



**CEPLAS**

Cluster of Excellence on Plant Sciences

# CEPLAS ARC Training

April, 2024

Dominik Brilhaus, CEPLAS

Sabrina Zander, MibiNet



# Tentative agenda

## Morning

Time	Topics
09:00 - 09:15	Welcome, Intro RDM and NFDI
09:15 - 09:45	Overview DataPLANT, and DataPLANT Tool chain
09:45 - 10:30	Demo ARCitect and Metadata annotation with SWATE
10:30 - 10:45	<i>Short break</i> ☕
10:45 - 11:45	Intro and Demo DataHUB
11:45 - 12:00	Q & A

## Afternoon

Time	Topics
12:00 - 13:00	<i>Lunch break</i> 🍕
13:00 - 15:00	Hands-on ARCitect, DataHUB and Swate
15:00 - 15:15	<i>Short break</i> ☕
15:15 - 16:30	Create your own ARC
16:30 - 17:00	Q & A and wrap-up

# 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.

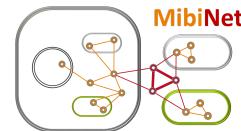
1. Slack Workspace for ad hoc support
2. Monthly user meeting (2nd Friday of the month)

→ [Details](#).



trr\_341

plant  
ecological  
genetics



# Introduce yourself

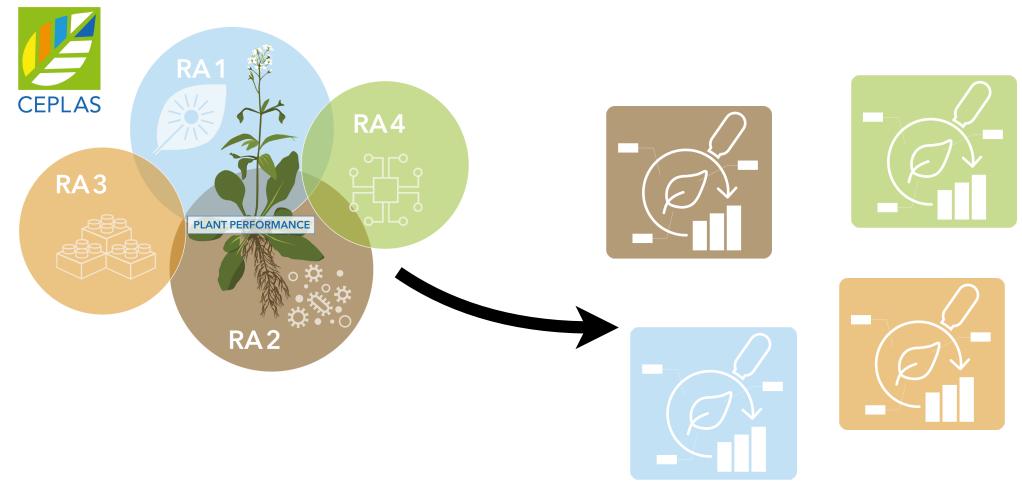
- Lab
- CEPLAS / TRR / MibiNet / ?
- Used code / programming language before
- Experience with Git / GitLab / GitHub?
- Have an ORCID
- My motivation / expectation
- My favorite lab assay

# Let's draw a typical lab workflow



# BYOD – Goals

- Move existing datasets into ARCs
- Share them via the DataHUB
- First few steps into ARCs
- You or collaborators can pick them up from there



# FAIR data stewardship

- Findable
- Accessible
- Interoperable
- Reusable

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

[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](#)

# The FAIR principles

<p><b>Findable</b> <b>Accessible</b> <b>Interoperable</b> <b>Reusable</b></p>	<p><b>Easier collaboration &amp; sharing</b></p> <pre>graph TD; A(( )) --&gt; B(( )); A(( )) --&gt; C(( )); A(( )) --&gt; D(( )); B(( )) --&gt; E(( )); C(( )) --&gt; E(( )); D(( )) --&gt; E(( ));</pre>	<p><b>Increased findability and visibility</b></p> <pre>graph LR; P1(( )) --- C1[Bar Chart]; P2(( )) --- C2[Code]; P3(( )) --- C3(Flower); C1 --- C2 --- C3;</pre>	<p><b>Reproducibility</b></p> <pre>graph LR; C1[Blue Bar Chart] --- C2[Yellow Bar Chart]; C2 --- Checkmark[Checkmark];</pre>
<p><b>Added-value to the research community</b></p> <p>nfdi      NCBI EMBL-EBI</p>	<p><b>Compliance with funding policies</b></p> <pre>graph LR; Doc[Document with Checkmark] --- DFG[DFG]; Doc --- EU[European Union Flag];</pre>	<p><b>Receive due credit</b></p> <pre>graph LR; Reuse[Reuse] --- Citations[Citations];</pre>	<p><b>Saves time &amp; workload</b></p> <p>FAIR</p> <p>Time wasted</p>

# Is your data FAIR?

Findable | Accessible | Interoperable | Reusable

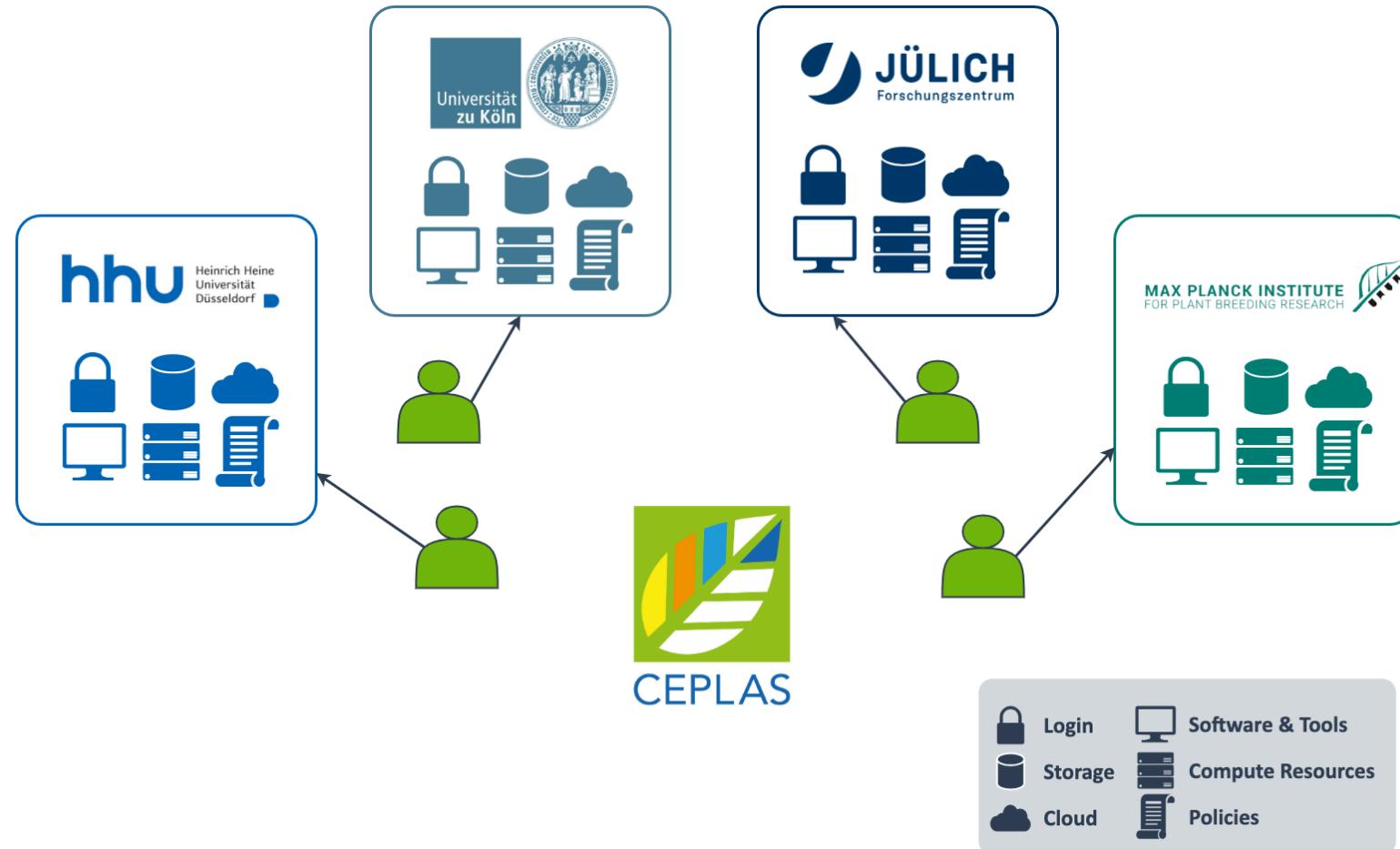
- Where do you store your data?
- How do you annotate your data?
- How do you share your data?
- What tools do you use to analyse your data?
- How do you reuse other people's data?



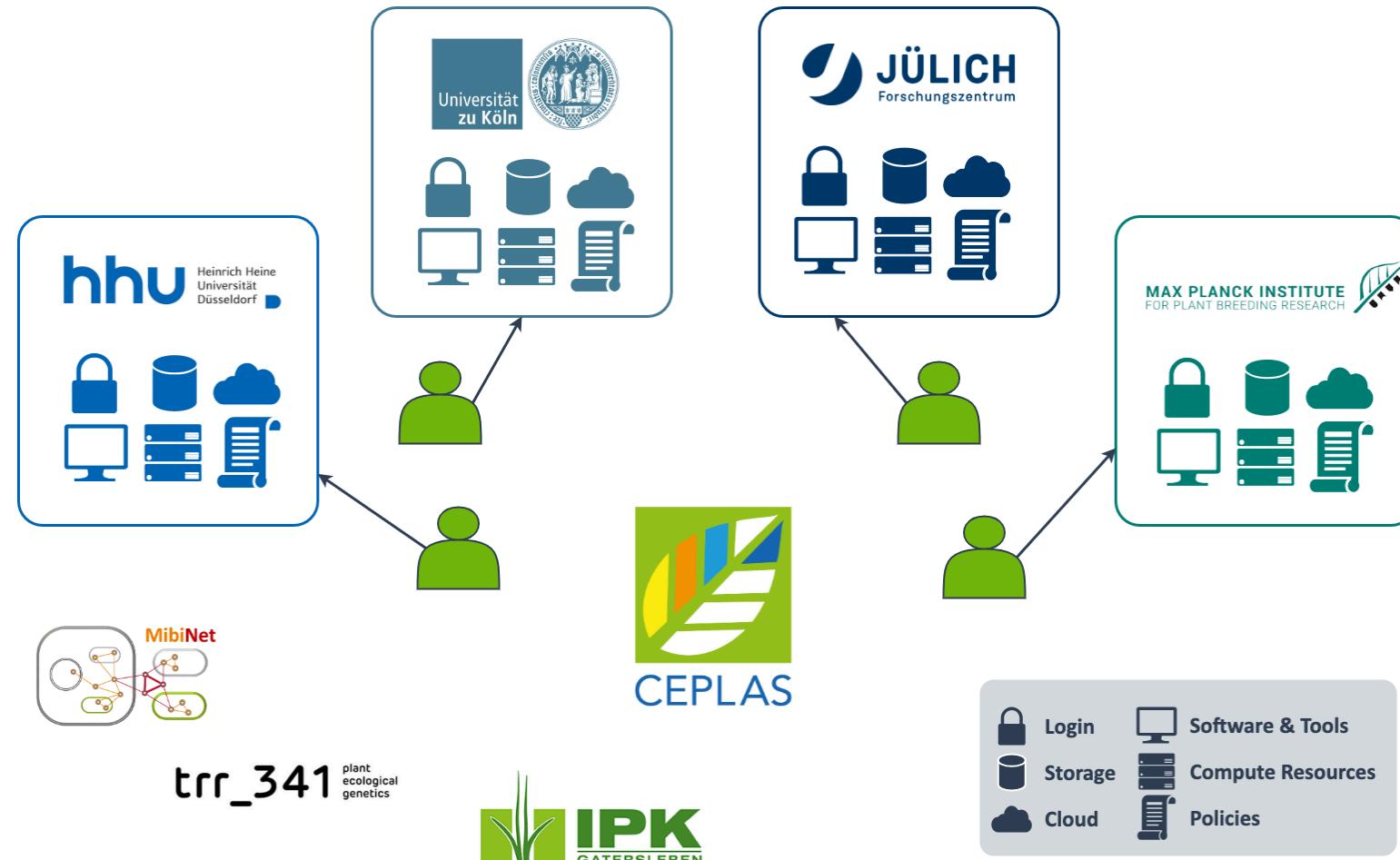
# CEPLAS – One cluster, four locations



# Data silos impede collaboration

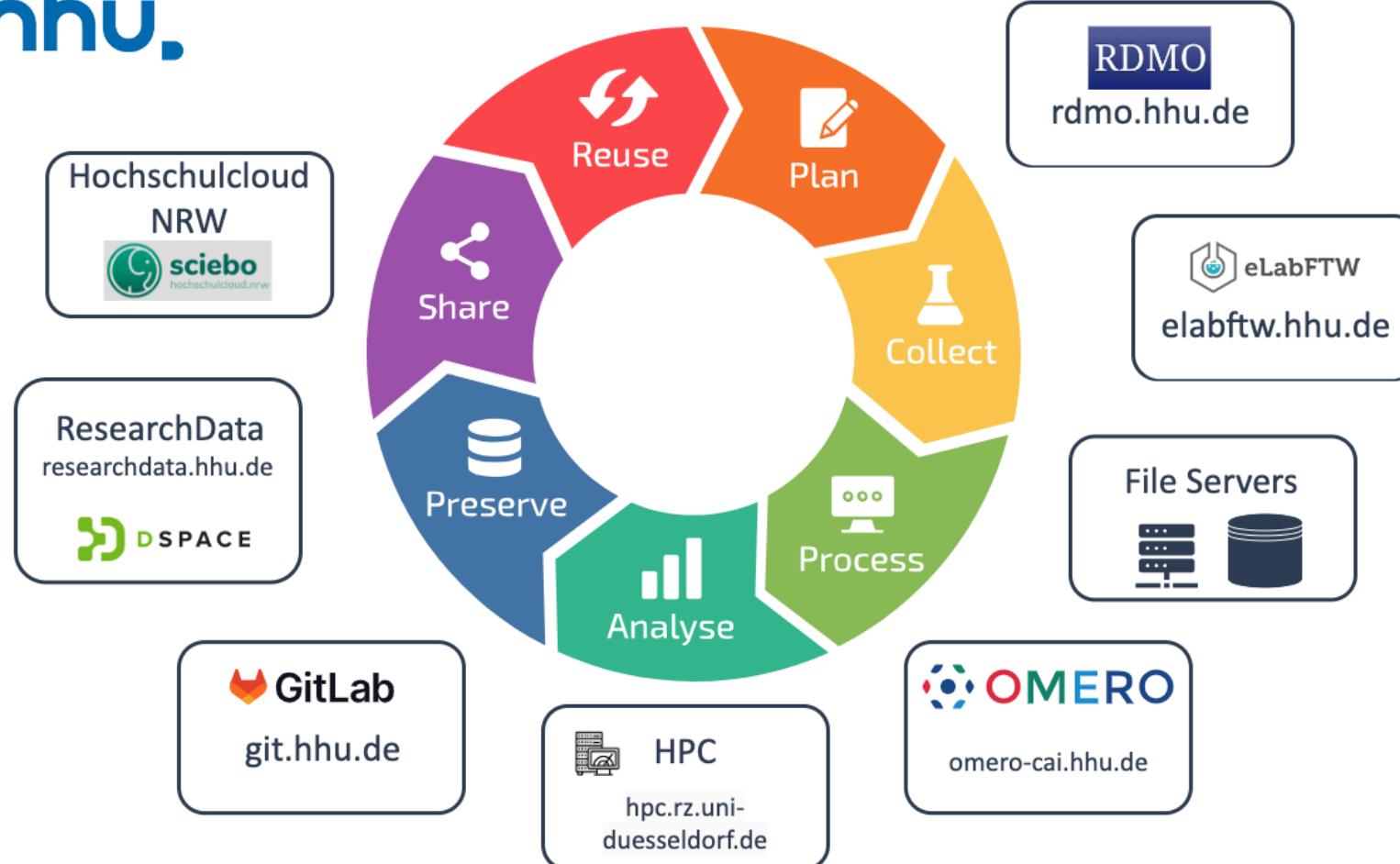


# Data silos impede collaboration



# Missing interfaces impede collaboration

hhu.



# Data Stewardship between DataPLANT and the community

*Community*

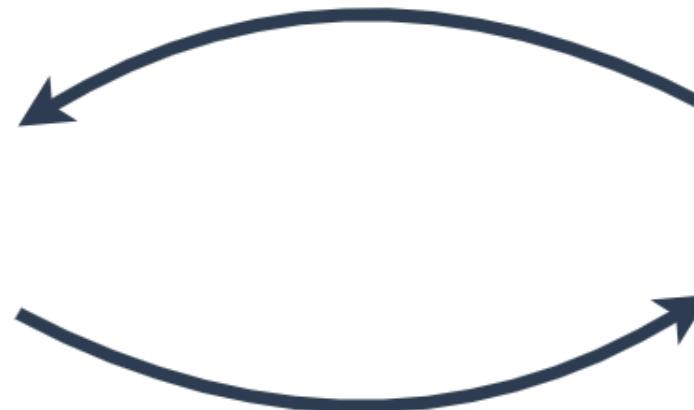


Domain experts  
User experience  
Training

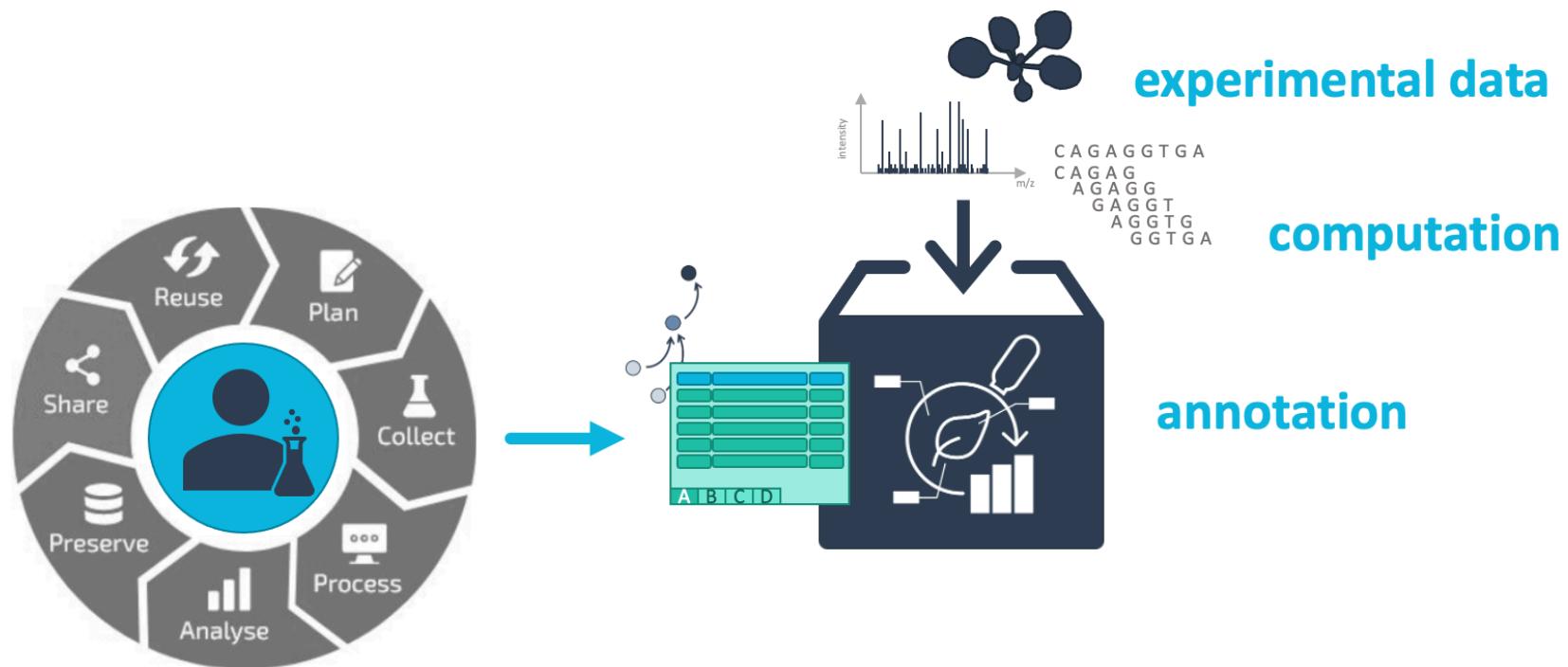
*nfdi4plants*



Service provider  
Developers  
Tech experts

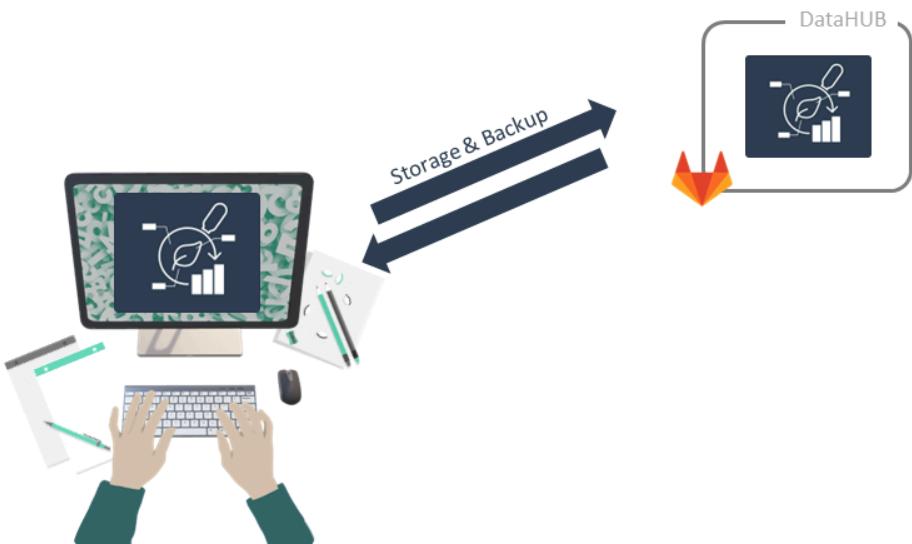


# Annotated Research Context (ARC)

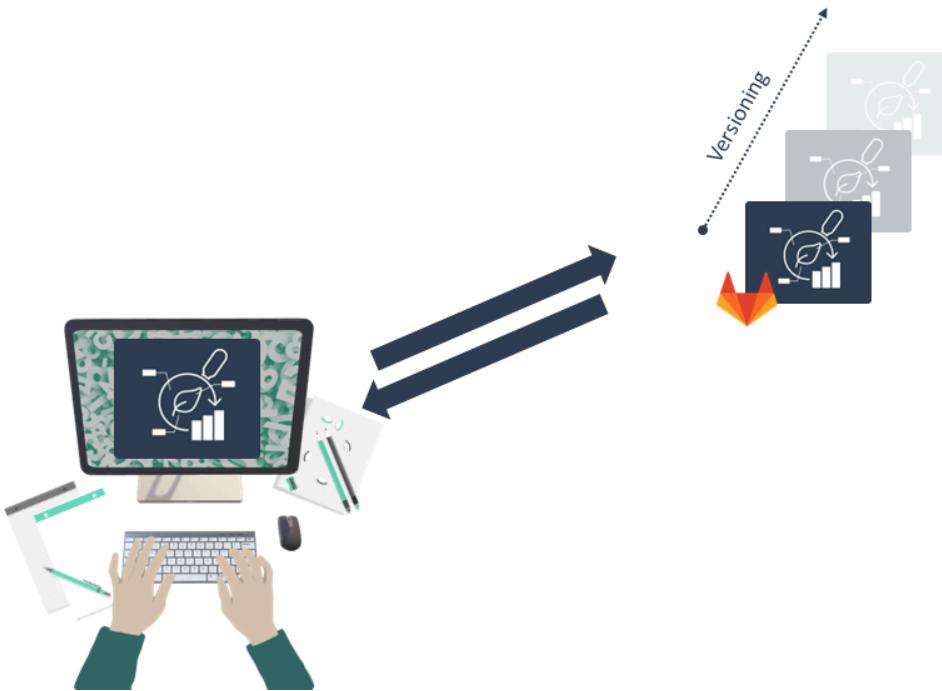


Your entire investigation in a single unified bag

# You can store your ARC in the DataHUB



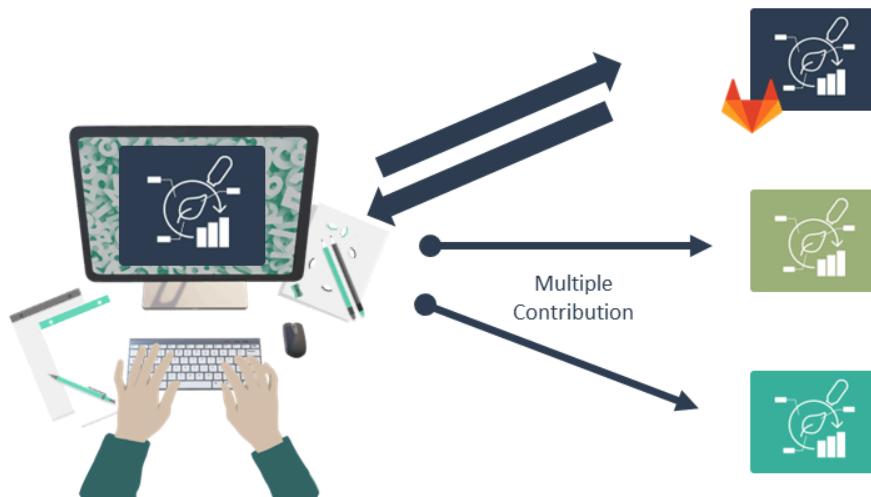
# ARCs are versioned



# You can invite collaborators



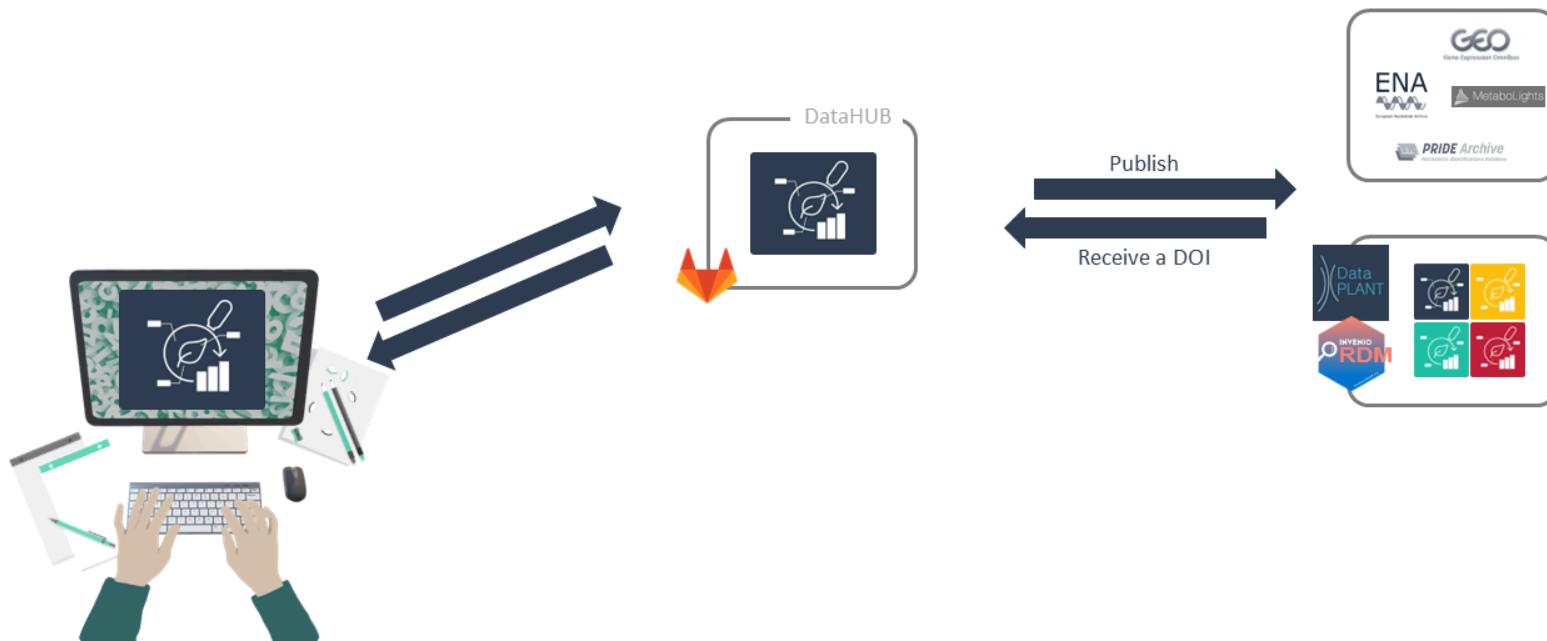
# Collaborate and contribute



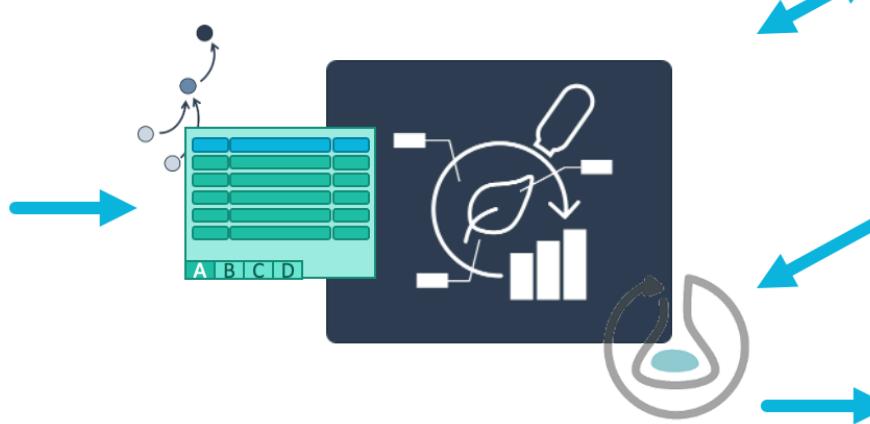
# Reuse data in ARCs



# 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

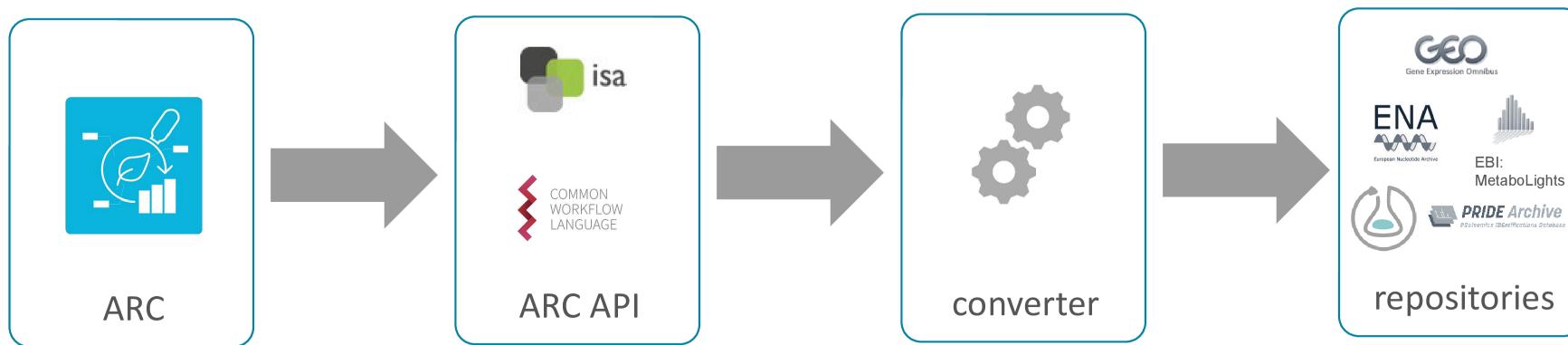
OpenAIRE

The Dataverse® Project

INVENIO

data publication

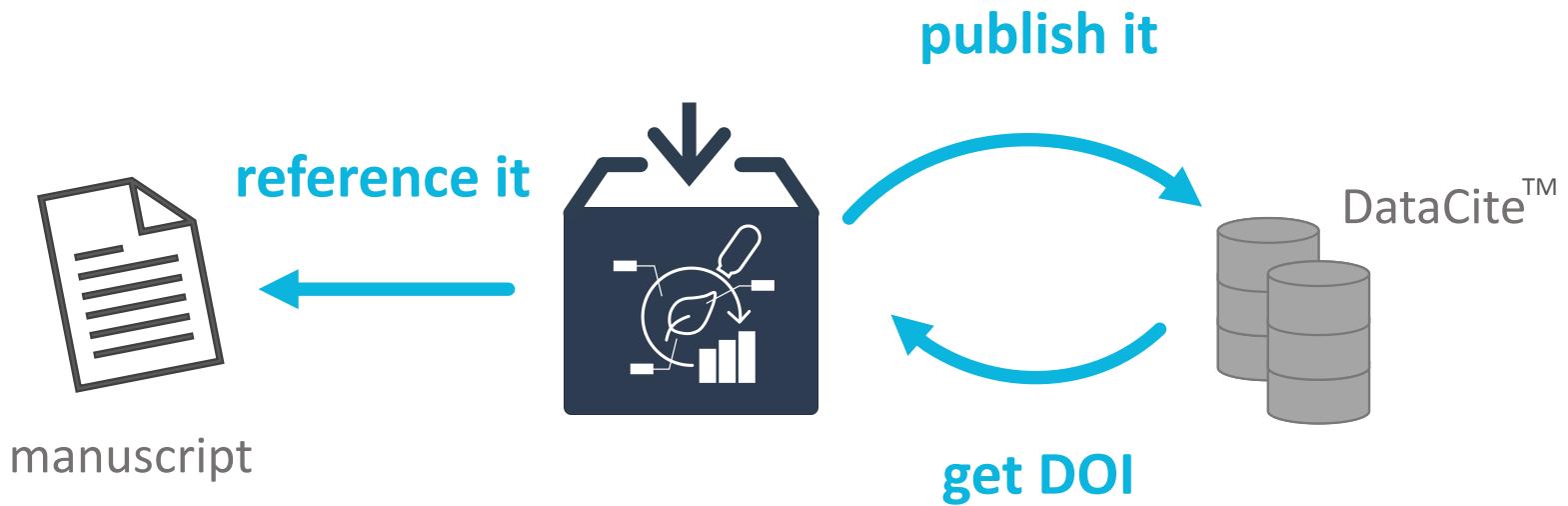
# From ARC to repositories



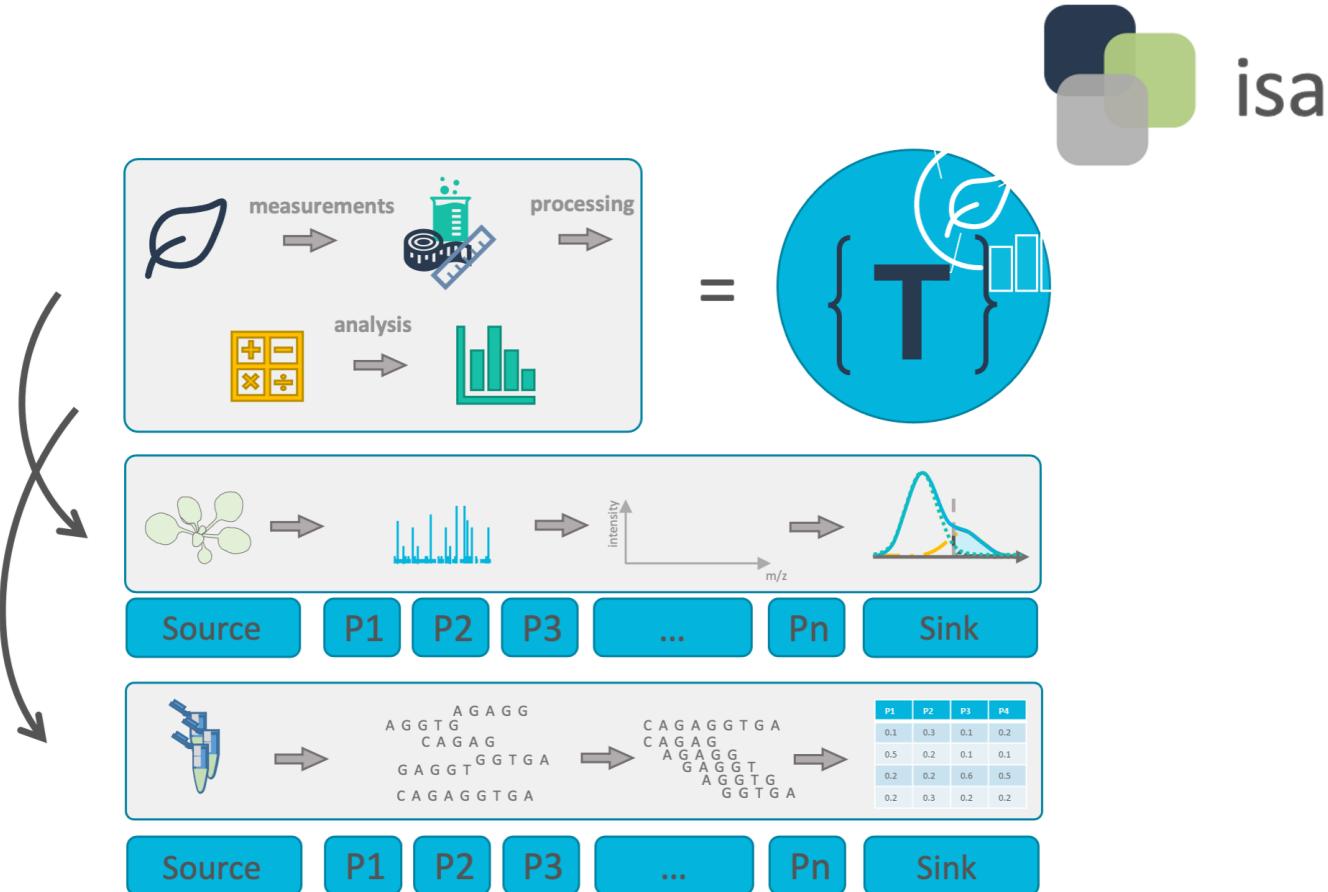
# Moving from paper to data publications



# Publish your ARC, get a DOI

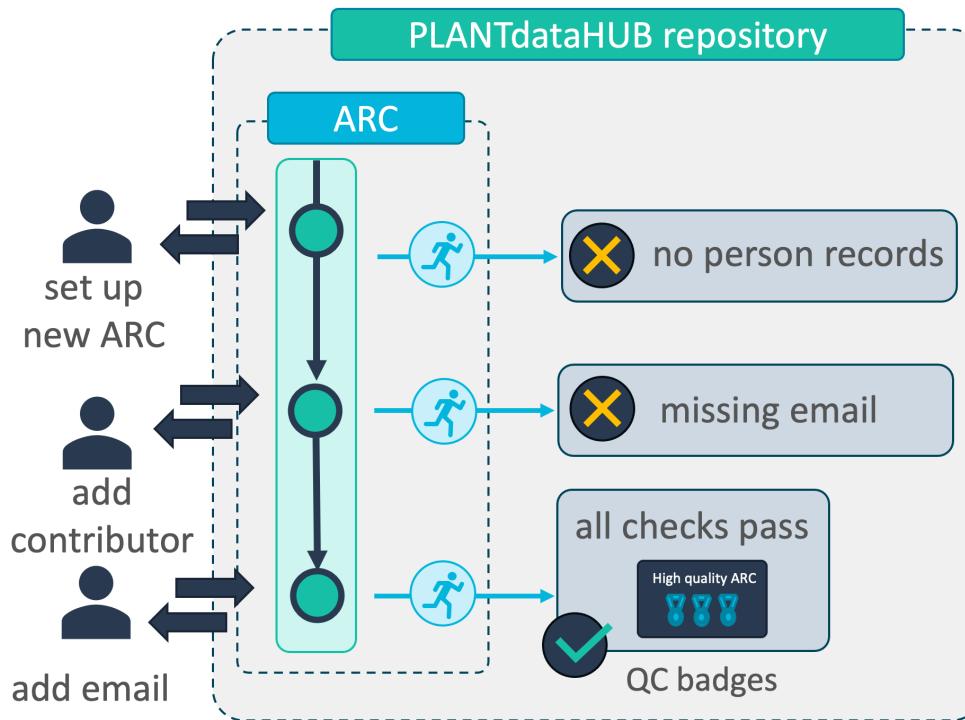


# Metadata templates



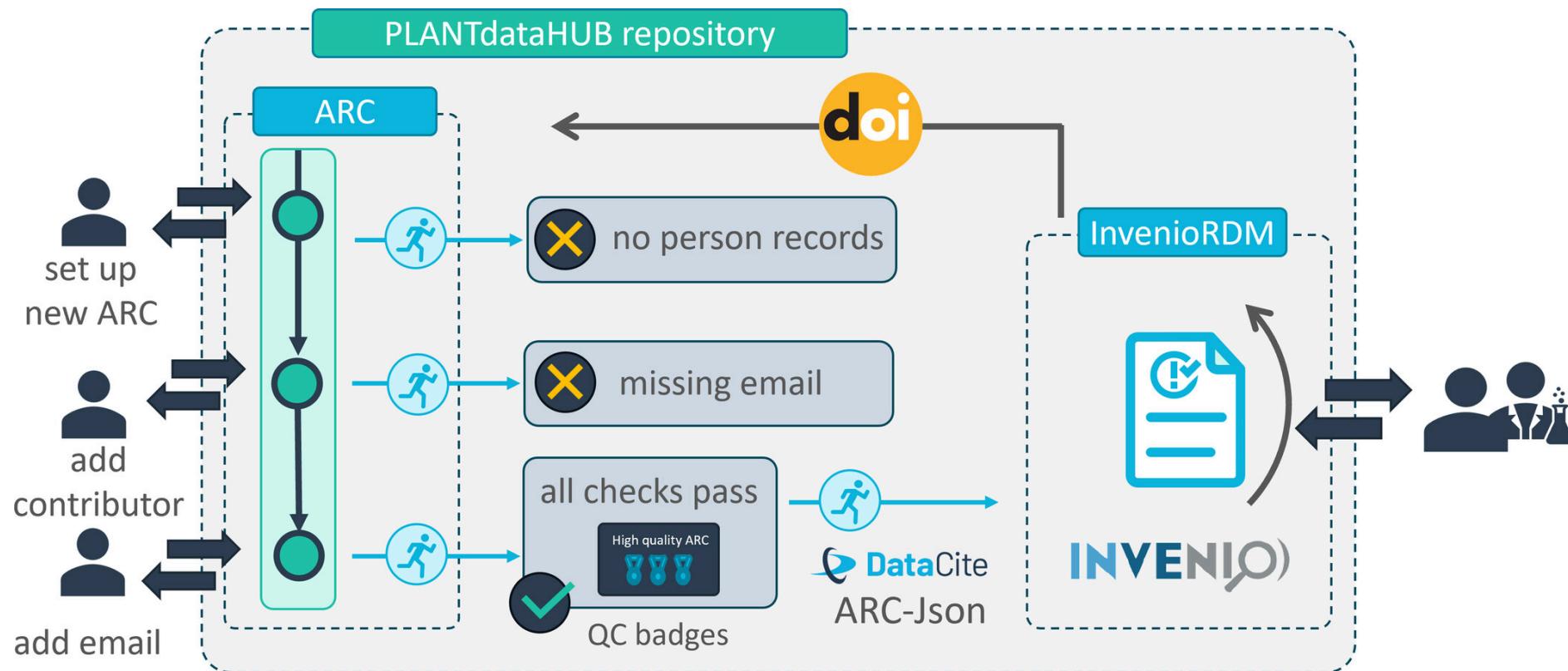
Facilities and labs can define their common workflows as templates

# Validation

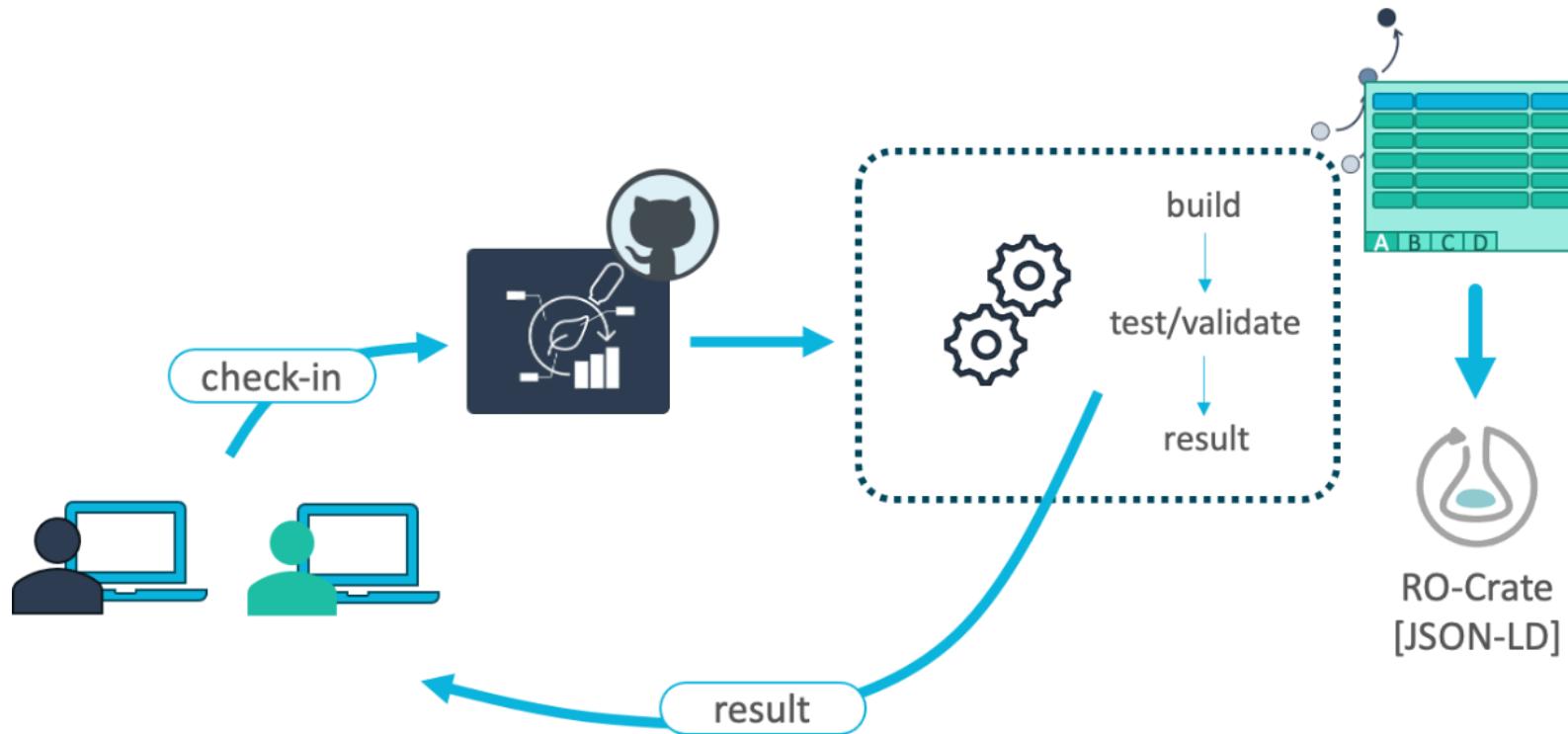


adapted from Weil, H.L., Schneider, K., et al. (2023), PLANTdataHUB: a collaborative platform for continuous FAIR data sharing in plant research. Plant J. <https://doi.org/10.1111/tpj.16474>

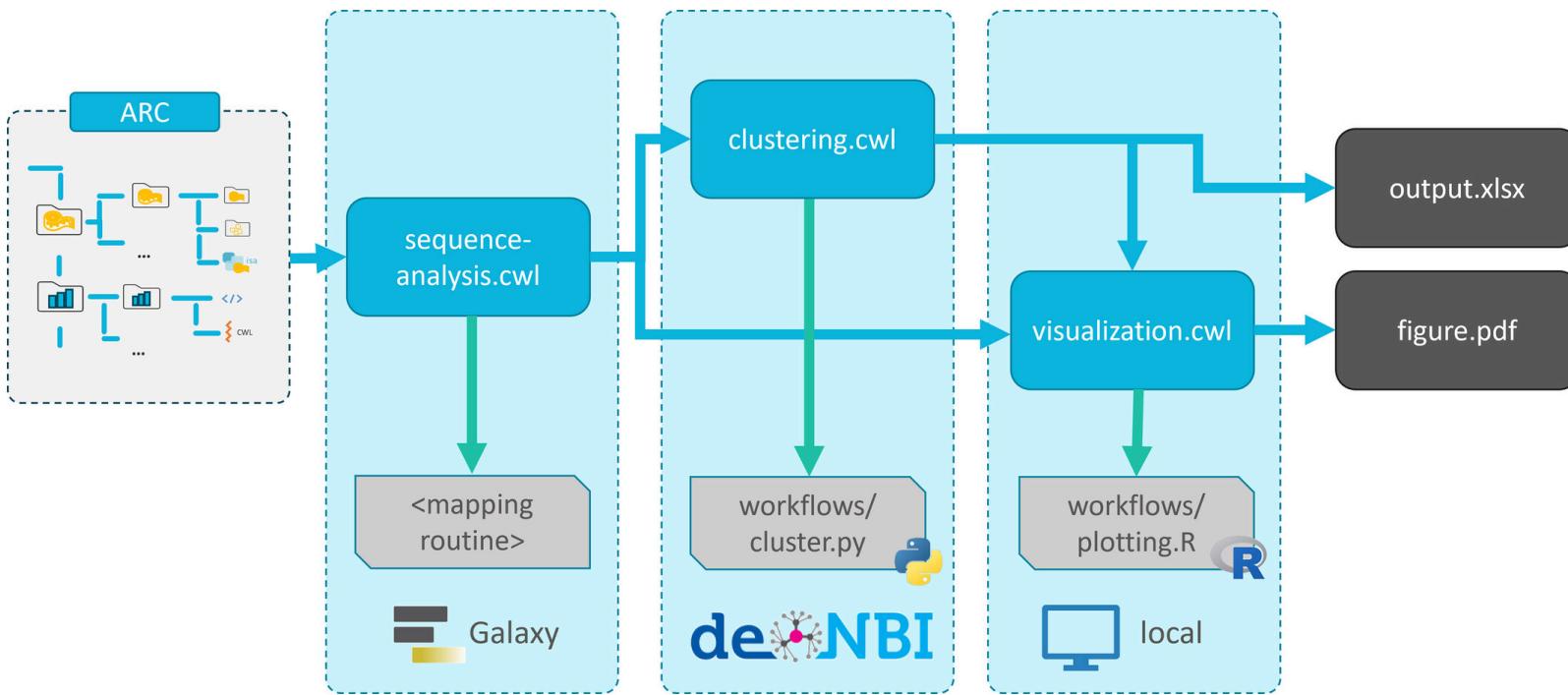
# Validate & publish



# Learning from Open-Source-Software development



# Data analysis and workflows



# Galaxy integration: Extra value for plant research

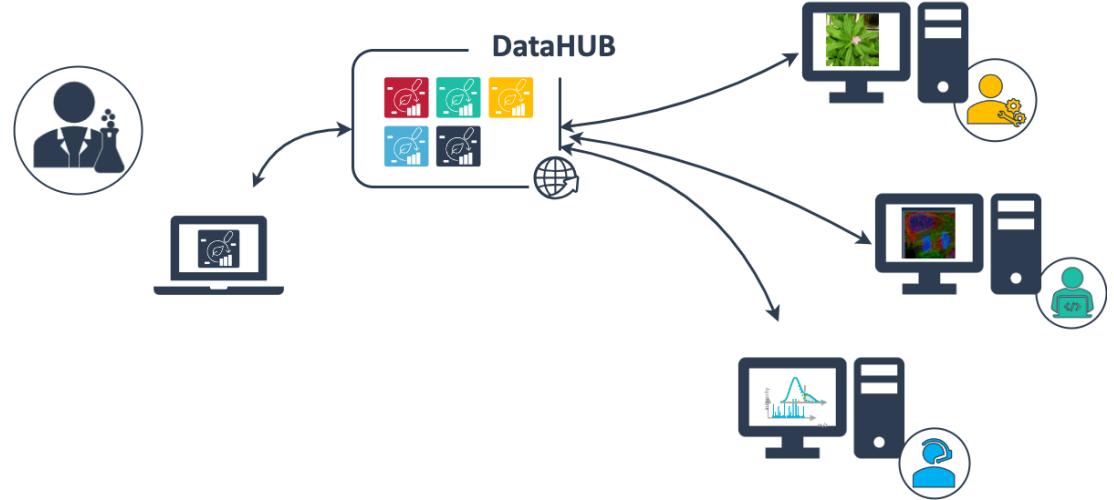


e.g. <https://plants.usegalaxy.eu>

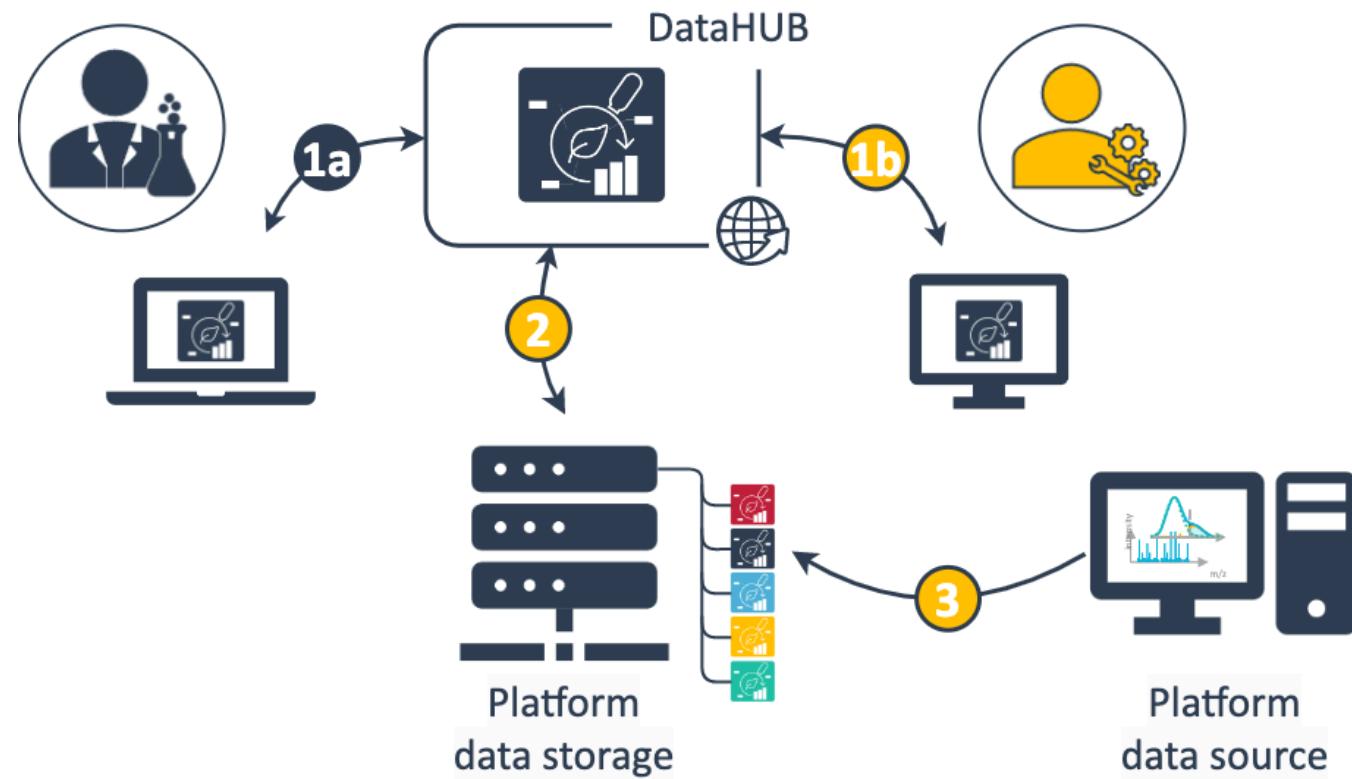
- Full ARC compatibility
- Automated metadata generation
- Specialized tools and workflows for ‘omics processing and analysis
- Public repository compatibility
- Galaxy teaching resource for data analysis

# Enabling platforms

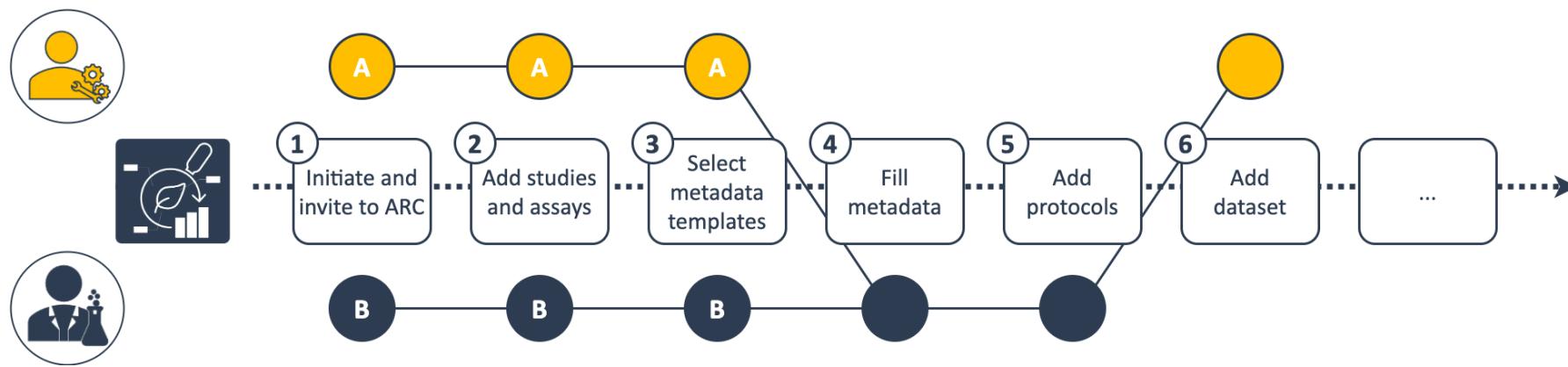
- Streamlined exchange of (meta)data
- Communication and project management



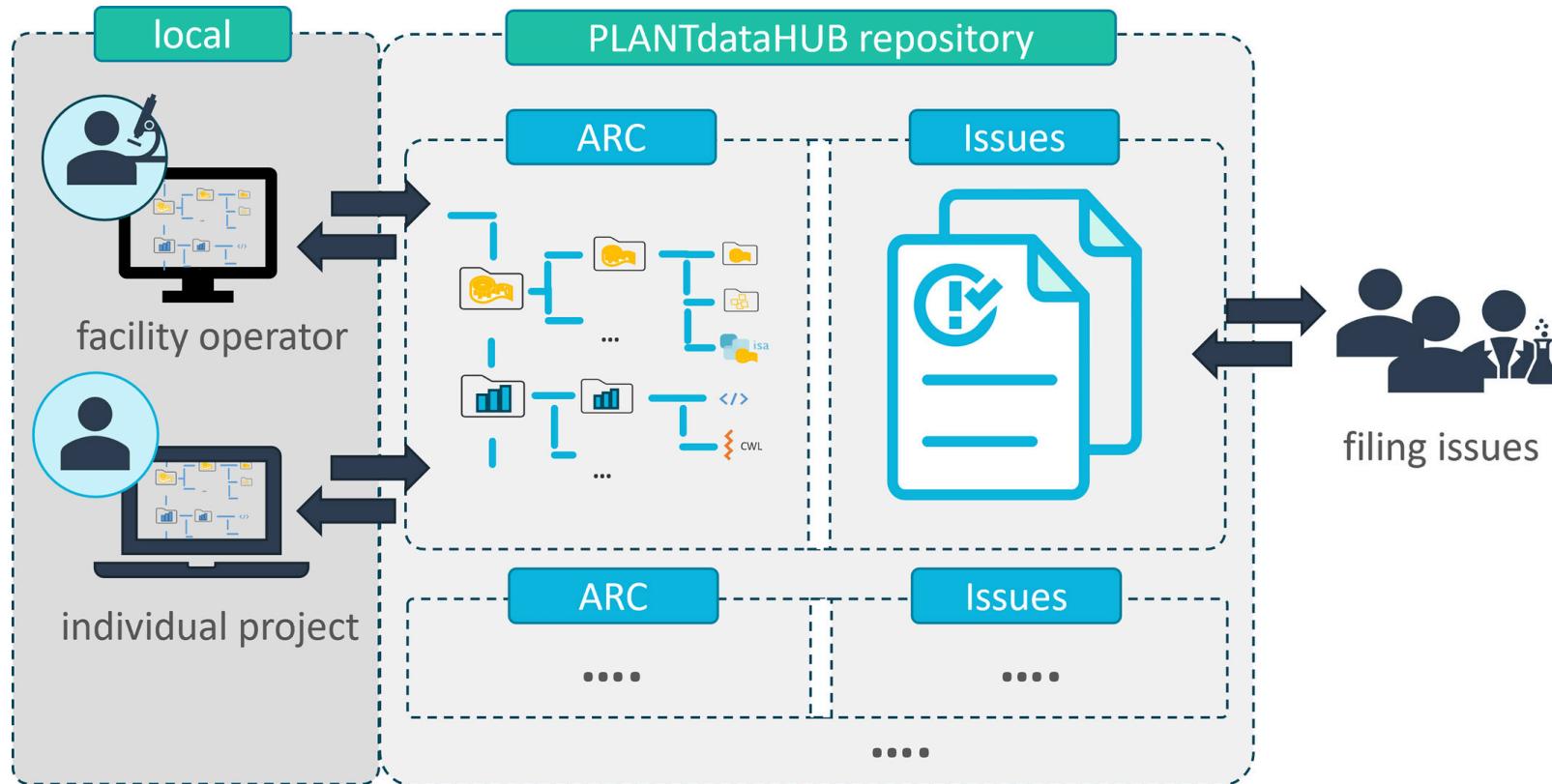
# Streamlined data exchange



# Meet your collaborators in an ARC

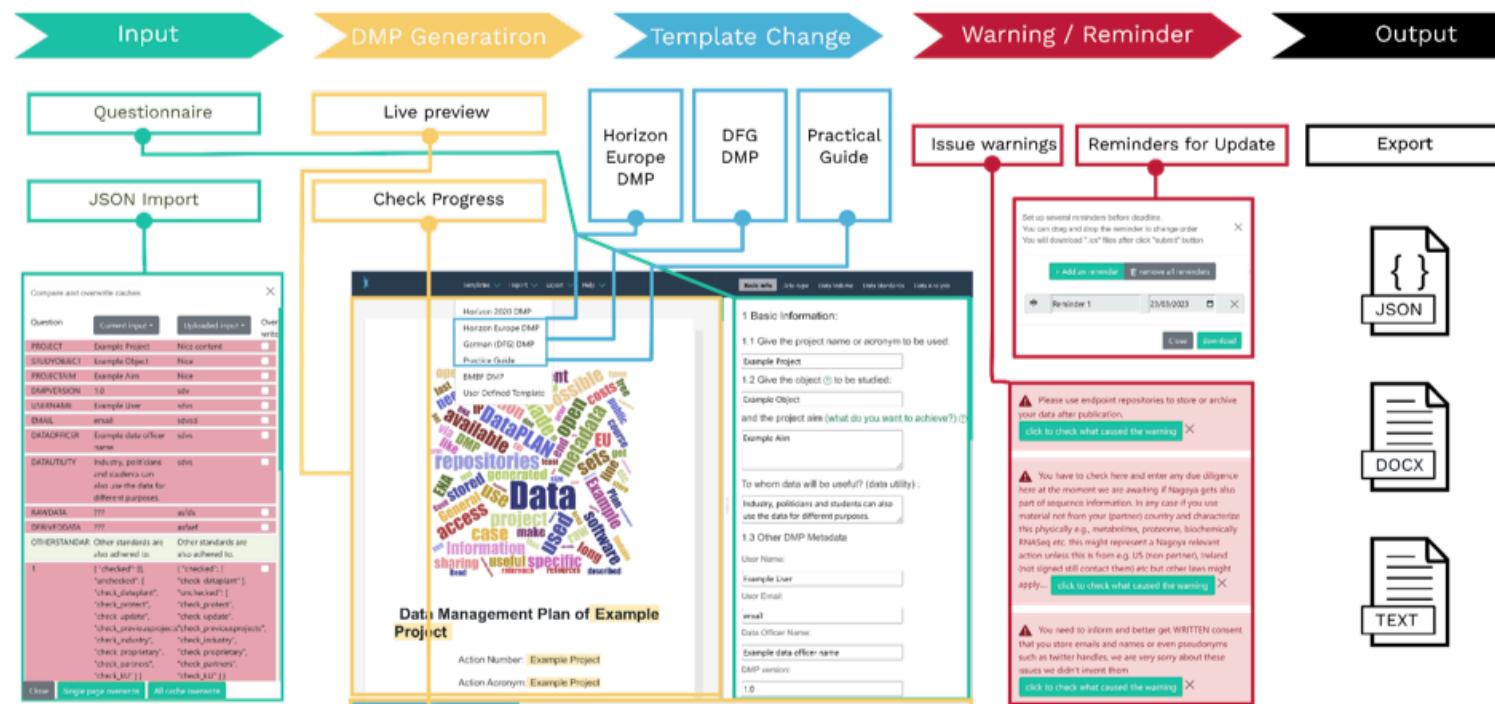


# Project management



# DataPLAN – a DataPLANT DMP generator

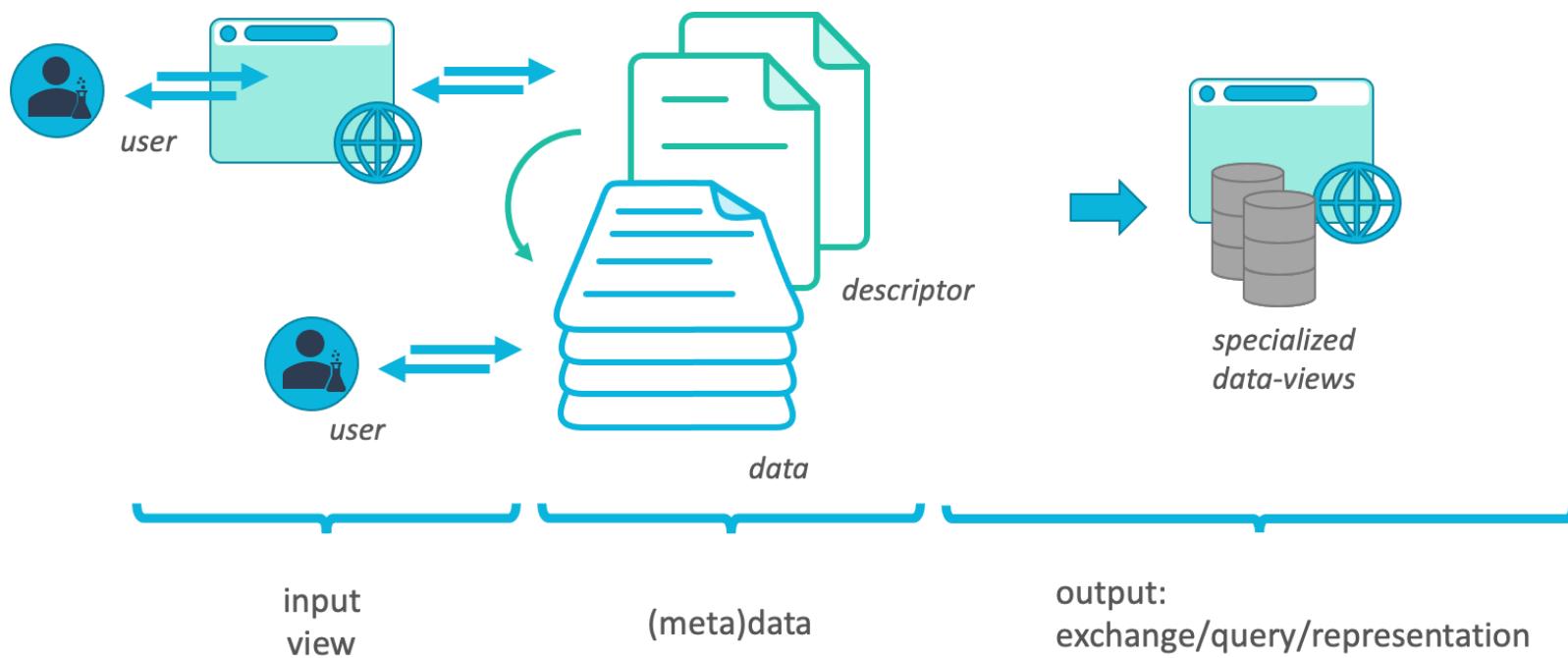
<https://dmpg.nfdi4plants.org>



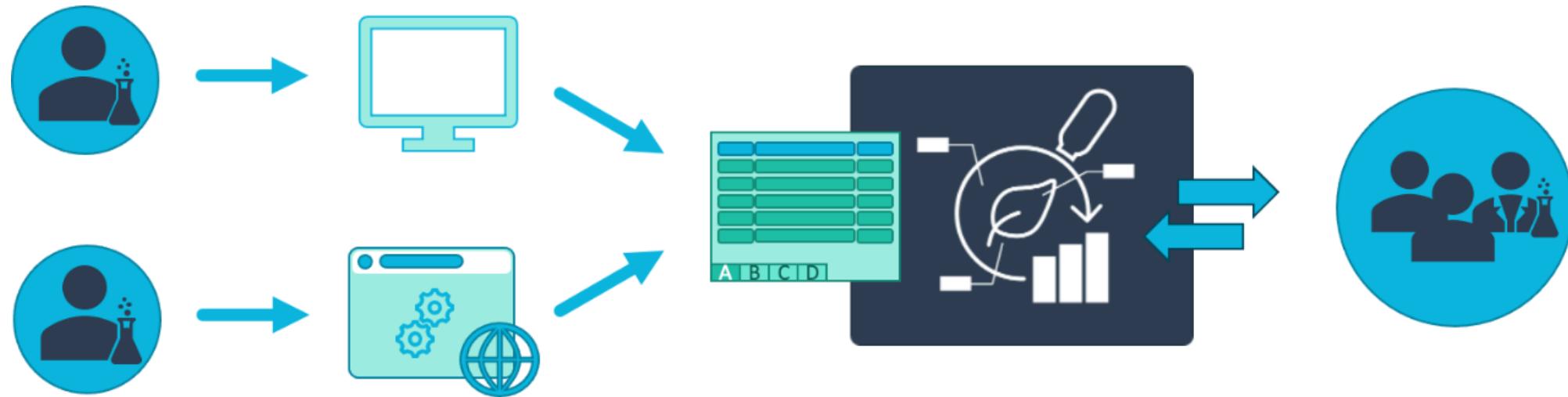
Zhou et al. (2023), DataPLAN: a web-based data management plan generator for the plant sciences, bioRxiv 2023.07.07.548147; doi: <https://doi.org/10.1101/2023.07.07.548147>

# Everything is a file

DataPLANT employs a **data-centric** approach to RDM

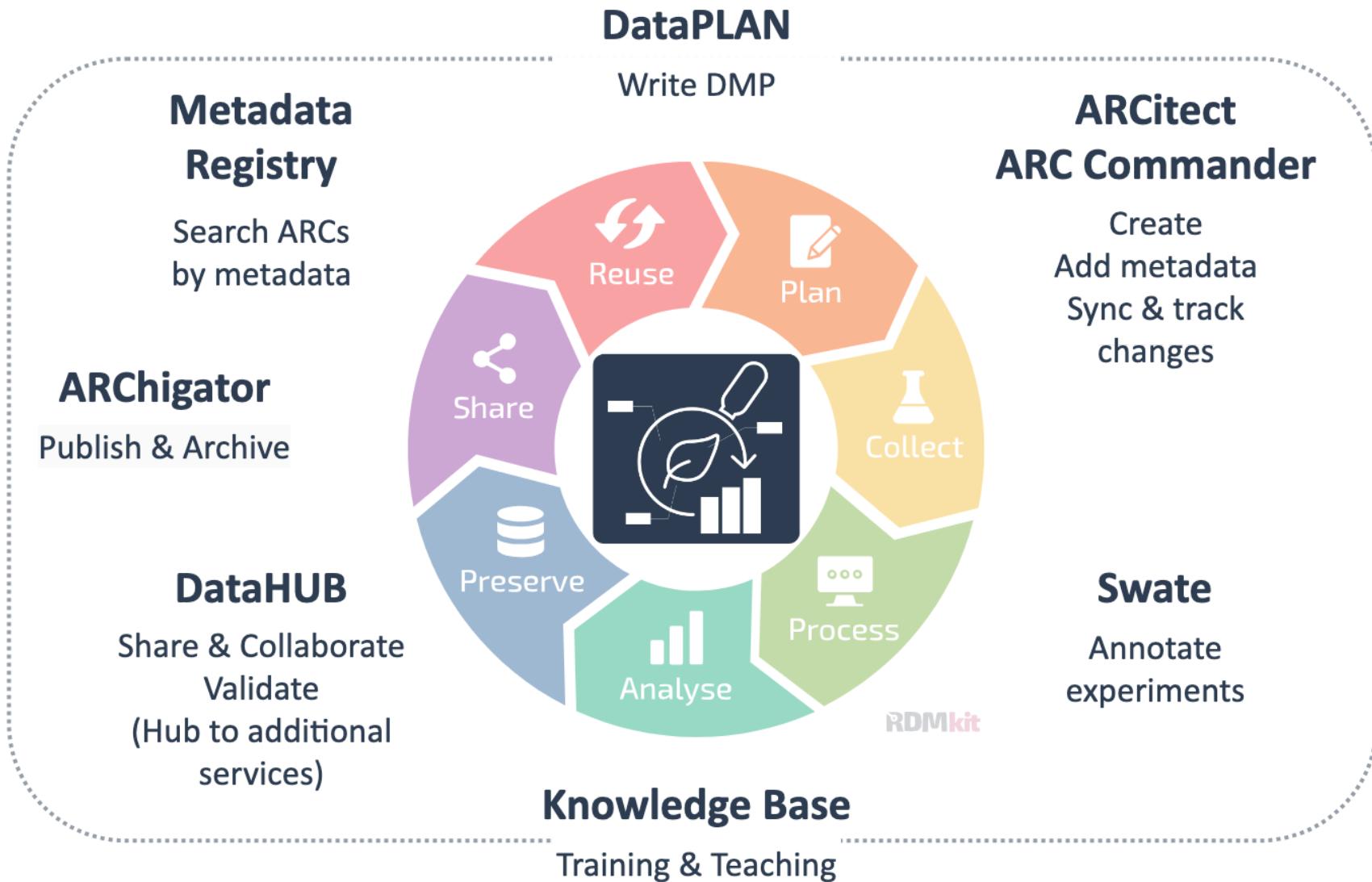


# No technical lock-in

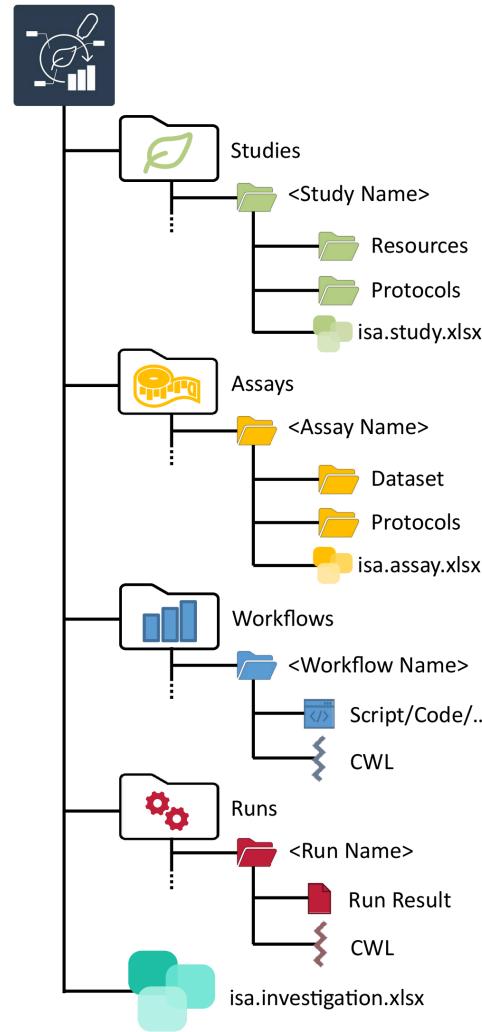


(Meta)data transparency with tool assistance but **no technical lock-in**

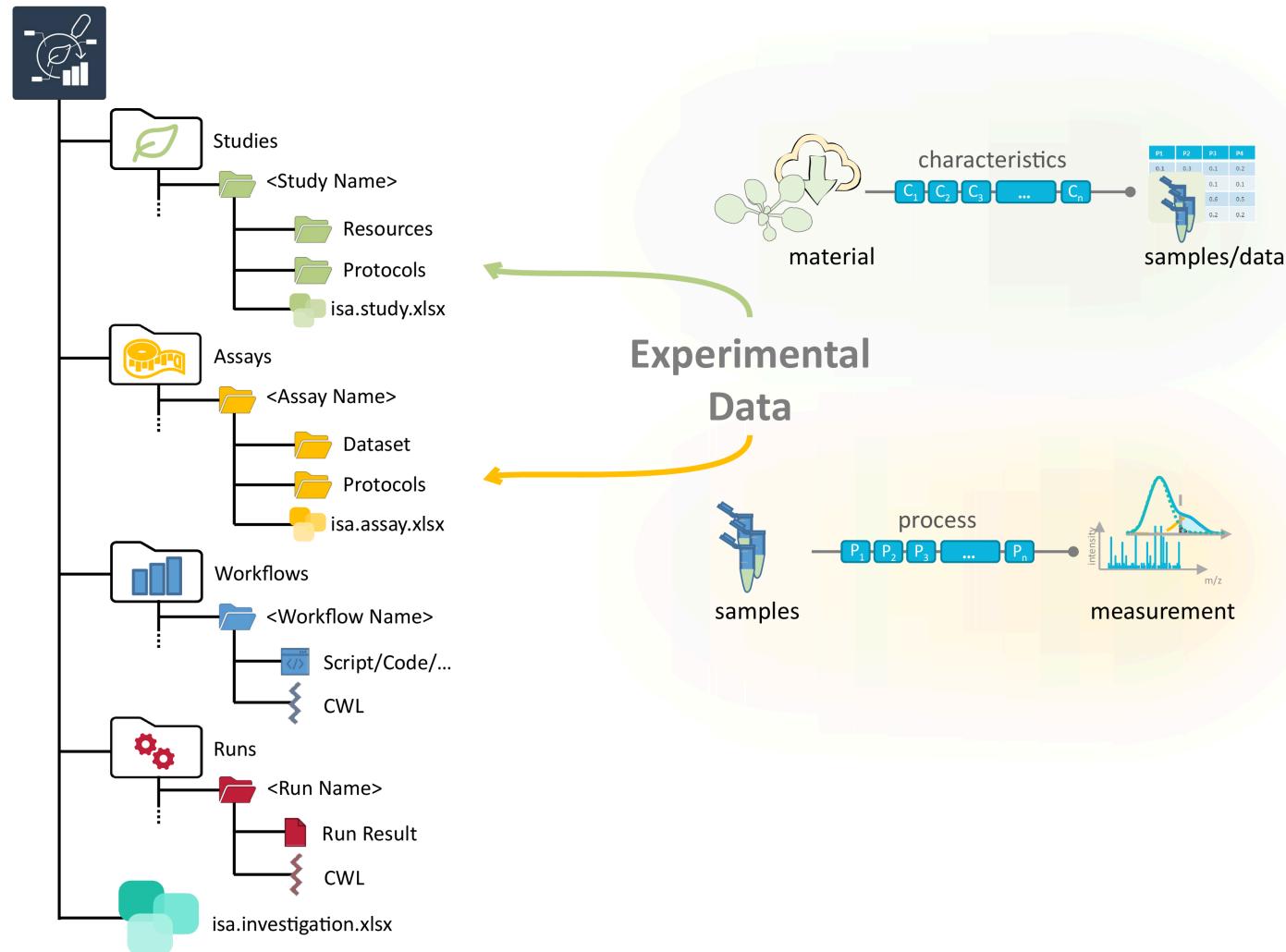
# The ARC ecosystem



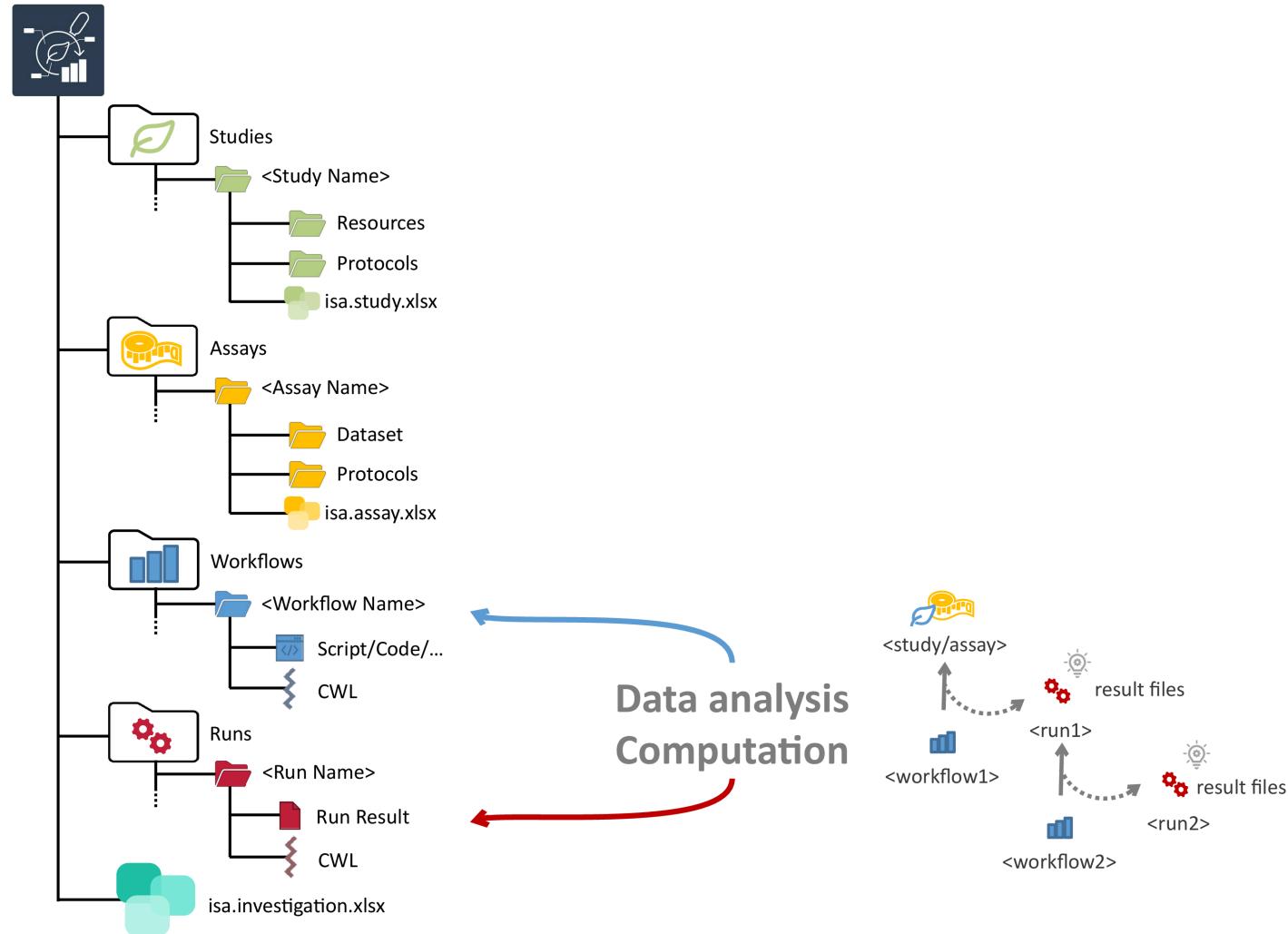
# What does an ARC look like?



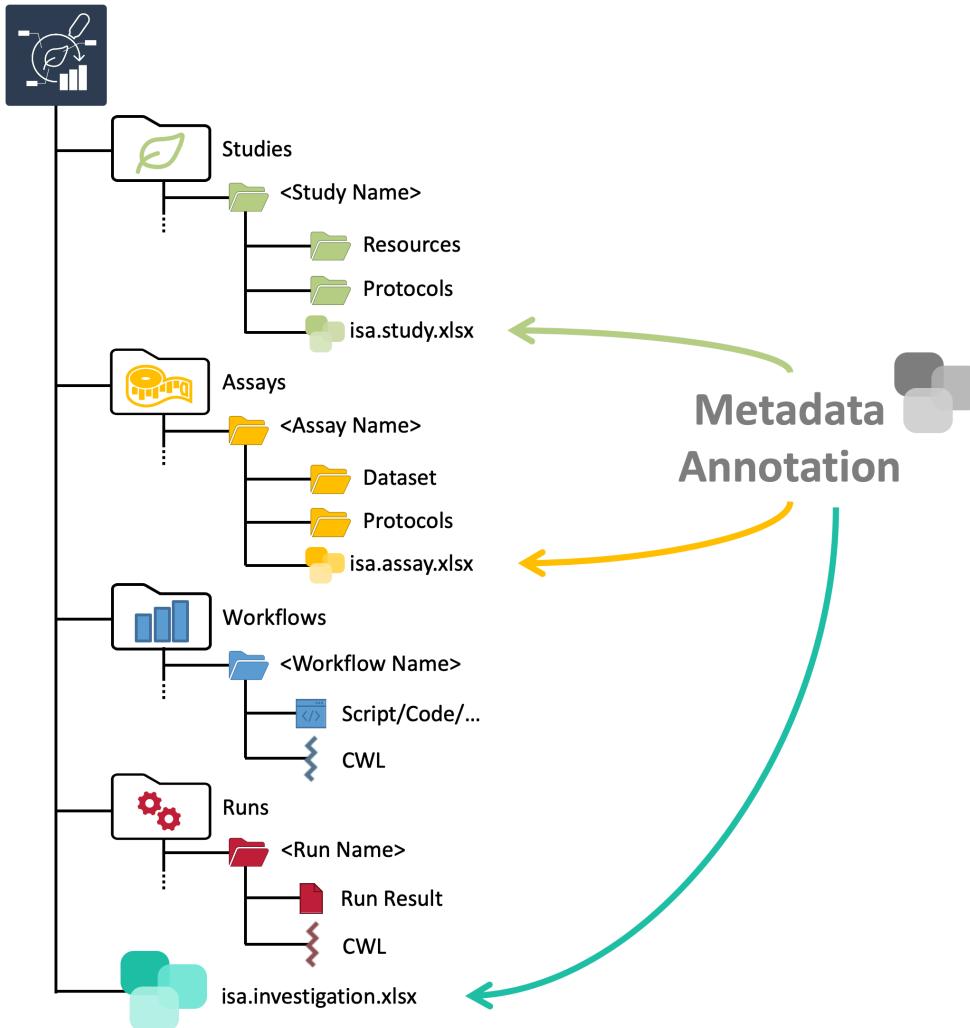
# ARCs store experimental data



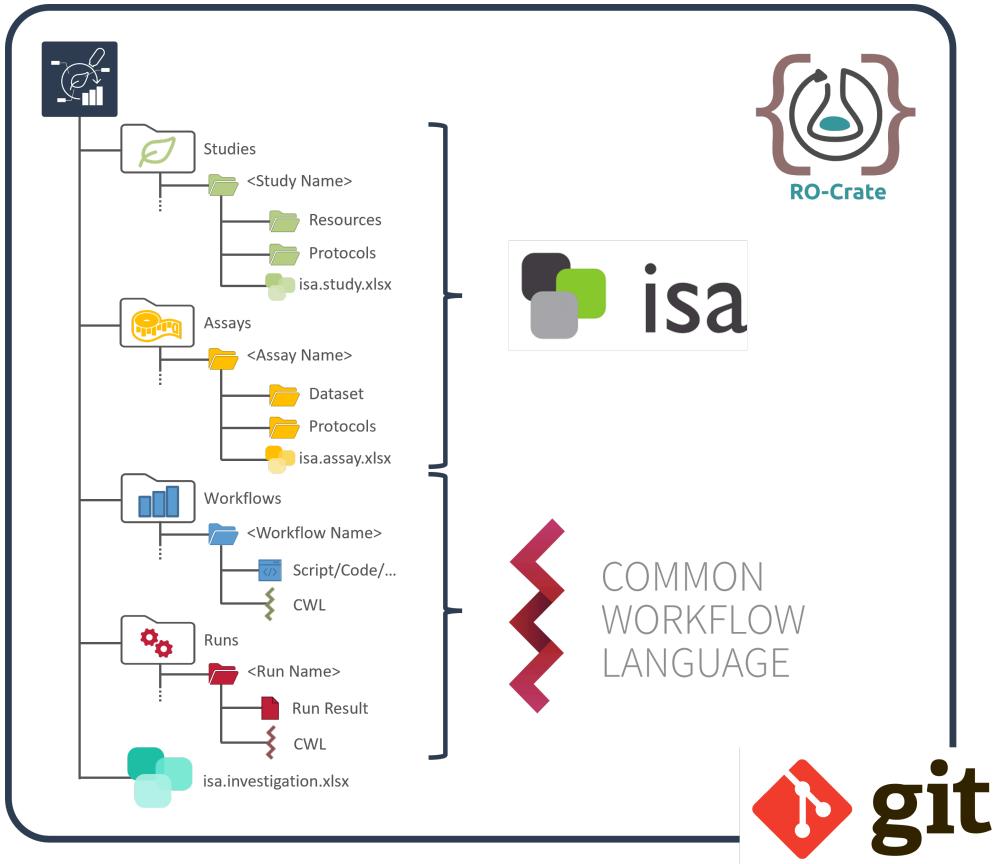
# Computations can be run inside ARCs



# ARCs come with comprehensive metadata

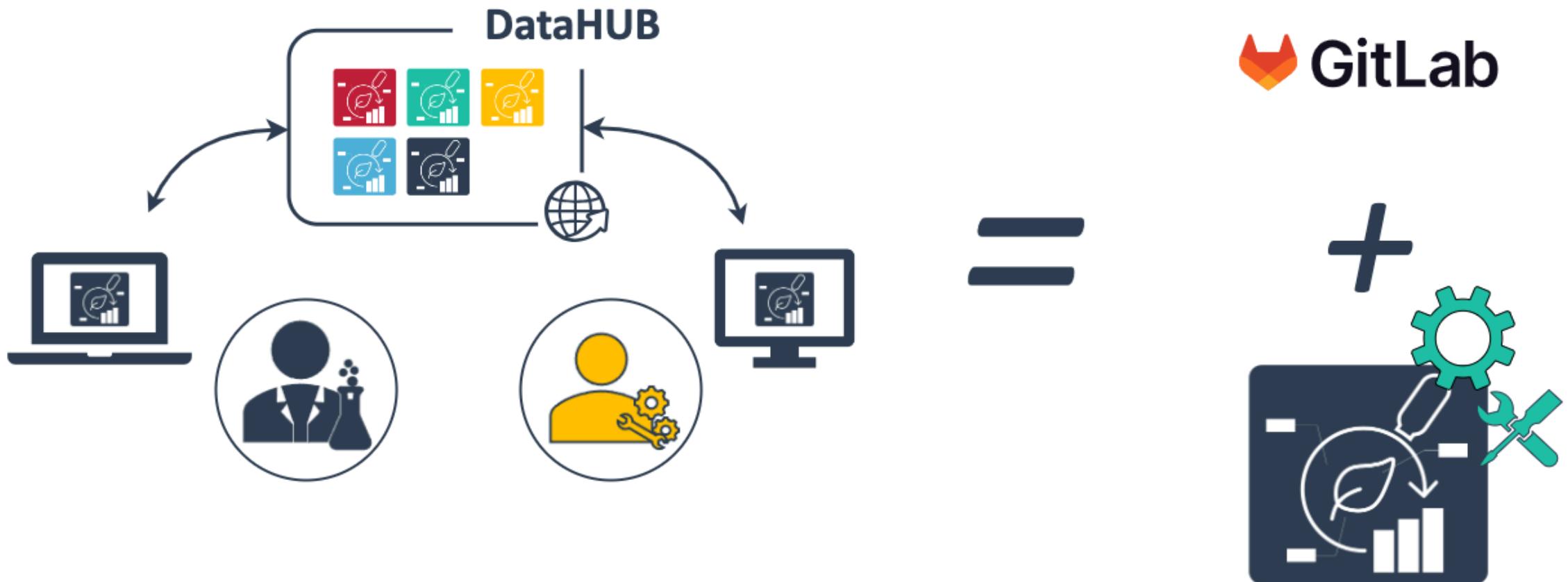


# ARC builds on standards

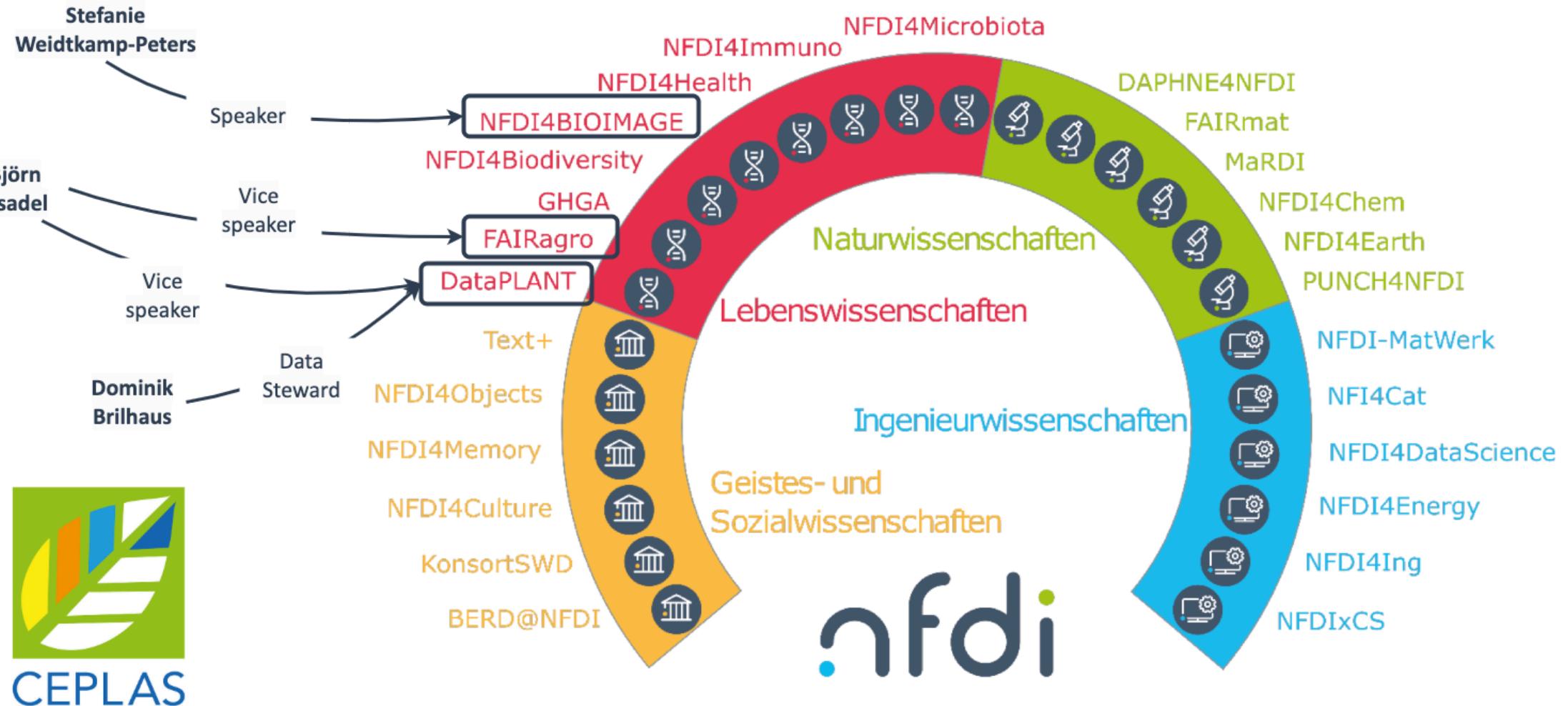


- RO-Crate: standardized exchange
- ISA: structured, machine-readable metadata
- CWL: reproducible, re-usable data analysis
- Git: version control

# The DataPLANT DataHUB – a GitLab *Plus*



# CEPLAS connection to the NFDI



# Data Stewardship between DataPLANT and the community

*Community*

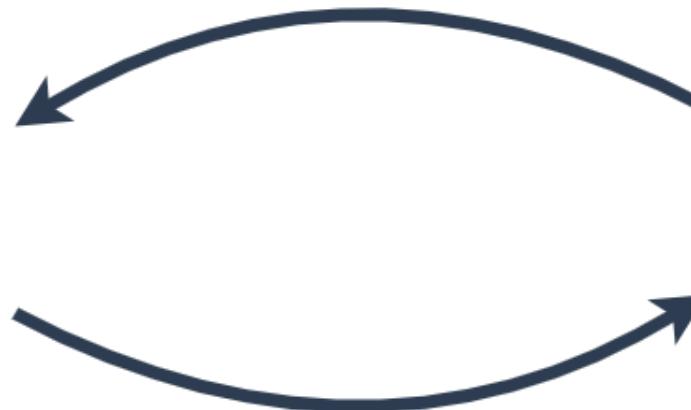


Domain experts  
User experience  
Training

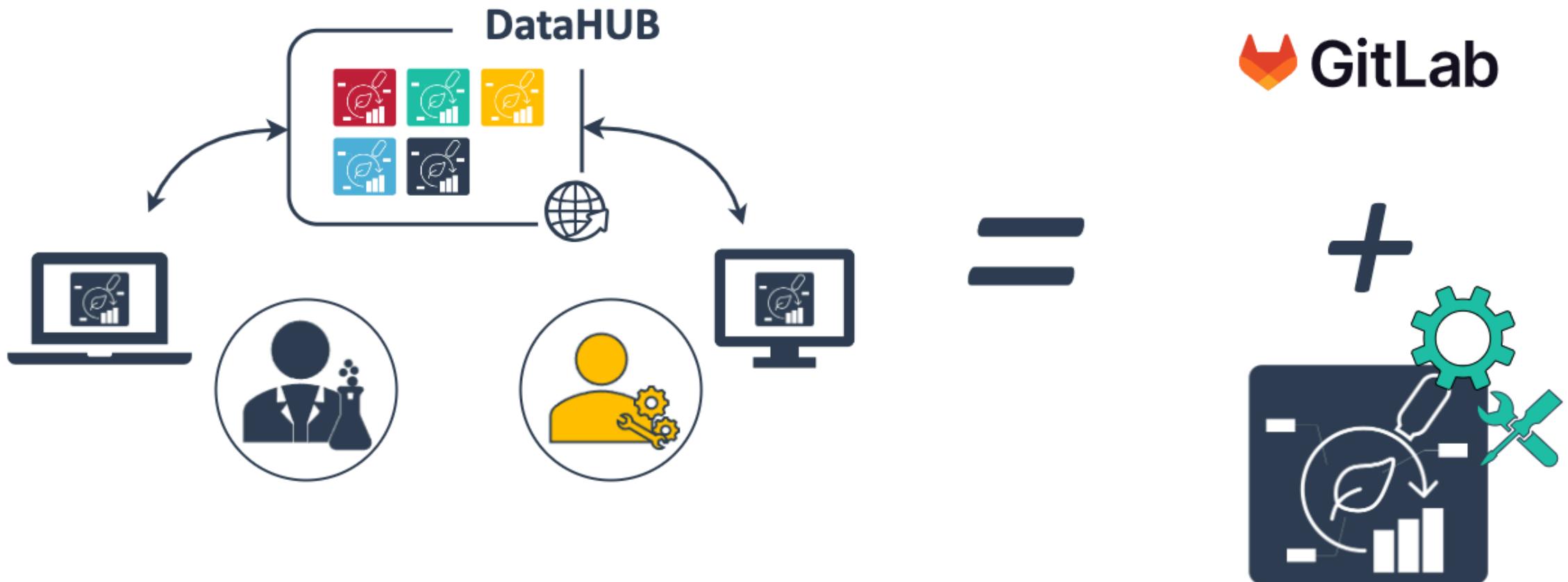
*nfdi4plants*



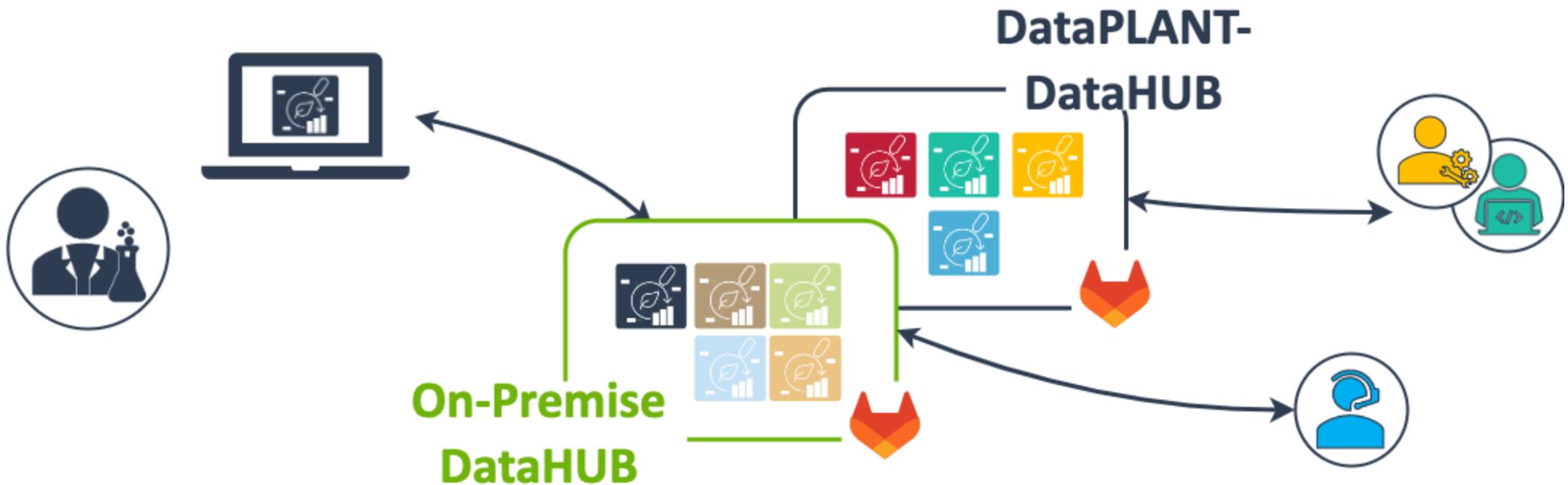
Service provider  
Developers  
Tech experts



# The DataPLANT DataHUB – a GitLab *Plus*

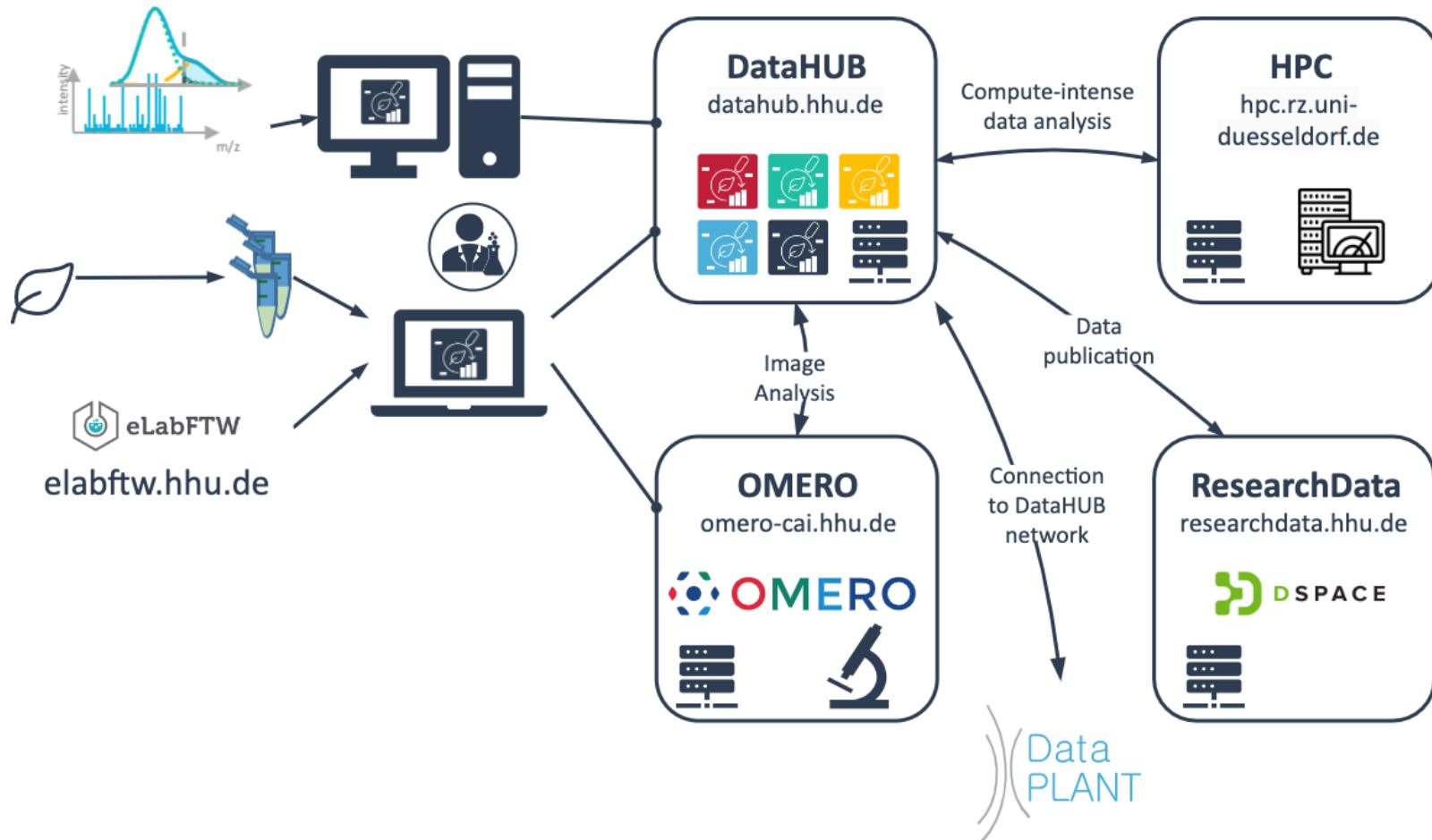


# On-premise DataHUBs



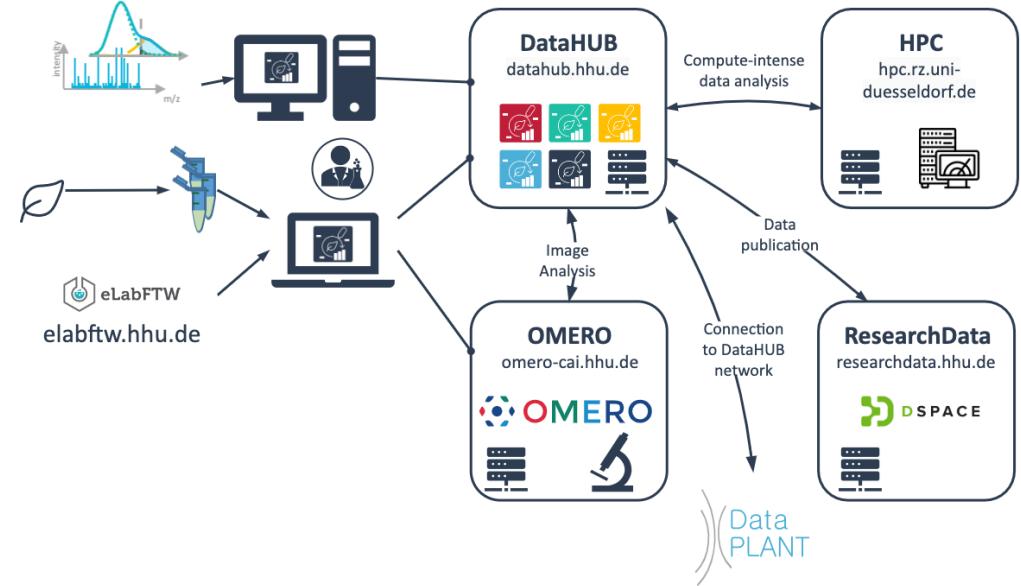
ARC services are available as on-premise option

# HHU-DataHUB

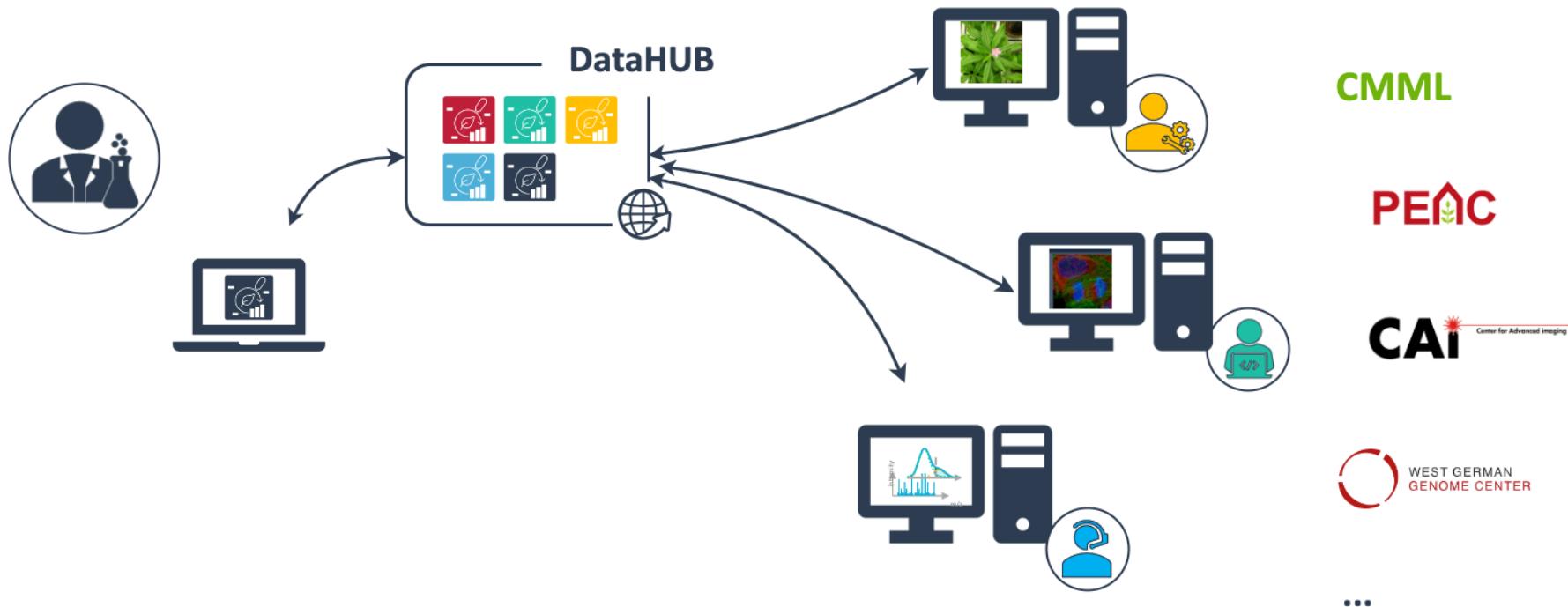


# HHU-DataHUB

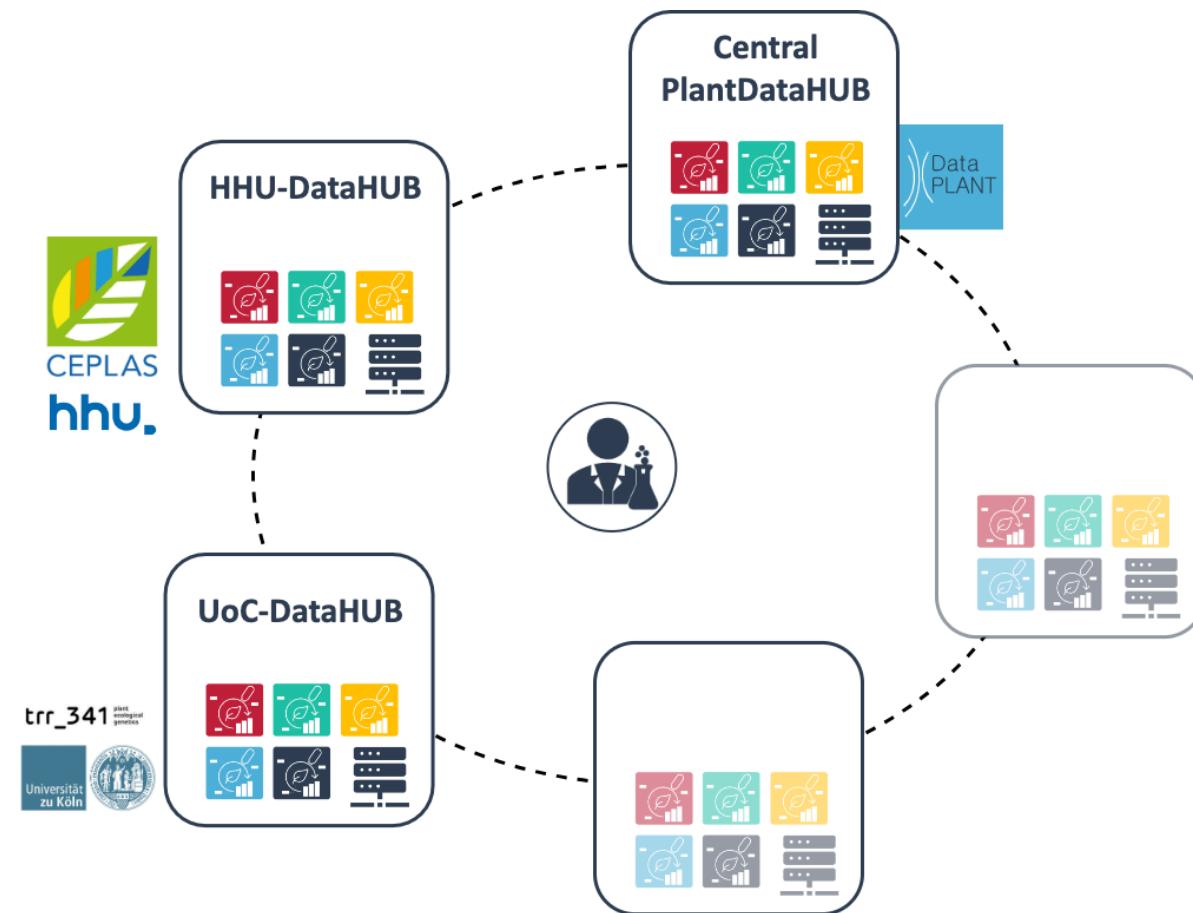
- FAIR data along the research data cycle
- Structured exchange of (meta)data
- Local (long-term) storage, backup, version control
- Integration of existing services and infrastructure
- External collaboration



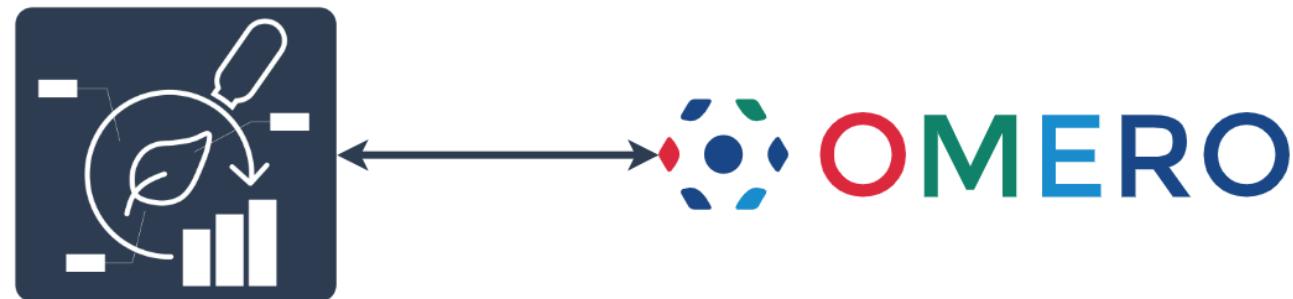
# HHU-DataHUB – Node for platforms



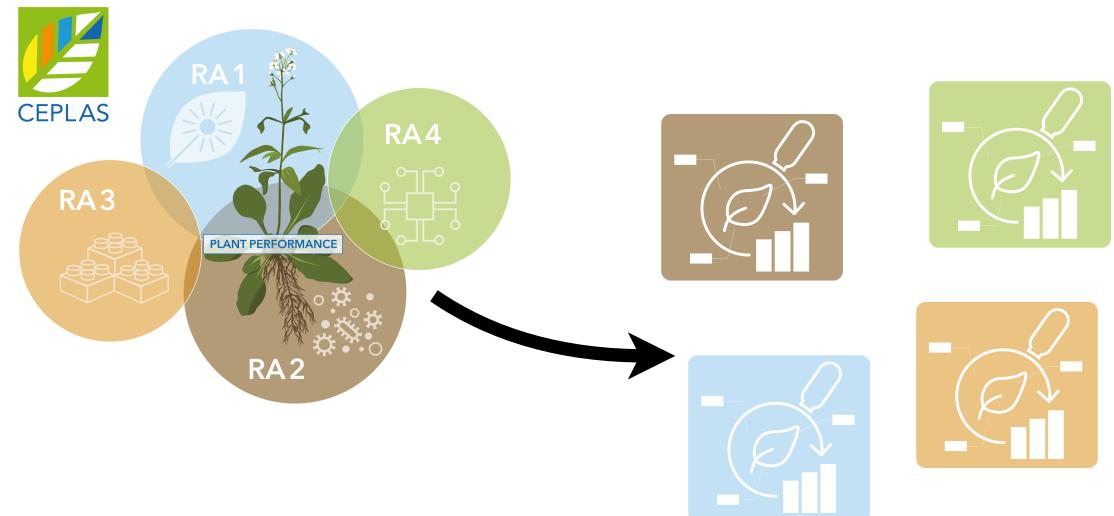
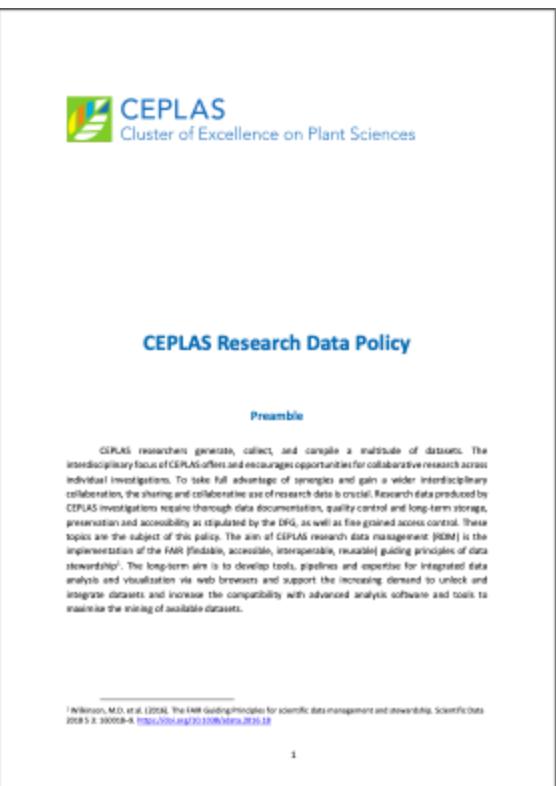
# Embedded into an RDM network



# CEPLAS collaborates with DataPLANT and NFDI4BIOIMAGE



# CEPLAS Research Data Policy



# ARCitect Hands-on

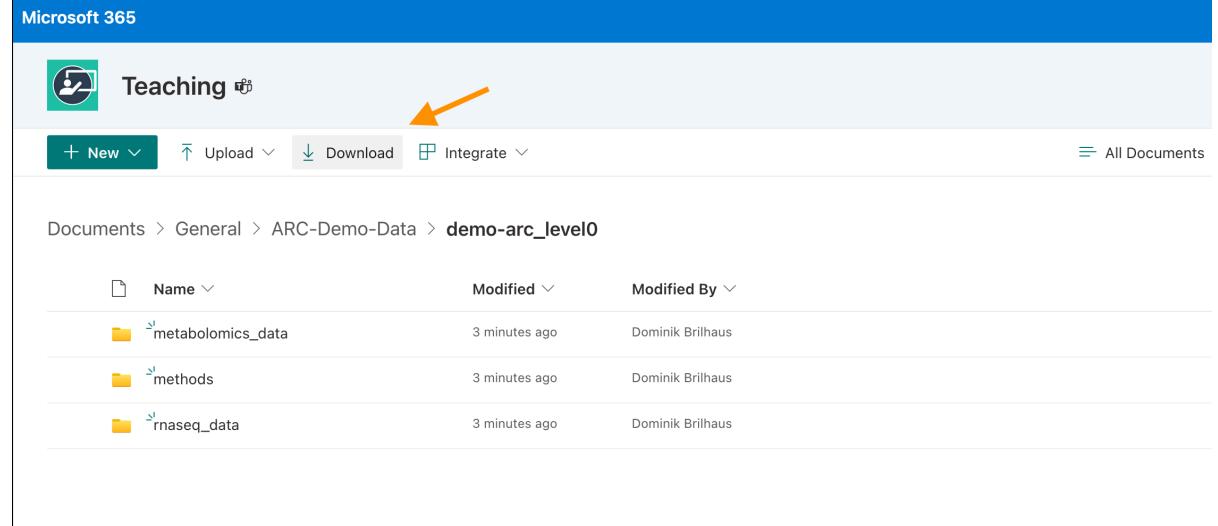
# ARCitect installation

Please install version **v0.0.31** (or later) of the ARCitect:

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

# Download the demo data

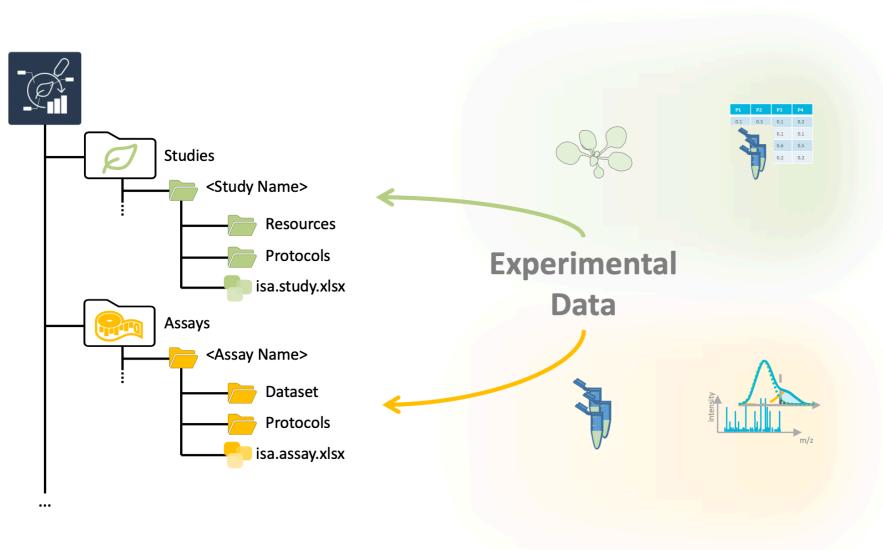
[https://nfdi4plant.sharepoint.com/:f/s/Teaching/Eik7koJiMREgZ24kt07sIYBGxHmmZIS\\_Kzf7psk-5w-xg?e=u0sADd](https://nfdi4plant.sharepoint.com/:f/s/Teaching/Eik7koJiMREgZ24kt07sIYBGxHmmZIS_Kzf7psk-5w-xg?e=u0sADd)



The screenshot shows a Microsoft 365 SharePoint interface. At the top, there's a blue header bar with the text "Microsoft 365". Below it is a navigation bar with icons for "New", "Upload", "Download" (which is highlighted with an orange arrow), and "Integrate". To the right of the navigation bar is a link to "All Documents". The main content area shows a file structure under "Documents > General > ARC-Demo-Data > demo-arc\_level0". There are three items listed: "metabolomics\_data", "methods", and "rnaseq\_data", all modified 3 minutes ago by Dominik Brilhaus.

Name	Modified	Modified By
metabolomics_data	3 minutes ago	Dominik Brilhaus
methods	3 minutes ago	Dominik Brilhaus
rnaseq_data	3 minutes ago	Dominik Brilhaus

# 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

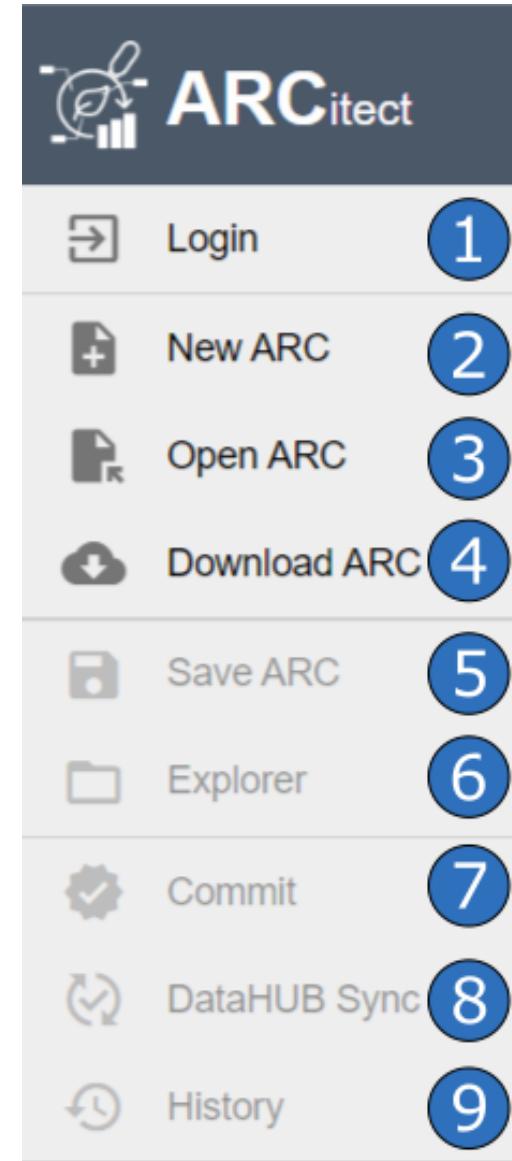
# Open ARCitect

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

Please Select a DataHub

Host  
git.nfdi4plants.org

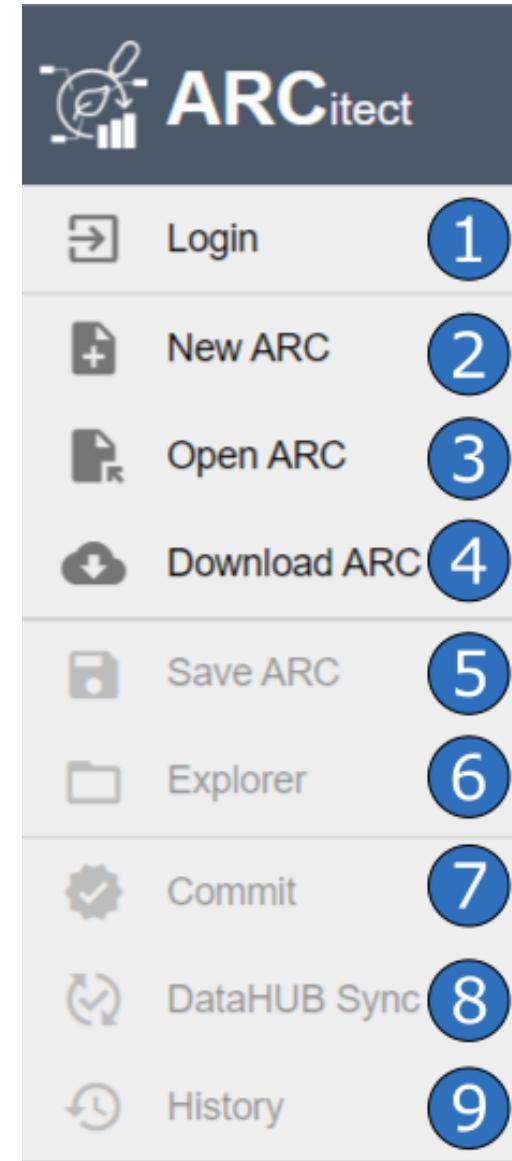
LOGIN CANCEL



# Initiate the ARC folder structure

1. Create a **New ARC** (2)
2. 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.
-  [TalinumPhotosynthesis](#)  
- ► assays  
- ► runs  
- ► studies  
- ► workflows

# 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 a contributor

- you could also add the contacts directly via ORCID

**Contacts**

Your First Name Your Last Name 6/10

Your ORCID

**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 X

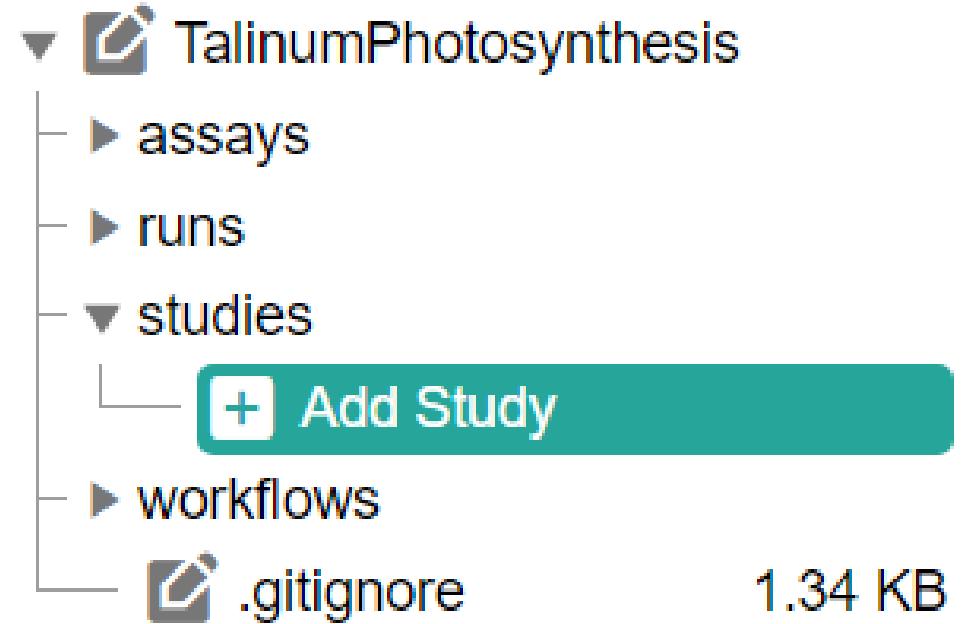
+

Delete

# 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 --:--

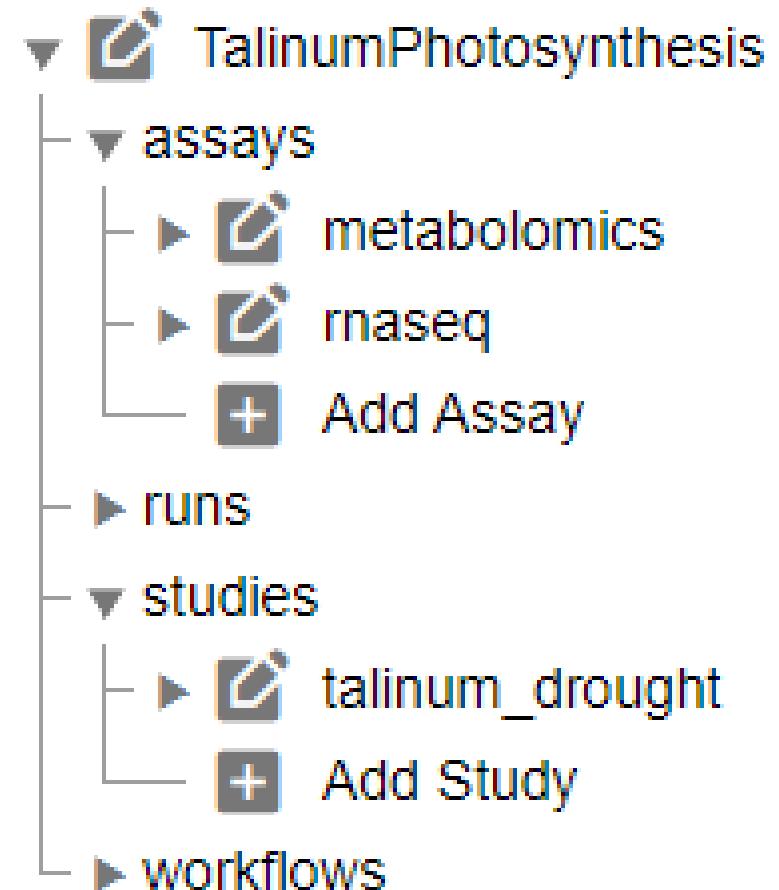
Public Release Date  
tt.mm.jjjj --:--

Study Design Descriptors

# Add an assay

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

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



# Link your assay to a study

You can either

- link your new assay to an existing study in your ARC or
- create a new one

Link your assays to your  
**talinum\_drought** study

Add Assay

Add Assay  
rnaseq

Study Identifiers  
talinum\_drought 

talinum\_drought

 ADD ASSAY CANCEL

# Add information about your assay

In the assay panel you can define the assay's

- measurement type
- technology type, and
- technology platform

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

# 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

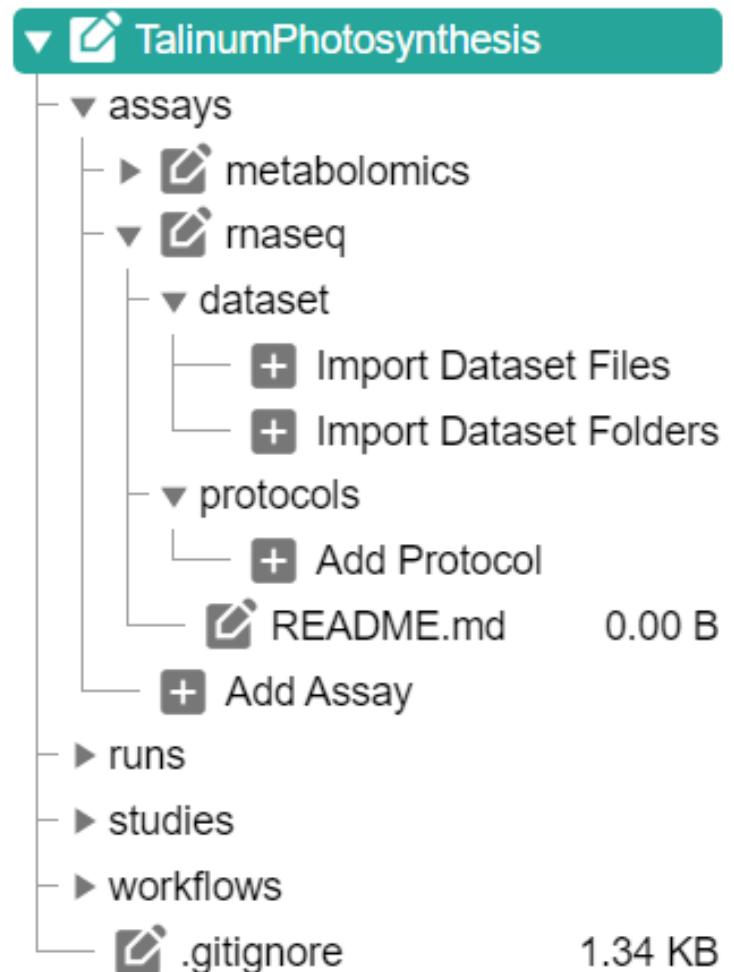
# Add protocols and datasets

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.

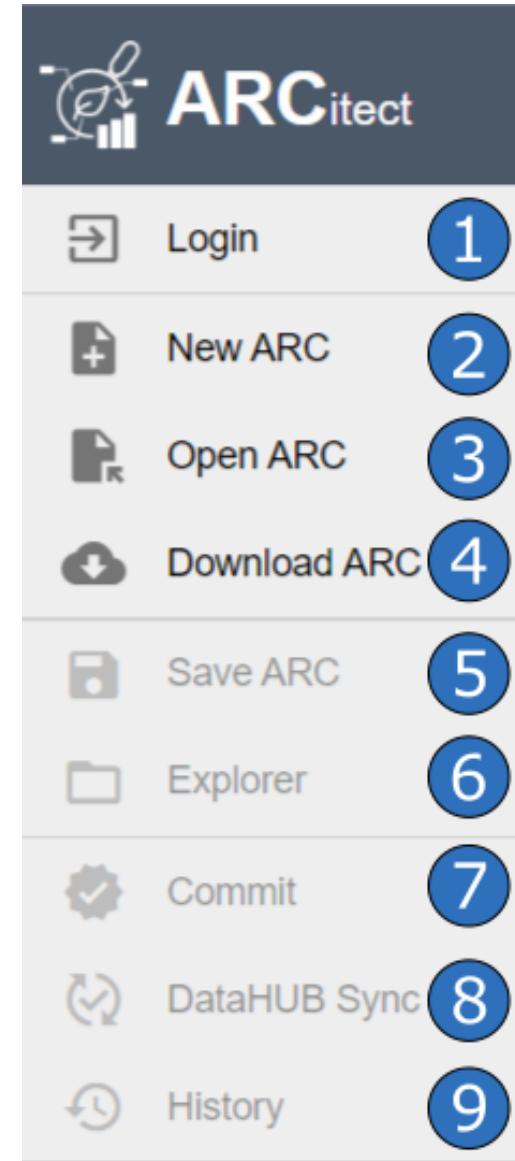


# Sort Demo Data to your ARC

 protocols and multiple datasets can directly be imported via ARCitect

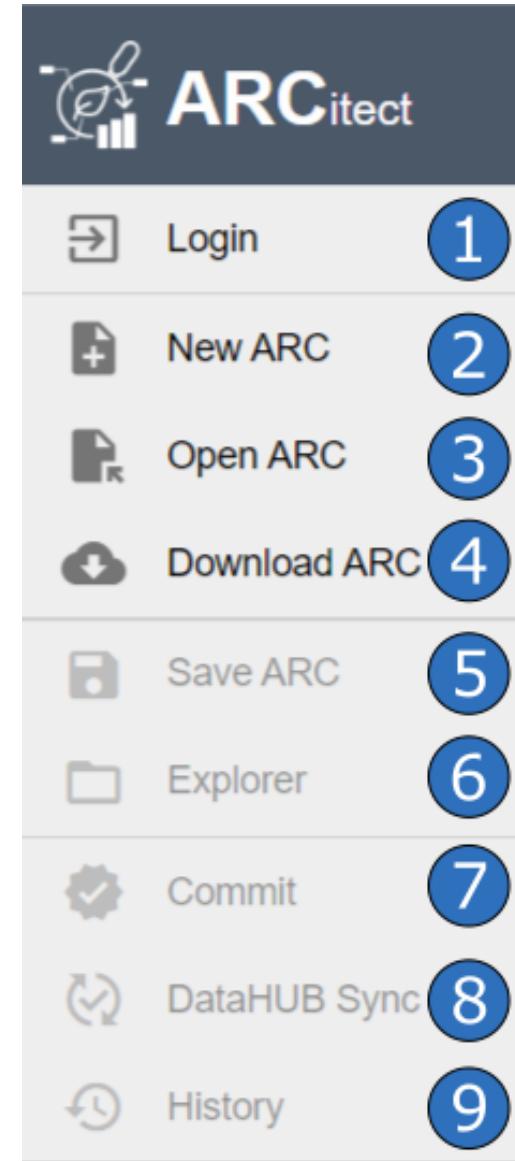
# Explorer

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



## Commit panel (7)

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



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

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

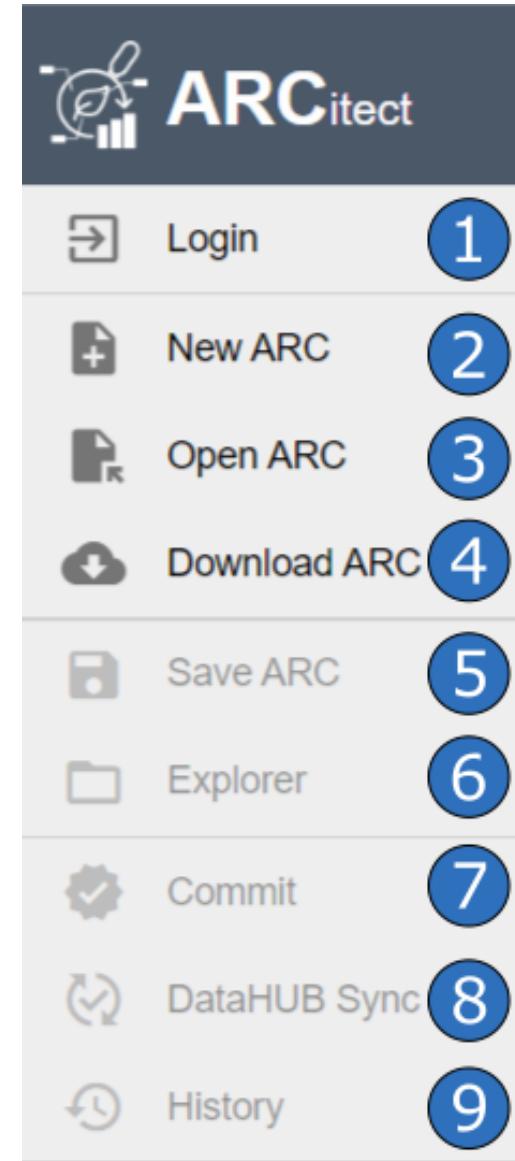
It allows you to

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

The screenshot shows the DataHUB Commit panel. At the top, there is a header with a gear icon and the text "Commit Changes" and "Track changes of the ARC with git". Below this, there are input fields for "Name" (Sabrina Zander) and "eMail" (sabrina.zander@uni-duesseldorf.de). There are also dropdowns for "Branch" and "Commit Message", and a field for "Large File Storage Limit in MB" set to 1. A horizontal line separates this from the "Changes" section. The "Changes" section shows a green trash bin icon and the text "No changes to commit". At the bottom right are two buttons: a teal "RESET" button with a circular arrow icon and a teal "COMMIT" button with a checkmark icon.

# Upload your local ARC to the DataHUB

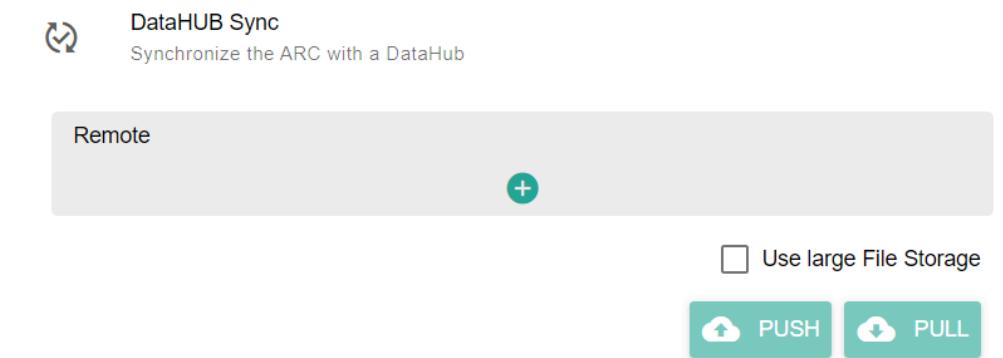
From the sidebar, navigate to **DataHUB Sync** (8)



# 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

# Your ARC is ready

 Initiated an ARC

 Structured and ...

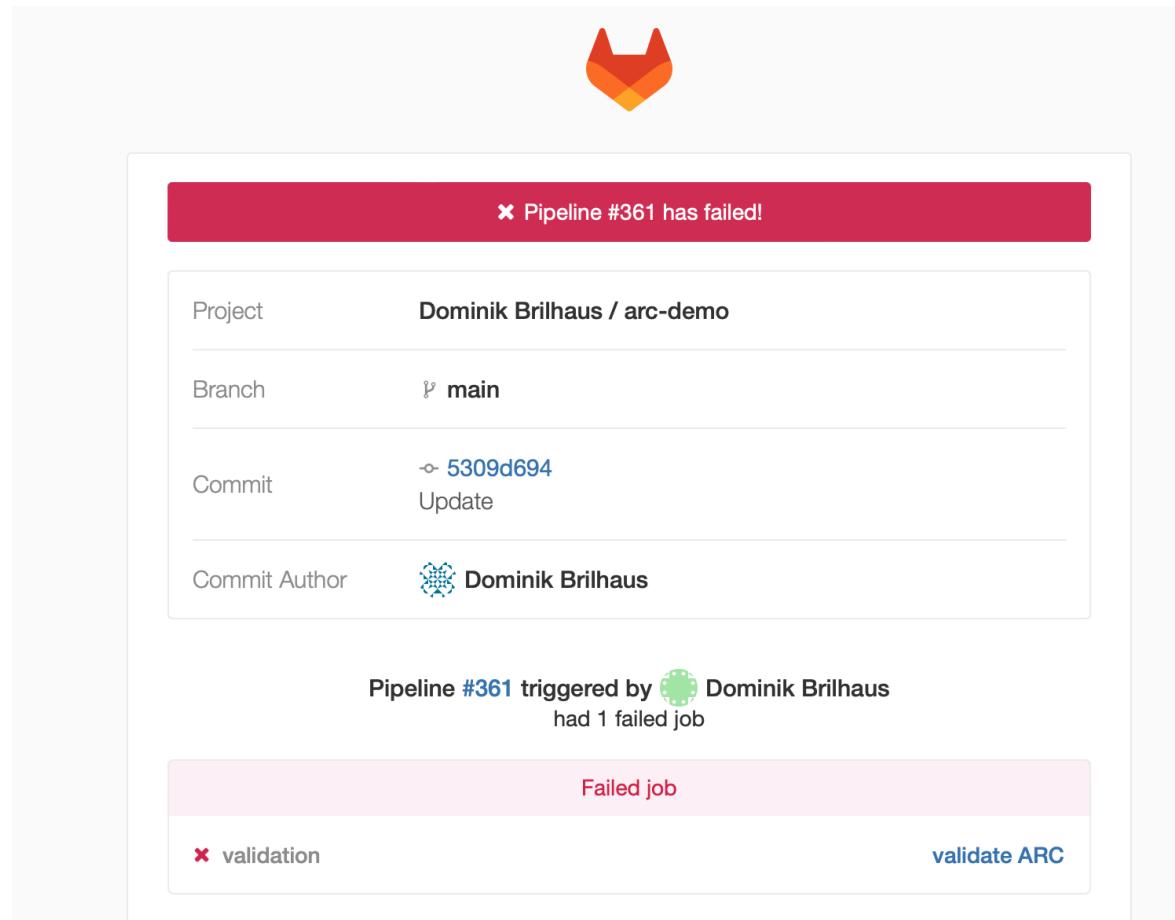
 ... annotated experimental data

 Shared with collaborators





# Received two emails from "GitLab" about a failed pipeline?

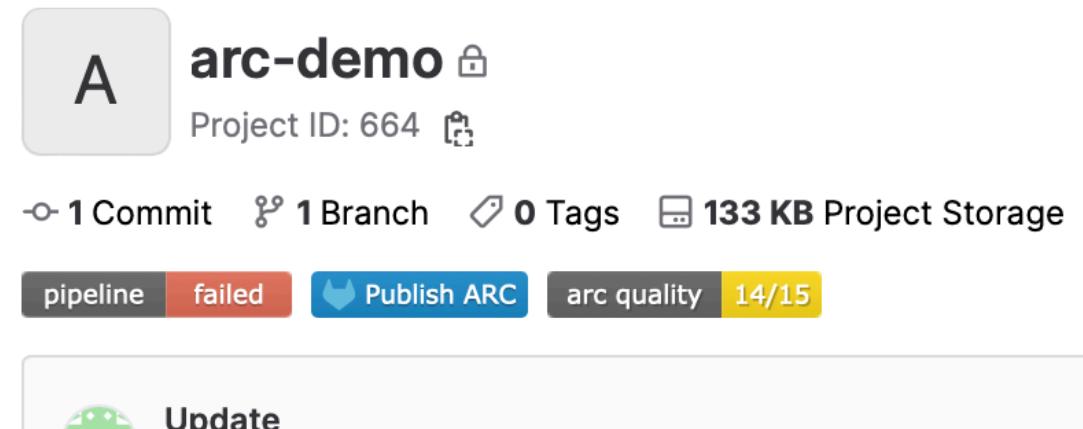


🔥 Don't worry 😊

# Pipeline Failed

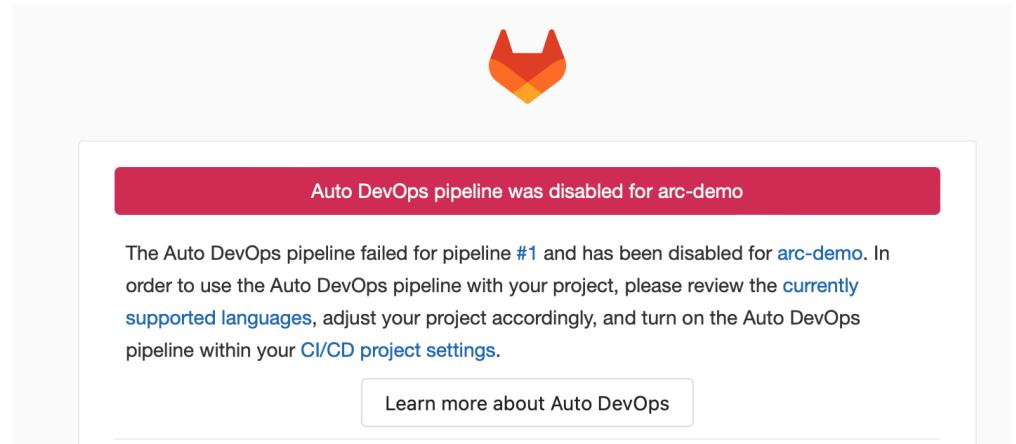
- a "continuous quality control" (CQC) pipeline validates your ARC
- This fails if one of the following metadata items is missing:

```
Investigation Identifier
Investigation Title
Investigation Description
Investigation Person Last Name
Investigation Person First Name
Investigation Person Email
Investigation Person Affiliation
```



# Pipeline Failed

If the pipeline has failed once, it is disabled by default



The screenshot shows a message about a disabled Auto DevOps pipeline. At the top right is a small orange and yellow logo. Below it, a red bar contains the text "Auto DevOps pipeline was disabled for arc-demo". The main message area is white with black text, stating: "The Auto DevOps pipeline failed for pipeline #1 and has been disabled for [arc-demo](#). In order to use the Auto DevOps pipeline with your project, please review the [currently supported languages](#), adjust your project accordingly, and turn on the Auto DevOps pipeline within your [CI/CD project settings](#)." At the bottom right of the message area is a button labeled "Learn more about Auto DevOps".

# Reactivate the CQC pipeline

To reactivate it and let the DataHUB validate your ARC again:

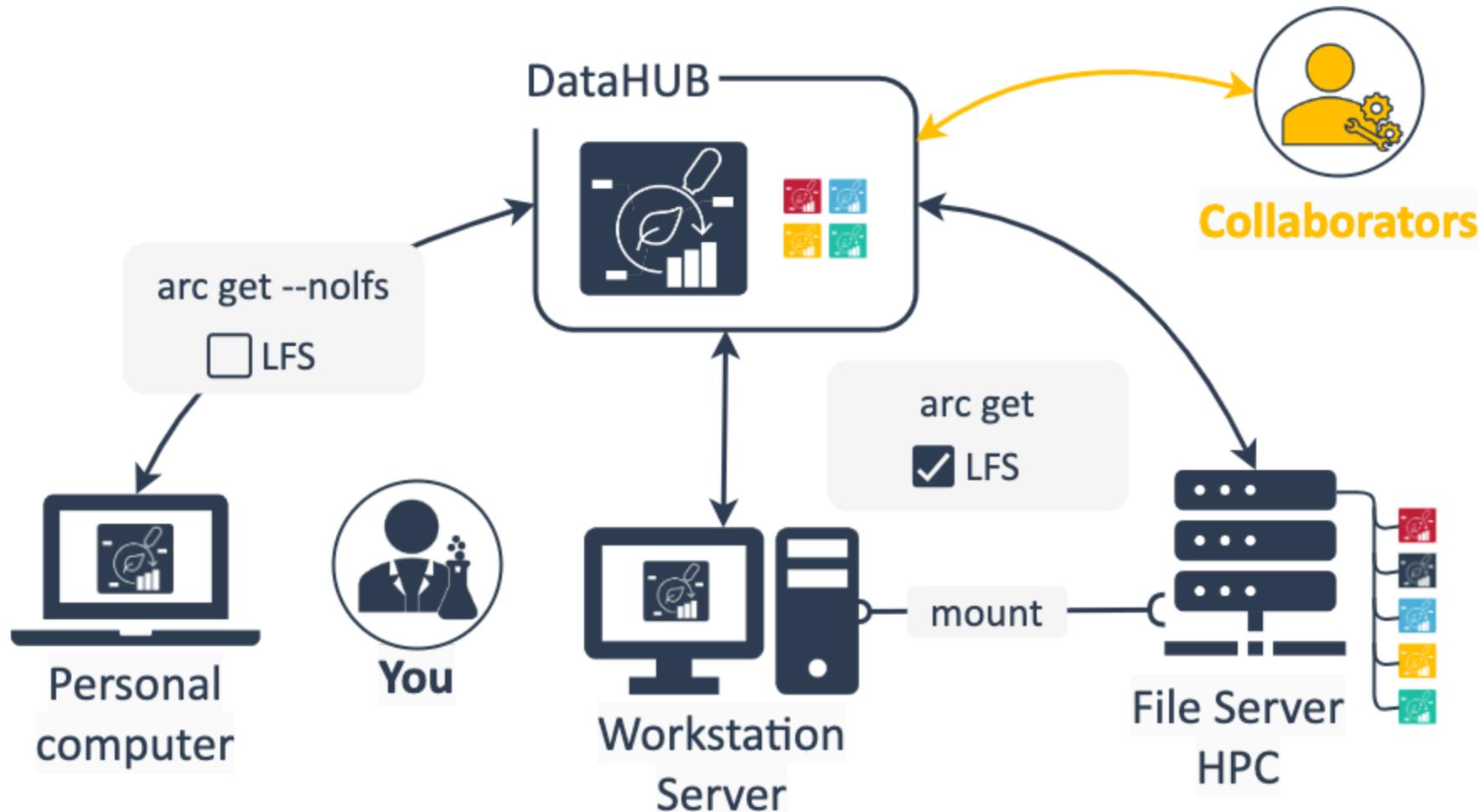
1. navigate to CI/CD setting `<arc-url>/-/settings/ci_cd`
2. expand "Auto DevOps"
3. check box "Default to Auto DevOps pipeline"
4. Save changes

The screenshot shows the GitLab CI/CD settings interface. On the left, there is a sidebar with various project management and development tools. The 'CI/CD' option is highlighted, indicating the current section. The main content area is titled 'Auto DevOps' and contains the following information:

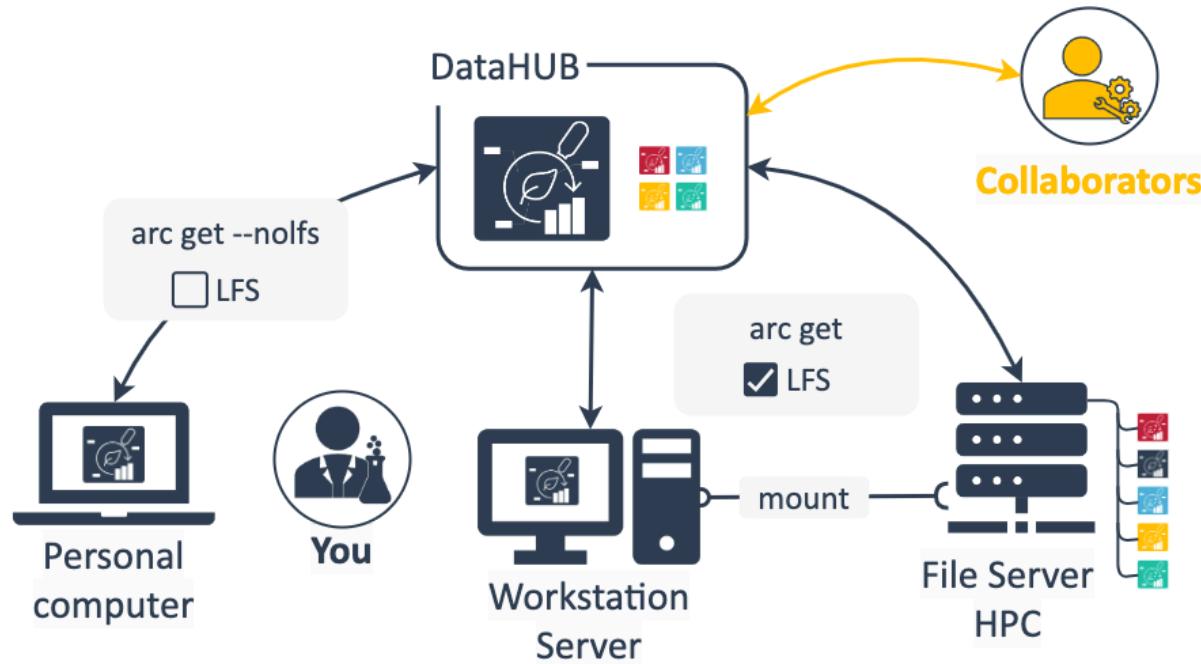
- A sub-section titled 'Default to Auto DevOps pipeline' with a checked checkbox labeled 'instance enabled'. A note below states: 'The Auto DevOps pipeline runs if no alternative CI configuration file is found.' with a 'Learn more' link.
- A note: 'Add a Kubernetes cluster integration with a domain, or create an AUTO\_DEVOPS\_PLATFORM\_TARGET CI variable.'
- A 'Deployment strategy' section with three radio button options:
  - Continuous deployment to production (selected)
  - Continuous deployment to production using timed incremental rollout
  - Automatic deployment to staging, manual deployment to production
- A 'Save changes' button at the bottom of the section.
- Below this, there are sections for 'Runners' and 'Artifacts', each with an 'Expand' button.

# DataHUB concepts and terminology

# 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

# Projects and Groups are not the same

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

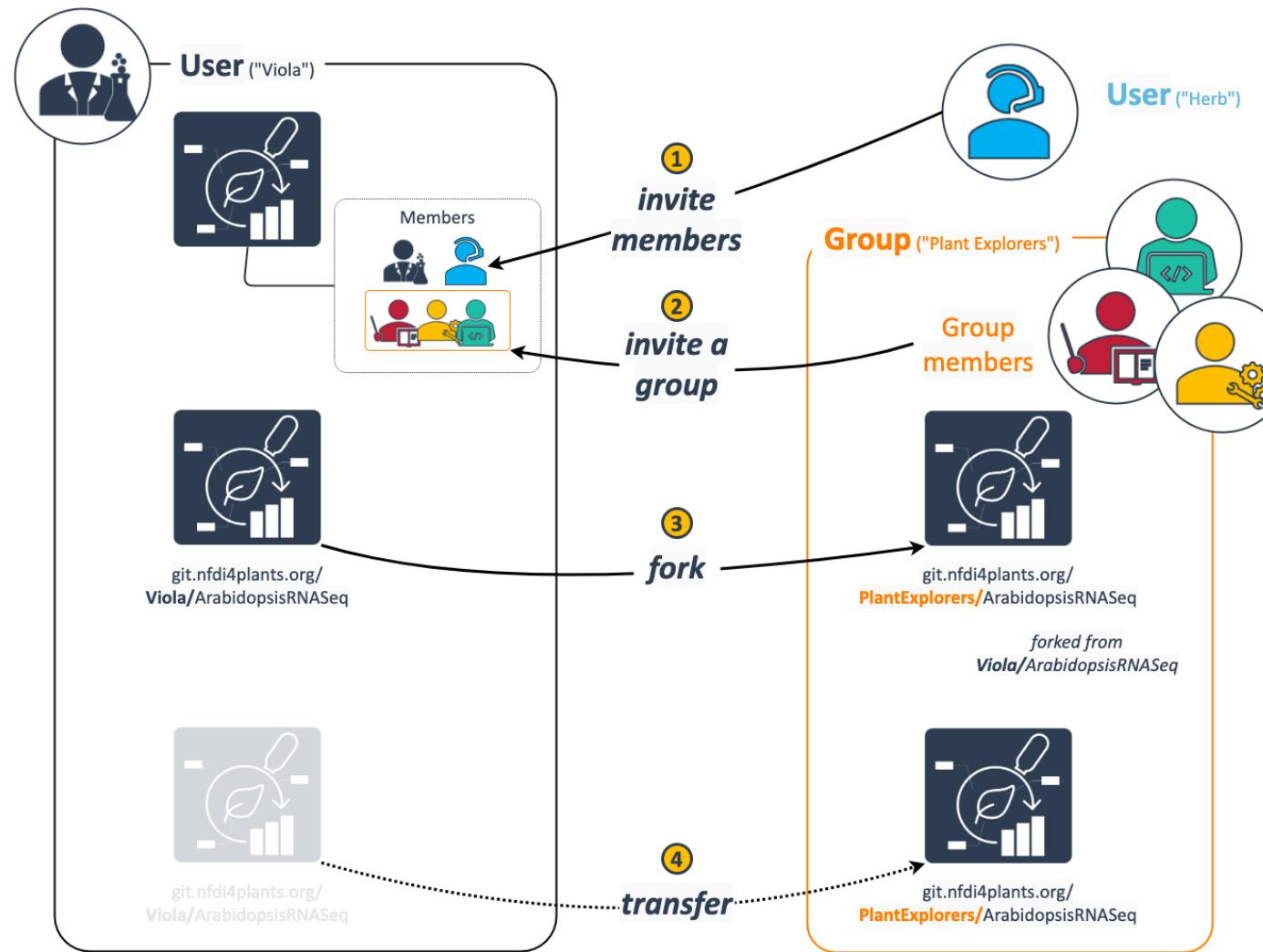
# Project = ARC

- 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

- 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**.

# Permissions & Roles

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.

# ARC DataHUB members // ARC Investigation contacts

The screenshot shows the 'Members' section of a GitLab project. It lists four project members with their accounts, source, max role, expiration date, and activity details. The members are:

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 'Investigation Contacts' section of the ARCitect interface. It displays a list of contacts with their names, accounts, and ratings. The contacts are:

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

ARCitect: Investigation Contacts

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

# DataHub Hands-On

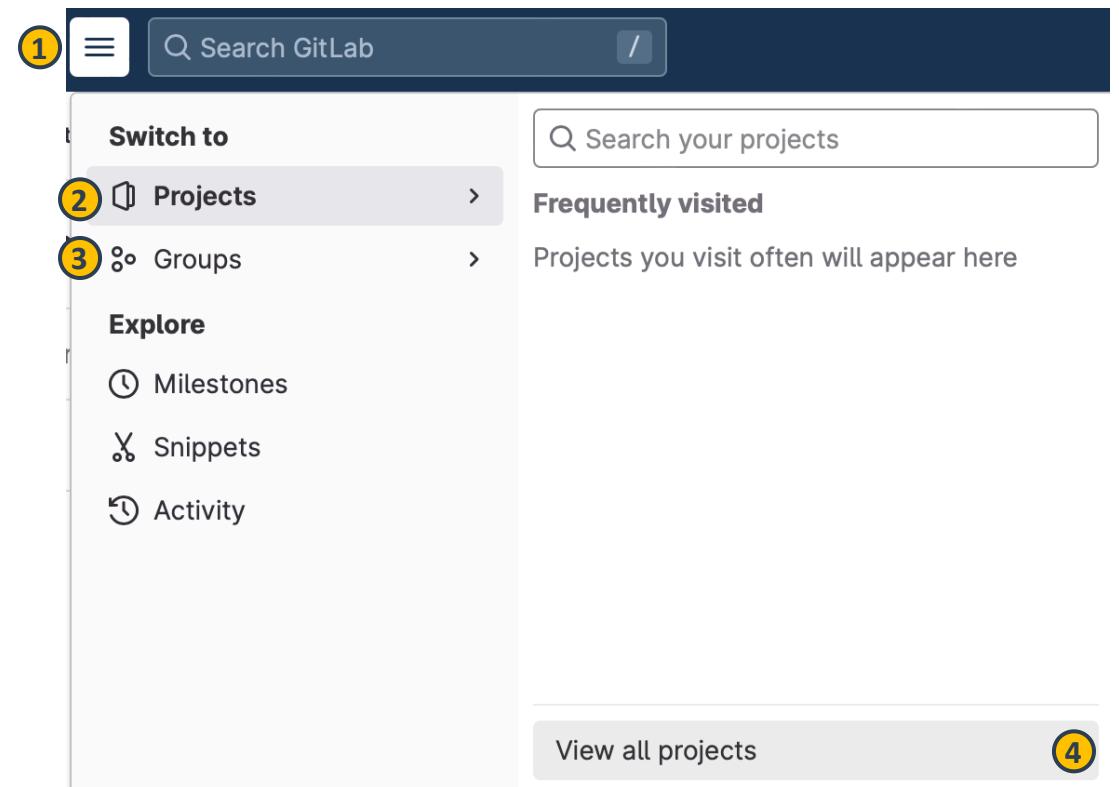
# Navigation Bar



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)

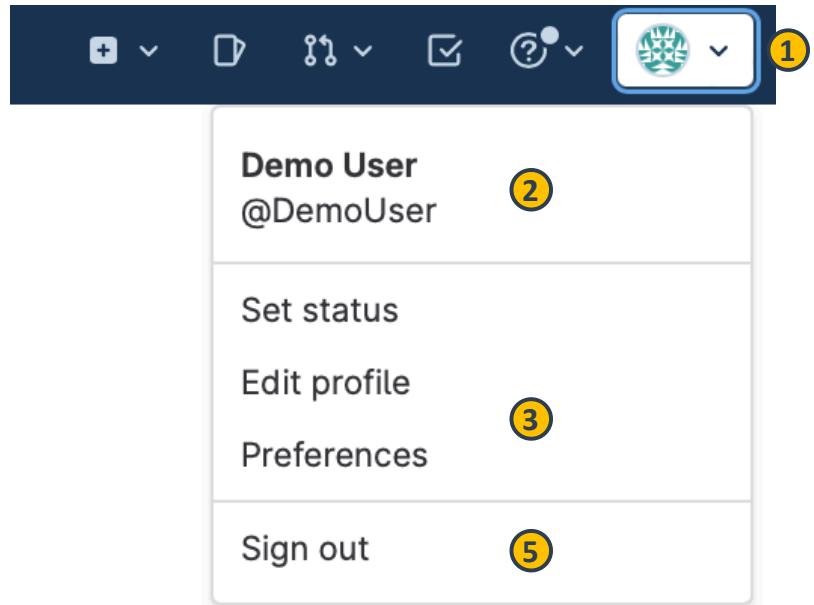
# Hamburger Menu

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

The screenshot shows the GitLab Projects panel. At the top, there is a navigation bar with a search bar containing 'Search GitLab' and a '/' icon. To the right of the search bar are several icons: a plus sign, a document, a gear, a checkmark, a question mark, and a user profile. Below the navigation bar, the main title 'Projects' is displayed, followed by a 'New project' button with a '5' badge indicating pending projects. A 'Yours 1' tab is selected, showing one starred project. Other tabs include 'Starred 0', 'Explore', and 'Topics'. There are filters for 'Filter by name' and 'Name'. Below the tabs, there are buttons for 'All' and 'Personal'. The main content area displays a single project card for 'Demo User / Demo ARC'. The card includes the project name 'D', the owner 'Demo User / Demo ARC', a lock icon indicating it's private, and an 'Owner' badge. Two small circles, labeled 3 and 4, are positioned below the card. To the right of the card are statistics: 0 stars, 0 forks, 0 issues, and 0 merge requests. The card was last updated 9 minutes ago.

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.
5. You can create a **New Project** in the top-right corner (5).

# ARC Panel

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

The screenshot shows the ARC Panel interface for the 'Demo ARC' project. The left sidebar contains navigation links: Project information (1), Repository (2), Issues (3) with 0 notifications, Merge requests (4) with 0 notifications, CI/CD, Security & Compliance, Deployments, Packages and registries, Infrastructure, Monitor, Analytics, Wiki (5), Snippets, and Settings. The main area displays the 'Demo ARC' project details: Project ID: 494, 1 Commit, 1 Branch, 0 Tags, and 51 KB Project Storage. A recent commit titled 'create investigation' by 'Demo User' is shown with a timestamp of 'just now' and a commit hash of '3c538706'. Below the commit, there's a file tree with 'main' selected, showing 'demo-arc /' and a plus sign icon (9). There are also buttons for 'Find file', 'Web IDE', 'Clone' (10), and 'Auto DevOps enabled'. A dashed box highlights 'Add README', 'Add LICENSE', 'Add CHANGELOG', 'Add CONTRIBUTING', 'Add Kubernetes cluster', and 'Configure Integrations'. A table (8) lists project files: '.arc' (Last commit: 'create investigation', Last update: 'just now'), 'assays' (Last commit: 'arc init', Last update: '1 minute ago'), 'runs' (Last commit: 'arc init', Last update: '1 minute ago'), 'studies' (Last commit: 'arc init', Last update: '1 minute ago'), and 'workflows' (Last commit: 'arc init', Last update: '1 minute ago').

Name	Last commit	Last update
.arc	create investigation	just now
assays	arc init	1 minute ago
runs	arc init	1 minute ago
studies	arc init	1 minute ago
workflows	arc init	1 minute ago

# 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 illustrates the ARC Panel interface. On the left, a sidebar menu is open, showing various project management options: Project information (1), Repository (2), Issues (3), Merge requests, CI/CD, Security & Compliance, Deployments, Packages and registries, Infrastructure, Monitor, Analytics, Wiki (4), Snippets, and Settings (5). The main area displays the 'Demo ARC' project details, including the project ID (494), 1 commit, 1 branch, 0 tags, and 51 KB of project storage. A recent commit titled 'create investigation' by 'Demo User' is shown. The repository navigation bar includes 'main' and 'demo-arc /' with a dropdown menu containing 'Auto DevOps enabled' and several 'Add' buttons for README, LICENSE, CHANGELOG, CONTRIBUTING, Kubernetes cluster, and Configure Integrations. Below this is a table of recent commits:

Name	Last commit	Last update
.arc	create investigation	just now
assays	arc init	1 minute ago
runs	arc init	1 minute ago
studies	arc init	1 minute ago
workflows	arc init	1 minute 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).

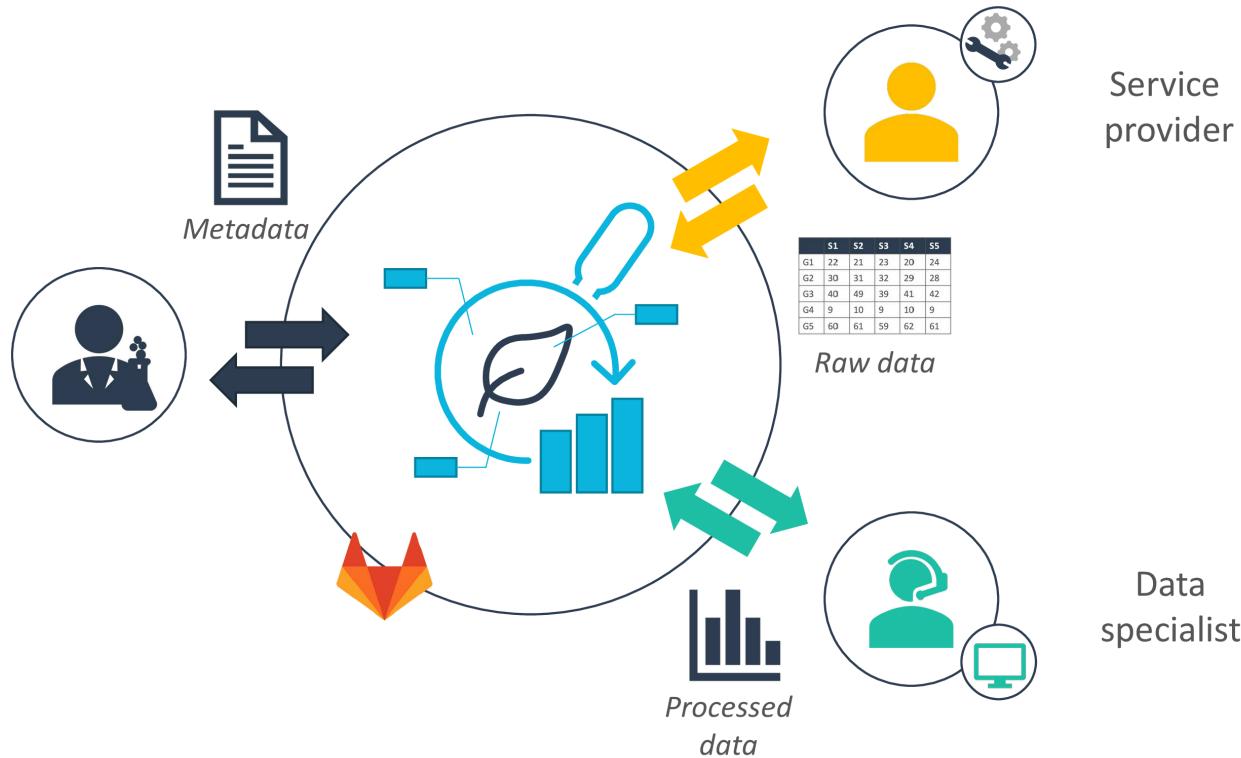
The screenshot displays the main interface of the ARC Panel. On the left is a sidebar with the following items:

- 1 Project information
- 2 Repository
- 3 Issues
- 4 CI/CD
- 5 Security & Compliance
- 6 Deployments
- 7 Packages and registries
- 8 Infrastructure
- 9 Monitor
- 10 Analytics
- 11 Wiki
- 12 Snippets
- 13 Settings

The main content area shows the following details for the "Demo ARC" project:

- Demo User > Demo ARC
- Project ID: 494
- 1 Commit, 1 Branch, 0 Tags, 51 KB Project Storage
- create investigation (Demo User authored just now)
- main / demo-arc /
- Auto DevOps enabled
- Add README, Add LICENSE, Add CHANGELOG, Add CONTRIBUTING
- Add Kubernetes cluster, Configure Integrations
- File list:
  - .arc (Last commit: create investigation, Last update: just now)
  - assays (arc init, 1 minute ago)
  - runs (arc init, 1 minute ago)
  - studies (arc init, 1 minute ago)
  - workflows (arc init, 1 minute ago)

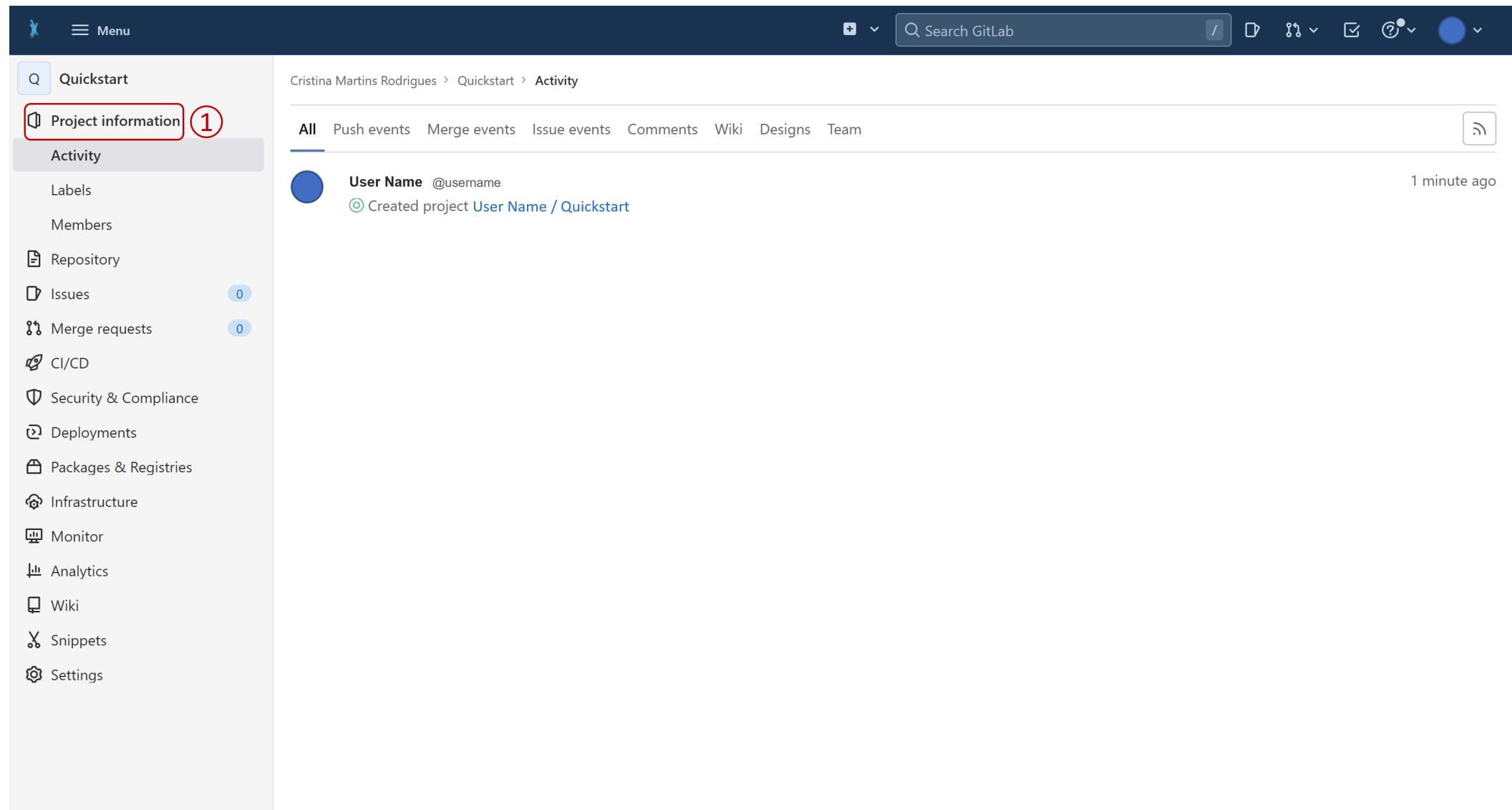
# Collaborate and share



# Invite collaborators

- Unless changed, your ARC is set to private by default.
- To collaborate, you can invite lab colleagues or project partners to your ARC by following the steps on the subsequent slides.
- To get started [sign in](#) to the DataHUB and open the ARC you want to share.

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



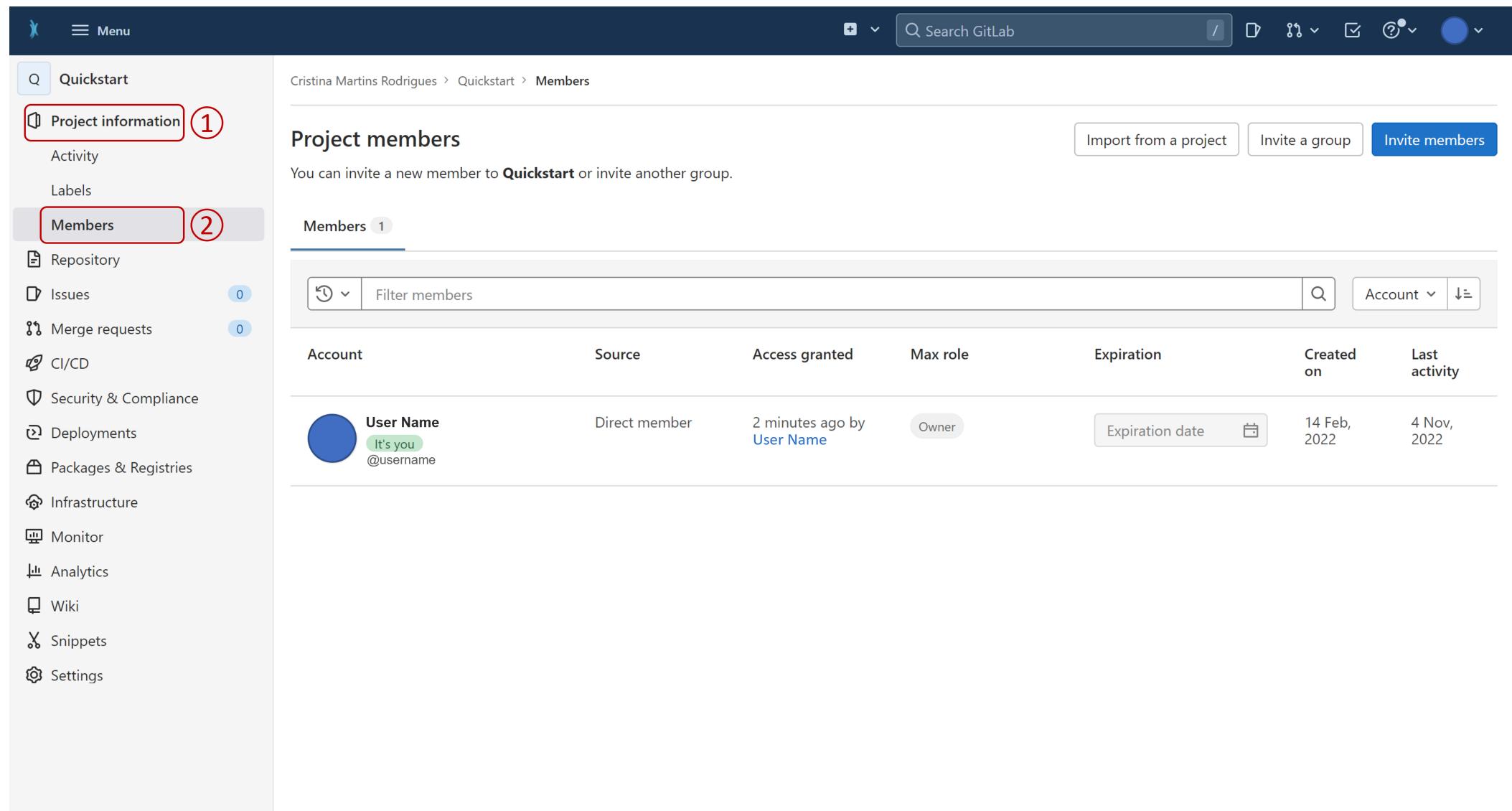
The screenshot shows the GitLab interface. The left sidebar has a list of project management sections: Quickstart, Project information (highlighted with a red box and circled with a red number 1), Activity, Labels, Members, Repository, Issues (0), Merge requests (0), CI/CD, Security & Compliance, Deployments, Packages & Registries, Infrastructure, Monitor, Analytics, Wiki, Snippets, and Settings. The main content area shows the activity feed for the 'Quickstart' project. The top navigation bar includes a 'Menu' icon, search bar, and various user and project management icons.

Cristina Martins Rodrigues > Quickstart > Activity

All Push events Merge events Issue events Comments Wiki Designs Team

User Name @username  
Created project User Name / Quickstart 1 minute ago

## 2. Click on Members



The screenshot shows the 'Members' page in the GitLab interface. The left sidebar has a 'Members' item highlighted with a red box and the number 2. The main content area is titled 'Project members' and displays one member: 'User Name' (highlighted with a red box and the number 1). The member was granted access 2 minutes ago by 'User Name' and has the role 'Owner'. The table includes columns for Account, Source, Access granted, Max role, Expiration, Created on, and Last activity.

Account	Source	Access granted	Max role	Expiration	Created on	Last activity
User Name It's you @username	Direct member	2 minutes ago by User Name	Owner	Expiration date <input type="button" value="calendar icon"/>	14 Feb, 2022	4 Nov, 2022

### 3. Click on Invite members

The screenshot shows the 'Members' page in GitLab. On the left, a sidebar lists various project management options like 'Project information', 'Members' (which is selected and highlighted with a red box and number 2), and 'Invite members'. The main area displays the 'Project members' section with a table showing one member named 'User Name'. A large blue button labeled 'Invite members' is highlighted with a red box and number 3.

Members 1

Account	Source	Access granted	Max role	Expiration	Created on	Last activity
User Name It's you @username	Direct member	2 minutes ago by User Name	Owner	Expiration date <input type="button" value="calendar icon"/>	14 Feb, 2022	4 Nov, 2022

## 4. Search for potential collaborators

The screenshot shows the GitLab interface for managing project members. On the left sidebar, under the 'Project information' section, the 'Members' option is highlighted with a red box and the number '2'. In the main content area, the 'Project members' section displays a message: 'You can invite a new member to **Quickstart** or invite another group.' Below this, there's a table with columns for 'Expiration', 'Created on', and 'Last activity'. A single member is listed: '14 Feb, 2022' (Expiration), '14 Feb, 2022' (Created on), and '4 Nov, 2022' (Last activity). An 'Import from a project' button is available. At the top right of the page, there's a search bar and several navigation icons. A modal window titled 'Invite members' is open over the main content. It contains the message 'You're inviting members to the **Quickstart** project.' and a form with the following fields:

- 'Username or email address' input field, which is empty and has a red box around it with the number '4'.
- 'Select members or type email addresses' placeholder text.
- 'Select a role' dropdown menu set to 'Guest'.
- 'Read more about role permissions' link.
- 'Access expiration date (optional)' input field with a calendar icon.
- 'Cancel' and 'Invite' buttons at the bottom.

## 5. Select a role

The screenshot shows the GitLab 'Project members' page for the 'Quickstart' project. The sidebar on the left is visible with various project management options like 'Project information' (1), 'Members' (2), and 'Repository'. The main area displays the 'Project members' list, which currently has one member: 'Cristina Martins Rodrigues' (1 member). A modal window titled 'Invite members' is open, prompting the user to invite members to the 'Quickstart' project. The modal includes fields for 'Username or email address' (with an empty input field) and a dropdown menu for 'Select a role' (5). The role dropdown is set to 'Guest' and is highlighted with a red box. Other available roles listed are 'Reporter', 'Developer', 'Maintainer', and 'Owner'. At the bottom of the modal are 'Cancel' and 'Invite' buttons.

# Choosing the proper role

## Guests

Have the least rights. They will not be able to see the content of your ARC (only the wiki page).

## Reporters

Have **read access** to your ARC. This is recommended for people you ask for consultancy.

## Developers

The choice for most people you want to invite to your ARC. Developers have **read and write access**, but cannot maintain the project on the DataHUB, e.g. inviting others.

## Maintainers

Gives the person the same rights as you have (except of removing you from your own project). This is recommended for inviting PIs or group leaders allowing them to add their group members for data upload or analysis to the project as well.

*A detailed list of all permissions for the individual roles can be found [here](#)*

# Congratulations!

You have just shared your ARC with a collaborator.



# Version control

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

The screenshot shows the 'Demo User > Demo ARC' interface. On the left, a sidebar lists project management features: Project information (1), Repository (2), Issues (3), Merge requests (0), CI/CD, Security & Compliance, Deployments, Packages and registries, Infrastructure, Monitor, Analytics, Wiki (4), Snippets, and Settings (5). The main area displays the 'Demo ARC' project details: Project ID: 494, 1 Commit, 1 Branch, 0 Tags, and 51 KB Project Storage. A recent commit is shown: 'create investigation' by Demo User, authored just now, with commit hash 3c538706. Below this is a file tree with 'main' selected, showing 'demo-arc /' and a '+' button (9). To the right are buttons for Find file, Web IDE, Clone (10), and Configure Integrations. At the bottom is a table of files with their last commits and update times.

Name	Last commit	Last update
.arc	create investigation	just now
assays	arc init	1 minute ago
runs	arc init	1 minute ago
studies	arc init	1 minute ago
workflows	arc init	1 minute ago

# Project Management

## Add issues to your ARC (3)

The screenshot shows the 'Demo User > Demo ARC' interface. On the left, a sidebar lists project management features with numbered callouts:

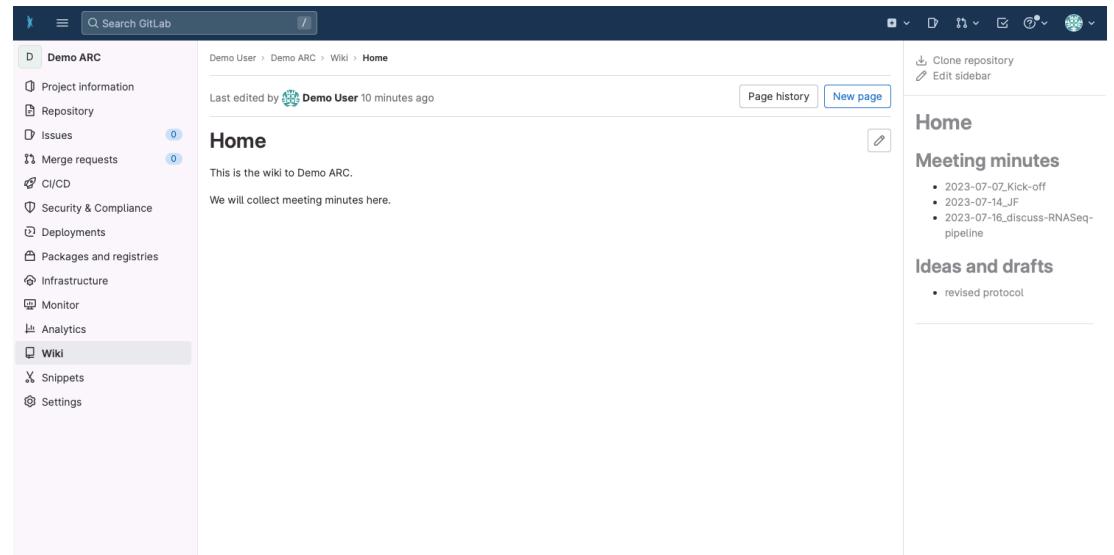
- 1 Project information
- 2 Repository
- 3 Issues (highlighted with a yellow circle)
- 4 Wiki
- 5 Settings

The main area displays the 'Demo ARC' project details:

- Demo ARC (Project ID: 494)
- 6 Issues (highlighted with a yellow circle)
- 7 1 Commit, 1 Branch, 0 Tags, 51 KB Project Storage
- 8 File list:
  - .arc (Last commit: create investigation, Last update: just now)
  - assays (Last commit: arc init, Last update: 1 minute ago)
  - runs (Last commit: arc init, Last update: 1 minute ago)
  - studies (Last commit: arc init, Last update: 1 minute ago)
  - workflows (Last commit: arc init, Last update: 1 minute ago)
- 9 Main branch dropdown (highlighted with a yellow circle)
- 10 Clone button (highlighted with a yellow circle)

# 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"



# Publish your ARC with a few clicks

The screenshot shows a bioRxiv project page for 'Ru\_ChlamyHeatstress'. The project icon features a green cell with a鞭毛 (flagella) and a red thermometer icon. The project ID is 122. The page displays 53 commits, 1 branch, 0 tags, and 293.9 GB of project storage. Topics listed include Chlamydomonas, abiotic stress, proteomics, and more. A note 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 are three numbered steps: 1. pipeline (status passed), 2. Publish ARC (button), and 3. arc quality (status 301/301).

Ru\_ChlamyHeatstress

Project ID: 122

53 Commits 1 Branch 0 Tags 293.9 GB Project Storage

Topics: Chlamydomonas, abiotic stress, proteomics, + 1 more

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.

1 pipeline passed 2 Publish ARC 3 arc quality 301/301

# Receive a DOI

Published September 7, 2023 | Version v1

Dataset  Edit  
New version  
Share

Versions

Version v1	Sep 7, 2023
10.60534/9e5jx-75d83	

Details

DOI  
[DOI 10.60534/9e5jx-75d83](https://doi.org/10.60534/9e5jx-75d83)

Resource type  
Dataset

Publisher  
DataPLANT

Export

JSON 

1

## Citation

Style

APA

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., Schröder, 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>

2

## Description

hosted on: <https://git.nfd4plants.org/projects/122>

## Files

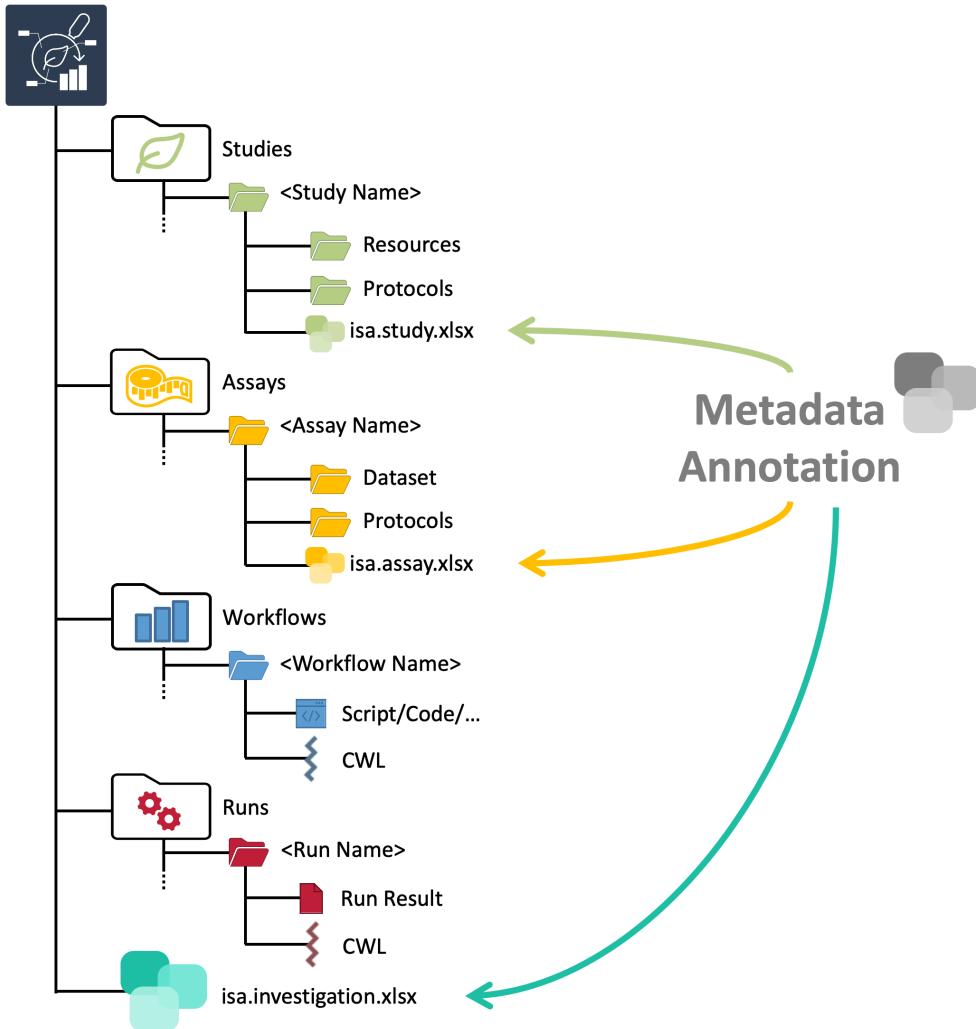
arc-summary.md

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

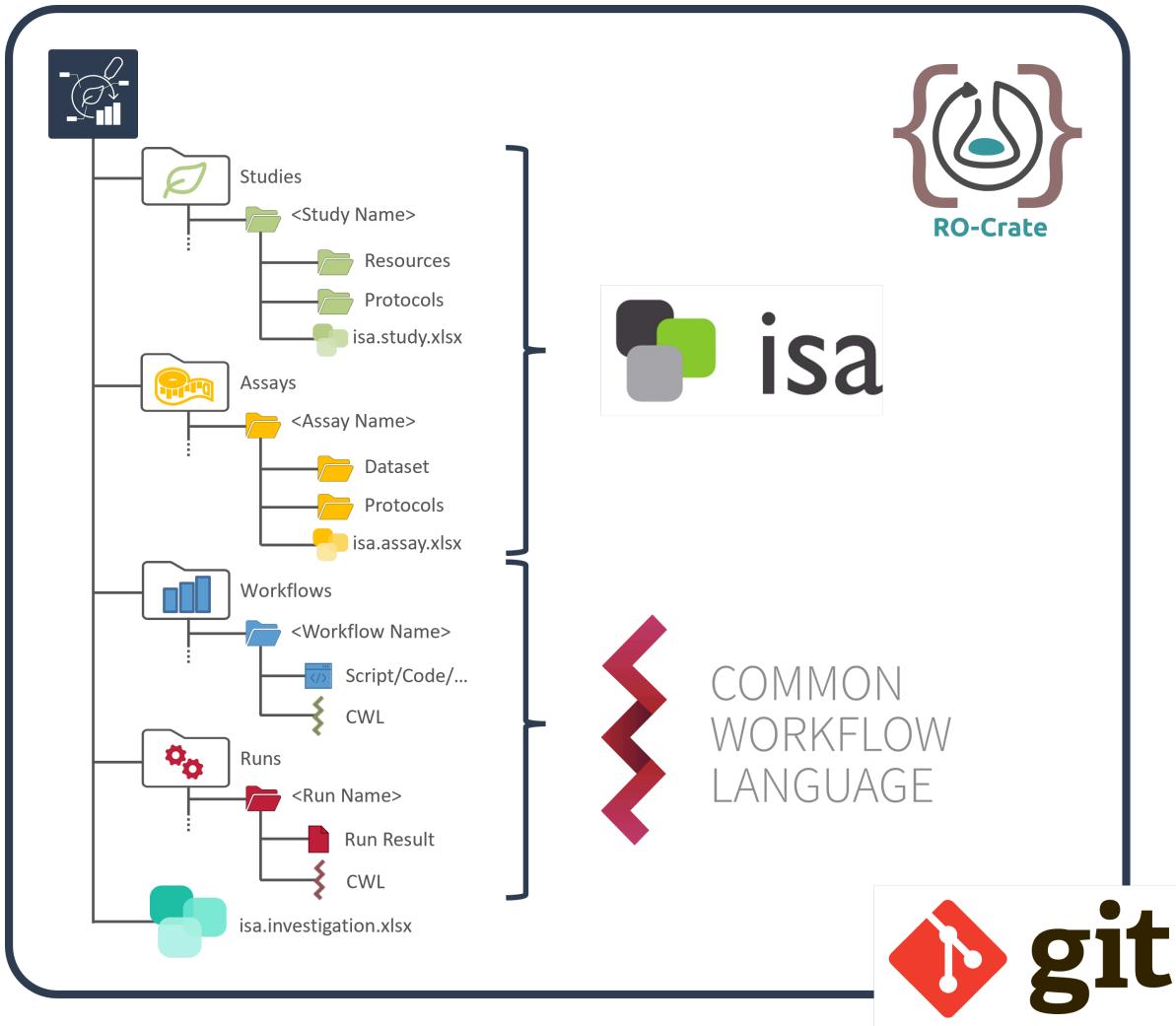
### File contents:

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

# ARCs come with comprehensive metadata

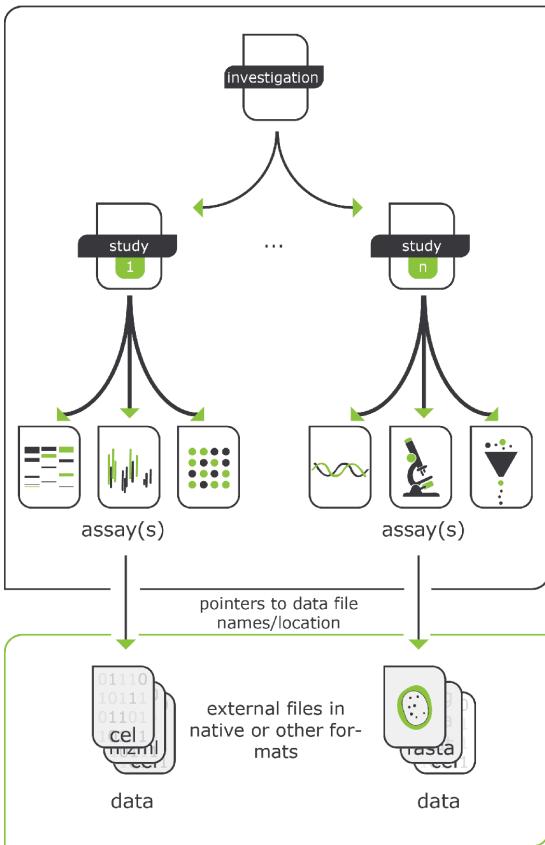


# ARC builds on standards



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

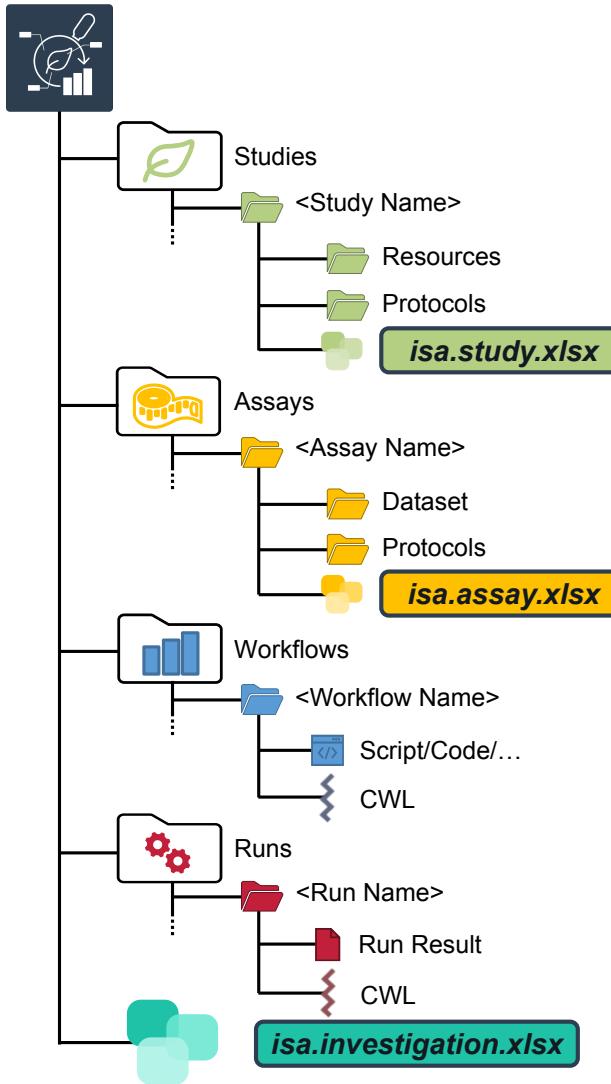
# ARC builds on ISA



**Investigation**  
Overall goals  
Scientific context

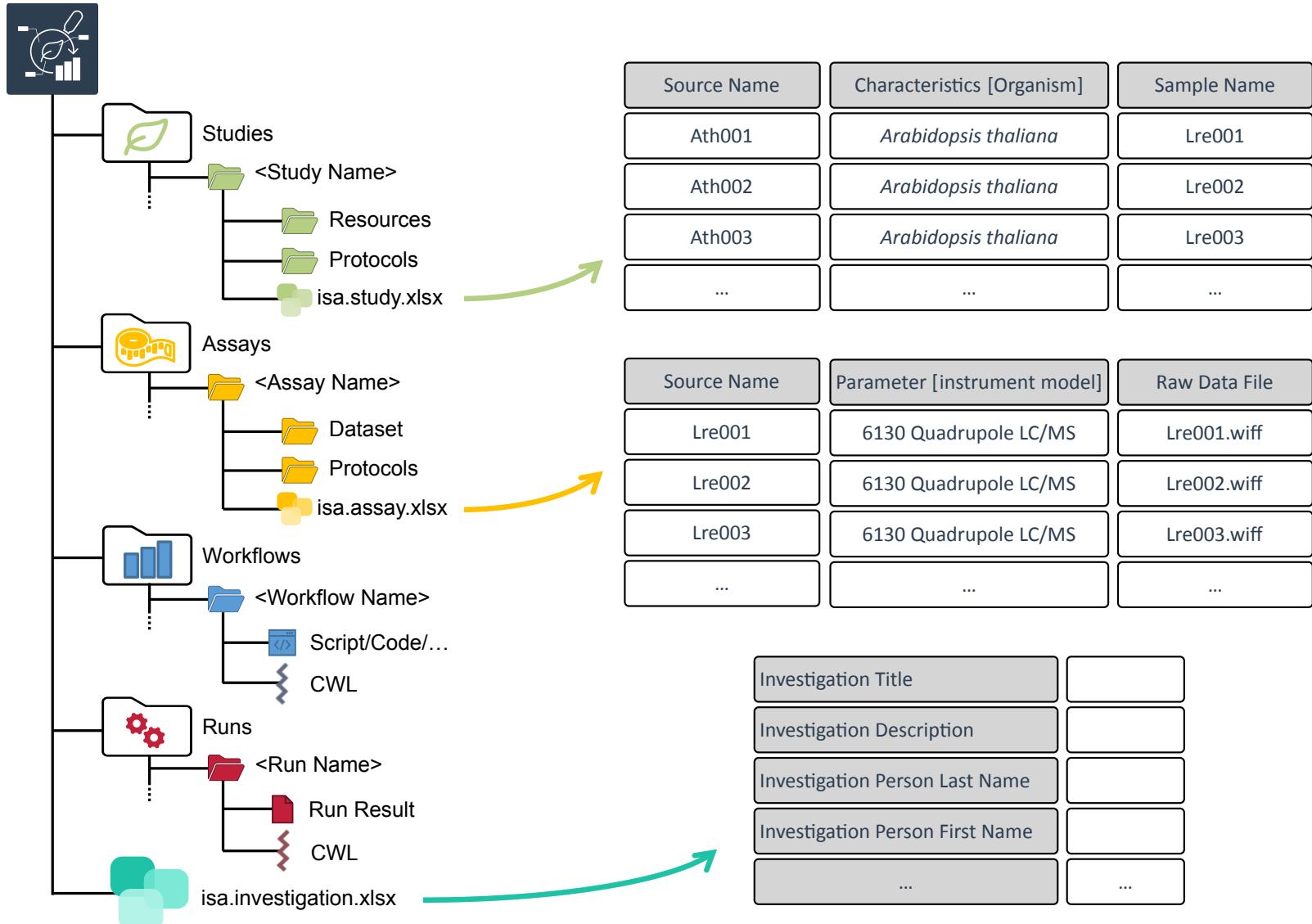
**Study**  
Experimental design

**Assay**  
Leading to (raw) data



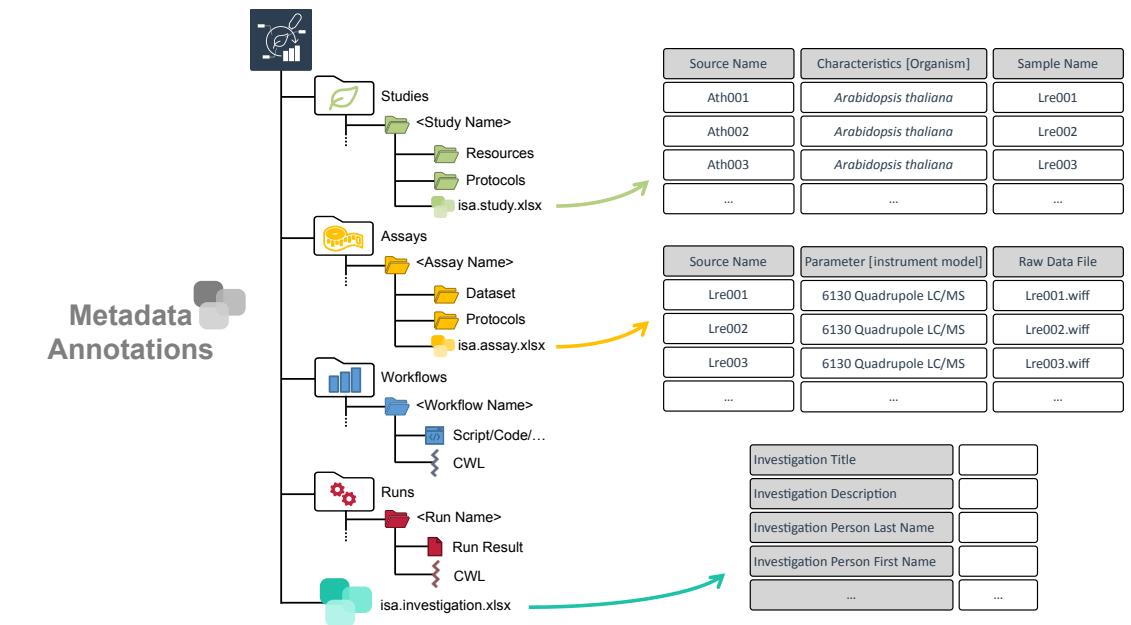
# ARC builds on ISA to connect data

Metadata Annotations

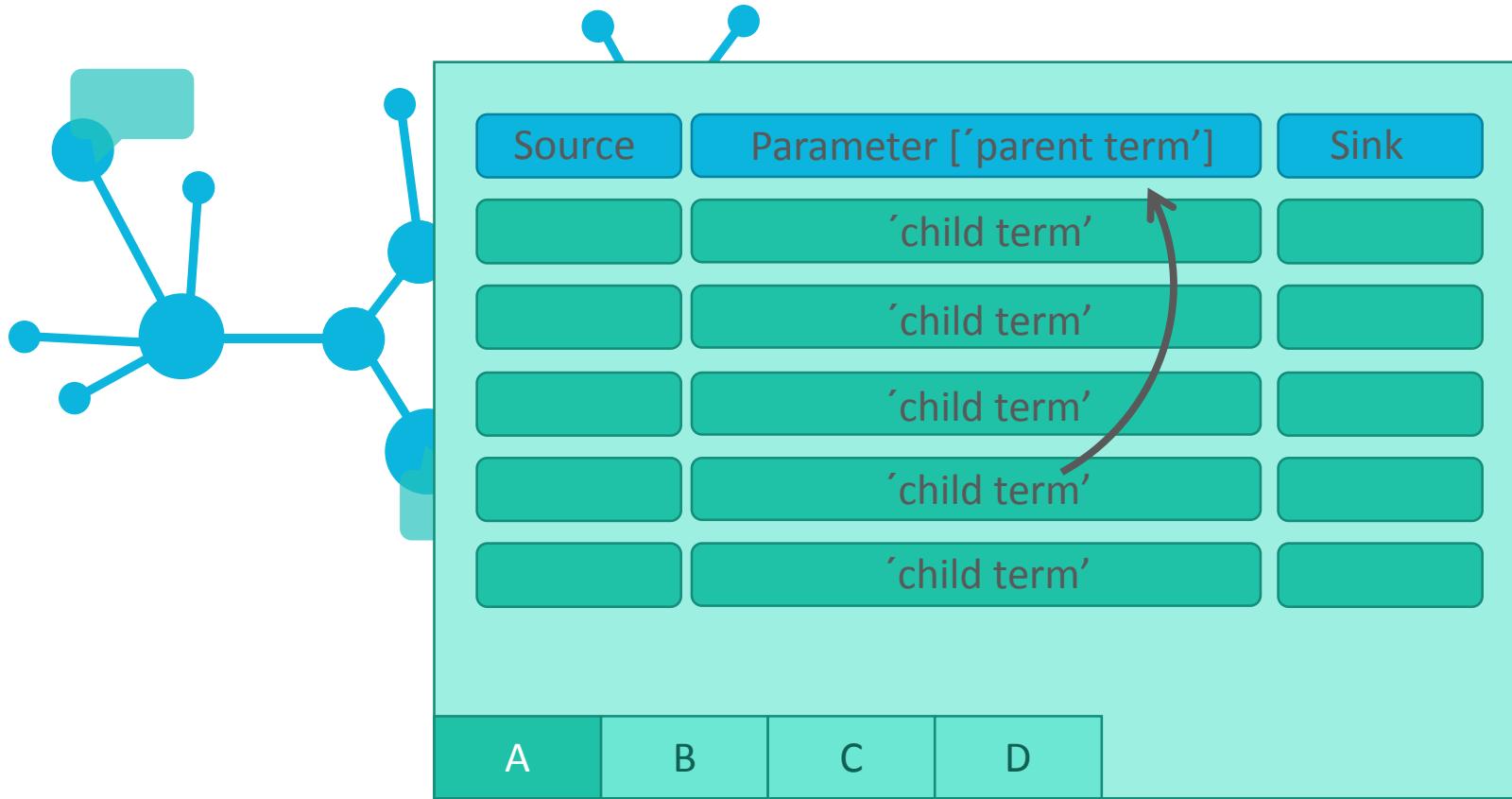


# ARC builds on ISA to link data

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



# Annotation by flattening the knowledge graph



- Low-friction metadata annotation
- Familiar spreadsheet, row/column-based environment

# Annotation principle

Sample	Parameter [instrument model]	Data
	'TripleTOF4600'	

A      B      C      D



- Low-friction metadata annotation
- Familiar spreadsheet, row/column-based environment

# Adding new building blocks

The screenshot shows the Swate software interface. On the left, there is a table with six rows of experimental data:

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

Below the table is a numeric input field with the value '1' and a '+' button. To the right of the table is a modal dialog box titled 'New building blocks'. The dialog has a search bar containing 'instrument model' and a dropdown menu with options: Input (selected), Parameter, Factor, Characteristic, Component, More, and Output. A green button labeled 'Add Column' is highlighted. At the bottom of the dialog are 'Help' and 'info' links.

At the bottom of the interface, there are tabs for 'Metadata' and 'plant\_material' (which is selected), followed by a '+' button.

Swate can be used for the annotation of **isa.study.xlsx** and **isa.assay.xlsx** files

# Annotation Building Block types

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

The screenshot shows a software interface for managing experimental data. 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. The columns are labeled "Input [Source Name]", "Characteristic [organism]", "Factor [watering exposure]", and "Output [Sample Name]". The "Factor" column is highlighted with a blue background and has the word "Factor" written over it. The "Output" column also has a blue background and has the word "Output" written over it. Arrows point from these labels to their respective columns. A "New Parameter" dialog box is open on the right side, containing fields for "Parameter", "Factor", "Characteristic", "Component", and "Output". The "Output" field is currently set to "instrument model". A "Sidebar" button is located at the bottom right of the dialog. At the bottom of the main window, there are tabs for "Metadata" and "plant\_material", with a "+" button between them.

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

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

# Ontology term search

The screenshot shows the Swate interface with a data grid and an ontology search sidebar.

**Data Grid:**

	Input [Source Name]	Characteristic [Organism]	»	Factor [watering exposure]	»	Output [Sample Name]
1	DB_097		Q	12 days drought		CAM_01
2	DB_099			12 days drought		CAM_02
3	DB_103			12 days drought		CAM_03
4	DB_161			12 days drought + 2 days rewated		reC3_01
5	DB_163			12 days drought + 2 days rewated		reC3_02
6	DB_165			12 days drought + 2 days rewated		reC3_03

**Search Sidebar:**

Ontology term search  
Search for an ontology term to fill into the selected field(s)

Parent: Organism, MIAPPE:0041   [Use advanced search](#)

[Fill selected cells with this term](#)

Metadata   plant\_material   +

Swate Release Version v1.0.0-beta.03 Host Browser

Enable related term directed search to directly fill cells with child terms

# Fill your table with ontology terms

The screenshot shows the Swate interface with a table on the left and a search sidebar on the right.

**Table Data:**

	Input [Source Name]	Characteristic [Organism]	Factor [watering exposure]	Output [Sample Name]
1	DB_097	Talinum fruticosum	✓ 12 days drought	CAM_01
2	DB_099	Talinum fruticosum	✓ 12 days drought	CAM_02
3	DB_103	Talinum fruticosum	✓ 12 days drought	CAM_03
4	DB_161	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_01
5	DB_163	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_02
6	DB_165	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_03

**Search Sidebar:**

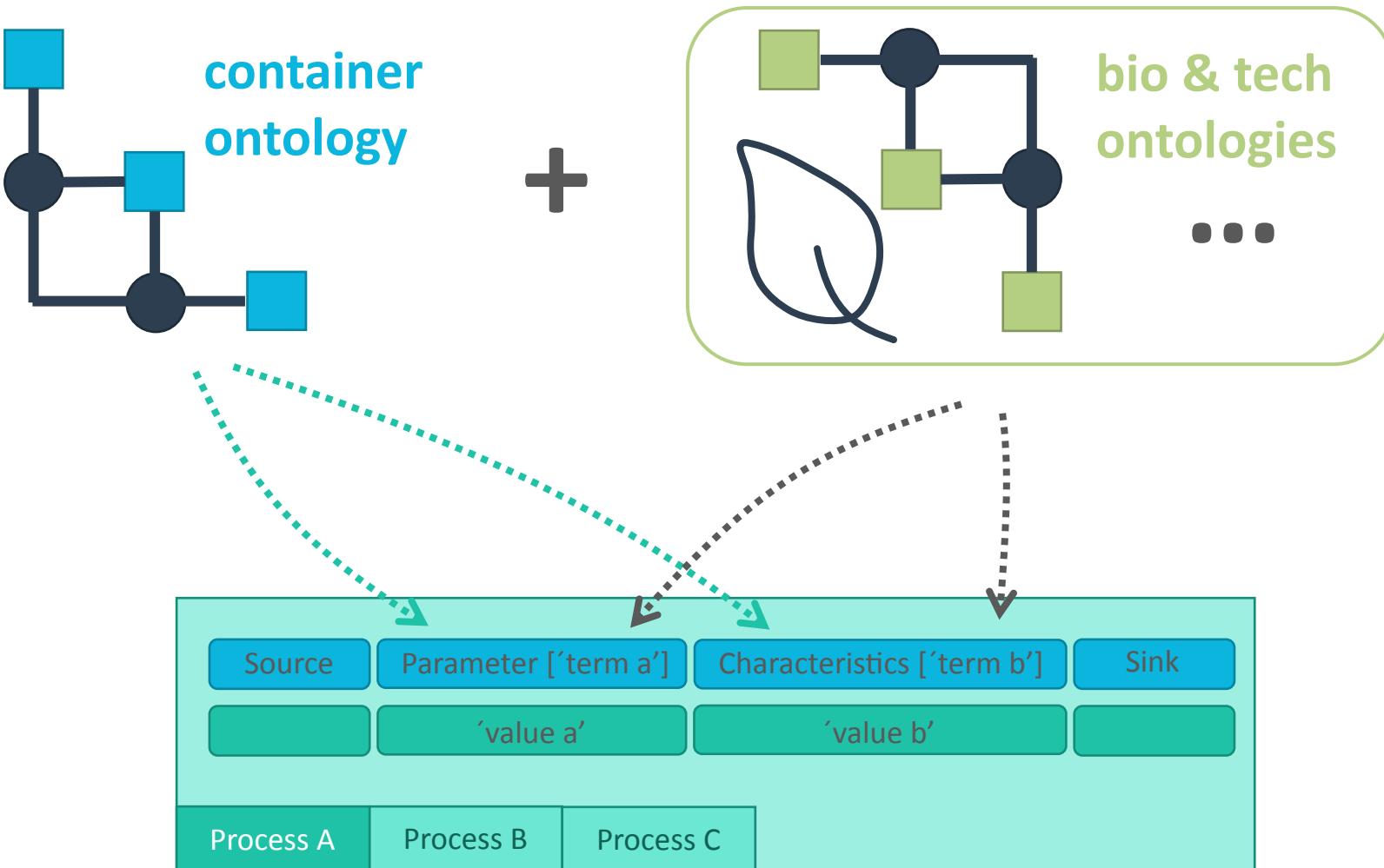
Ontology term search  
Search for an ontology term to fill into the selected field(s)

Parent: Organism, MIAPPE:0041 Use advanced search

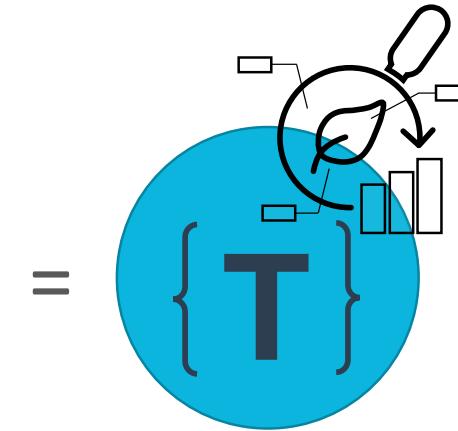
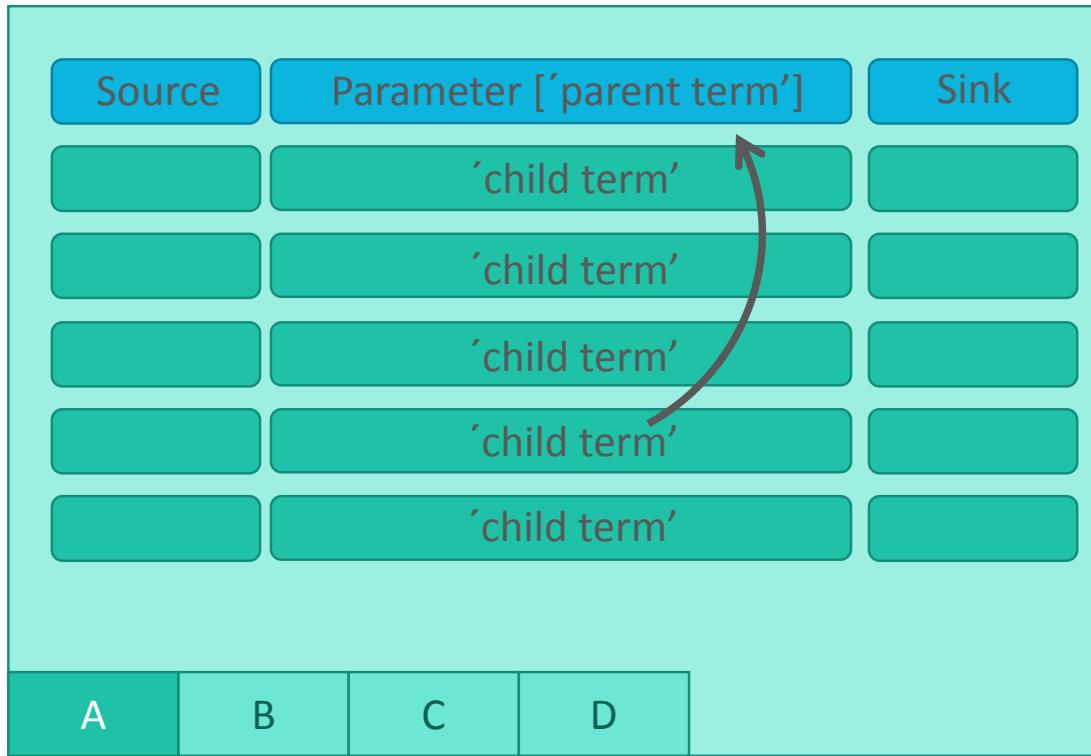
**Bottom Navigation:**

Metadata plant\_material + > Swate Release Version v1.0.0-beta.03 Host Browser

# Hierarchical combination of ontologies

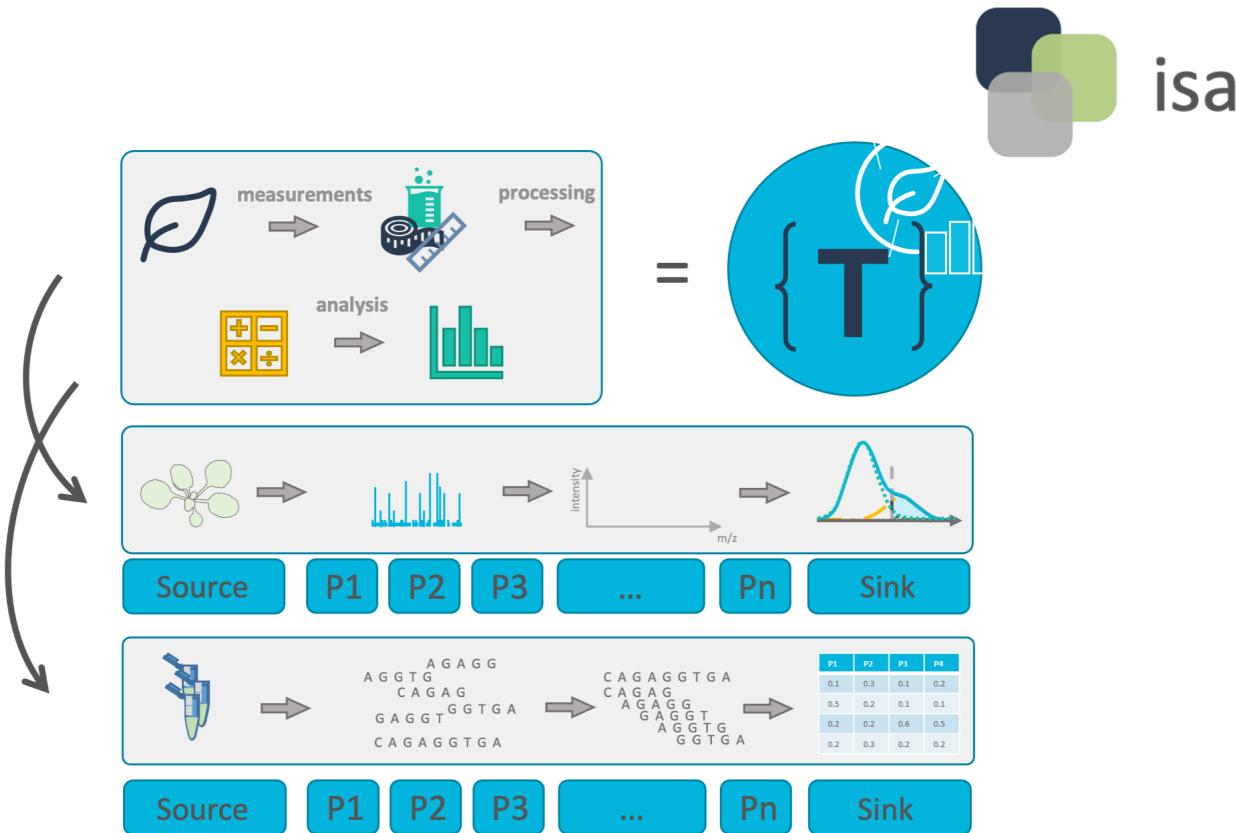


# Checklists and Templates



Metadata standards or repository requirements can be represented as templates

# Realization of lab-specific metadata templates



Facilities can define their most common workflows as templates

# Directly import templates via Swate

- DataPLANT curated
- Community templates

The screenshot shows the Swate web application interface. At the top, there is a dark header bar with several small icons. Below the header is a search bar with two input fields: "Search by template name" and "Search for tags". Underneath the search bar is a dropdown menu labeled "Select community" with the option "DataPLANT official" selected. The main content area is a table listing various templates. The columns are "Template Name", "Community", and "Template Version". Each row contains a link icon at the end. The table lists the following templates:

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

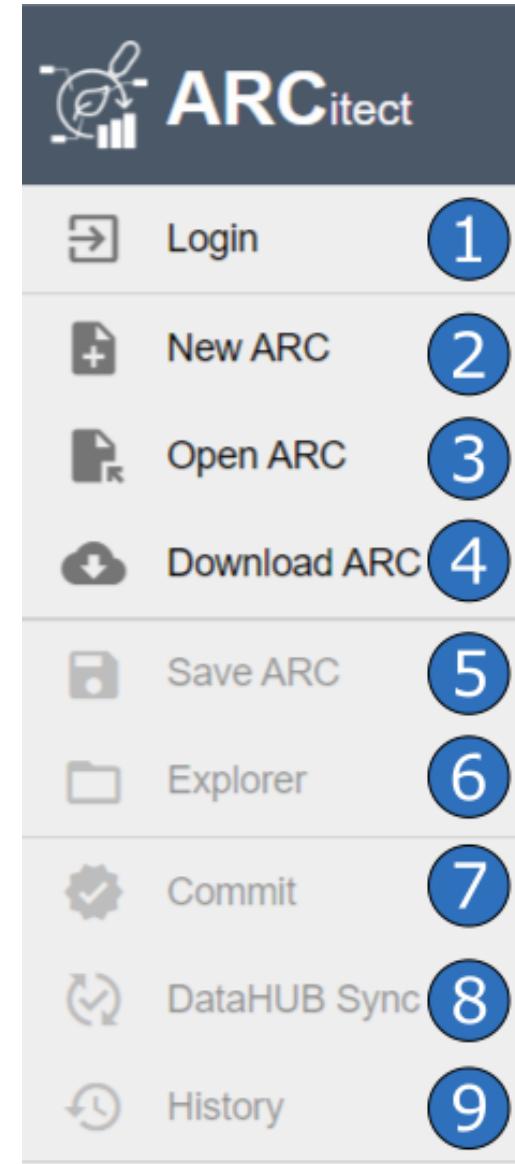
# Swate hands-on with demo data

# Goals

- Get familiar with ISA metadata and Swate
- Annotate data in your ARC

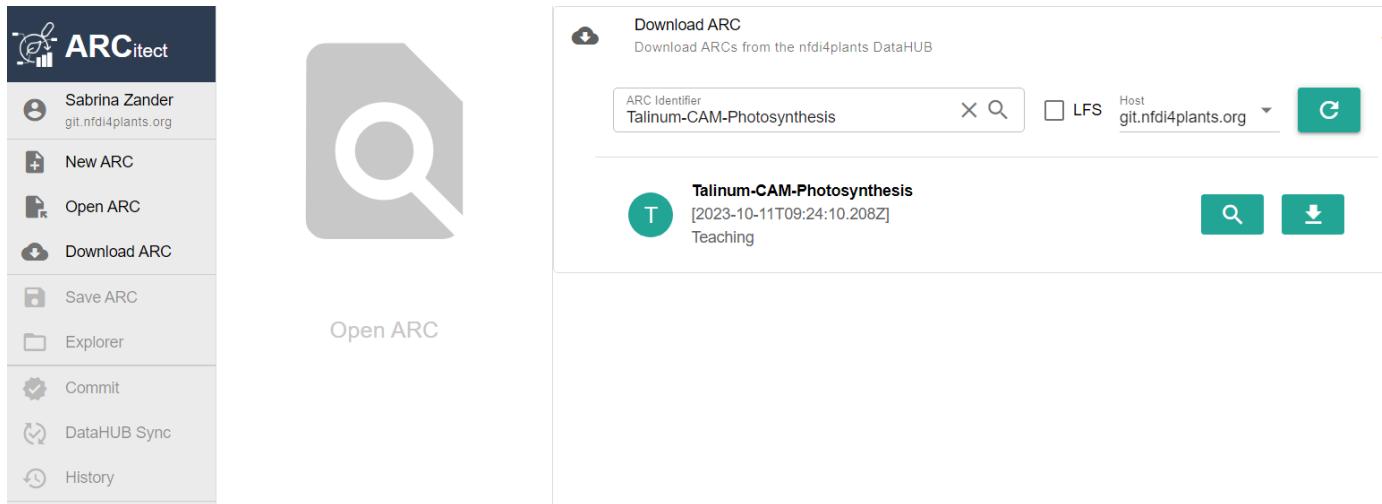
# Download the demo data

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



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

## Alternative: Use Swate standalone

Open [Swate standalone](#) in web browser

💡 Alternatively, you can use Swate as [Excel Plugin](#), but this is outdated

# Swate Overview

The screenshot illustrates the Swate software interface with several key components highlighted:

- Widgets**: A header bar containing various icons.
- Input**: A column in the data grid labeled "Input [Source Name]".
- Characteristic**: A column in the data grid labeled "Characteristic [organism]".
- Factor**: A column in the data grid labeled "Factor [watering exposure]".
- Output**: A column in the data grid labeled "Output [Sample Name]".
- New Parameter**: A modal dialog titled "New Parameter" with a search bar for "instrument model".
- Sidebar**: A vertical bar on the right side of the interface.

The data grid contains the following rows:

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

At the bottom left, there are buttons for "Metadata" and "plant\_material" with a plus sign. The sidebar has a back arrow icon.

# Let's annotate the plant samples first

1. Check out the lab notes

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

*ARCitect*

2. Select the study `talinum_drought`

3. Add a new table

*Swate standalone*

- Import the empty

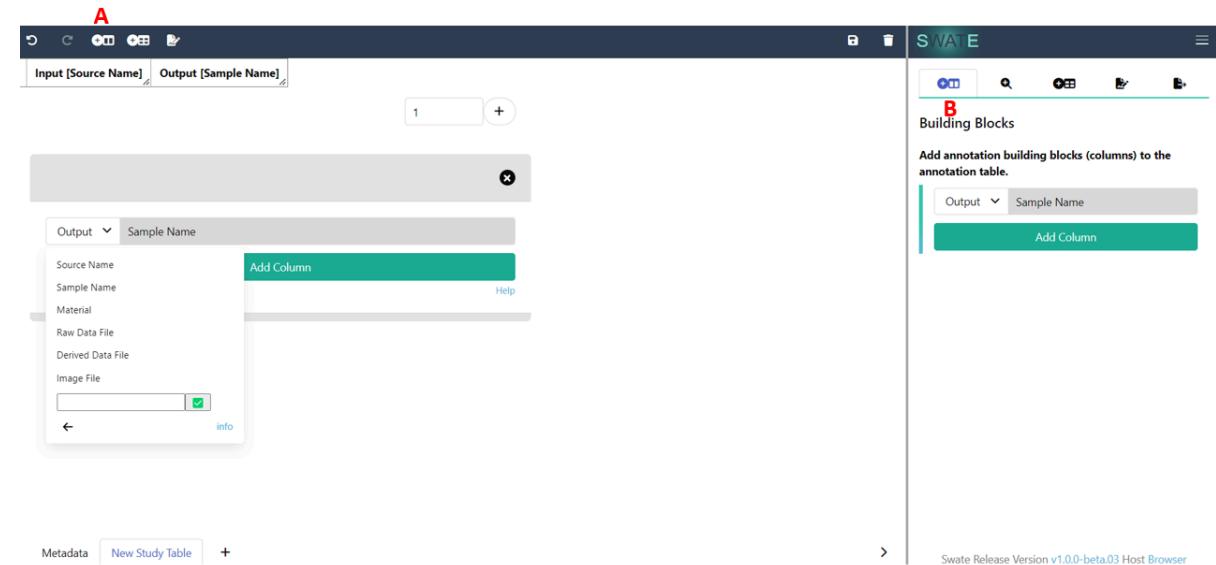
```
studies/talinum_drought/isa.study.xlsx
```



# Create an annotation table

Create a Swate annotation table by adding **Building blocks** via the widget (A) or the sidebar (B)

- 💡 Each table can contain only one *Input* and one *Output* column
- 💡 different *Input* and *Output* options:  
Source Name , Sample Name ,  
Material , Raw Data File ,  
Derived Data File , Image File



# Add more Building blocks

1. Add an *Input* ( Source Name ) and *Output* ( Sample Name ) column

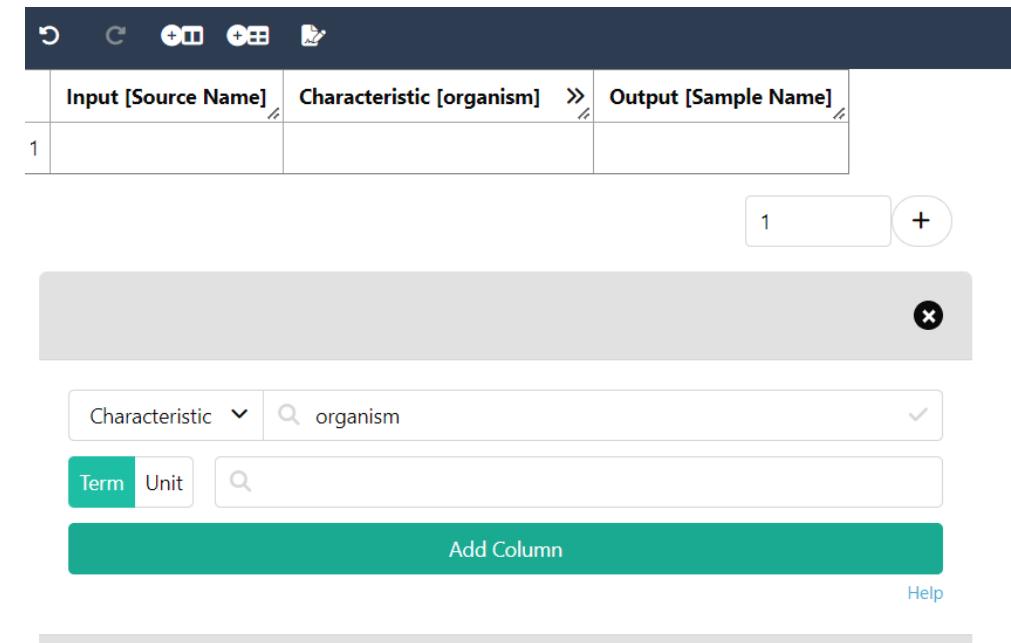
2. Select *Characteristic* from the drop-down menu

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

4. Select the Term with the id OBI:0100026 and,

5. Click Add Column

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



The screenshot shows a software interface for managing building blocks. At the top, there is a toolbar with various icons. Below it is a table with three columns: 'Input [Source Name]', 'Characteristic [organism]', and 'Output [Sample Name]'. The first row contains the number '1' under the 'Input' column. To the right of the table is a counter '1' and a '+' button. Below the table is a search bar with the placeholder 'Characteristic' and a dropdown arrow, followed by a search input field containing 'organism' with a checkmark. Underneath the search bar are two buttons: 'Term' (highlighted in green) and 'Unit', and a magnifying glass icon. At the bottom is a teal-colored button labeled 'Add Column'.

# Insert values to annotate your data

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

Input [Source Name]	Characteristic [organism]	Output [Sample Name]
1	Talinum fruticosum	
	↳ Talinum fruticosum NCBITaxon:110664	
	↳ Talinum paniculatum NCBITaxon:107604	

## Add a 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 `microeinsteин 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 to annotate your data

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

## 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

# Your ISA 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.

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

## 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 (DB\_\*) as **Sample Name**

## Let's annotate the RNA Seq data

1. Navigate to the demo ARC.
2. Open the lab notes `assays/rnaseq/protocols/RNA_extraction.txt` in a text editor.
3. Import the empty `assays/rnaseq/isa.assay.xlsx` workbook in Swate.

# Use a template

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

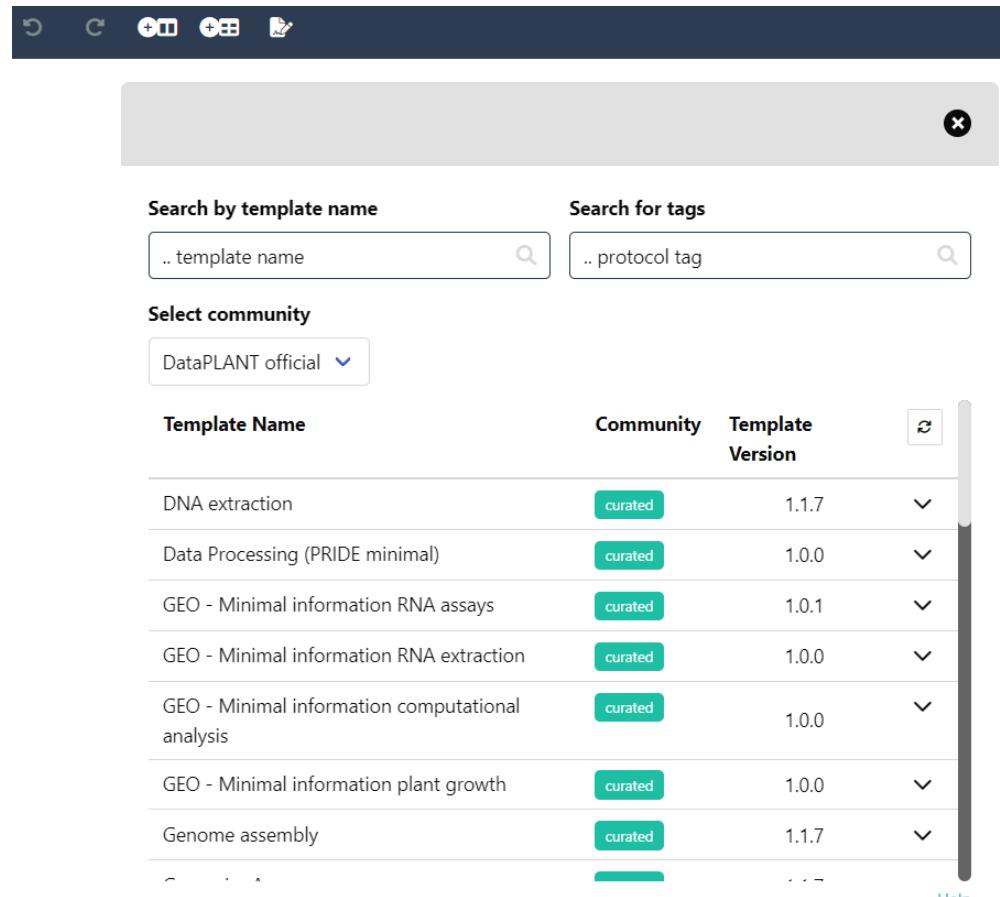
💡 Here you can find DataPLANT and community created workflow annotation templates

### 1. Search for **RNA extraction** and click

select

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

### 2. Click **Add template** to add all *Building Blocks* from the template to your table, which do not exist yet.



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×

PreviewUpdate 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

## 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.

**Your ISA table is ready** 

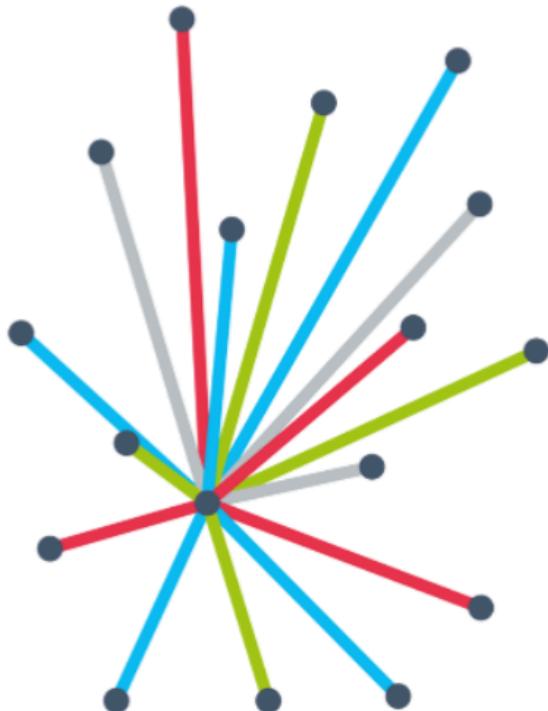
Go ahead, adjust the *Building Blocks* you want to use to describe your experiment as you see fit.

Insert values using Swate Term search and add input and output.

## Known issues with ARCitect and Swate (April 2024)

 Swate Desktop outdated

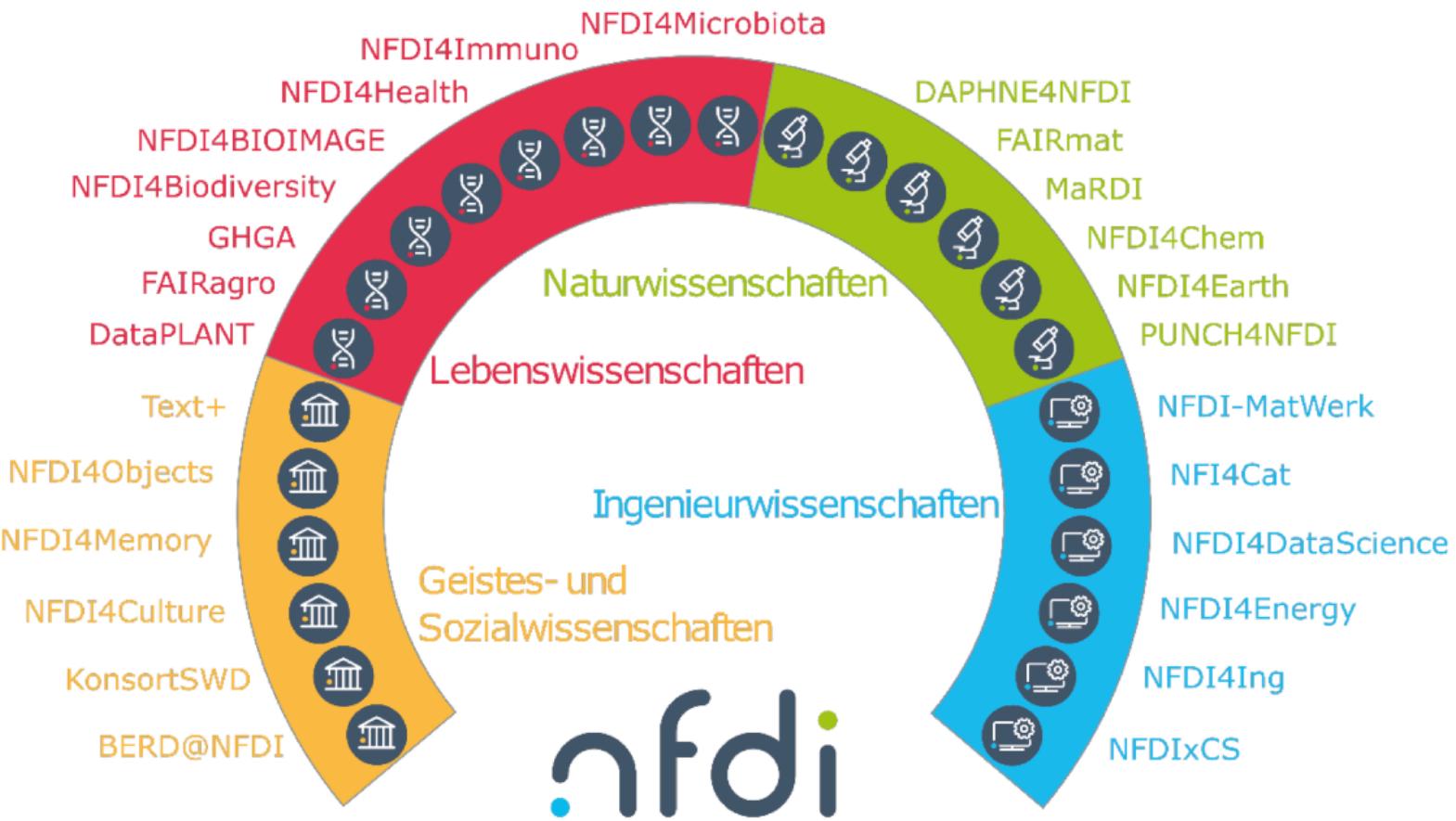
# NFDI – Nationale Forschungsdaten Infrastruktur



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](http://nfdi.de)

# NFDI consortia



nfdi.de

# NFDI sections



## Section-metatada

(Meta)data, Terminologies, Provenance



## Section-infra

Common Infrastructures



## Section-edutrain

Training and Education



## Section-ELSA

Ethical, Legal and Social Aspects



## Section-industry

Industry Engagement

nfdi.de

# Base4NFDI



[www.base4nfdi.de](http://www.base4nfdi.de) ●●●●●●

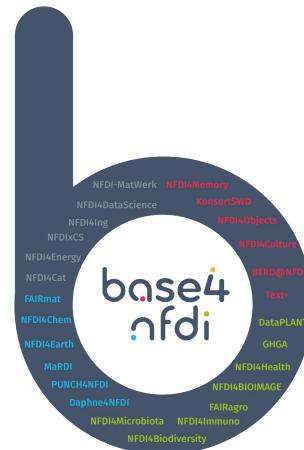
## What is Base4NFDI?

- Base4NFDI is a joint initiative of all 26 NFDI-**consortia**.
- Our goal is to establish reliable NFDI-wide **basic services** for **FAIR** research data management.
- Such basic and technical services will **benefit all communities and domains**.
- We ensure **interoperability** with national and international infrastructures and engage with science policy bodies.

## What is a basic service?

A **basic service** is a technical-organizational solution which typically includes:

- storage and computing services
- software, processes and workflows
- personnel support for different service desks



## Basic services under development

**PID4.nfdi**

Persistent Identifier Services  
for the German National Research  
Data Infrastructure



**IAM4.nfdi**

Identity and Access Management  
for the German National Research  
Data Infrastructure



**TS4.nfdi**

Terminology Services  
for the German National Research  
Data Infrastructure

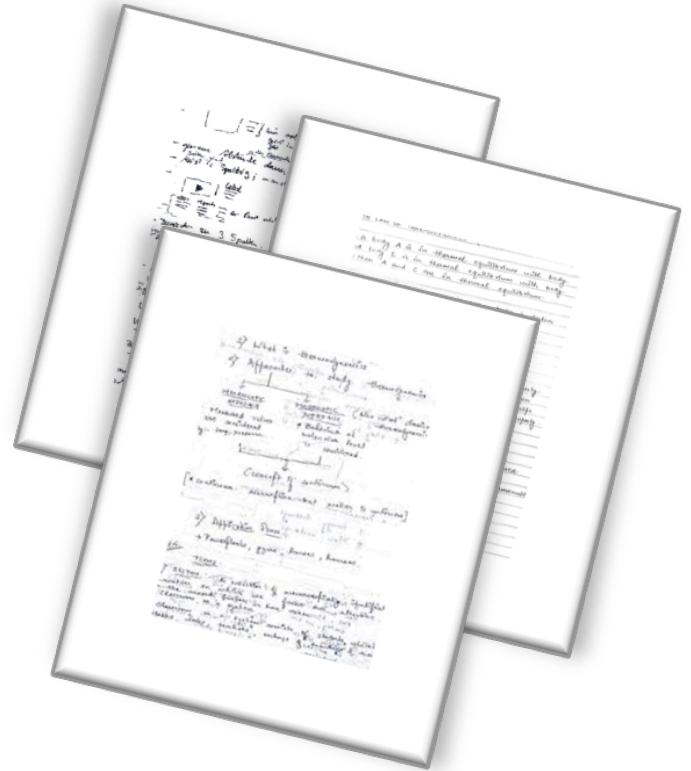


# Electronic Lab Notebooks (ELNs)

# ELN ≈ Digital Protocol Editors

- Documenting daily lab routine
- Lab methods & protocols
- Lab inventory (biologicals, chemicals, instruments)
- Local sharing & collaboration
- Backup (locally)

 ELNs help to digitalize research documentation



# Are ELNs FAIR?

FAIR indicator*	ELN
<b>Findable</b>	
F1. (Meta)data are assigned a globally unique and persistent identifier.	
F2. Data are described with rich metadata (defined by R1 below).	
F3. Metadata clearly and explicitly include the identifier of the data they describe.	
F4. (Meta)data are registered or indexed in a searchable resource.	
<b>Accessible</b>	
A1. (Meta)data are retrievable by their identifier using a standardised	

# Findable

FAIR indicator*	elabFTW
F1. (Meta)data are assigned a globally unique and persistent identifier.	 / 
F2. Data are described with rich metadata (defined by R1 below).	
F3. Metadata clearly and explicitly include the identifier of the data they describe.	 / 
F4. (Meta)data are registered or indexed in a searchable resource.	

# Accessible

FAIR indicator*	elabFTW
A1. (Meta)data are retrievable by their identifier using a standardised communications protocol	<span style="color: red;">●</span>
A1.1 The protocol is open, free, and universally implementable	<span style="color: green;">●</span> / <span style="color: orange;">●</span>
A1.2 The protocol allows for an authentication and authorisation procedure, where necessary	<span style="color: red;">?</span>
A2. Metadata are accessible, even when the data are no longer available	<span style="color: red;">?</span>

# Interoperable

FAIR indicator*	elabFTW
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	●
I2. (Meta)data use vocabularies that follow FAIR principles.	●
I3. (Meta)data include qualified references to other (meta)data.	●

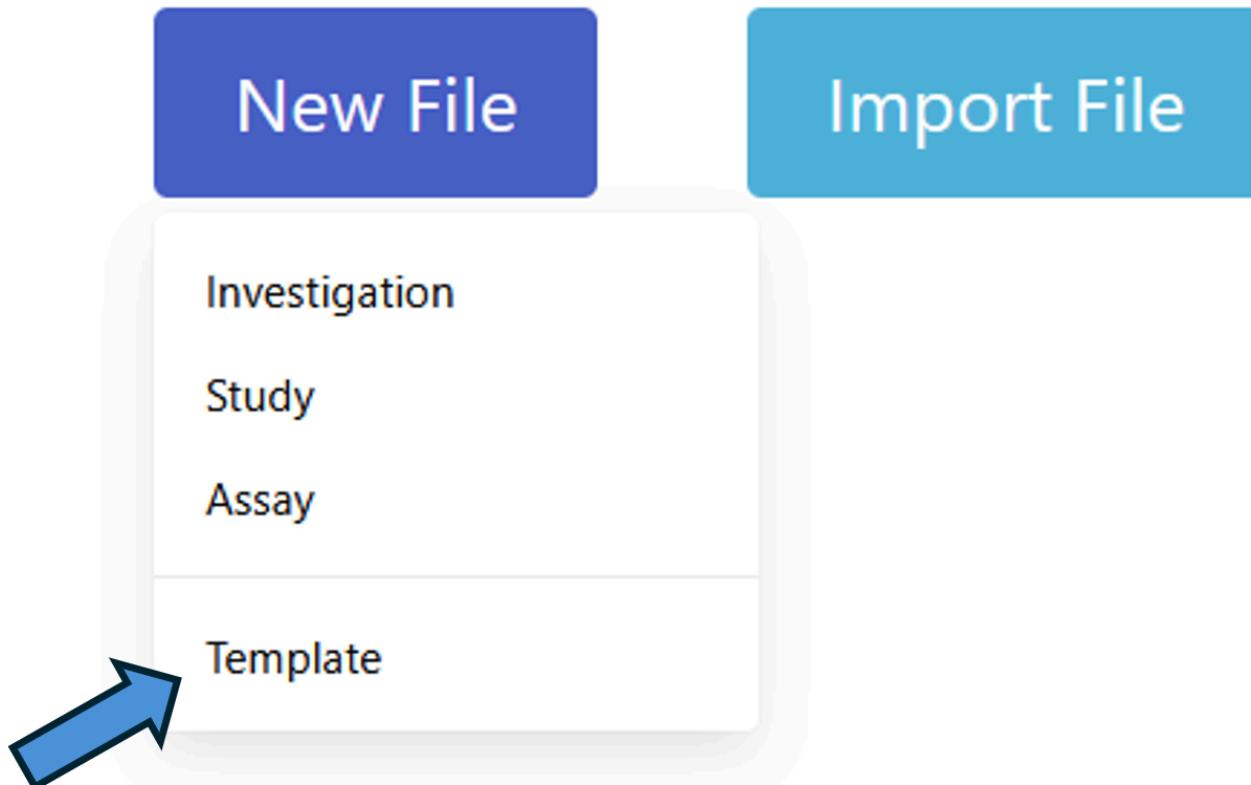
# Reusable

FAIR indicator*	elabFTW
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	
R1.1. (Meta)data are released with a clear and accessible data usage license	
R1.2. (Meta)data are associated with detailed provenance	
R1.3. (Meta)data meet domain-relevant community standards	

# Creating Swate Templates

# Create or open a Swate template

Open [Swate-alpha](#). To create a new template click "New File" and select "Template". If you want to edit an existing template, click "Import" and select the respective file.



# Add Template Metadata

- Once you open a new template, you will see a "Metadata" sheet.
- Enter a name for the template
- Add a description about the template

 Do not change the **Identifier**

# Template Metadata

Key	Definition	Tip 
Identifier		<p>Do not change this field. It maps your template to a database entry</p> <p></p>
Name	This is the first info Swate users see about your template	<p>Try using a short, descriptive and human readable name. Capitalize the first letter of the first word.</p>
Description	Here you can describe your template.	<p>Users interested in your template can read this in Swate, but not</p> <p></p>

# Template Metadata Example

Example template metadata and how it helps in Swate's template search

<b>Identifier</b>		
a927fd4c-851f-4a69-8aa0-fc680f495a64		
Guid should contain 32 digits with 4 dashes following: xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx. Allowed are a-f, A-F and numbers.		
<b>Name</b>		
DNA extraction		
<b>Description</b>		
Template to describe the extraction of DNA.		
<b>Organisation</b>		
DataPLANT		
<b>Version</b>		
1.1.7		
<b>Last Updated</b>		
21.03.2024, 02:53		
<b>Tags</b>		
1. Term Name	TSR	TAN
<input type="text" value="extraction"/>	<input type="text" value="OBI"/>	<input type="text" value="OBI:0302884"/>

Template Name	Community	Template Version	
DNA extraction	curated	1.1.7	▼

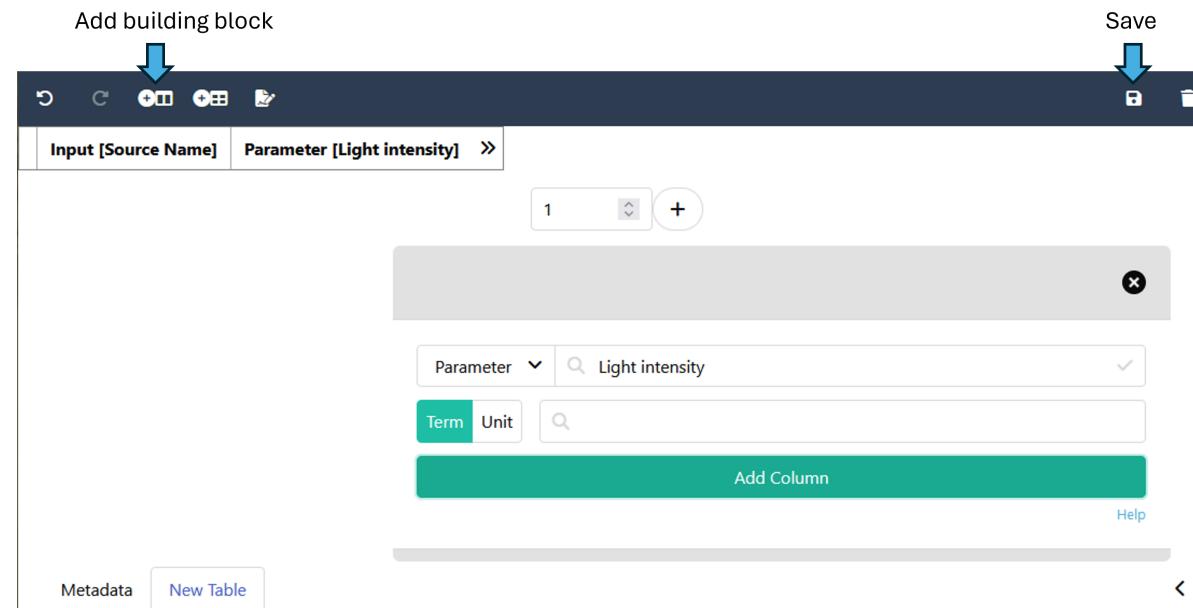
Template to describe the extraction of DNA.  
**Author:** Angela Kranz, Dominik Brilhaus  
**Created:** 2024/03/19  
**Organisation:** DataPLANT

extraction DNA

select

# Add building blocks

- To add building blocks to the actual template, switch from the "Metadata" sheet to the "New Table" sheet on the bottom left.
- Using the "Add Building Block" button, you can add your desired building blocks with or without a unit.



# Recommendations for template design

- Keep the template as concise as possible
- If you miss a term or ontology, please follow the [DPBO contribution guide](#) to let us know
- If you add a template to address a missing method, try to add building blocks that cover experimental procedures (as Parameters) and features of the sample (as Characteristics) that the experimenter would use when working on an experiment of that type
- The typical order of the columns is: **Input** -> (all the Parameters and Characteristics in between in chronological order) -> **Output** -or- **Raw Data File** -or- **Derived Data File**. Try to think about in which order the experimenter in the lab will do their work. Try to match this chronological order from left to right. This step is optional and only meant to increase readability.

# Recommendations for template design (continued)

- Avoid using the building block type Factor in templates. Any given characteristic or parameter in one study or assay can become a factor in another study or assay depending on the experimental context or scientific question.
- Background info on your template can be shared alongside the template via the [Swate-templates GitHub repository](#)

# Recommendations for endpoint repository templates

- Please follow the following naming pattern: "Repository" - "Assay", e.g. MetaboLights - MS measurement
- The endpoint repository tag should be added ONLY in the "Endpoint repository" category in the metadata sheet
- By default repository templates should contain **only mandatory** information. In this case, please add "mandatory" as a tag.
- Optional or recommended information can be added with an extension template that contains only the optional/recommended information. In this case, please add "-extension" to the template name.
- If applicable, templates should be split into different assays.
- Templates should be checked for validity of requirements every ~ 6 months.

# Save your template

- You can download and save your template as xlsx-file by clicking on the disc symbol in the top right corner.
- Please adjust the name of the file to correspond to the name of the template, but use underscores instead of spaces.
- Save the file in a suitable folder within your local clone of the Swate templates repository.

# Your template is ready for upload



- Well done! You created a new template.
- You can now submit your template via GitHub
- Once your pull request is merged, you will receive an Email from "Swobup Commit Report"

# GitHub Workflow to add or edit templates

1. (Recommended) [Create an issue](#) with background information about the template you want to add. This also serves as a place for discussion.
2. [Fork the Swate-templates repository](#).
3. (Recommended) Create a feature branch (e.g. "template-xy") on your fork.
4. [Clone](#) your fork-branch
5. Add/update **ONE** template ([see steps below](#))
6. Commit, push and [sync](#) your branch.
7. Open a [pull request](#)
  - | You can **reference your issue** typing `#` in the pull request's commenting dialog

## Known pitfalls with Swate Templates

- Opening and saving a Swate template .xlsx file with a program other than Microsoft Excel (e.g. LibreOffice, python script, R script) often destroys the template (backend). Please, avoid to upload this file into the GitHub repository.

# 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>

