



CEPLAS

Cluster of Excellence on Plant Sciences

CEPLAS ARC Training

May 24th, 2024

Dominik Brilhaus (CEPLAS)



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.

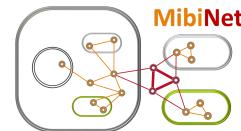
1. Slack Workspace for ad hoc support
2. Monthly user meeting (2nd Friday of the month)

→ [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



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>Findable Accessible Interoperable Reusable</p>	<p>Easier collaboration & sharing</p> <pre>graph TD; A(()) --> B(()); A(()) --> C(()); A(()) --> D(()); B(()) --> E(()); C(()) --> E(()); D(()) --> E(());</pre>	<p>Increased findability and visibility</p>	<p>Reproducibility</p>
<p>Added-value to the research community</p> <p>nfdi NCBI EMBL-EBI</p>	<p>Compliance with funding policies</p> <pre>graph LR; A[Checkmark Document] --- B[DFG]; A --- C[EU Flag]</pre>	<p>Receive due credit</p> <pre>graph LR; A[Reuse] --- B[Citations]</pre>	<p>Saves time & workload</p> <p>FAIR</p> <p>Time wasted</p>

Is your data FAIR?

Findable | Accessible | Interoperable | Reusable

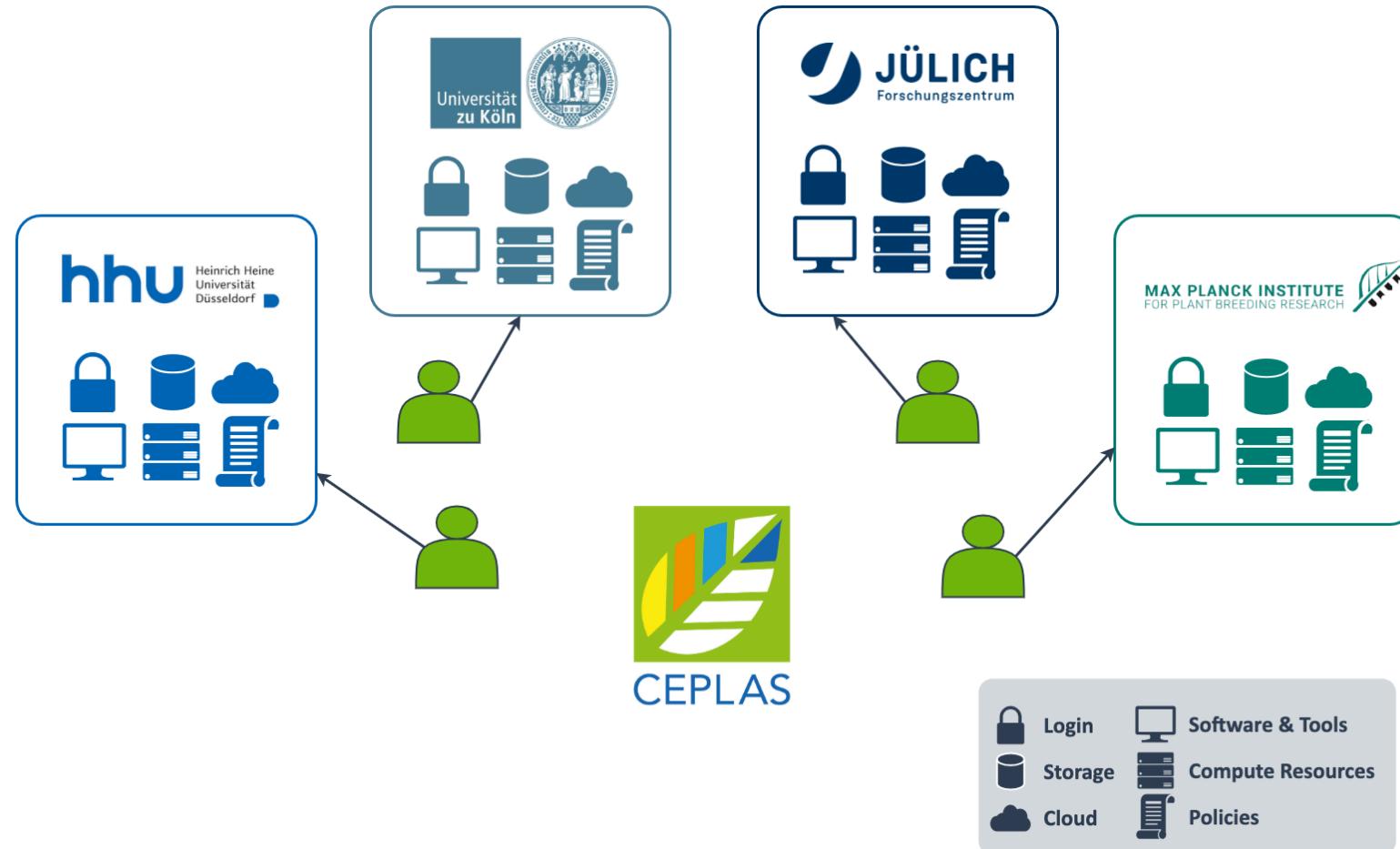
- Where do you store your data?
- How do you annotate your data?
- How do you share your data?
- What tools do you use to analyse your data?
- How do you reuse other people's data?



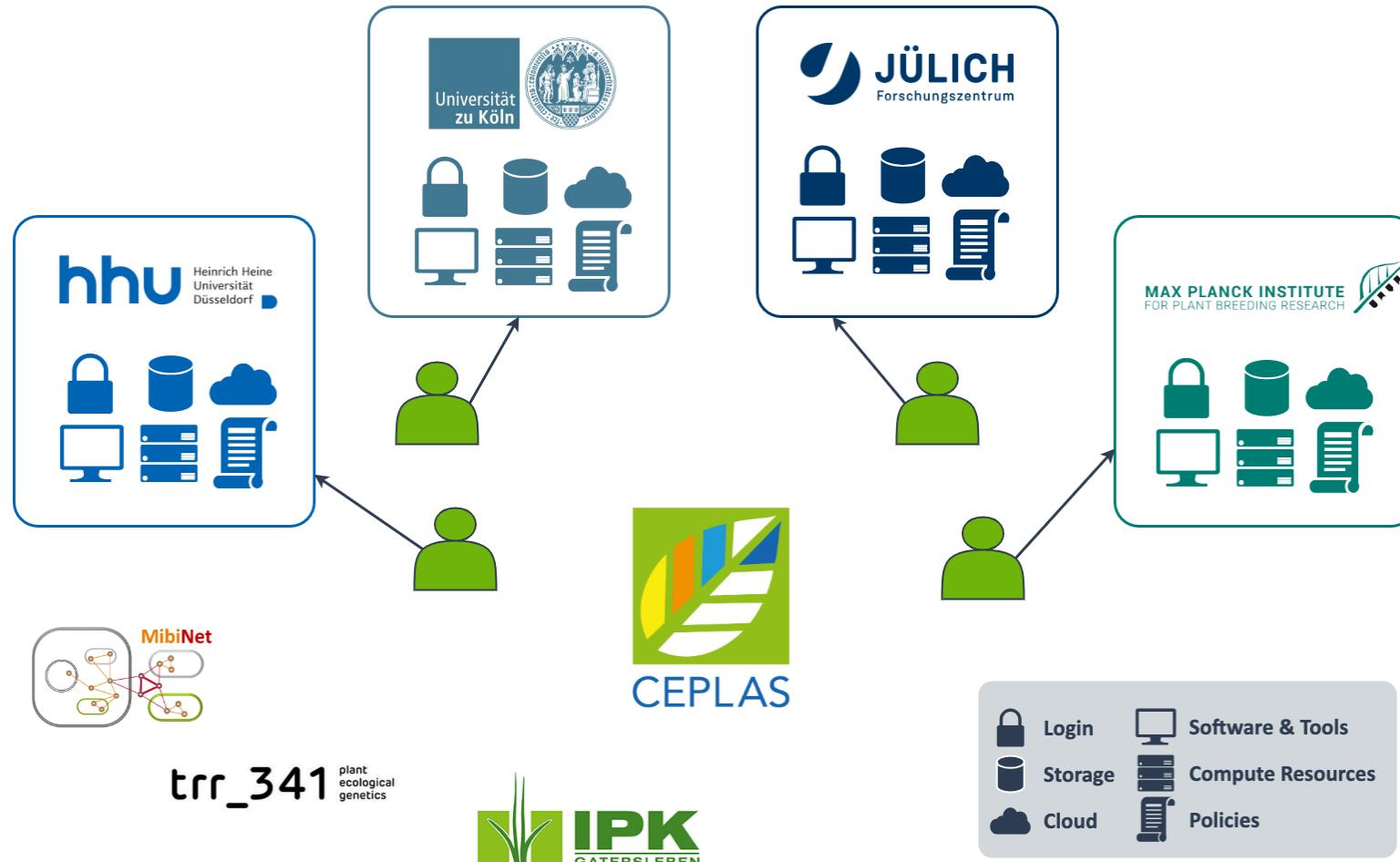
CEPLAS – One cluster, four locations



Data silos impede collaboration

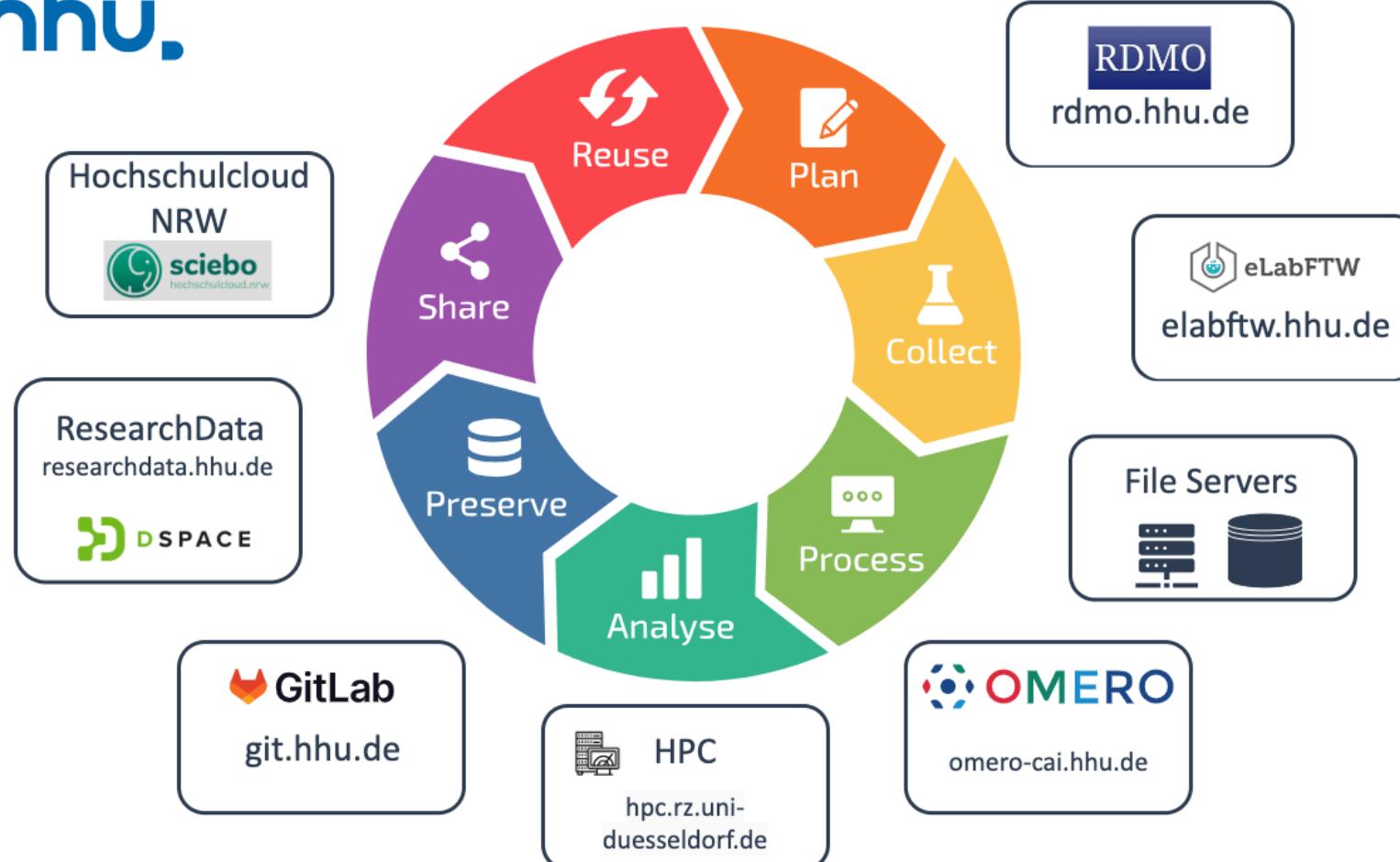


Data silos impede collaboration



Missing interfaces impede collaboration

hhu.



Data Stewardship between DataPLANT and the community

Community

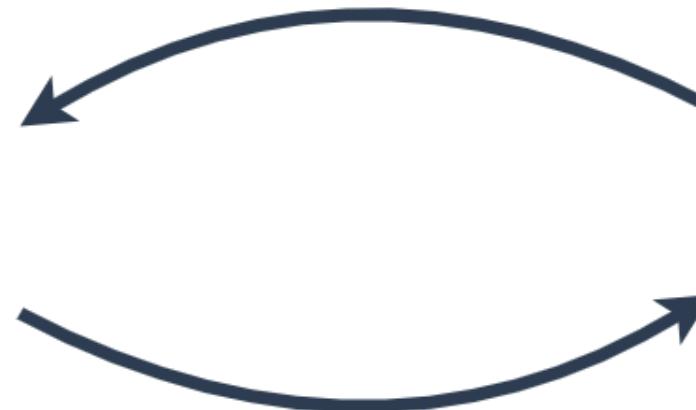


Domain experts
User experience
Training

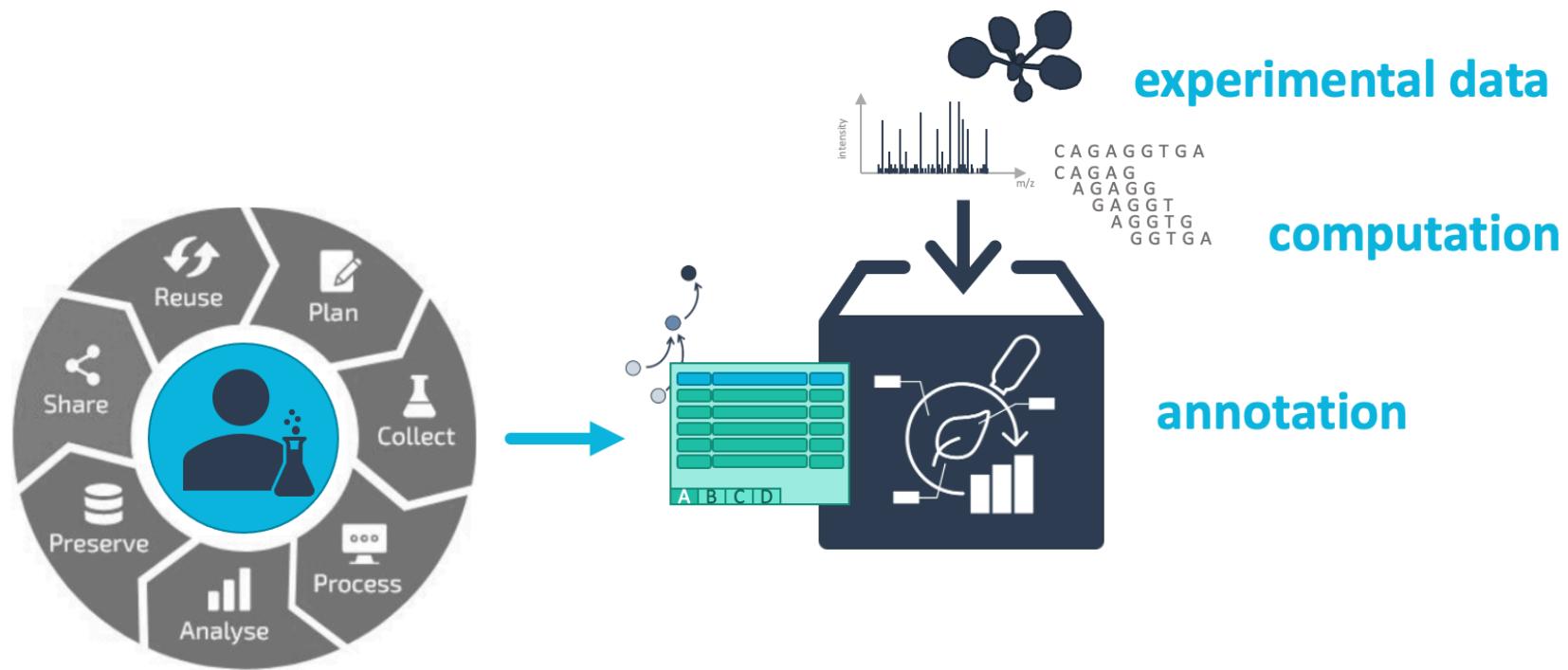
nfdi4plants



Service provider
Developers
Tech experts

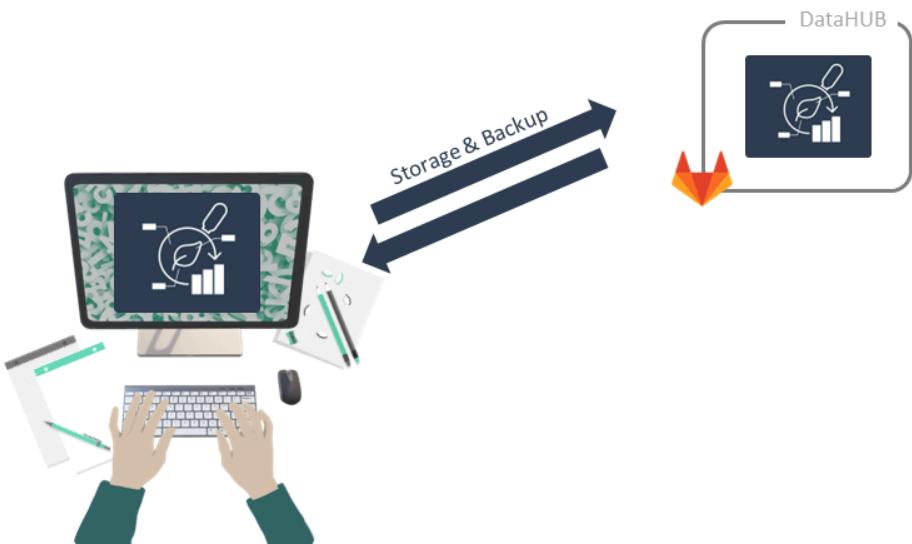


Annotated Research Context (ARC)

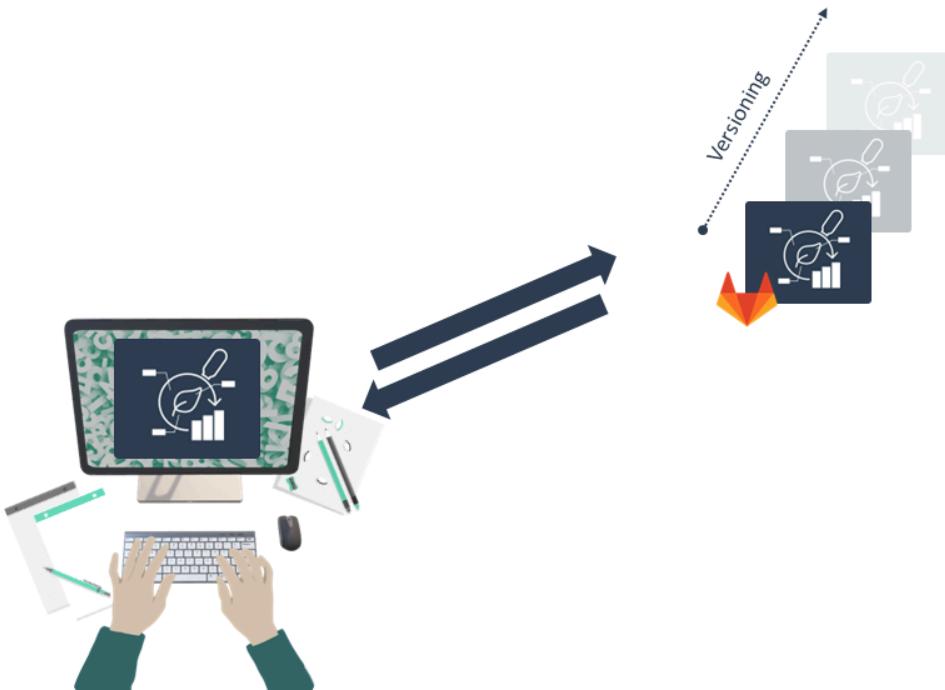


Your entire investigation in a single unified bag

You can store your ARC in the DataHUB



ARCs are versioned



You can invite collaborators



Collaborate and contribute



Reuse data in ARCs



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

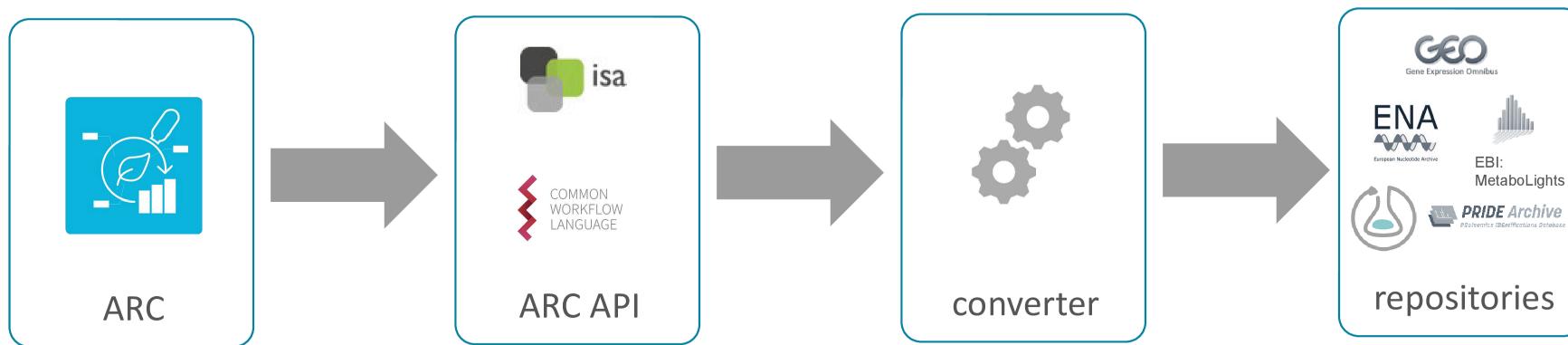
OpenAIRE

data publication

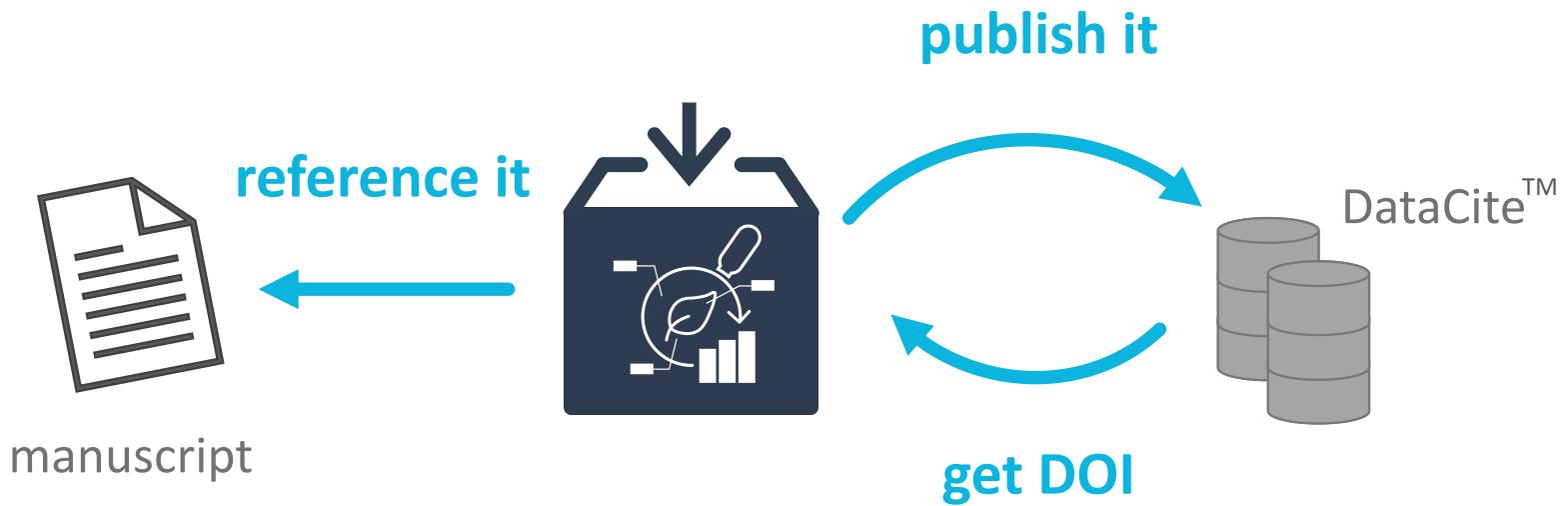
The Dataverse® Project

INVENIO

From ARC to repositories



Publish your ARC, get a DOI



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 buttons: 1. pipeline (passed), 2. Publish ARC (blue button with a heart icon), and 3. arc quality (yellow button showing 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

The screenshot shows a dataset page with the following interface elements:

- Top right:** Dataset, Open buttons.
- Header:** Edit, New version, Share buttons.
- Versions:** Version v1 (10.60534/9e5jx-75d83), Sep 7, 2023.
- Details:** DOI (10.60534/9e5jx-75d83), Resource type (Dataset), Publisher (DataPLANT).
- Export:** JSON, Export button.

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

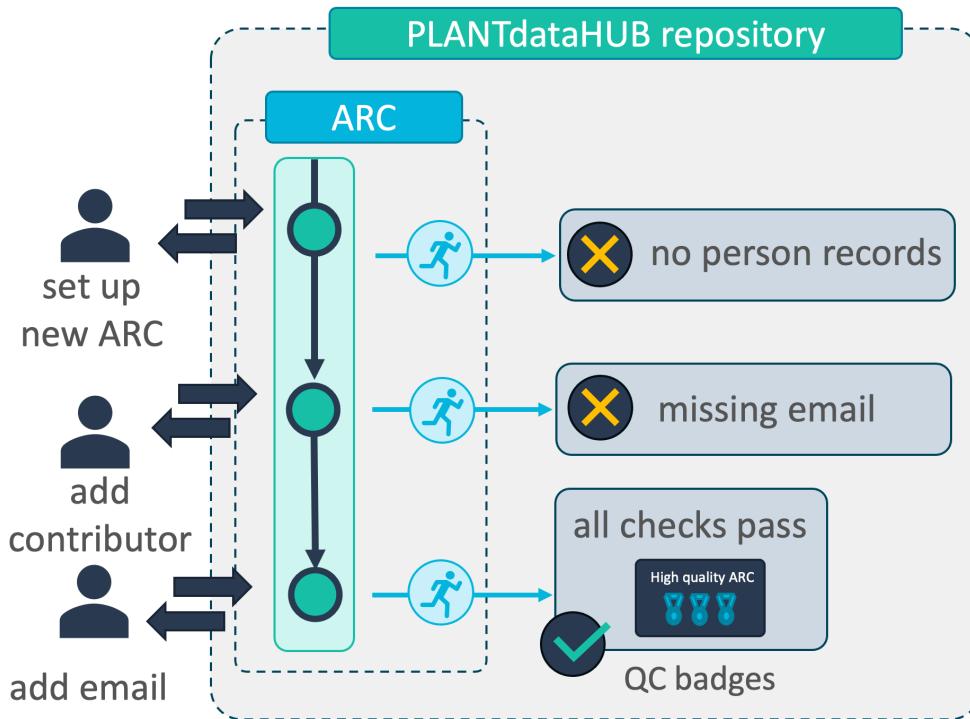
hosted on: <https://git.nfdi4plants.org/projects/122>

Files

The screenshot shows a file summary page with the following content:

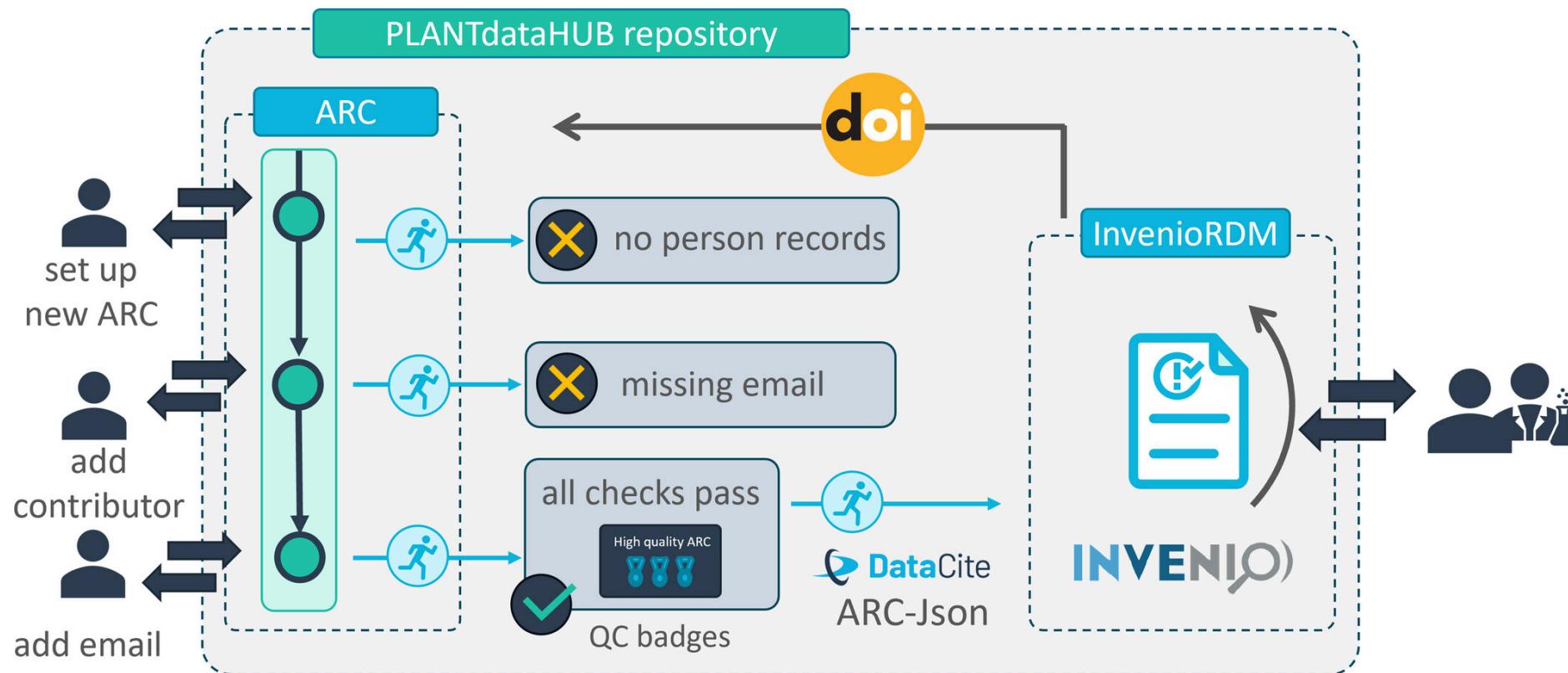
- File:** arc-summary.md
- Description:** [Data set] Systems-wide investigation of responses to moderate and acute high temperatures in the green alga Chlamydomonas reinhardtii.
- File contents:**
 - root
 - isa.investigation.xlsx
 - README.md
 - runs

Validation

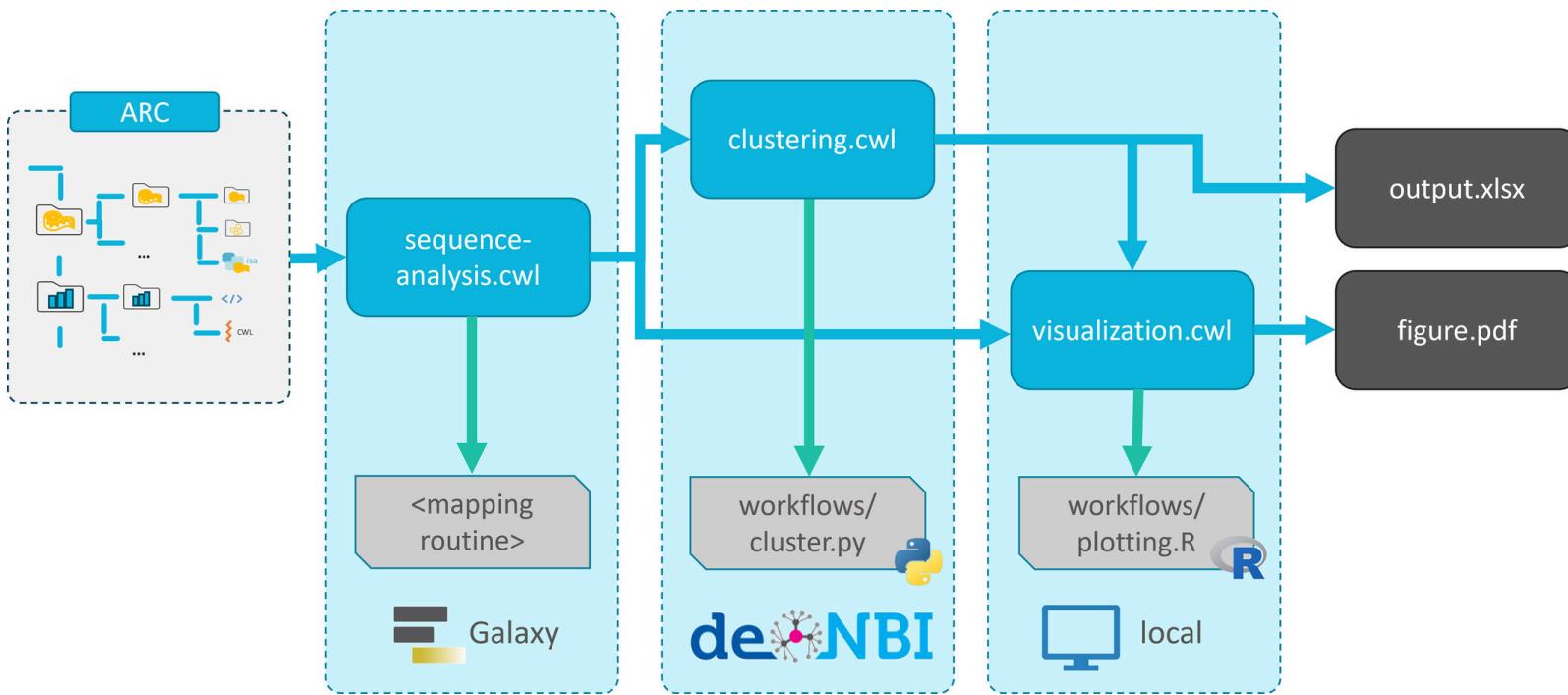


adapted from Weil, H.L., Schneider, K., et al. (2023), PLANTdataHUB: a collaborative platform for continuous FAIR data sharing in plant research. Plant J. <https://doi.org/10.1111/tpj.16474>

Validate & publish



Data analysis and workflows



Galaxy integration: Extra value for plant research

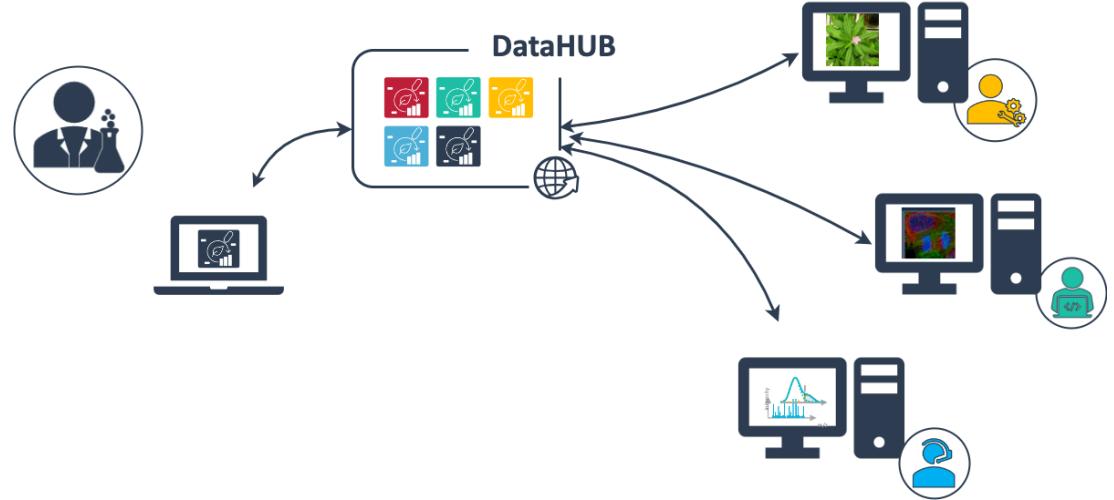


e.g. <https://plants.usegalaxy.eu>

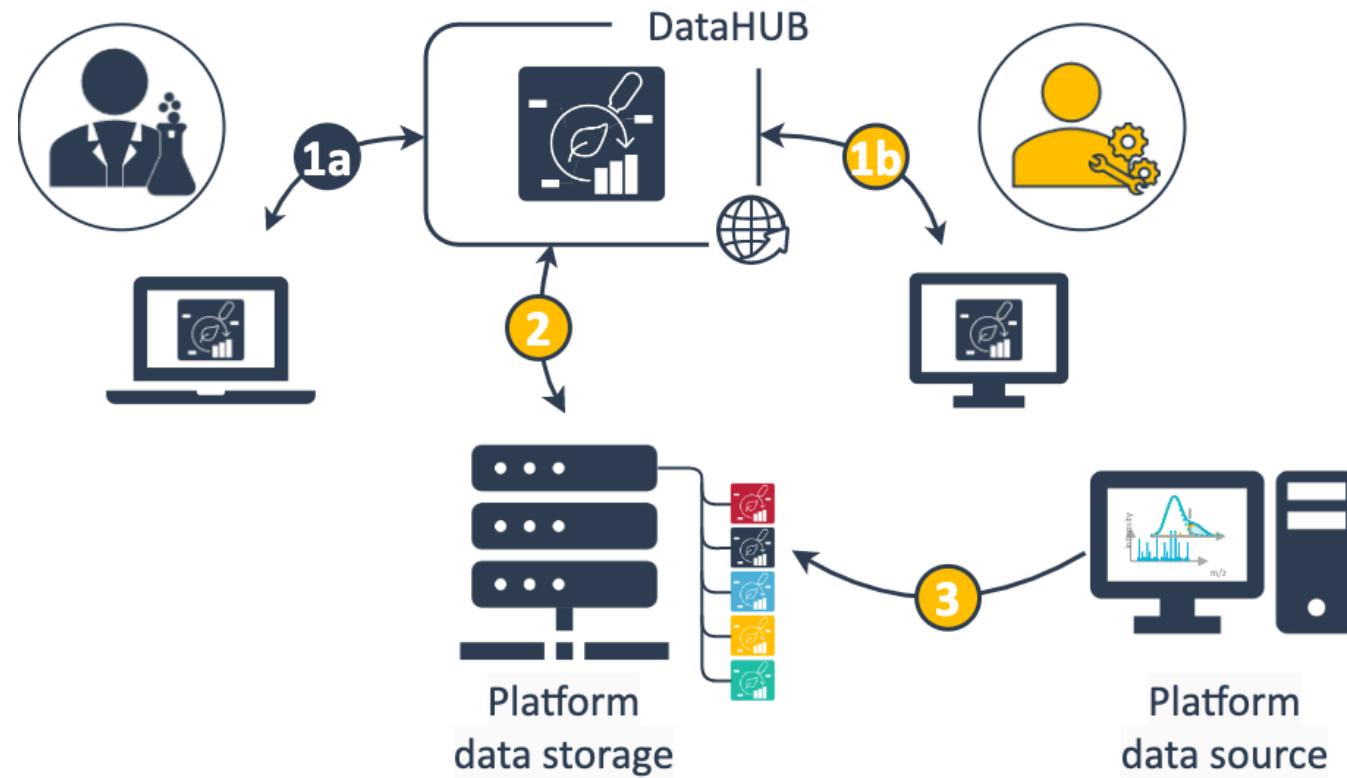
- Full ARC compatibility
- Automated metadata generation
- Specialized tools and workflows for ‘omics processing and analysis
- Public repository compatibility
- Galaxy teaching resource for data analysis

Enabling platforms

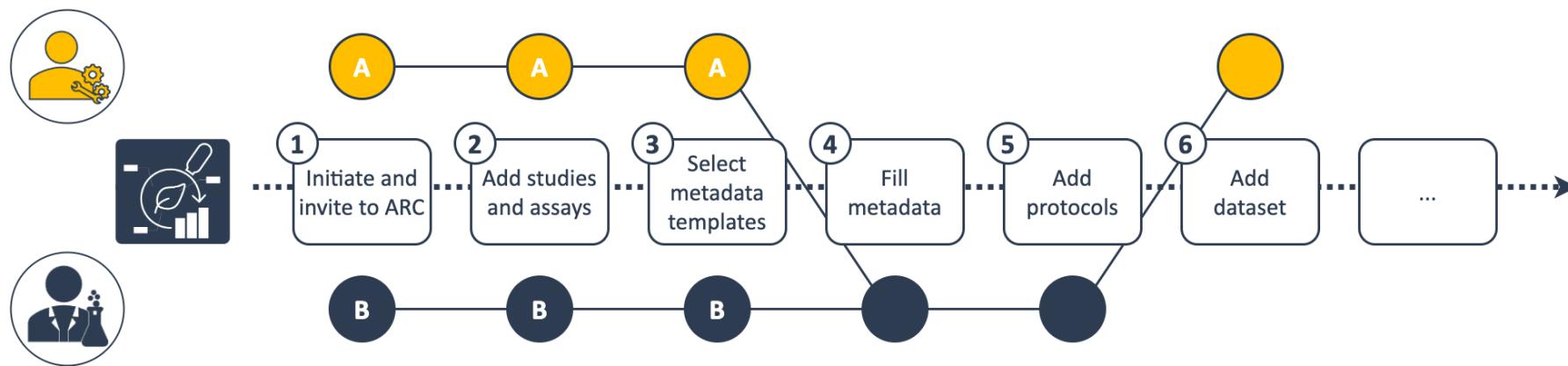
- Streamlined exchange of (meta)data
- Communication and project management



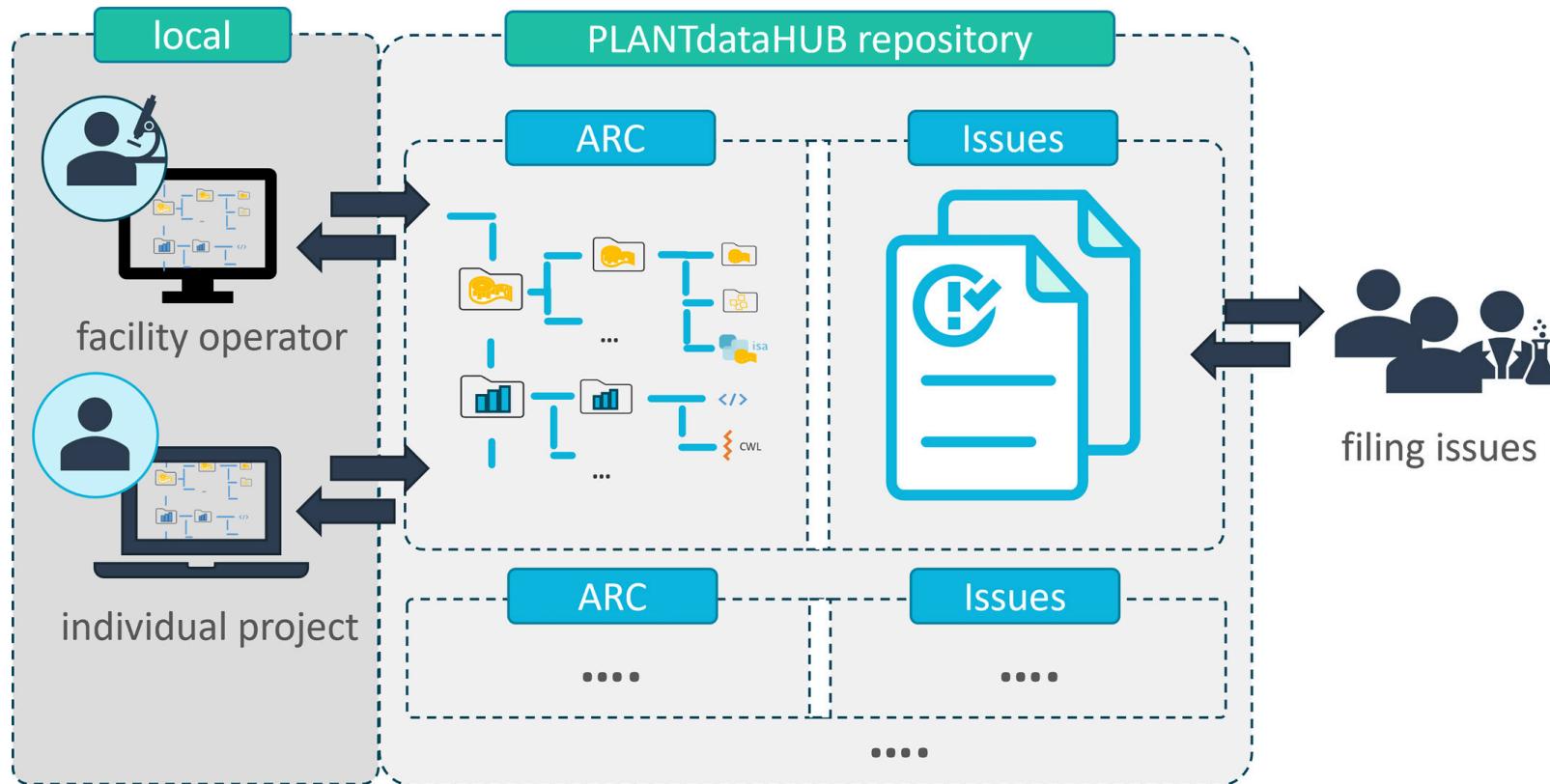
Streamlined data exchange



Meet your collaborators in an ARC

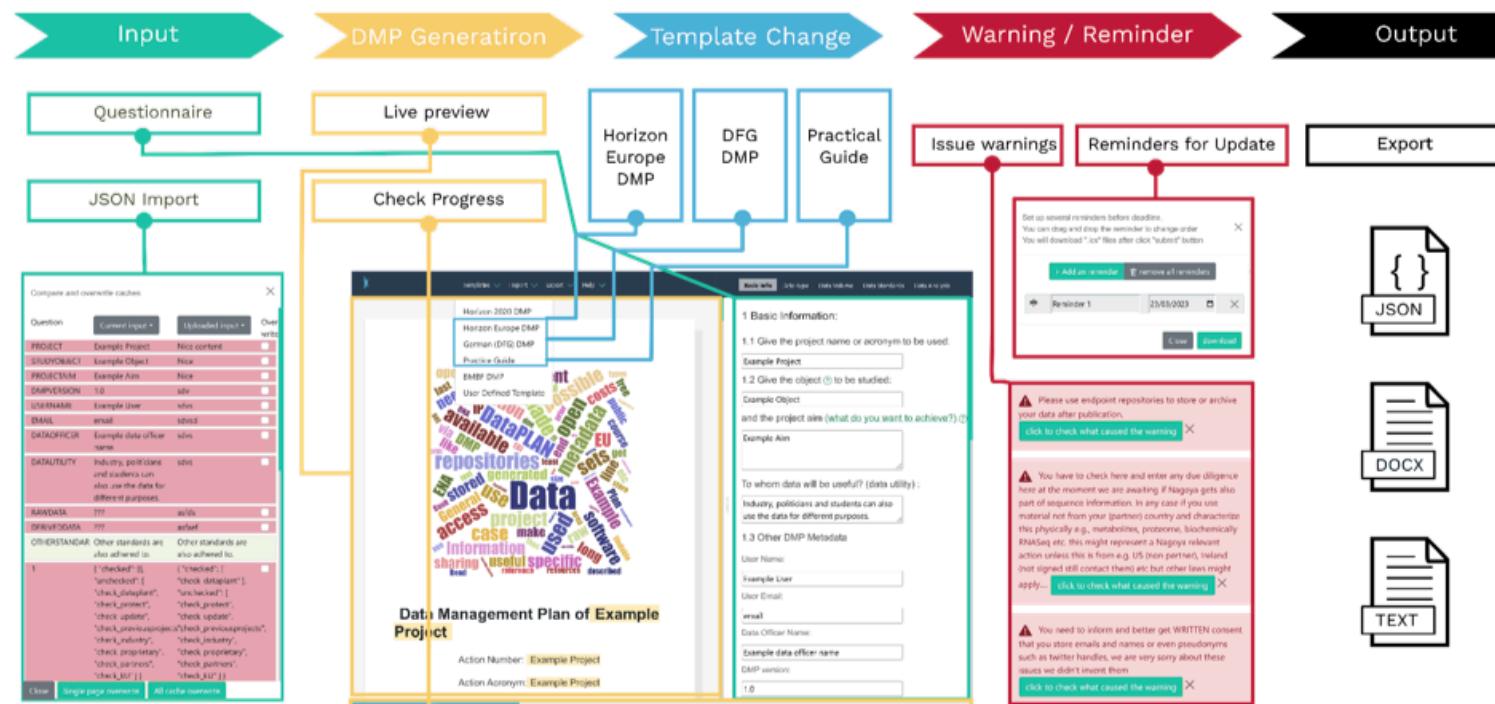


Project management



DataPLAN – a DataPLANT DMP generator

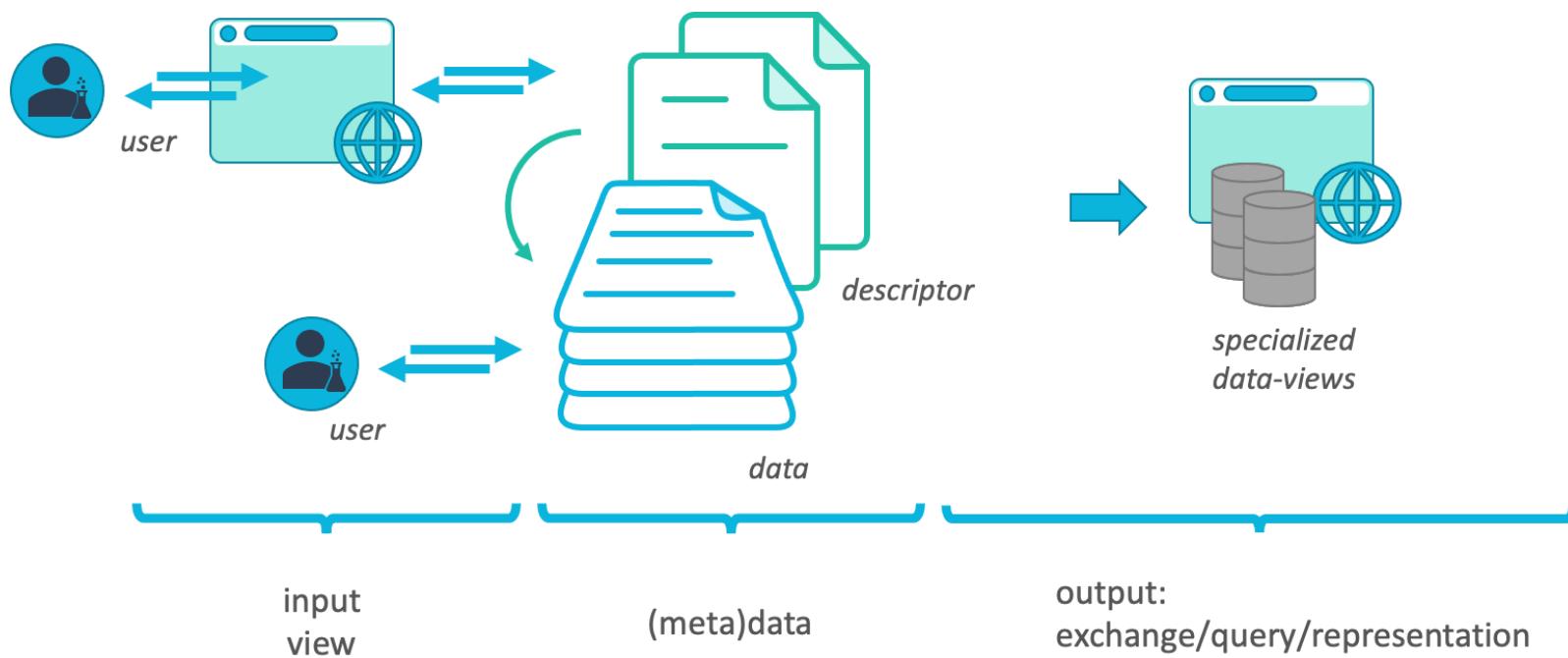
<https://dmpg.nfdi4plants.org>



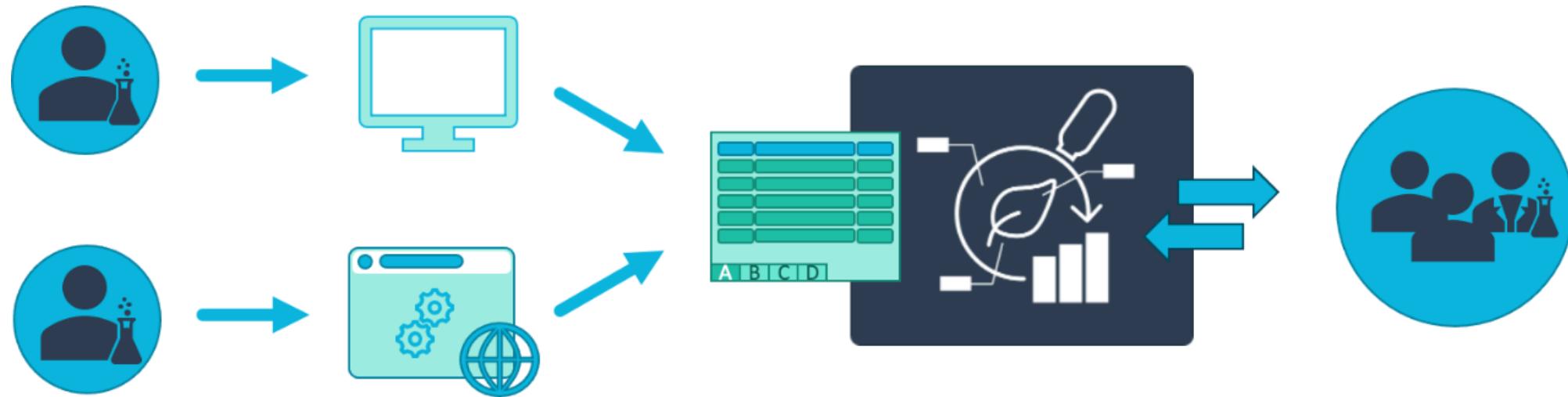
Zhou et al. (2023), DataPLAN: a web-based data management plan generator for the plant sciences, bioRxiv 2023.07.07.548147; doi: <https://doi.org/10.1101/2023.07.07.548147>

Everything is a file

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

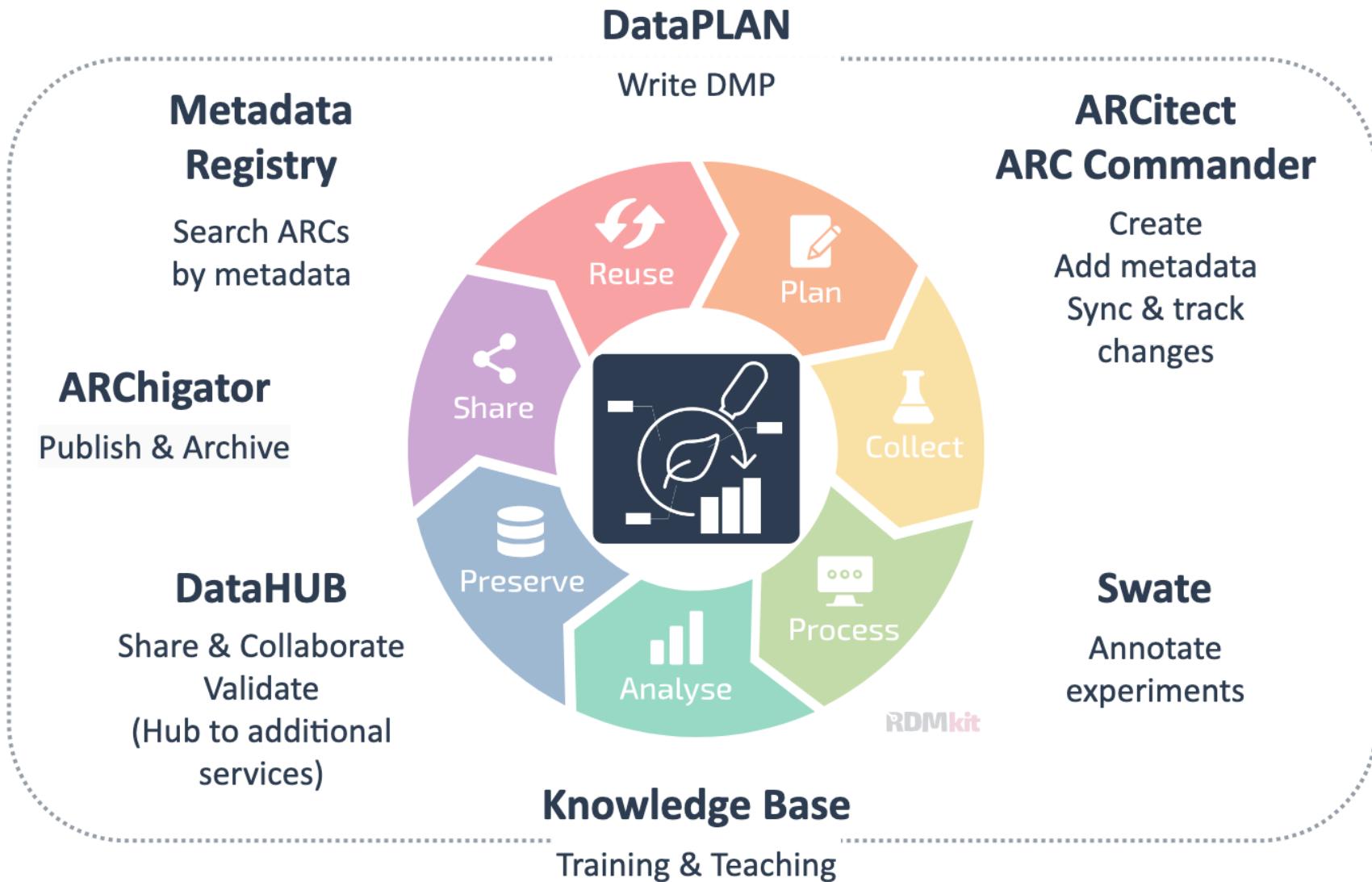


No technical lock-in

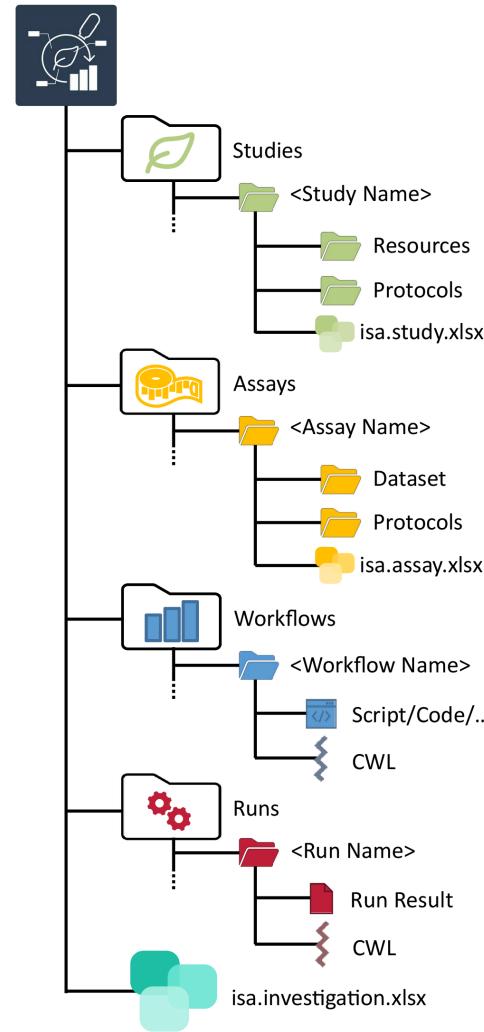


(Meta)data transparency with tool assistance but **no technical lock-in**

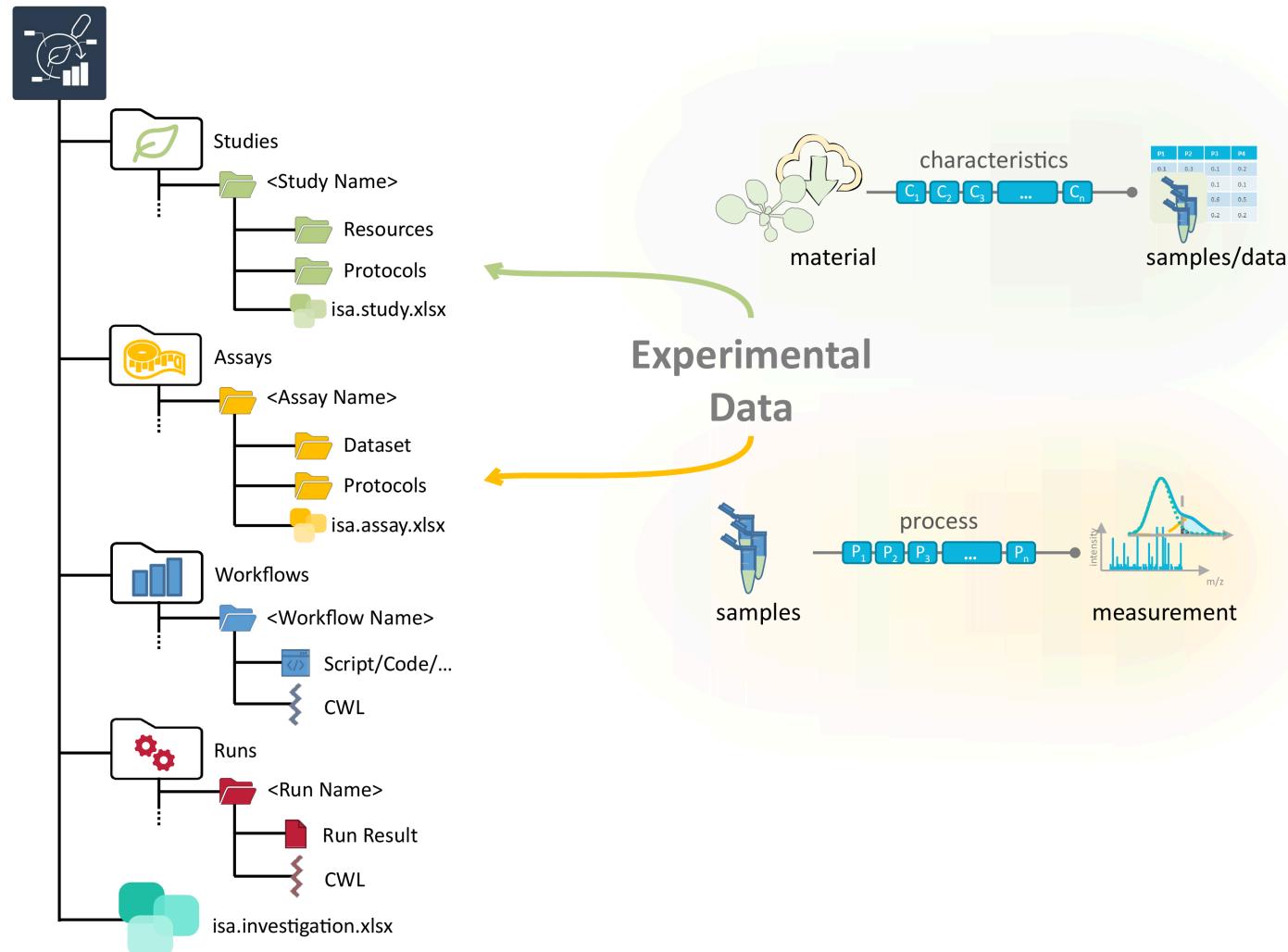
The ARC ecosystem



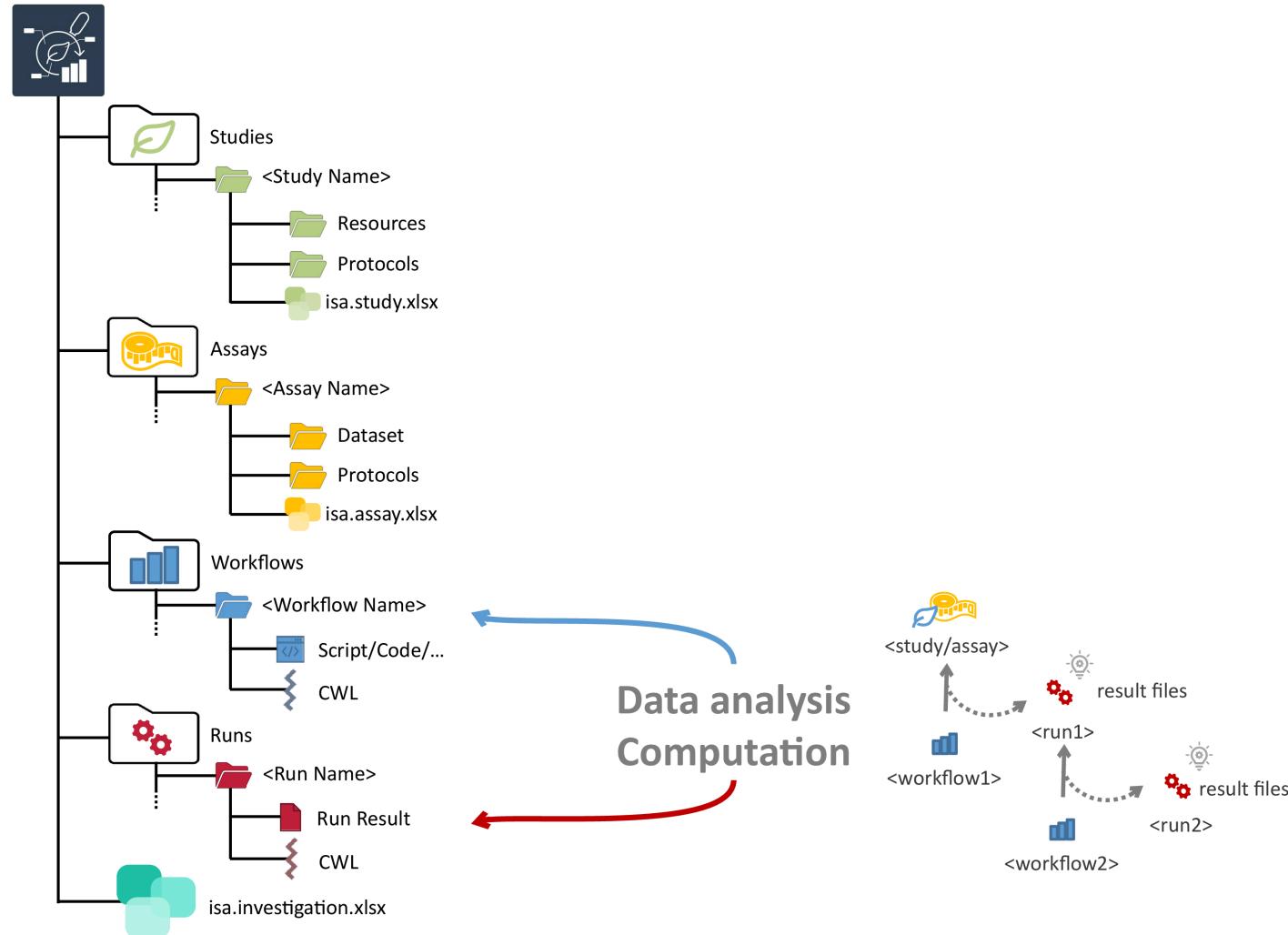
What does an ARC look like?



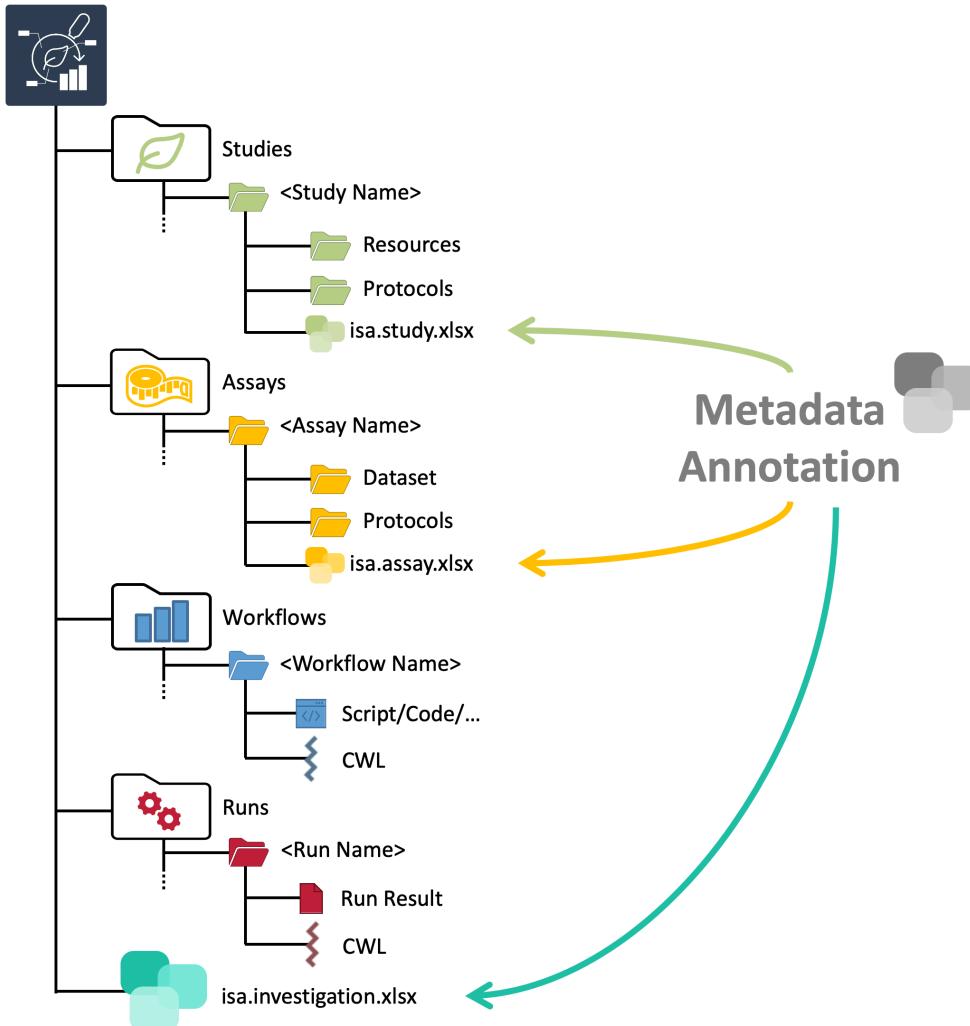
ARCs store experimental data



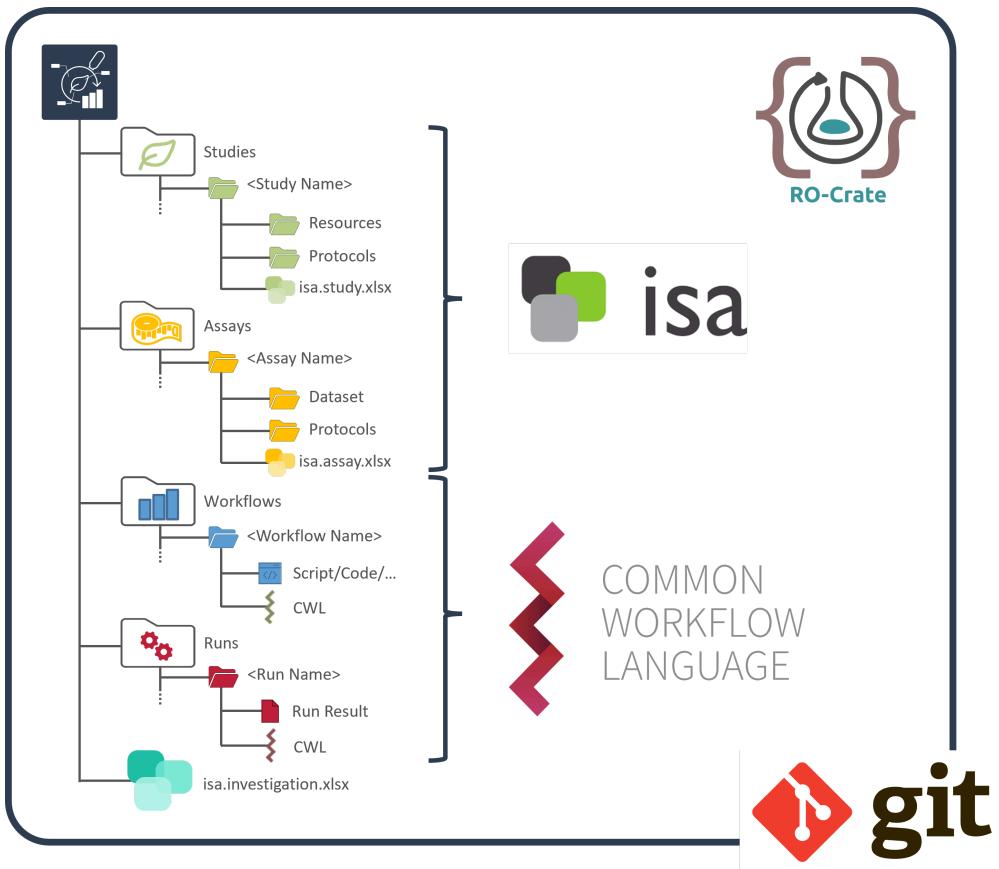
Computations can be run inside ARCs



ARCs come with comprehensive metadata

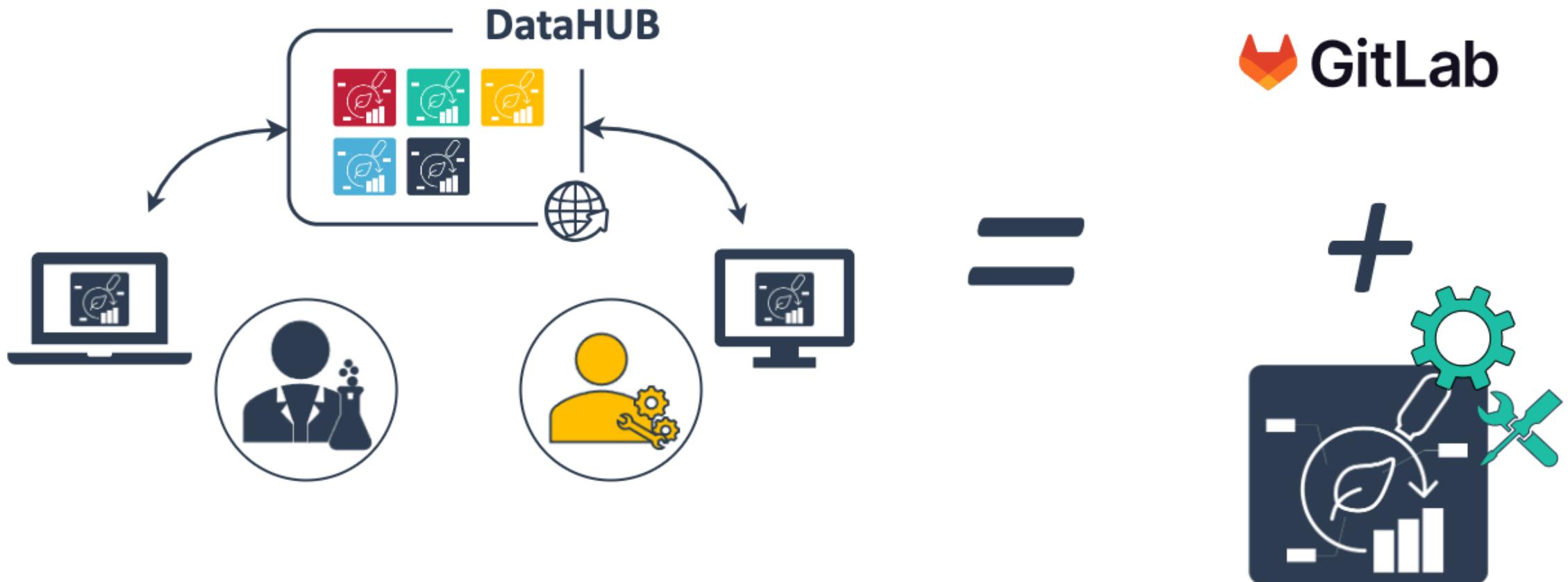


ARC builds on standards

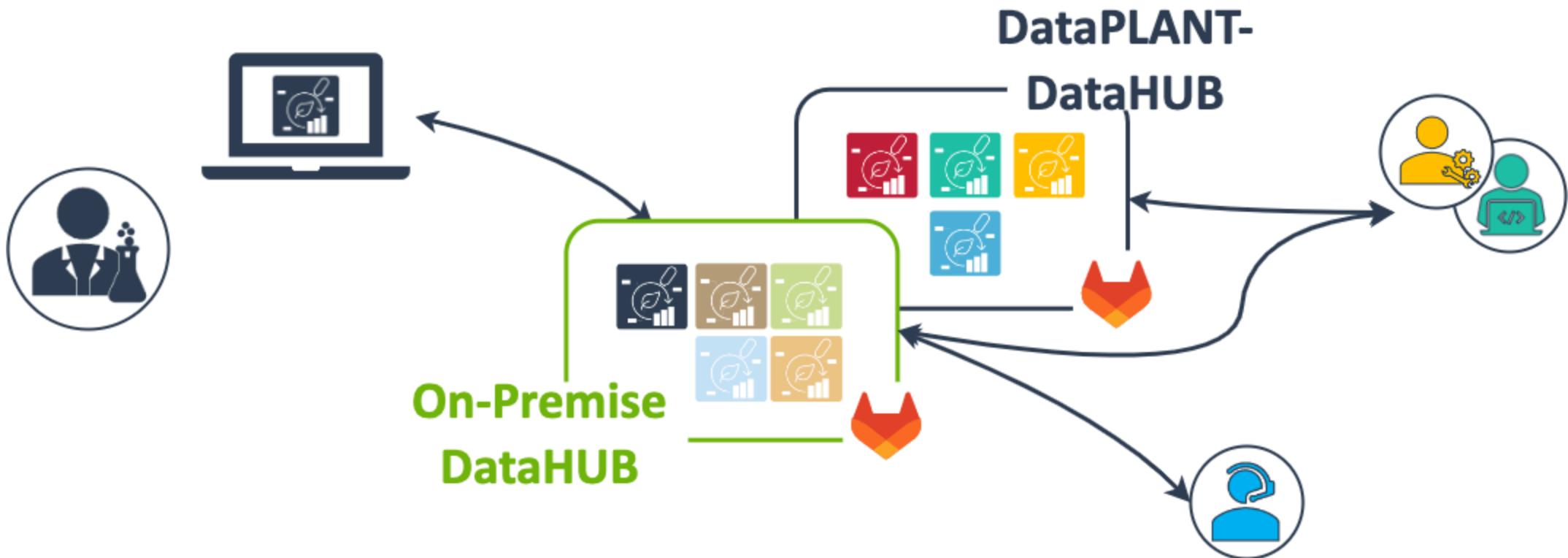


- RO-Crate: standardized exchange
- ISA: structured, machine-readable metadata
- CWL: reproducible, re-usable data analysis
- Git: version control

The DataPLANT DataHUB – a GitLab *Plus*

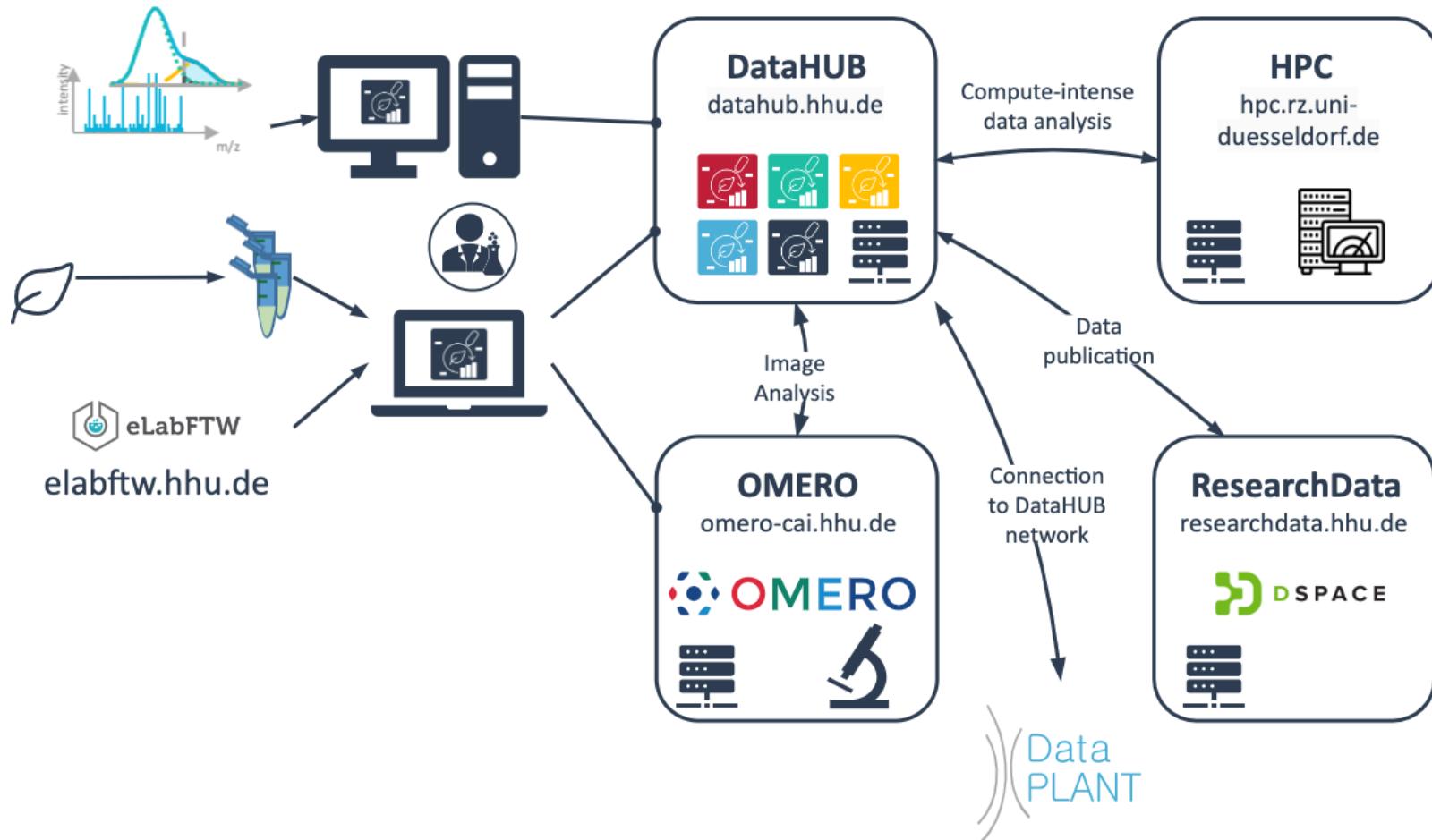


On-premise DataHUBs



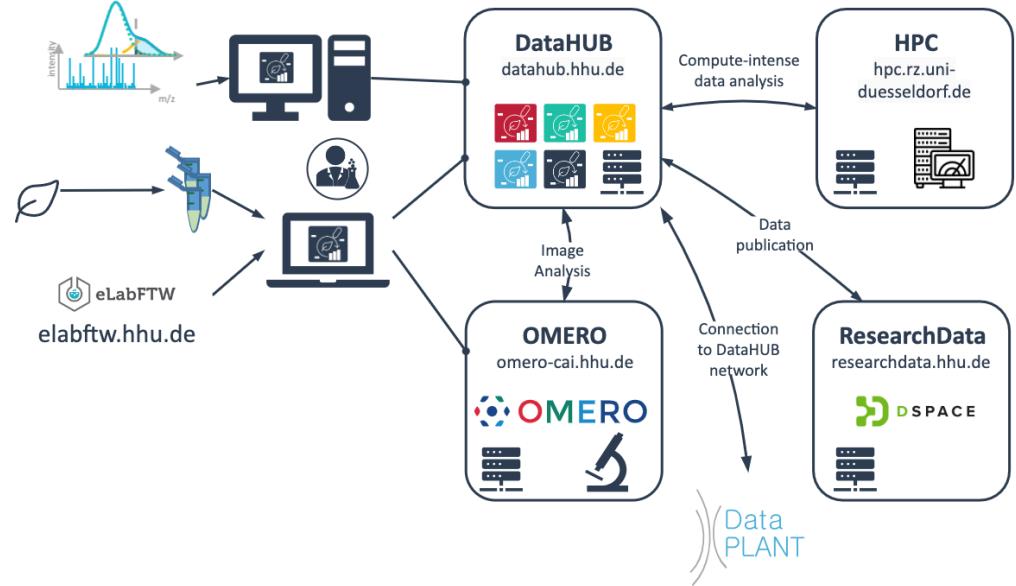
ARC services are available as on-premise option

HHU-DataHUB

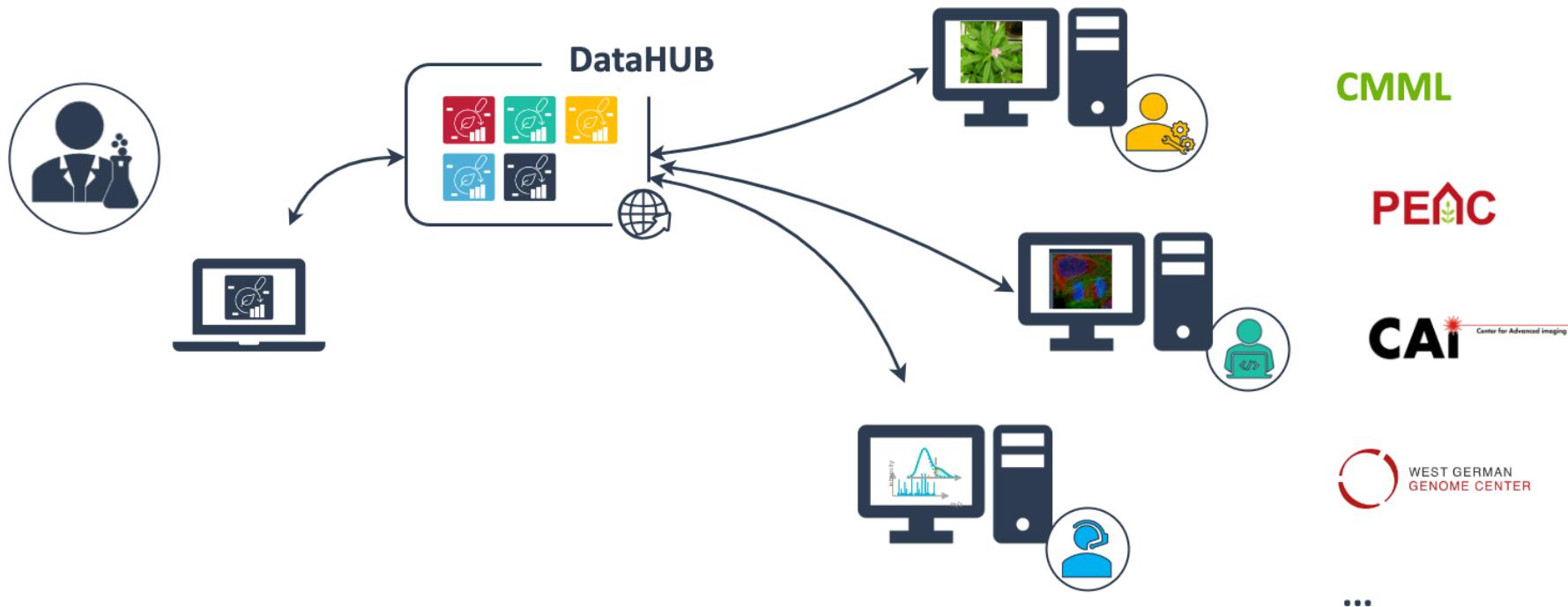


HHU-DataHUB

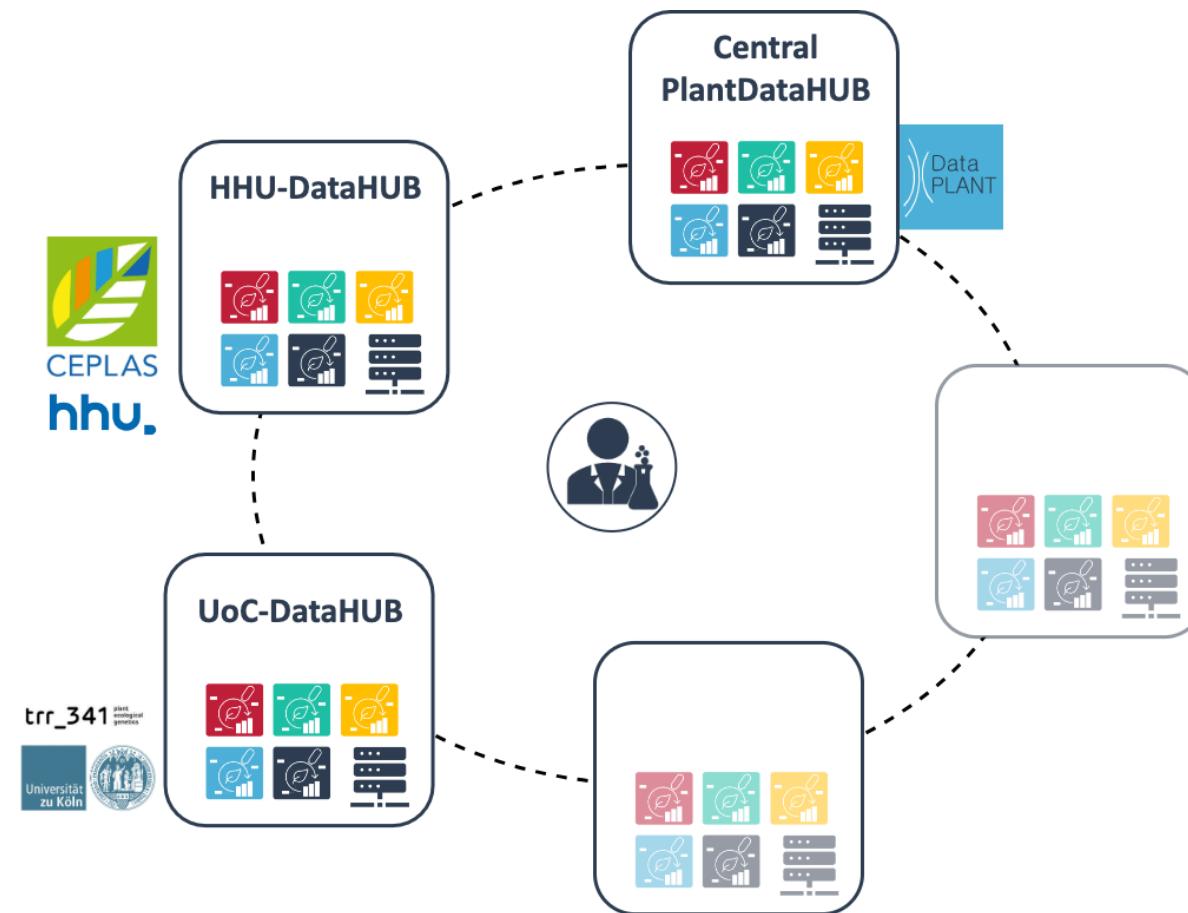
- FAIR data along the research data cycle
- Structured exchange of (meta)data
- Local (long-term) storage, backup, version control
- Integration of existing services and infrastructure
- External collaboration



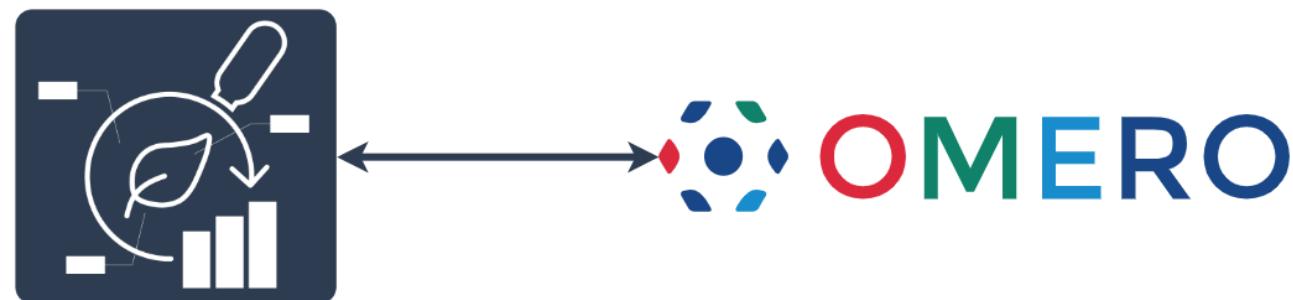
HHU-DataHUB – Node for platforms



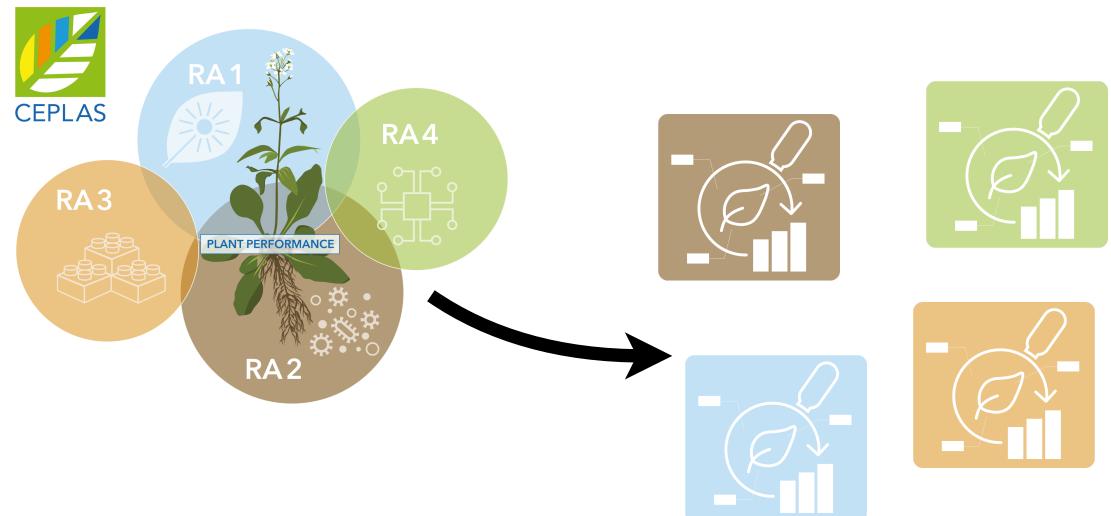
Embedded into an RDM network



CEPLAS collaborates with DataPLANT and NFDI4BIOIMAGE



CEPLAS Research Data Policy



ARCitect Hands-on

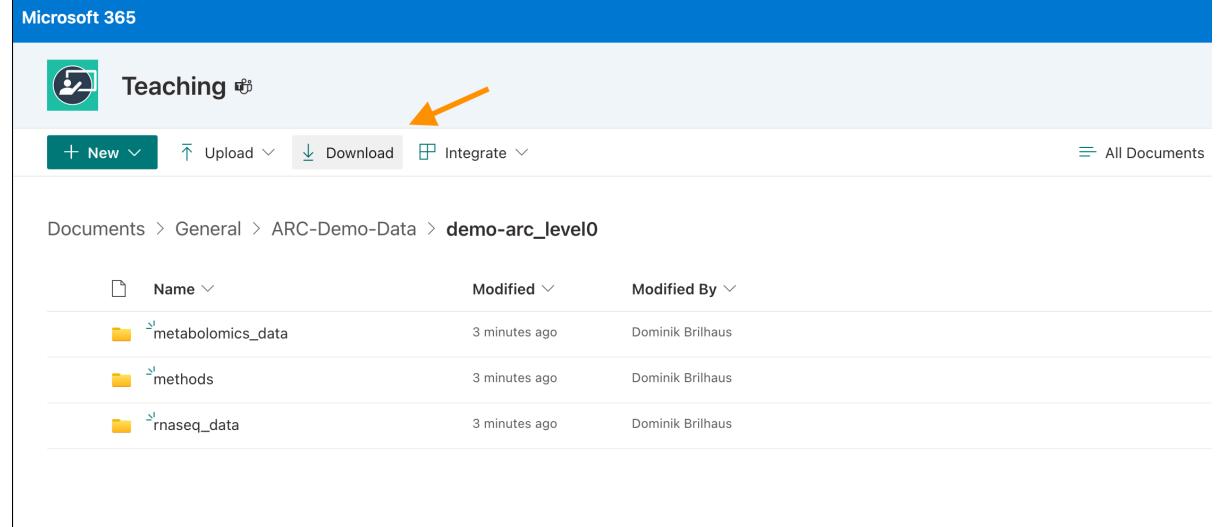
ARCitect installation

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

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

Download the demo data

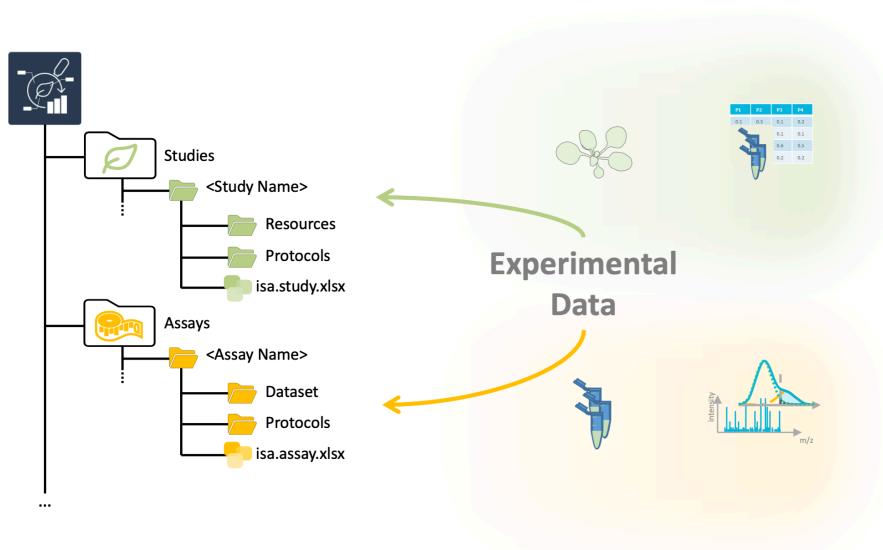
https://nfdi4plant.sharepoint.com/:f/s/Teaching/Eik7koJiMREgZ24kt07sIYBGxHmmZIS_Kzf7psk-5w-xg?e=u0sADd



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

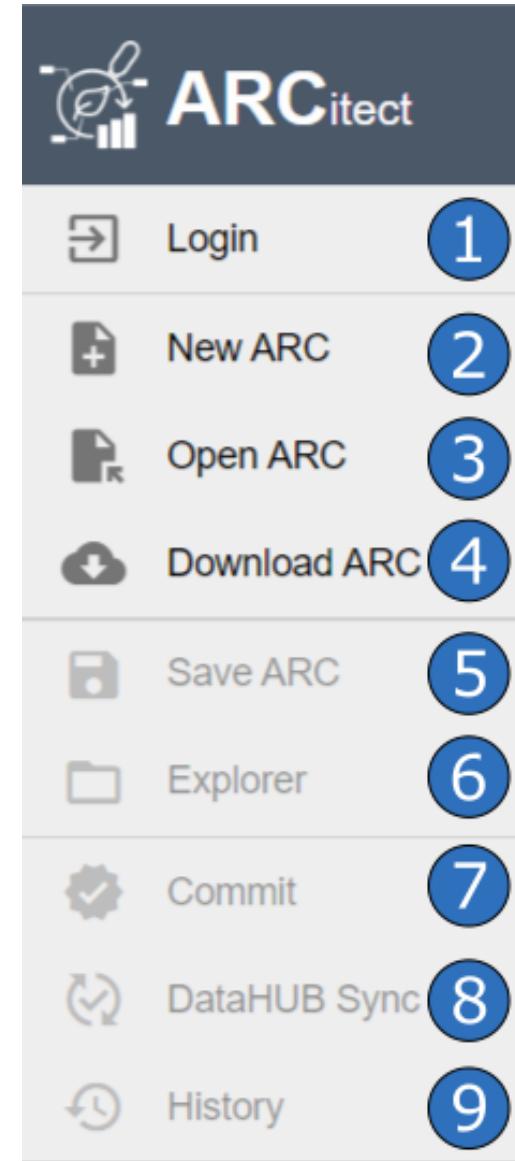
Open ARCitect

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

Please Select a DataHub

Host
git.nfdi4plants.org

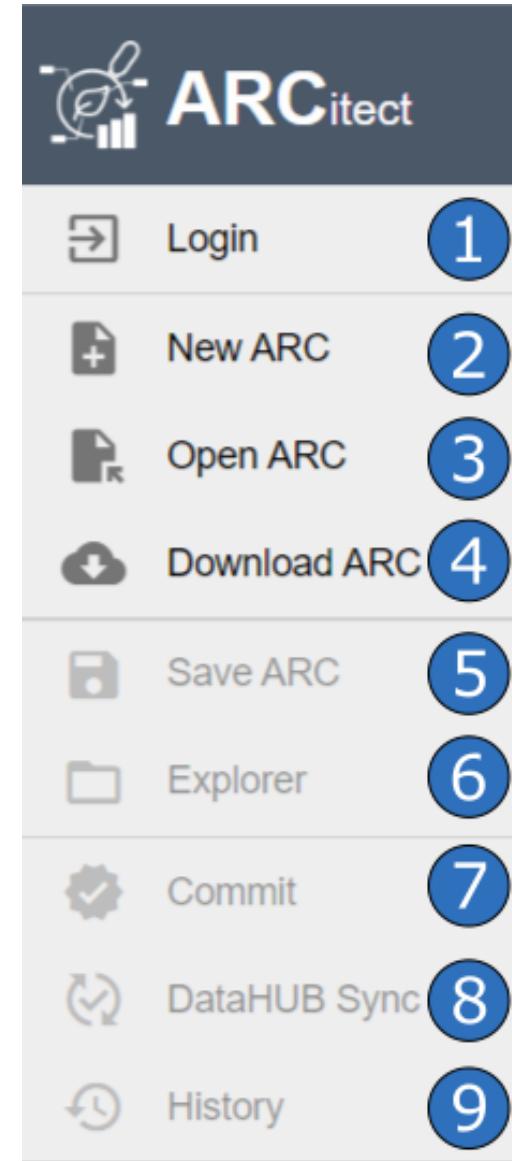
LOGIN CANCEL



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

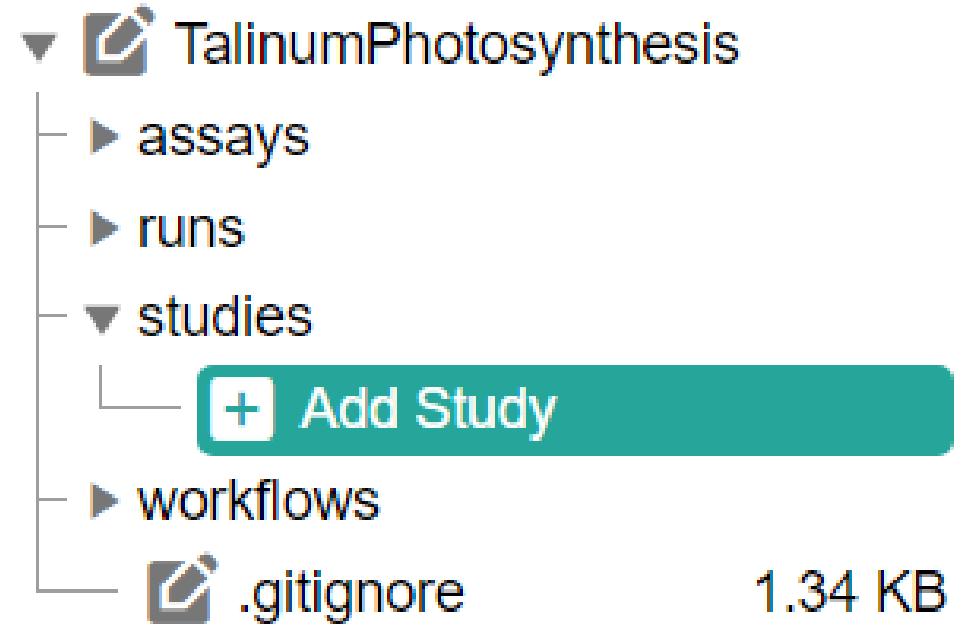
+

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 --:--

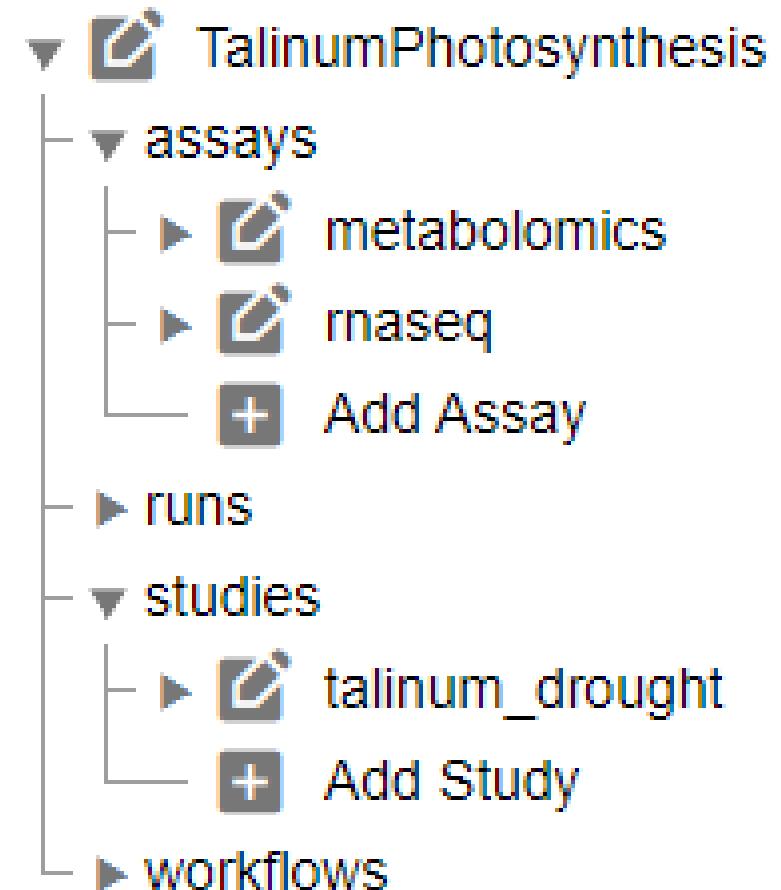
Public Release Date
tt.mm.jjjj --:--

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

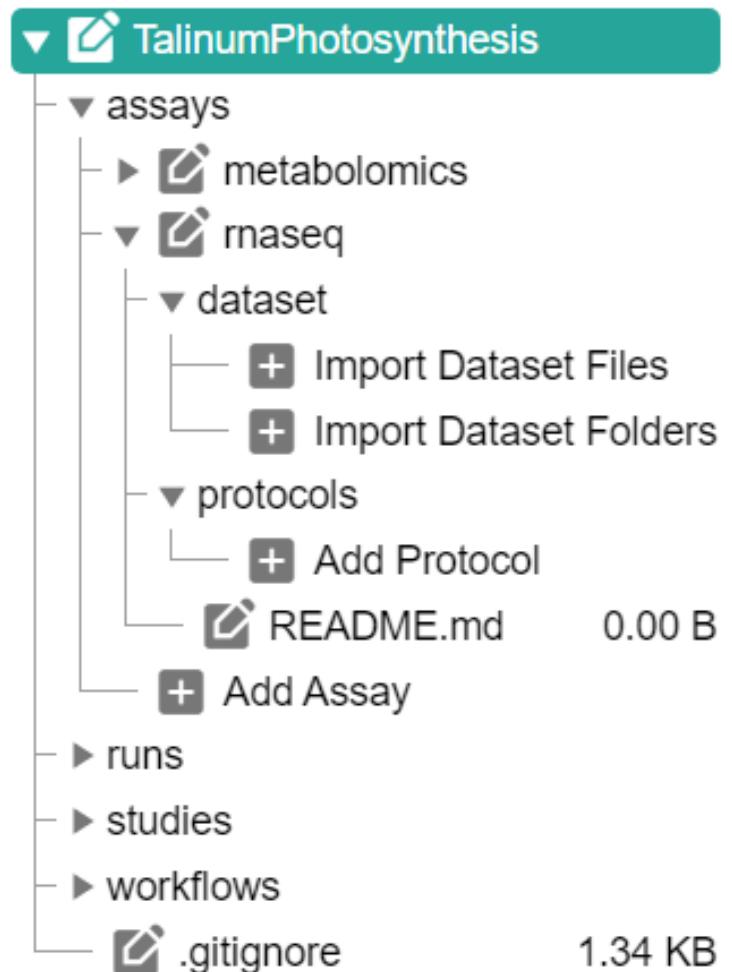
Add protocols and datasets

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.

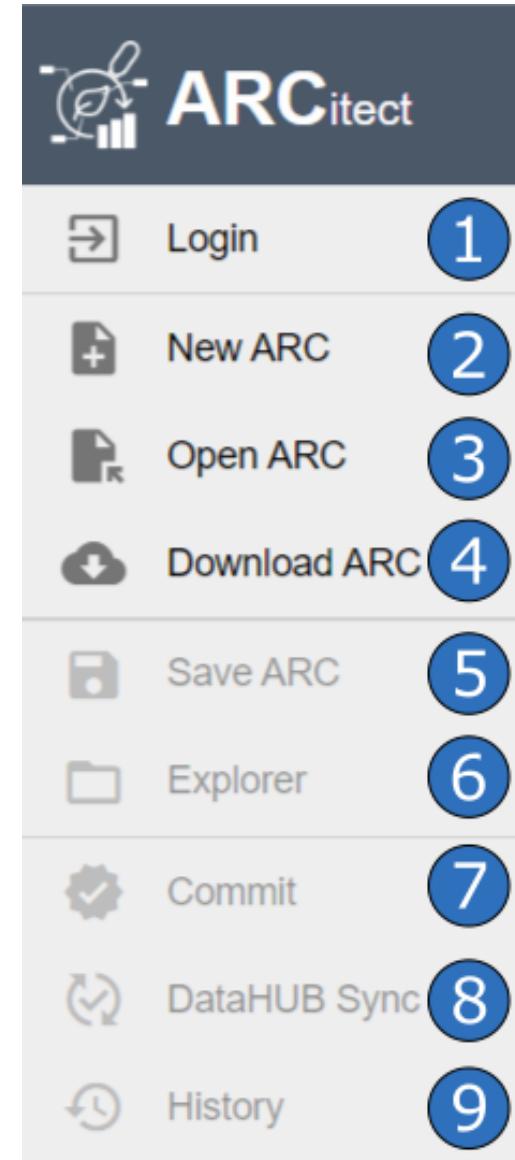


Sort Demo Data to your ARC

 protocols and multiple datasets can directly be imported via ARCitect

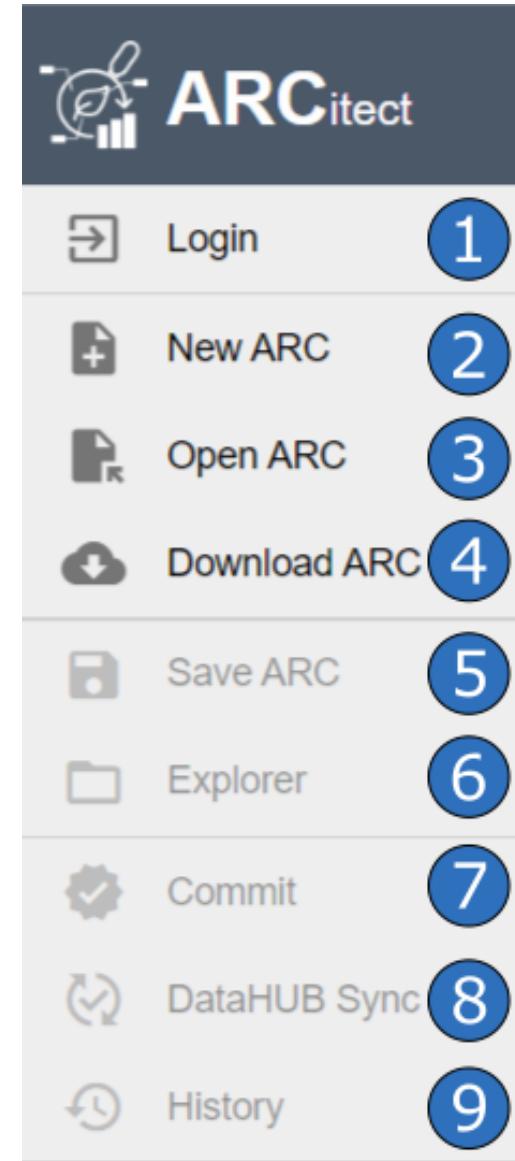
Explorer

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*

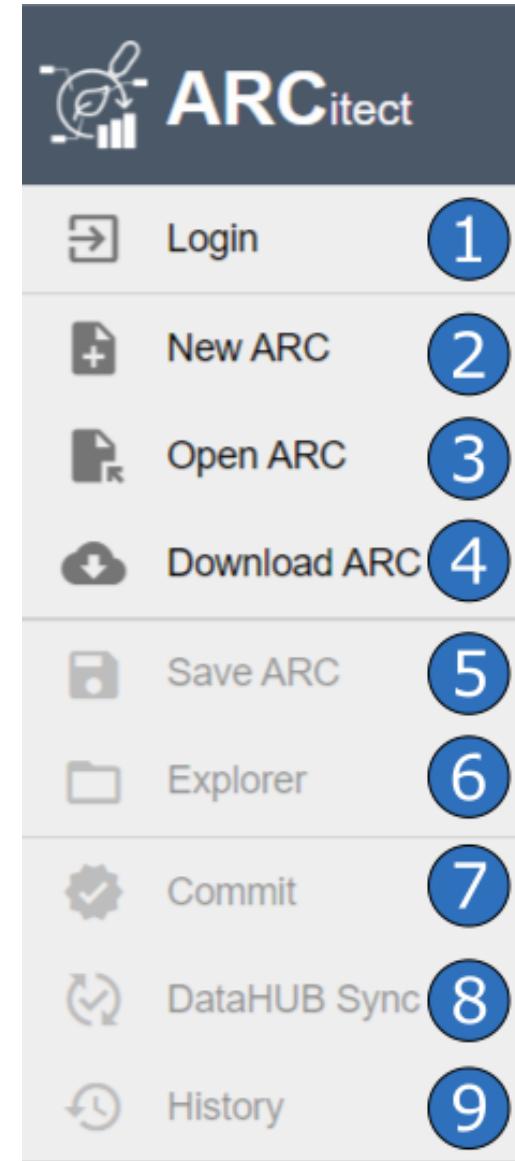
It allows you to

- 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

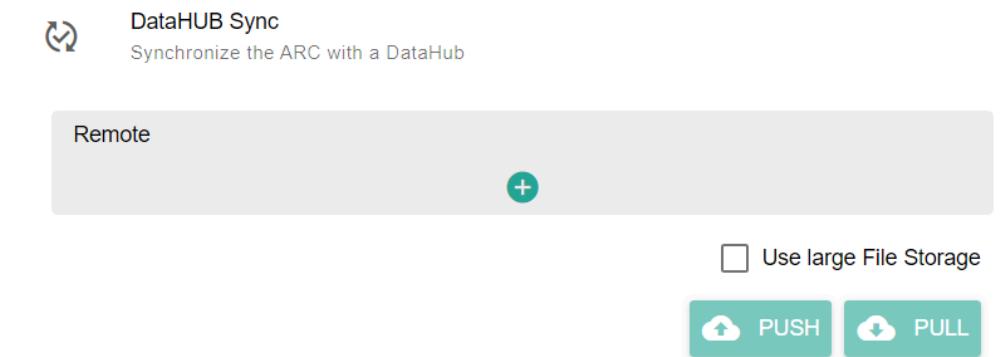
From the sidebar, navigate to **DataHUB Sync** (8)



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

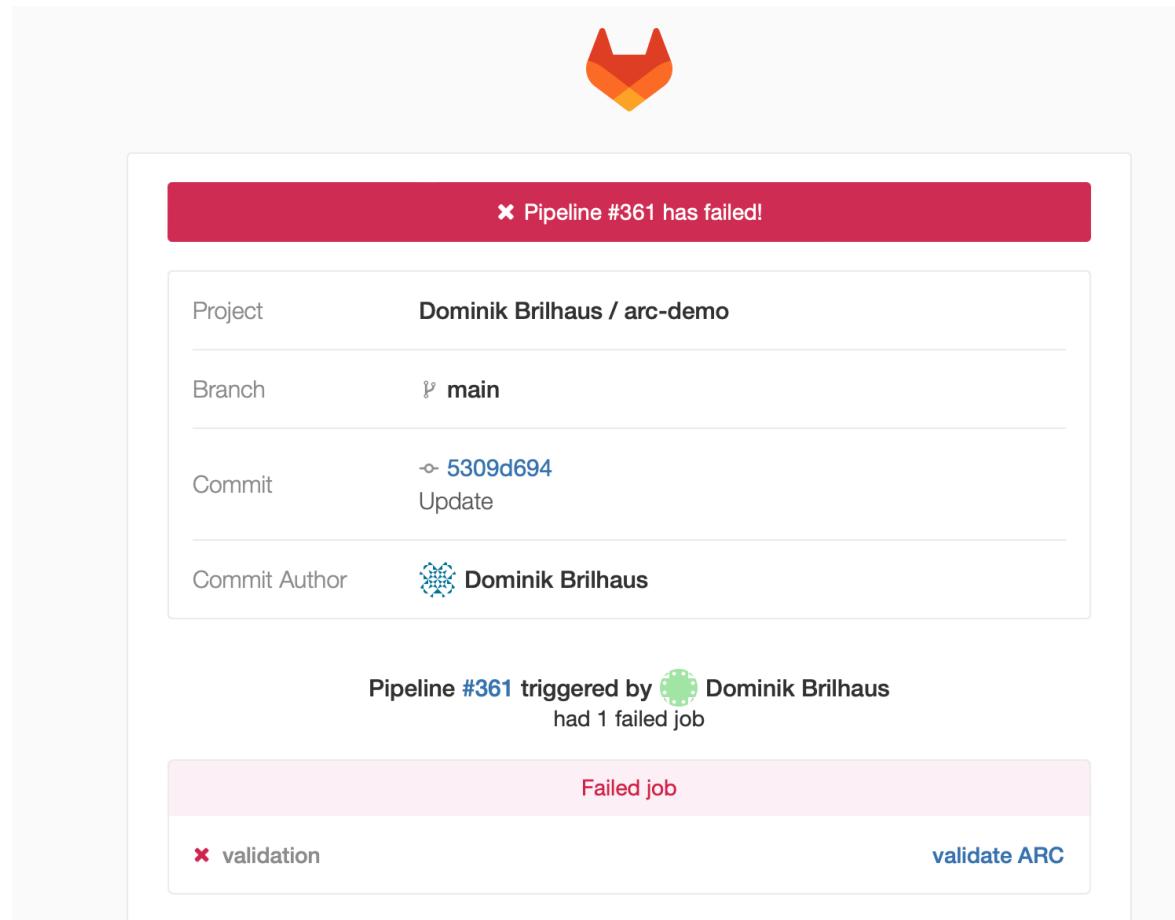
 Structured and ...

 ... annotated experimental data

 Shared with collaborators



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

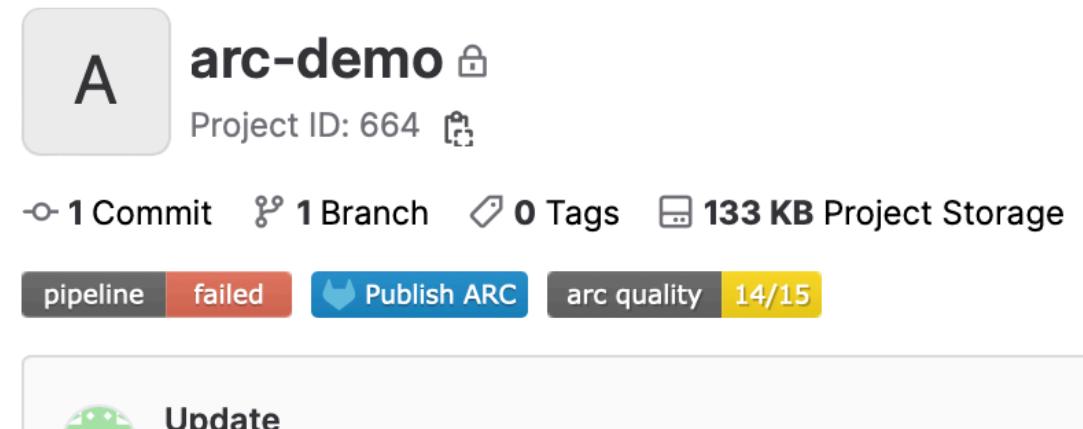


🔥 Don't worry 😊

Pipeline Failed

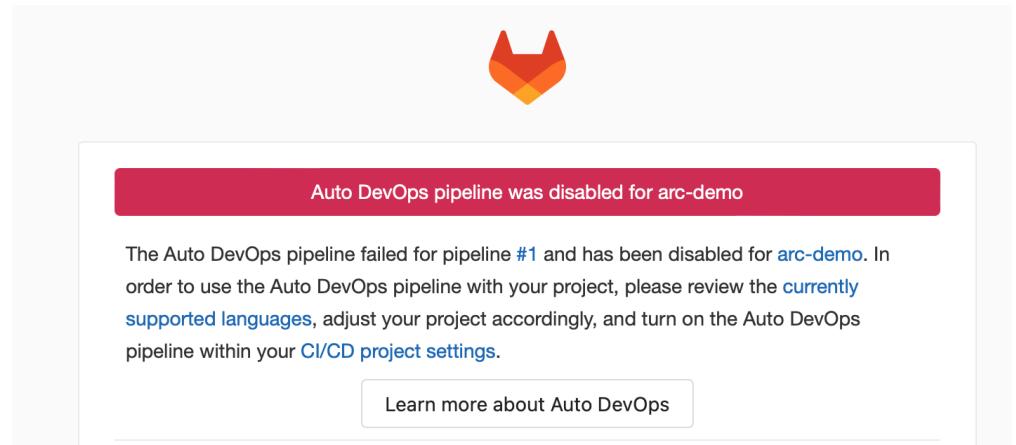
- a "continuous quality control" (CQC) pipeline validates your ARC
- This fails if one of the following metadata items is missing:

```
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

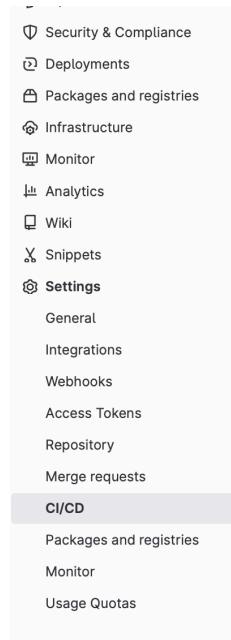


The screenshot shows a message about a disabled Auto DevOps pipeline. At the top right is a small orange and yellow logo. Below it, a red bar contains the text "Auto DevOps pipeline was disabled for arc-demo". The main message area is white with black text, stating: "The Auto DevOps pipeline failed for 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



Auto DevOps

Automate building, testing, and deploying your applications based on your continuous integration and delivery configuration.
How do I get started?

Default to Auto DevOps pipeline instance enabled

The Auto DevOps pipeline runs if no alternative CI configuration file is found. [Learn more.](#)

Add a [Kubernetes cluster integration](#) with a domain, or create an AUTO_DEVOPS_PLATFORM_TARGET CI variable.

Deployment strategy

- Continuous deployment to production [?](#)
 Continuous deployment to production using timed incremental rollout [?](#)
 Automatic deployment to staging, manual deployment to production [?](#)

[Save changes](#)

Runners

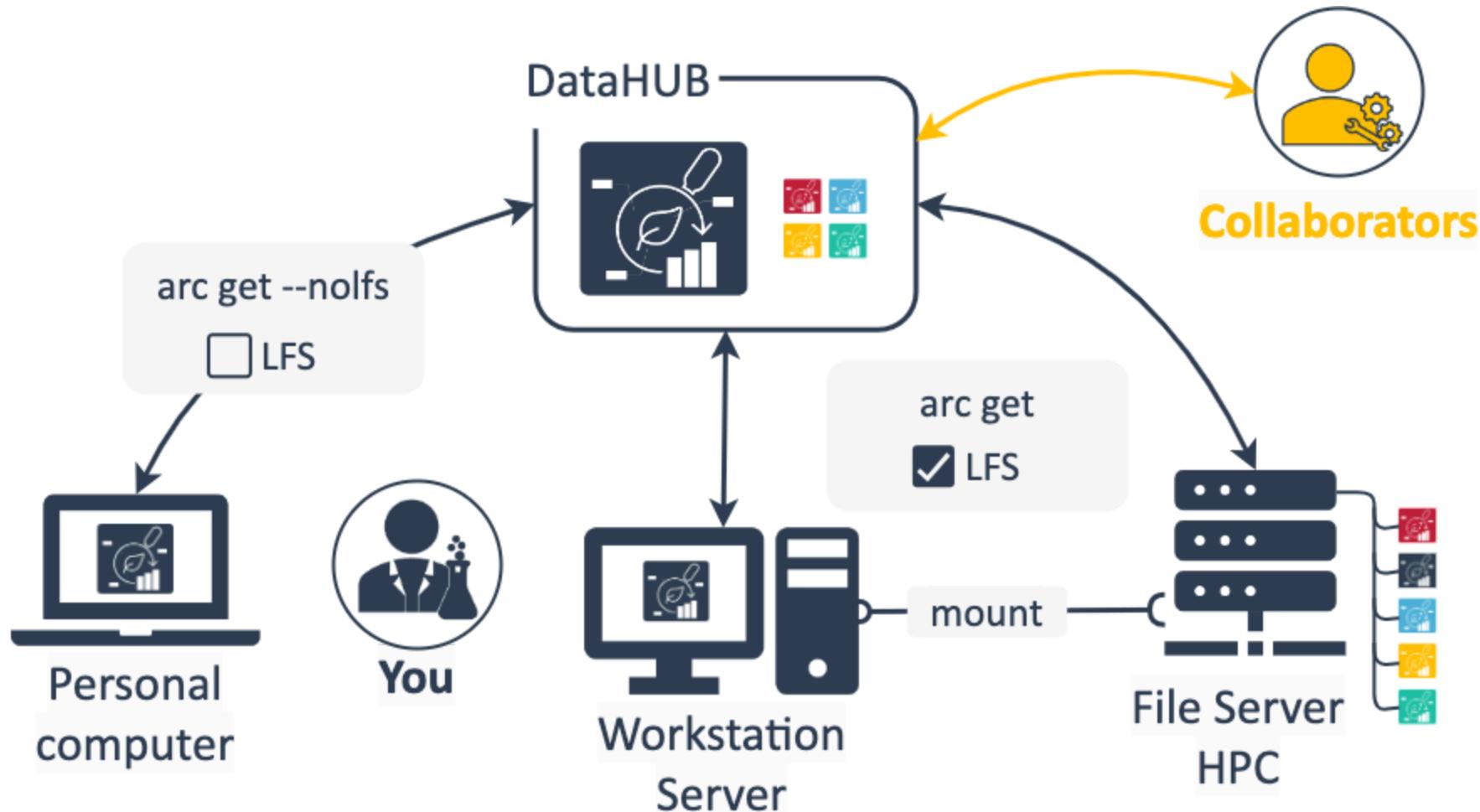
Runners are processes that pick up and execute CI/CD jobs for GitLab. [What is GitLab Runner?](#)

Artifacts

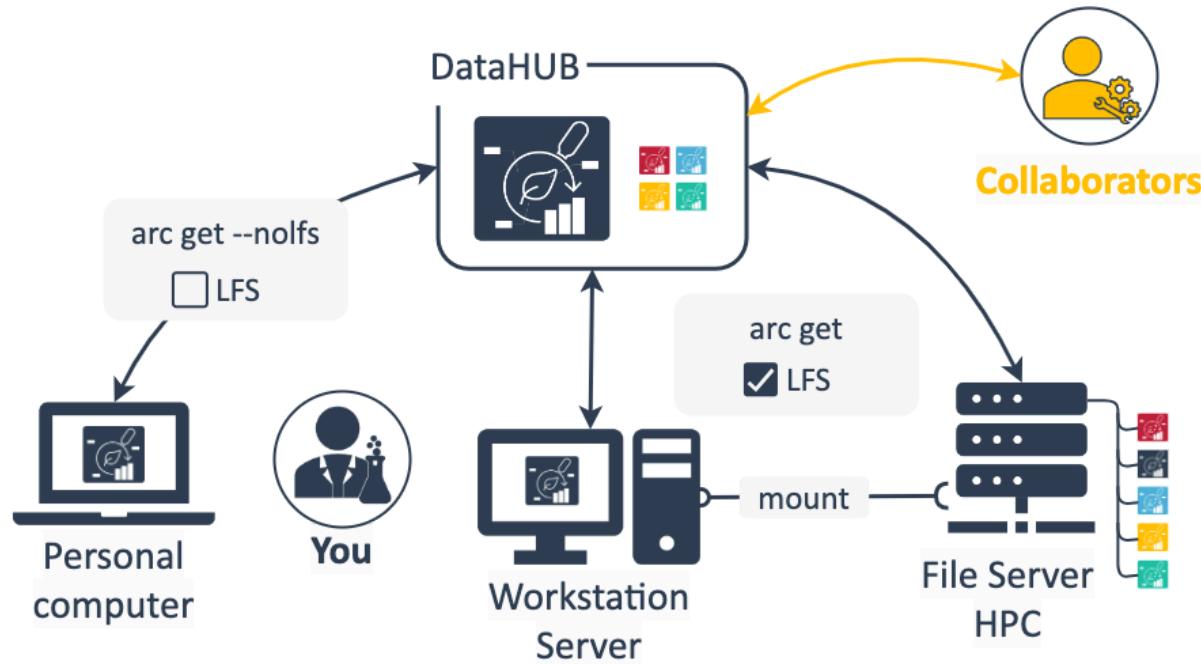
A job artifact is an archive of files and directories saved by a job when it finishes.

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

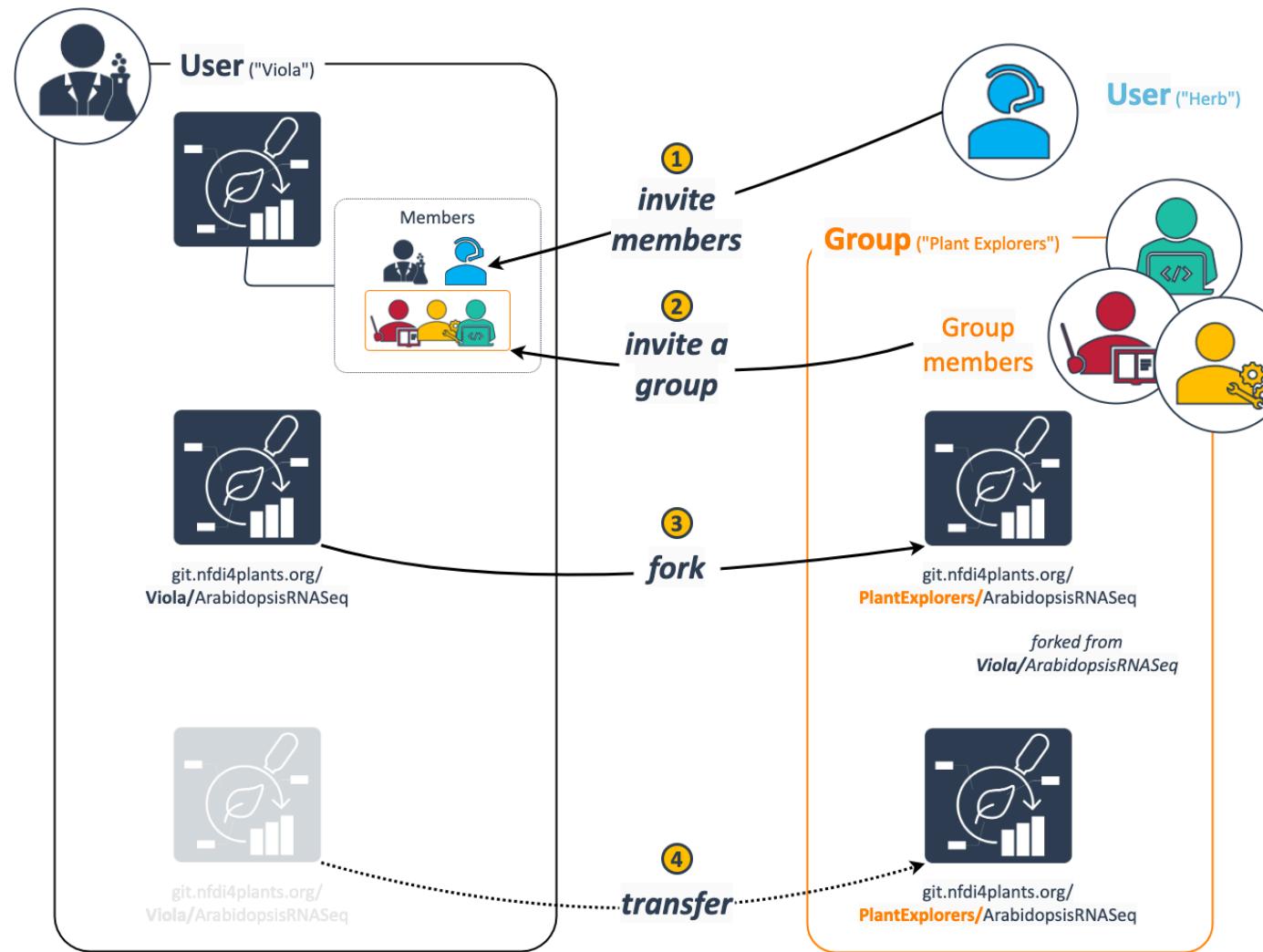
Project = ARC

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

 **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' page of a project named 'Samuilov-2018-BOU-PSP'. It includes a search bar, buttons for 'Import from a project', 'Invite a group', and 'Invite members', and a table listing four members with their accounts, source, max role, expiration date, and activity details.

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

The screenshot shows the 'Samuilov-2018-BOU-PSP' project in ARCIctect. It displays the project's identifier, title, and description, along with a file tree and a list of investigation contacts with their scores and links.

Identifier: Samuilov-2018-BOU-PSP

Title:

Description:

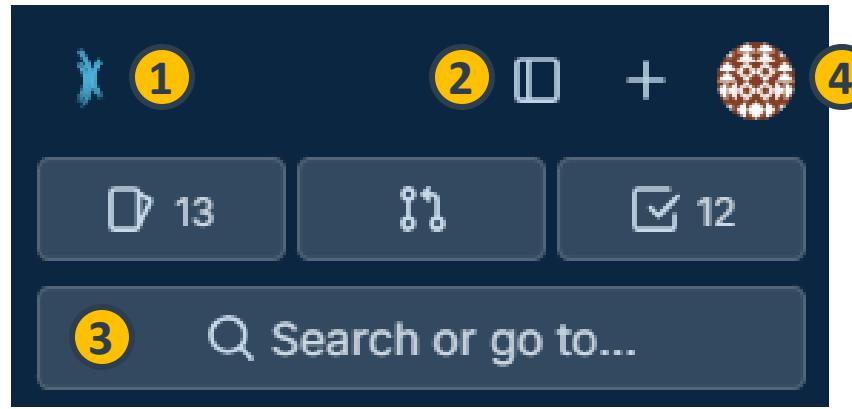
Contacts:

Contact	Score	Link
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	

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

DataHub Hands-On

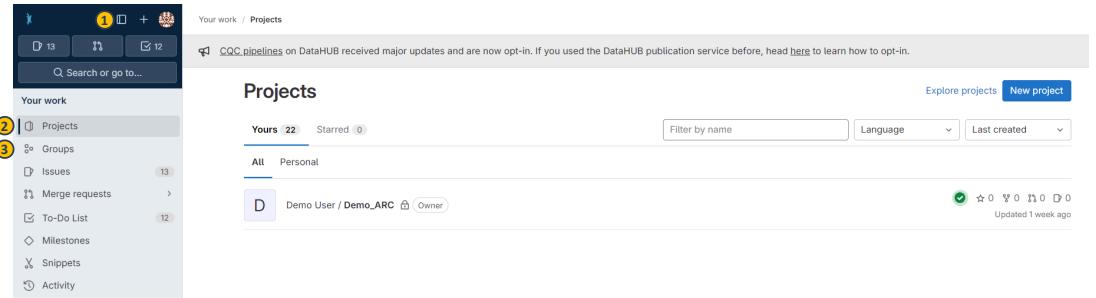
Navigation Bar



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)

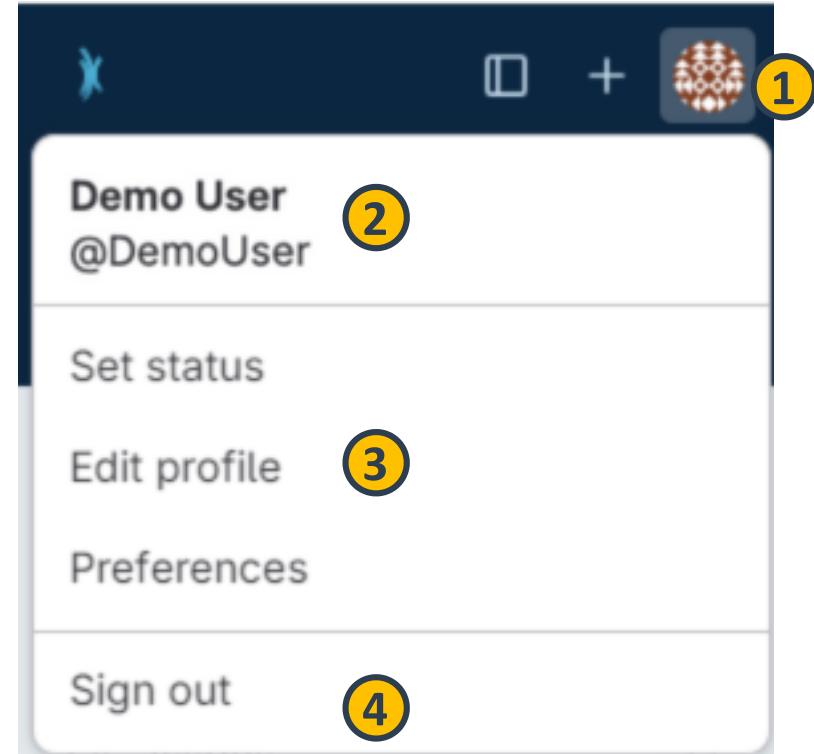
Hamburger Menu

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.



The screenshot shows the DataHUB interface for managing ARCs (Application Readiness Checkpoints). The top navigation bar includes icons for user profile, search, and project creation, followed by the path 'Your work / Projects'. A banner at the top right informs users about CQC pipelines updates. The main area is titled 'Projects' and displays a list of ARCs. On the left, a sidebar lists 'Your work' sections: Projects (selected), Groups, Issues (13), Merge requests (1), To-Do List (12), Milestones, Snippets, and Activity. The main content area has tabs for 'Yours' (22) and 'Starred' (0). It includes filters for 'Filter by name', 'Language', and 'Last created'. A specific ARC is highlighted with yellow circles numbered 1 through 6. Circle 1 points to the 'Yours' tab. Circle 2 points to the main list item 'Demo User / Demo_ARC'. Circle 3 points to the visibility icon (eye). Circle 4 points to the permission or role icon (key). Circle 5 points to the 'Explore projects' button. Circle 6 points to the 'New project' button.

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.

Demo User / Demo_ARC

D Demo_ARC 5

main / Demo_ARC / + 7

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 Publish ARC

- 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

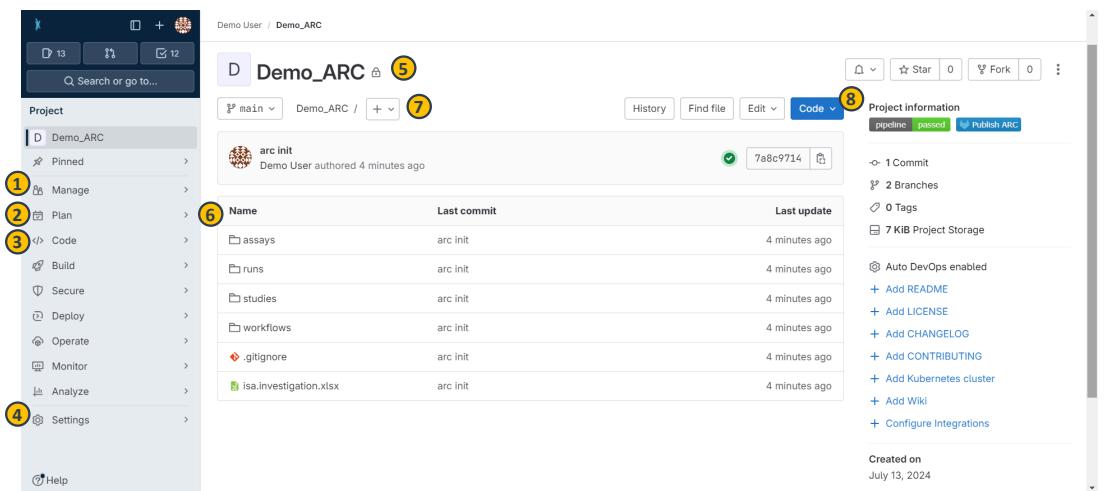
Manage 1
Plan 2
</> Code 3
Build 4
Secure
Deploy
Operate
Monitor
Analyze
Settings
Help

arc init
Demo User authored 4 minutes ago

7a8c9714

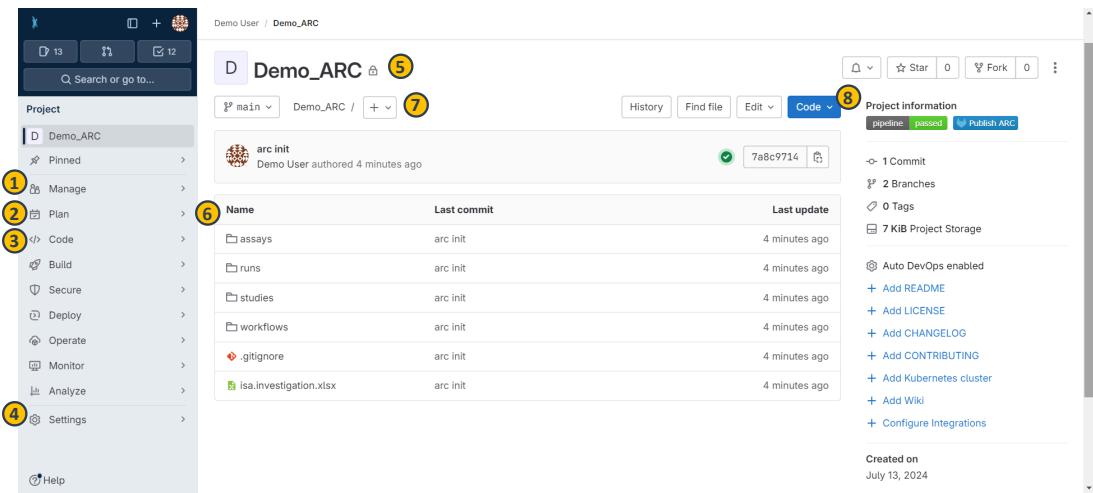
ARC Panel – sidebar

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.

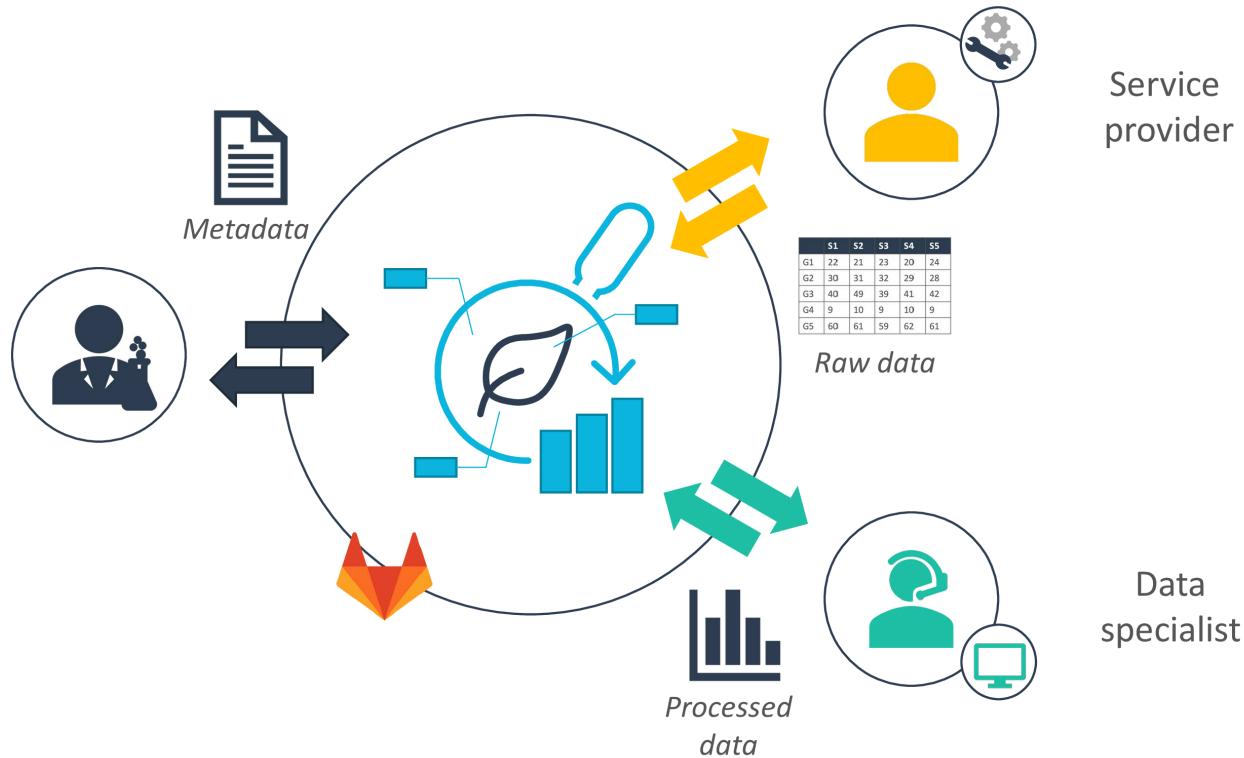


ARC Panel – main panel

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).



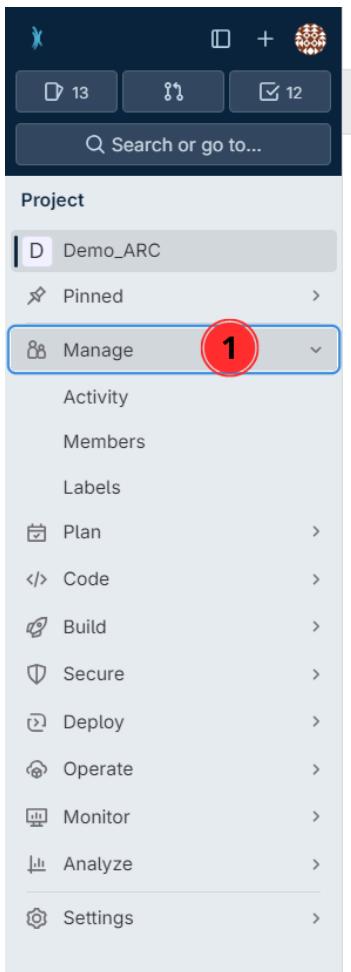
Collaborate and share



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 DataHUB project management interface. On the left, a sidebar lists various project management sections: Project, Pinned, Manage (highlighted with a red circle), Activity, Members, Labels, Plan, Code, Build, Secure, Deploy, Operate, Monitor, Analyze, and Settings. The main content area displays the 'Demo_ARC' project details. At the top, there's a message about CQC pipelines. Below it, the project name 'Demo_ARC' is shown with a lock icon, and a 'Code' dropdown menu is open. A commit history table lists files like 'assays', 'runs', 'studies', 'workflows', '.gitignore', and 'isa.investigation.xlsx' with their last commits and update times. To the right, a 'Project information' section provides summary statistics: 1 Commit, 2 Branches, 0 Tags, 21 KiB Project Storage, and various enablement options like Auto DevOps, Wiki, README, LICENSE, CHANGELOG, CONTRIBUTING, Kubernetes cluster, and Integrations. It also shows the creation date as July 13, 2024.

Demo User / Demo_ARC

CQC 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.

D Demo_ARC

main Demo_ARC / +

History Find file Edit Code

Project information

pipeline passed [Publish ARC](#)

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

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

Auto DevOps enabled
Wiki
+ Add README
+ Add LICENSE
+ Add CHANGELOG
+ Add CONTRIBUTING
+ Add Kubernetes cluster
+ Configure Integrations

Created on
July 13, 2024

2. Click on Members

The screenshot shows the DataHub interface for the project 'Demo_ARC'. The left sidebar has a 'Project' section with 'Demo_ARC' pinned, 'Manage' (circled 1), 'Activity', 'Members' (circled 2), 'Labels', 'Plan', 'Code', 'Build', 'Secure', 'Deploy', 'Operate', 'Monitor', 'Analyze', and 'Settings'. The main area is titled 'Project members' and says 'You can invite a new member to Demo_ARC or invite another group.' It shows one member: 'Demo User @DemoUser' (It's you). The member details include 'Source: Direct member by Demo User', 'Max role: Owner', and an 'Expiration' section with 'Expiration date: Sep 27, 2023', 'Jul 13, 2024', and 'Jul 21, 2024'. There are 'Import from a project', 'Invite a group', and 'Invite members' buttons.

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 'Members' section of the DataHub interface for the 'Demo_ARC' project. The left sidebar has 'Members' selected (2). The main area shows one member ('Demo User') with an 'Invite members' button (3) highlighted.

Demo User / Demo_ARC / Members

CQC_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.

Project members

You can invite a new member to Demo_ARC or invite another group.

Members 1

Filter members Account ▾

Account	Source	Max role	Expiration	Activity
Demo User @DemoUser It's you	Direct member by Demo User	Owner	Expiration date <input type="button" value="Sep 27, 2023"/>	✓ Jul 13, 2024 ✗ Jul 21, 2024

Invite members



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





Guest

Reporter

Developer

Maintainer

Owner

Guest

ARC project.

4

ses

5

[Read more about role permissions](#)

Access expiration date (optional)

YYYY-MM-DD



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 Data PLANT CEPLAS interface with several numbered callouts highlighting specific features:

- 1** Manage: A sidebar menu item.
- 2** Plan: A sidebar menu item.
- 3** Code: A sidebar menu item.
- 4** Settings: A sidebar menu item.
- 5** Demo_ARC: The project name in the header.
- 6** Name: A column header in the file list table.
- 7** History: A button in the top navigation bar.
- 8** Project information: A section on the right side of the screen.

The main area displays a file list table with the following data:

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

The "Project information" section includes:

- pipeline: passed
- Published ARC
- 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

Project Management

Add issues to your ARC (3)

The screenshot shows the Arc project management interface. The left sidebar contains the following items:

- 1 Manage
- 2 Plan
- 3 Code
- 4 Settings
- 5 Help

The main project details page for "Demo_ARC" is displayed. The top navigation bar includes:

- 13 Issues
- 12 Pull Requests
- Search or go to...

The project title "Demo_ARC" is shown with a lock icon and a yellow circle containing the number 5. The navigation bar below the title includes:

- main
- Demo_ARC /
- + (yellow circle 7)
- History
- Find file
- Edit
- Code (blue button)
- 8 (yellow circle)

The code editor area displays a single commit:

arc init
Demo User authored 4 minutes ago

The commit hash is 7a8c9714. The right side of the screen shows "Project information" with the following details:

- pipeline passed
- 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.

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 a DataHub interface. At the top, there's a header with user information and navigation links. Below it is a sidebar with project management options like 'Pinned', 'Manage', 'Plan', 'Issues', 'Issue boards', 'Milestones', and 'Wiki'. The 'Wiki' option is highlighted. To the right of the sidebar, the main content area displays the 'Home' page of the 'Wiki' for 'Demo_ARC'. The page includes a message about CQC pipelines, an 'Edit' button, and a 'Pages' section showing 2 items. Further down, sections for 'Meeting Schedule' and 'Ideas and drafts' are visible, each containing a bulleted list of items.

Demo User / Demo_ARC / Wiki / Home

CQC 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.

Pages 2

Home

Last edited by Demo User just now

This is the wiki to Demo_ARC. We will announce meeting schedules here.

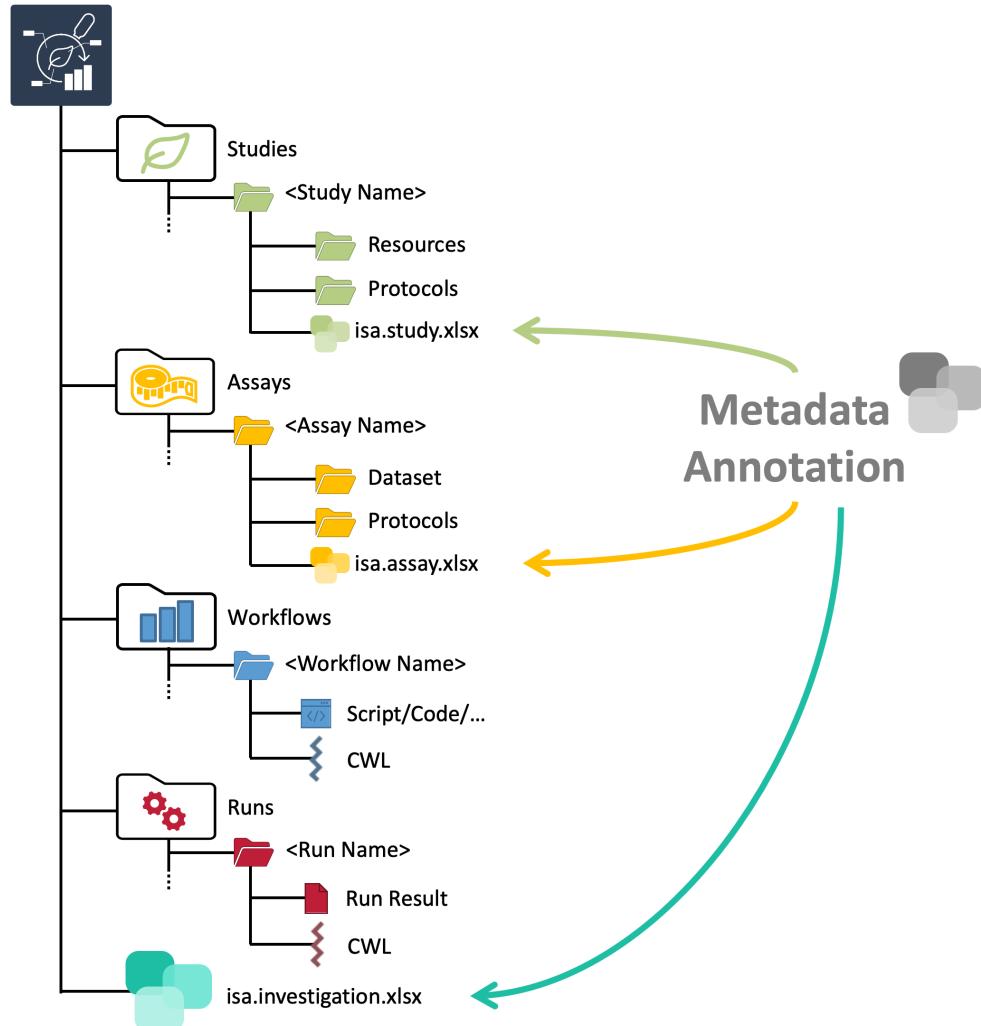
Meeting Schedule

- 2024-06-12 Kick-off
- 2024-06-27 Proposal discussion
- 2024-07-04 Discuss RNA-seq pipeline

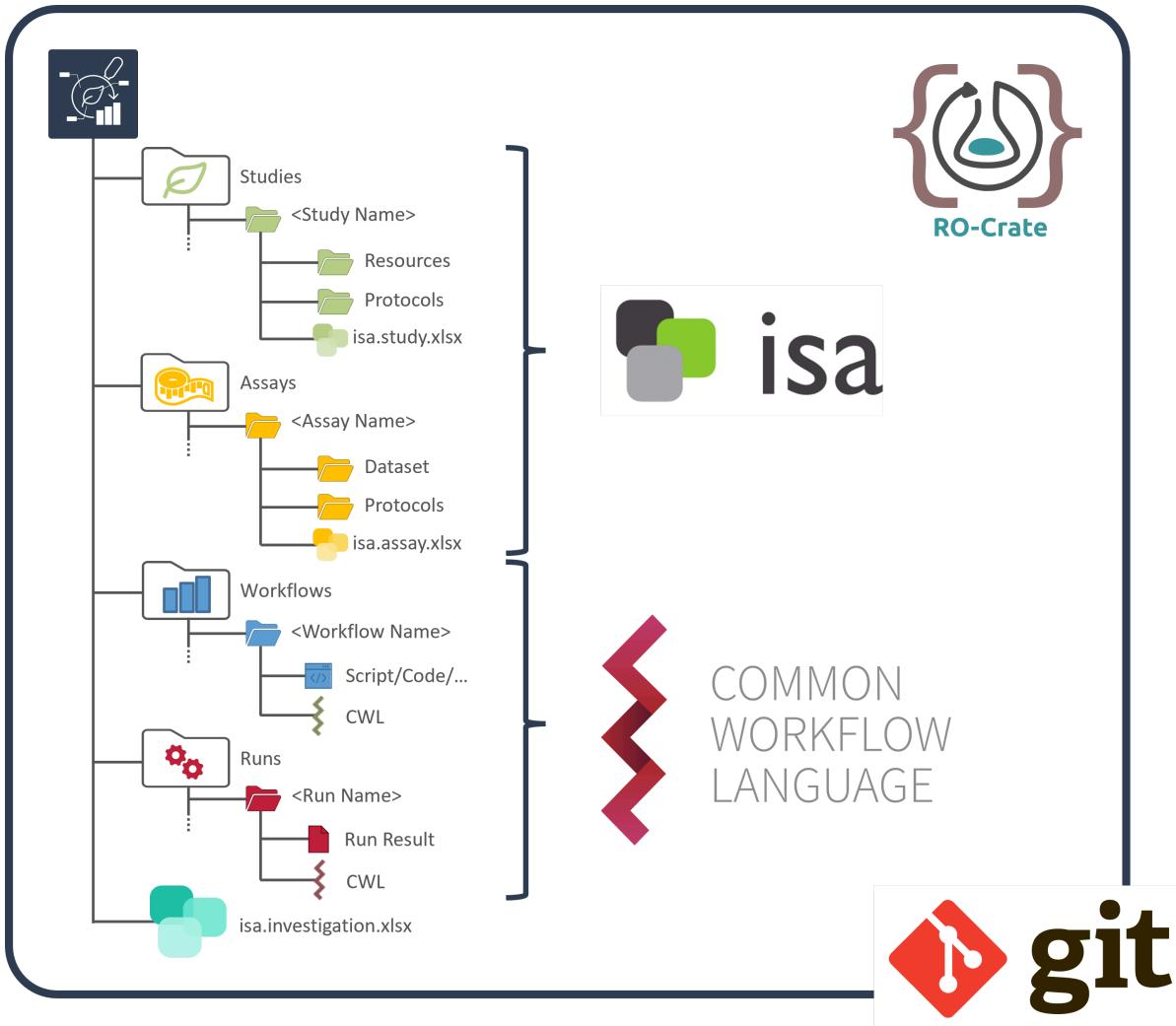
Ideas and drafts

- Golden Gate protocol

ARCs come with comprehensive metadata

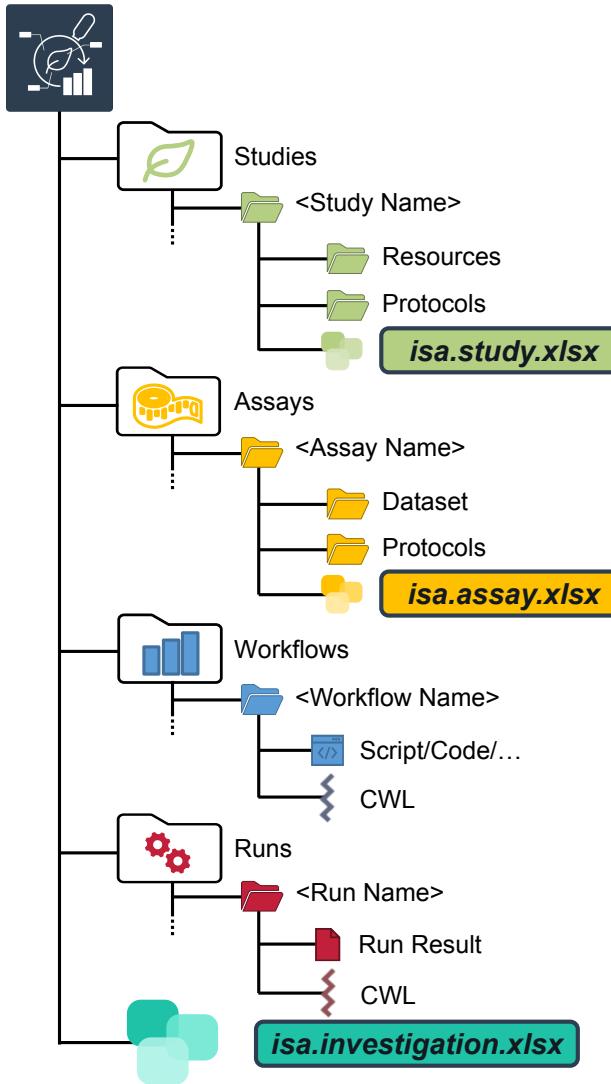
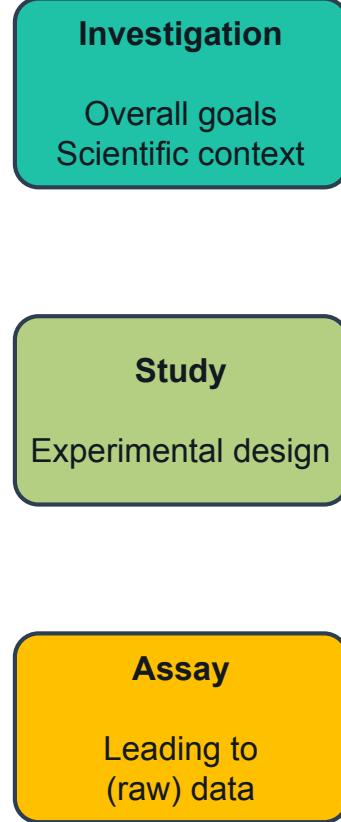
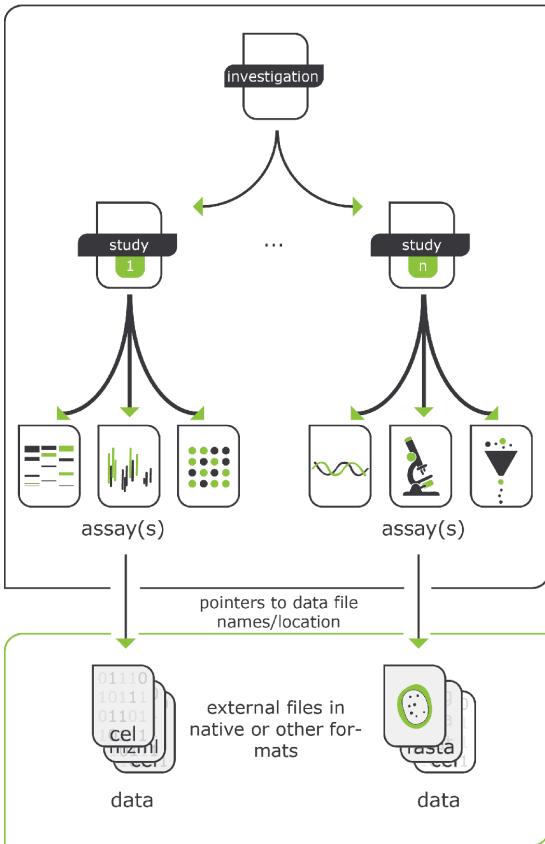


ARC builds on standards



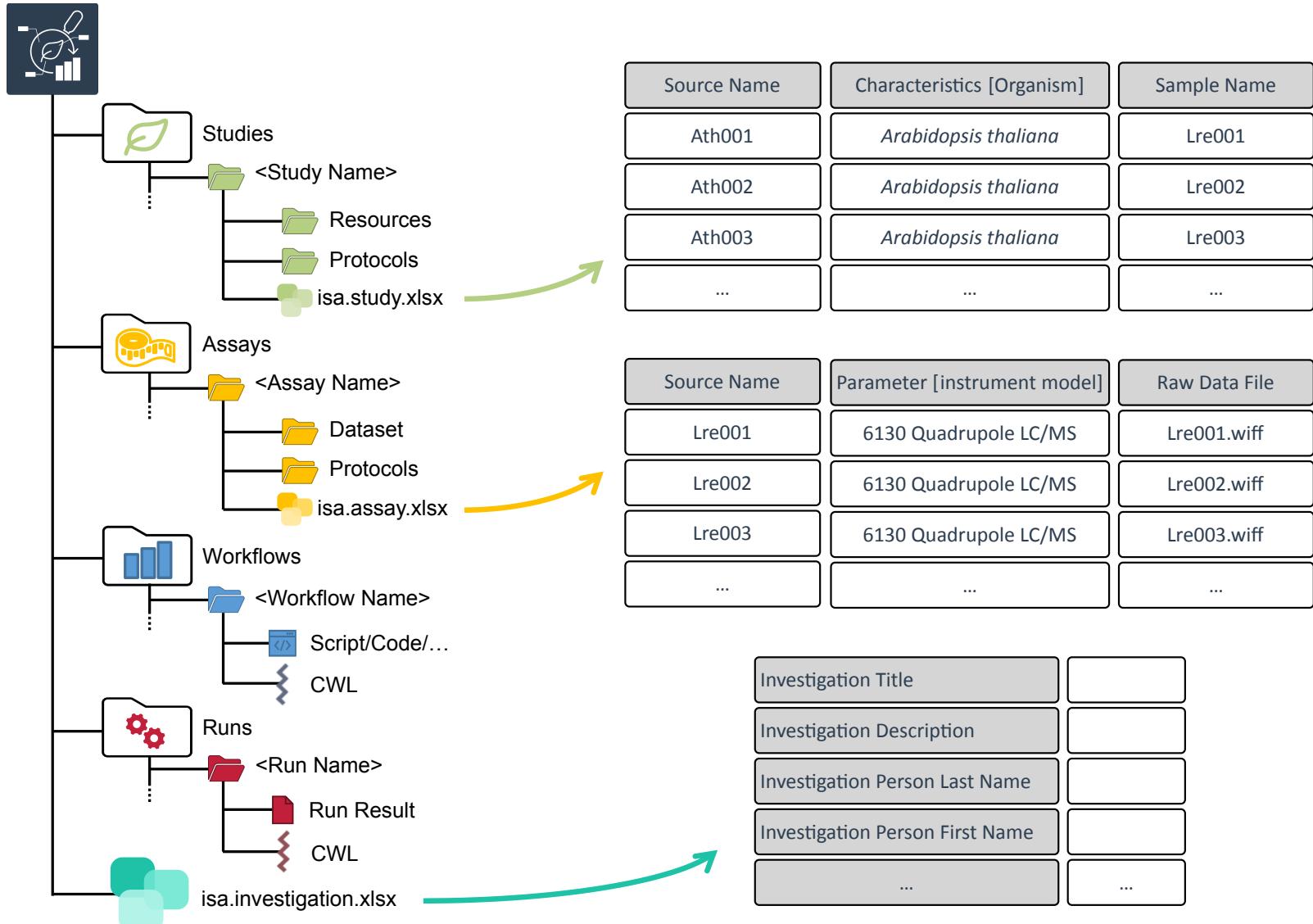
<https://isa-tools.org/> | <https://www.commonwl.org/>
<https://www.researchobject.org/ro-crate/> | <https://git-scm.com>

ARC builds on ISA



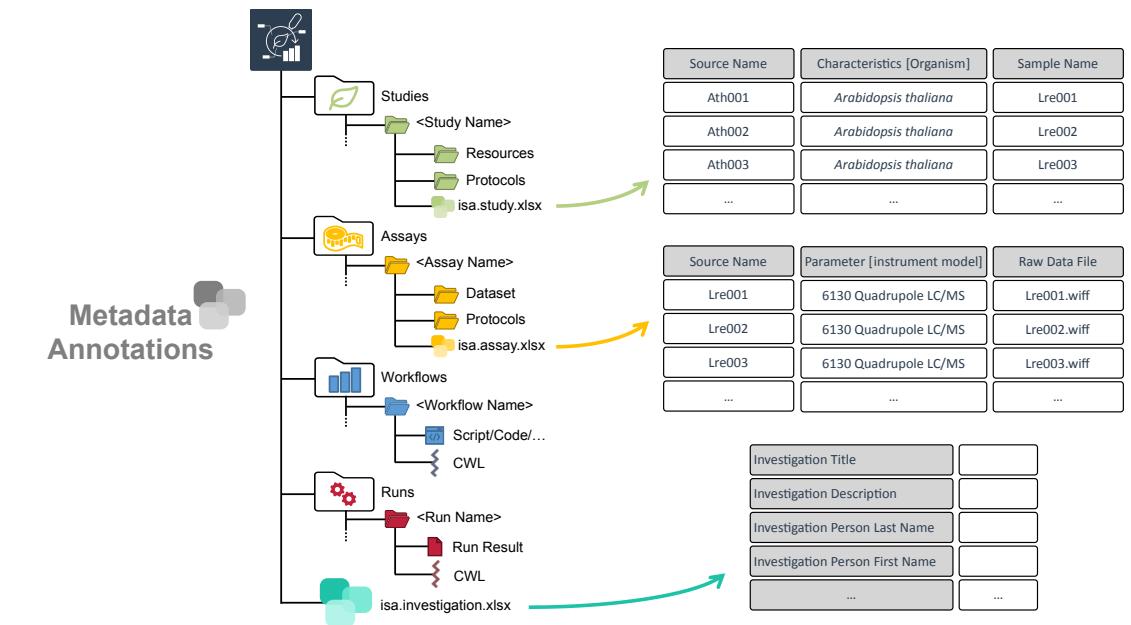
ARC builds on ISA to connect data

Metadata Annotations

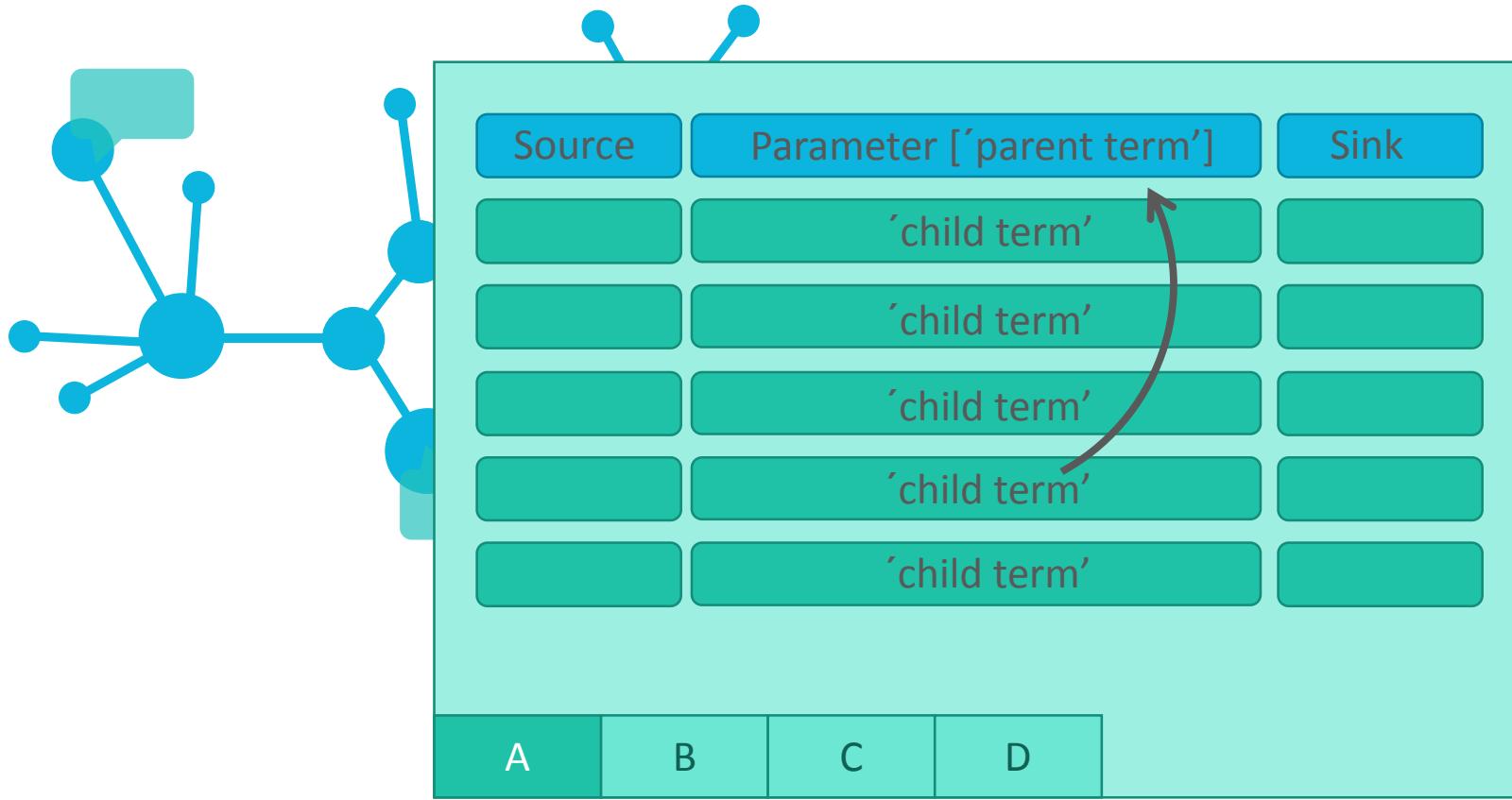


ARC builds on ISA to link data

- 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. On the left is a table titled 'Widgets' with the following 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 rewetted	reC3_01
DB_163	Talinum fruticosum	✓ 12 days drought + 2 days rewetted	reC3_02
DB_165	Talinum fruticosum	✓ 12 days drought + 2 days rewetted	reC3_03

Annotations are present in the 'Factor' column:

- Row 1: '12 days drought'
- Row 2: '12 days drought'
- Row 3: '12 days drought'
- Row 4: '12 days drought + 2 days rewetted'
- Row 5: '12 days drought + 2 days rewetted'
- Row 6: '12 days drought + 2 days rewetted'

Labels with arrows point to specific elements:

- 'Input' points to the 'Input [Source Name]' column.
- 'Characteristic' points to the 'Characteristic [organism]' column.
- 'Factor' points to the 'Factor [watering exposure]' column.
- 'Output' points to the 'Output [Sample Name]' column.
- 'Sidebar' points to the right-hand panel containing a 'New Parameter' dialog and a list of components.

The sidebar includes tabs for 'Parameter', 'Factor', 'Characteristic', 'Component', 'More', and 'Output'. A search bar at the top of the sidebar says 'instrument model'.

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 an ontology search sidebar on the right.

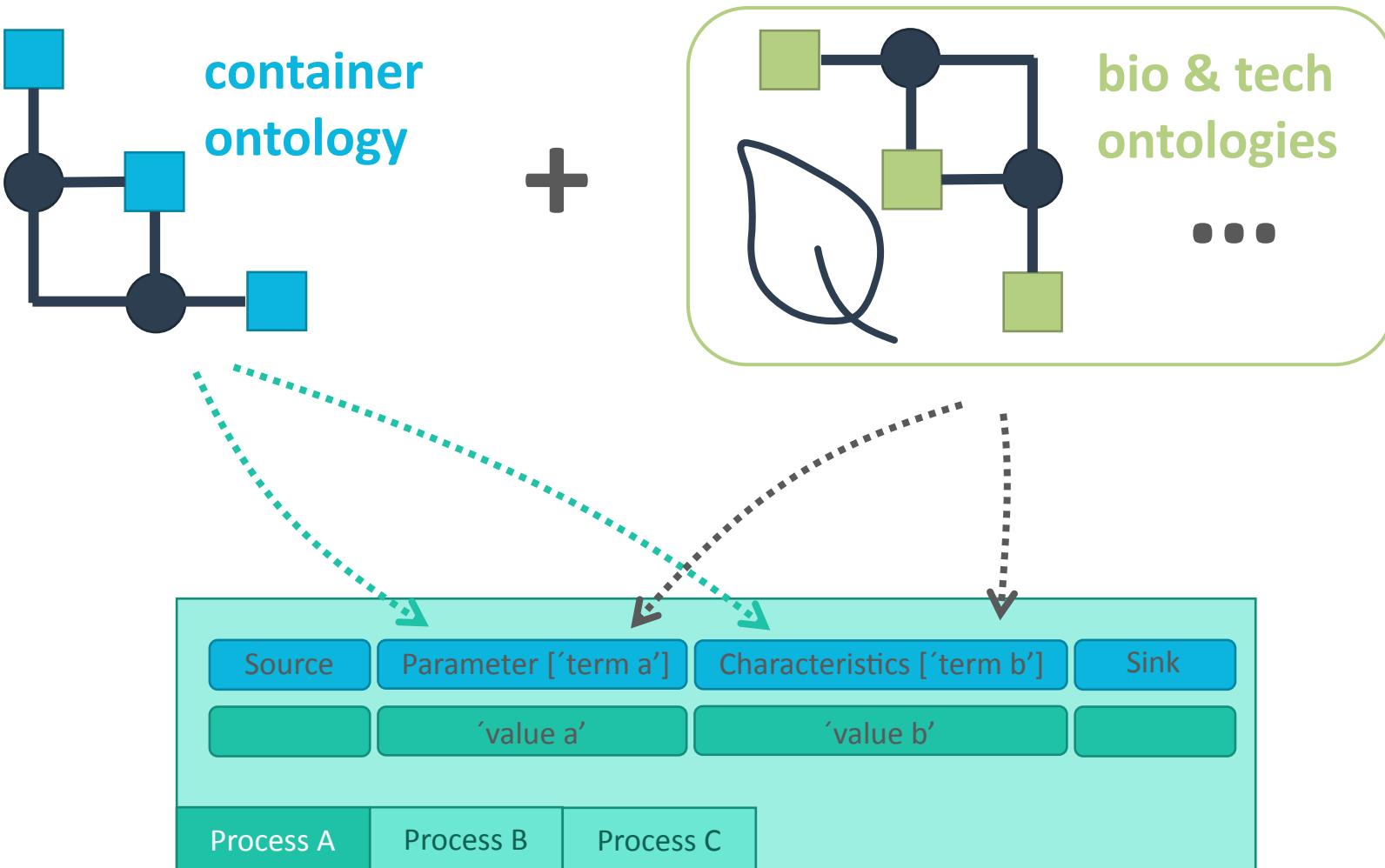
Table Data:

	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

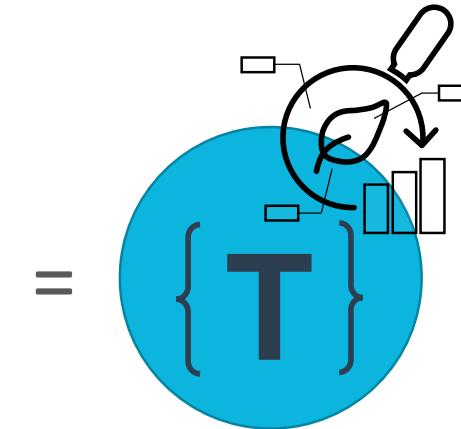
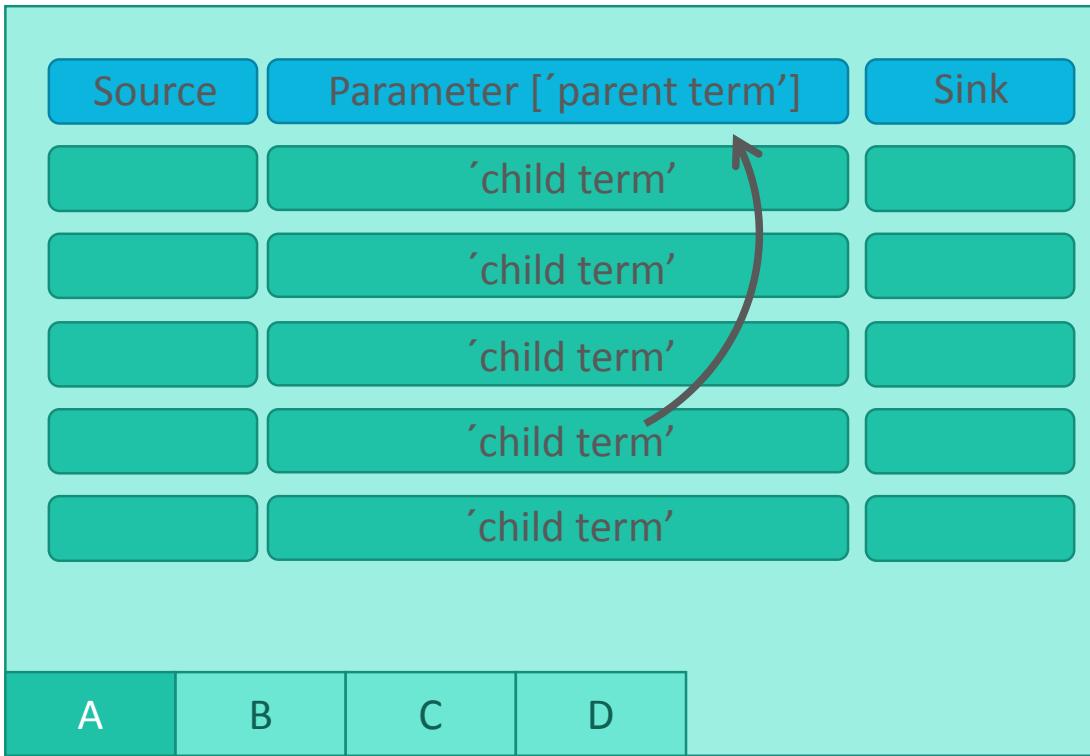
Ontology Search Sidebar:

- Search bar: Talinum fruticosum
- Parent: Organism, MIAPPE:0041
- Use advanced search
- Fill selected cells with this term

Hierarchical combination of ontologies

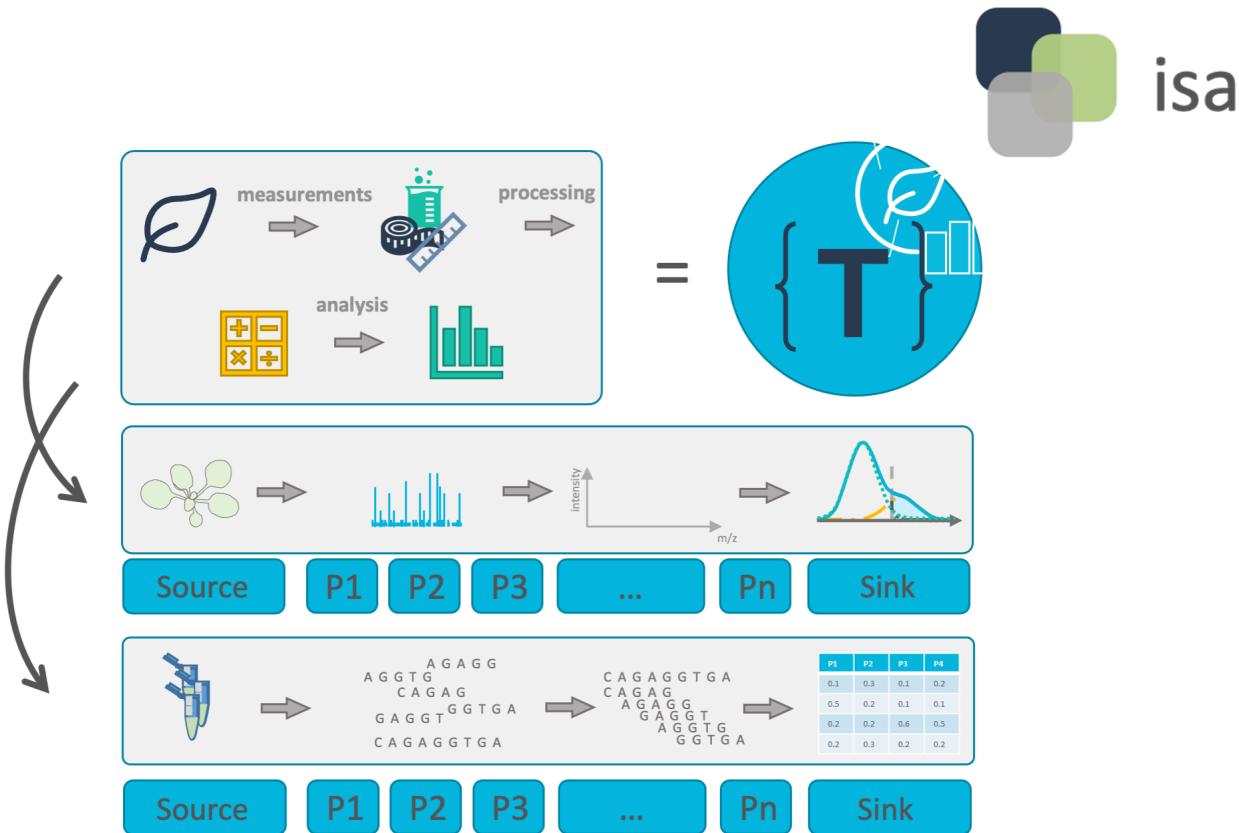


Checklists and Templates



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

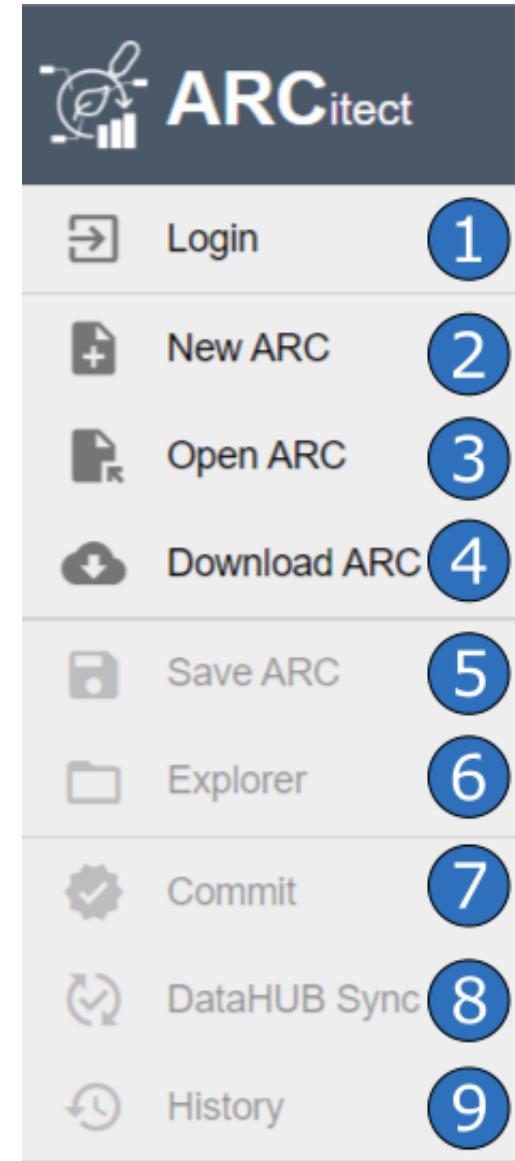
Swate hands-on with demo data

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)

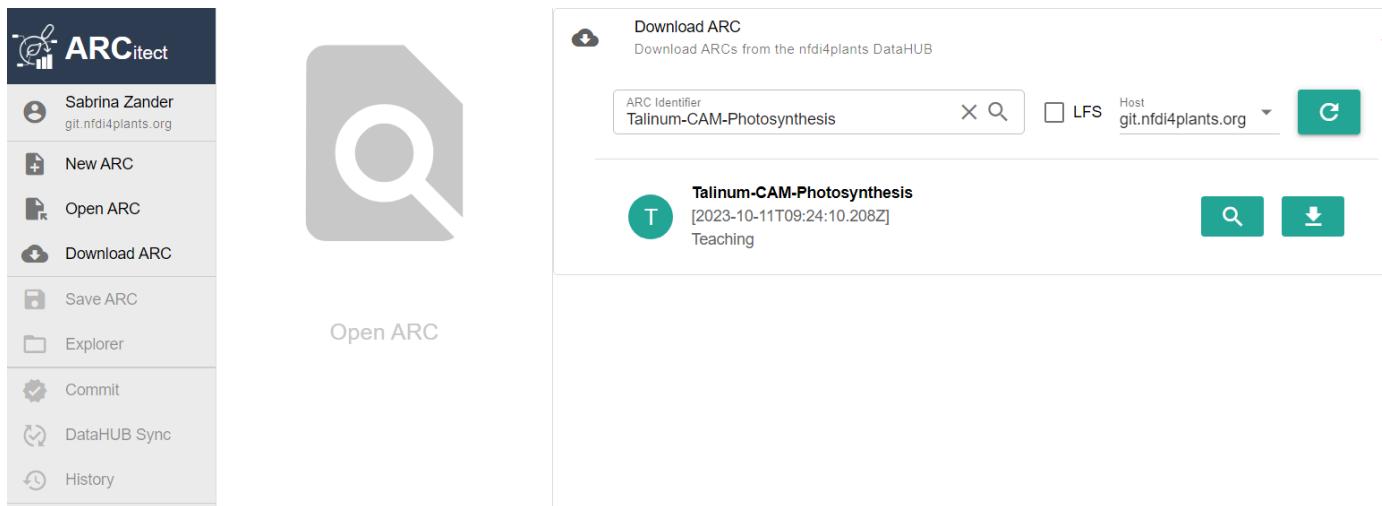


Download the demo data

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 shows the Swate software interface with several labeled components:

- Widgets**: The main title bar.
- Input**: A blue box pointing to the "Input [Source Name]" column in the data grid.
- Characteristic**: A blue box pointing to the "Characteristic [organism]" column in the data grid.
- Factor**: A blue box pointing to the "Factor [watering exposure]" column in the data grid.
- Output**: A blue box pointing to the "Output [Sample Name]" column in the data grid.
- New Parameter**: A modal dialog titled "New Parameter" with a search bar for "instrument model".
- Sidebar**: A blue box pointing to the right sidebar area.

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 bottom right corner shows the page number "132".

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

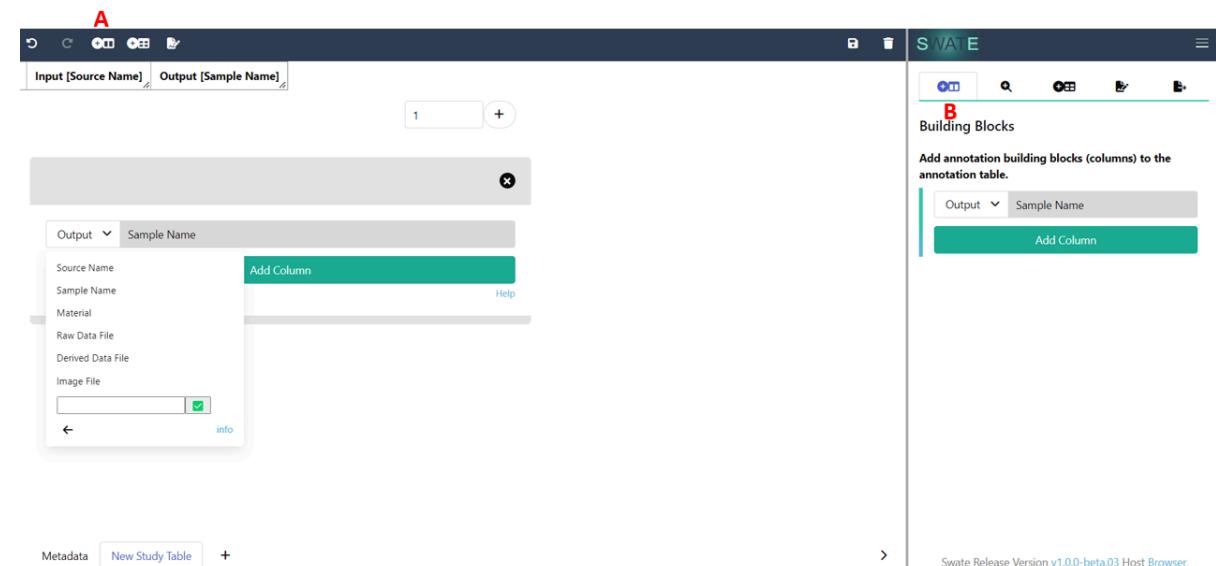
```
studies/talinum_drought/isa.study.xlsx
```



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

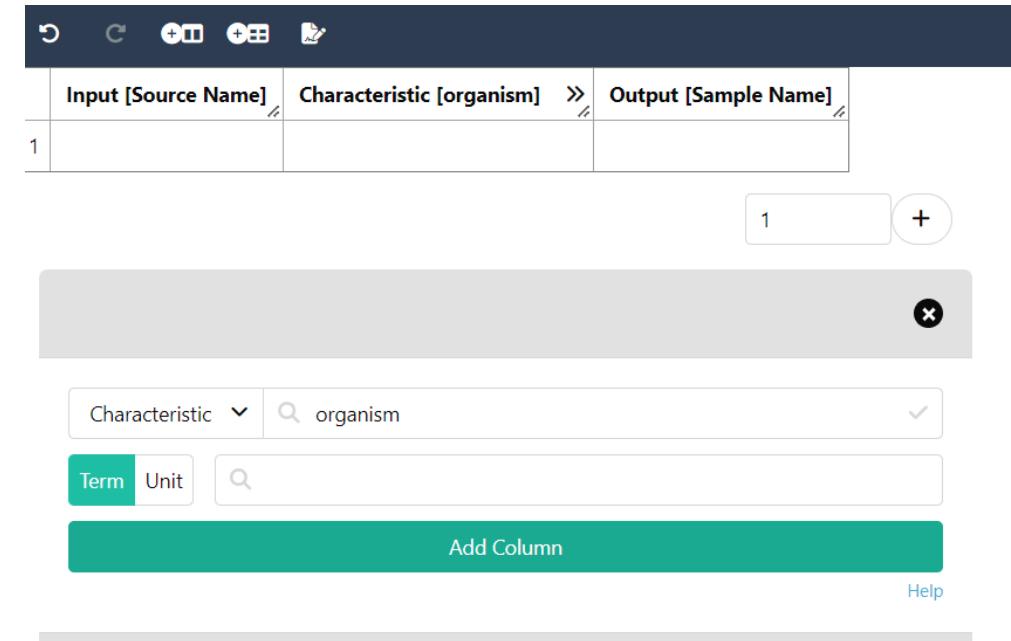
2. Select *Characteristic* from the drop-down menu

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's 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 value '1' under 'Input [Source Name]'. To the right of the table is a numeric input field with '1' and a '+' button. Below the table is a modal dialog. The dialog has a dropdown menu set to 'Characteristic'. It also contains a search bar with the word 'organism' and a checkmark icon. There are tabs for 'Term' and 'Unit', and a magnifying glass icon. At the bottom of the dialog is a large green 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	<input type="text"/>
	↳ Talinum fruticosum NCBITaxon:110664	<input checked="" type="checkbox"/>
	↳ Talinum paniculatum NCBITaxon:107604	<input type="checkbox"/>

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	
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.

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>