



**CEPLAS**

Cluster of Excellence on Plant Sciences

# CEPLAS ARC Training

October, 2024

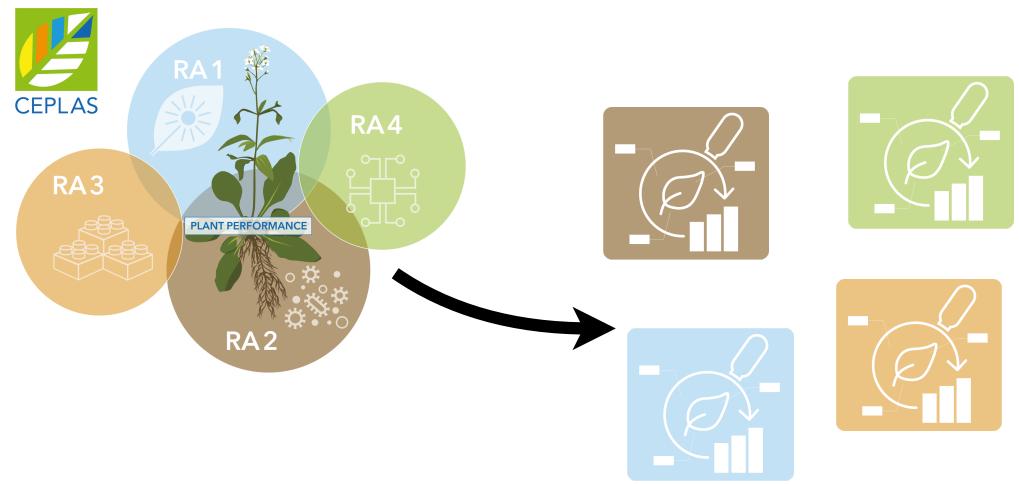
Dominik Brilhaus (CEPLAS)

Sabrina Zander (MibiNet)



# BYOD – Goals

- First few steps into ARC ecosystem
- Move existing datasets into ARCs
- Share them via the DataHUB
- Annotate with metadata



# House-keeping

- Use the pad to raise questions and feedback
- Copy / paste links (hands-on)



# Training Materials

Slides are shared via [DataPLANT knowledge base](#)

# Resources – join the open source movement



## DataPLANT (nfdi4plants)

Website: <https://nfdi4plants.org/>

Knowledge Base: <https://nfdi4plants.org/nfdi4plants.knowledgebase/>

DataHUB: <https://git.nfdi4plants.org>

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

HelpDesk: <https://helpdesk.nfdi4plants.org>



You can help us by raising issues, bugs, ideas...

# Continuous support

Data managers in Düsseldorf, Cologne, Jülich and close by (CEPLAS, MibiNet, TRR341) offer support.

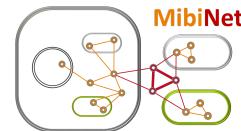
1. Slack Workspace for ad hoc support
2. Monthly user meeting (2nd Friday of the month)

→ [Details](#).



trr\_341

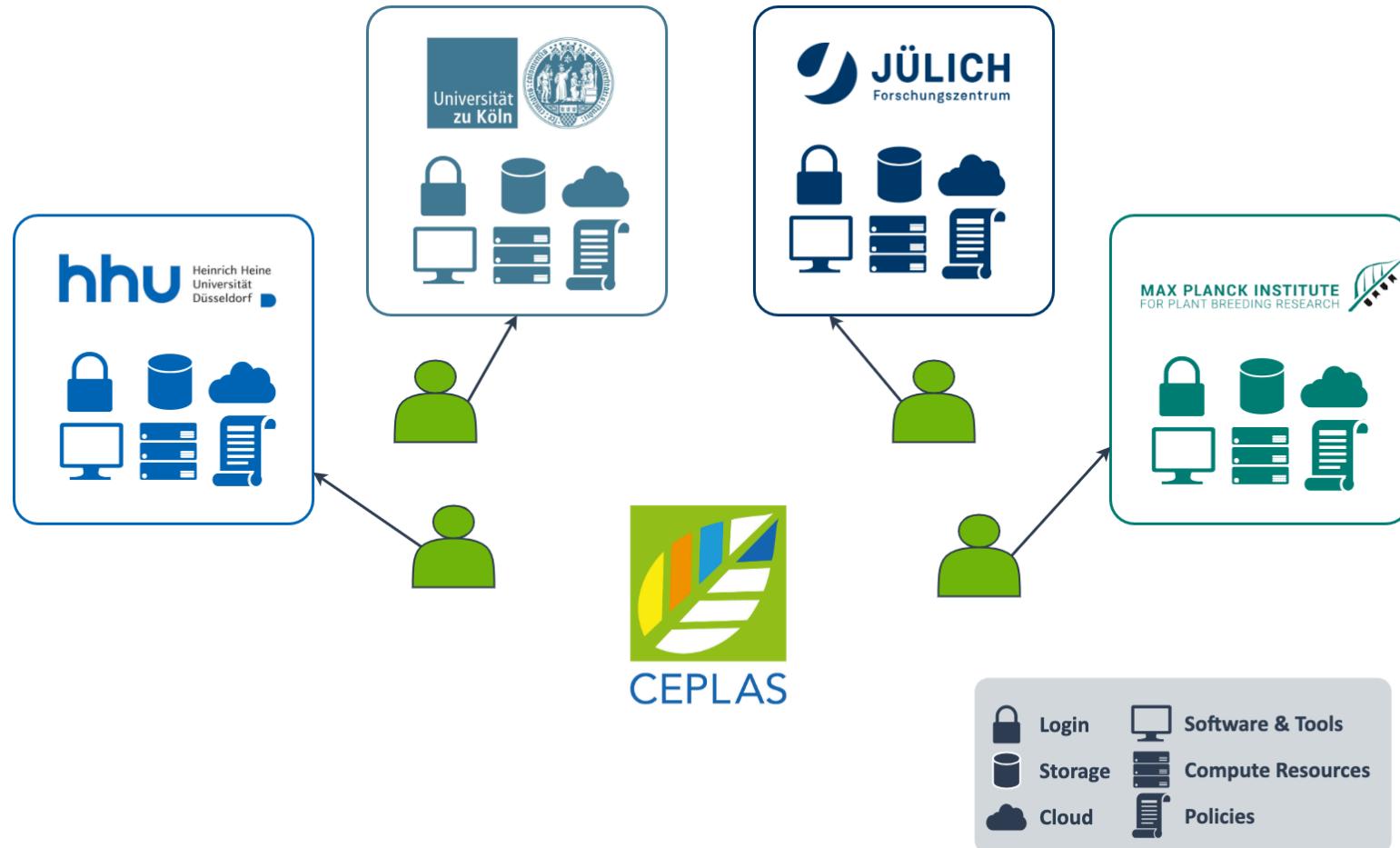
plant  
ecological  
genetics



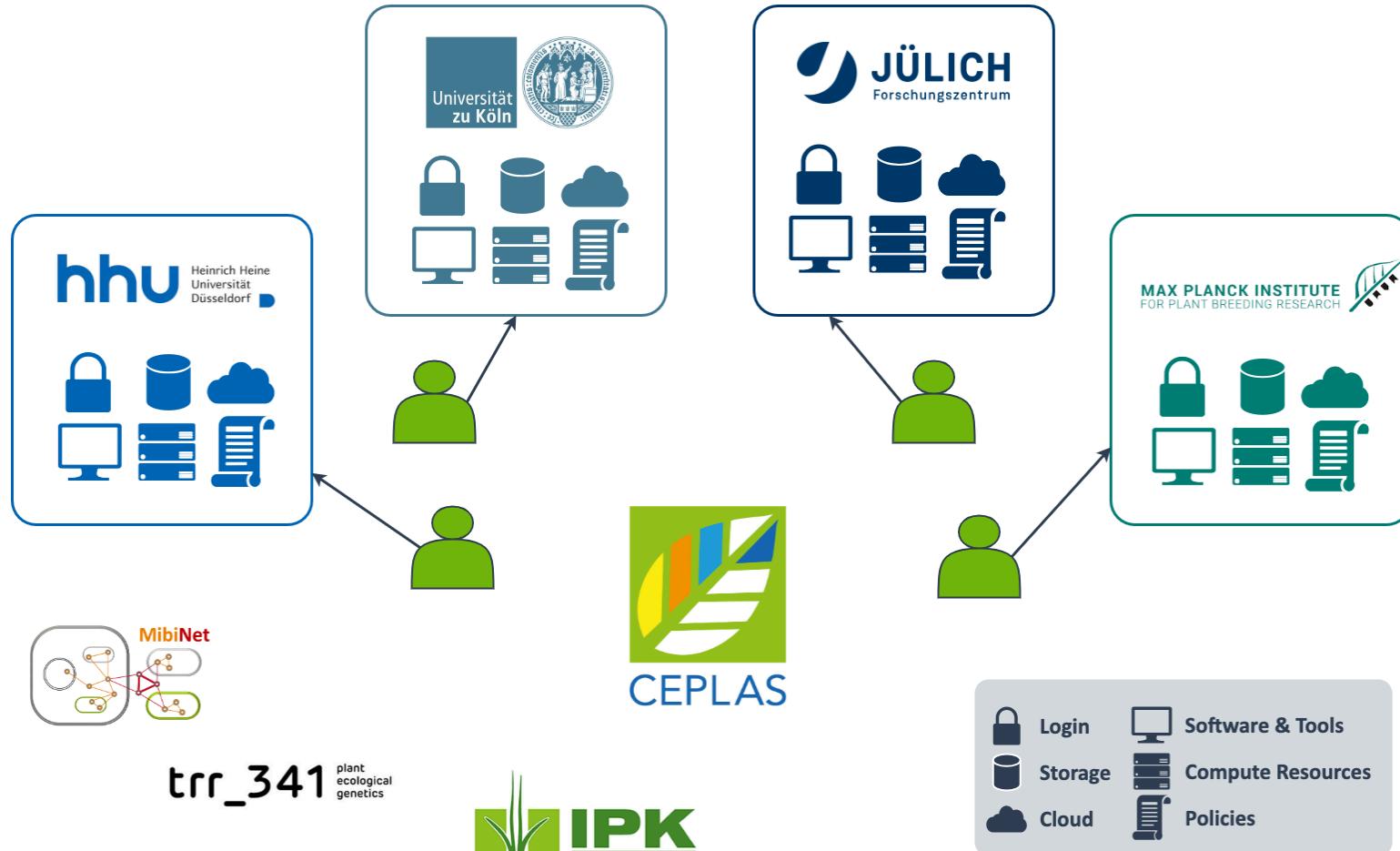
# CEPLAS – One cluster, four locations



# Data silos impede collaboration

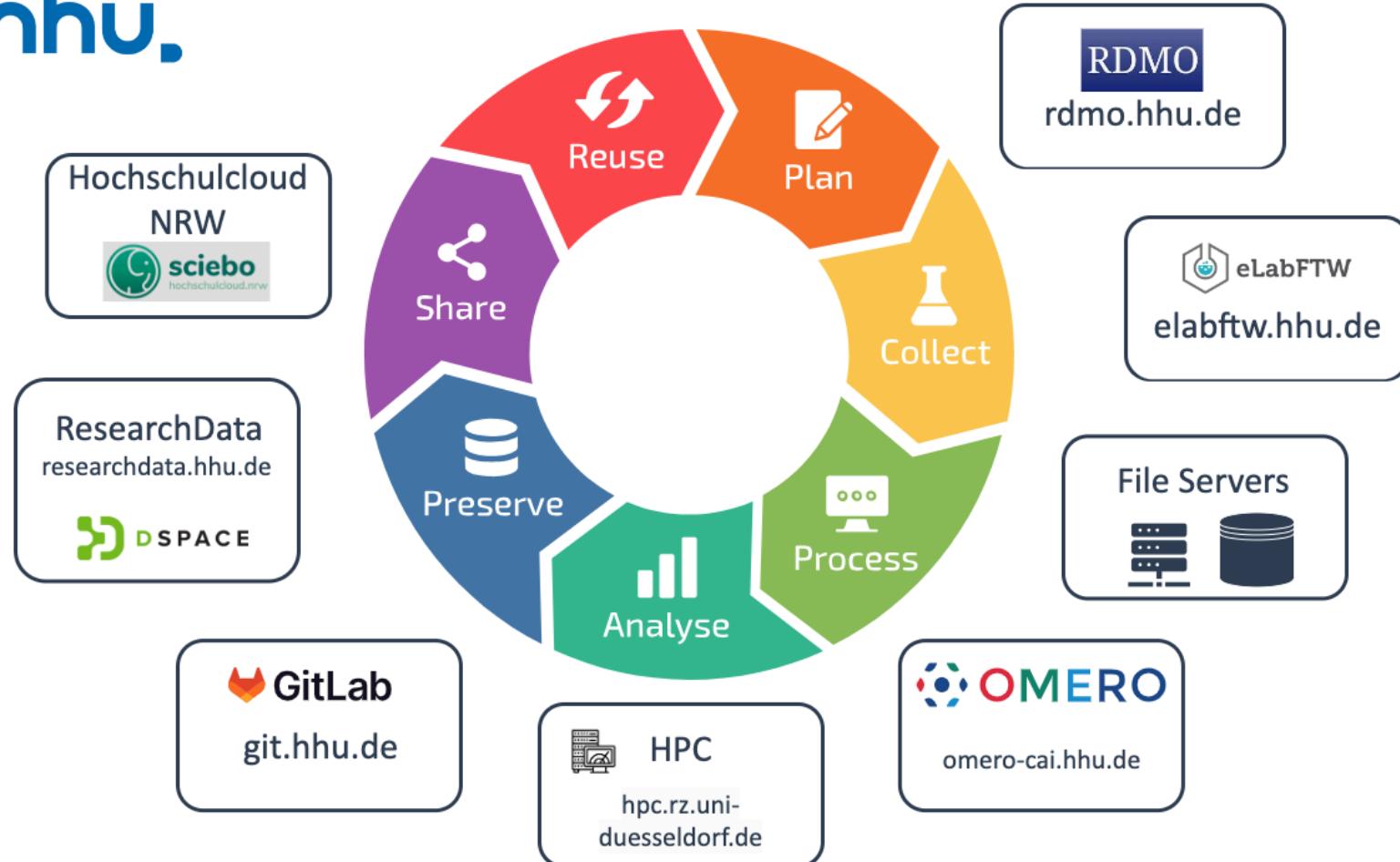


# Data silos impede collaboration



# Missing interfaces impede collaboration

hhu.



# Data Stewardship between DataPLANT and the community

*Community*

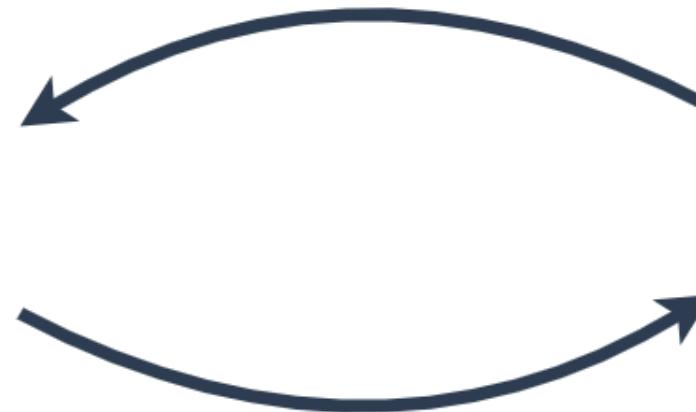


Domain experts  
User experience  
Training

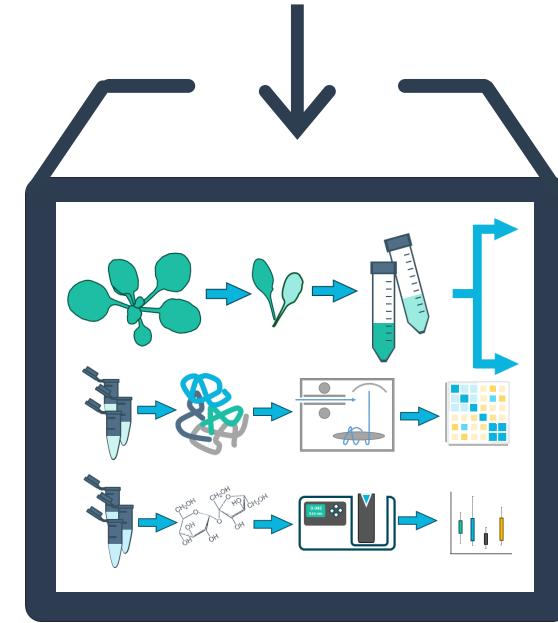
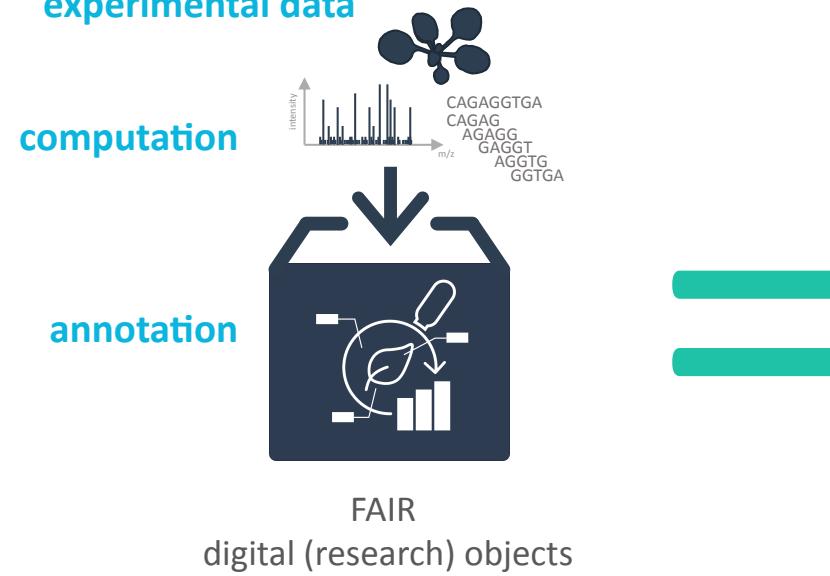
*nfdi4plants*



Service provider  
Developers  
Tech experts

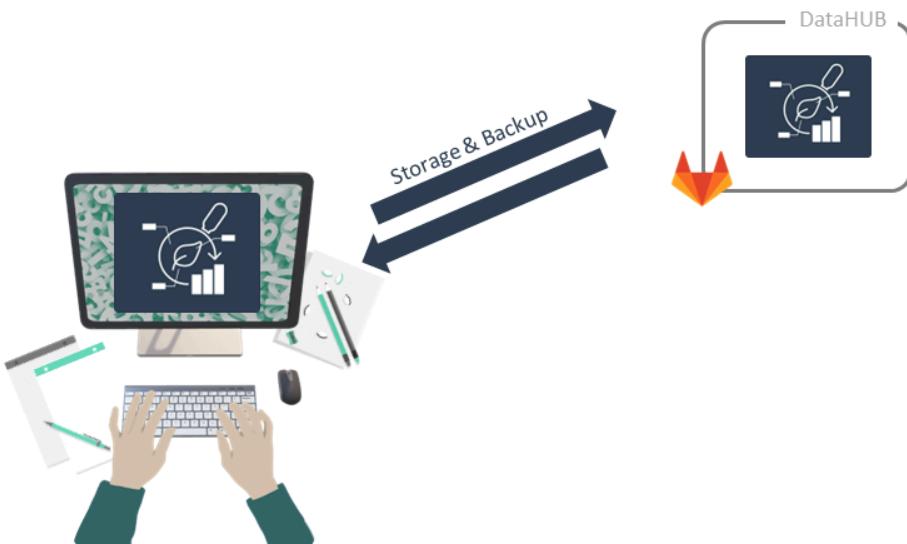


# Annotated Research Context (ARC)

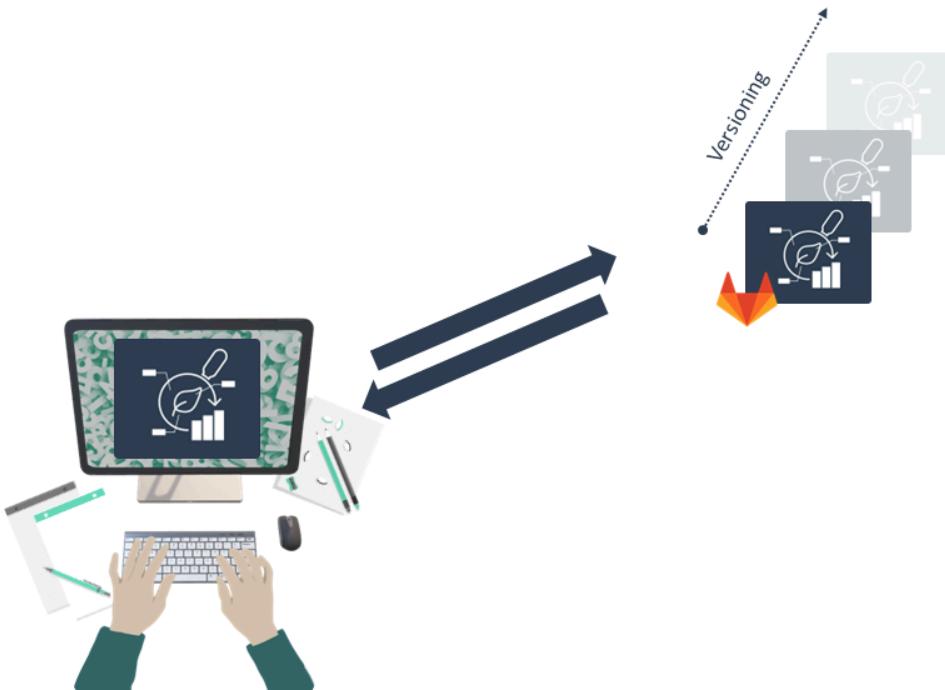


Your entire investigation in a single unified bag

# You can store your ARC in the DataHUB



# ARCs are versioned



# You can invite collaborators



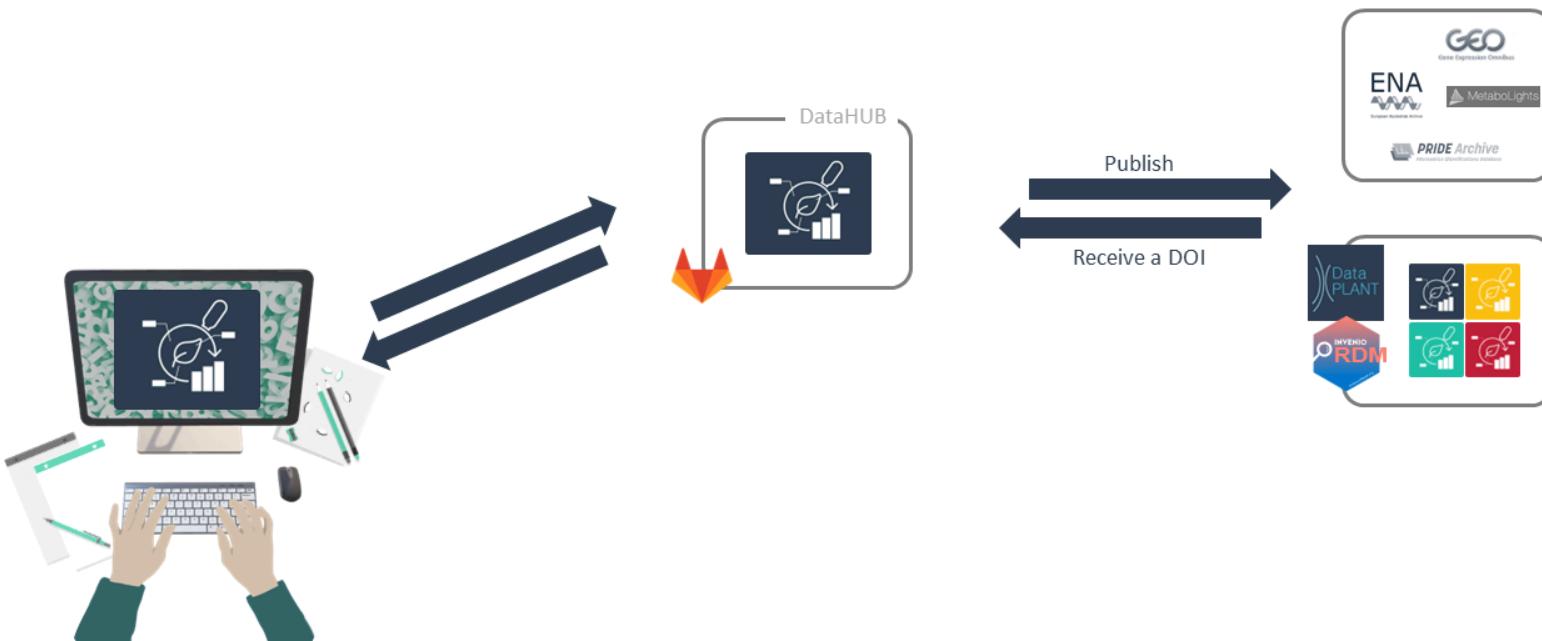
# Collaborate and contribute



# Reuse data in ARCs



# Publish your ARC



# ARC as single-entry point



**specialized endpoints**

ENA  
European Nucleotide Archive

GEO  
Gene Expression Omnibus

PRIDE Archive  
Proteomics Identifications database

EBI: MetaboLights

BioImage Archive

**dataset search**

Google

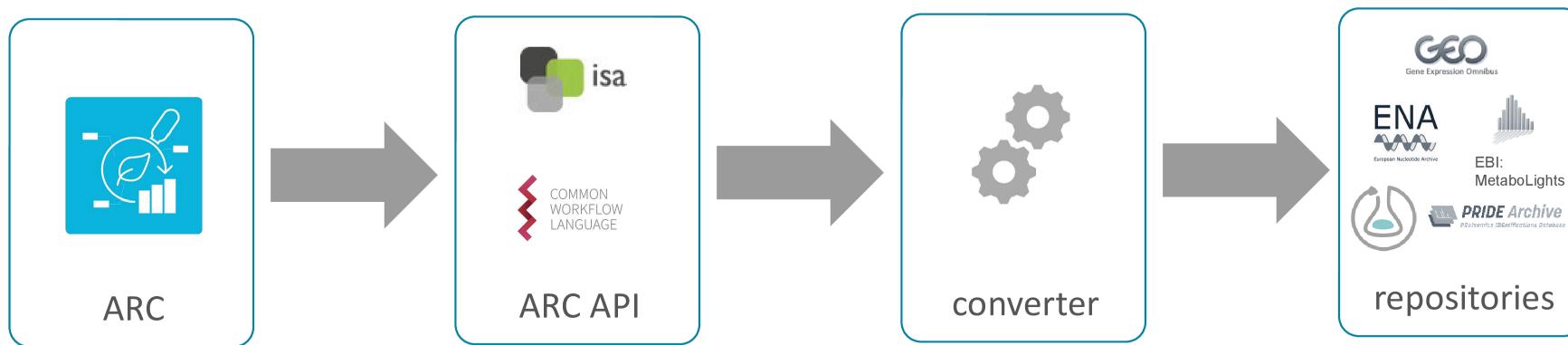
OpenAIRE

**data publication**

The Dataverse® Project

INVENIO

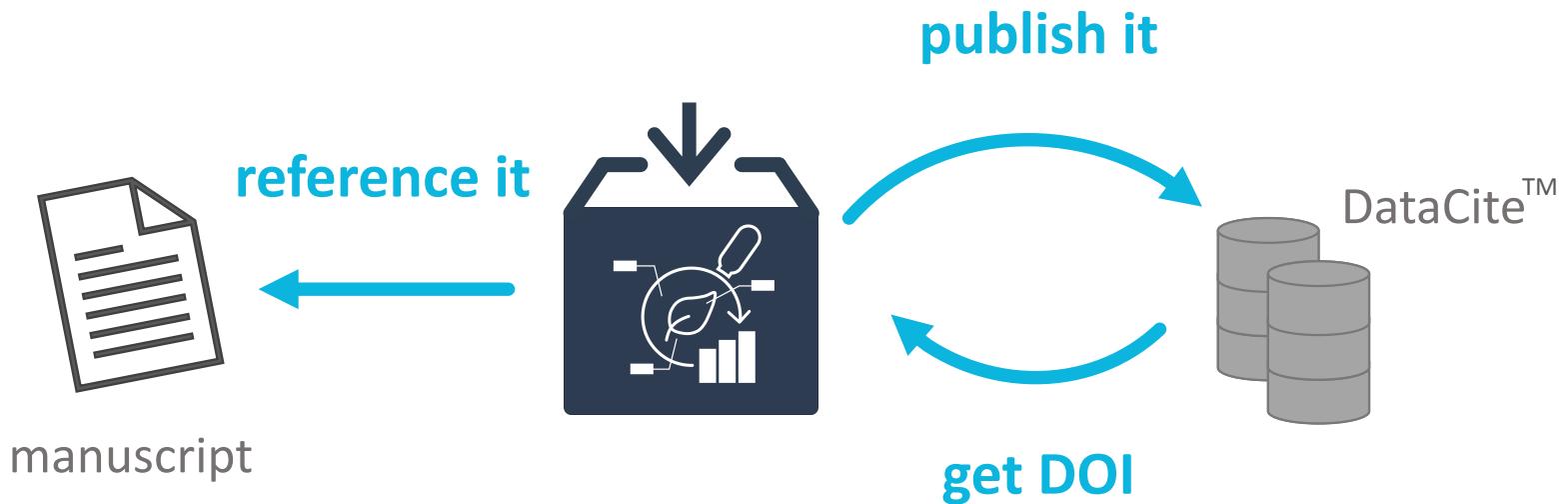
# From ARC to repositories



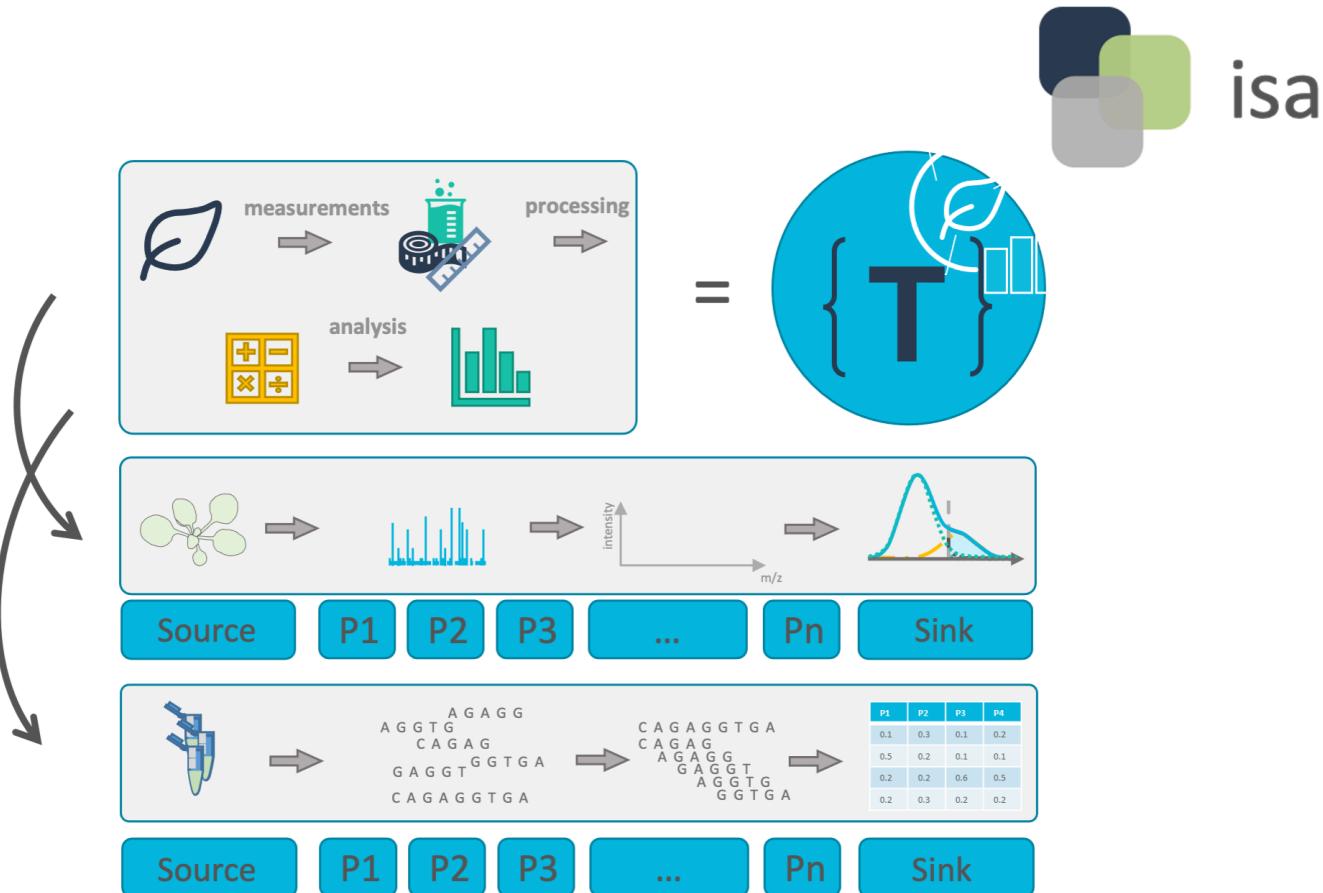
# Moving from paper to data publications



# Publish your ARC, get a DOI

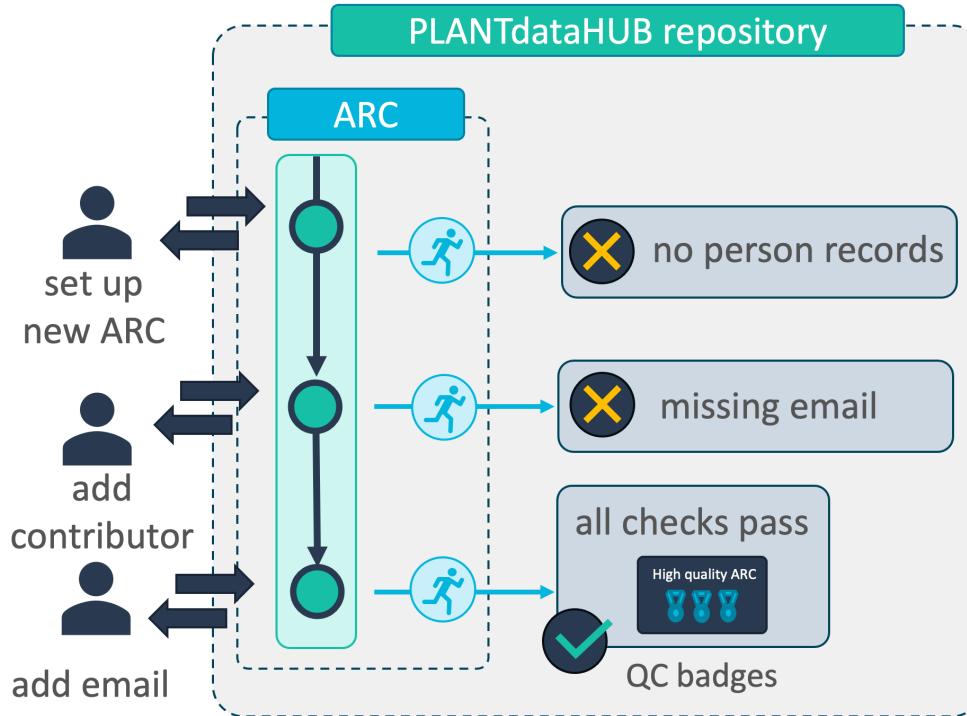


# Metadata templates



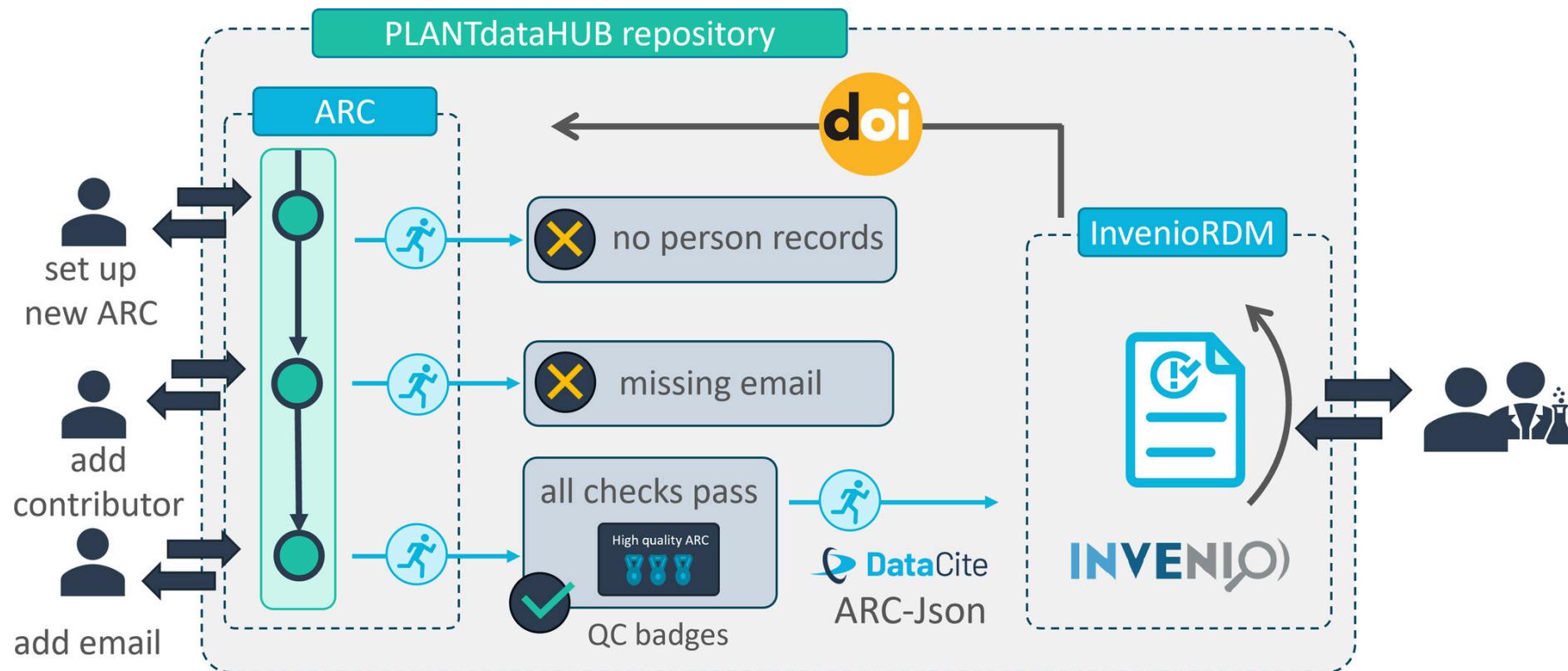
Facilities and labs can define their common workflows as templates

# Validation

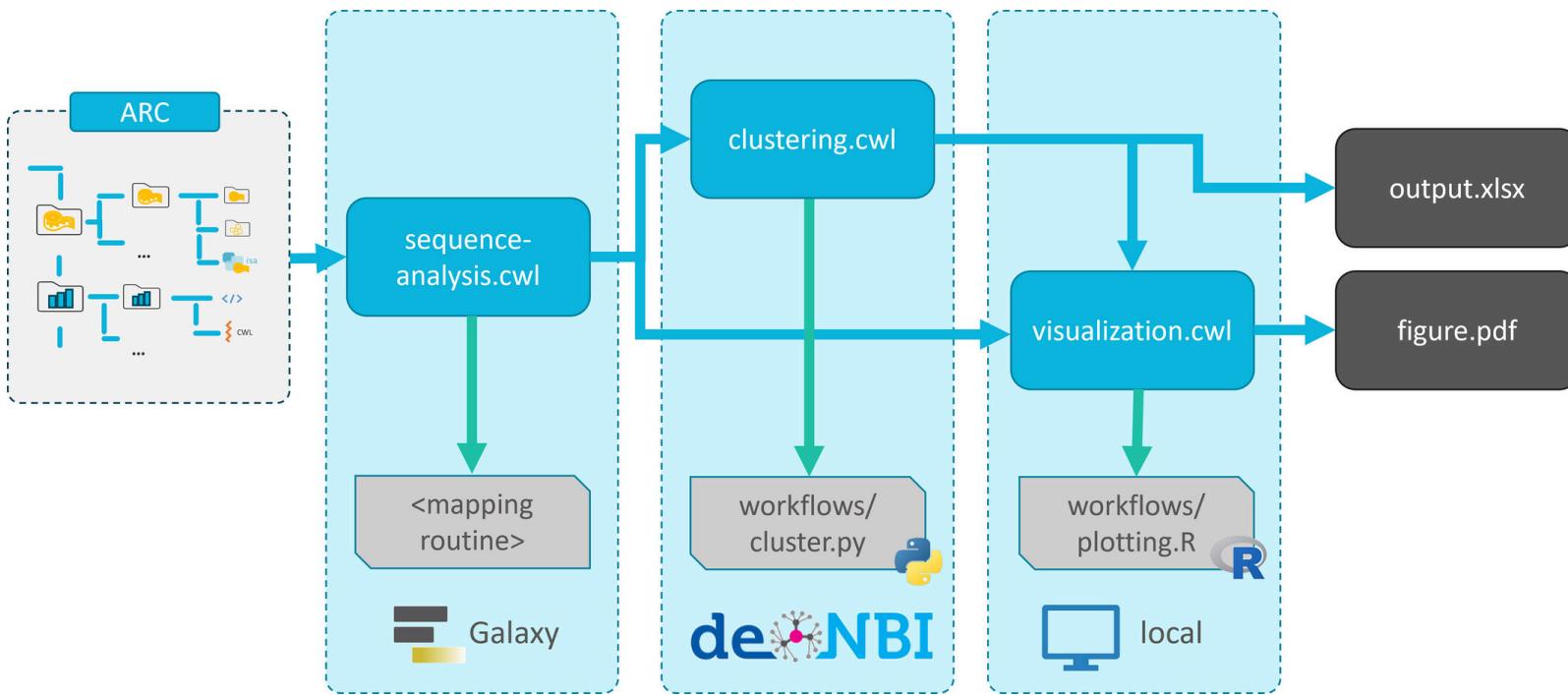


adapted from Weil, H.L., Schneider, K., et al. (2023), PLANTdataHUB: a collaborative platform for continuous FAIR data sharing in plant research. Plant J. <https://doi.org/10.1111/tpj.16474>

# Validate & publish



# Data analysis and workflows



# Galaxy integration: Extra value for plant research

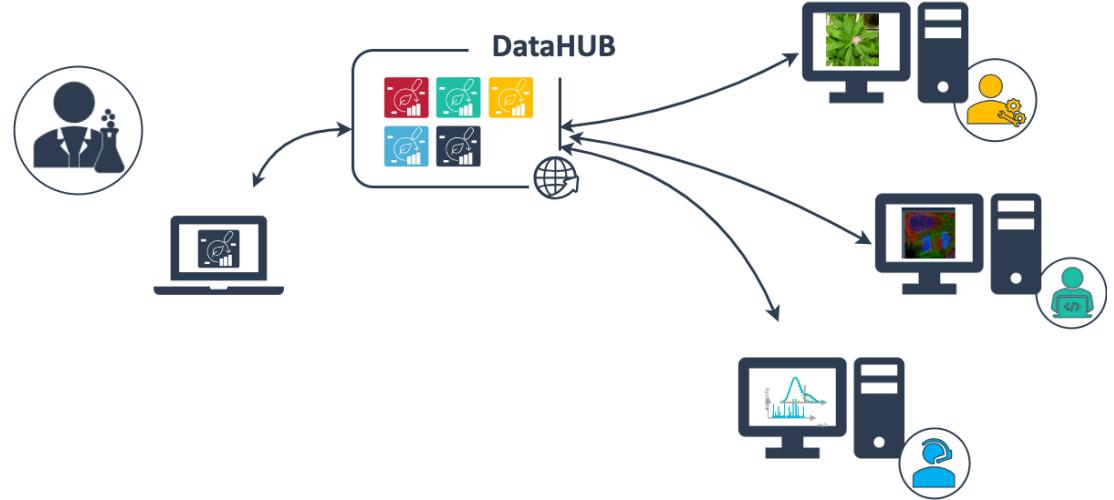


e.g. <https://plants.usegalaxy.eu>

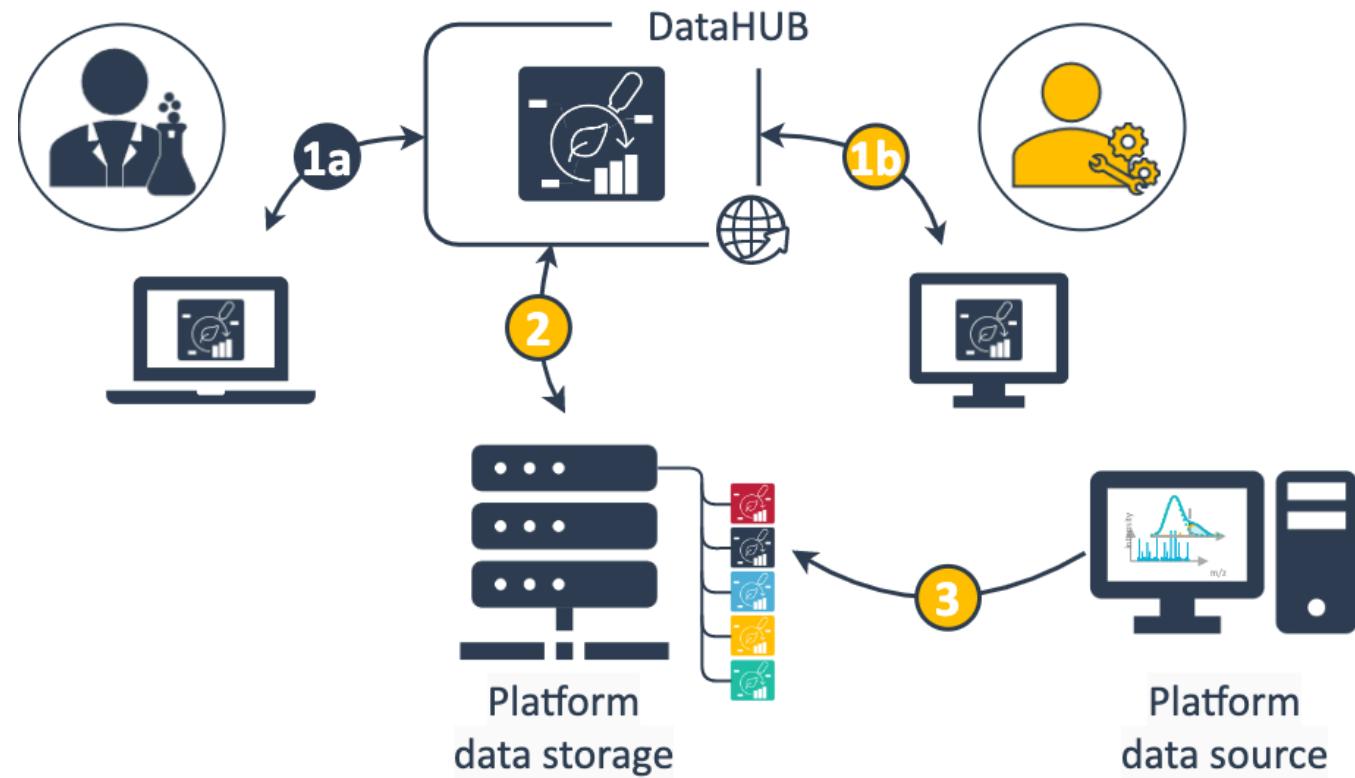
- Full ARC compatibility
- Automated metadata generation
- Specialized tools and workflows for ‘omics processing and analysis
- Public repository compatibility
- Galaxy teaching resource for data analysis

# Enabling platforms

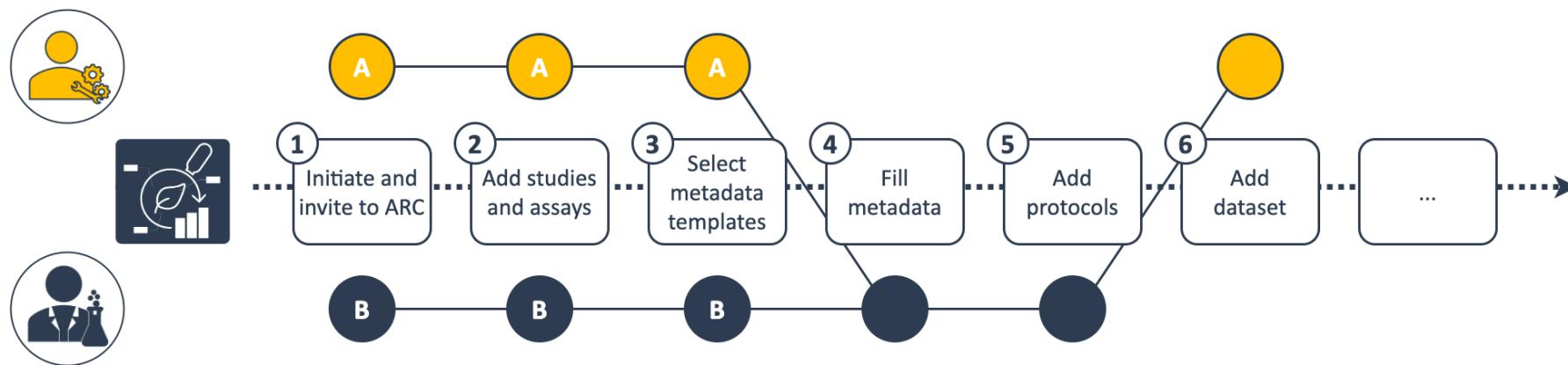
- Streamlined exchange of (meta)data
- Communication and project management



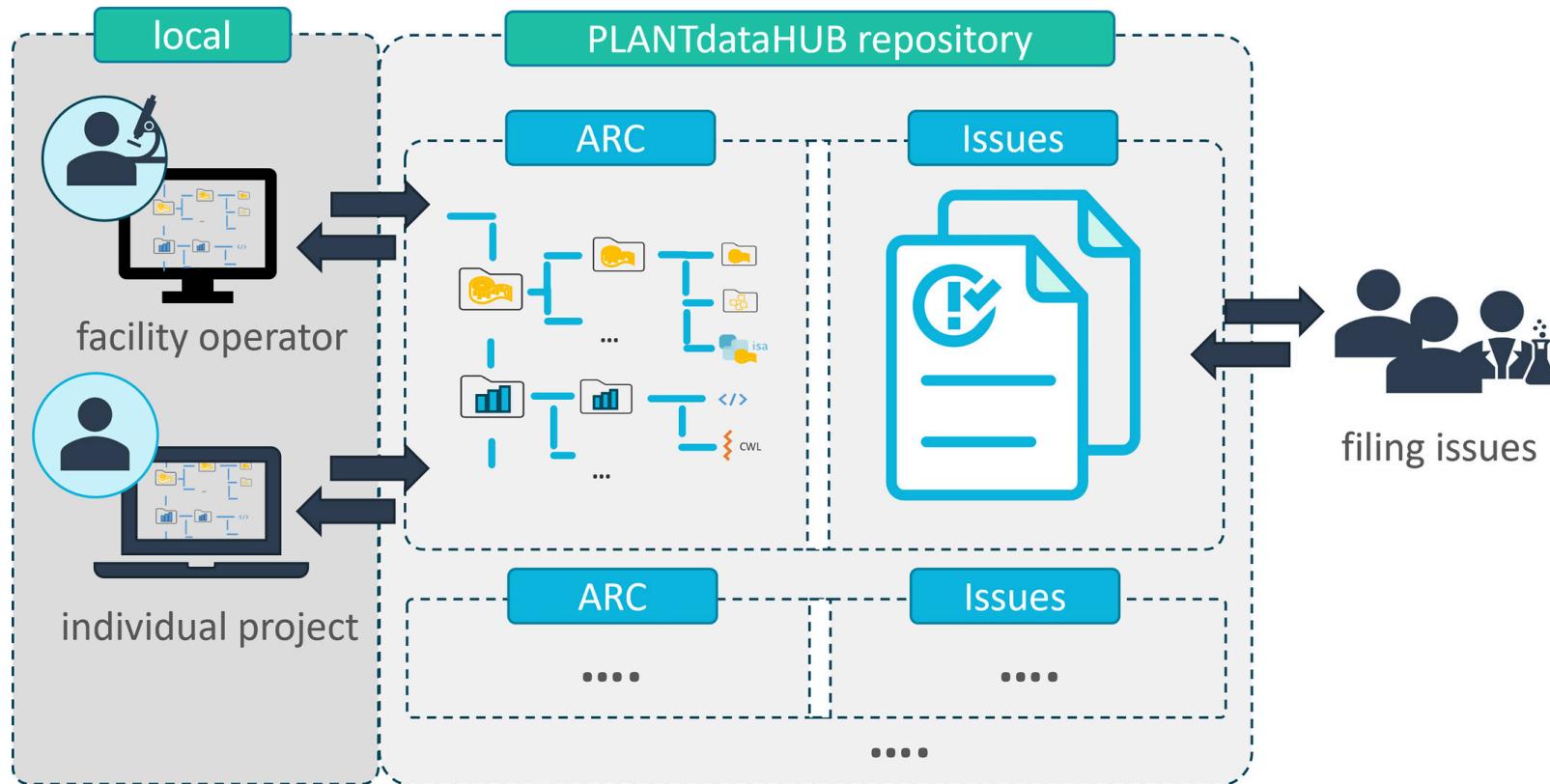
# Streamlined data exchange



# Meet your collaborators in an ARC

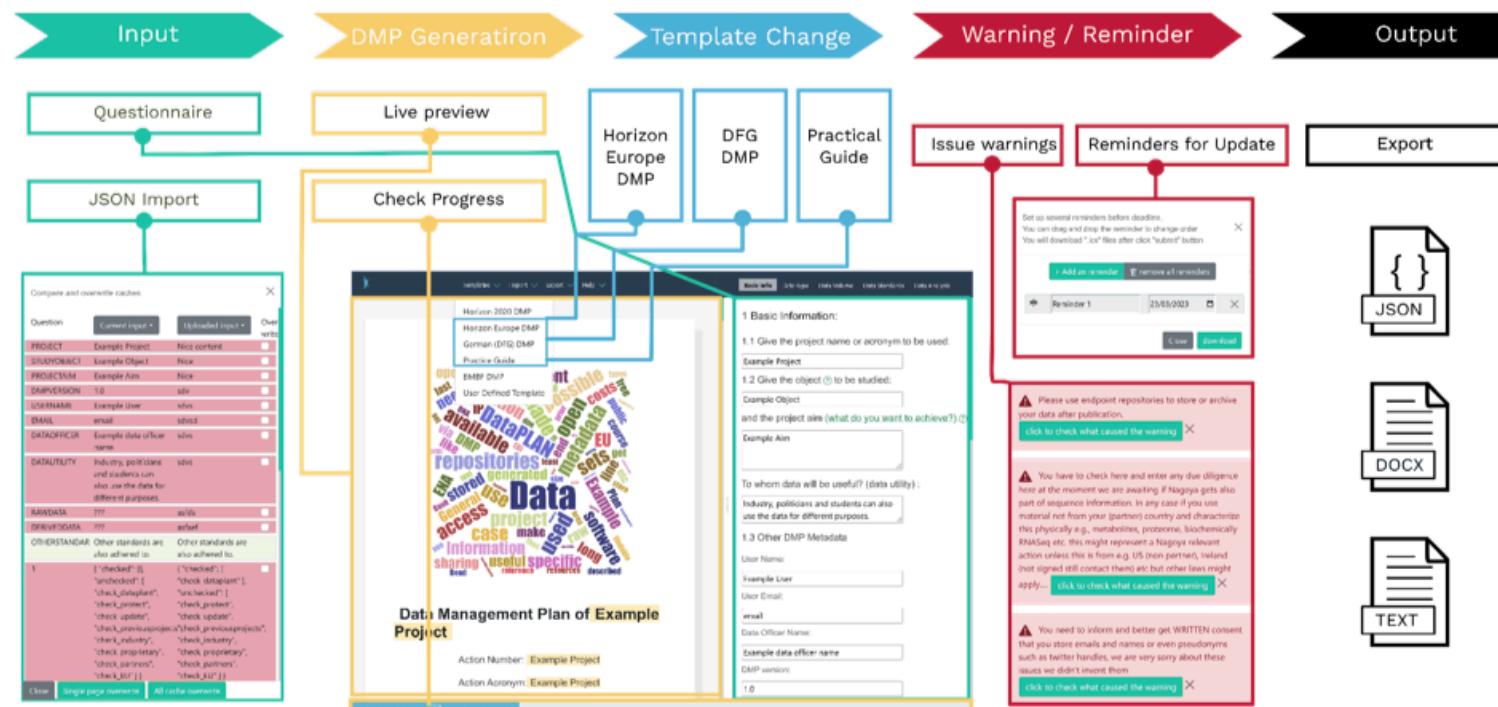


# Project management



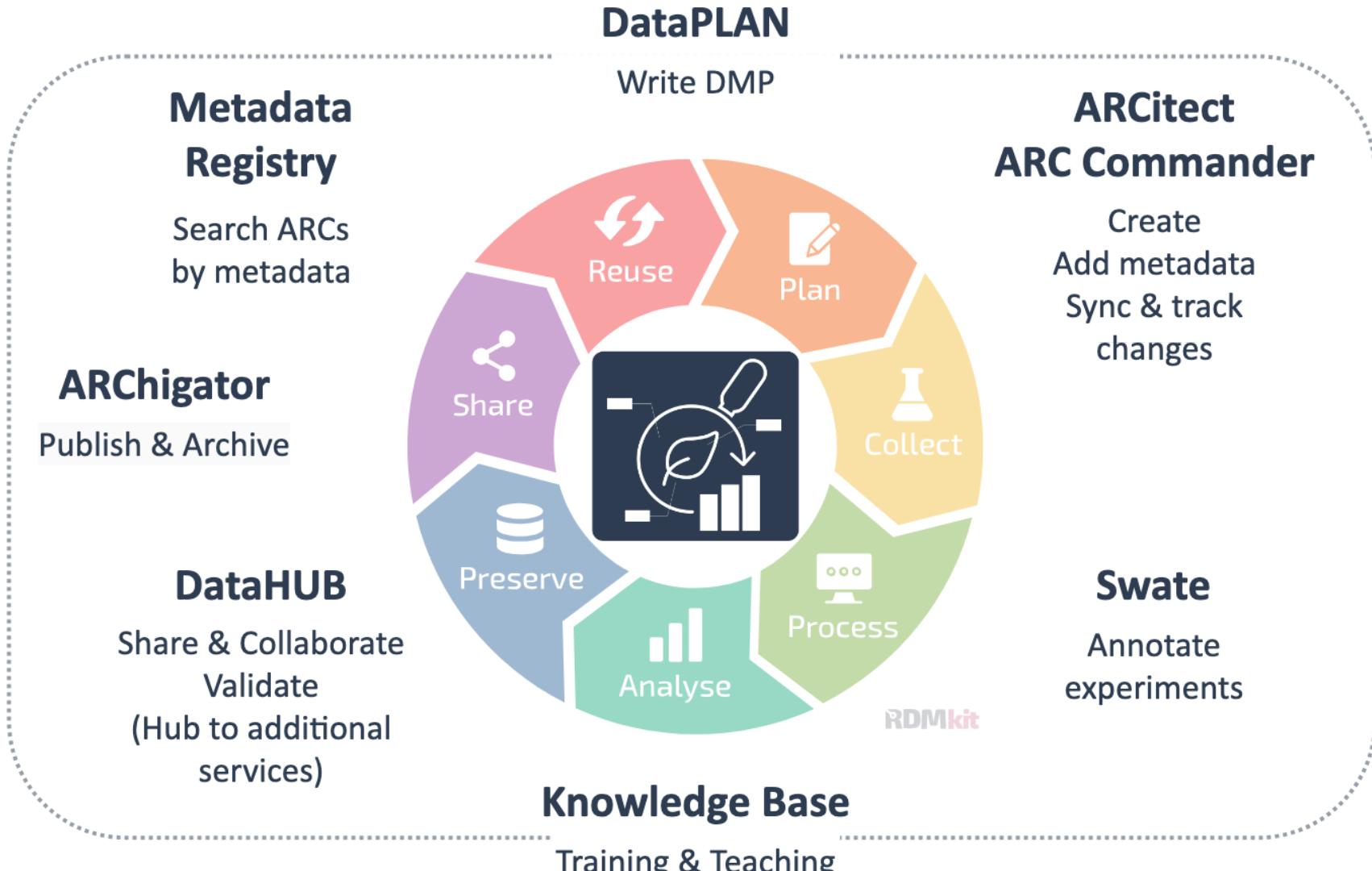
# DataPLAN – a DataPLANT DMP generator

<https://dmpg.nfdi4plants.org>

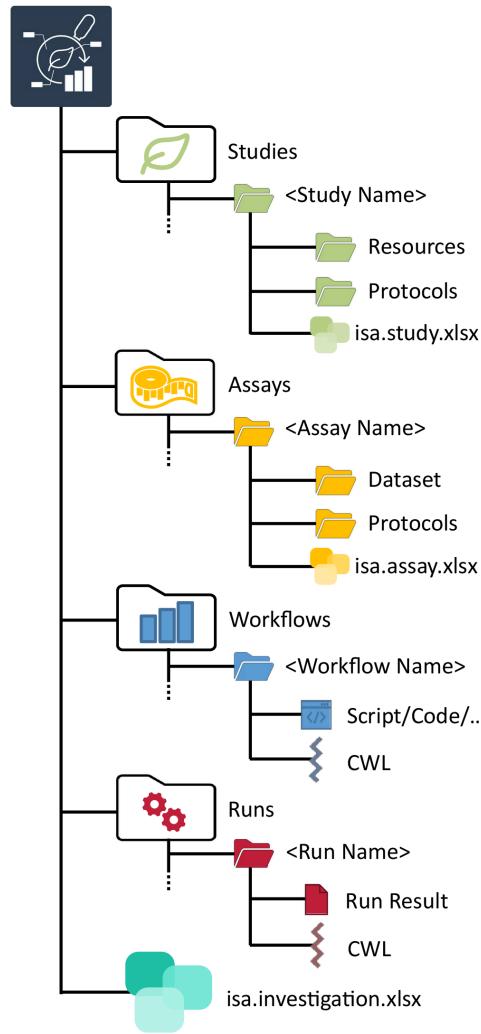


Zhou et al. (2023), DataPLAN: a web-based data management plan generator for the plant sciences, bioRxiv 2023.07.07.548147; doi: <https://doi.org/10.1101/2023.07.07.548147>

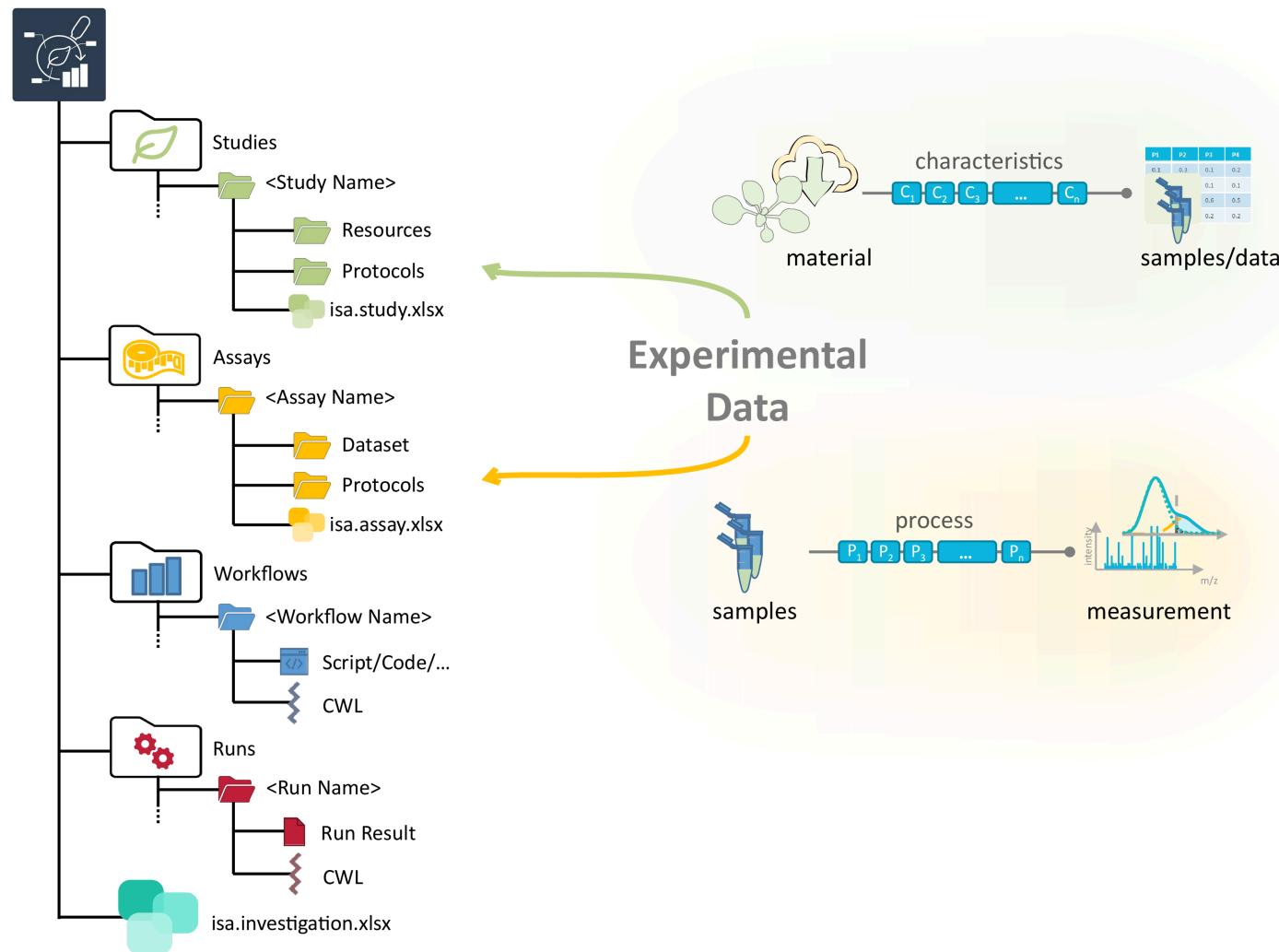
# The ARC ecosystem



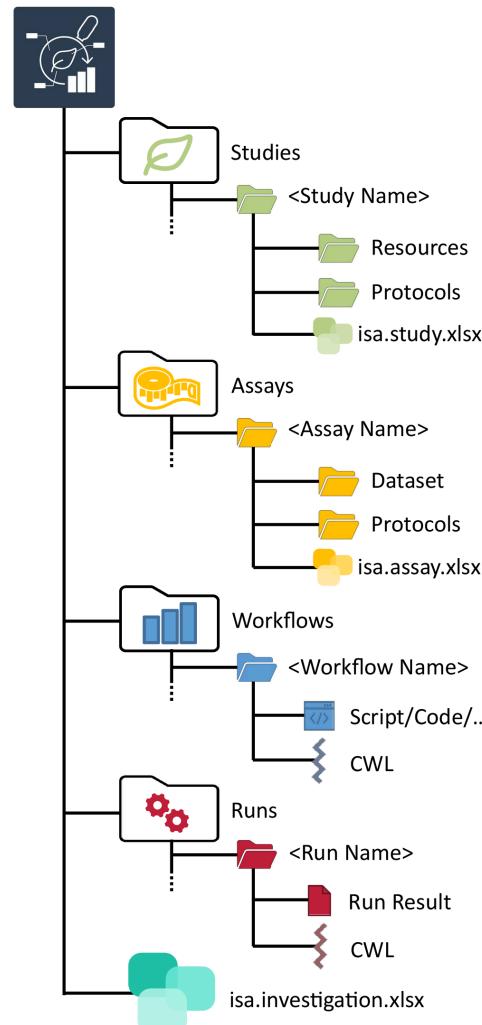
# What does an ARC look like?



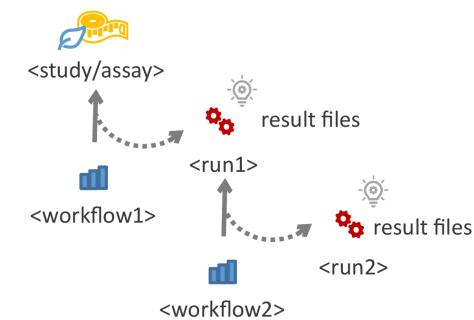
# ARCs store experimental data



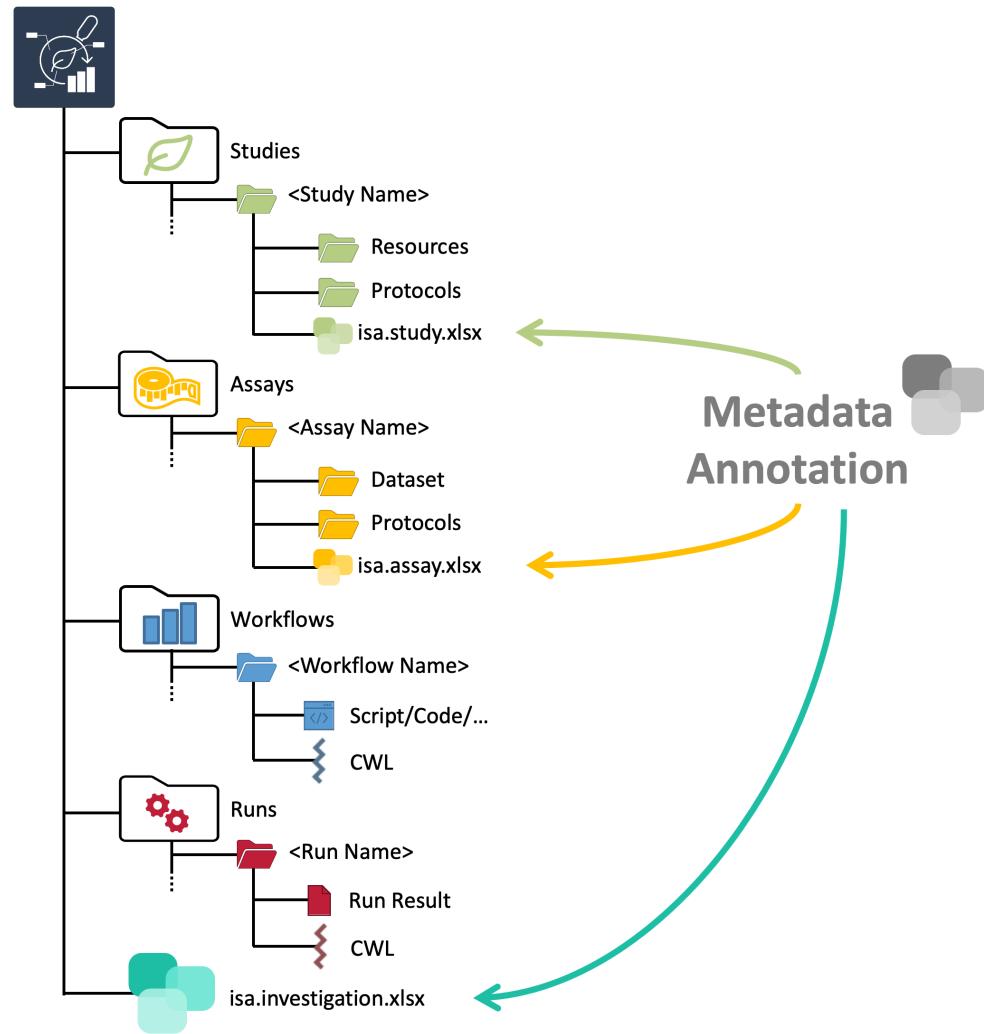
# Computations can be run inside ARCs



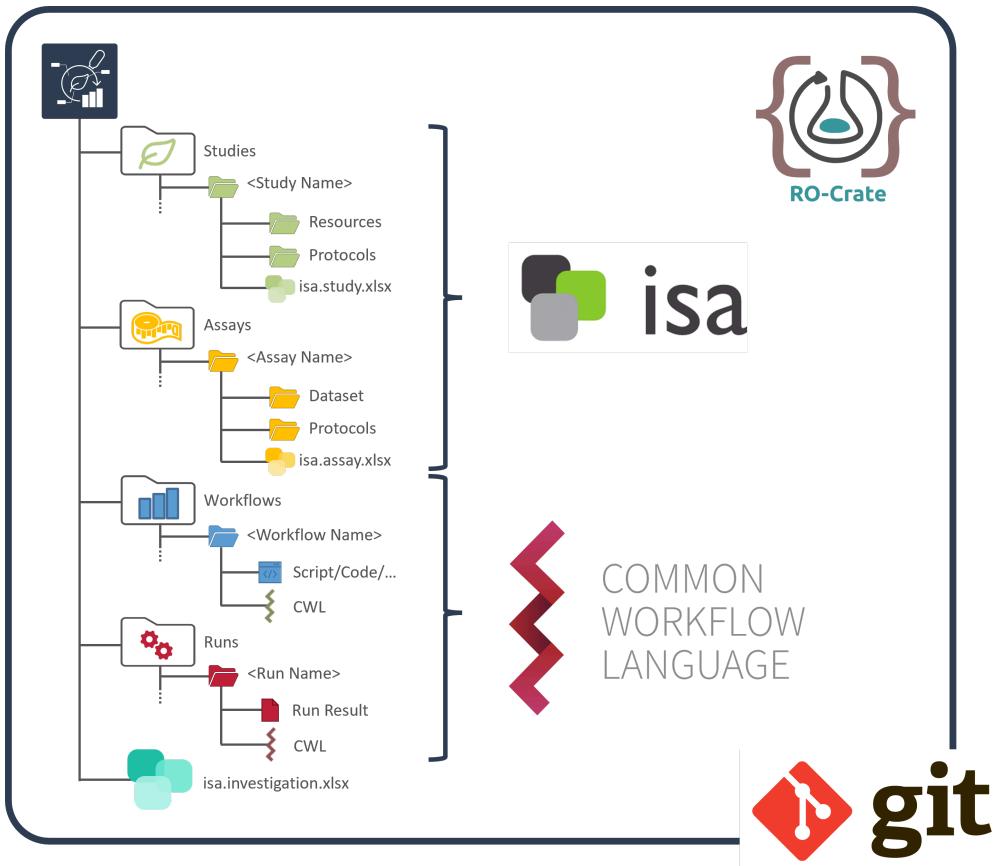
Data analysis  
Computation



# ARCs come with comprehensive metadata



# ARC builds on standards



- RO-Crate: standardized exchange
- ISA: structured, machine-readable metadata
- CWL: reproducible, re-usable data analysis
- Git: version control

## **Hands-on demo**

## ARCitect installation

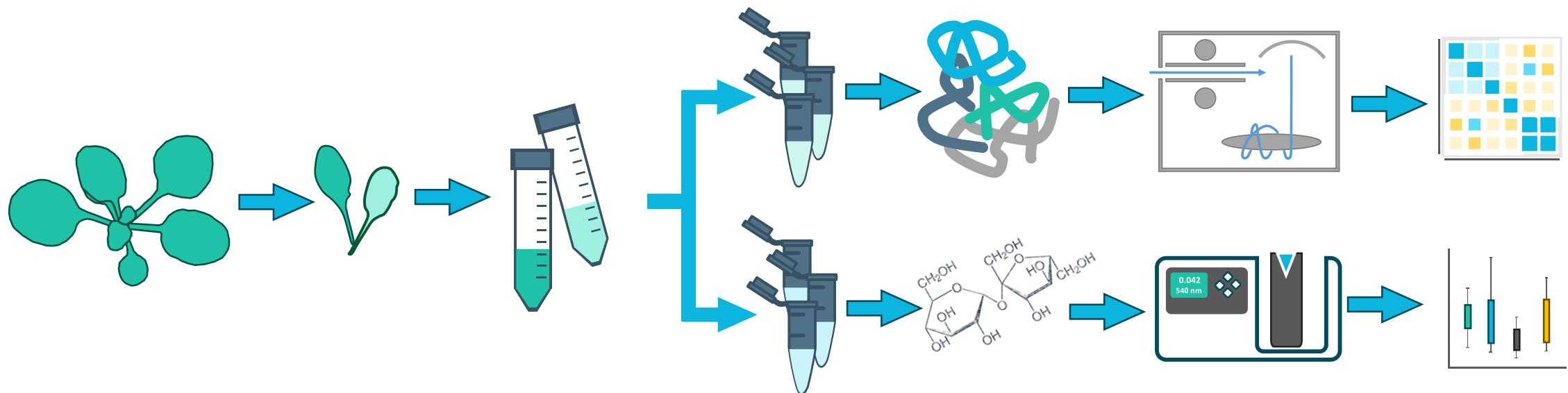
Please install version **v0.0.48** (or later) of the ARCitect:

<https://github.com/nfdi4plants/ARCitect/releases/latest>

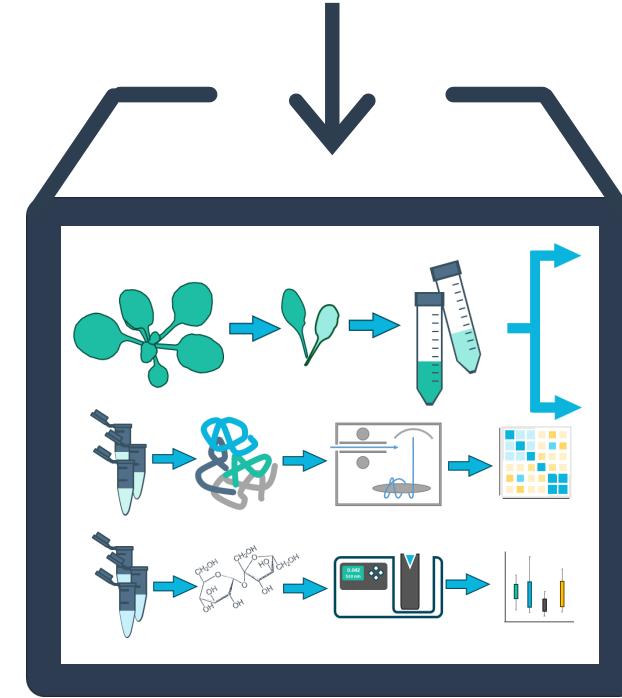
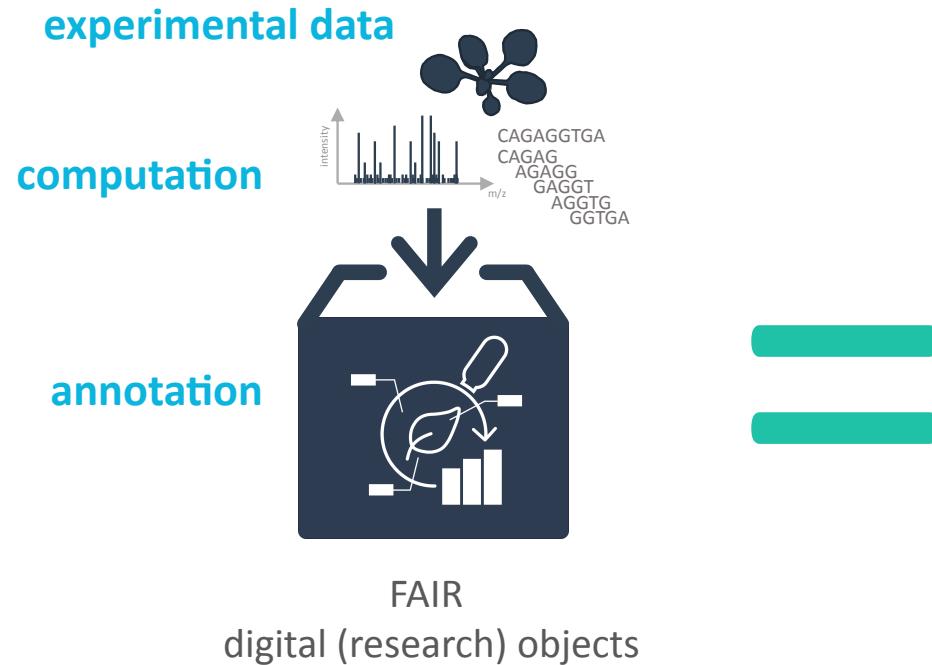
## Download the demo data

<https://uni-duesseldorf.sciebo.de/s/C7ms3QA6q7OZnU2>

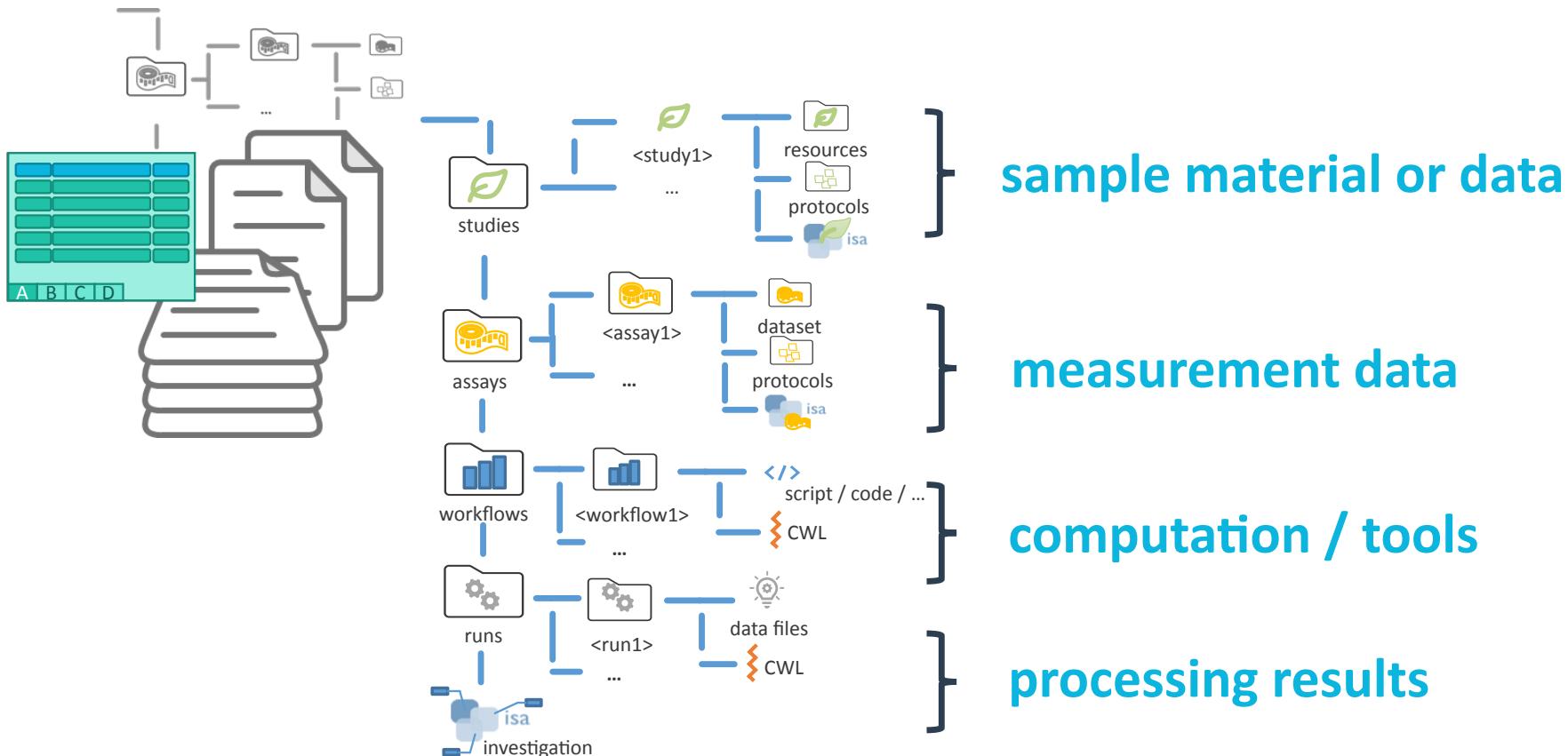
# A small prototypic project



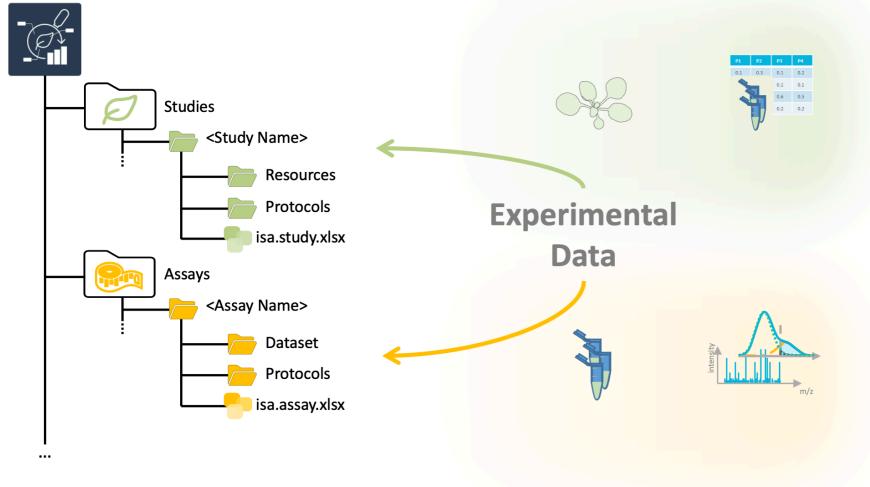
# ARC: Annotated research context



# The ARC scaffold structure



# Sort Demo data in an ARC

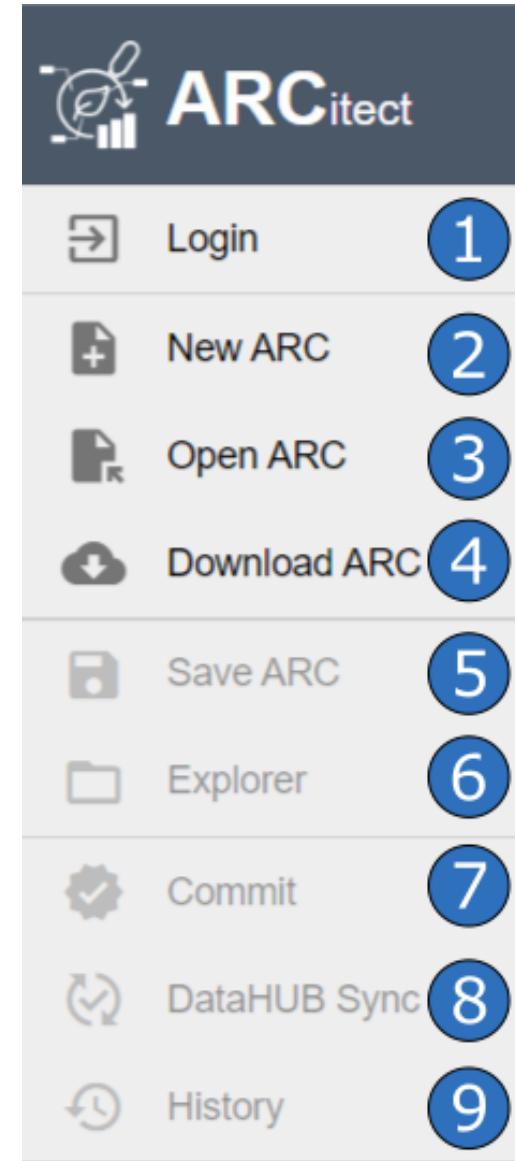


metabolomics_data
150112_56.D
150112_62.D
150112_66.D
150115_12.D
150115_14.D
150115_16.D
gcms_samplelist.tsv
method_gcms.txt
sample_submission_gcms.csv
methods
Illumina_libraries.txt
metabolite_extraction.txt
plant_material.txt
RNA_extraction.txt
rnaseq_data
DB_097_CAGATC_L001_R1_001.fastq.gz
DB_099_CTTGTA_L001_R1_001.fastq.gz
DB_103_AGTCAA_L001_R1_001.fastq.gz
DB_161_GTCCGC_L001_R1_001.fastq.gz
DB_163_GTGAAA_L001_R1_001.fastq.gz
DB_165_GTGAAA_L002_R1_001.fastq.gz
NGS_SampleSheet.xlsx

## Initiate the ARC folder structure

1. Open ARCitect
2. Create a **New ARC** (2)
3. Select a location and name it

**TalinumPhotosynthesis**



## Your ARC's name

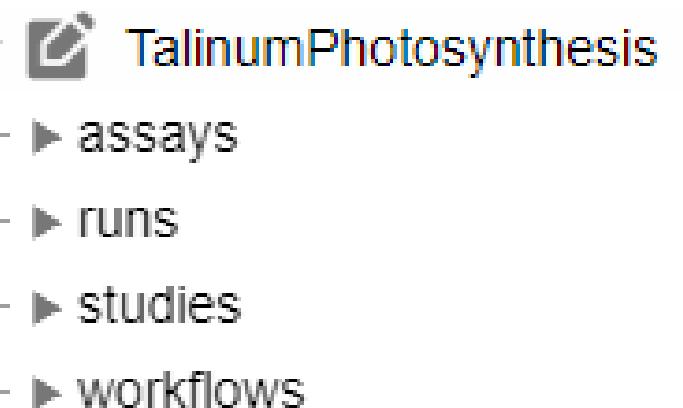
💡 By default, your ARC's name will be used

- for the ARC folder on your machine
- to create your ARC in the DataHUB at  
<https://git.nfdi4plants.org/<YourUserName>/<YourARC>>  
(see next steps)
- as the identifier for your investigation

💡 Make sure that no ARC exists at

<https://git.nfdi4plants.org/<YourUserName>/<YourARC>> .

Otherwise you will sync to that ARC.



# Add a description to your investigation

C:/Users/Sabrina/Desktop/Workshops ARCs  
/TalinumPhotosynthesis

## TalinumPhotosynthesis

- assays
- runs
- studies
- workflows
-  .gitignore

1.34 KB

Identifier

Title

Description

This is a very interesting investigation about life and photosynthesis

# Add contributors

- you could also add the contacts directly via ORCID

## Contacts

Your First Name Your Last Name

Your ORCID

6/10 ▾

First Name

Last Name

Your First Name

Your Last Name

Mid Initials

ORCID

Your ORCID

Search

Affiliation

Address

Your Affiliation

Email

Phone

Fax

yourEmailAdress@uni.de

Roles

1.

Author



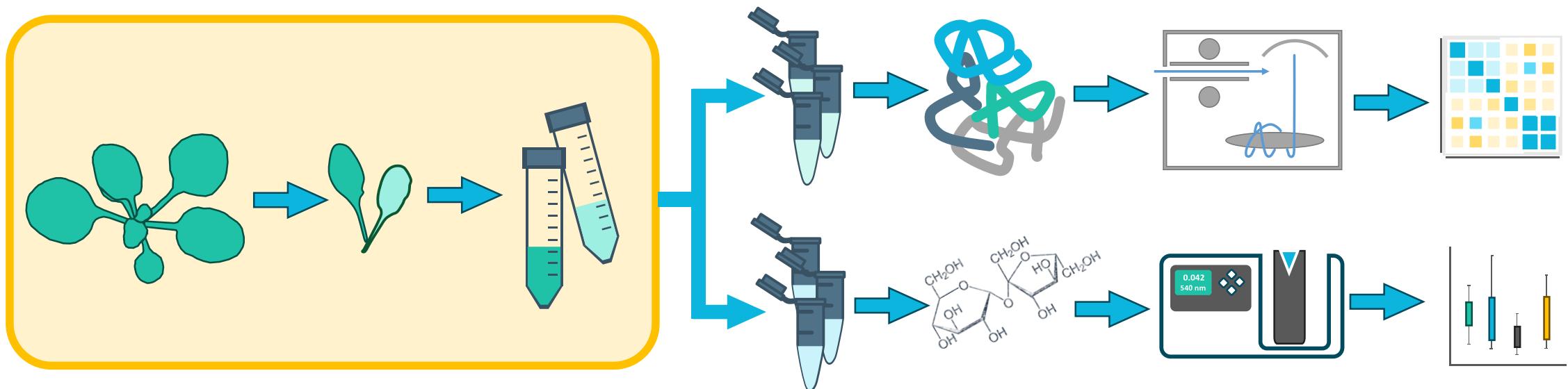
NCIT

NCIT:C42781

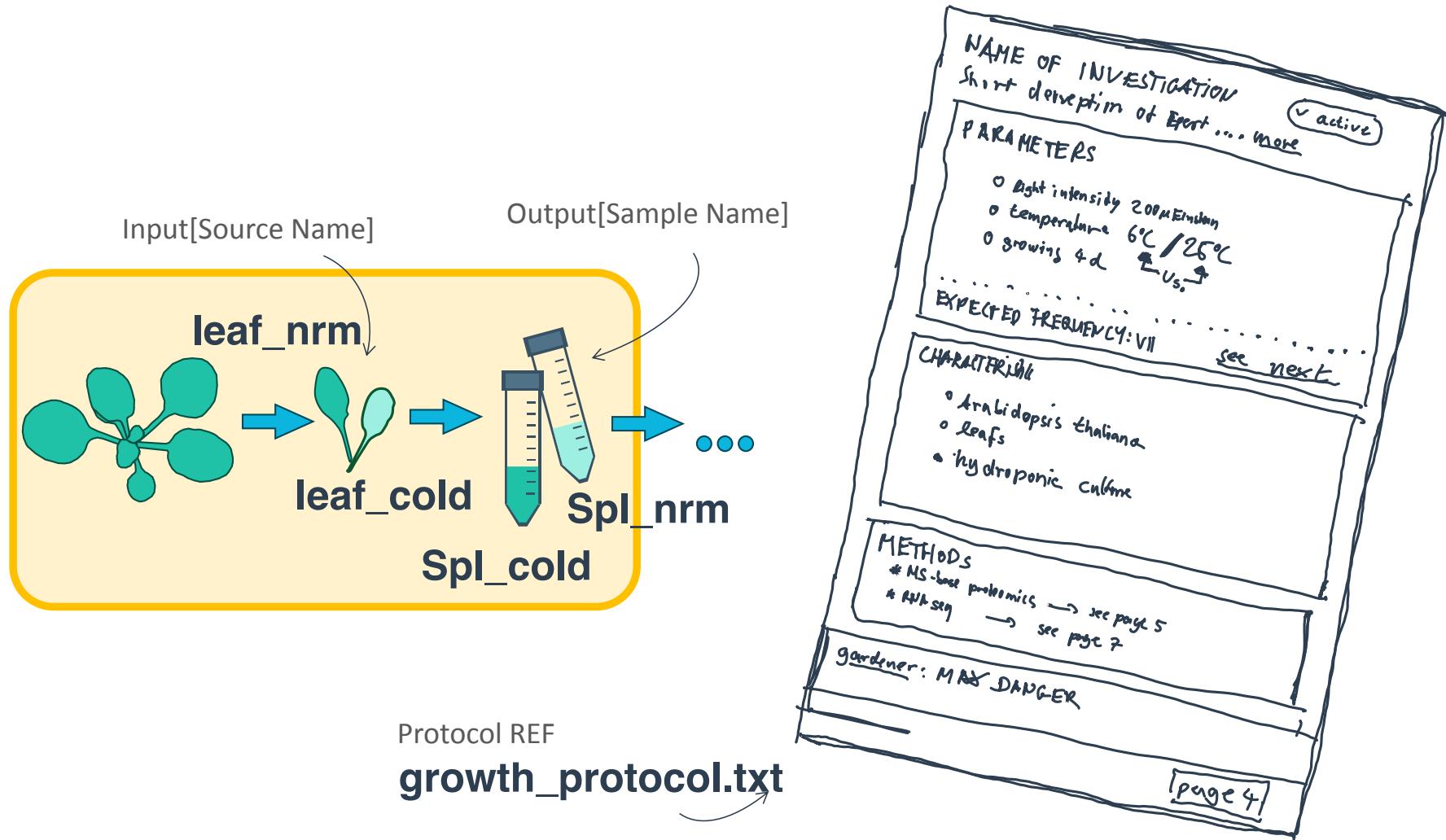


Delete

# Divide and conquer for reproducibility



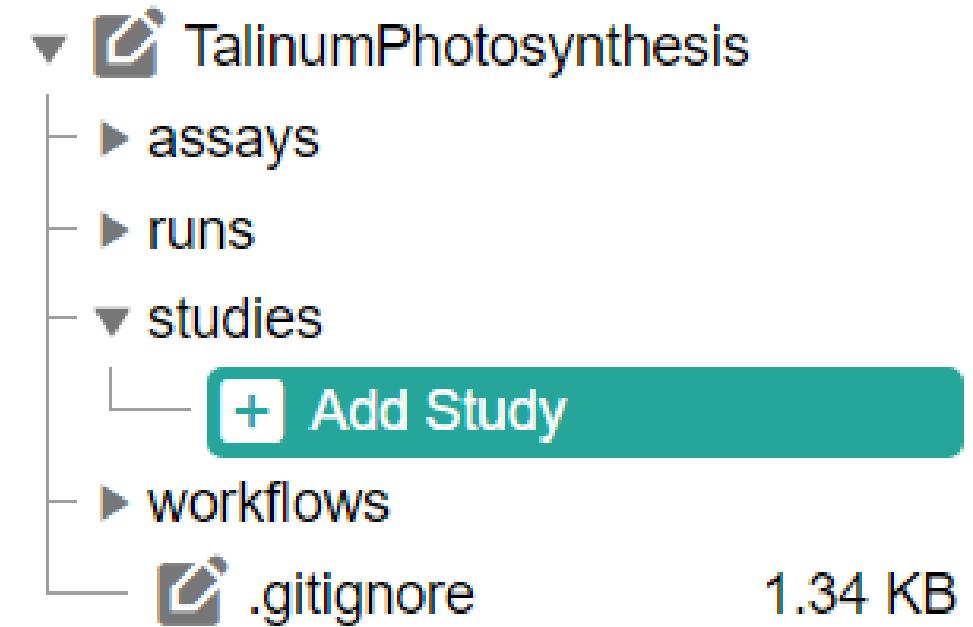
# Identifying the 'study' part



## Add a study

by clicking "Add Study" and entering an identifier for your study

Use **talinum\_drought** as an identifier



# Study panel

In the study panel you can add

- general metadata,
- people, and
- publications
- data process information

**Identifier**  
talinum\_drought

**Description**

**Contacts**

**Publications**

**Submission Date**  
tt.mm.jjjj --:--

**Public ReleaseDate**  
tt.mm.jjjj --:--

**Study Design Descriptors**

## Let's annotate the plant samples

1. Check out the lab notes

```
studies/talinum_drought/protocols/plant_material.txt
```

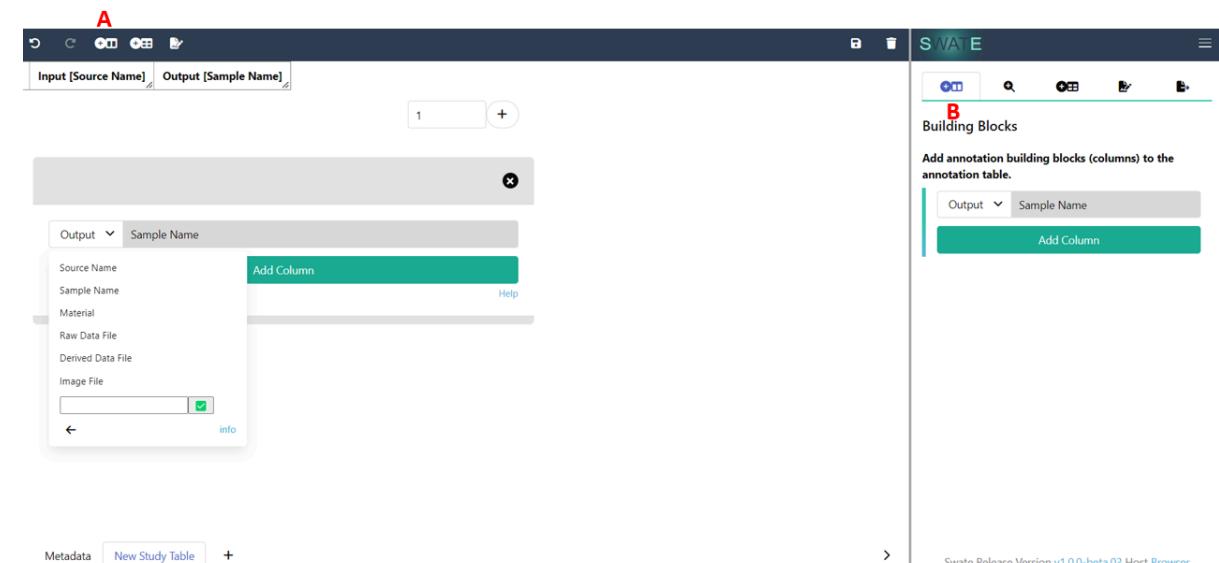
2. Select the study talinum\_drought

3. Add a new table sheet at the bottom

# Create an annotation table

1. Create a annotation table by adding **Building blocks** via the widget (A) or the sidebar (B)
2. Add an *Input* ( Source Name ) and *Output* ( Sample Name ) column

 Each table can contain only one *Input* and one *Output* column



# A table-based organization schema

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

## Fill out source name and sample name

Transfer the sample ids from the protocol.

1. Invent names for **Source Name** (we do not have this information)
2. Use the sample names from the protocol (DB\_\*) as **Sample Name**

# Add protocols

You can either

- directly write a **new protocol** within the ARCitect or
- import an existing one from your computer

## Create or Import Protocol

Protocol Name



NEW PROTOCOL



IMPORT PROTOCOL

CANCEL

## Link the protocol to the isa table

1. In the *Building Blocks* tab, select *More* -> *Protocol REF*.
2. Click  *Add Column*.
3. Add the name of the protocol file (`plant_material.txt`) to the *Protocol REF* column.

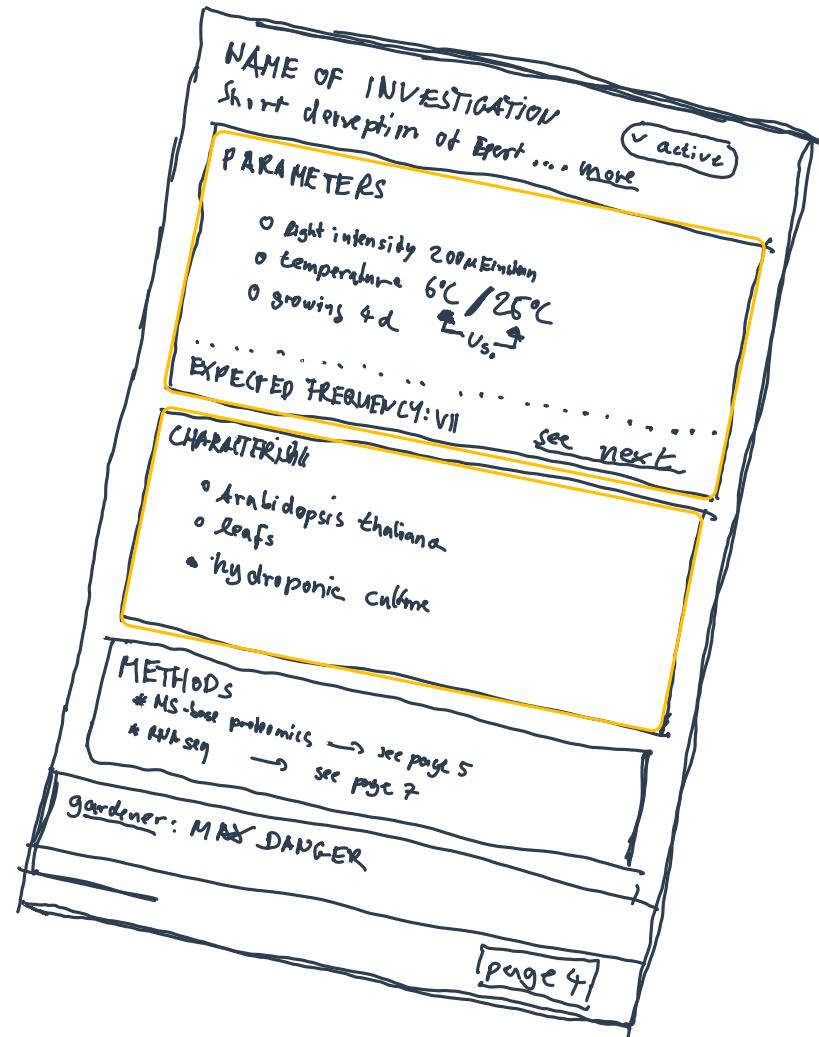
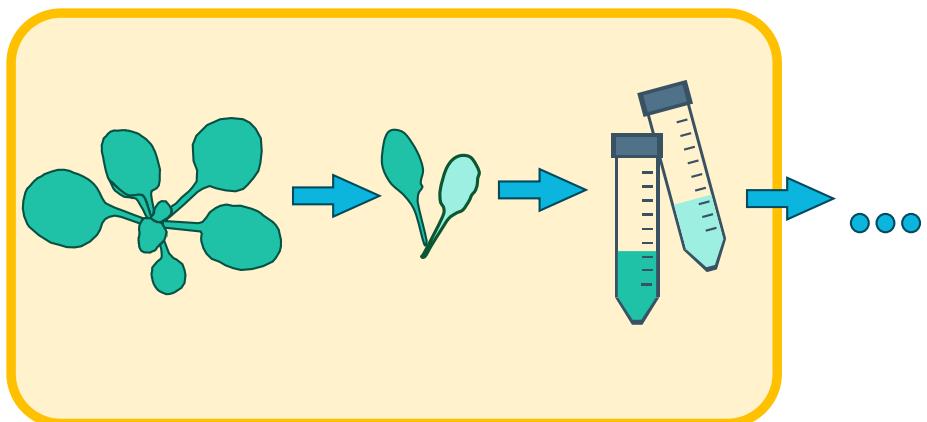
# Referencing a protocol

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

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		

- 💡 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

# 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](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](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](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

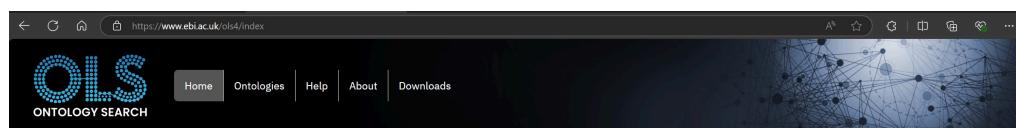
temperature PATO:0000146 

[http://purl.obolibrary.org/obo/PATO\\_0000146](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](#)

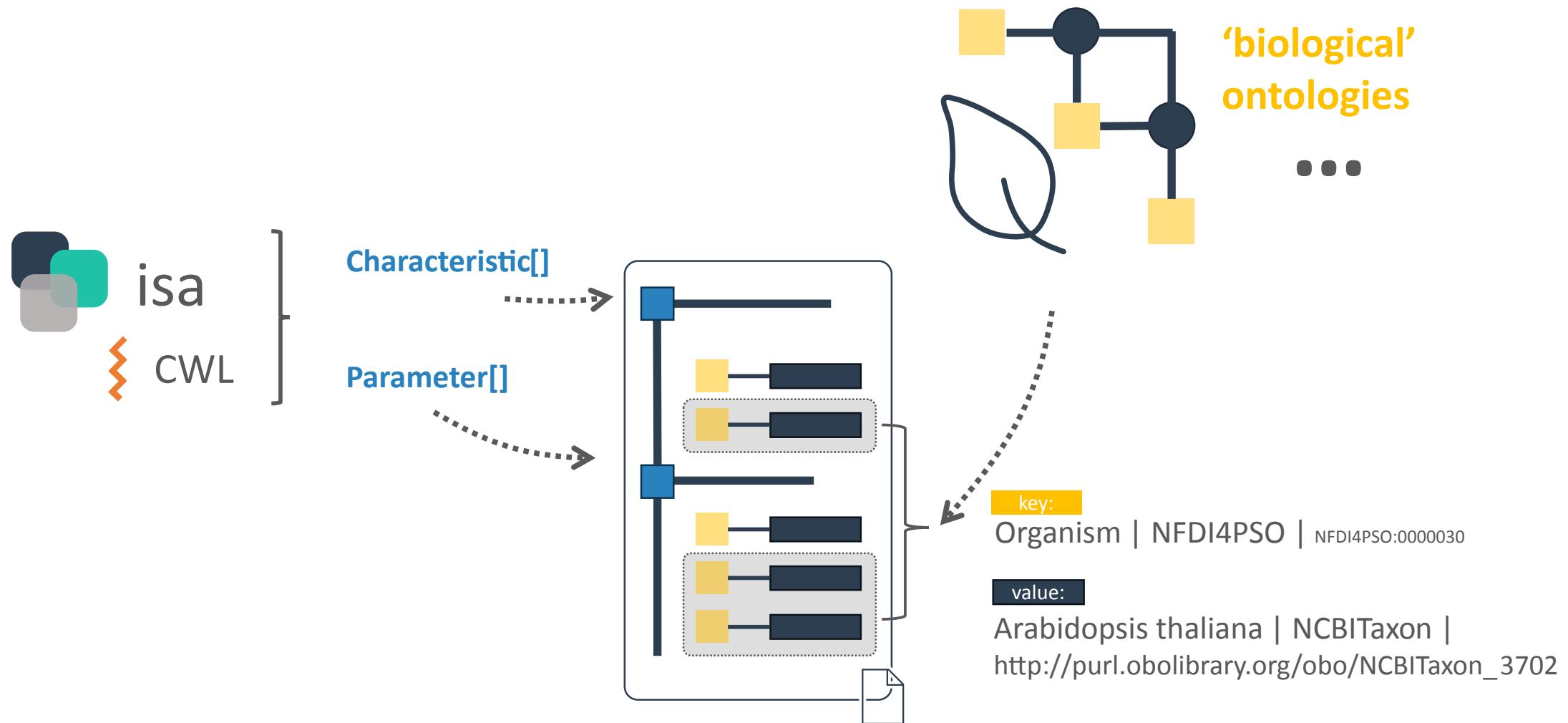
In addition to OLS the SPOT team also provides the OxO and ZOOMA services. OxO provides cross-ontology mappings between terms from different ontologies. ZOOMA is a service to assist in mapping data to ontologies in OLS.

 [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.



# 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 µEinstein
- █ Temperature
  - █ 6°C / 25°C
- █ Growth time
  - █ 4d

## Characteristics []

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

# Annotation Building Block types

- Input (e.g. Source Name, Sample Name)
- Protocol columns
- Characteristic // Parameter // Factor //
- Component
- Output (e.g. Sample Name, Raw Data File, Derived Data File)

The screenshot shows a software interface for managing annotation building blocks. At the top, there's a toolbar with various icons. Below it is a header bar labeled "Widgets". The main area contains a table with six rows of data:

Input [Source Name]	Characteristic [organism]	Factor [watering exposure]	Output [Sample Name]
DB_097	Talinum fruticosum	✓ 12 days drought	CAM_01
DB_099	Talinum fruticosum	✓ 12 days drought	CAM_02
DB_103	Talinum fruticosum	✓ 12 days drought	CAM_03
DB_161	Talinum fruticosum	✓ 12 days drought + 2 days rewetared	reC3_01
DB_163	Talinum fruticosum	✓ 12 days drought + 2 days rewetared	reC3_02
DB_165	Talinum fruticosum	✓ 12 days drought + 2 days rewetared	reC3_03

Annotations are present in the "Factor" column:

- Row 1: "Factor" is highlighted in blue.
- Row 2: "12 days drought" has a red arrow pointing to the "Characteristics" label below the table.
- Row 3: "12 days drought" has a red arrow pointing to the "Characteristics" label below the table.
- Row 4: "12 days drought + 2 days rewetared" has a red arrow pointing to the "Characteristics" label below the table.
- Row 5: "12 days drought + 2 days rewetared" has a red arrow pointing to the "Characteristics" label below the table.
- Row 6: "12 days drought + 2 days rewetared" has a red arrow pointing to the "Characteristics" label below the table.

To the right of the table is a "New Parameter" dialog box with tabs for "Parameter", "Factor", "Characteristic", "Component", "More", and "Output". The "Output" tab is selected. A "Sidebar" button is located at the bottom right of the dialog. The sidebar itself is visible on the far right, showing sections for "Parameter", "Instrument model", "Add Column", and "Help".

Let's take a detour on [Annotation Principles](#)

# Add Characteristics

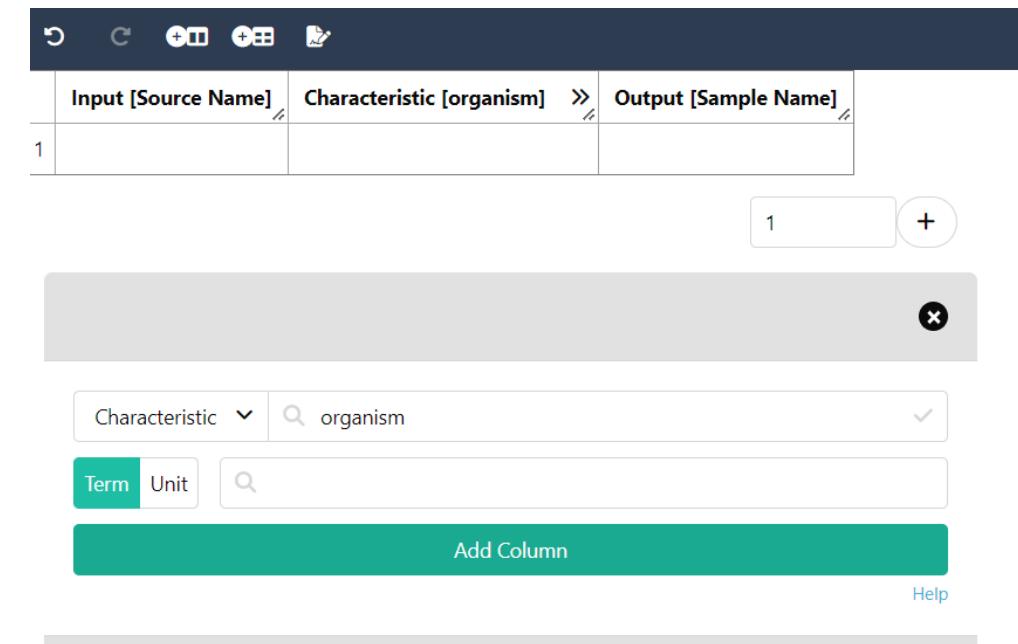
1. Select *Characteristic* from the drop-down menu

2. Enter **organism** in the search bar. This search looks for suitable *Terms* in our *Ontology* database.

3. Select the Term with the id **OBI:0100026** and,

4. Click **Add Column**

 This adds four columns to your table, one visible and **three** hidden.



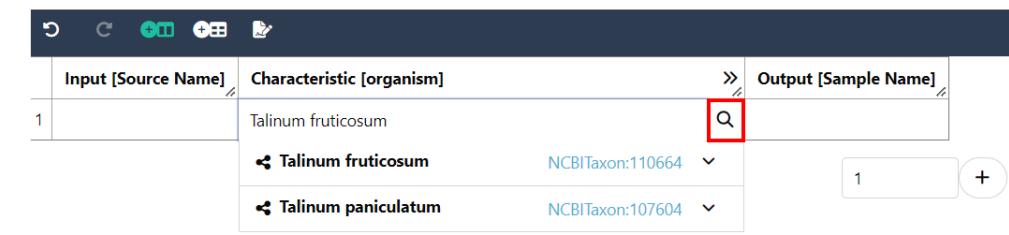
## Insert ontology terms

1. Insert values by selecting any cell below

Characteristic [organism]

2. Use free text or use the magnifying glass to activate *Term* search

3. Write "Talinum fruticosum" and enable *Term* search
4. Select the hit



## Add a Parameter building block with a unit

1. In the *Building Blocks* widget, select *Parameter*, search for `light intensity exposure` and select the term with id `PEC0:0007224`.
2. Check the box for *Unit* and search for `microeinsteins per square meter per second` in the adjacent search bar.
3. Select `U0:0000160`.
4. Click Add Column.

 This also adds four columns to your table, one visible and **three** hidden.

## Insert unit-values

In the annotation table, select any cell below **Parameter [light intensity exposure]** and add "425" as light intensity.

 You can see the numbers being complemented with the chosen unit, e.g. 425 microeinstein per square meter per second

# Your annotation table is growing

At this point. Your table should look similar to this:

	Input [Source Name]	Characteristic [organism] >>	Parameter [light intensity exposure] >>	Output [Sample Name]
1		Talinum fruticosum ✓	425 microeinsteins per square meter per second ✓	
2		Talinum fruticosum ✓	425 microeinsteins per square meter per second ✓	
3		Talinum fruticosum ✓	425 microeinsteins per square meter per second ✓	
4		Talinum fruticosum ✓	425 microeinsteins per square meter per second ✓	
5		Talinum fruticosum ✓	425 microeinsteins per square meter per second ✓	
6		Talinum fruticosum ✓	425 microeinsteins per square meter per second ✓	

1+

## Exercise

Try to add suitable *Building Blocks* for other pieces of metadata from the plant growth protocol ( `studies/talinum_drought/protocols/plant_material.txt` ).

## Add a factor building block

1. In the *Building Blocks* widget, select *Factor*, search for `watering exposure` and select the term with id `PEC0:0007383`.
  2. Click `Add Column`.
  3. Add the drought treatment ("no water for 12 days", "re-water for 2 days") to the respective samples
-  There are different options to add the drought treatment.

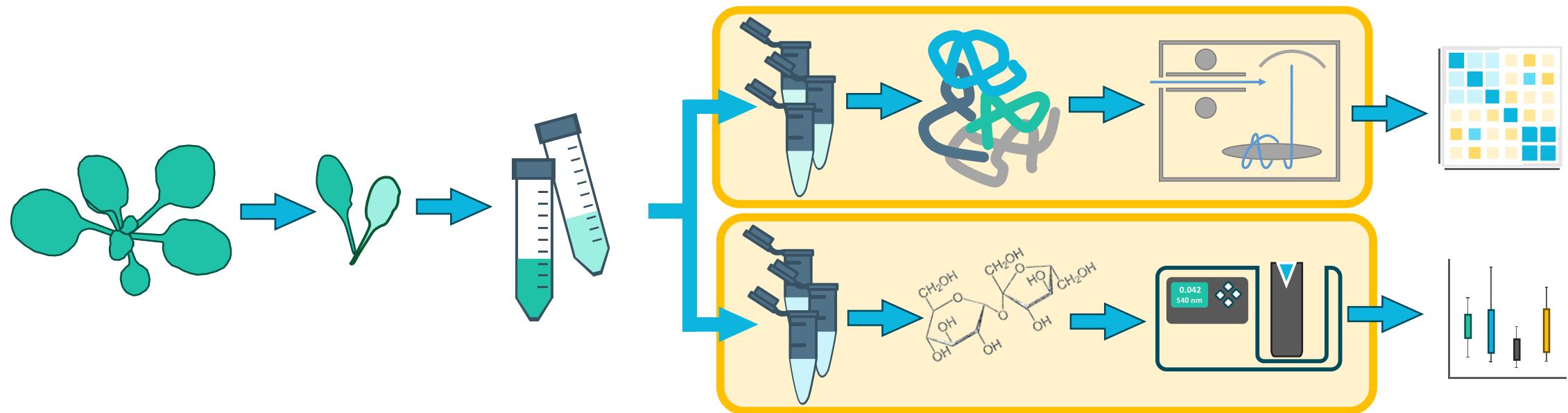
## Showing ontology reference columns

Use double pointed quotation mark to un-hide hidden columns.

- 💡 You can see that your organism of choice was added with id and source Ontology in the reference (hidden) columns.

Characteristic [organism]	Unit	TSR (OBI:0100026)	TAN (OBI:0100026)
Talinum fruticosum	-	NCBITaxon	NCBITaxon:110664

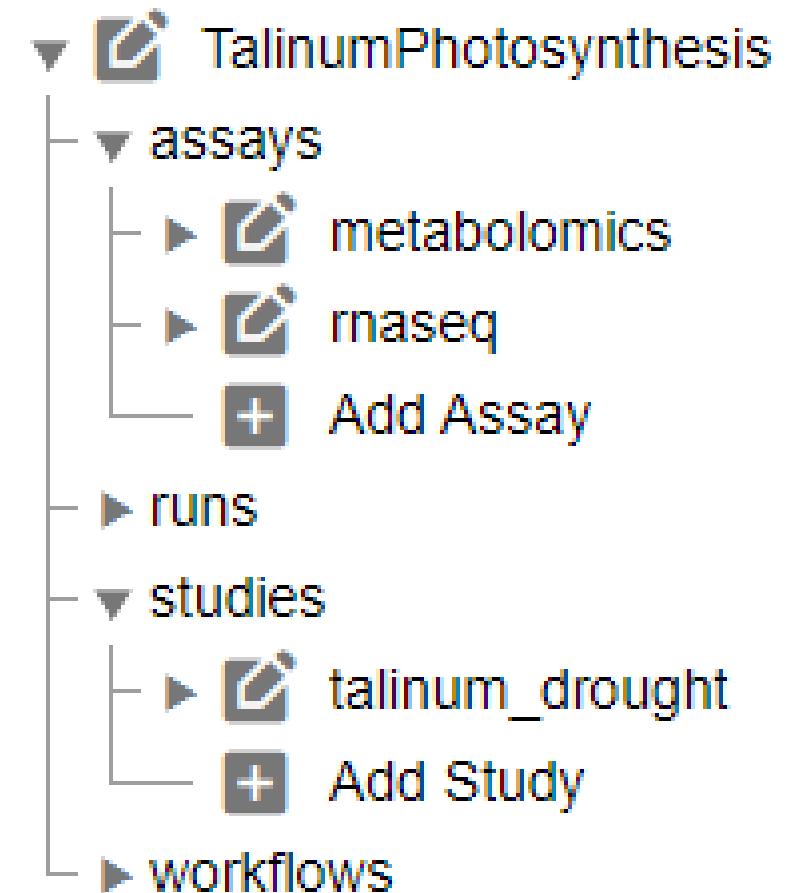
# Identifying assays



## Add an assay

by clicking "Add Assay" and entering an identifier for your assay

Add two assays with **rnaseq** and **metabolomics** as an identifier



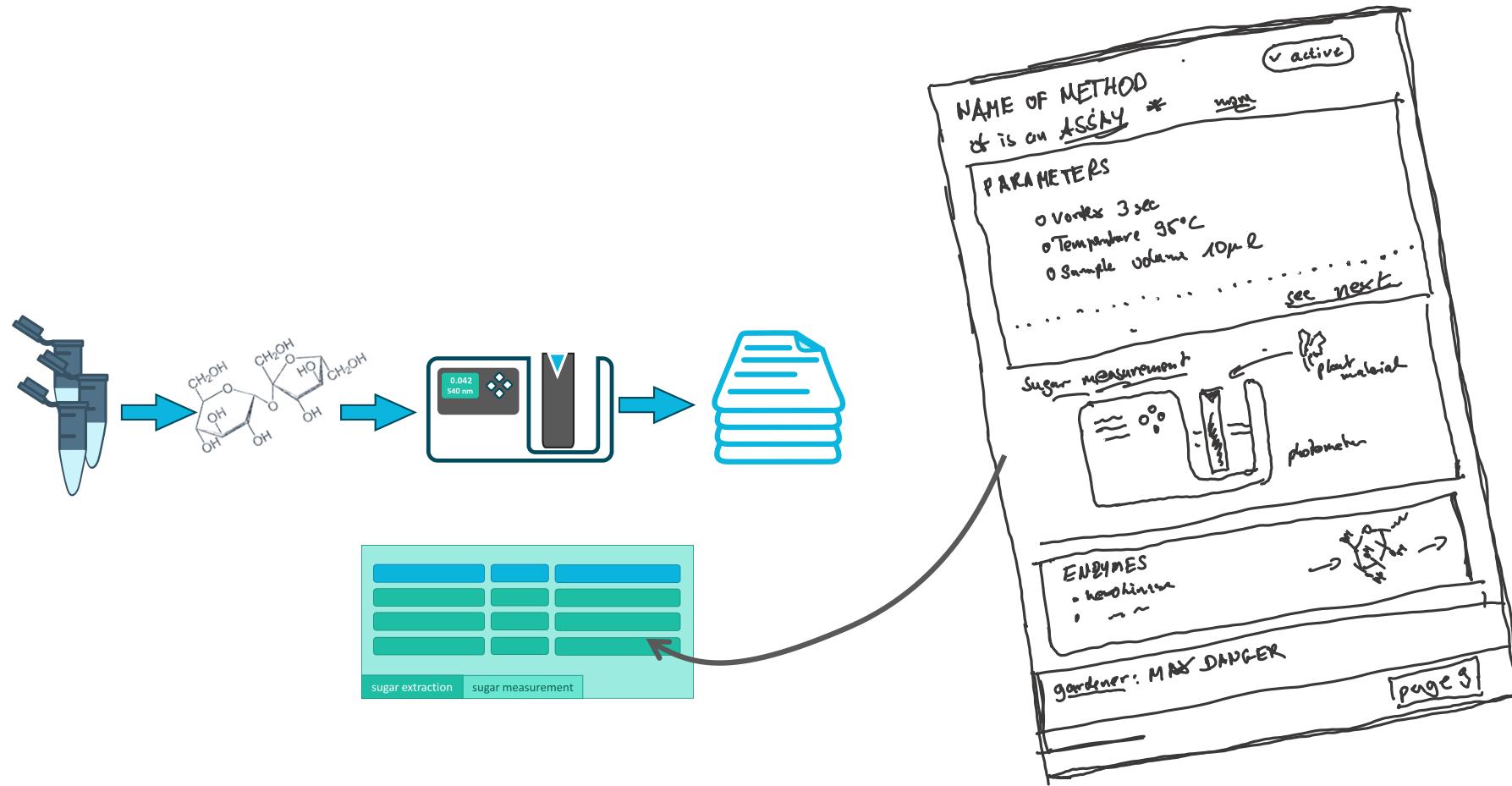
## Add information about your assay

In the assay panel you can define the assay's

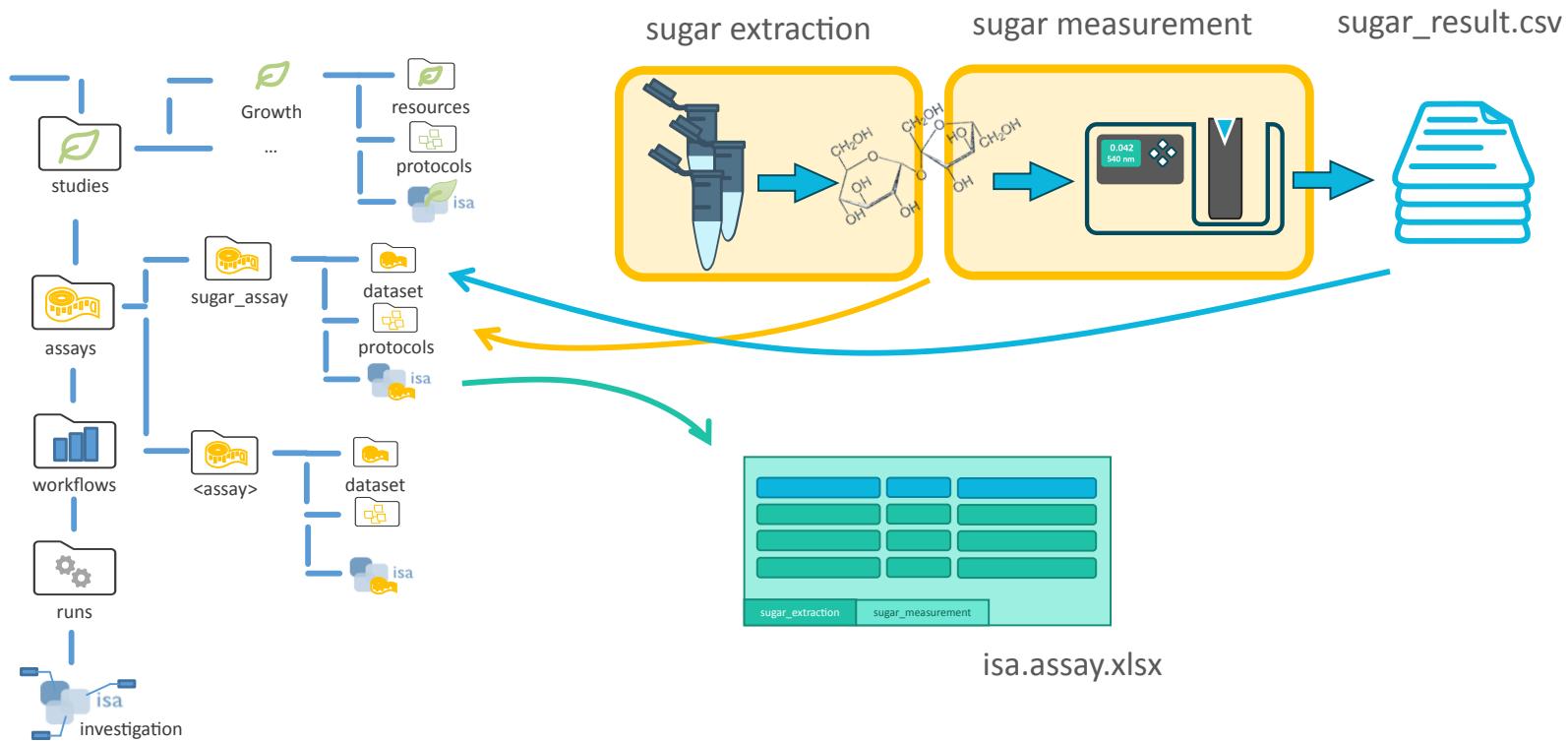
- measurement type
- technology type, and
- technology platform

<b>Identifier</b>		
rnaseq		
<b>Measurement Type</b>		
Term Name	TSR	TAN
<input type="text"/>	<input type="text"/>	<input type="text"/>
<b>Technology Type</b>		
Term Name	TSR	TAN
<input type="text"/>	<input type="text"/>	<input type="text"/>
<b>Technology Platform</b>		
Term Name	TSR	TAN
<input type="text"/>	<input type="text"/>	<input type="text"/>
<b>Performers</b>		
<input type="button" value="+"/>		
<b>Comments</b>		
<input type="button" value="+"/>		

# Assay for sugar measurement



# Separating different assay elements



# Parametric description of the lab process

## sugar extraction

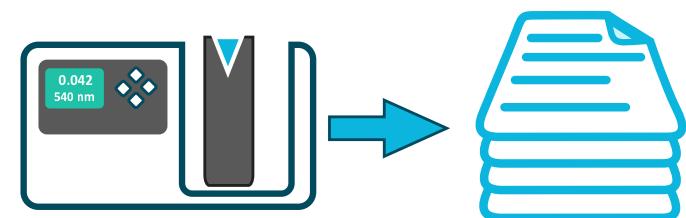
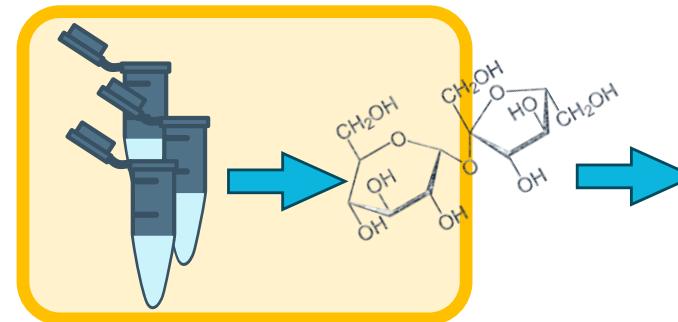
- Parameter []

- Temperature

- 95 °C

- Vortex Mixer

- 3 s



# Parametric description of the lab process

- Parameter []

- technical replicate

- 1-3

- sample volume

- 10 µl

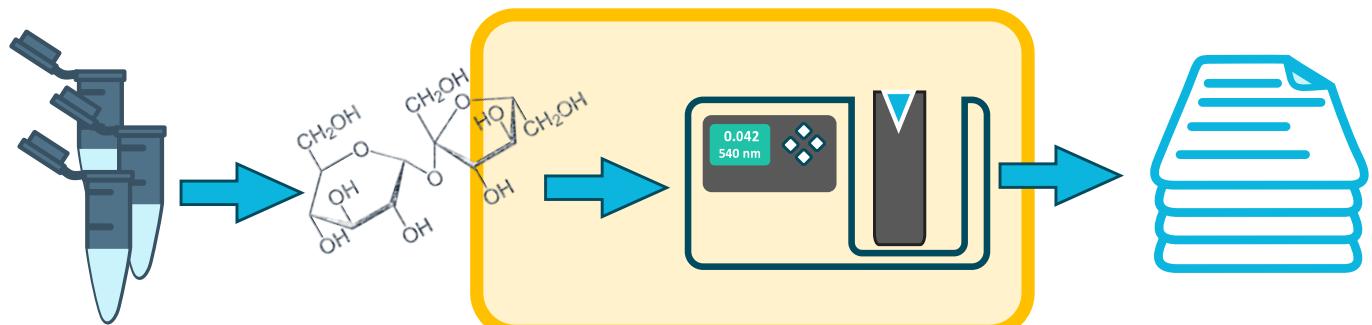
- Buffer

- 190 µl

- Cycle

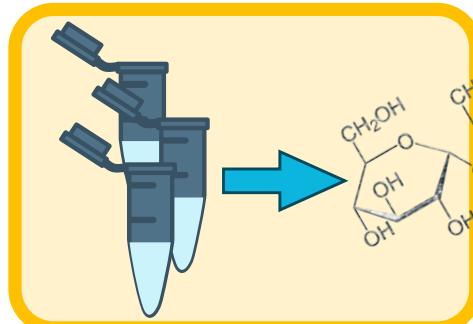
- 5

sugar measurement

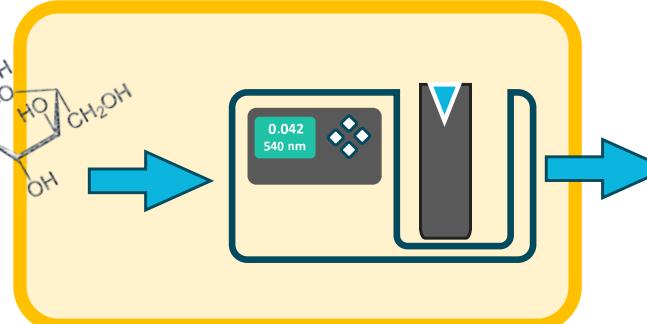


# Isolating the lab processes in an assay

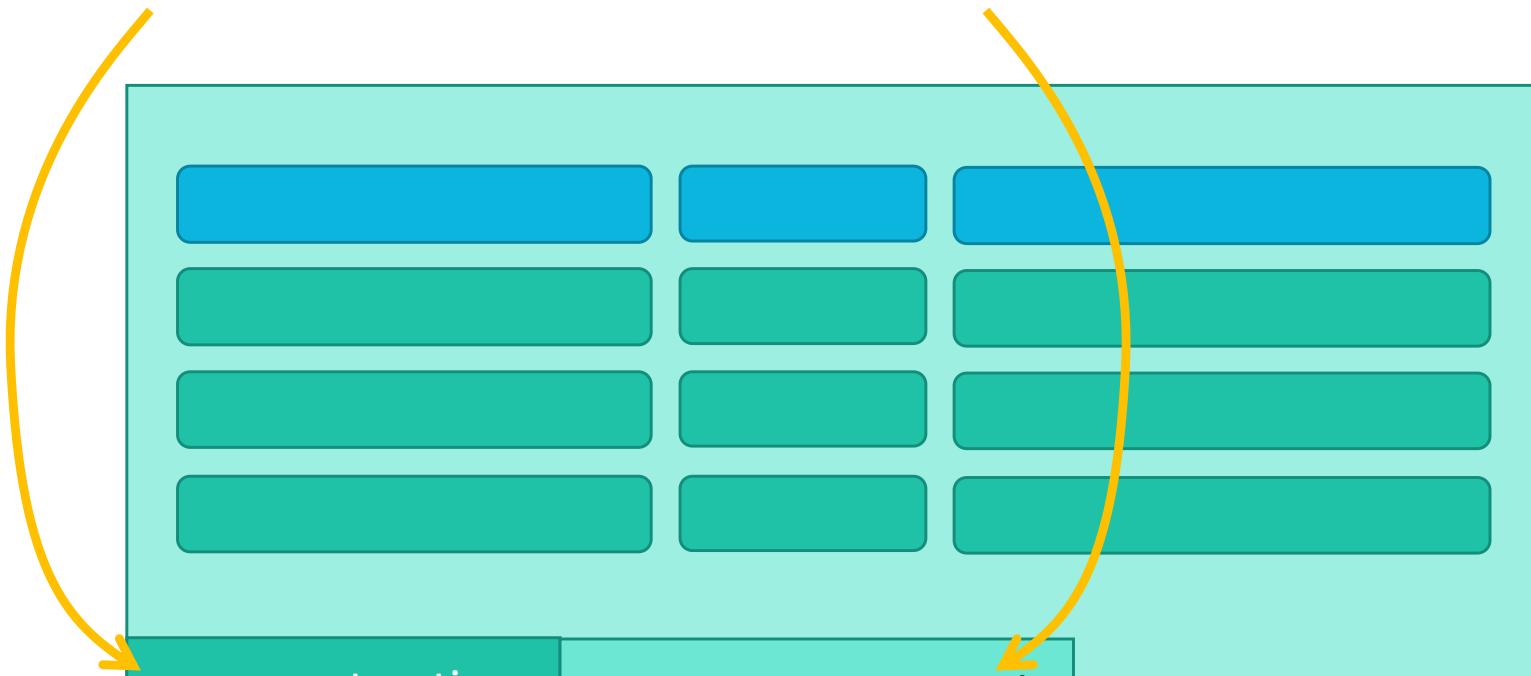
sugar extraction



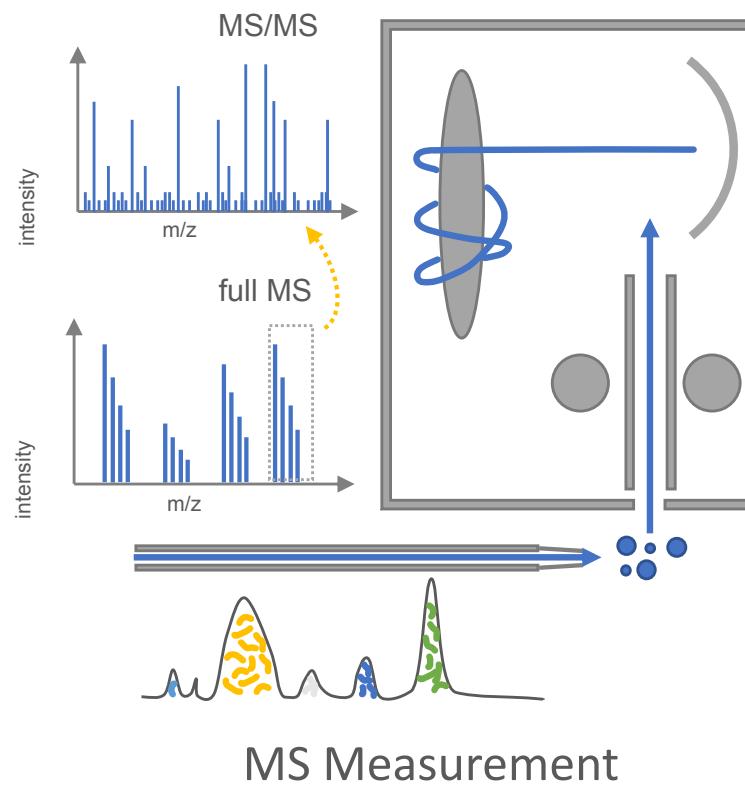
sugar measurement



sugar\_result.csv

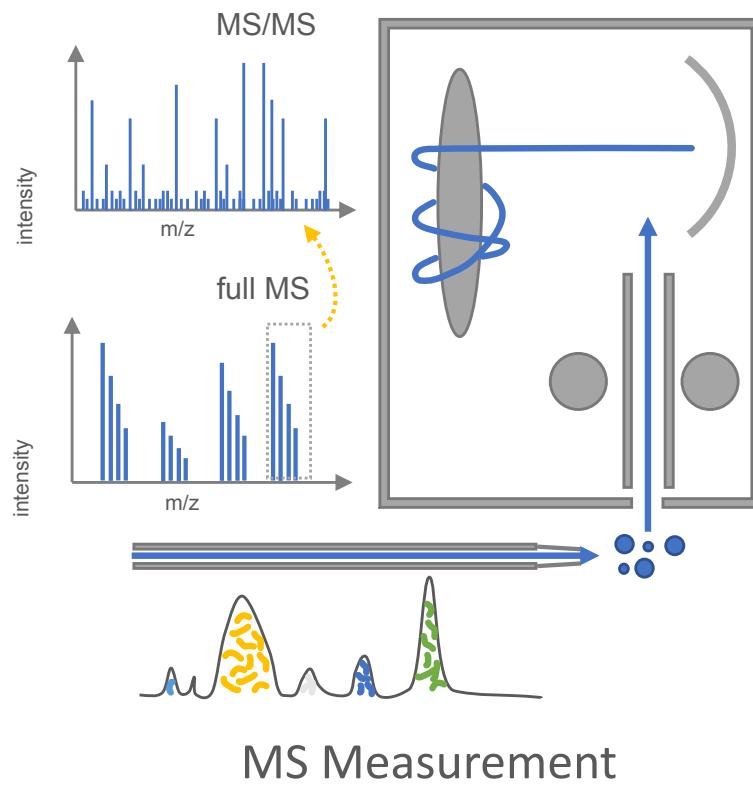


# Save time using standard methods and SOPs



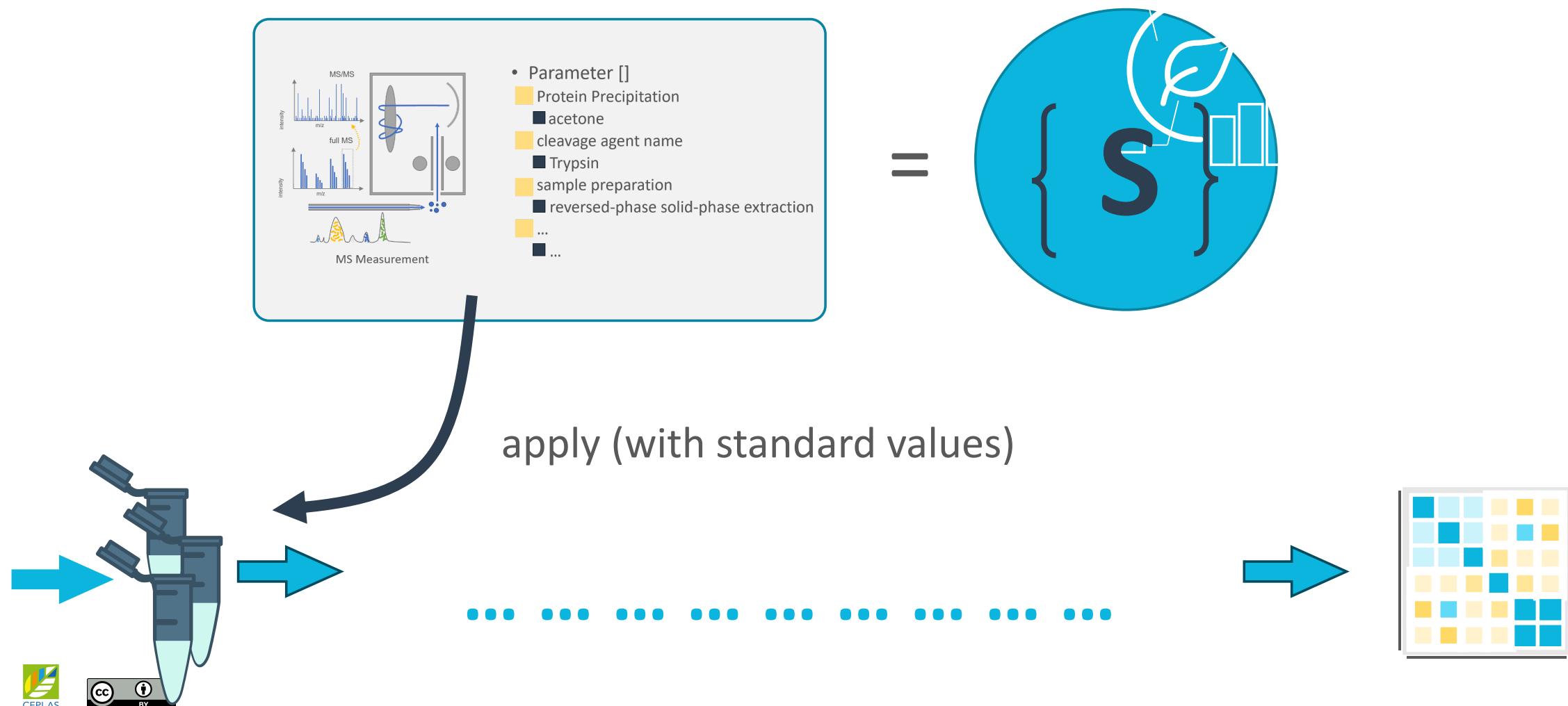
- Parameter []
  - Protein Precipitation
  - acetone
  - cleavage agent name
  - Trypsin
  - sample preparation
  - reversed-phase solid-phase extraction
  - ...
  - ...

# Save time using standard methods and SOPs

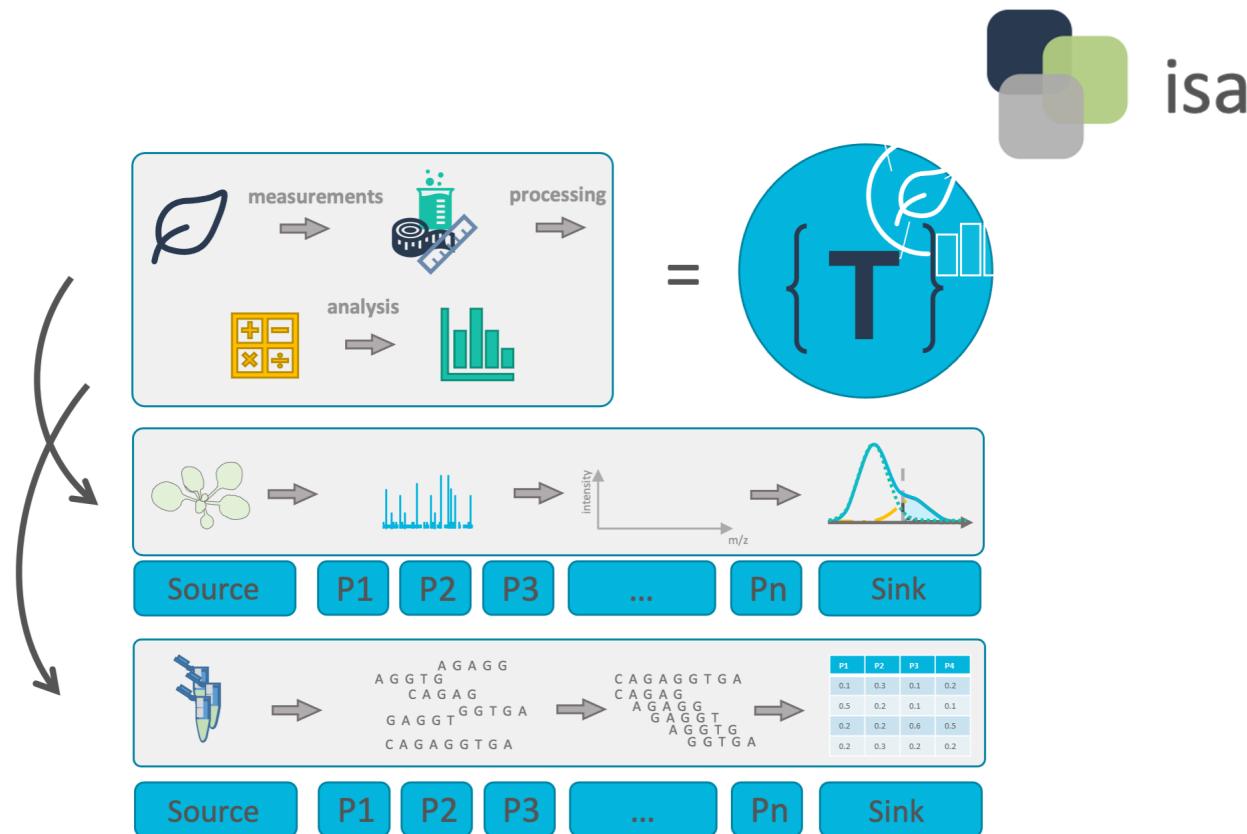


- Component []
  - chromatography instrument model
  - nanoElute2
  - chromatography column model
  - PepSep C18 1.9u, 25cm x 75um
  - ...
  - ...
  - ...
  - ...

# Applying standard procedures to sample record



# Realization of lab-specific metadata templates



Facilities can define their most common workflows as templates

# Import templates from a database

- DataPLANT curated
- Community templates

The screenshot shows a user interface for managing templates. At the top, there is a dark header bar with several small icons. Below it is a light gray search bar containing two input fields: "Search by template name" and "Search for tags". Underneath these are dropdown menus for "Select community" and "DataPLANT official". The main content area is a table with the following columns: "Template Name", "Community", and "Template Version". The table lists several templates, each with a "curated" status indicator and a dropdown arrow. The templates listed are:

Template Name	Community	Template Version
DNA extraction	curated	1.1.7
Data Processing (PRIDE minimal)	curated	1.0.0
GEO - Minimal information RNA assays	curated	1.0.1
GEO - Minimal information RNA extraction	curated	1.0.0
GEO - Minimal information computational analysis	curated	1.0.0
GEO - Minimal information plant growth	curated	1.0.0
Genome assembly	curated	1.1.7

## Let's annotate the RNA Seq assay

Open the lab notes `assays/rnaseq/protocols/`

# Use a template

## 1. Open the *Templates* widget in the Bar

💡 Here you can find DataPLANT and community created workflow annotation templates

## 2. Search for **RNA extraction** and click **select**

- You will see a preview of all *Building Blocks* which are part of this template.

## 3. Click **Add template** to add all *Building Blocks* from the template to your table

The screenshot shows the 'Templates' widget interface. At the top, there are search bars for 'Search by template name' and 'Search for tags', both with placeholder text ('.. template name' and '.. protocol tag'). Below these are dropdown menus for 'Select community' (set to 'DataPLANT official') and 'Template Version'. The main area is a table listing various templates:

Template Name	Community	Template Version	Actions
DNA extraction	curated	1.1.7	▼
Data Processing (PRIDE minimal)	curated	1.0.0	▼
GEO - Minimal information RNA assays	curated	1.0.1	▼
GEO - Minimal information RNA extraction	curated	1.0.0	▼
GEO - Minimal information computational analysis	curated	1.0.0	▼
GEO - Minimal information plant growth	curated	1.0.0	▼
Genome assembly	curated	1.1.7	▼
...			...

## Remove Building blocks

If there are any *Building Blocks* which do not fit to your experiment you can use right click --> "Delete Column" to remove it including all related (hidden) reference columns.

# Move Building blocks

If the order of the *Building Blocks* should be adjusted you can use right click --> "Move Column"

Move Column×

Preview1ApplyUpdate TableSubmit

Index	Column
0	Input [Source Name]
1	Characteristic [Organism]
2	Factor [watering exposure]
3	Output [Sample Name]

## Replace multiple names

Right click --> "Update Column" can be used to replace names in batches

 this only works on Input columns

Update Column ×

	Regex	Replacement
	DB	sample

Preview

	Before	After
0	DB_097	sample_097
1	DB_099	sample_099
2	DB_103	sample_103
3	DB_161	sample_161
4	DB_163	sample_163

Submit

## New process, new worksheet

1. Add a new sheet to the `assays/rnaseq/isa.assay.xlsx` workbook.
2. Add the template "RNASeq Assay"

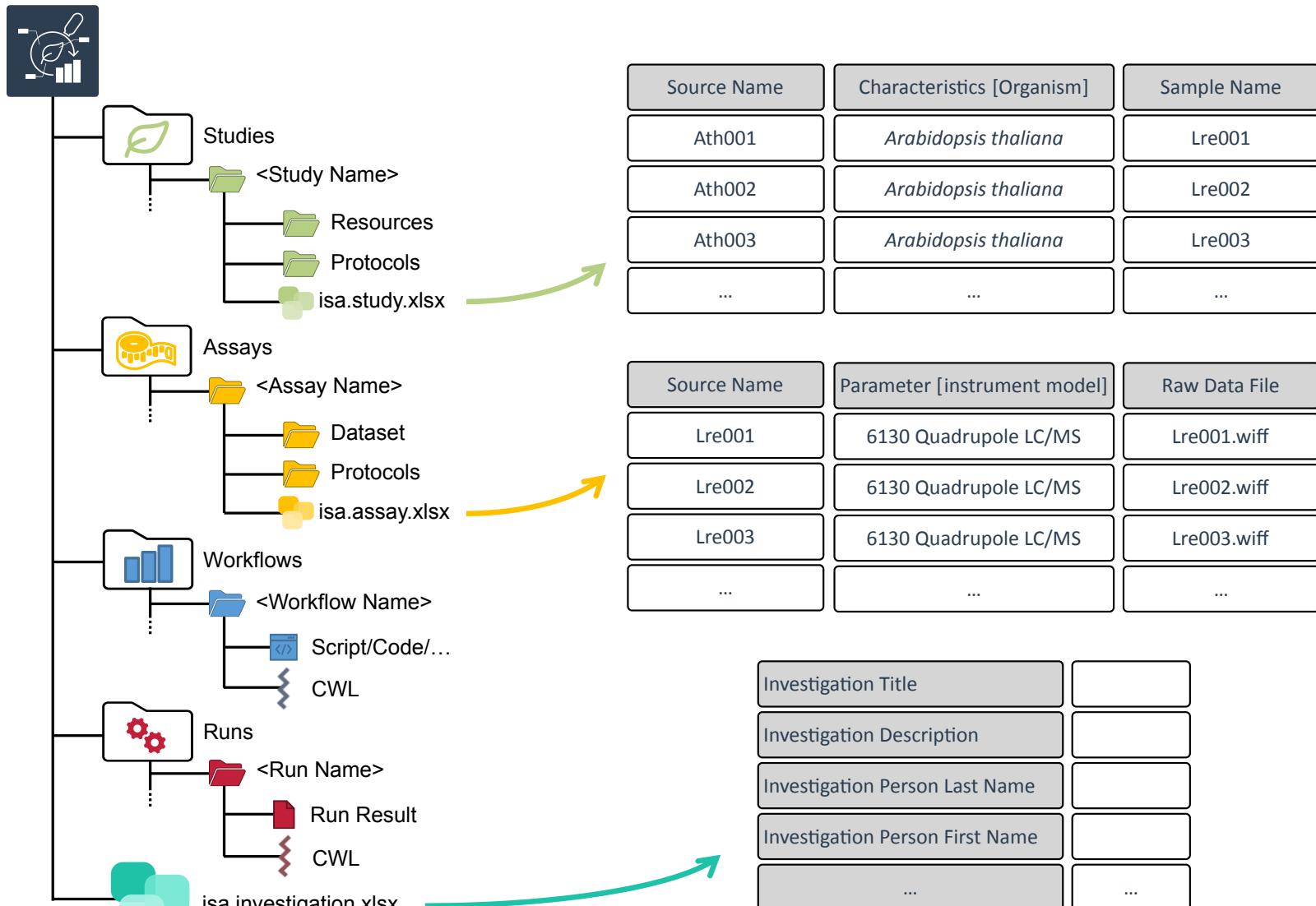
## Exercise

Try to fill the two sheets with the protocol details:

- assays/rnaseq/protocols/RNA\_extraction.txt and
- assays/rnaseq/protocols/Illumina\_libraries.txt

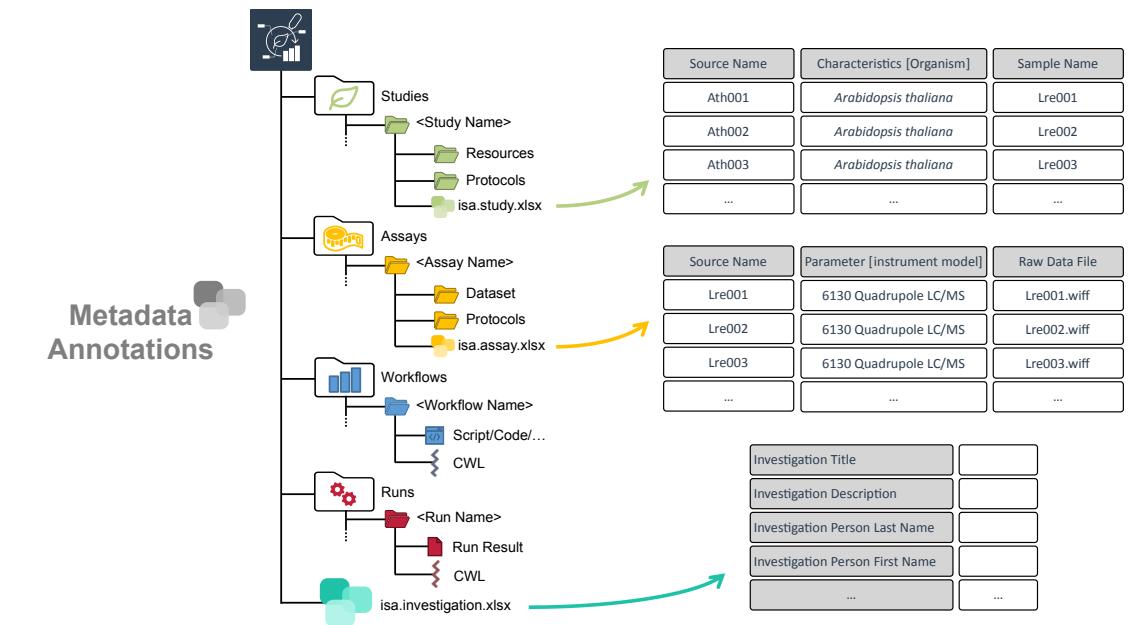
# ARC builds on ISA to connect data

Metadata Annotations



# ARC builds on ISA to link data

- Samples are linked study-to-assay, assay-to-assay
- Raw data is linked to assays
- Protocols can be referenced
- ...



## Link samples across studies and assays

1. Use the **Output [Sample Name]** of studies/talinum\_drought/isa.study.xlsx as the **Input [Sample Name]** to **rna-extraction**.
2. Use the **Output [Sample Name]** of **rna-extraction** as the **Input [Sample Name]** to **illumina-libraries**.

**Seeds –Plant growth→ Leaves –RNA Extraction→ RNA –Illumina→ fastq files**

# Link dataset files to samples

1. In the *Building Blocks* widget select *Output -> Raw Data File*.

2. Click **Add Column**.

 You see a warning about a changed output column.

3. Click **Continue**.

4. Go to the *File Picker* tab and click **Pick file names**.

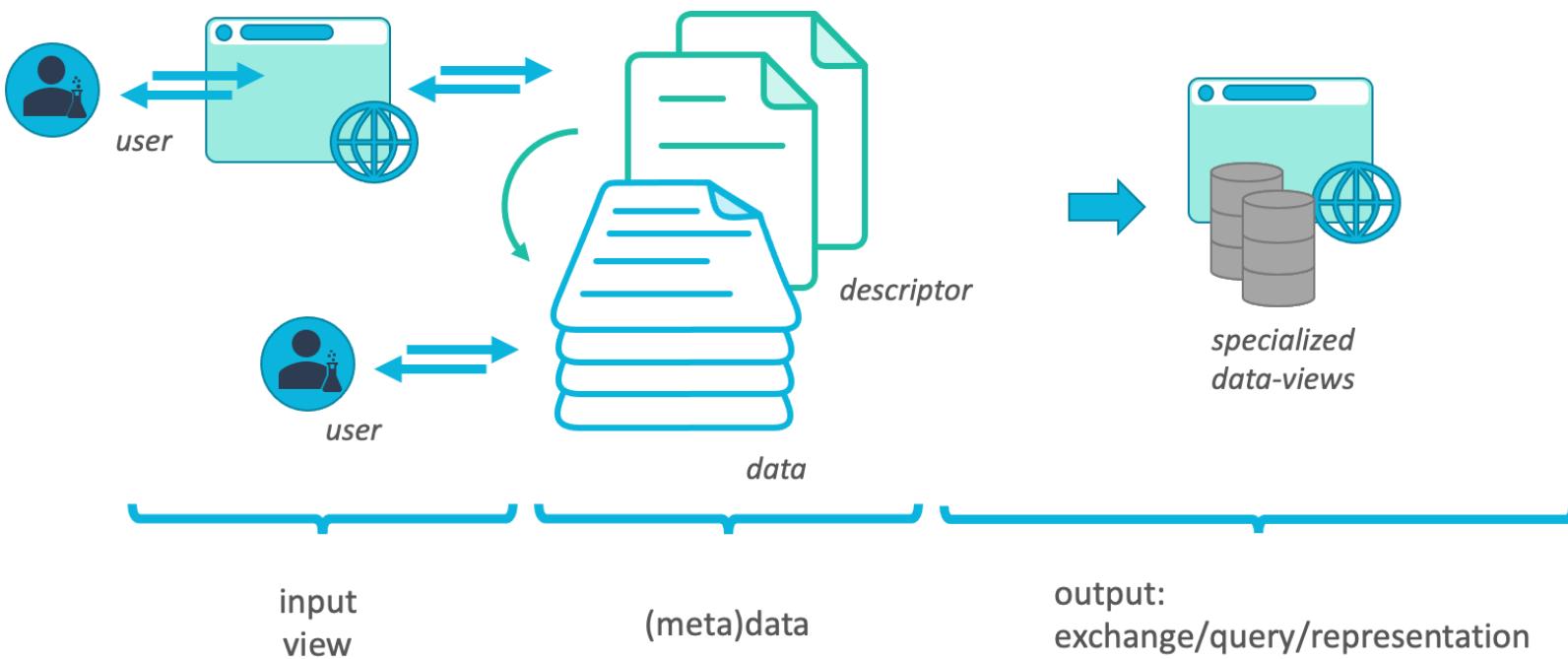
5. Select and open the \*fastq.gz files from the dataset folder.

6. Copy / paste them to the **Raw Data File**.

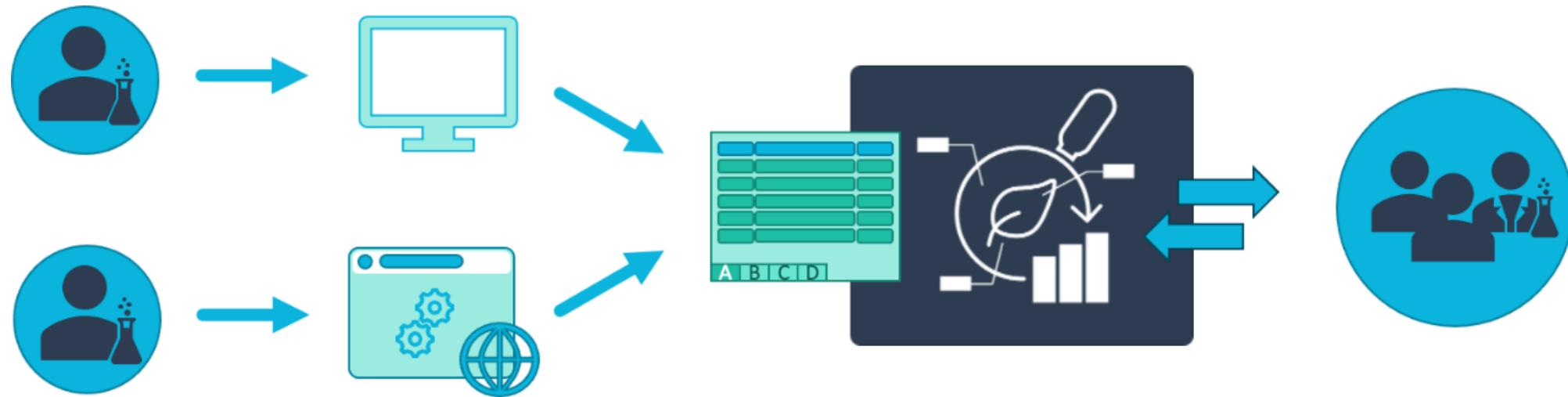
 This allows you to link your samples to the resulting raw data files.

# Everything is a file

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



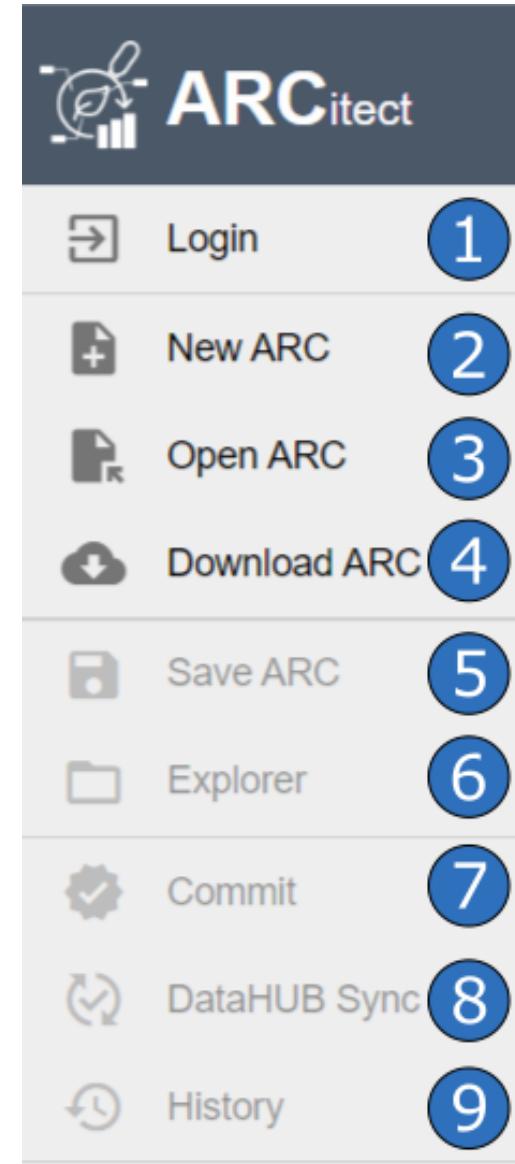
# No technical lock-in



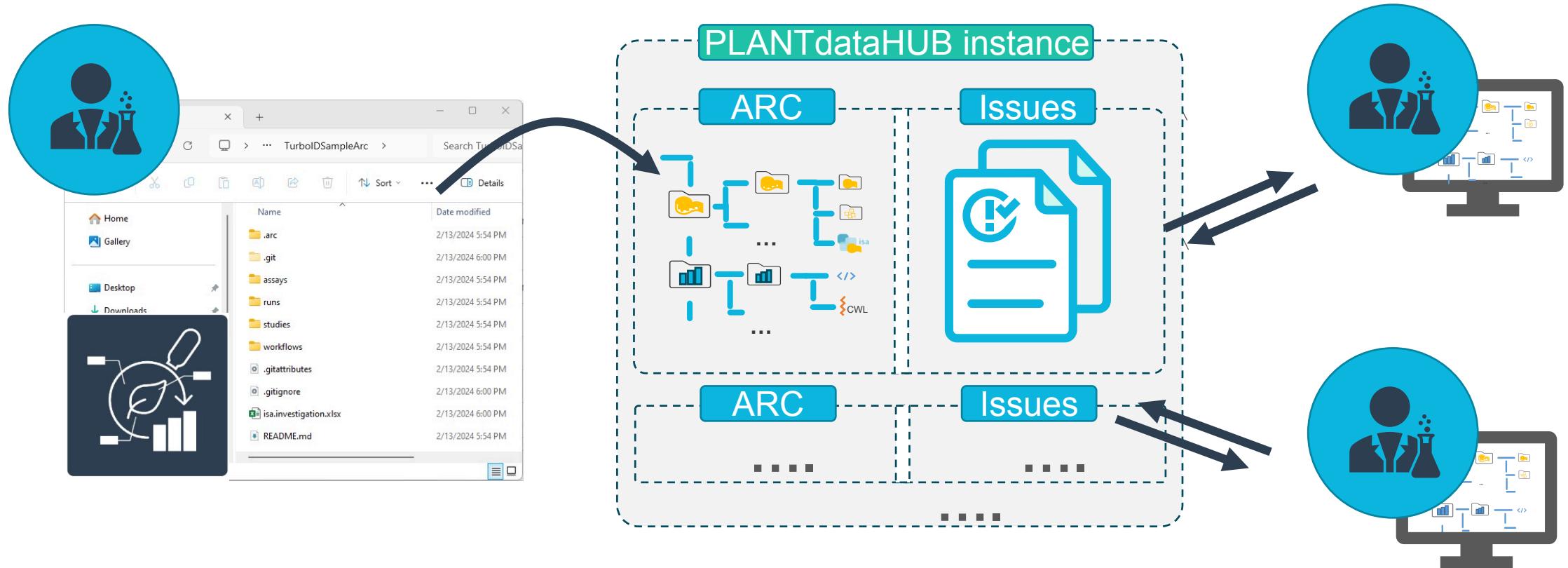
(Meta)data transparency with tool assistance but **no technical lock-in**

## Explorer

The **Explorer (6)** button directly opens your ARC locally



# Using the DataHUB to collaborate



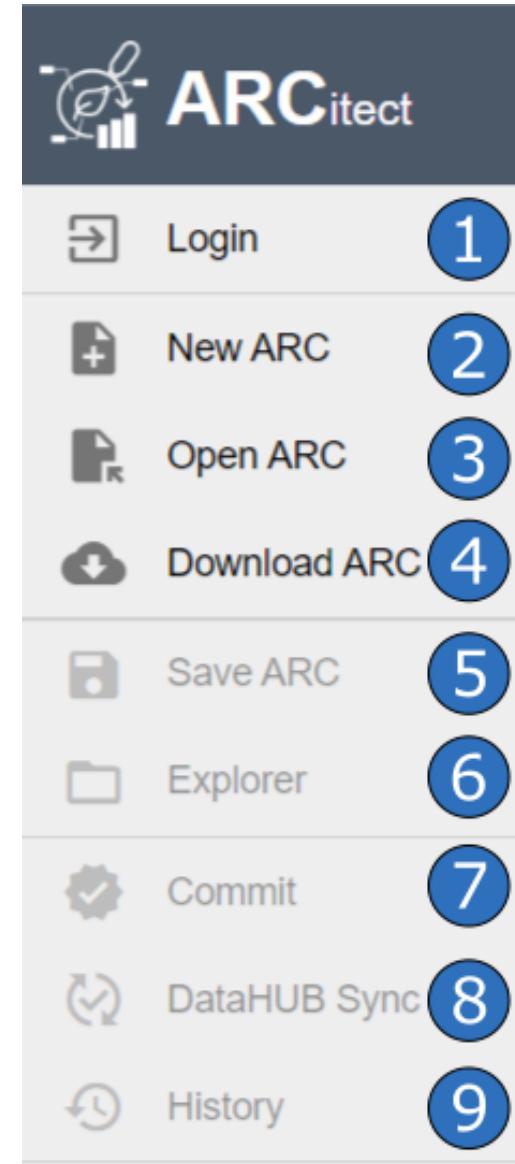
# Login to DataHUB

1. Login to DataHUB (1)
2. Select `git.nfdi4plants.org` as Host

Please Select a DataHub

Host  
`git.nfdi4plants.org`

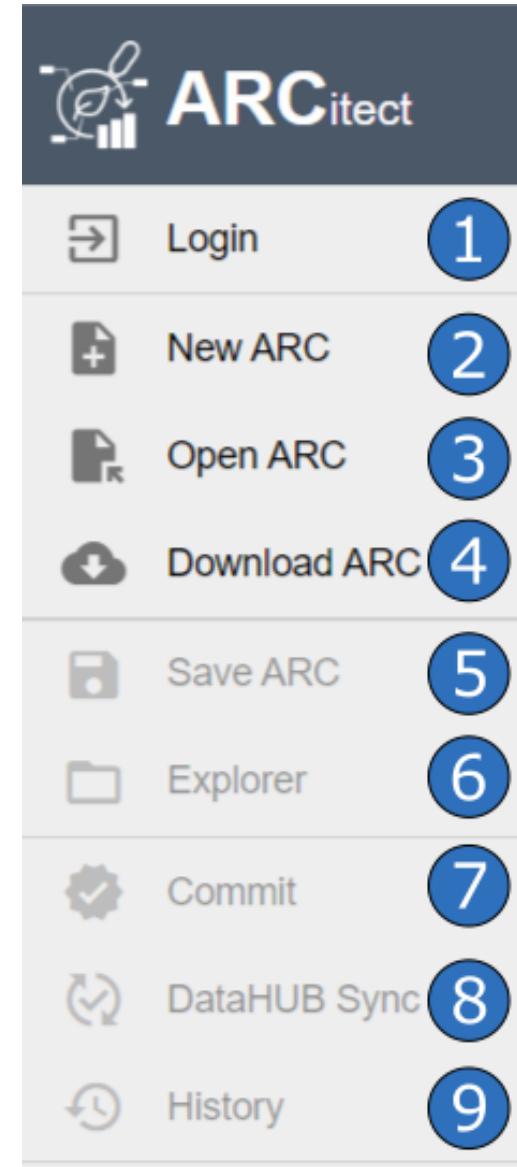
LOGIN CANCEL





## Commit panel (7)

You have to commit changes before you can upload to the DataHUB



# Commit your changes

If you are logged in, the **Commit panel** shows

- your DataHUB's *Full Name* and *eMail*

It allows you to

- track changes of the ARC with git
- add a commit message
- use different branches

The screenshot shows the 'Commit Changes' panel. At the top, it says 'Commit Changes' and 'Track changes of the ARC with git'. Below that, there are fields for 'Name' (Sabrina Zander) and 'eMail' (sabrina.zander@uni-duesseldorf.de). There are dropdowns for 'Branch' and 'Commit Message'. A field for 'Large File Storage Limit in MB' is set to 1. Below these, a section titled 'Changes' shows a message 'No changes to commit'. At the bottom are 'RESET' and 'COMMIT' buttons.

## History panel (9)

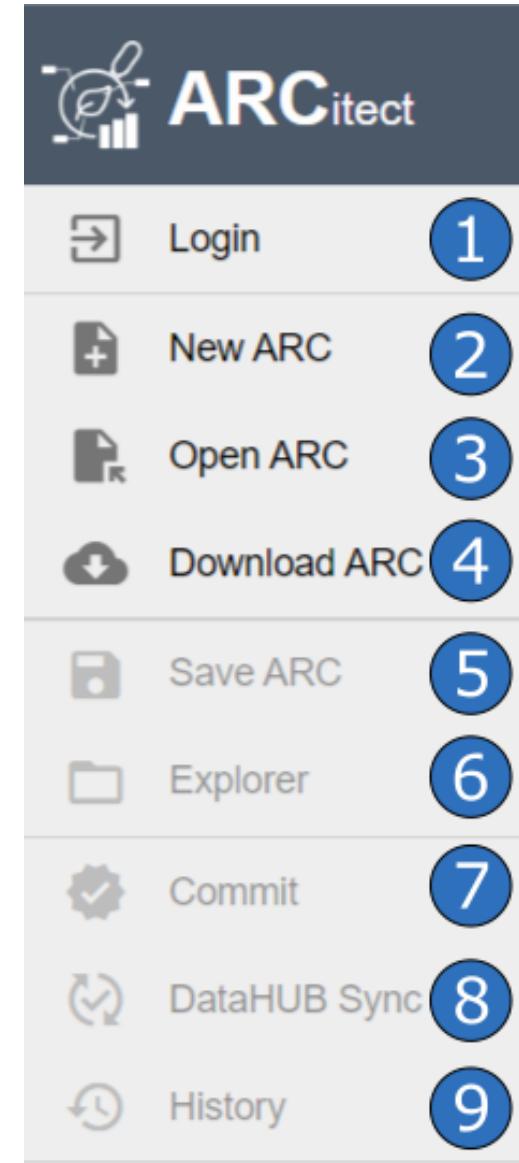
In the History panel you can inspect your ARCs history with all commits

The screenshot shows a 'History' panel with a title 'Inspect ARC history'. Below the title is a circular icon with a clock and a downward arrow. The main area displays a vertical timeline of commits, each marked with a teal circle. The commits are listed from top to bottom:

- 08.04.2024 12:44 - SABRINA ZANDER (SABRINA.ZANDER@UNI-DUESSELDORF.DE)  
add data and protocols to rnaseq
- 08.04.2024 12:43 - SABRINA ZANDER (SABRINA.ZANDER@UNI-DUESSELDORF.DE)  
add assay rnaseq
- 08.04.2024 12:43 - SABRINA ZANDER (SABRINA.ZANDER@UNI-DUESSELDORF.DE)  
add study talinum\_drought
- 08.04.2024 12:43 - SABRINA ZANDER (SABRINA.ZANDER@UNI-DUESSELDORF.DE)  
add contributors
- 08.04.2024 12:42 - SABRINA ZANDER (SABRINA.ZANDER@UNI-DUESSELDORF.DE)  
add description to investigation
- 08.04.2024 12:41 - SABRINA ZANDER (SABRINA.ZANDER@UNI-DUESSELDORF.DE)  
set up new ARC

## Upload your local ARC to the DataHUB

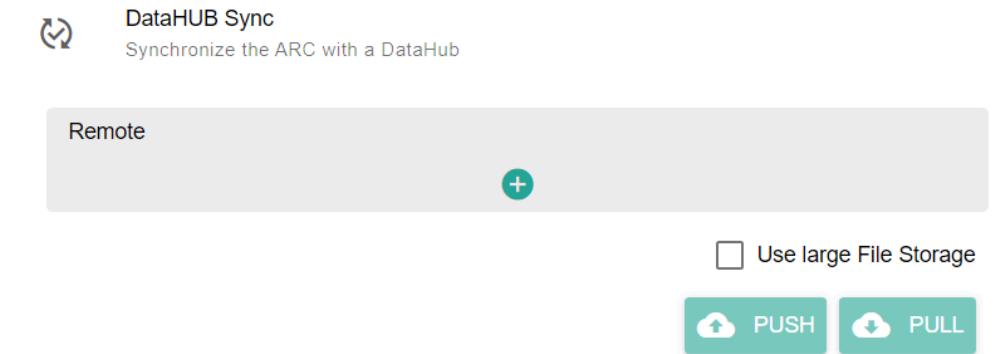
From the sidebar, navigate to **DataHUB Sync** (8)



# DataHUB Sync

The DataHUB Sync panel allows you to

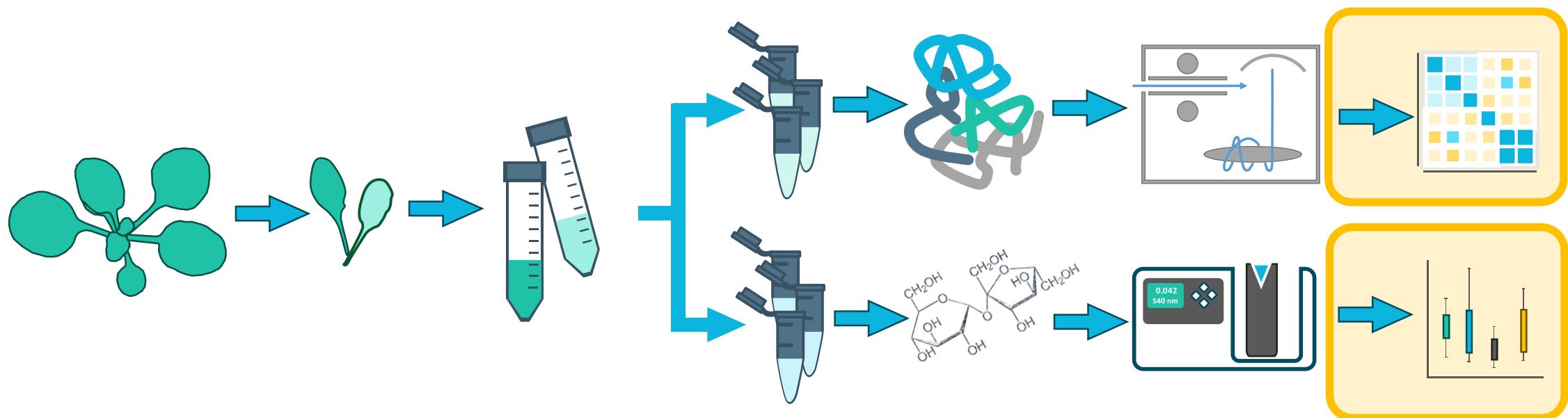
- sync the changes to the DataHUB: **Push**
- sync from the DataHUB: **Pull**, and
- change the Remote for the synchronization



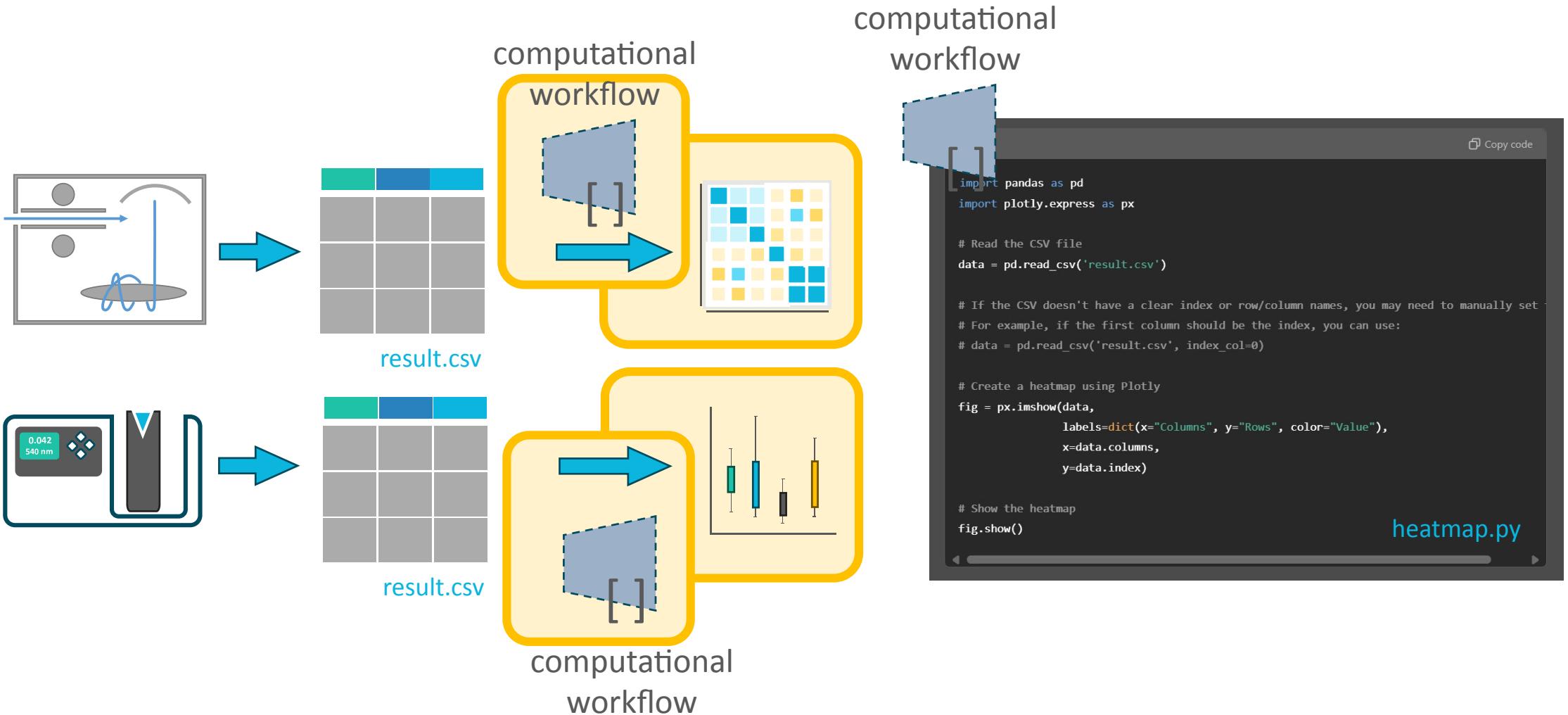
## Check if your ARC is successfully uploaded

1. [sign in](#) to the DataHUB
2. Check your projects

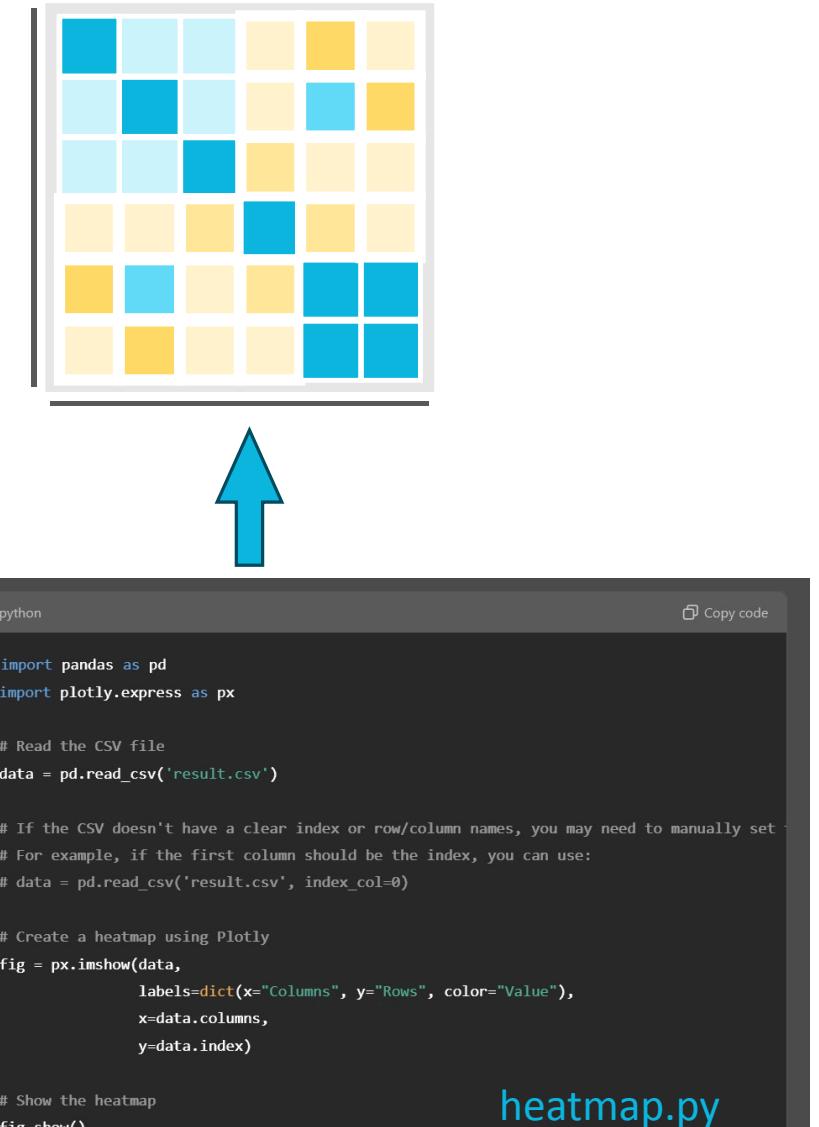
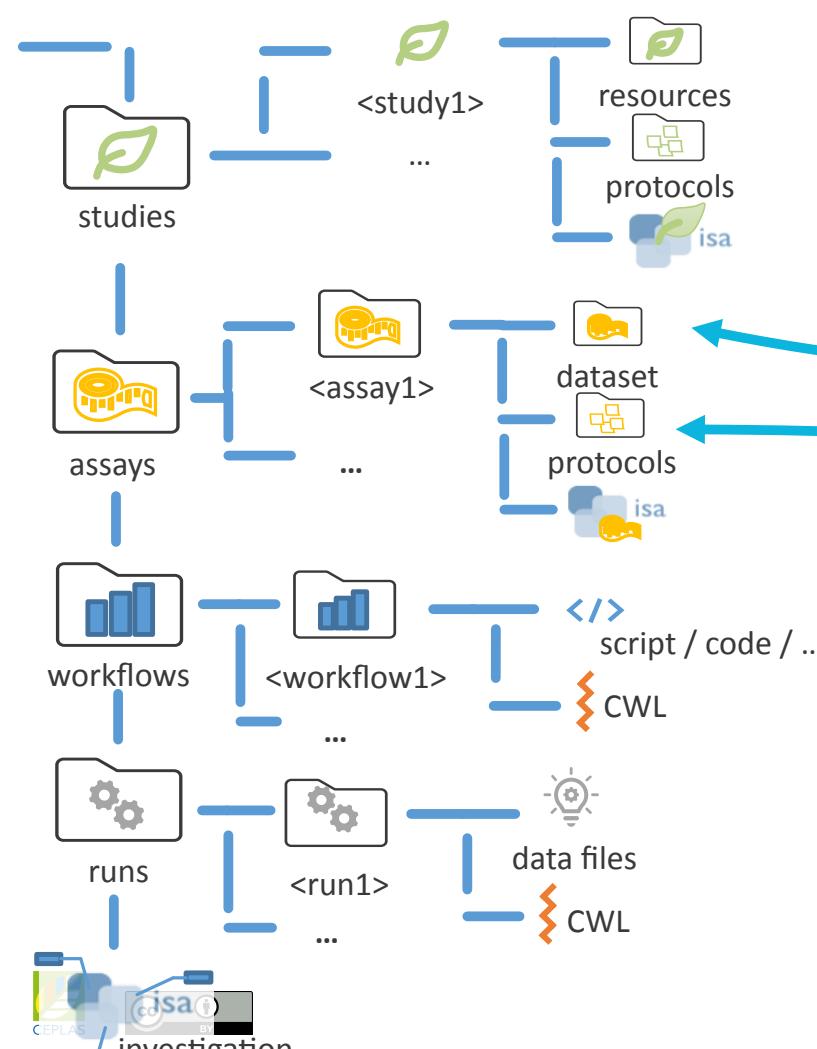
## Options to annotate the data analysis



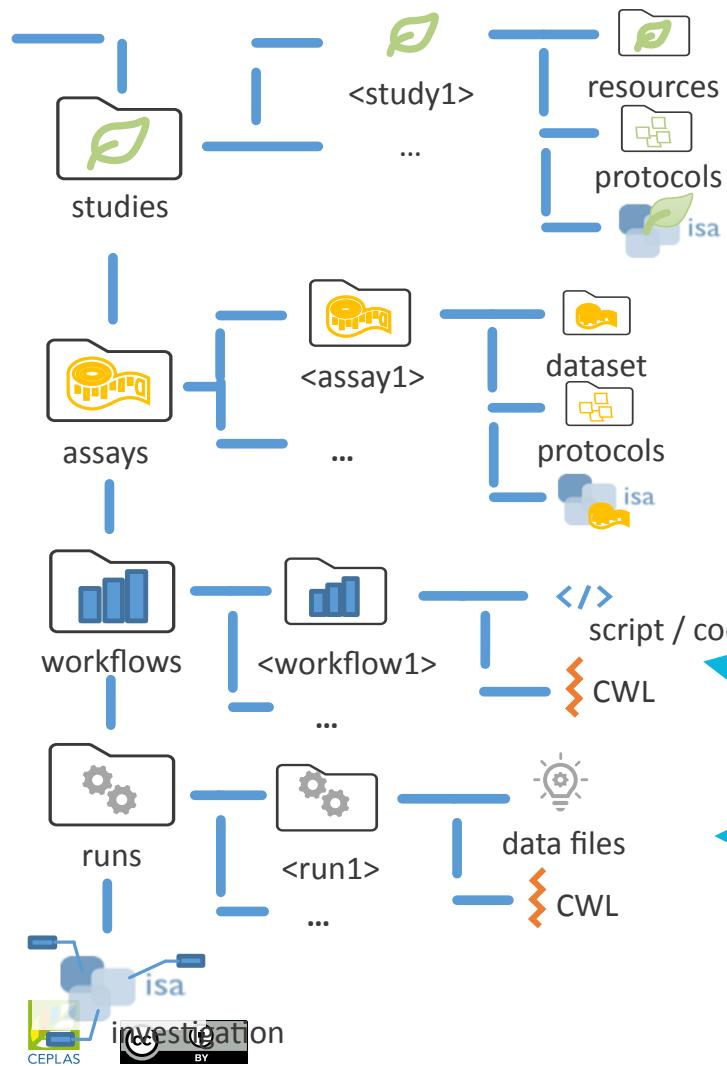
# A computational workflow is like a protocol



# Option I: Create a virtual assay



# Option II: Create a workflow and run



```
python
Copy code

import pandas as pd
import plotly.express as px

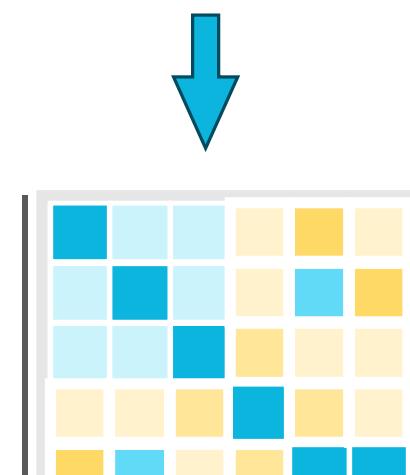
# Read the CSV file
data = pd.read_csv('result.csv')

# 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



# Use CWL to wrap your workflow

CWL workflow document (\*.cwl)

```
graph LR; A[1] --> B(( )); B --> C[3];
```

```
#!/usr/bin/env cwl-runner

cwlVersion: v1.0
class: Workflow
inputs:
  tarball: File
  name_of_file_to_extract: string

outputs:
  compiled_class:
    type: File
    outputSource: compile/classfile

steps:
  untar:
    run: tar-param.cwl
    in:
      tarfile: tarball
      extractfile: name_of_file_to_extract
    out: [extracted_file]

  compile:
    run: arguments.cwl
    in:
      src: untar/extracted_file
    out: [classfile]
```

CWL  
tool descriptors (\*.cwl)

CWL job parameter (\*.yaml)

```
job
yaml

file: fastq
param: 5
workflow: wf.cwl
output_folder:
  /temp
```

CWL runner

output  
folder

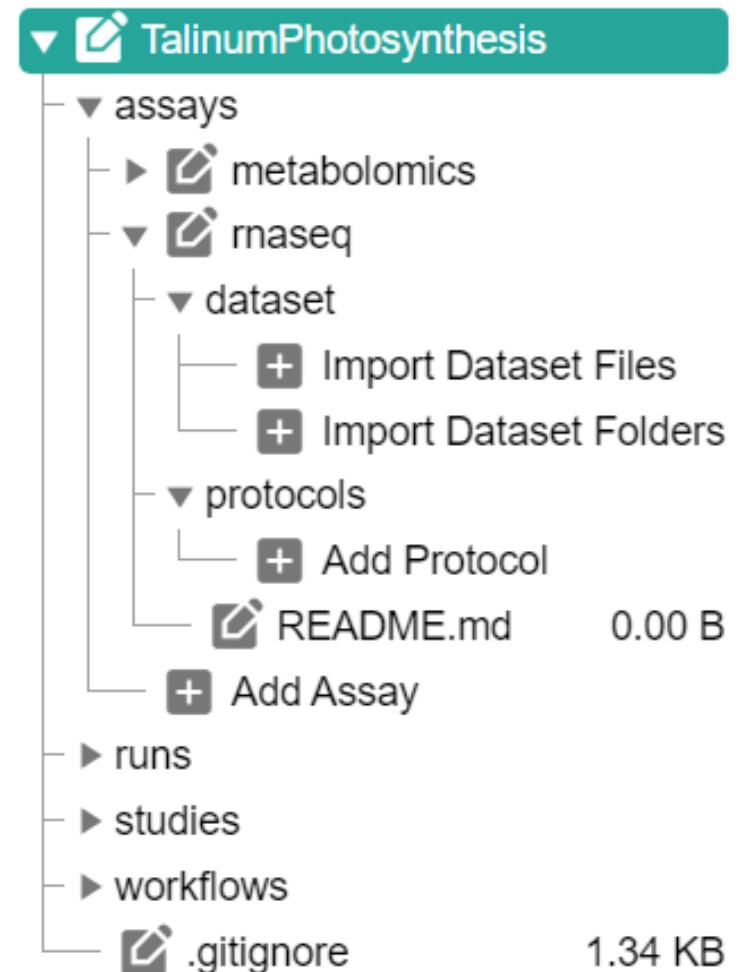
## Add datasets

In the file tree you can

- **import dataset files or folders** and
- **protocols** associated to that dataset.

 **Import Dataset** allows to import data from any location on your computer into the ARC.

 Depending on the file size, this may take a while. Test this with a small batch of files first.

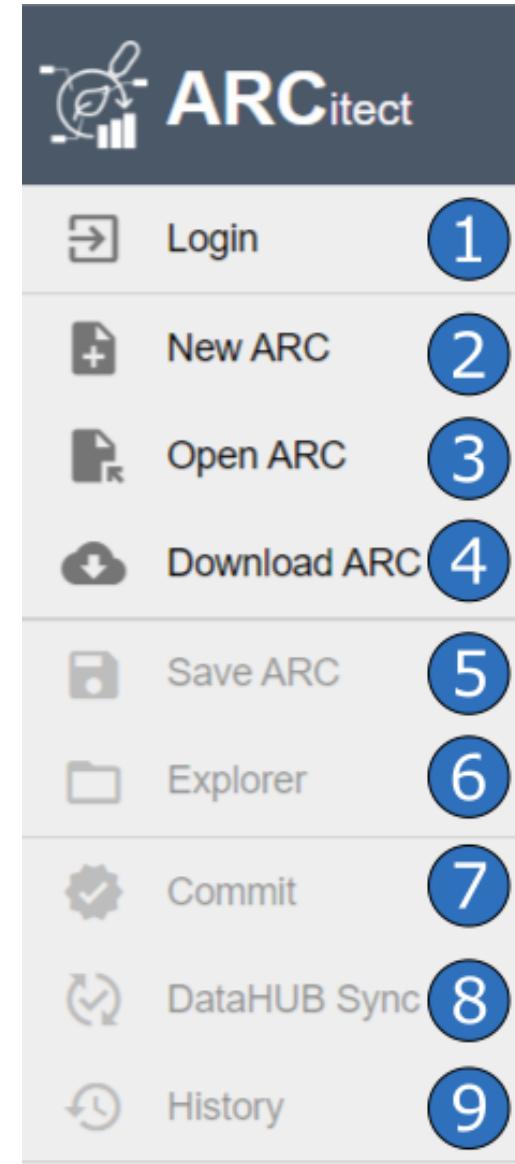


# Backup Slides



## Download the demo data

1. Open the ARCitect
2. Login (1) to your DataHUB account
3. Navigate to Download ARC (4)

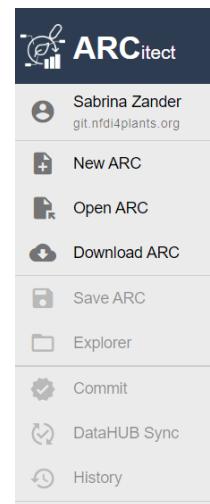


# Download the demo data

4. Search for **Talinum-CAM-Photosynthesis**

5. Click the download button, select a location and open the ARC.

6. Open the downloaded ARC



Open ARC

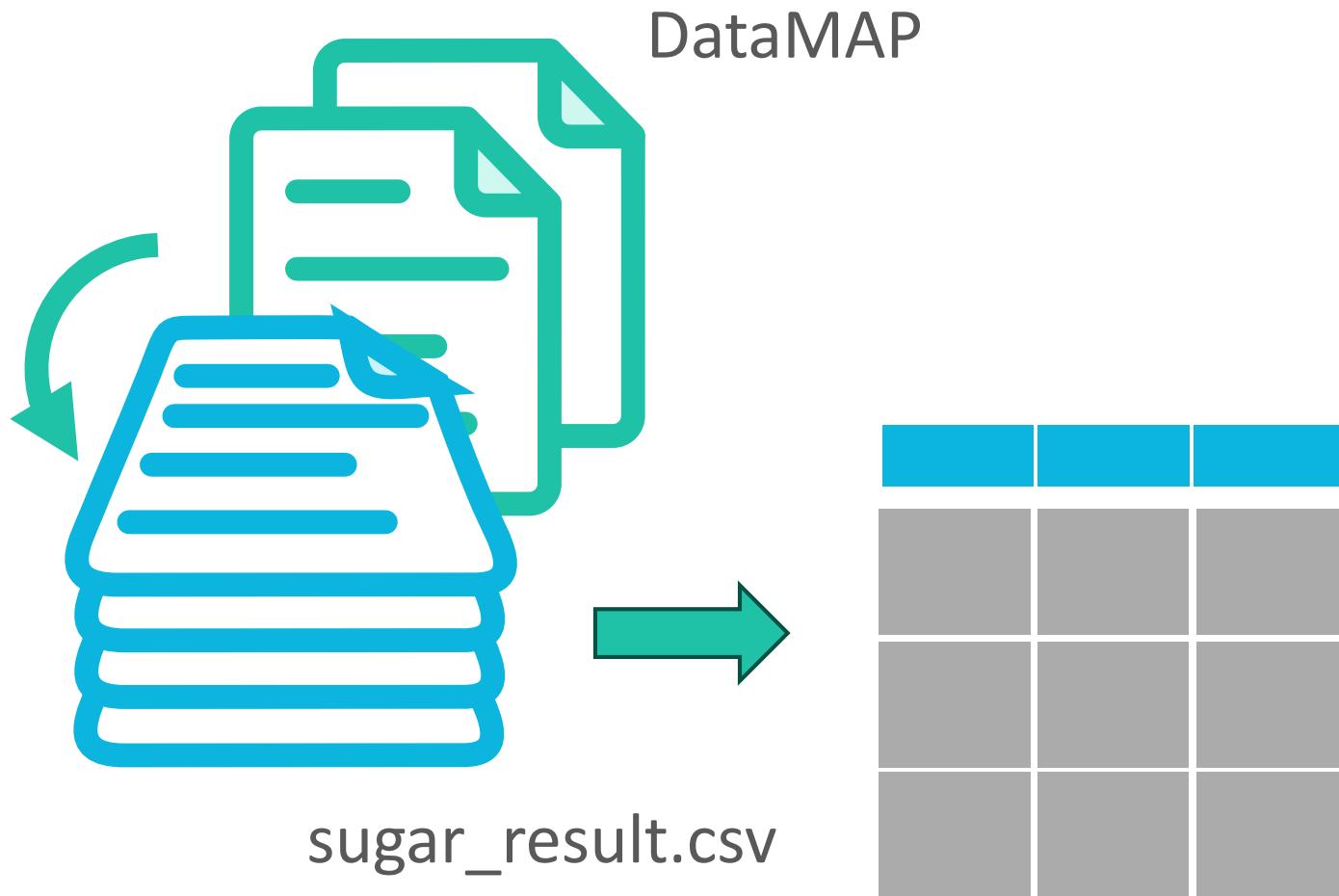
A screenshot of the 'Download ARC' page from the DataHUB. The search bar at the top contains the text 'Talinum-CAM-Photosynthesis'. Below the search bar, there are checkboxes for 'LFS' and 'Host git.nfdi4plants.org'. A green button labeled 'C' is visible. The results section shows a card for 'Talinum-CAM-Photosynthesis' with the identifier '[2023-10-11T09:24:10.208Z]' and the category 'Teaching'. There are search and download buttons next to the card.

 This is basically the ARC we created in the last session.

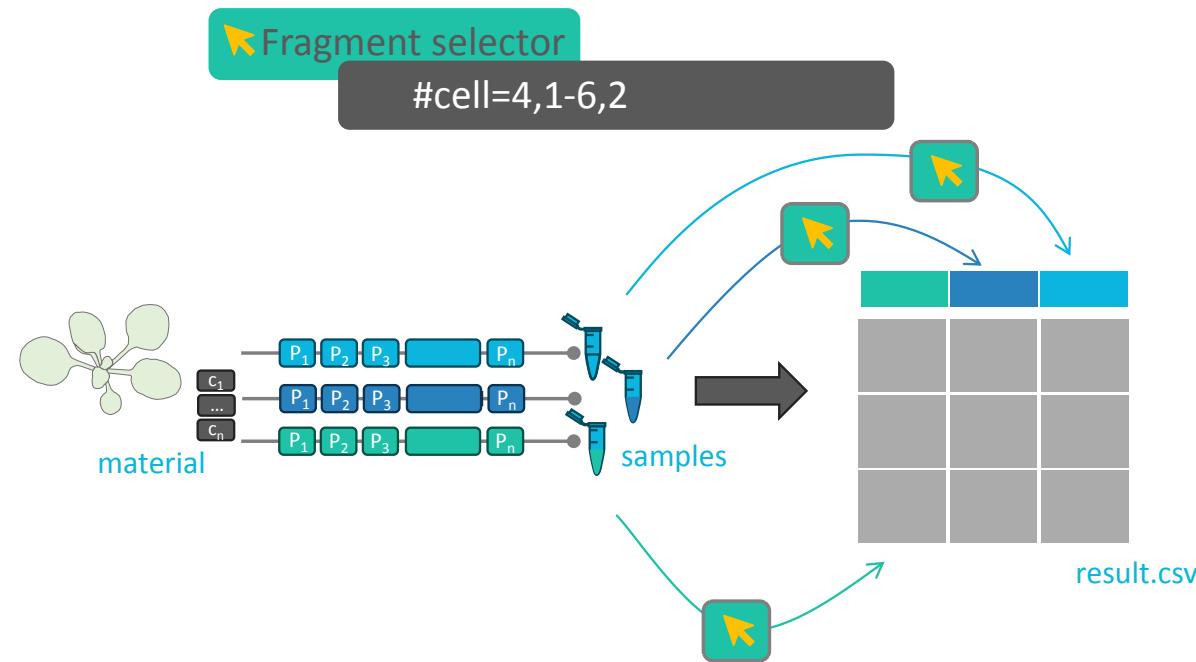




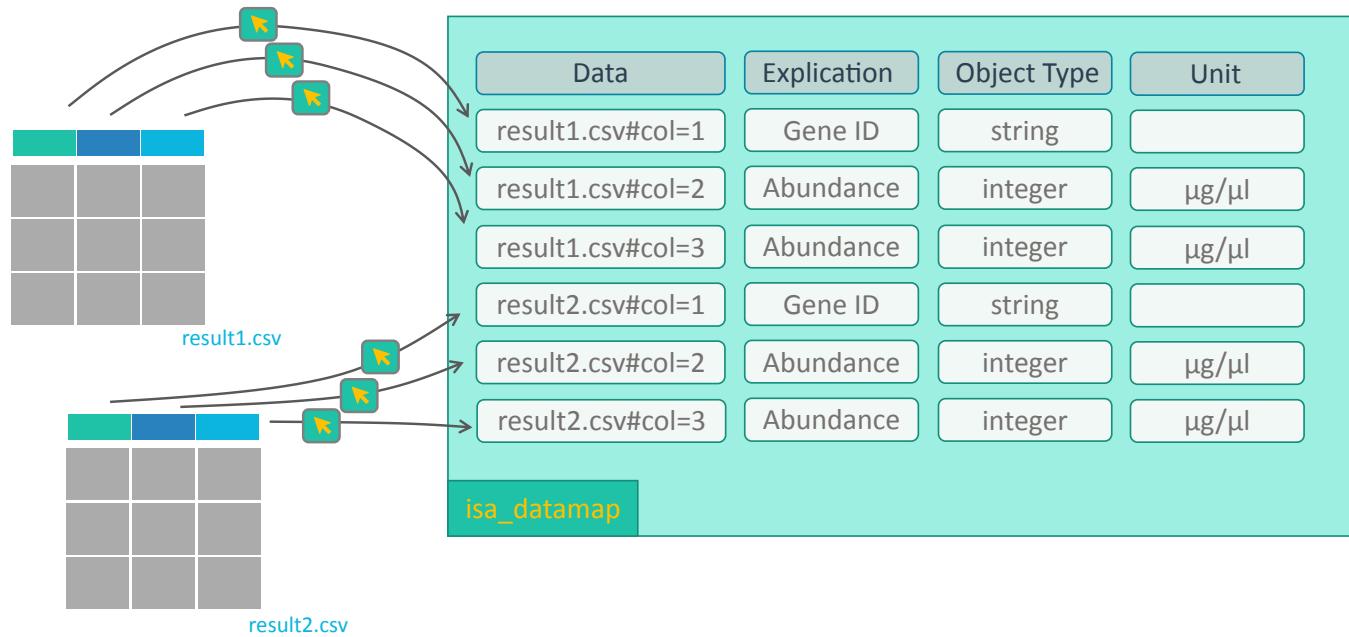
# Annotation of the result data points



# Point into results - Fragment selectors

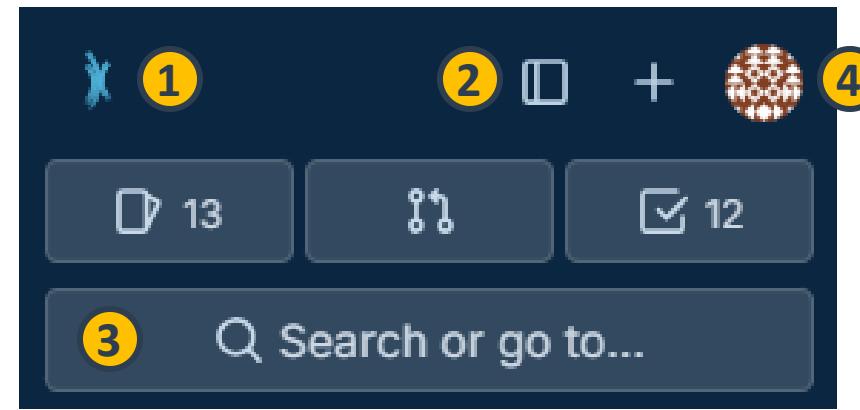


# DataMAP: Annotation for the fragment selectors



# DataHub Hands-On

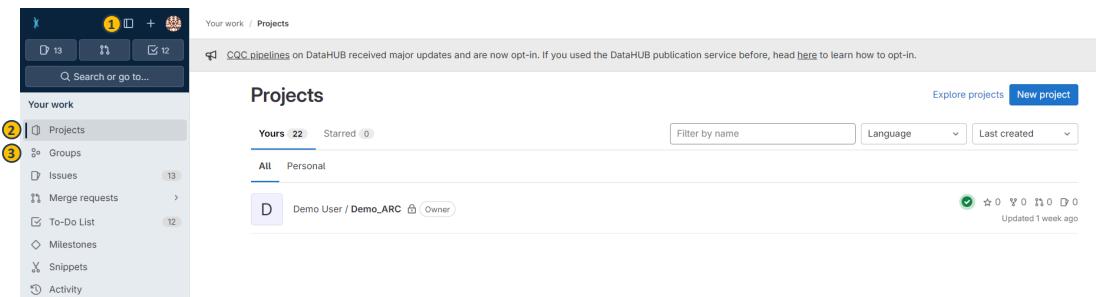
# Navigation Bar



1. navigate directly to the [projects panel](#) via the icon in the top-left (1)
2. open the [hamburger Menu](#) (2)
3. use the search field (3) to find ARCs, users and groups
4. open the [avatar Menu](#) (4)

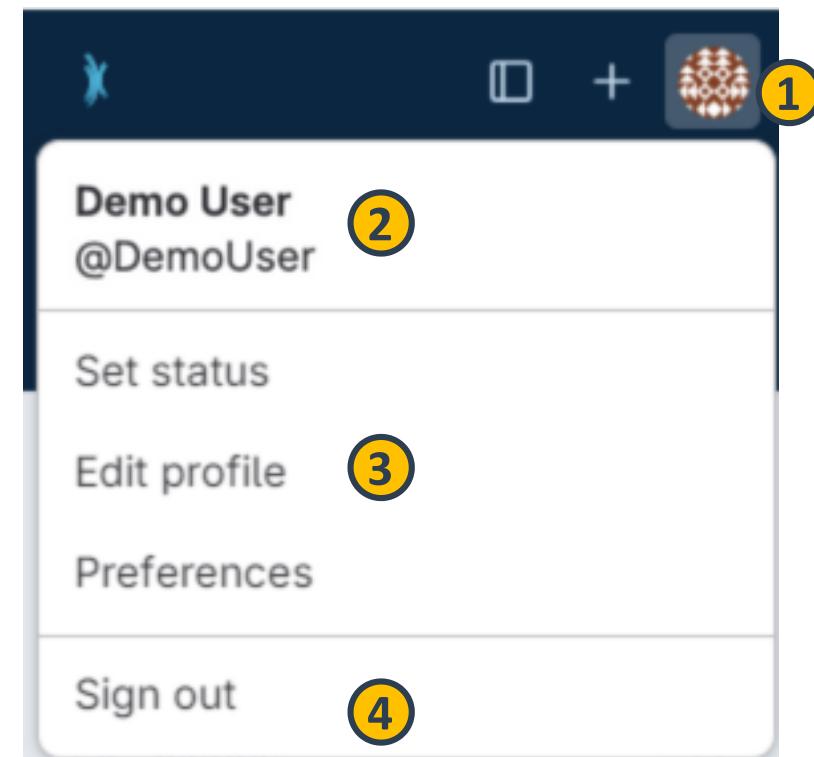
# Hamburger Menu

1. From the hamburger menu (1) you can
2. navigate to the **projects** (2)
3. or **groups** (3) panels



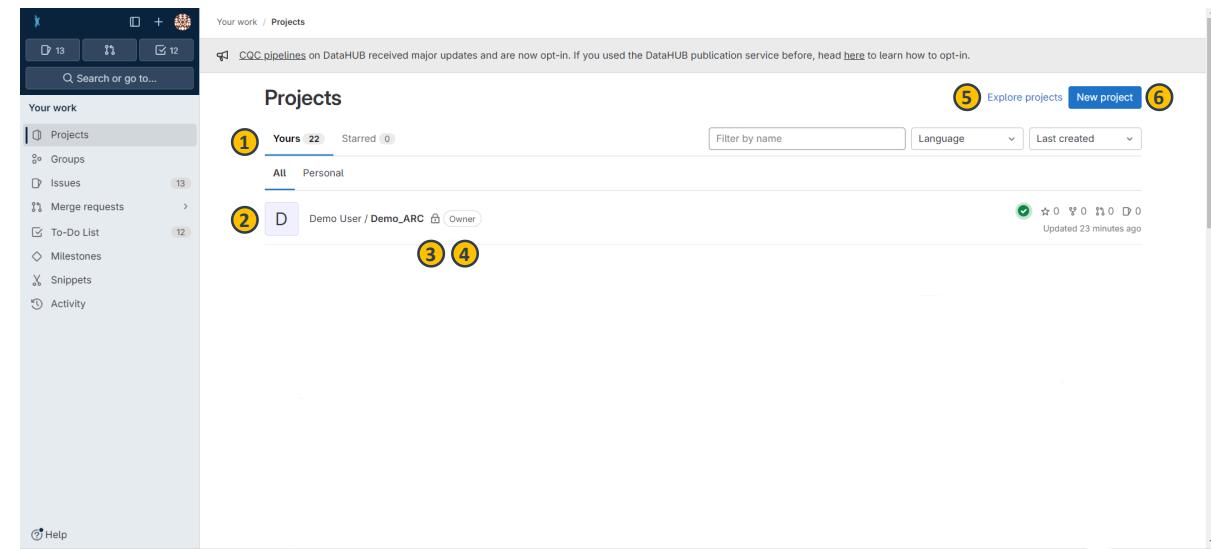
## Avatar Menu

1. In the avatar menu (1) you can
2. find your profile name and user name (2),
3. navigate to the [user settings](#) (3)
4. or sign out (4) of the DataHUB.



# Projects Panel

1. Choose a tab (1) to see only your ARCs, or explore other publicly available ARCs.
2. The main panel (2) lists all ARCs
3. Here you can also see, the visibility level (3), and
4. your permission or role (4) for the listed ARC.



# ARC Panel

The ARC Panel is the main working area for your ARC.

The screenshot shows the DataHUB ARC Panel interface. The left sidebar contains navigation links: Manage (1), Plan (2), Code (3), Build, Secure, Deploy, Operate, Monitor, Analyze, Settings (4), and Help. The main area displays a project named "Demo\_ARC" (5). The code editor shows a commit titled "arc init" (6) made by "Demo User" 4 minutes ago. A file list (7) includes "assays", "runs", "studies", "workflows", ".gitignore", and "isa.investigation.xlsx". The top right shows project information: pipeline passed (8), 1 Commit, 2 Branches, 0 Tags, 7 KiB Project Storage, Auto DevOps enabled, and various integration options like Add README, Add LICENSE, etc. The bottom right shows the project was created on July 13, 2024.

Demo User / Demo\_ARC

D Demo\_ARC locked (5)

main Demo\_ARC / + (7)

arc init  
Demo User authored 4 minutes ago

Name Last commit Last update

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

Project information

pipeline passed (8) Publish ARC

-o 1 Commit

2 Branches

0 Tags

7 KiB Project Storage

Auto DevOps enabled

+ Add README

+ Add LICENSE

+ Add CHANGELOG

+ Add CONTRIBUTING

+ Add Kubernetes cluster

+ Add Wiki

+ Configure Integrations

Created on

July 13, 2024

Data PLANT CEPLAS

CC BY

# ARC Panel – sidebar

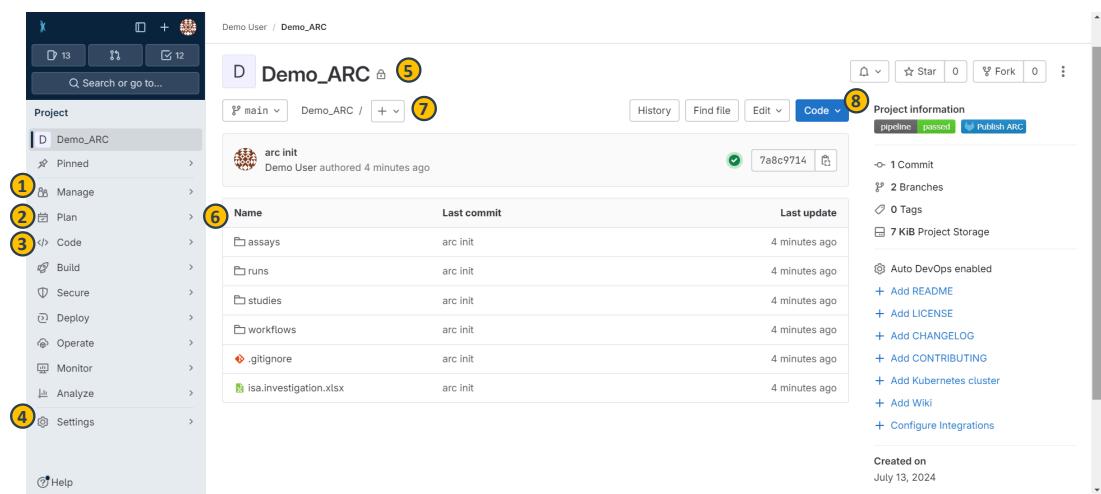
1. access the project information (1), e.g.  
invite members to the ARC
2. follow the progress of your ARC  
repository (2),
3. organize tasks in issue lists and boards  
(3),
4. take notes in a wiki to your ARC (4),
5. adapt the **settings** (5) of the ARC.

The screenshot shows the DataHUB interface. On the left is a sidebar with various project management options: Manage (1), Plan (2), Code (3), Build (4), Secure (5), Deploy (6), Operate (7), Monitor (8), Analyze (9), Settings (10), and Help (11). The main area displays project details for 'Demo User / Demo\_ARC'. It includes a header with project name, pipeline status (passed), and repository stats (1 commit, 2 branches, 0 tags, 7 KB storage). Below this is a table of files and their last commits:

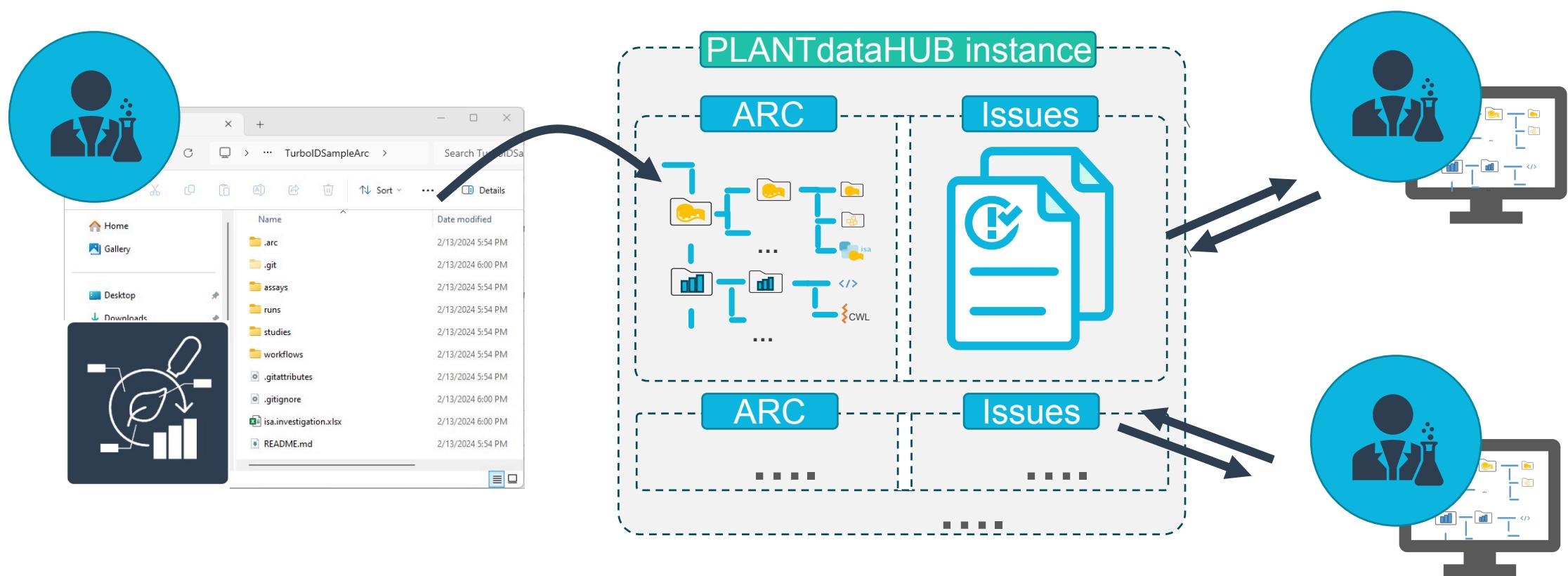
Name	Last commit	Last update
assays	arc init	4 minutes ago
runs	arc init	4 minutes ago
studies	arc init	4 minutes ago
workflows	arc init	4 minutes ago
.gitignore	arc init	4 minutes ago
isa.investigation.xlsx	arc init	4 minutes ago

## ARC Panel – main panel

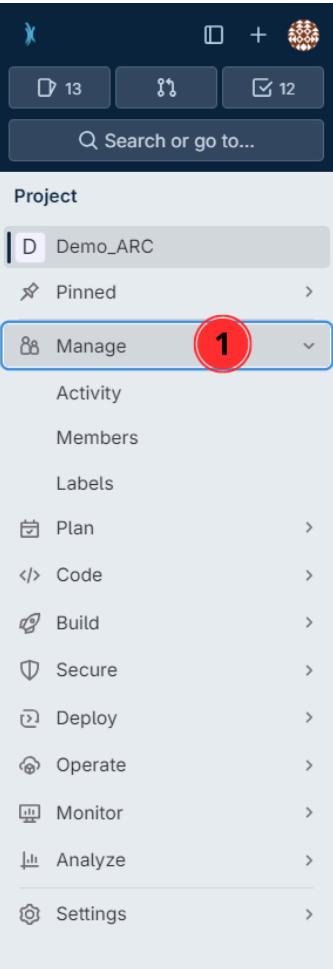
6. see the ARC's name and visibility level (6),
7. follow the ARC's commit history (7),
8. see files contained in your ARC just like on your computer (8),
9. add new files or directories (9), and
10. download or clone your ARC (10).



# Collaborate and share



# 1. Click on Project Information in the left navigation panel



The screenshot shows the DataHUB web interface. On the left, there is a vertical navigation panel with various project management options. One of these options, 'Manage', is highlighted with a red circle and the number '1' above it, indicating it is the target for the first step. The main content area displays a project named 'Demo\_ARC'. At the top of this area, there are several buttons: a bell icon, a star icon with '0' forks, a fork icon with '1' forks, and a three-dot menu icon. Below these buttons, there is a message about CQC pipelines. The central part of the screen shows a commit history for the 'main' branch, with one commit listed: 'arc init' by 'Demo User' a week ago. To the right of the commit history is a table listing files and their details. At the bottom right of the main content area, there is a 'Project information' section with a green 'pipeline passed' status and a 'Publish ARC' button. Below this, there is a list of project statistics and several blue 'Add' buttons for various features like DevOps, Wiki, README, LICENSE, CHANGELOG, CONTRIBUTING, Kubernetes cluster, and Integrations. At the very bottom, there is a note about the project's creation date: 'Created on July 13, 2024'.

Name	Last commit	Last update
assays	arc init	1 week ago
runs	arc init	1 week ago
studies	arc init	1 week ago
workflows	arc init	1 week ago
.gitignore	arc init	1 week ago
isa.investigation.xlsx	arc init	1 week ago

## 2. Click on Members

The screenshot shows the DataHUB interface for a project named 'Demo\_ARC'. On the left, there is a sidebar with various project management options like Plan, Code, Build, Secure, Deploy, Operate, Monitor, Analyze, and Settings. Two specific items are highlighted with red circles and numbers: 'Manage' (number 1) and 'Members' (number 2). The main content area is titled 'Project members' and displays a single member named 'Demo User' (@DemoUser). The member is listed under the 'Members' tab, which has a count of 1. There are buttons for 'Import from a project', 'Invite a group', and 'Invite members'. Below the member list, there are columns for Account, Source, Max role, Expiration, and Activity. The account is 'Demo User' (@DemoUser), the source is 'Direct member by Demo User', the max role is 'Owner', and the expiration date is set to 'Sep 27, 2023'.

Account	Source	Max role	Expiration	Activity
Demo User @DemoUser	Direct member by Demo User	Owner	Expiration date Sep 27, 2023 ✓ Jul 13, 2024 ✗ Jul 21, 2024	

### 3. Click on Invite members

The screenshot shows the DataHUB interface for a project named 'Demo\_ARC'. The left sidebar has a 'Members' tab selected (labeled 2). The main area shows 'Project members' with one member listed: 'Demo User @DemoUser' (labeled 1). A red circle highlights the 'Invite members' button in the top right corner of the main content area (labeled 3).

Account	Source	Max role	Expiration	Activity
Demo User @DemoUser <i>It's you</i>	Direct member by Demo User	Owner	8+ Sep 27, 2023 ✓ Jul 13, 2024 ✗ Jul 21, 2024	⋮

## 4. Search for potential collaborators

**Invite members** X

You're inviting members to the **Demo\_ARC** project.

**Username, name or email address** 4

Select members or type email addresses

**Select a role**

Guest ▼

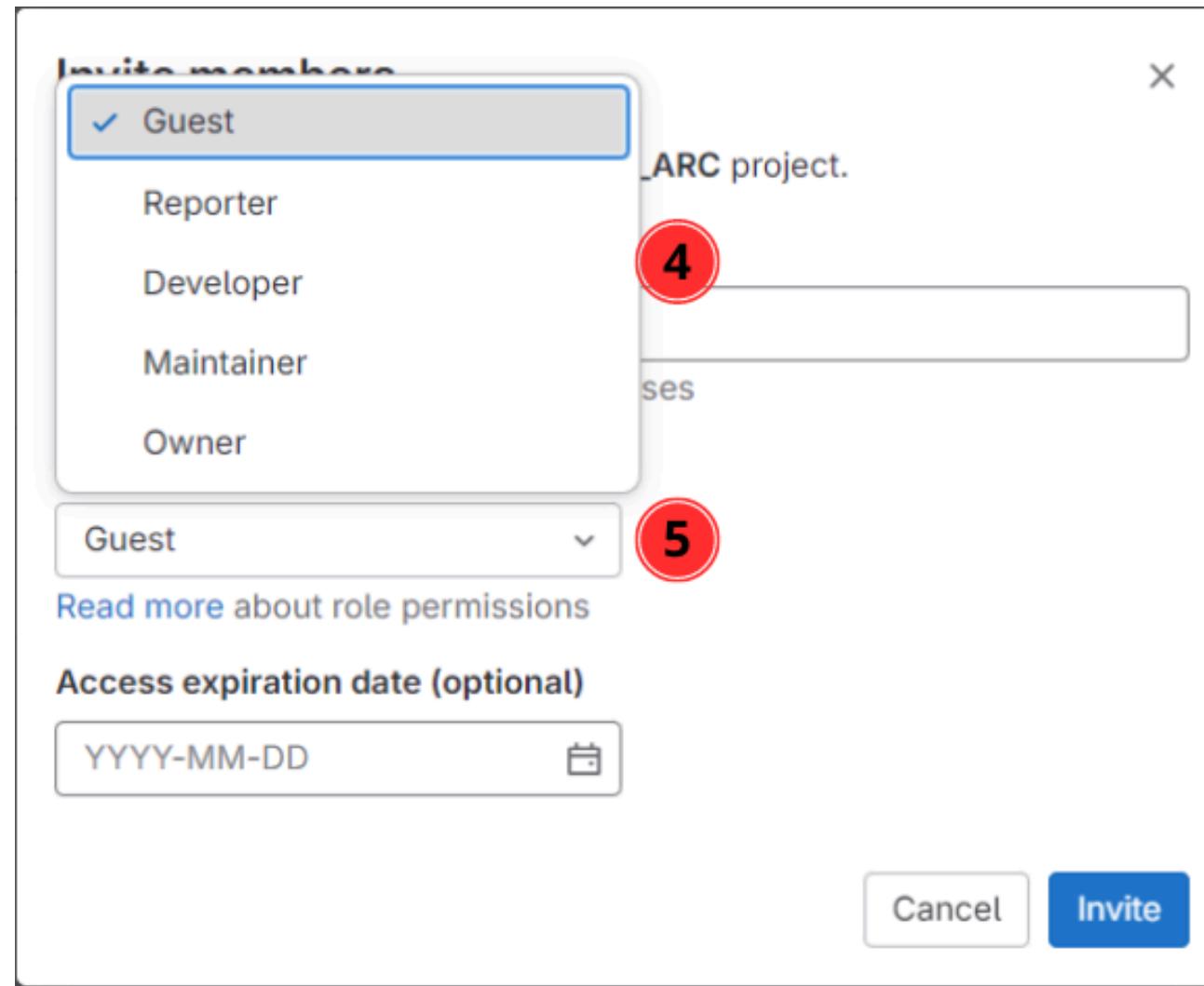
[Read more](#) about role permissions

**Access expiration date (optional)**

YYYY-MM-DD CALENDAR

Cancel Invite

## 5. Select a role



# 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

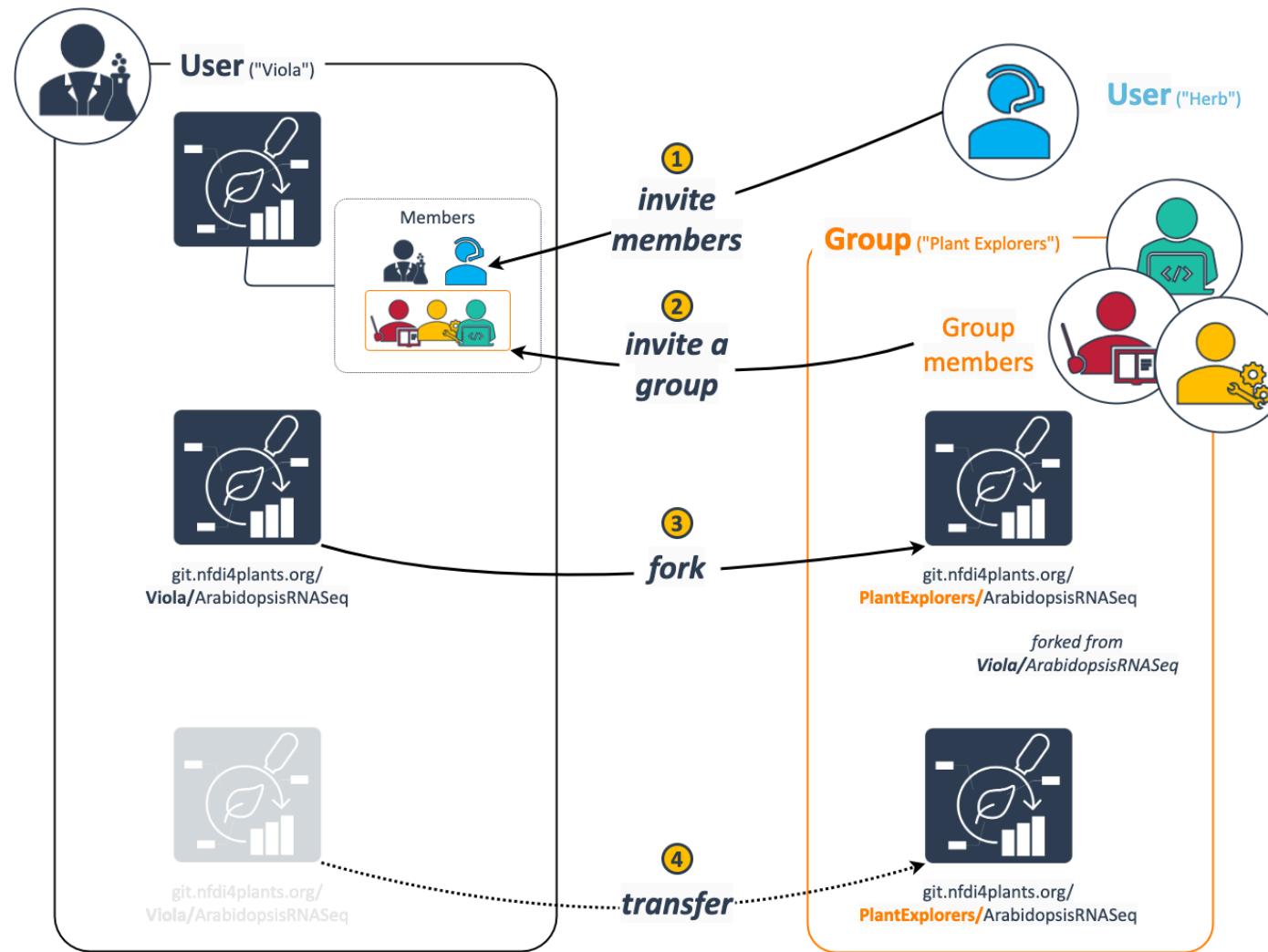
# Project = ARC

- 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	<a href="https://git.nfdi4plants.org/brilator/Facultative-CAM-in-Talinum">https://git.nfdi4plants.org/brilator/Facultative-CAM-in-Talinum</a>	brilator	Dominik Brilhaus
An group-shared ARC	<a href="https://git.nfdi4plants.org/hhu-plant-biochemistry/Samuilov-2018-BOU-PSP">https://git.nfdi4plants.org/hhu-plant-biochemistry/Samuilov-2018-BOU-PSP</a>	hhu-plant-biochemistry	HHU Plant Biochemistry

 **Personal** is not the same as **private**

# 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

The screenshot shows the 'Members' section of a GitLab project. It lists four project members:

Account	Source	Max role	Expiration	Activity
Adriano Nunes-Nesi @nunesnesi	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](https://git.nfdi4plants.org/hhu-plant-biochemistry/Samuilov-2018-BOU-PSP/-/project_members)

The screenshot shows the ARCIctect interface for the 'Samuilov-2018-BOU-PSP' project. It displays a file tree and a list of investigation contacts:

- /Users/dominikbrilhaus/datahub-dataplant/Samuilov-2018-BOU-PSP/
  - \_publications
  - assays
  - runs
  - studies
  - workflows
    - glitterignore 1.34 KB
    - LICENSE 18.21 KB
    - README.md 4.32 KB

Identifier	Title	Description
Samuilov-2018-BOU-PSP		
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 Kopriva <orcid>	3/10	
Andreas P. M. Weber <orcid>	4/10	
Tabea Mettler-Altmann <orcid>	3/10	

ARCIctect: 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 DataHUB interface with the following numbered callouts:

- 1** Manage: Project management menu.
- 2** Plan: Plan section.
- 3** Code: Code section.
- 4** Settings: Settings menu.
- 5** Demo\_ARC: Project name in the header.
- 6** Name: Column header for the file list table.
- 7** History: History tab in the repository view.
- 8** Project information: Project information sidebar.

**Demo User / Demo\_ARC**

**D Demo\_ARC** **5**

**main** Demo\_ARC / **7**

**arc init** Demo User authored 4 minutes ago **8**

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

**Project information**

- pipeline passed
- Published ARC**

-o 1 Commit  
2 Branches  
0 Tags  
7 KiB Project Storage

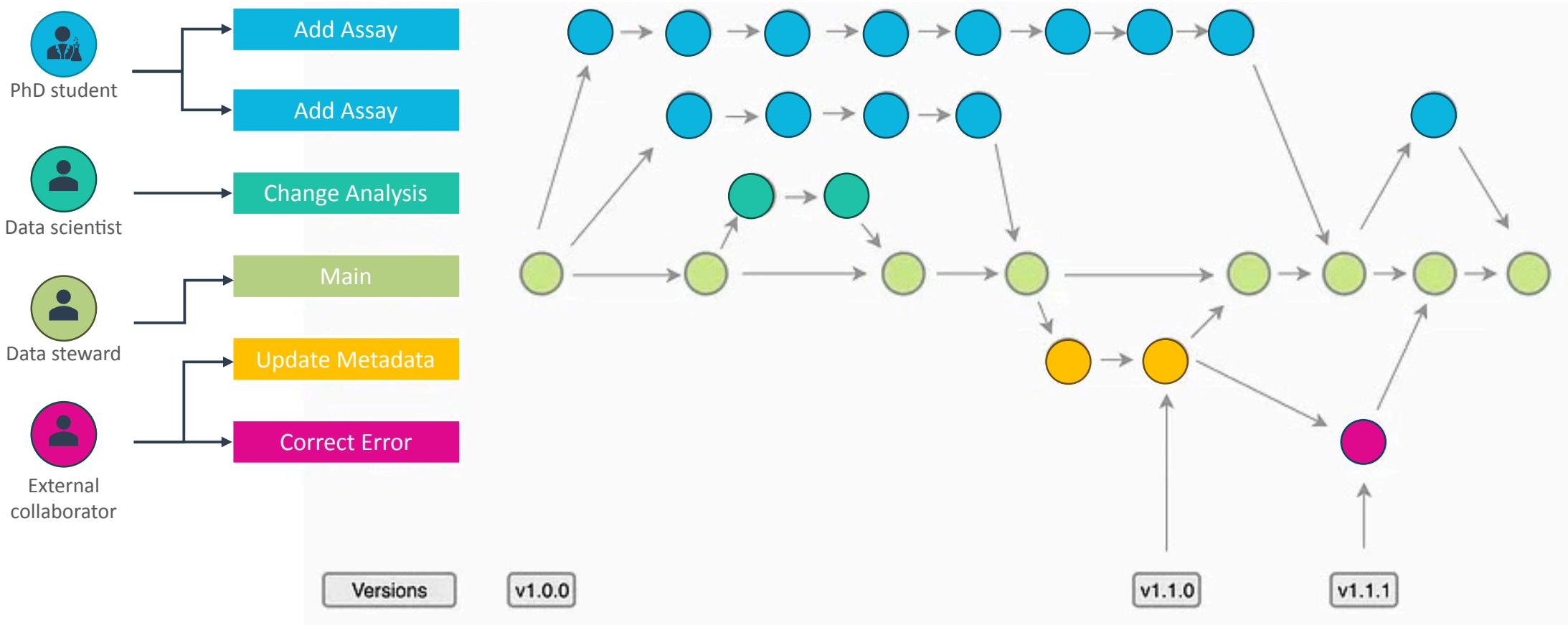
Auto DevOps enabled

+ Add README  
+ Add LICENSE  
+ Add CHANGELOG  
+ Add CONTRIBUTING  
+ Add Kubernetes cluster  
+ Add Wiki  
+ Configure Integrations

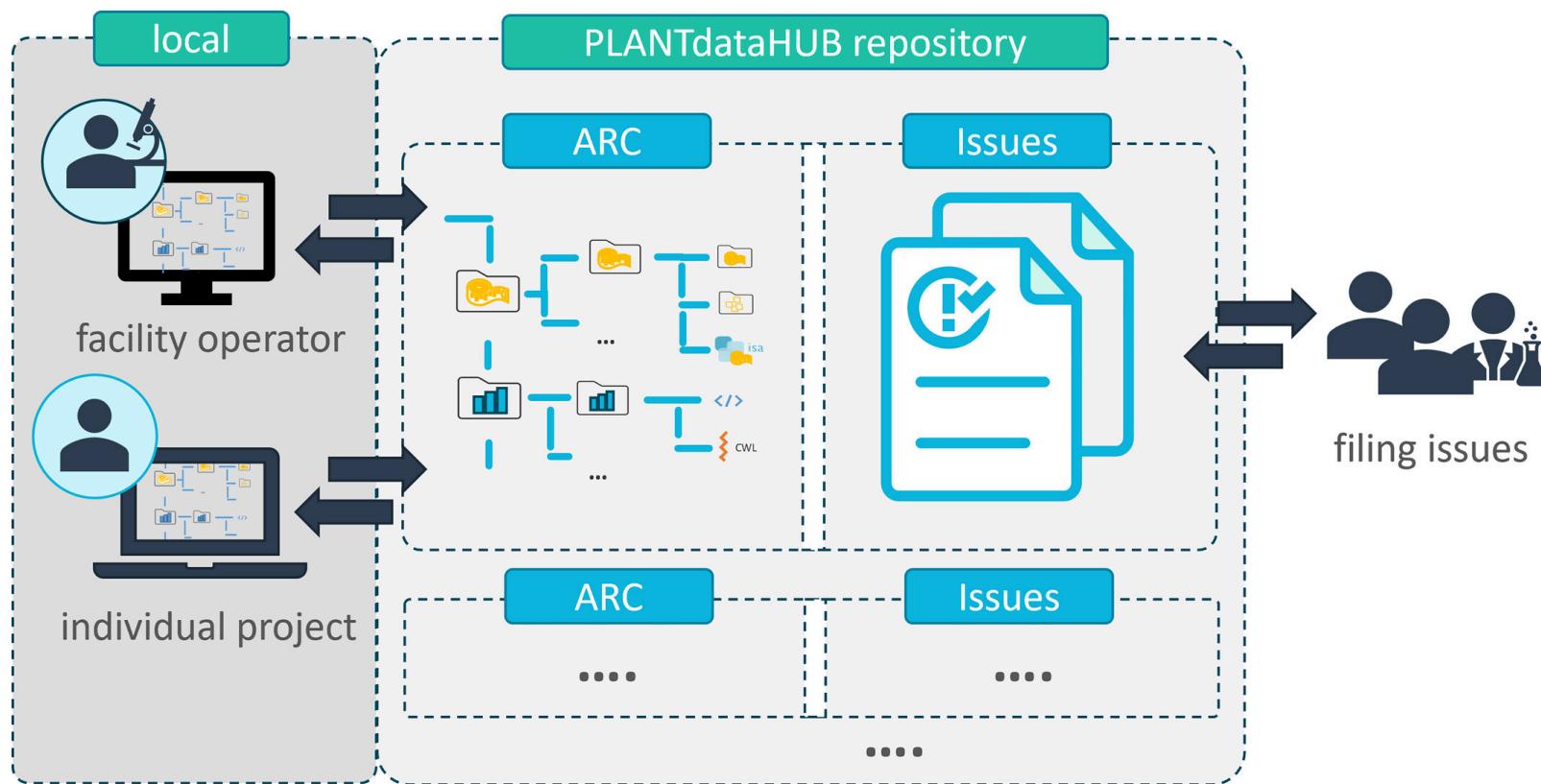
Created on July 13, 2024

**Data PLANT** **CEPLAS** **CC BY**

# Collaboration Workflows



# Project management



# Project Management

## Add issues to your ARC (3)

The screenshot shows the DataHUB Project Management interface. On the left, there is a sidebar with various project management options: Manage (1), Plan (2), Code (3), Build, Secure, Deploy, Operate, Monitor, Analyze, Settings (4), and Help. The main workspace is titled "Demo User / Demo\_ARC". It displays a file tree under "arc init" with the following structure:

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

At the top right, there are buttons for History, Find file, Edit, Code (selected), and Project information. The Project information section shows a green "pipeline passed" status and a blue "Publish ARC" button. Below this, it lists 1 Commit, 2 Branches, 0 Tags, and 7 KiB Project Storage. There are also links to Auto DevOps enabled and various integration options like Add README, Add LICENSE, etc.

# ARCs come with their own wiki space

- directly associated to your ARC
- same access rights as your ARC
- share meeting minutes or ideas with collaboration partners
- keep ARC clean of files that are not considered "research data"

The screenshot shows the DataHUB interface. On the left, a sidebar for the 'Demo\_ARC' project is visible, featuring sections for Project (with Demo\_ARC pinned), Manage, Plan, Issues (0), Issue boards, Milestones, and Wiki. The Wiki section is currently selected. The main content area displays a 'Home' page for the 'Wiki' of the 'Demo\_ARC'. The page header includes navigation links for Demo User, Demo\_ARC, Wiki, and Home. A note at the top states: 'QC.pipelines on DataHUB received major updates and are now opt-in. If you used the DataHUB publication service before, head [here](#) to learn how to opt-in.' Below this, the 'Home' section is titled 'Home' and last edited by 'Demo User just now'. It contains the message: 'This is the wiki to Demo\_ARC. We will announce meeting schedules here.' To the right of the 'Home' section, there are two more sections: 'Meeting Schedule' (listing events like Kick-off, Proposal discussion, and RNA-seq pipeline) and 'Ideas and drafts' (listing a Golden Gate protocol). At the bottom right of the main content area, there are 'Pages 2' and a gear icon.

# Publish your ARC with a few clicks

The screenshot shows a project page for "Ru\_ChlamyHeatstress". The header includes a small icon of a green cell with a鞭毛 (flagella) and two temperature markers (red and blue), followed by the project name and a globe icon. Below the name is "Project ID: 122" and a copy icon. To the right is a "Star" button with the number "1".

Key statistics below the header: 53 Commits, 1 Branch, 0 Tags, and 293.9 GB Project Storage.

Topics listed include Chlamydomonas, abiotic stress, proteomics, and one more topic.

A descriptive text block states: "Algae cultures were grown mixotrophically (TAP). After 24h of 35°C/40°C the cells were shifted back to room temperature for 48h. 'omics samples were taken."

At the bottom, there is a navigation bar with three items: "pipeline" (status passed), "Publish ARC" (button with a blue arrow icon), and "arc quality" (status 301/301).

# Receive a DOI

Published September 7, 2023 | Version v1

Dataset

Open

## Systems-wide investigation of responses to moderate and acute high temperatures in the green alga Chlamydomonas reinhardtii.

Zhang, Ningning<sup>1</sup>; Mattoon, Erin<sup>1</sup>; McHargue, Will<sup>1</sup> ; Venn, Benedict<sup>2</sup> ; Zimmer, David<sup>2</sup> ; Pecani, Kresti<sup>3</sup>; Jeong, Jooyeon<sup>1</sup>; Anderson, Cheyenne<sup>1</sup>; Chen, Chen<sup>4</sup>; Berry, Jeffrey<sup>1</sup>; Xia, Ming<sup>1</sup>; Tzeng, Shin-Cheng<sup>1</sup> ; Becker, Eric<sup>1</sup>; Pazouki, Leila<sup>1</sup>; Evans, Bradley<sup>1</sup>; Cross, Fred<sup>3</sup>; Cheng, Jianlin<sup>4</sup>; Czymmek, Kirk<sup>1</sup> ; Schröda, Michael<sup>5</sup> ; Mühlhaus, Timo<sup>2</sup> ; Zhang, Ru<sup>1</sup>

Show affiliations

Style

APA

①  
Citation

Zhang, N., Mattoon, E., McHargue, W., Venn, B., Zimmer, D., Pecani, K., Jeong, J., Anderson, C., Chen, C., Berry, J., Xia, M., Tzeng, S.-C., Becker, E., Pazouki, L., Evans, B., Cross, F., Cheng, J., Czymmek, K., Schröda, M., ... Zhang, R. (2023). Systems-wide investigation of responses to moderate and acute high temperatures in the green alga Chlamydomonas reinhardtii. [Data set]. DataPLANT. <https://doi.org/10.60534/9e5jx-75d83>

### Description

hosted on: <https://git.nfdl4plants.org/projects/122>

### Files

arc-summary.md

[Data set] Systems-wide investigation of responses to moderate and acute high temperatures in the green alga Chlamydomonas reinhardtii.

#### File contents:

- root
  - issa.investigation.xlsx
  - README.md
  - runs

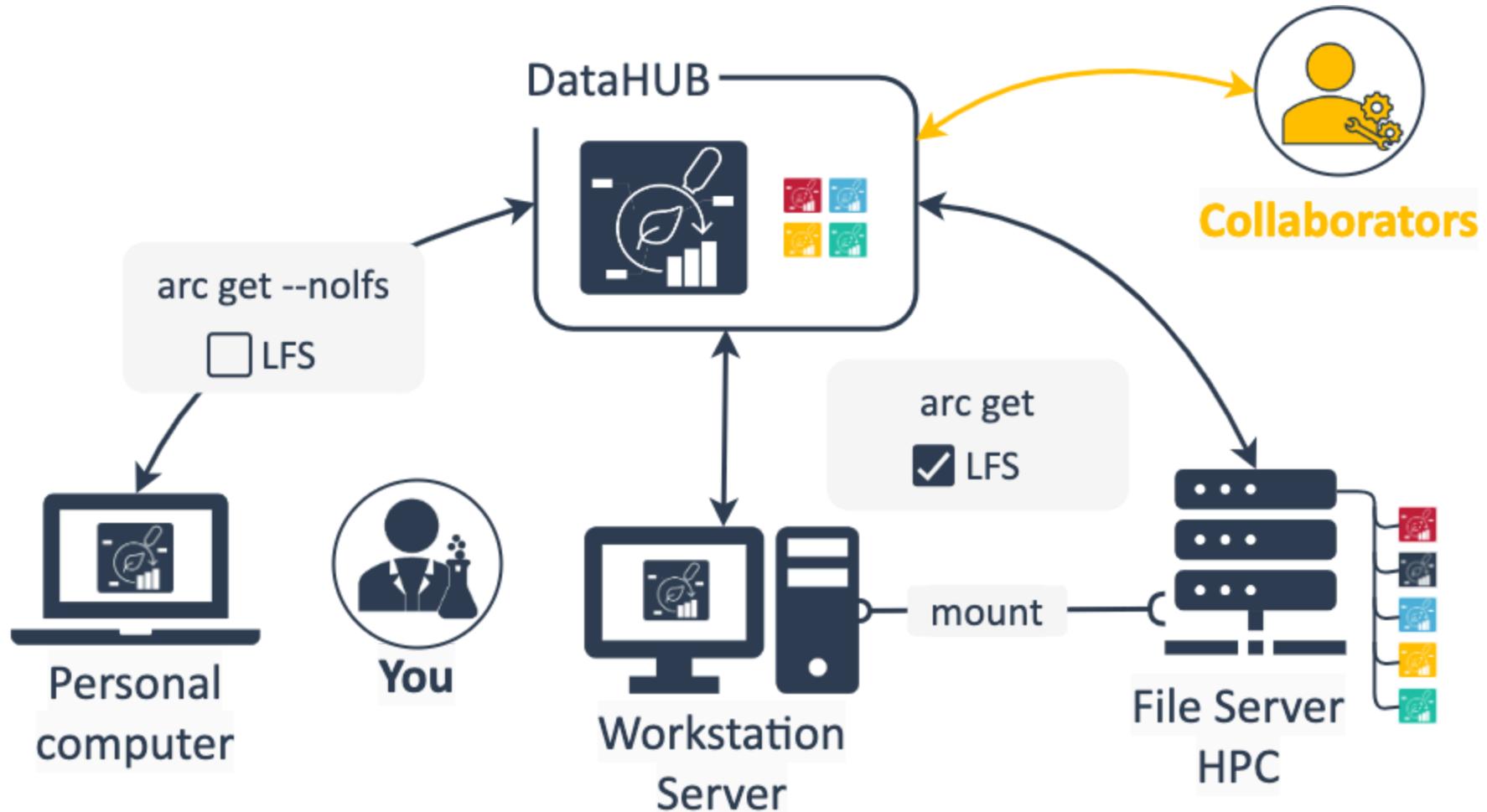
②

The screenshot shows the DataHUB interface for a dataset. At the top right, there are buttons for 'Dataset' and 'Open'. Below the title, the citation information is displayed. On the left, there's a 'Description' section with a link to the project page. The 'Files' section lists 'arc-summary.md' with its content displayed below it. The right side of the screen is divided into several sections: 'Versions' (listing 'Version v1' with a DOI), 'Details' (showing the DOI as 10.60534/9e5jx-75d83, resource type as 'Dataset', and publisher as 'DataPLANT'), and an 'Export' section with a 'JSON' dropdown and an 'Export' button.

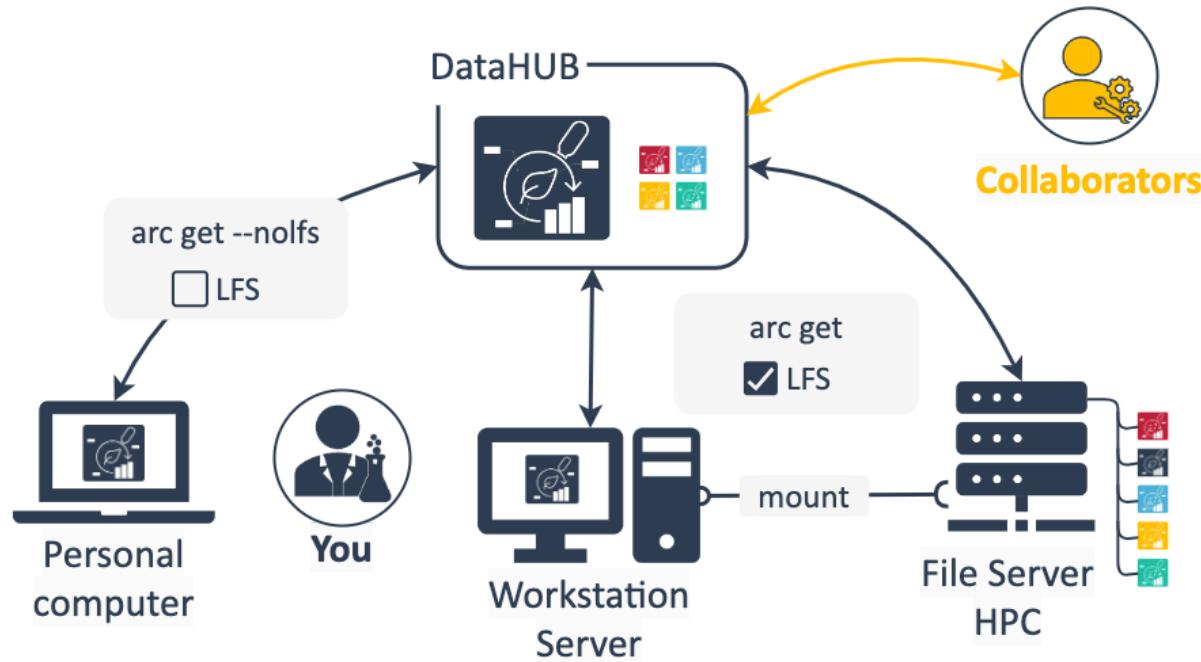


# Backup

# Where do I store my ARC?



# ARC storage and sharing



- DataHUB as "ground truth" / original clone
- You can sync and communicate all changes to your ARC via the DataHUB

💡 ARCitect and ARC commander provide options to avoid syncing large files (LFS = Large file storage)

# Example setup to store and use ARCs

## Personal computer

- work on small files
- annotate metadata
- add scripts, protocols

## Workstation / Server

- work on large files
- run computations

## FileShare

- mount to local machine, sync ARC from there

## HPC

- direct connection HPC to DataHUB (depends on security settings)
- or mount to local machine and sync, ARC from there



# Contributors

If not referenced otherwise, figures and slides presented here were created by members of DataPLANT (<https://nfdi4plants.org>).

Additional slides were contributed by

- name: Dominik Brilhaus  
github: <https://github.com/brilator>  
orcid: <https://orcid.org/0000-0001-9021-3197>
- name: Cristina Martins Rodrigues  
github: <https://github.com/CMR248>  
orcid: <https://orcid.org/0000-0002-4849-1537>
- name: Kevin Frey  
github: <https://github.com/Freymaurer>  
orcid: <https://orcid.org/0000-0002-8493-1077>
- name: Sabrina Zander  
orcid: <https://orcid.org/0009-0000-4569-6126>
- name: Martin Kuhl  
github: <https://github.com/Martin-Kuhl>  
orcid: <https://orcid.org/0000-0002-8493-1077>

