#/%23arc-user-support:matrix.org>
- User Support Meeting (2nd Friday of the month | 1 – 2pm):
<https://nfdi4plants.github.io/events/arc-user-support/>
- User Support Mailing List: [Click here to subscribe](#)

Open Source Development

- GitHub: <https://github.com/nfdi4plants>

Interaction during the course

Google Doc

- Please use the [google doc](#) to interact, raise questions and feedback
- Links for hands-on sections

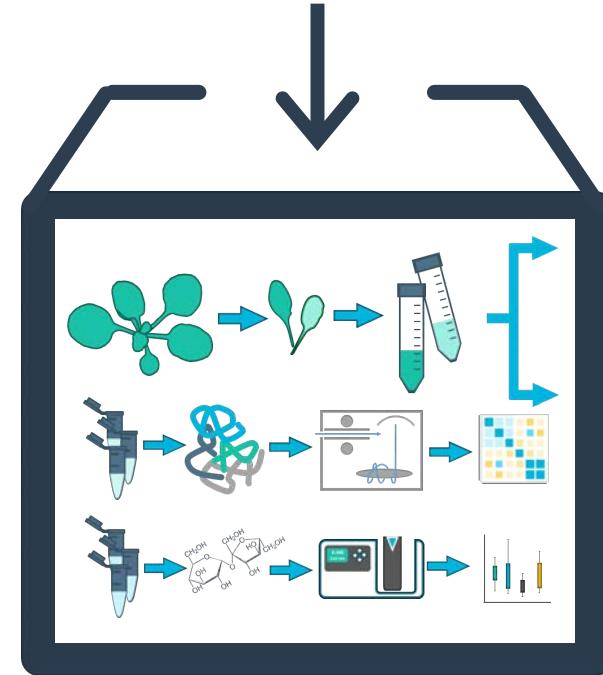
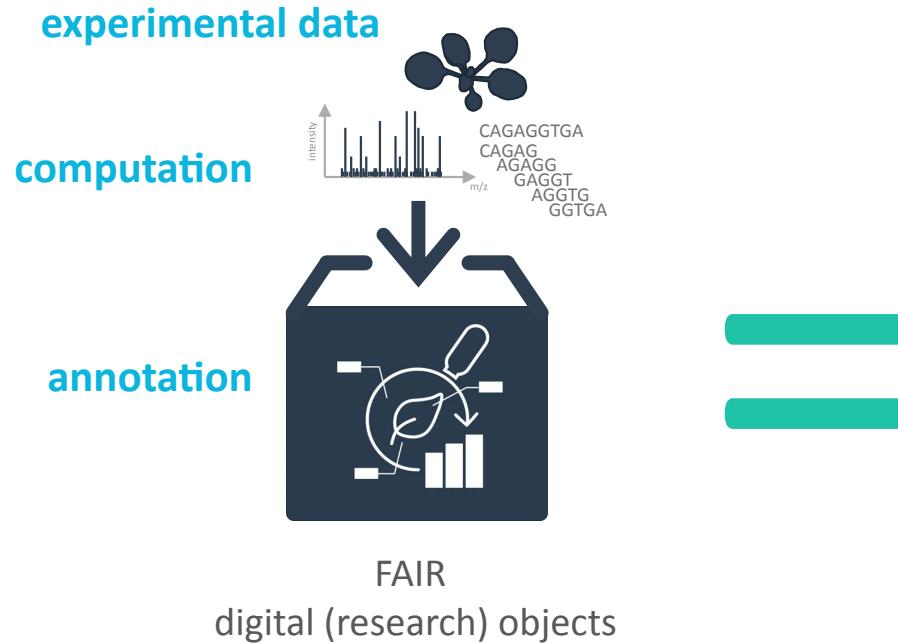
"I'm ready"

During the self-guided parts, please use Zoom reactions to tell me you're done with that part.

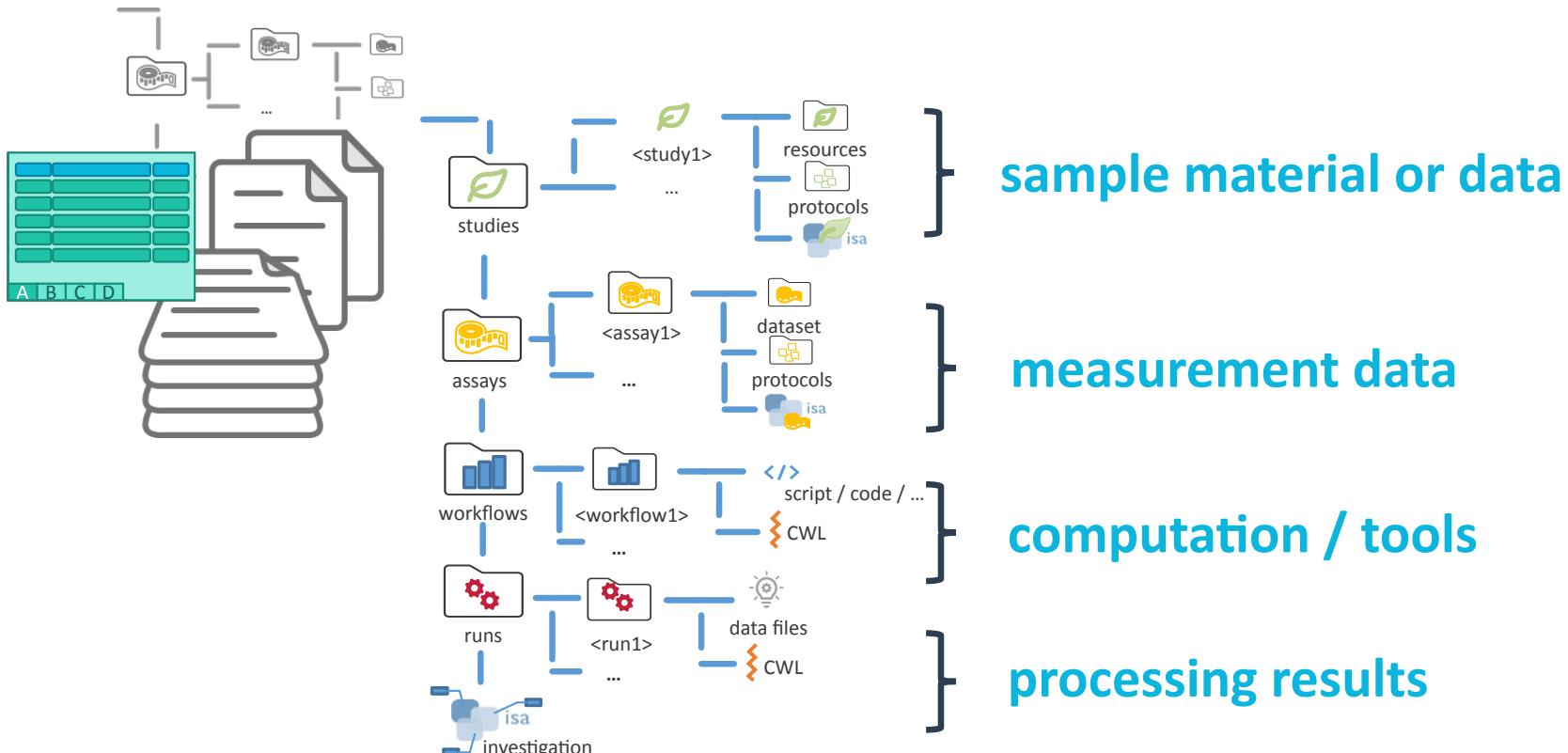


Yes

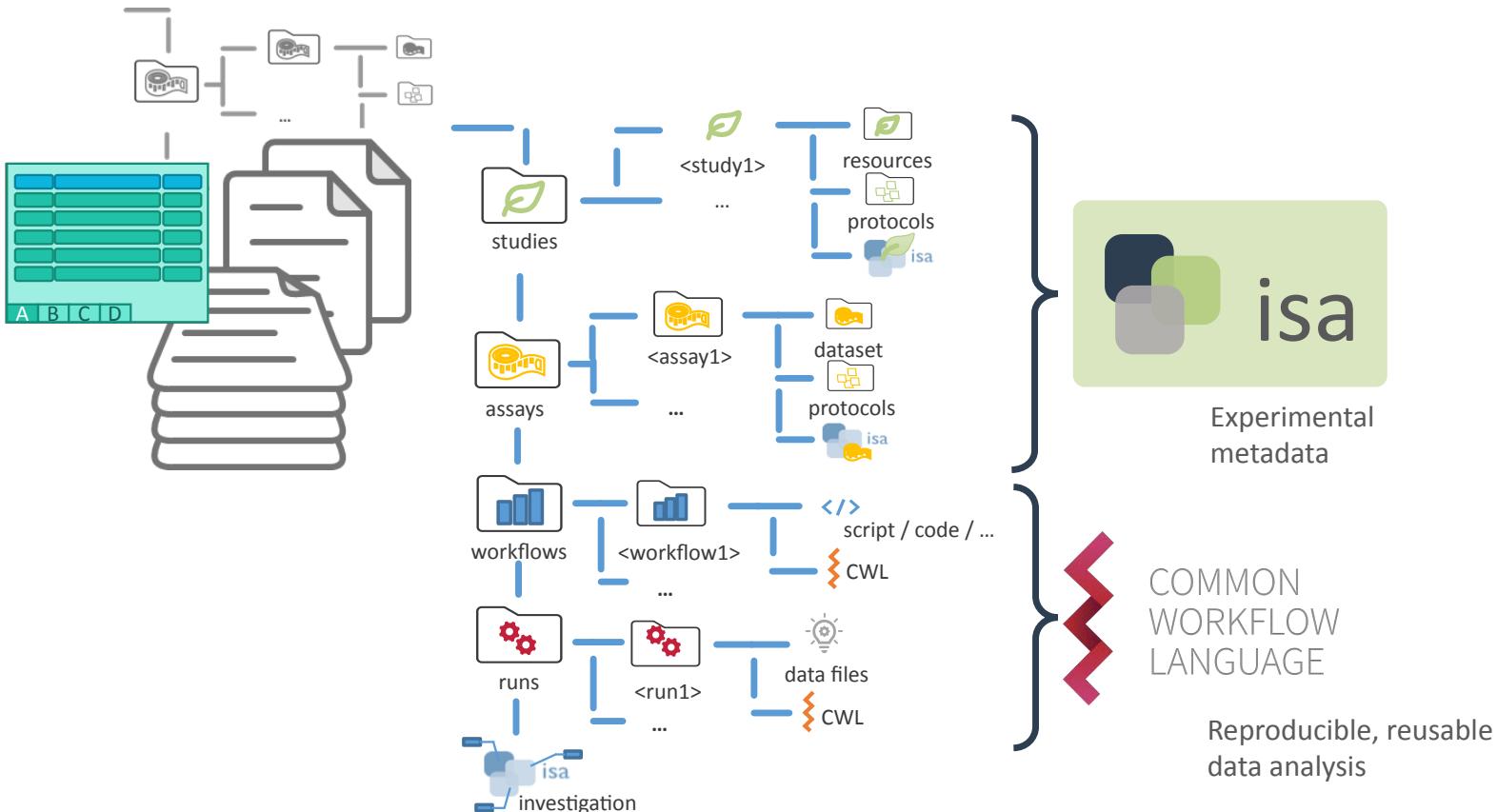
Annotated Research Context (ARC)



The ARC scaffold structure



ARC builds on standards



<https://isa-tools.org> | <https://www.commonwl.org> | <https://www.researchobject.org/ro-crate> | <https://git-scm.com>

ISA abstract model in a nutshell



isa

Investigation
administrative (meta)data

- Summary
 - Titel
 - Description
- Person
- Organisation
- Publication reference

Study
descriptive (meta)data
information on the subject

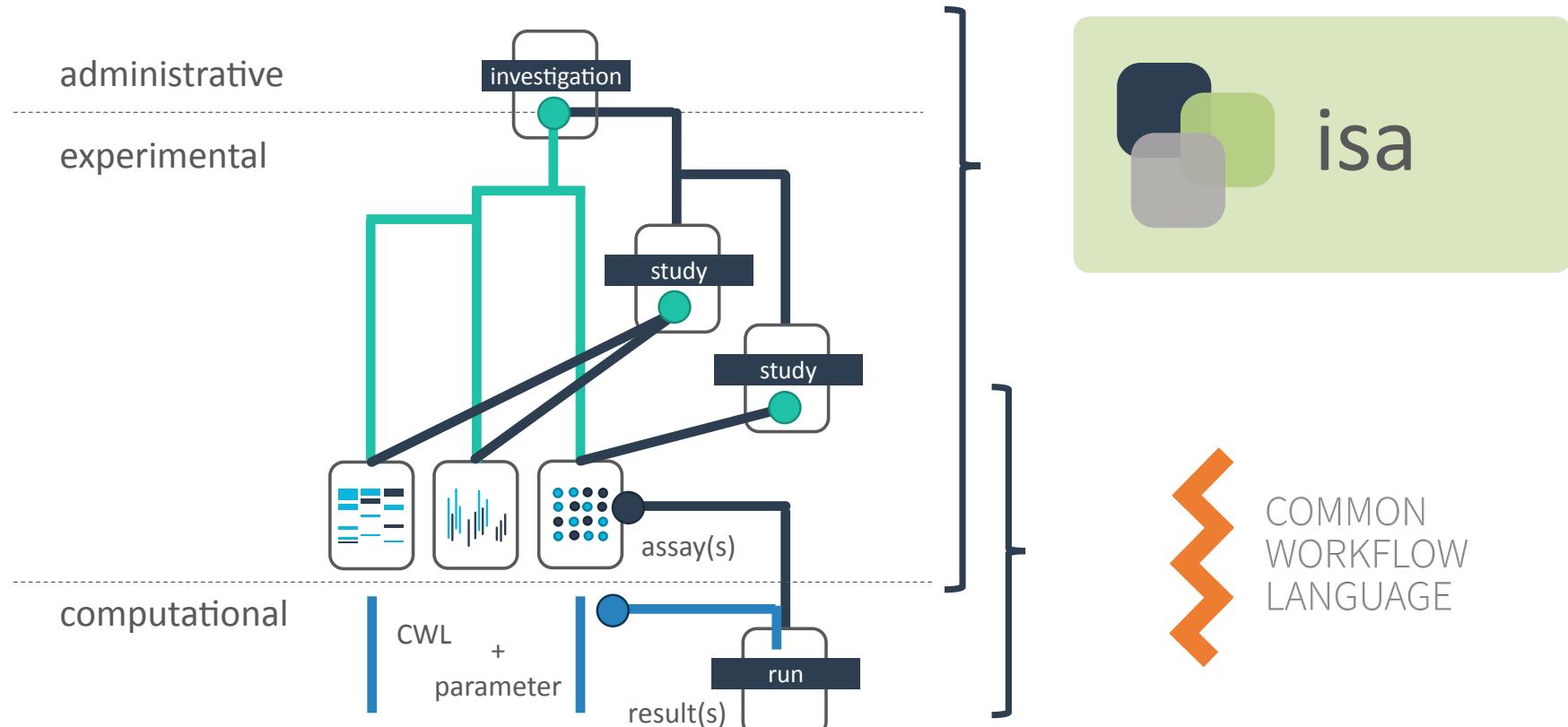
- Characteristics
- Parameters
- Components
- Factors

Assay

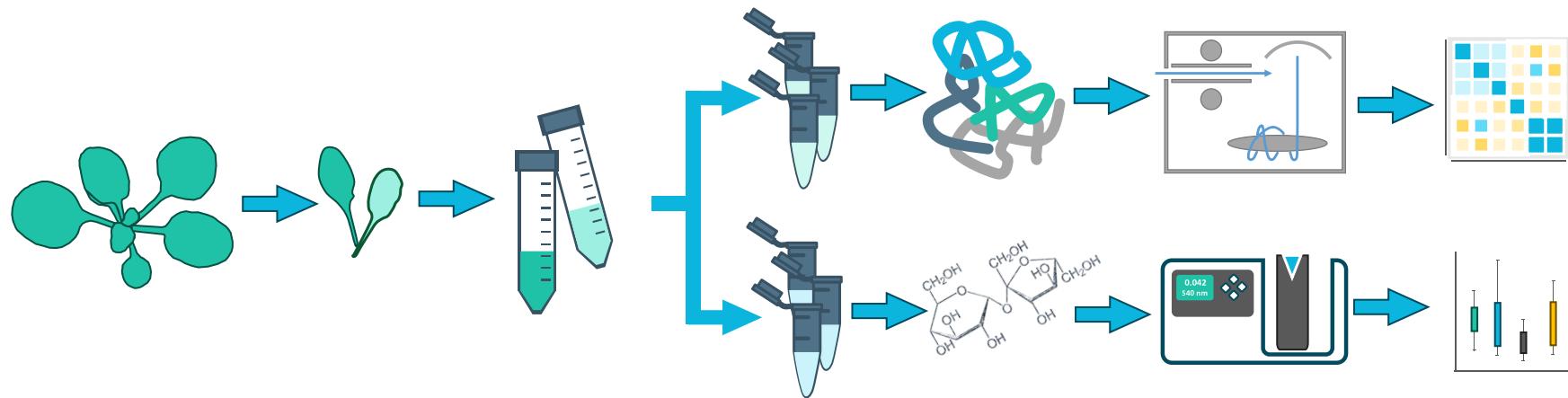
descriptive (meta)data
information on the measurement

- Characteristics
- Parameters
- Components
- Factors

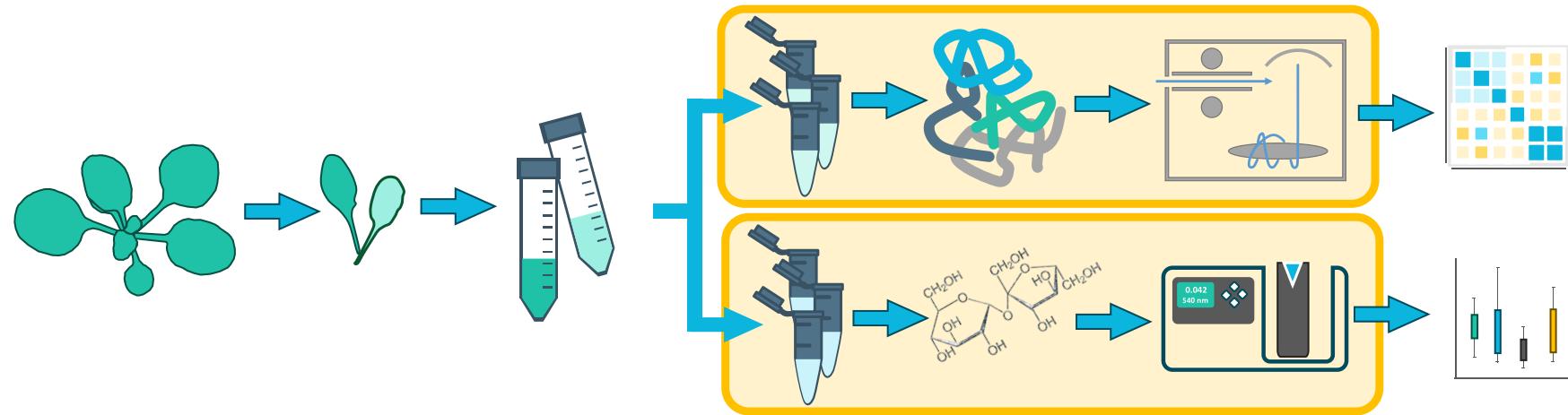
ISA and CWL – Connected by similarity



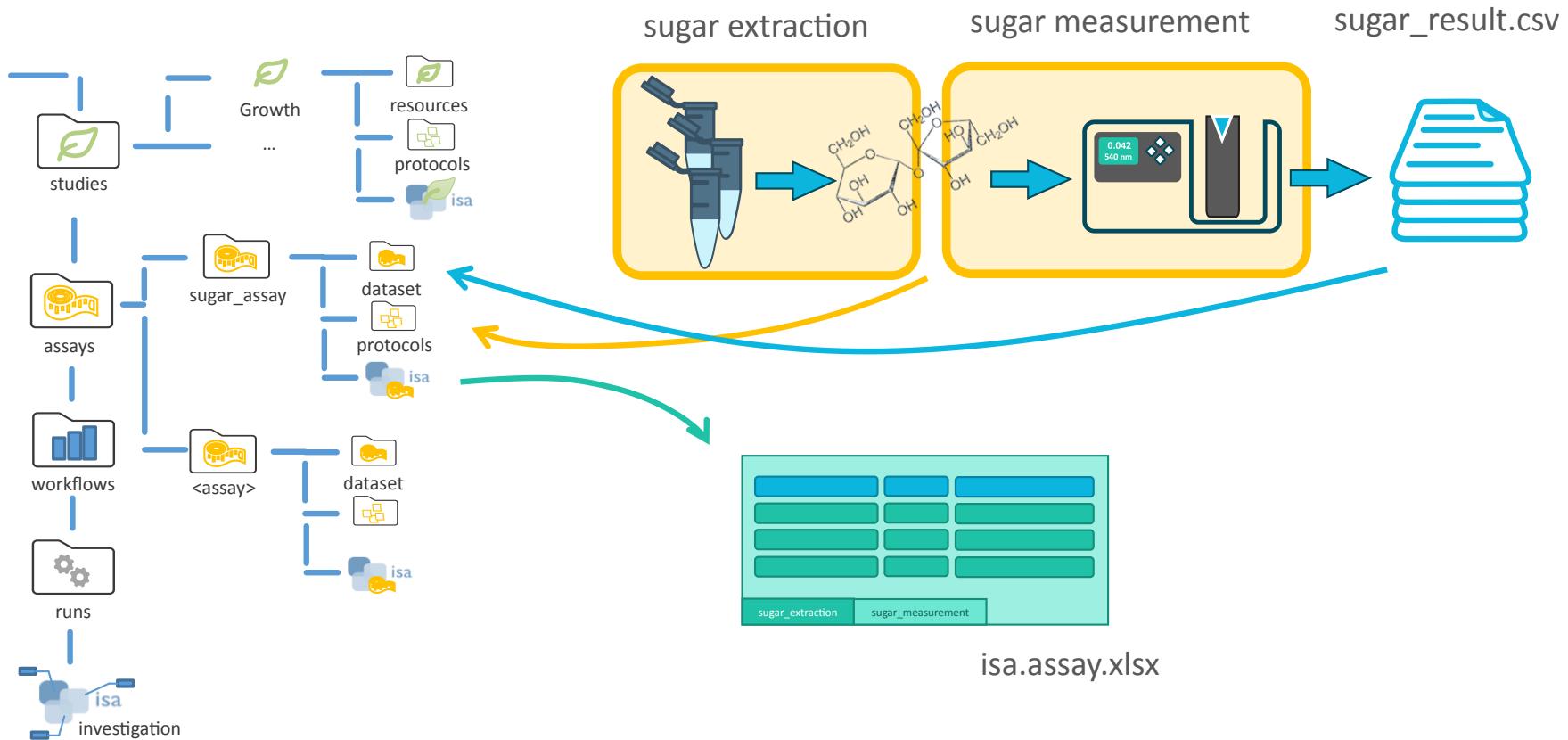
Metadata annotation – from sample to data



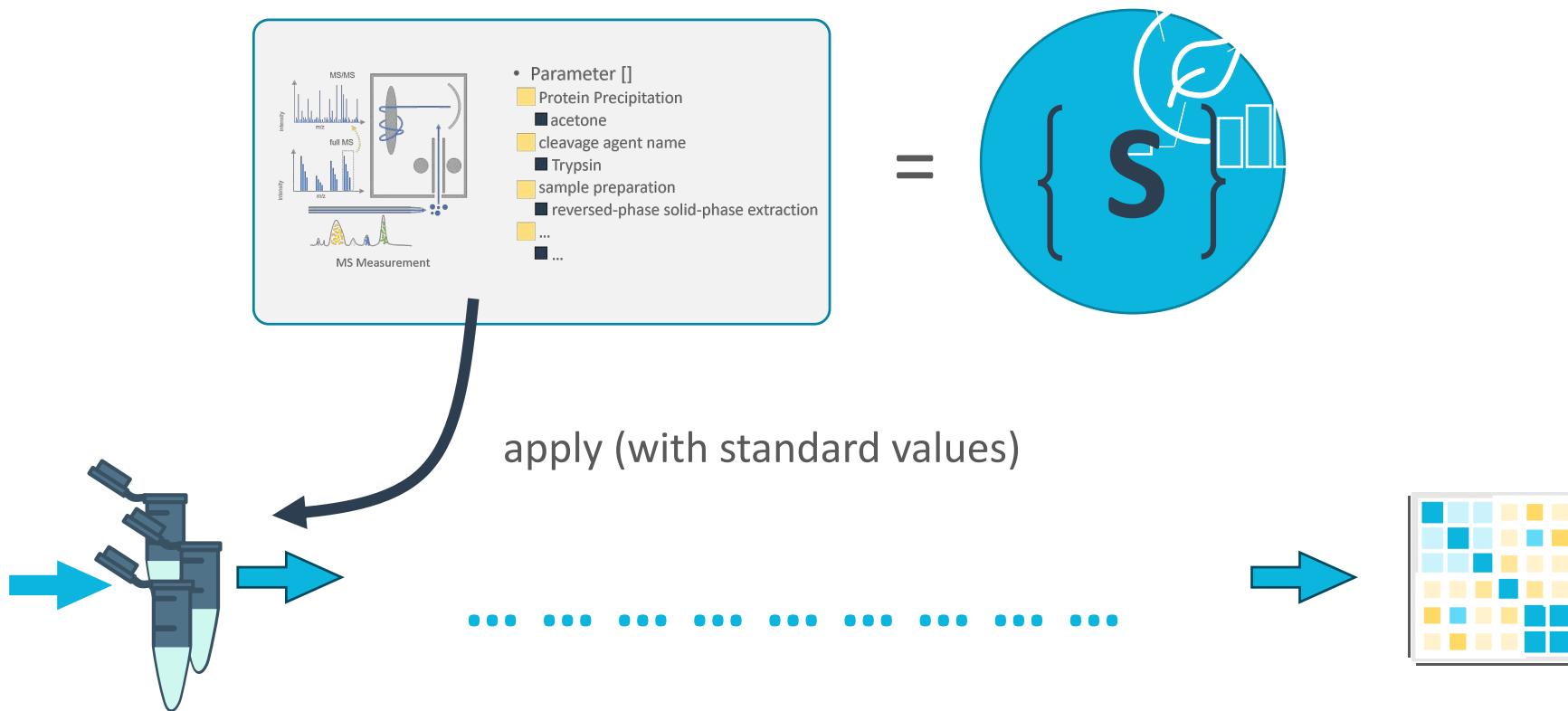
Modular separation of experimental processes



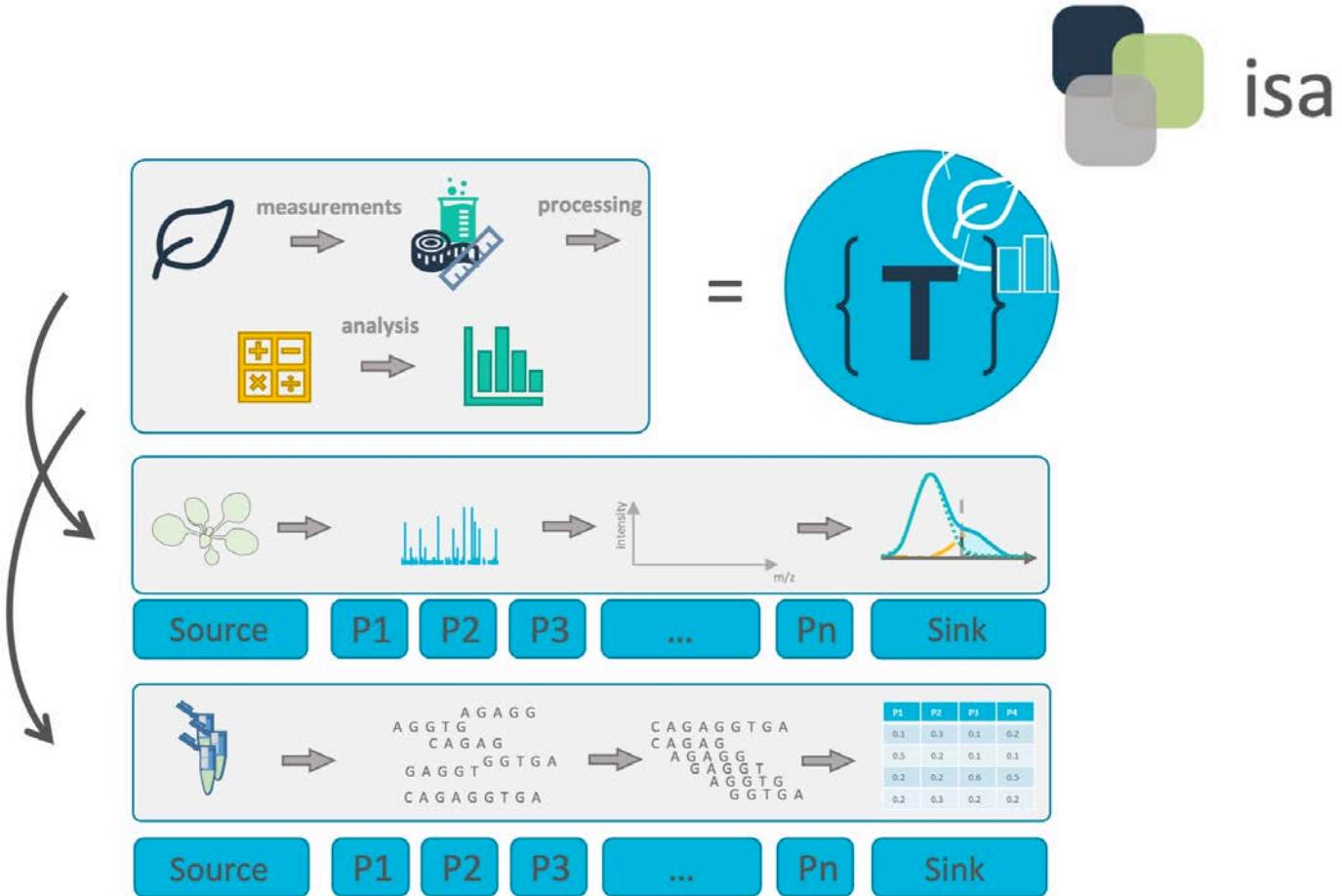
Modular separation of experimental processes



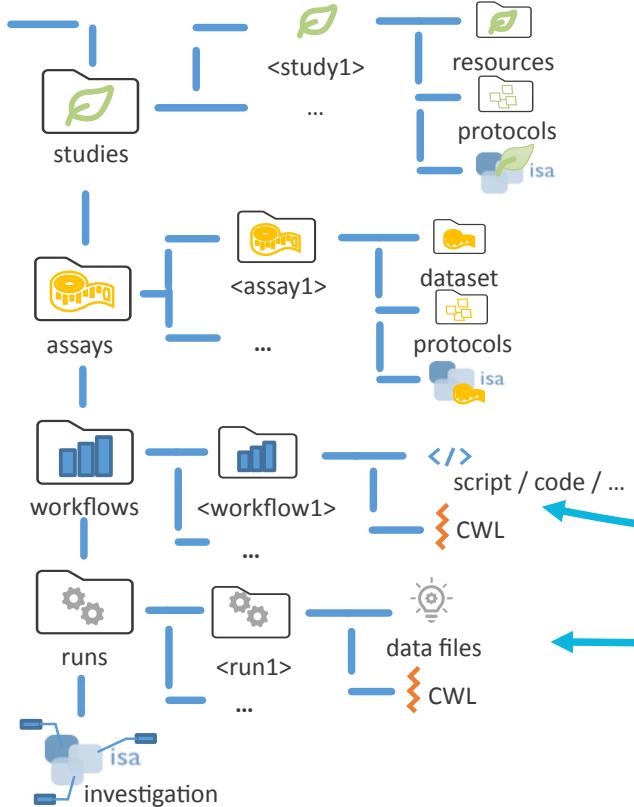
Applying standard procedures to sample record



Realization of lab-specific metadata with templates



Data analysis: CWL workflows and runs



```
python
Copy code

import pandas as pd
import plotly.express as px

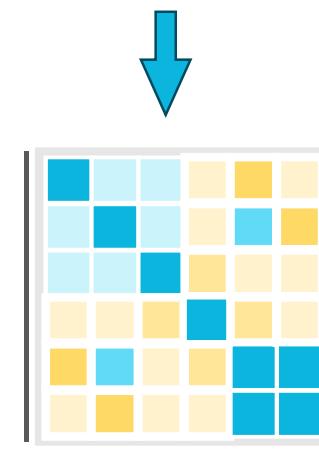
# Read the CSV file
data = pd.read_csv('result.csv')

# If the CSV doesn't have a clear index or row/column names, you may need to manually set
# For example, if the first column should be the index, you can use:
# data = pd.read_csv('result.csv', index_col=0)

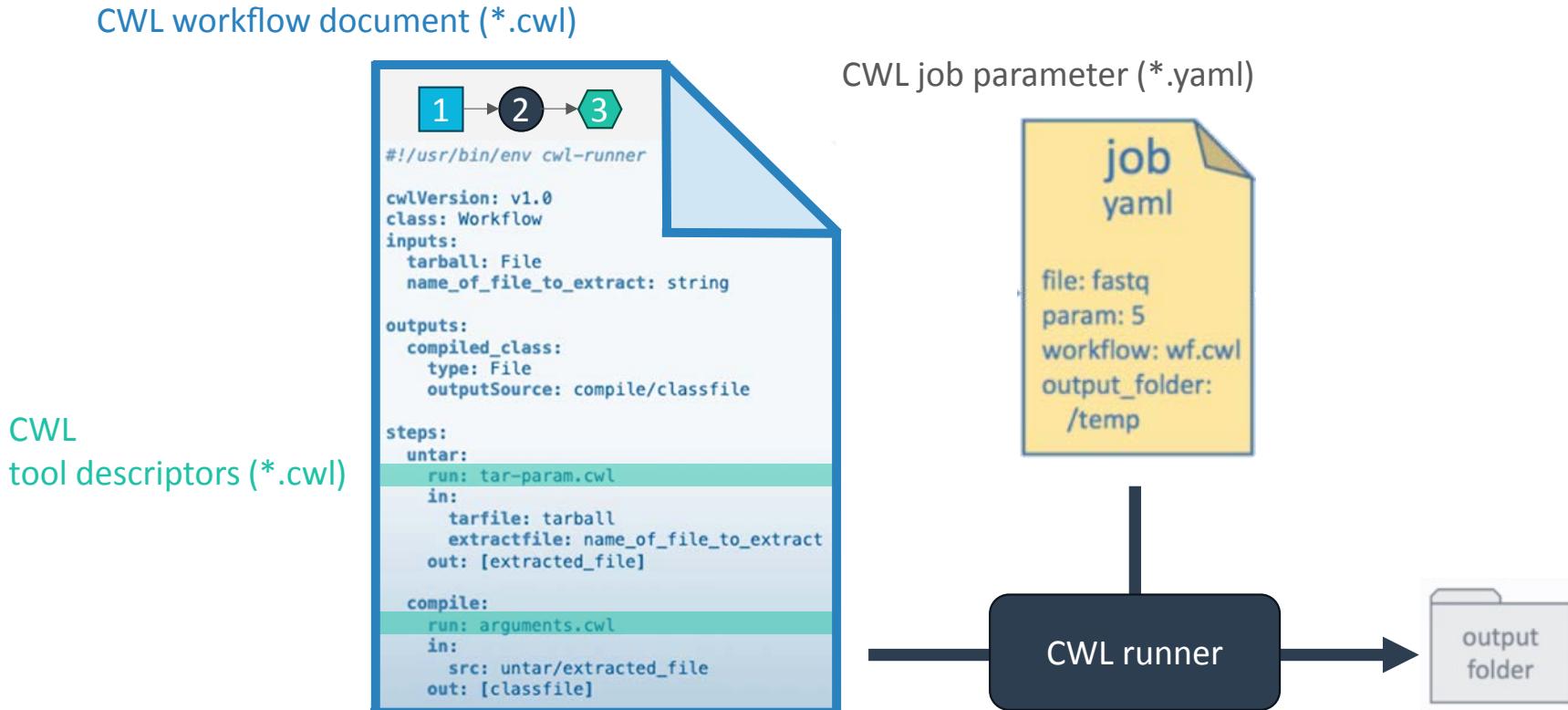
# Create a heatmap using Plotly
fig = px.imshow(data,
                 labels=dict(x="Columns", y="Rows", color="Value"),
                 x=data.columns,
                 y=data.index)

# Show the heatmap
fig.show()
```

heatmap.py

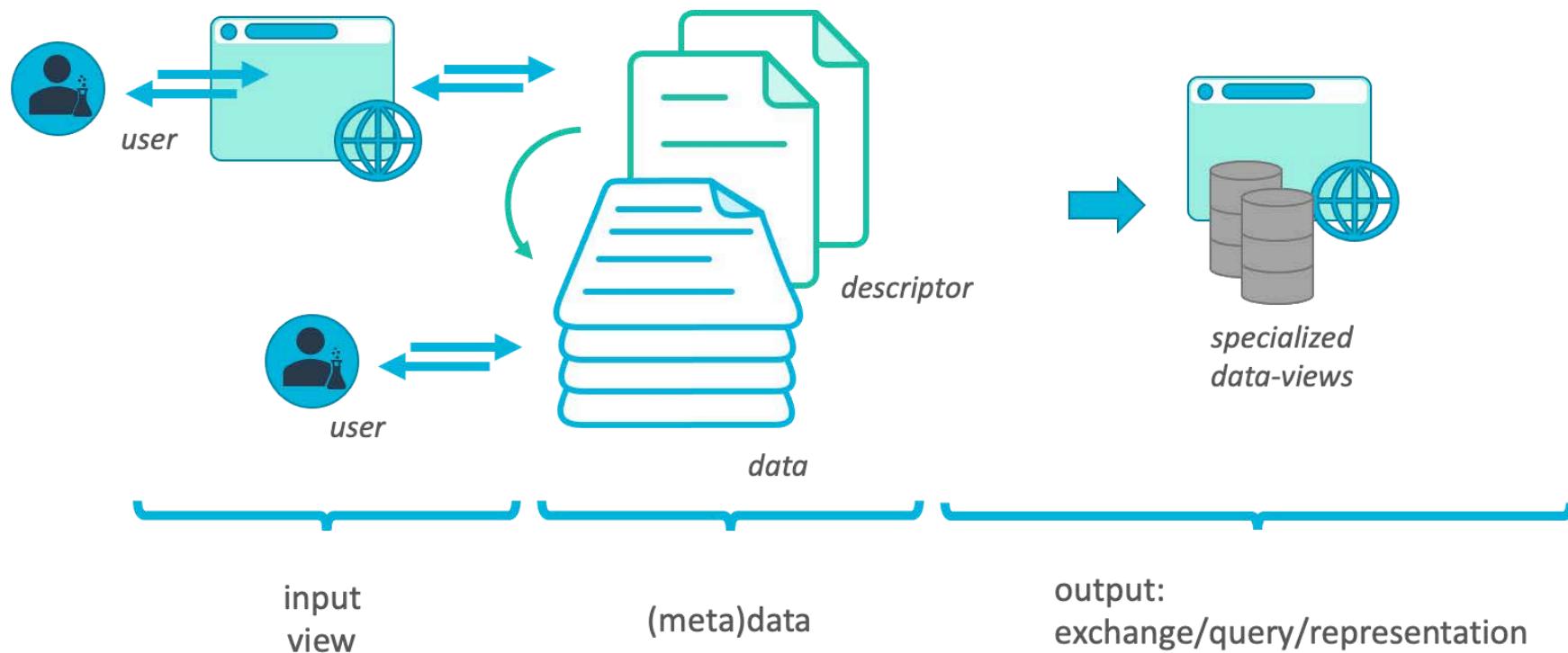


Metadata annotation – from data to result

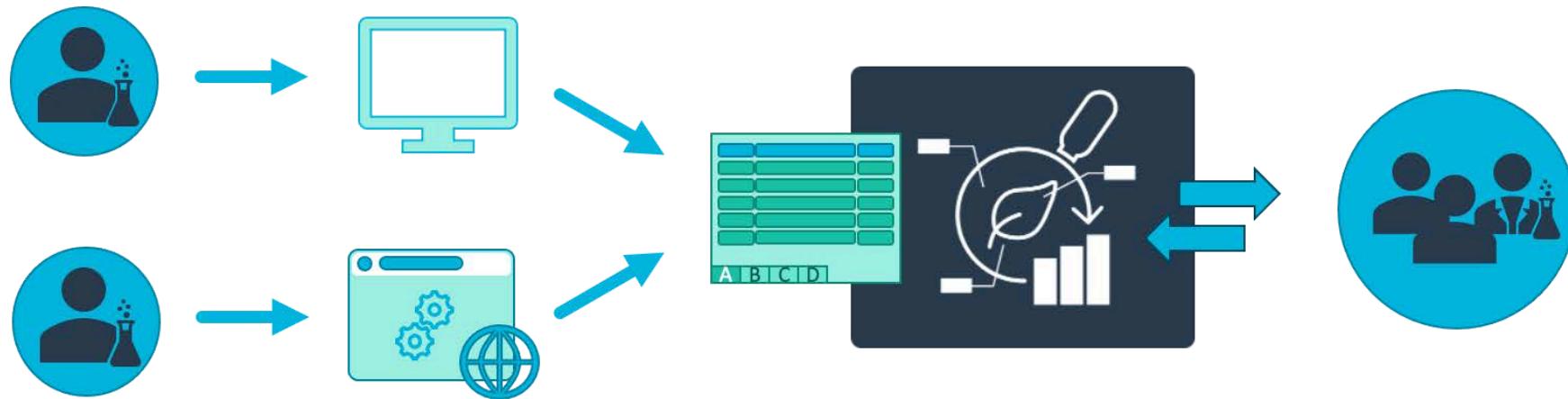


Everything is a file

The ARC is a **data-centric** approach to RDM

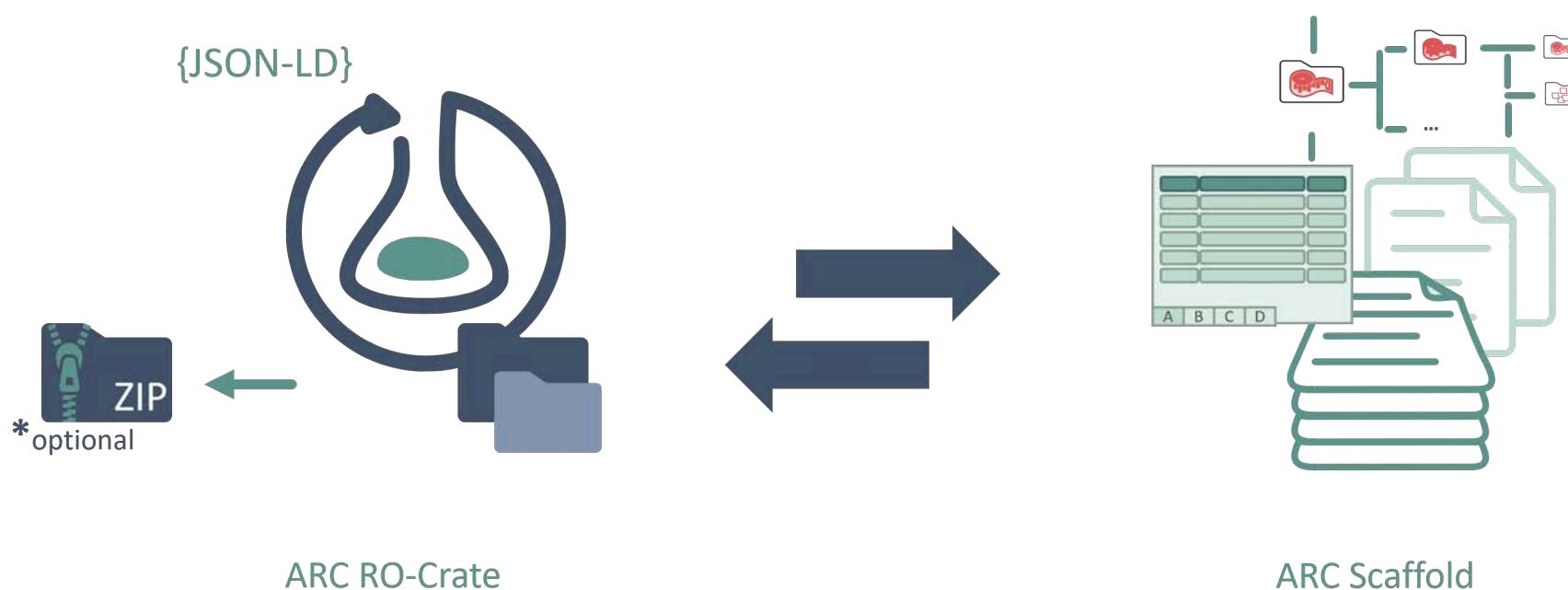


No technical lock-in



(Meta)data transparency with tool assistance but **no technical lock-in**

Two representations of the ARC



Two sides of the same coin

"Developer View": RO-Crate

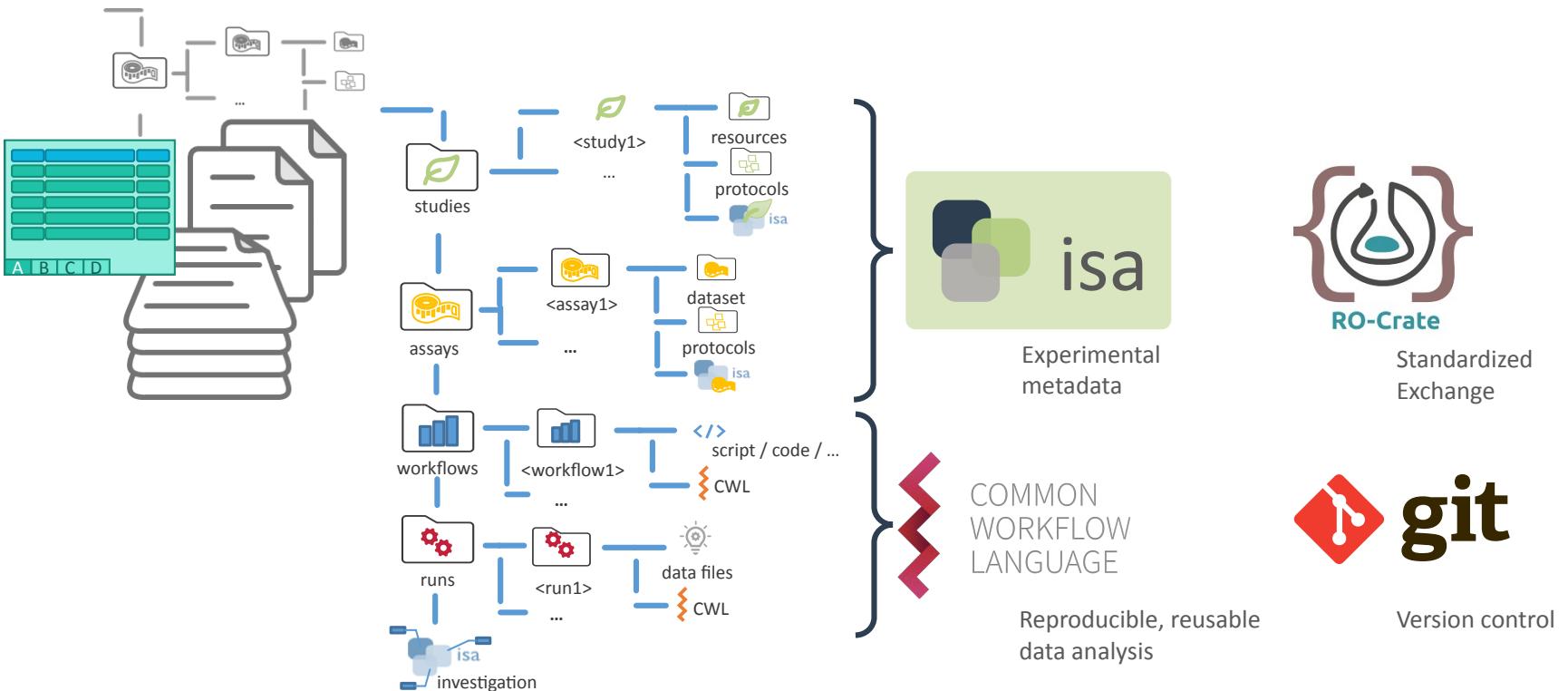
```
{  
  "Identifier": "Proteomics_MS",  
  "MeasurementType": {  
    "annotationValue": "Proteomics_MS",  
    "termSource": "MS",  
    "termAccession": "https://purl.obolibrary.org/obo/FMS_1003348"  
  },  
  "TechnologyType": ...  
  ...  
  "Tables": [  
    {  
      "name": "ProtDigest",  
      "header": [  
        {  
          "headertype": "Parameter",  
          "values": [  
            {  
              "annotationValue": "sample mass",  
              "termSource": "MS",  
              "termAccession": "https://purl.obolibrary.org/obo/FMS_1003348"  
            }  
          ]  
        }  
      ]  
    }  
  ]  
}
```

"User View": ARC Scaffold and metadata tables

The screenshot shows the ARCIctect software interface. On the left, the ARCIctect sidebar menu includes options like Login, New ARC, Open ARC, Download ARC, Explorer, Commit, DataHUB Sync, History, Validation, Services, Settings, and Toggle Sidebar. The version v0.0.55 is displayed at the bottom. The main window shows a file structure under 'AthalianaColdStressSugar' with 'assays' expanded, showing 'Proteomics_DataAnalysis' and 'Proteomics_MS'. 'Proteomics_MS' is highlighted with a teal border. Other items include 'dataset', 'protocols', 'README.md', 'SugarMeasurement', 'Visualization', 'workflows', 'runs', and another 'README.md'. To the right, the 'Assay Metadata' panel displays fields for Identifier (Proteomics_MS), Measurement Type (Proteomics_MS), Technology Type (Mass Spectrometry), Technology Platform (timsTOF Pro 2), and Performers (Assay, PeptideMS_Bruker, ProtDigest). Below this, a table lists parameters: Parameter [sample mass] (10 microgram), Parameter [Protein Precipitation] (acetone), Parameter [alkylating agent] (Chloroacetamide), and Parameter [red] (TCEP). The 'acetone' and 'Chloroacetamide' rows have green checkmarks.

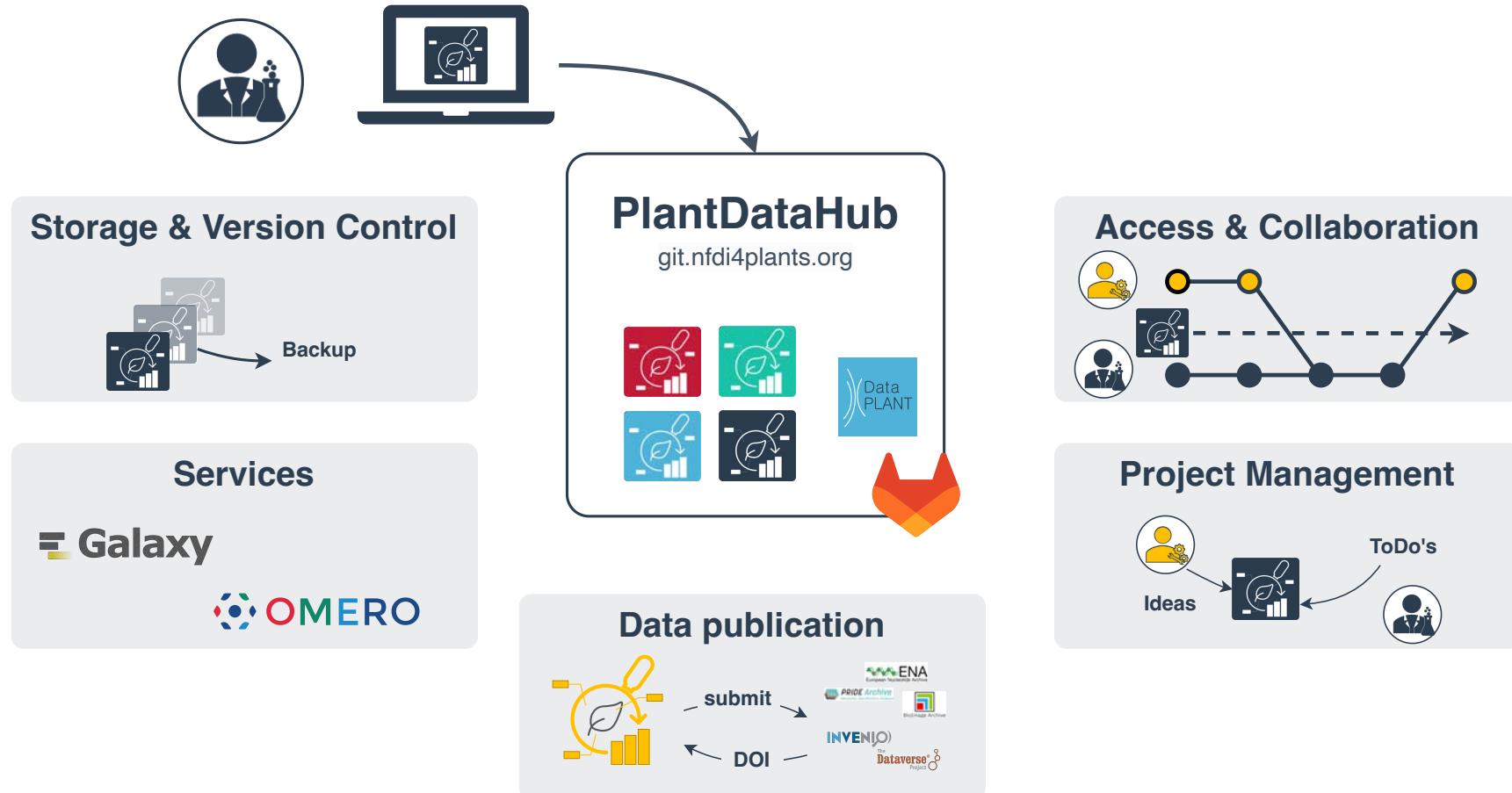
Parameter [sample mass] >>	Parameter [Protein Precipitation] >>	Parameter [alkylating agent] >>	Parameter [red] >>
10 microgram	acetone	Chloroacetamide	TCEP

ARC builds on standards

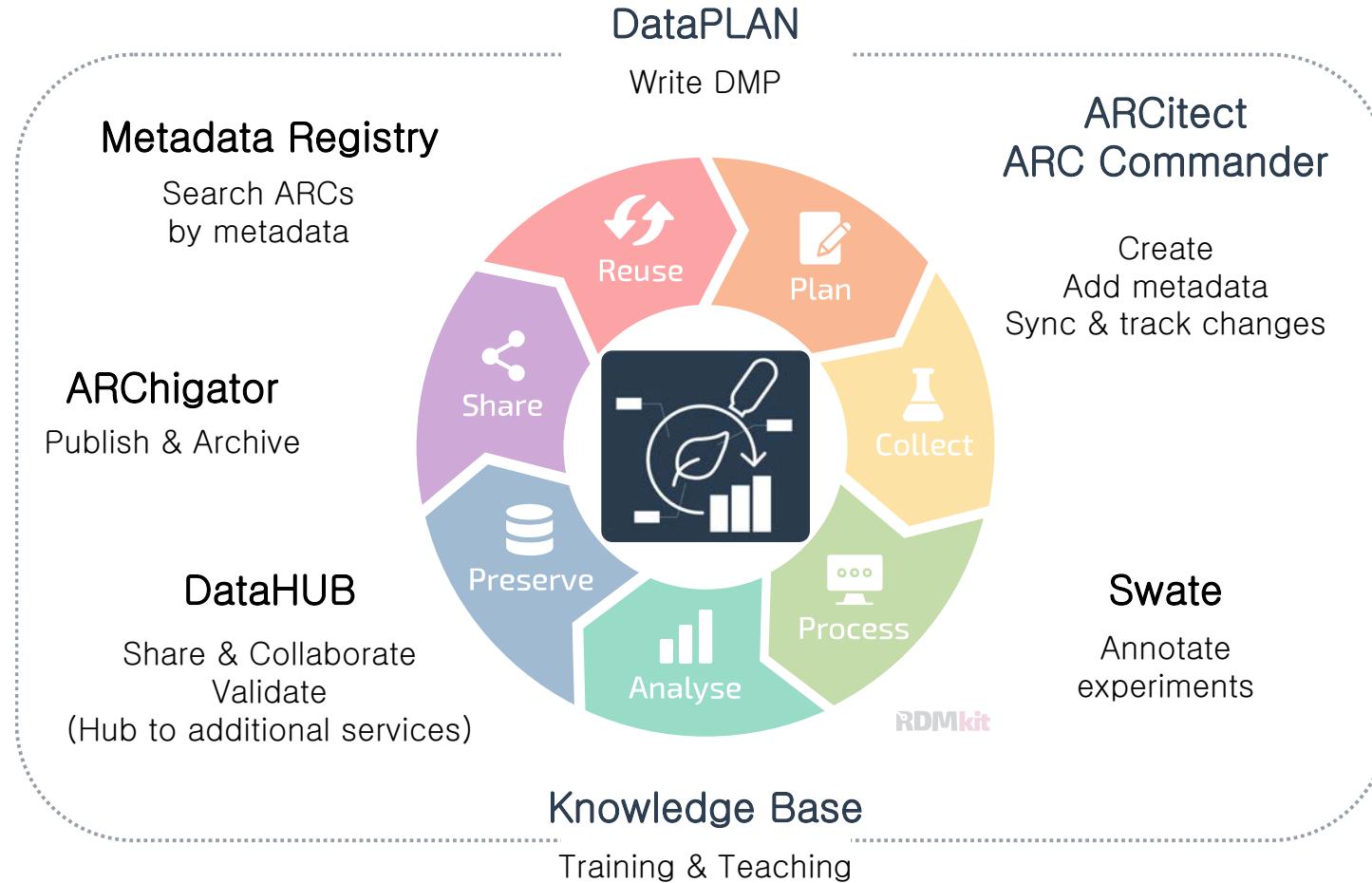


<https://isa-tools.org> | <https://www.commonwl.org> | <https://www.researchobject.org/ro-crate> | <https://git-scm.com>

ARC and DataHUB as entry point



The ARC ecosystem



Acknowledgements



Team Kaiserslautern

- Timo Mühlhaus
- Lukas Weil
- Kevin Frey
- Kevin Schneider
- Jonas Lukasczyk

Team Freiburg

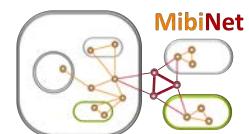
- Dirk von Suchodeletz
- Jonathan Bauer
- Marcel Tschöpe
- Julian Weidhase

Team Jülich

- Stella Eggels
- Angela Kranz



- Björn Usadel
- Vittorio Tracanna
- Yaser Alashloo



- Sabrina Zander

Goals for today

- Get an idea of the ARC concept
- Create and upload a simple ARC
- Explore and discuss ARC features
- ...

Start an ARC for your investigation



Hands-on

Follow the [Start Here guide](#) until Describe the investigation

Interaction during the course

Google Doc

- Please use the [google doc](#) to interact, raise questions and feedback
- Links for hands-on sections

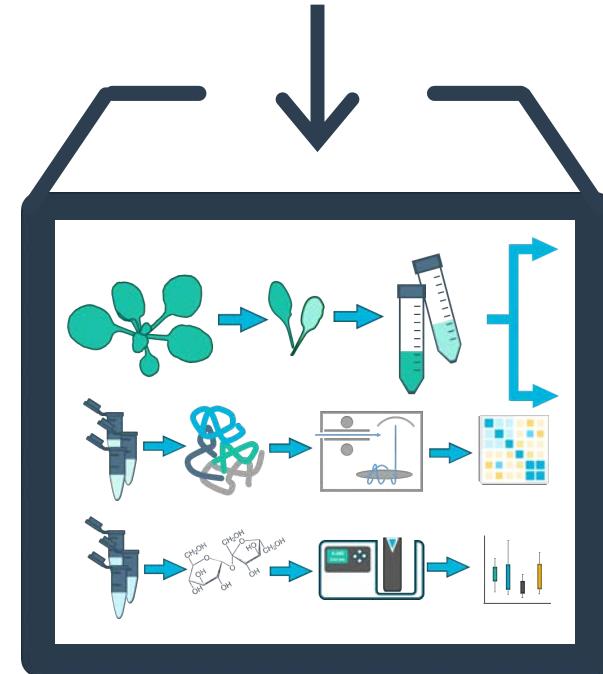
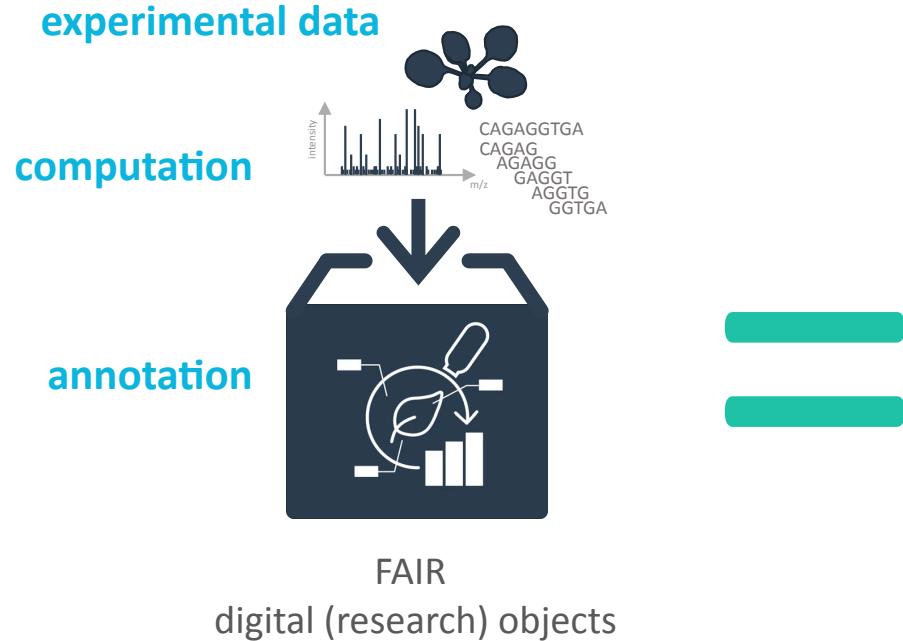
"I'm ready"

During the self-guided parts, please use Zoom reactions to tell me you're done with that part.

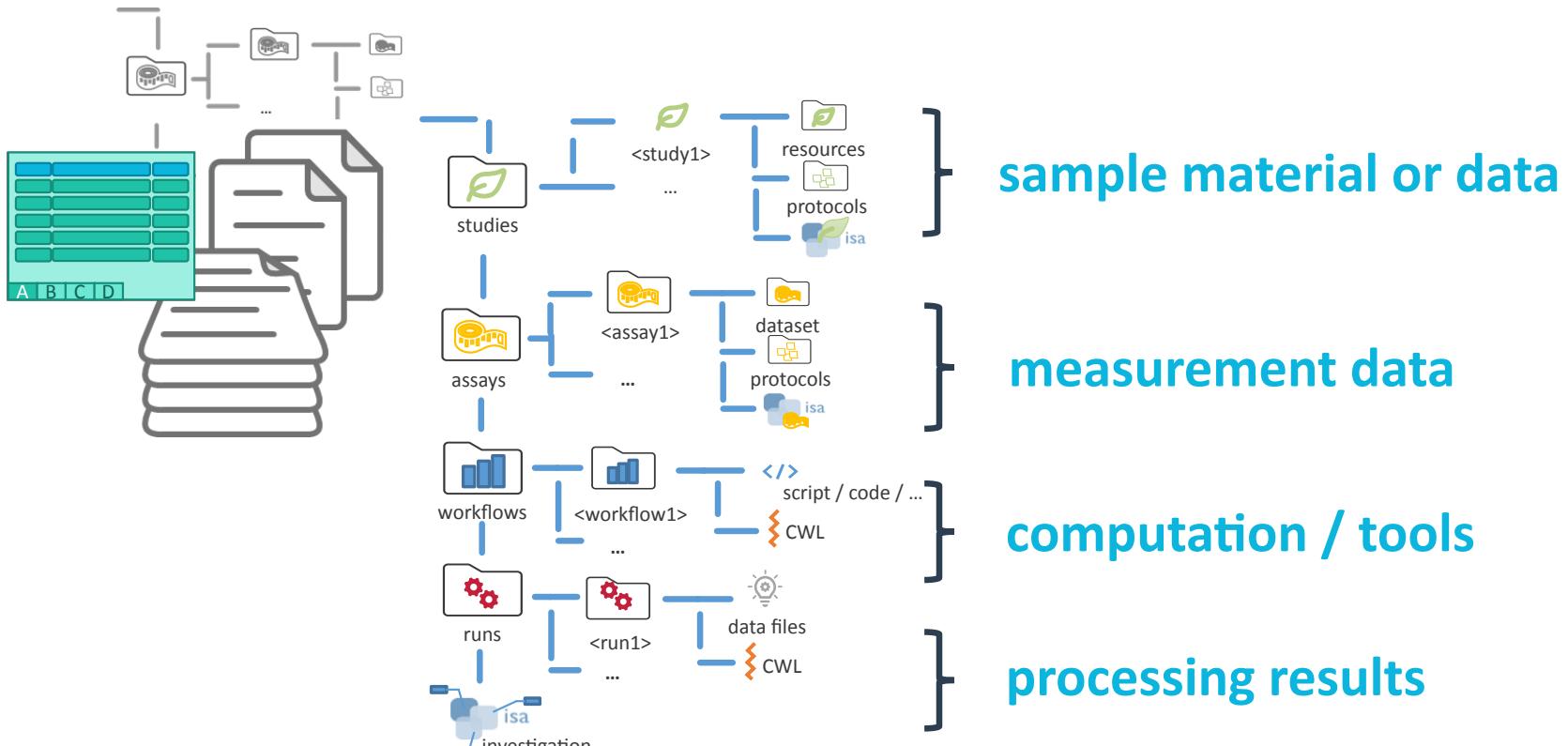


Yes

ARC: Annotated research context



The ARC scaffold structure



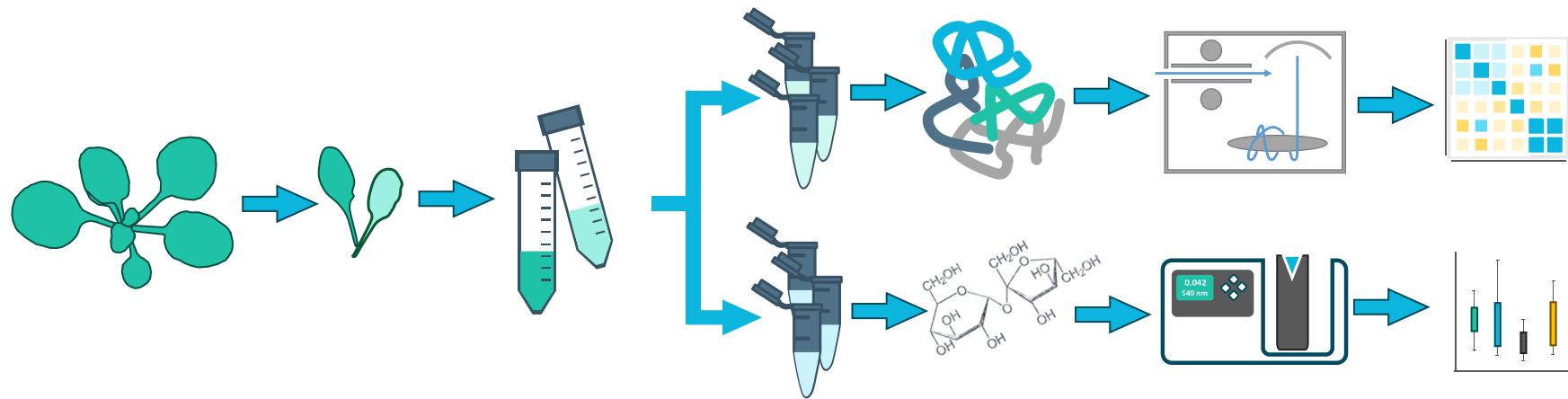
Add and describe a study



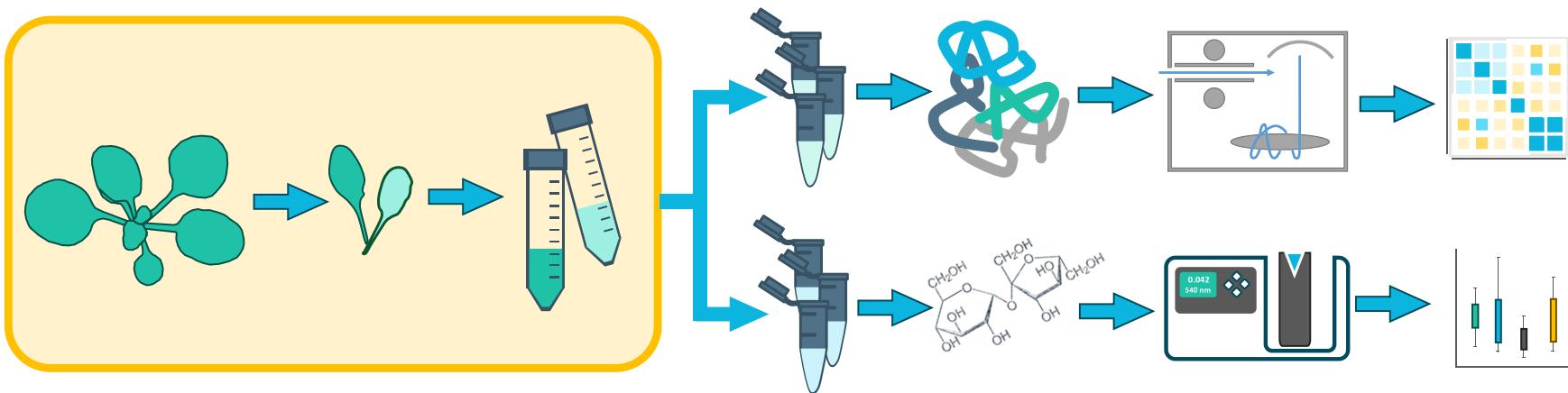
Hands-on

Follow the [Start Here guide](#) until Add a study

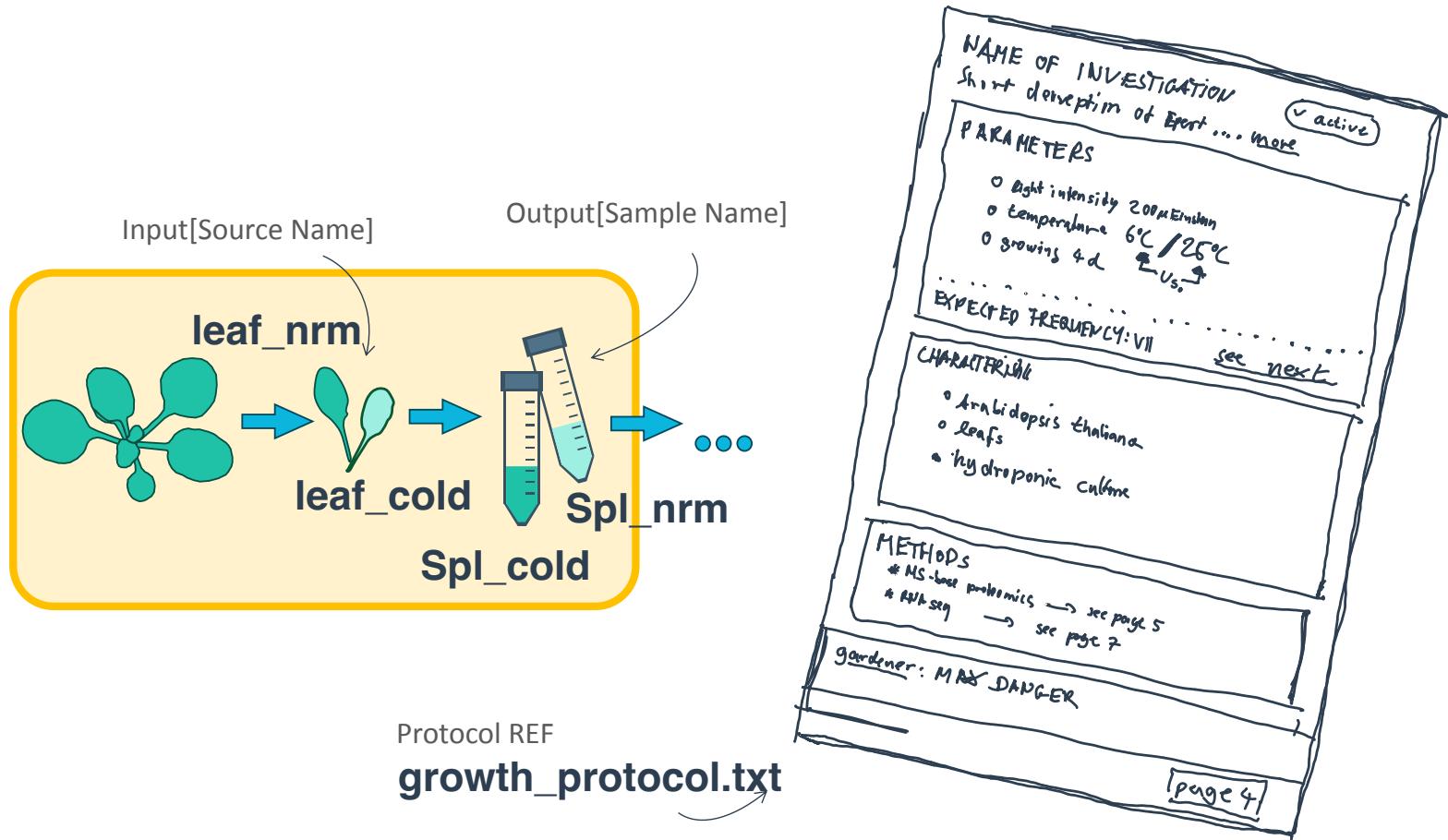
A small prototypic project



Divide and conquer for reproducibility



Identifying the ‘study’ part



A table-based organization schema

Input[Source Name]		Output[Sample Name]
leaf_nrm		spl_nrm
leaf_cold		spl_cold
A	B	C
D		

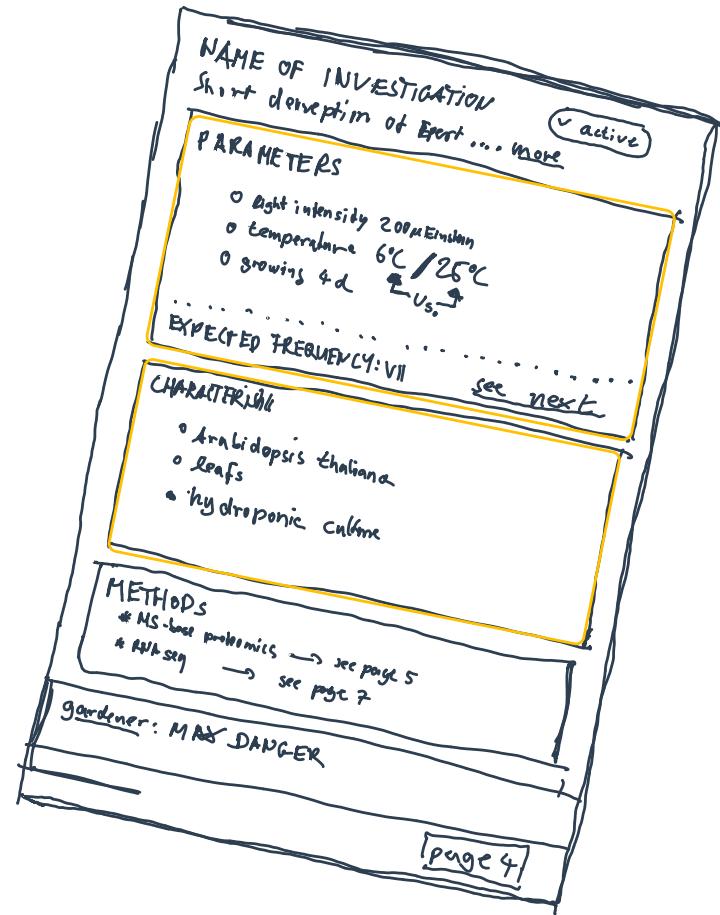
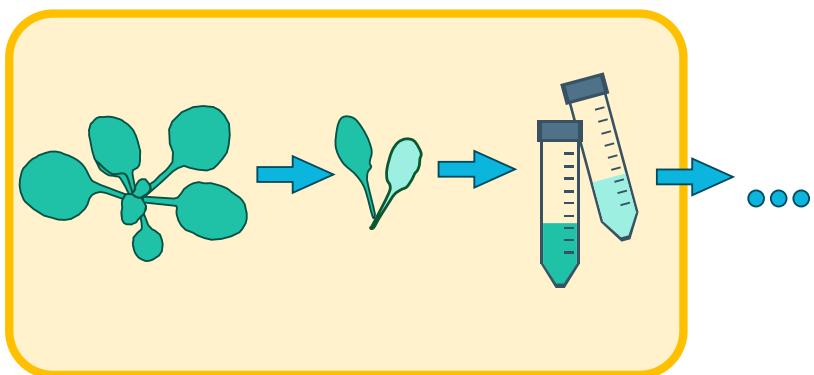
Referencing a protocol

This allows you to reference the free-text, human-readable protocol.

- 💡 It is recommended that the protocol is in an open format (.md|.txt|.docx|…)
- 💡 But everything is possible also an URI to an electronic lab notebook

Input[Source Name]	Protocol REF	Output[Sample Name]
leaf_nrm	growth_protocol.txt	spl_nrm
leaf_cold	growth_protocol.txt	spl_cold
A	B	C
D		

Parameterizing the ‘study’



Finding the right metadata vocabulary

Parameters []

- Light intensity 200 µEinstein
- Temperature 6°C / 25°C
- Growing 4d

Characteristics []

- *Arabidopsis thaliana*
- Leaf
- Hydroponic culture
- Columbia

OLS: Finding the right metadata vocabulary

Temperature Dependence

Temperature:Dependence_Annotation

http://purl.uniprot.org/core/Temperature_Dependence_Annotation

Indicates the optimum temperature for enzyme activity and/or the variation of enzyme activity with temperature variation; the thermostability/thermolability of the enzyme is also mentioned when it is known.

Ontology: UNIPROT RDFS

temperature

AFO:/result#AFR_0001584

http://purl.allotrope.org/ontologies/result#AFR_0001584

A temperature (datum) is a quantity facet that quantifies some temperature. [Allotrope]

Ontology: AFO

temperature

FBcv:0000466

http://purl.obolibrary.org/obo/FBcv_0000466

Mutation caused by exposure to a temperature that is higher or lower than 25 degrees Celsius.

Ontology: FBCV

temperature

PATO:0000146

http://purl.obolibrary.org/obo/PATO_0000146

A physical quality of the thermal energy of a system.

Ontology: PATO

Also appears in:

NGBO

HTN

CAO

ZP

AGRO

OMIABIS

OBIB

MONDO

TXPO

MCO

+

Welcome to the EMBL-EBI Ontology Lookup Service

temperature

Exact match Include obsolete terms Include imported terms

Examples: diabetes, GO:0098743

Looking for a particular ontology?

About OLS

The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the Samples, Phenotypes and Ontologies Team (SPOT) at EMBL-EBI.

Related Tools

In addition to OLS the SPOT team also provides the OxD and ZOOMA services. OxD provides cross-ontology mappings between terms from different ontologies. ZOOMA is a service to assist in mapping data to ontologies in OLS.

Report an Issue

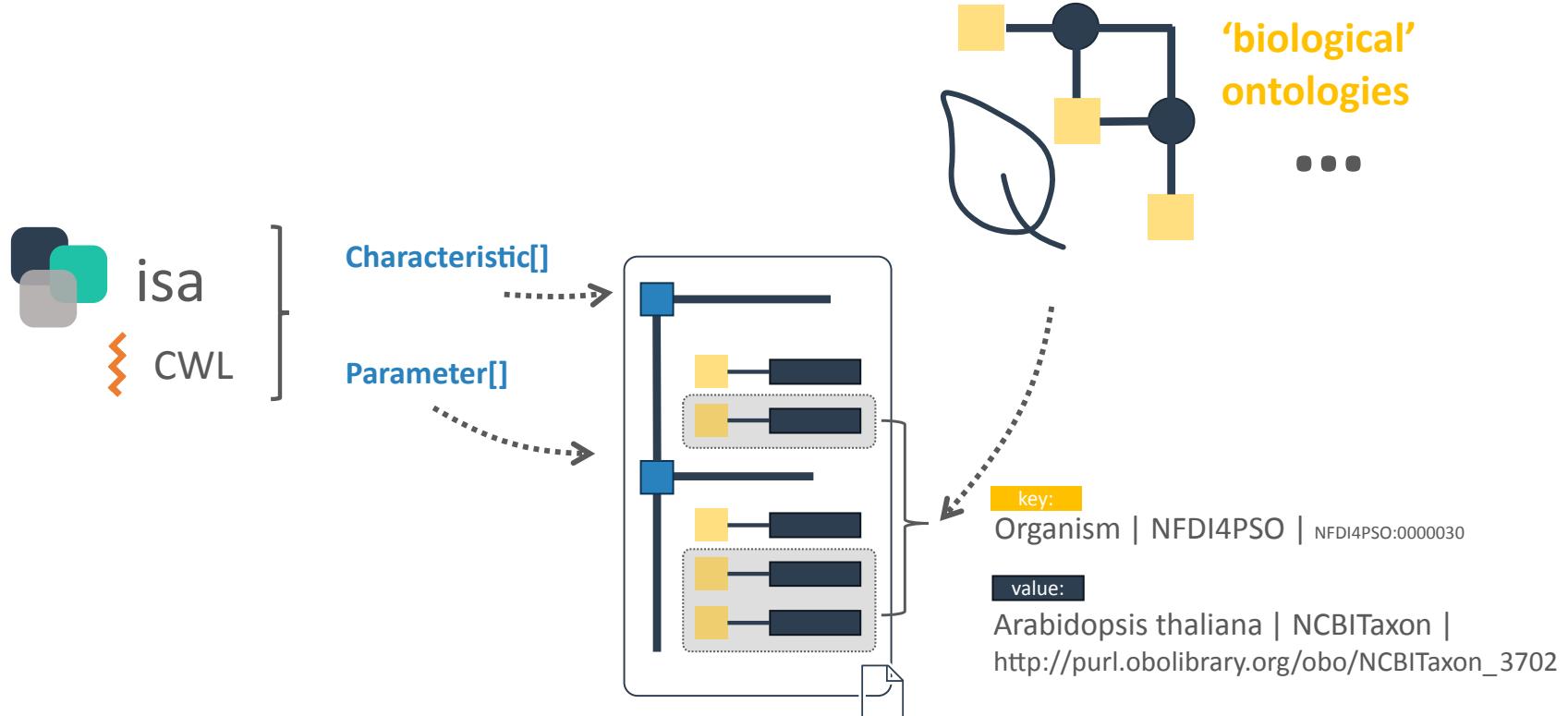
For feedback, enquiries or suggestion about OLS or to request a new ontology please use our GitHub issue tracker. For announcements relating to OLS, such as new releases and new features sign up to the OLS announce mailing list.

FOLLOW US

EMBL-EBI 2023 Licensing

Ontology Lookup Service (OLS): <https://www.ebi.ac.uk/ols4/>

Finding the metadata vocabulary and descriptors



Finding the metadata vocabulary and descriptors

Parameters []

- Light intensity 200 µEinstein
- Temperature 6°C / 25°C
- Growing 4d

Characteristics []

- *Arabidopsis thaliana*
- Leaf
- Hydroponic culture
- Columbia

Finding the metadata vocabulary and descriptors

Parameters []

- Light intensity
 - 200 μ Einstein
- Temperature
 - 6°C / 25°C
- Growth time
 - 4d

Characteristics []

- Organism
 - *Arabidopsis thaliana*
- Tissue
 - Leaf
- Growth medium
 - Hydroponic culture
- Ecotype
 - Columbia

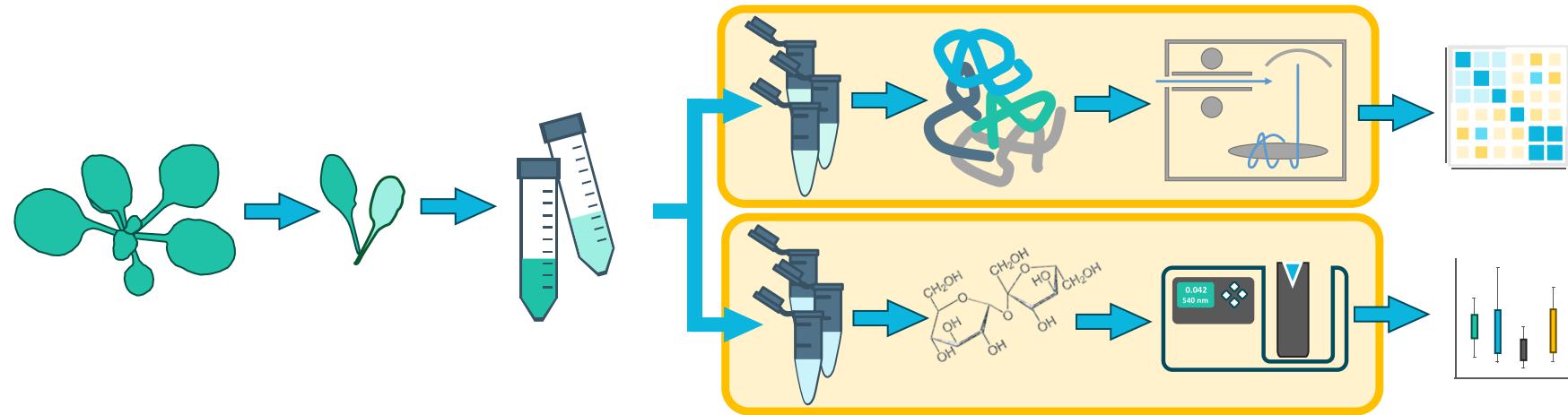
Add and describe an assay



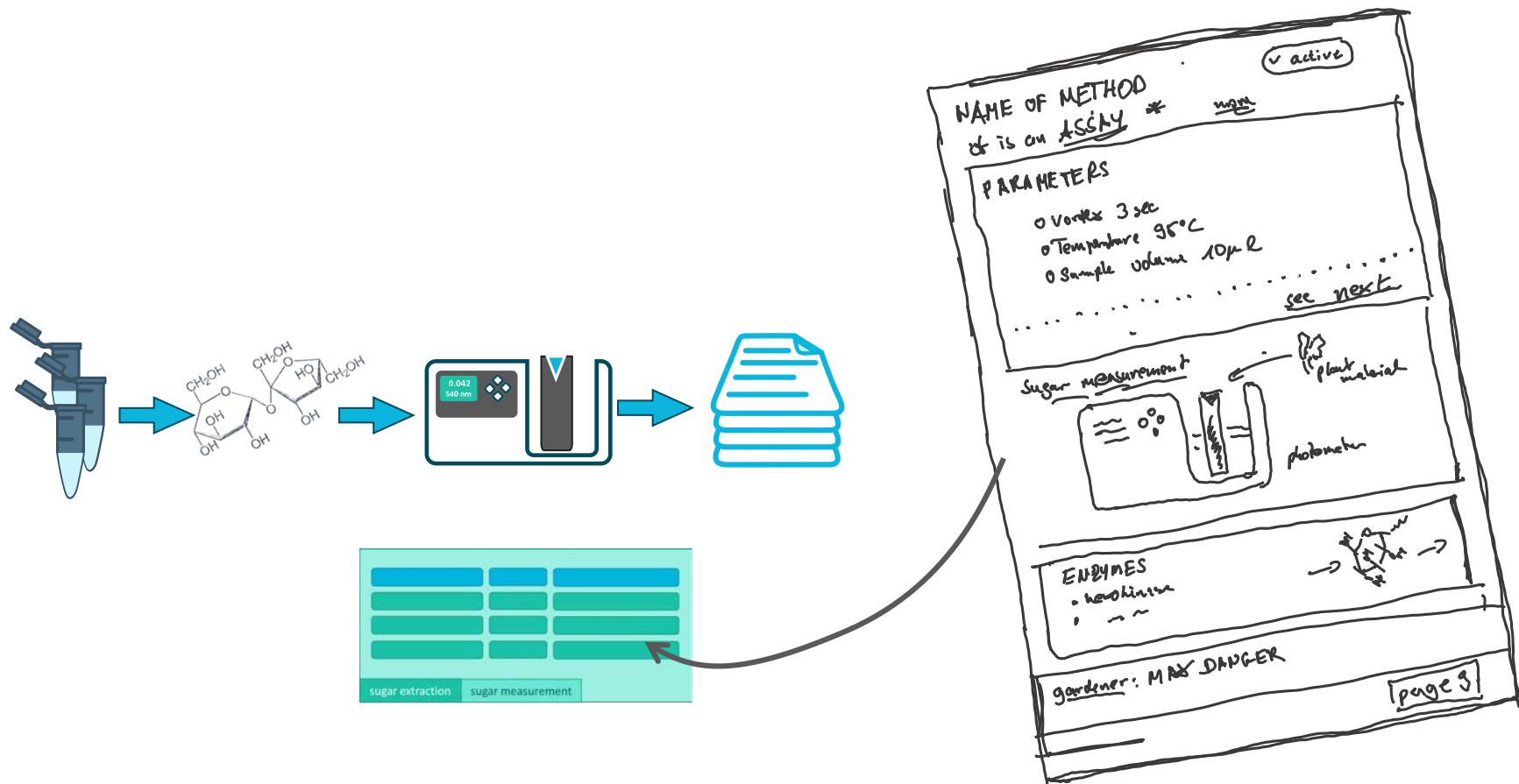
Hands-on

Follow the [Start Here guide](#) until Add assay data

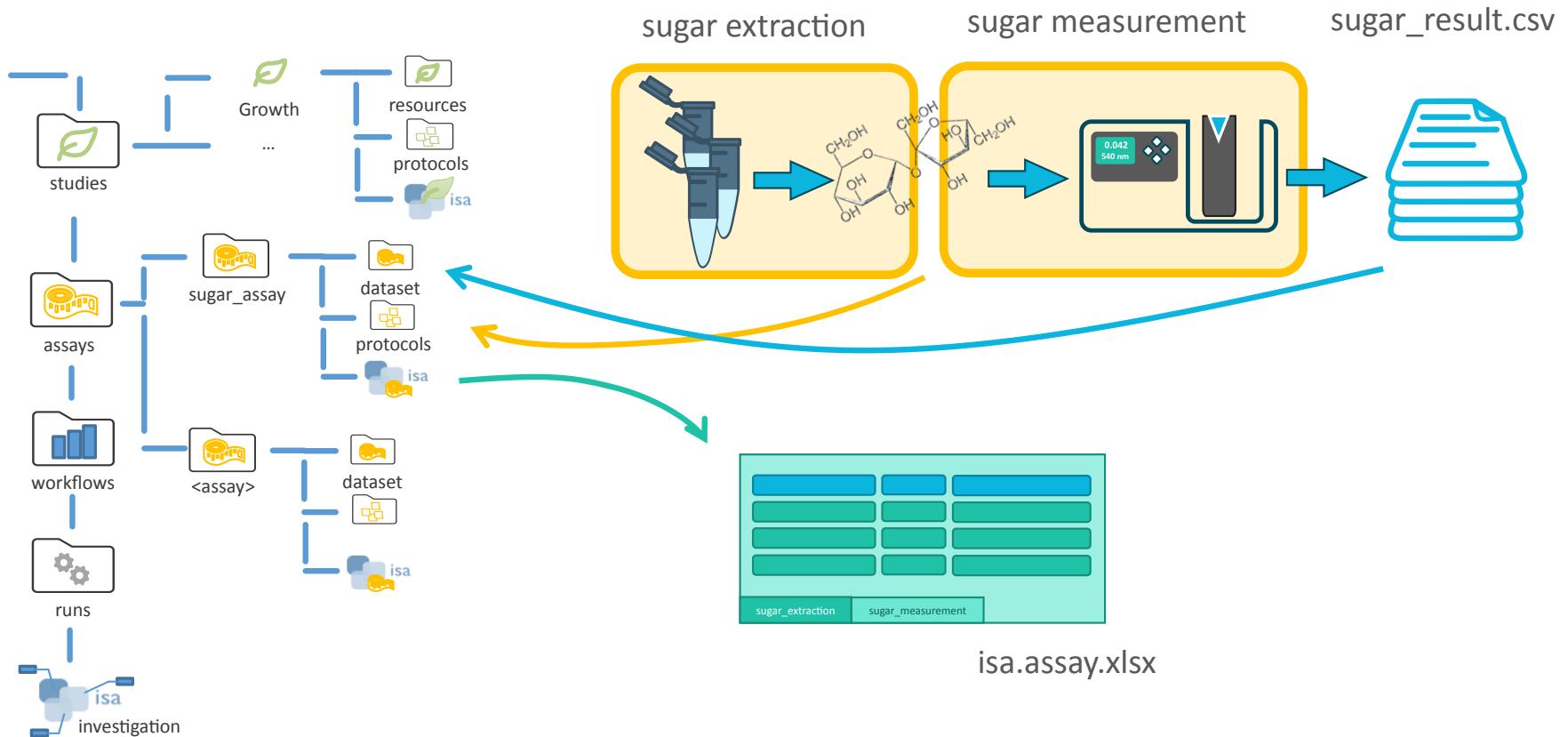
Identifying assays



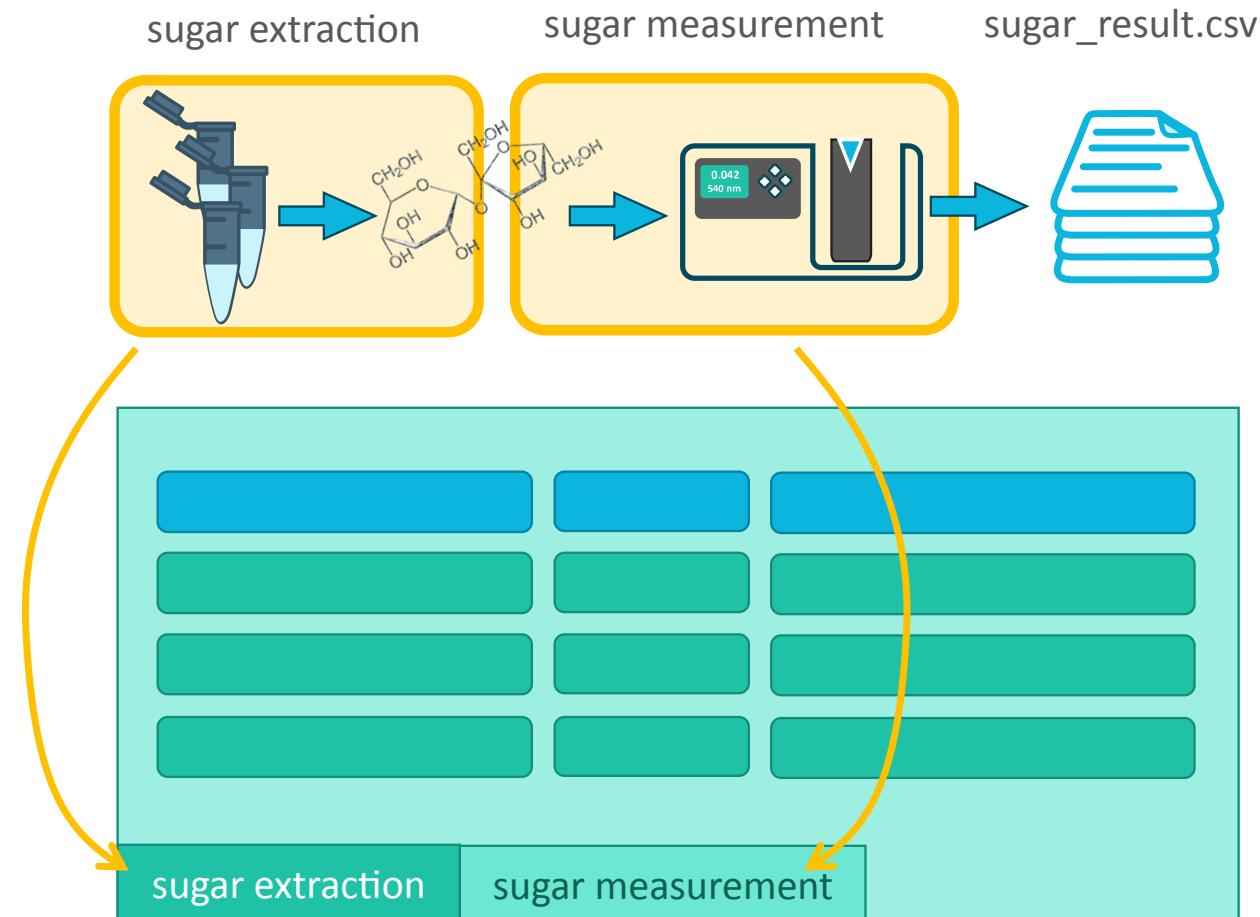
Assay for sugar measurement



Separating different assay elements



Isolating the lab processes in an assay



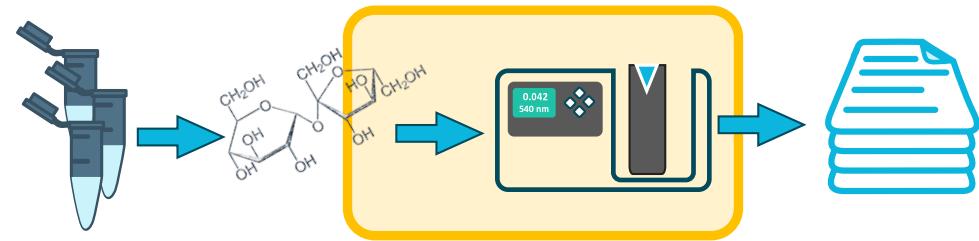
Parameteterization: sugar extraction

- Vortex Mixer
 - 3 seconds
- Temperature
 - 95 degree celsius



Parameteterization: sugar measurement

- ■ technical replicate
 - ■ 1,2,3,...
- ■ sample volume
 - ■ 10 microliter
- ■ buffer volume
 - ■ 190 microliter
- ■ cycle count
 - ■ 5



Share your ARC



Hands-on

Follow the [Start Here guide](#) until Share your ARC via the DataHUB

DataHUB terminology and data sharing

Choosing the proper role

When inviting new members to an ARC or group, you can choose between different levels.

Permissions & Roles

Roles are assigned when adding a user to an ARC or to a group. This is a very simplified summary.

Guest – Can only see the ARC's wiki

Reporter – Can read, but not add or edit data

Developer – Reporter permissions + can read, add, and edit data

Maintainer – Developer permissions + can add new members

Owner – Maintainer + can delete ARC, manage memberships and permissions

💡 By default you are **Owner** of an ARC you create or upload to the DataHUB.

Projects and Groups are not the same

- "Project" = ARC
- "Groups" = Group of users

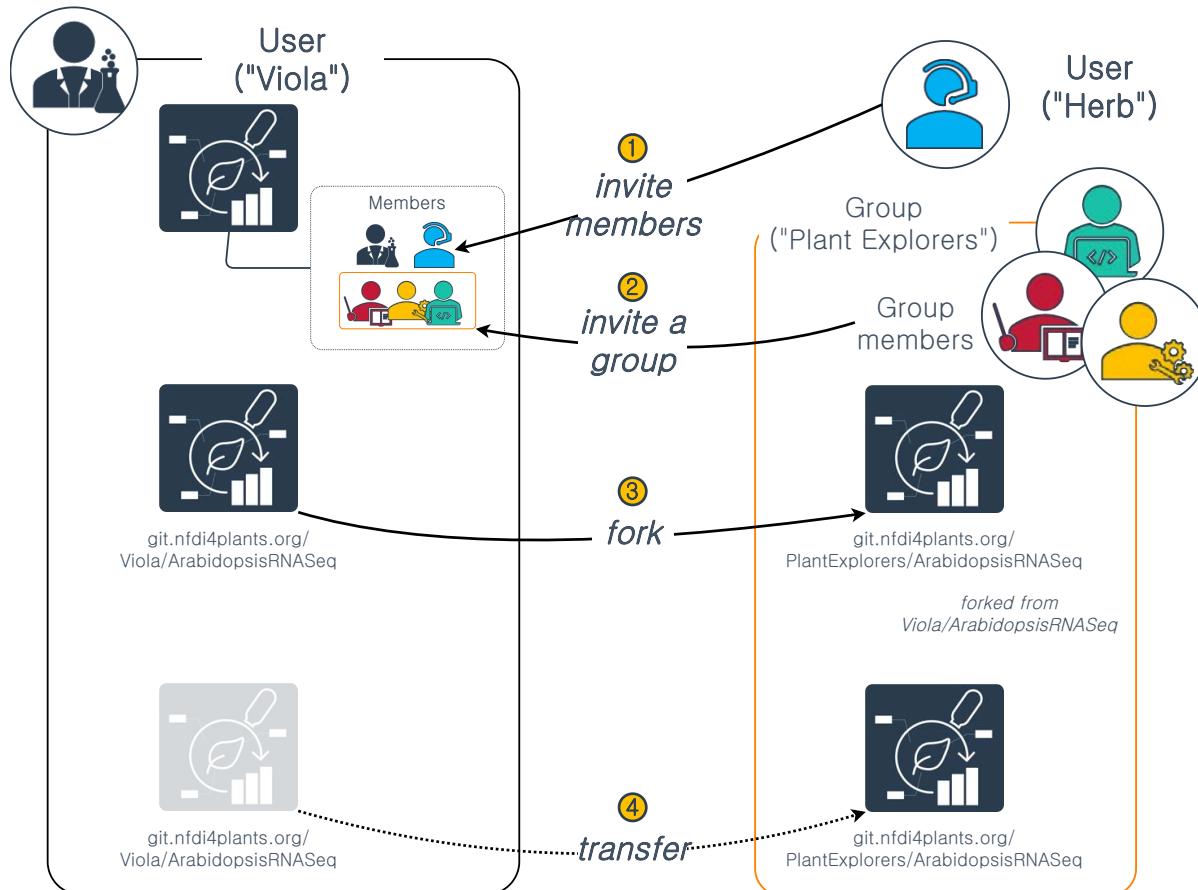
Project = ARC

- In the DataHUB, ARCs are called "projects"; they are the same.
- An ARC can be shared with individual users (invited as "members") or a group.

DataHUB Groups = Group of members (e.g. lab)

- A "Group" is a group of users with specific [permissions](#)
- A group can share ARCs
- A group can be invited to an ARC
- Groups can have subgroups

Options to share an ARC via the DataHUB



Namespaces

- Every user has a personal namespace, where they can upload or create new ARCs
- Every group and subgroup has its own namespace

Type	URL	Namespace	Name
A personal ARC	https://git.nfdi4plants.org/brilator/Facultative-CAM-in-Talinum	brilator	Dominik Brilhaus
An group-shared ARC	https://git.nfdi4plants.org/hhu-plant-biochemistry/Samuilov-2018-BOU-PSP	hhu-plant-biochemistry	HHU Plant Biochemistry

Visibility

The visibility of ARCs and groups can be managed individually for each ARC or group

Visibility

The visibility of each ARC can be managed in the settings of the ARC

Private – ARC access must be granted explicitly to each user or group.

Internal – ARC can be accessed by any logged in user.

Public – ARC can be accessed without authentication.

💡 By default every ARC and every group is set to **private**.

ARC DataHUB members // ARC Investigation contacts

Project members

You can invite a new member to Samuilov-2018-BOU-PSP or invite another group.

Members 18 Groups 1

Account	Source	Max role	Expiration	Activity
Adriano Nunes-Nesi @unesnesi	HHU Plant Biochemistry by Sebastian Triesch	Maintainer	Expiration date	User created: Jul 05, 2023 Access granted: Jul 10, 2023 Last activity: Aug 09, 2023
Andreas Weber @andreas.weber	HHU Plant Biochemistry by Sebastian Triesch	Maintainer	Expiration date	User created: Mar 10, 2023 Access granted: Jul 31, 2023 Last activity: Sep 11, 2023
Dominik Brilhaus It's you @brilator	Direct member by Dominik Brilhaus	Owner	Expiration date	User created: Feb 21, 2022 Access granted: Dec 06, 2023 Last activity: Mar 26, 2024
Franziska Fichtner @franziska.fichtner	HHU Plant Biochemistry by Sebastian Triesch	Maintainer	Expiration date	User created: Aug 11, 2023 Access granted: Aug 11, 2023 Last activity: Aug 11, 2023

DataHUB: ARC members

https://git.nfdi4plants.org/hhu-plant-biochemistry/Samuilov-2018-BOU-PSP/-/project_members

Identifier

Title

Description

Contacts

Contact	Score
Sladjana Samuilov <orcid>	4/10
Nadine Rademacher <orcid>	3/10
Samantha Flachbart <orcid>	3/10
Leila Arab <orcid>	3/10
Saleh Alfarraj <orcid>	3/10
Franziska Kuhnert <orcid>	3/10
Stanislav Kopriwa <orcid>	3/10
Andreas P. M. Weber <orcid>	4/10
Tabea Mettler-Altmann <orcid>	3/10

ARCIct: Investigation Contacts

- 💡 Investigation contacts are not automatically invited as members to the ARC.

Version control

Check out the **commit history** of your ARC via Repository (2) or directly via commits (7)

The screenshot shows the DataPLANT CEPLAS interface with the following numbered callouts:

- Manage
- Plan
- Code
- Build
- Secure
- Deploy
- Operate
- Analyze
- Settings
- Help

Key UI elements include:

- D Demo_ARC**: Project name in the header.
- Code**: Selected tab in the top navigation bar.
- arc init**: Commit message by Demo User, authored 4 minutes ago.
- Project information**: Pipeline status (passed), Publish ARC button, 1 Commit, 2 Branches, 0 Tags, 7 KIB Project Storage.
- Auto DevOps enabled**: Options to add README, LICENSE, CHANGELOG, CONTRIBUTING, Kubernetes cluster, Wiki, and Integrations.
- Created on**: July 13, 2024.

Name	Last commit	Last update
assays	arc init	4 minutes ago
runs	arc init	4 minutes ago
studies	arc init	4 minutes ago
workflows	arc init	4 minutes ago
.gitignore	arc init	4 minutes ago
isa.investigation.xlsx	arc init	4 minutes ago

Resources



Info & materials

- DataPLANT Website: <https://nfdi4plants.org/>
- ARC website: <https://arc-rdm.org>
- Knowledge Base:
<https://nfdi4plants.org/nfdi4plants.knowledgebase/>

Tools and Services

- ARCitect: <https://github.com/nfdi4plants/arcitect>
- DataHUB: <https://git.nfdi4plants.org>

Continuous support

- HelpDesk: <https://helpdesk.nfdi4plants.org>
- Matrix for ad hoc support: <https://matrix.to/#/%23arc-user-support:matrix.org>
- User Support Meeting (2nd Friday of the month | 1 – 2pm):
<https://nfdi4plants.github.io/events/arc-user-support/>
- User Support Mailing List: [Click here to subscribe](#)

Open Source Development

- GitHub: <https://github.com/nfdi4plants>

Acknowledgements



Team Kaiserslautern

- Timo Mühlhaus
- Lukas Weil
- Kevin Frey
- Kevin Schneider
- Jonas Lukasczyk

Team Freiburg

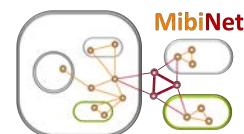
- Dirk von Suchodeletz
- Jonathan Bauer
- Marcel Tschöpe
- Julian Weidhase

Team Jülich

- Stella Eggels
- Angela Kranz



- Björn Usadel
- Vittorio Tracanna
- Yaser Alashloo



- Sabrina Zander