

Block 0 – Preparation

before September 29th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

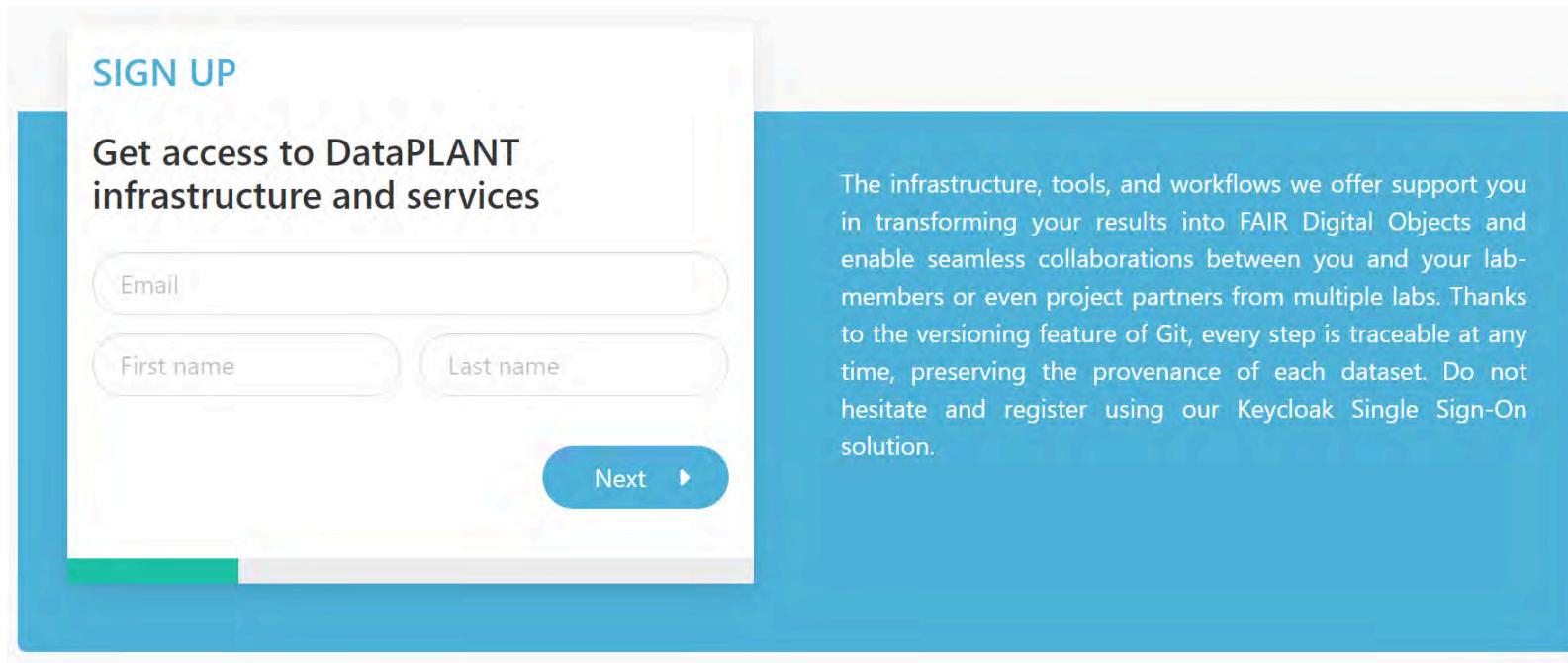
Checklist hands-on sessions

💡 Please prepare the following before the workshop:

- ✓ Register at DataPLANT
- ✓ Find your command line
- ✓ Install and configure Git on your computer
- ✓ Install ARCitect on your computer
- ✓ Install Swate on your computer
- ✓ (optional) Install VS Code

DataPLANT Registration

If you do not have a DataPLANT account, please register at the [DataPLANT website](#).



Role and consortium

Please add your Project/consortium (e.g. CEPLAS, SFB, TRR) and choose the role Guest

The screenshot shows a 'SIGN UP' interface. At the top, it says 'SIGN UP'. Below that, 'Affiliation details' is listed. There are two input fields: one for 'Project/consortium' and another for 'Research interests. Multiple interests need to be separated with a comma.' A dropdown menu titled 'Choose your Role in DataPLANT' is open, listing 'DataSteward', 'Developer', 'Member', and 'Guest'. The 'Guest' option is highlighted with a blue background and a checked checkbox icon.

SIGN UP

Affiliation details

Project/consortium

Research interests. Multiple interests need to be separated with a comma.

✓ Choose your Role in DataPLANT

- DataSteward
- Developer
- Member
- Guest

The command line

Find the **command-line interface (CLI)** on your system.

- On Windows: Enter `powershell` into the explorer path
- On MacOS: Search `terminal` via spotlight (`⌘ + ⌂`) or navigate to `Applications` -> `Utilities` -> `Terminal`

 In our tutorials we sometimes use *terminal*, *command-line interface (CLI)* and *powershell* interchangeably.

Git Installation

Please install [Git](#) and [Git LFS](#) on your system

- 💡 Git LFS may already be installed with your Git installation (at least on Windows)
- 💡 For macOS we recommend to install via homebrew as described on the site above

Configuration of Git

Check the git user configuration on your system, by executing

```
git config --global --get-regexp user
```

This should prompt two lines

```
user.name <Your Name>
```

```
user.email <Your Email>
```

 Configuration needs to be done once after installation of git on your system.

Git configuration

Set the git user configuration on your system, by executing

1. Your name

```
git config --global user.name "Your Name"
```

2. Your email address

```
git config --global user.email "Your Email"
```

ARCitect Installation

Please follow the instructions to install the latest version of ARCitect.

- [macOS](#)
- [Windows](#)

Swate Installation

Please follow [these instructions](#) to install the latest version of Swate.

Have a simple text editor ready

- Windows Notepad
- MacOS TextEdit

Recommended text editor with code highlighting, git support, terminal, etc: [Visual Studio Code](#)

Resources



DataPLANT (nfdi4plants)

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

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

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

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

Contributors

Slides presented here include contributions by

- name: Dominik Brilhaus
github: <https://github.com/brilator>
orcid: <https://orcid.org/0000-0001-9021-3197>

Start your ARC Workshop

for CSCS

October 5th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

Block 1 – Welcome and Intro

Welcome

Please introduce yourselves

- Who are you?
- Where are you?
- What was your **motivation** to join this workshop?
- Summarize your **study design**
- Name your **favorite assay** or measurement technique

House-keeping

This workshop session will be recorded

I will cut out all participant chat, video, audio, etc. or ask for permission before sharing any recordings

Let's make this an interactive workshop

Please feel free to use the chat, raise hands, discuss, etc.

 Alt text

 Let's try to collect questions and answers in the Q&A panel =>

Goal

Create ARCs to share research data

- 💡 In this workshop we focus more on **how** and less on **why**

Tentative agenda

Time	Topics
13:00 - 14:00	Welcome and intro
14:00 - 14:15	<i>Short break</i>
14:15 - 16:00	ARC and ARCitect Hands-on
16:00 - 16:15	<i>Short break</i>
16:15 - 17:00	Q & A

 Please try to prepare your own ARC until the next session

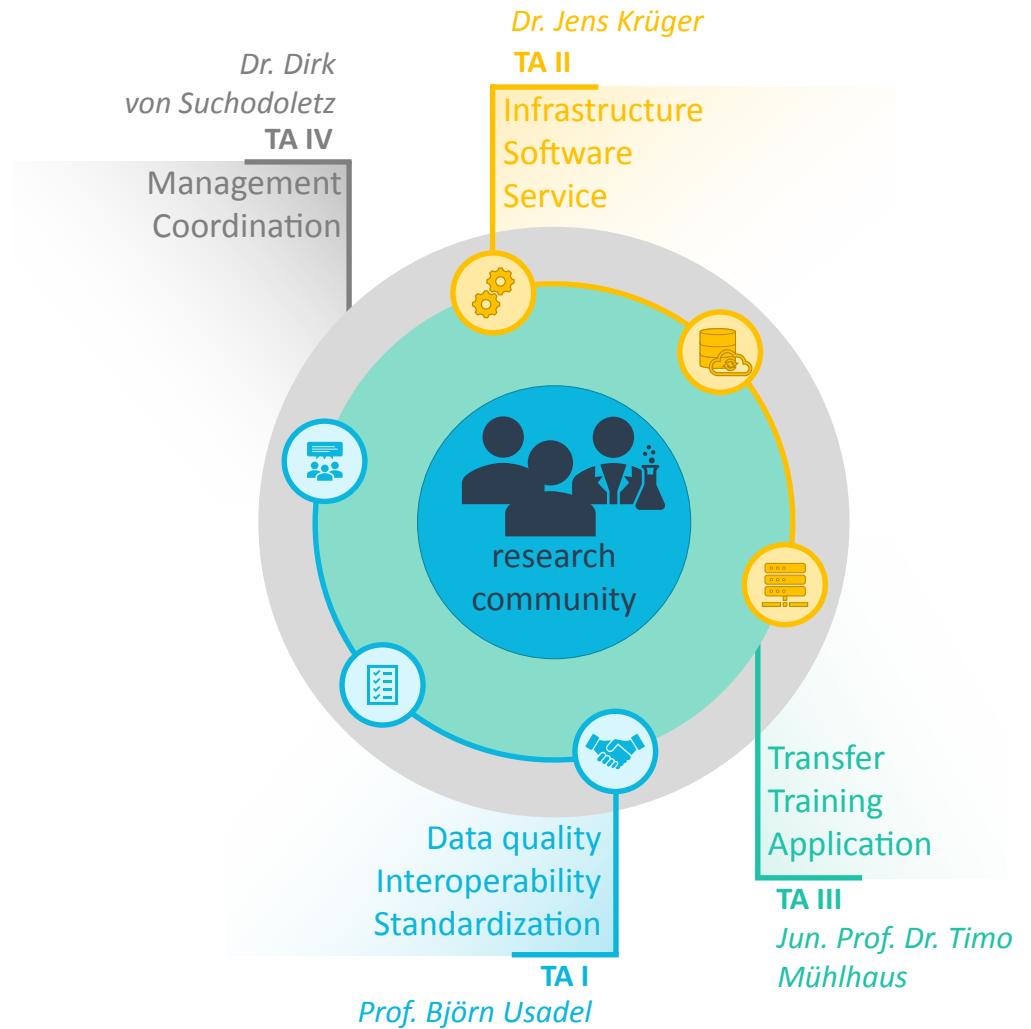
Block 2 – Intro to DataPLANT and ARC

October 5th, 2023

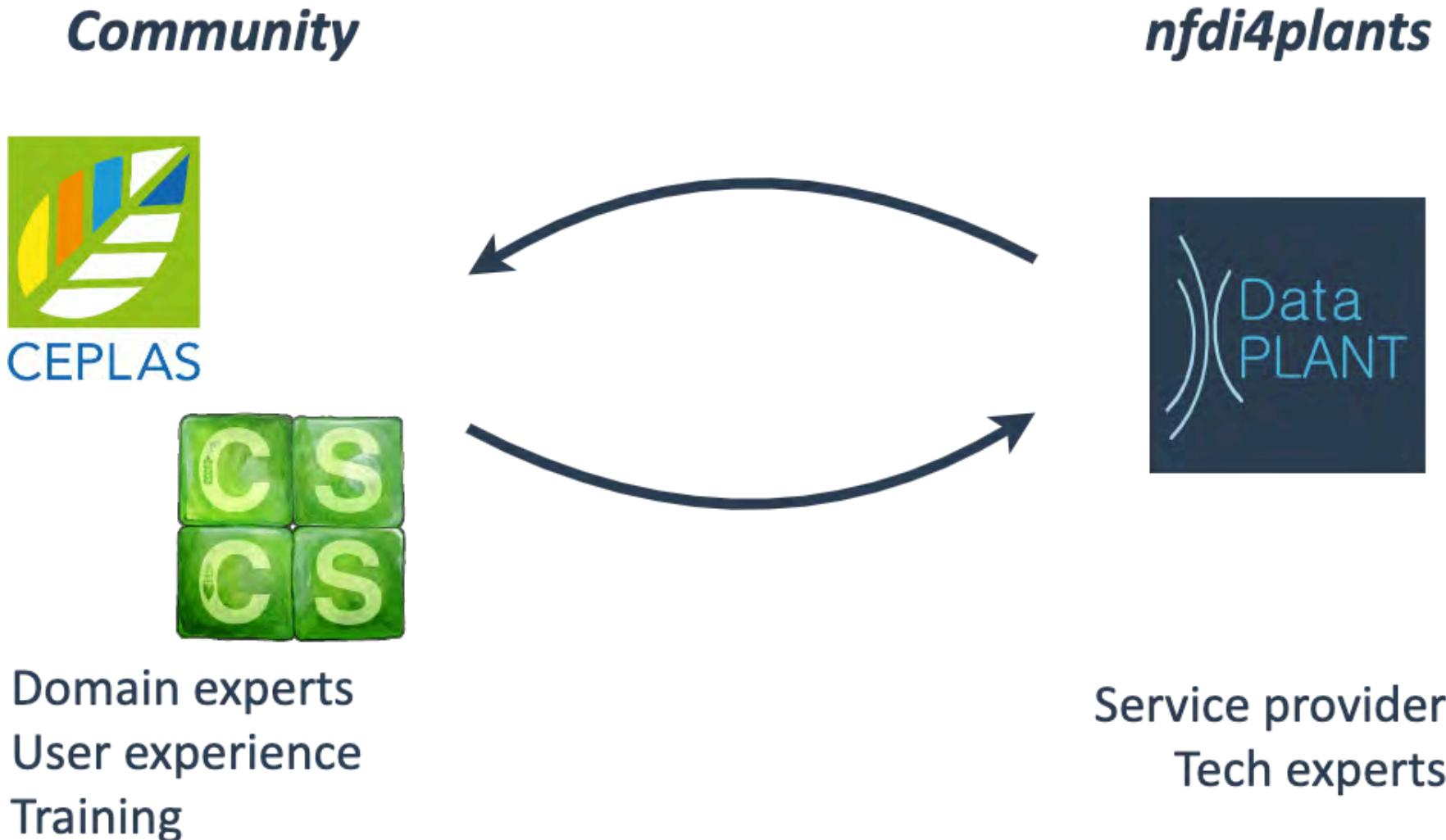
Dominik Brilhaus, [CEPLAS Data Science](#)

DataPLANT – The NFDI4Plants

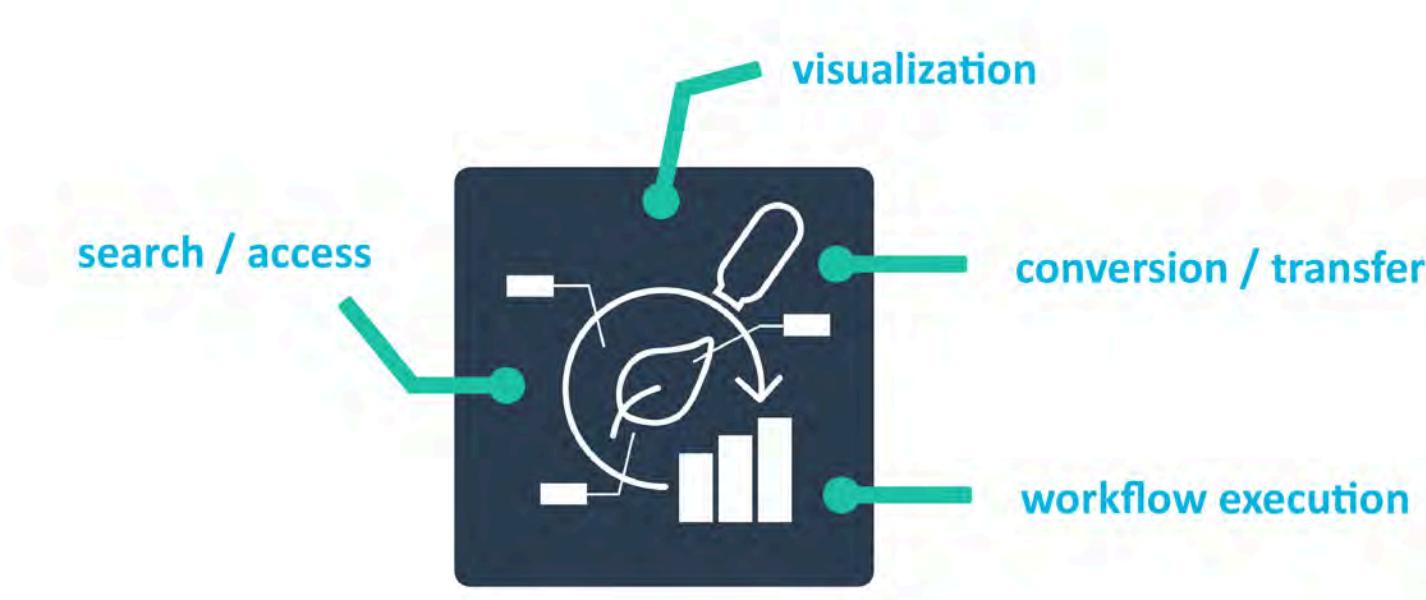
- NFDI: "Nationale Forschungsdaten Infrastruktur" – www.nfdi.de
- Funded since end of 2020



Data Stewardship between DataPLANT and the community



Annotated Research Context (ARC)





FINDABLE

ACCESSIBLE

INTEROPERABLE

REUSABLE

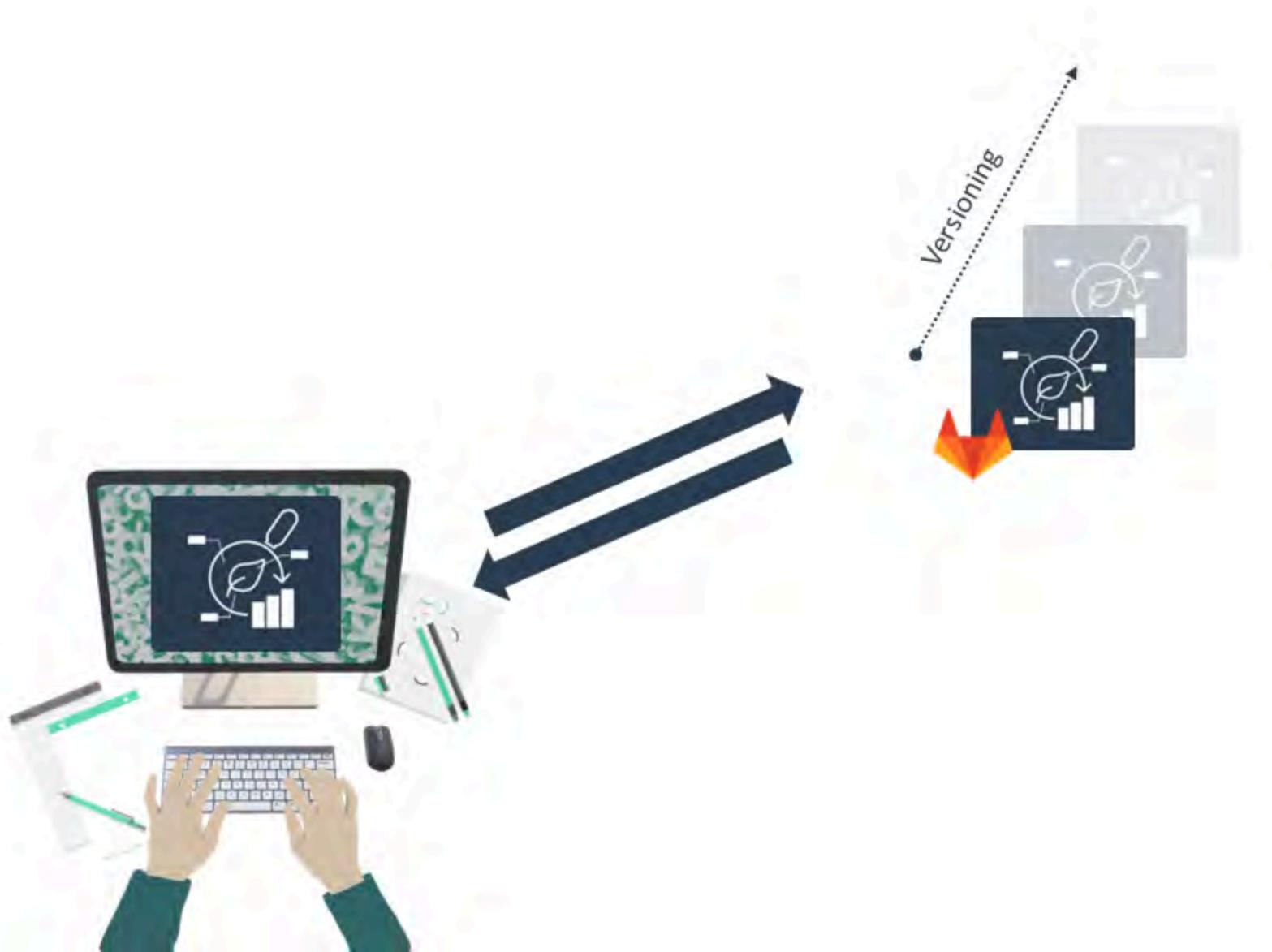
Metadata
Templates

Controlled
Vocabularies

Standards



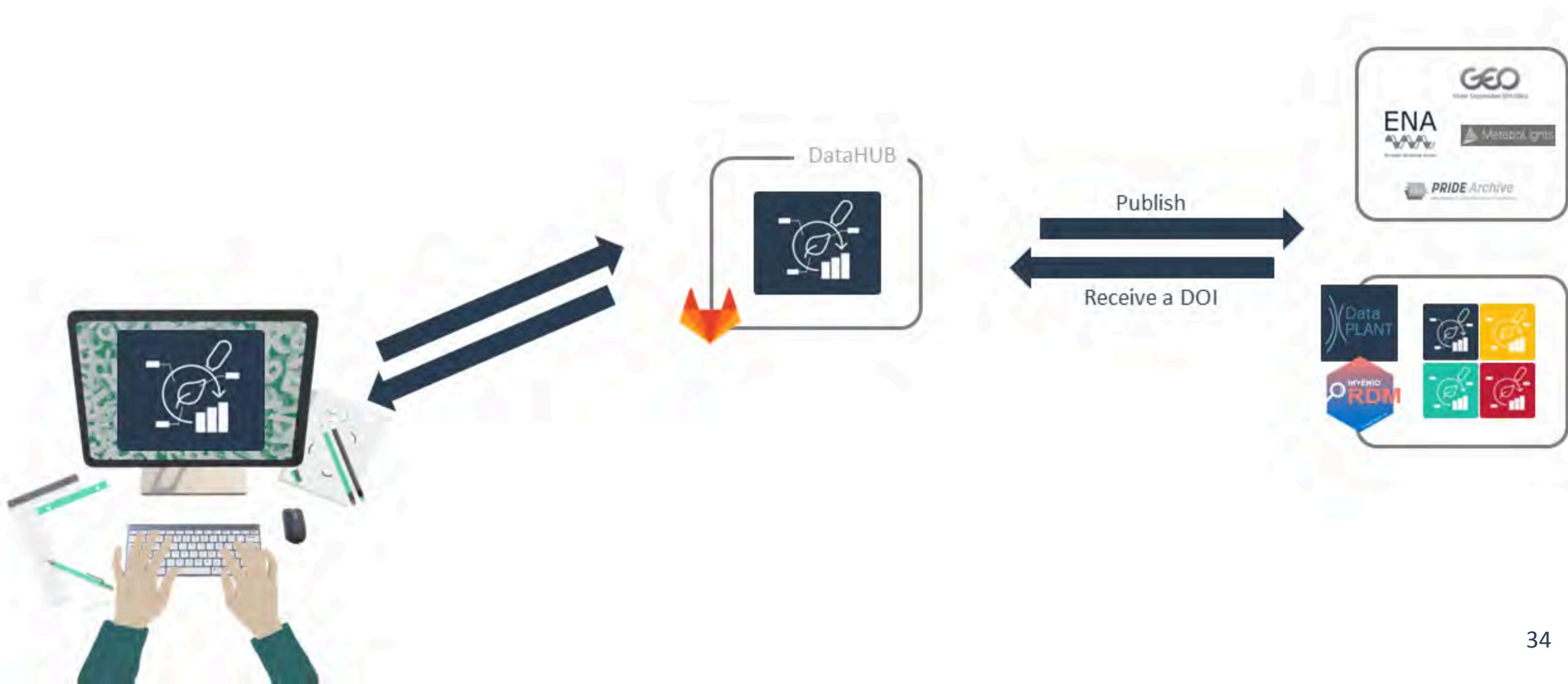


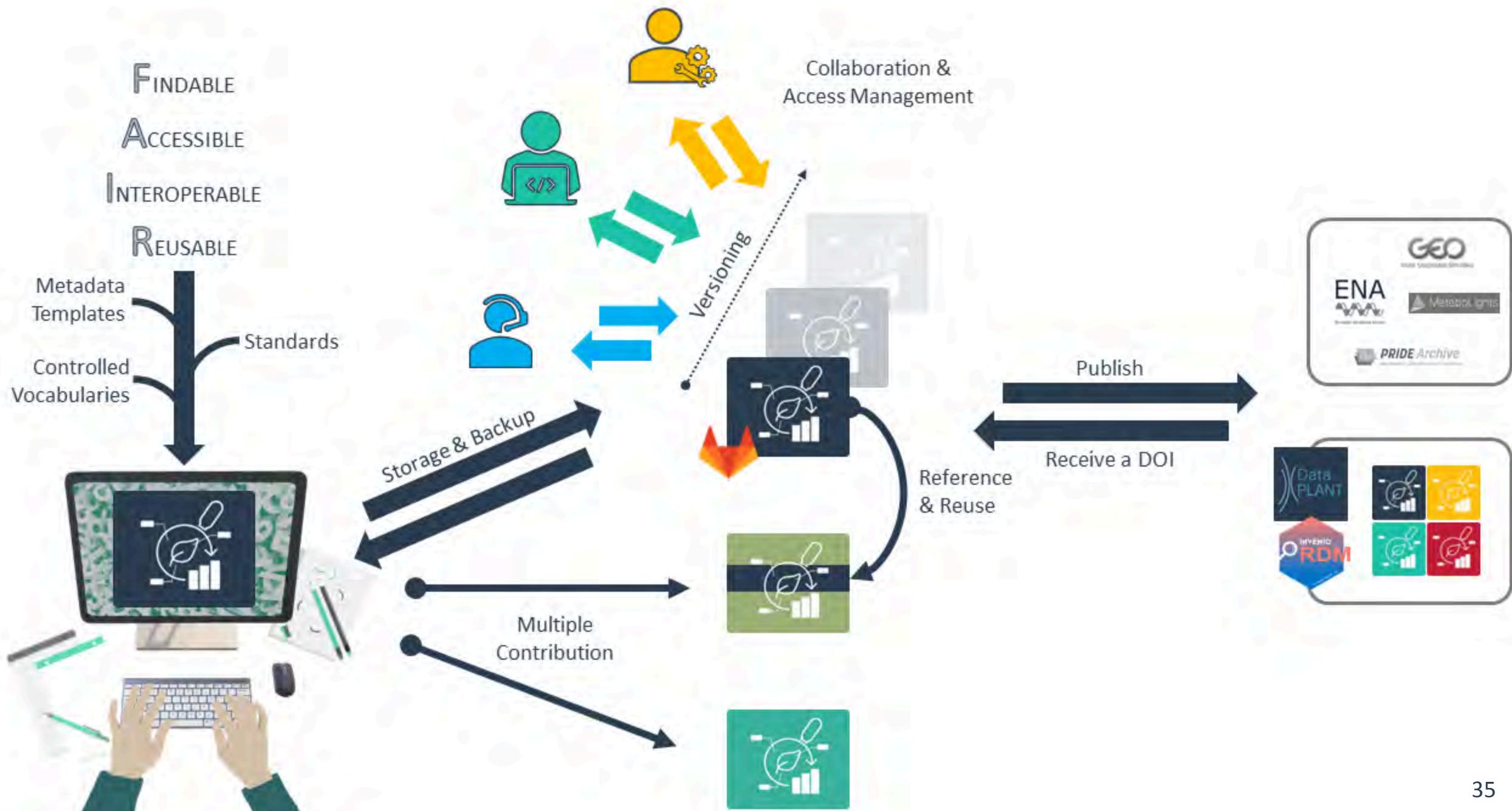




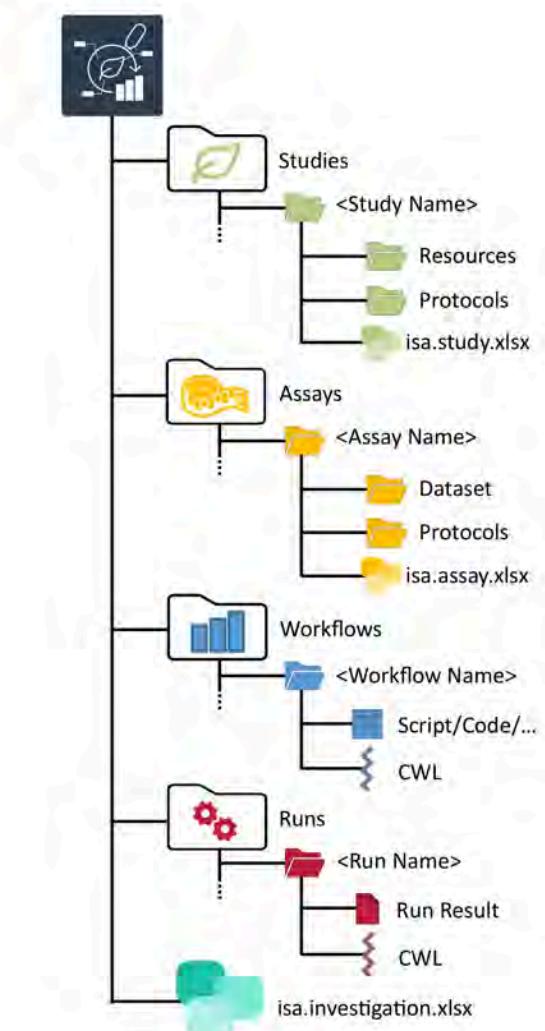




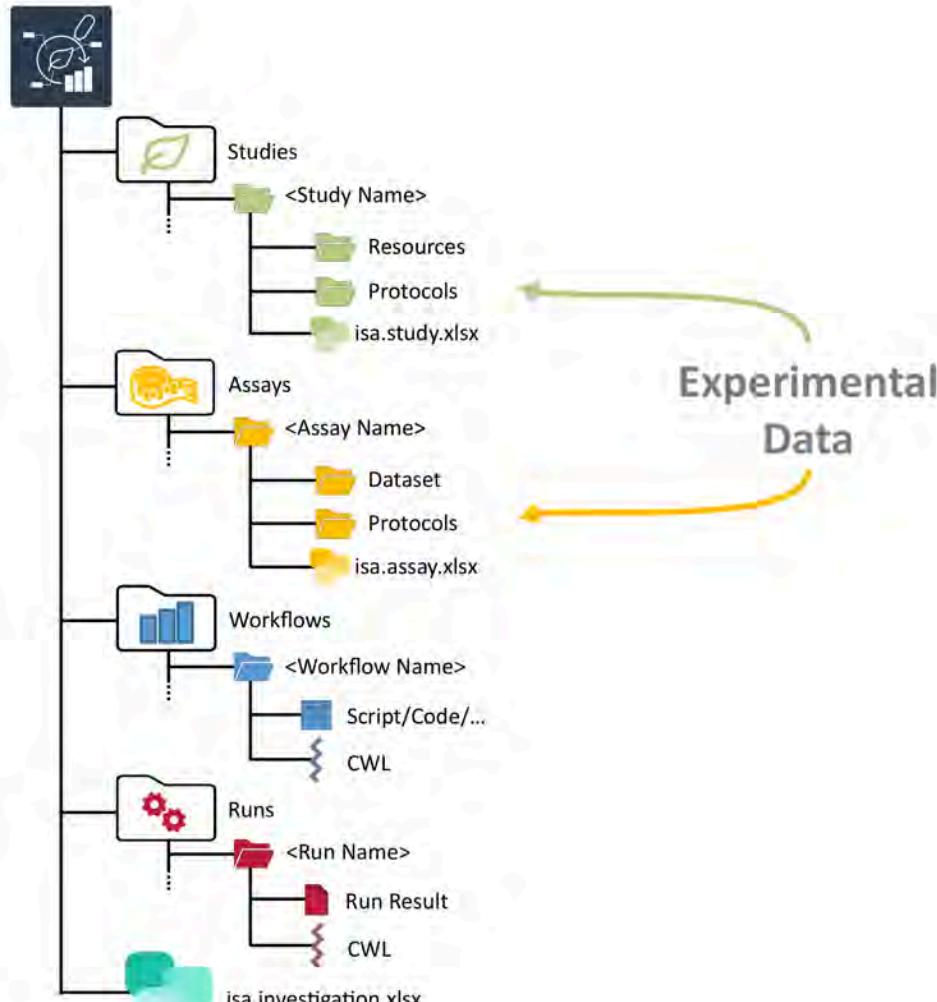




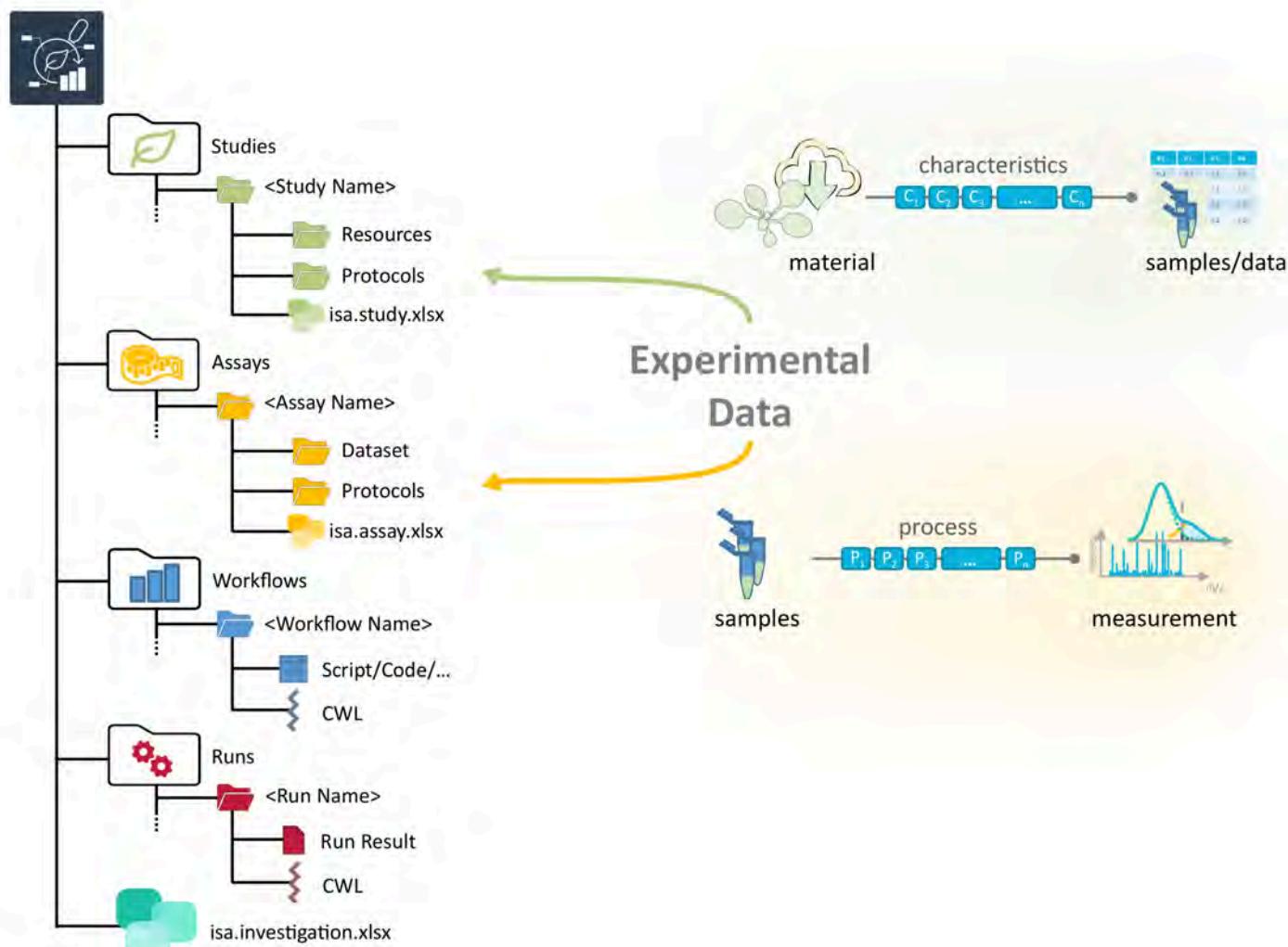
What does an ARC look like?



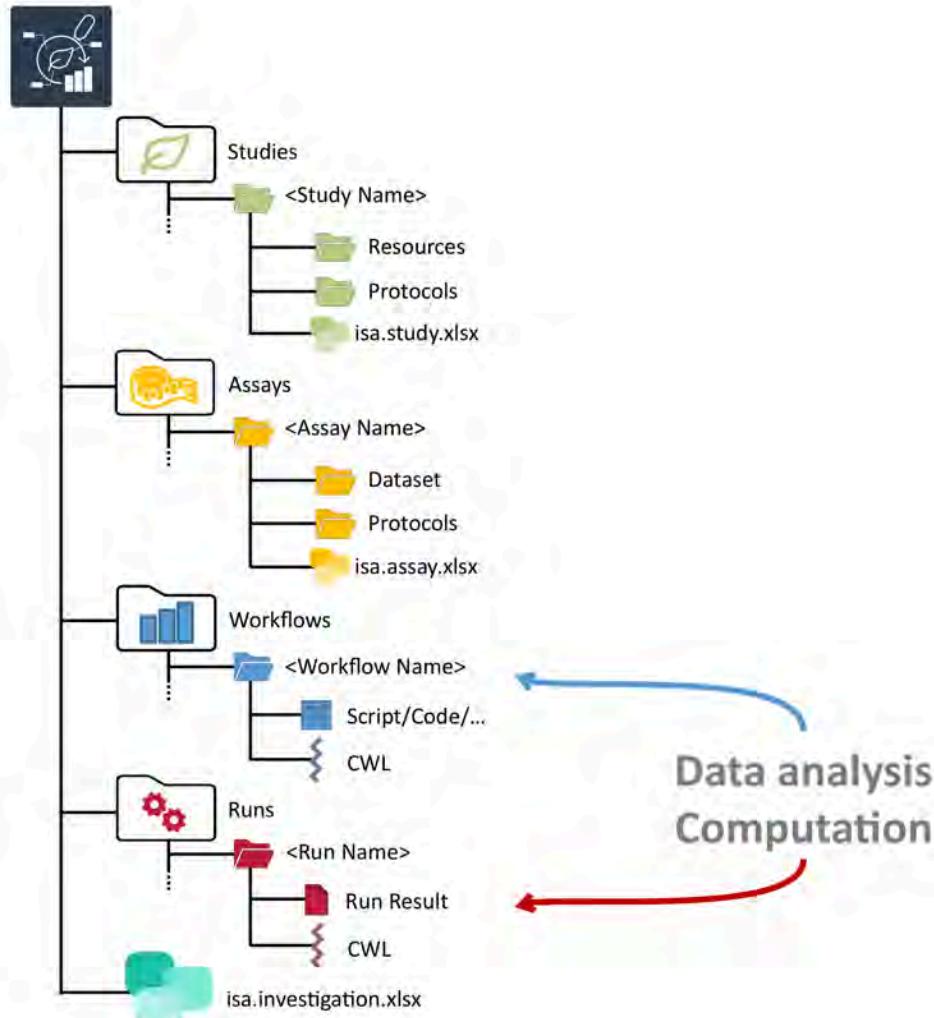
What does an ARC look like?



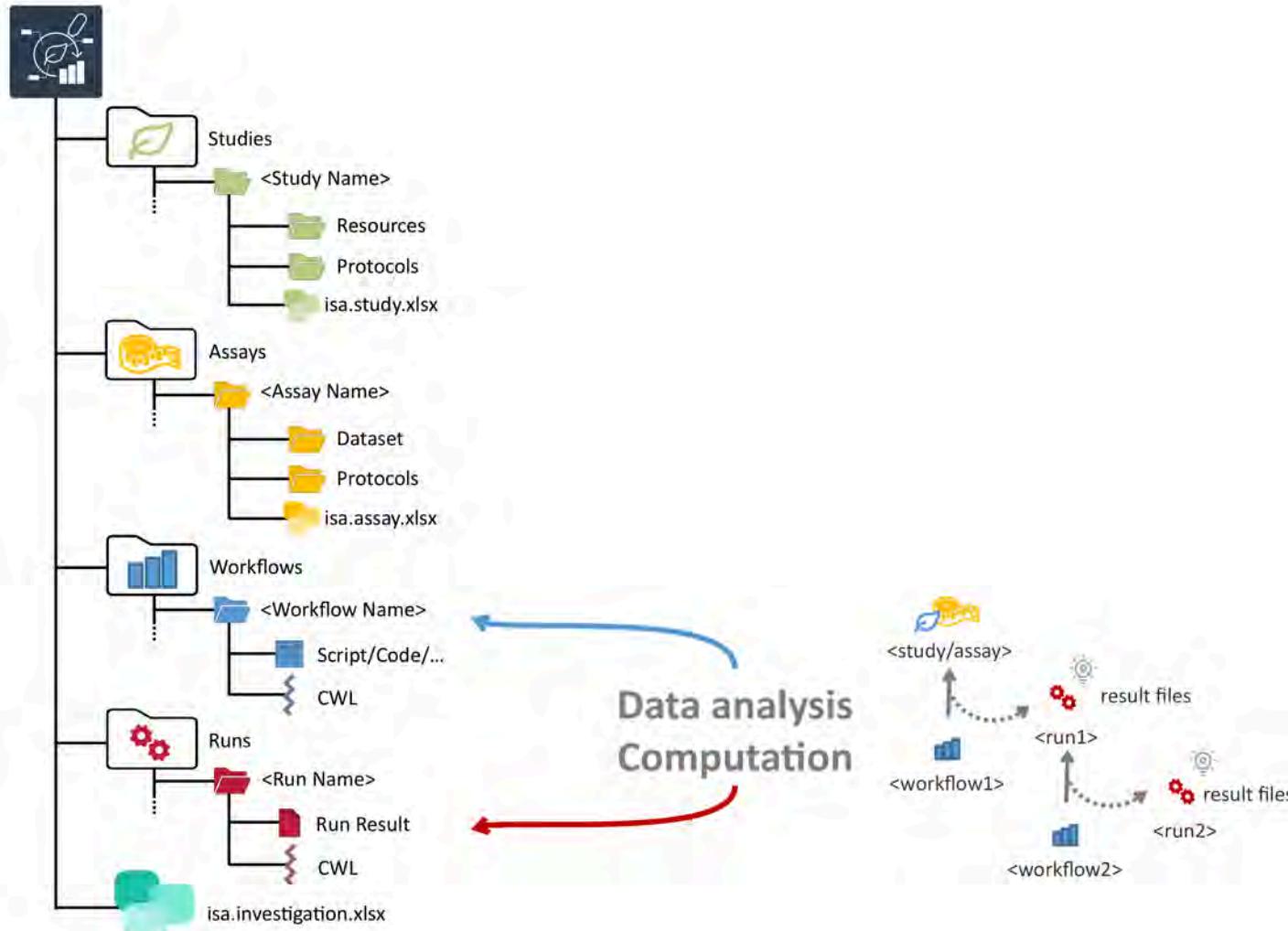
What does an ARC look like?



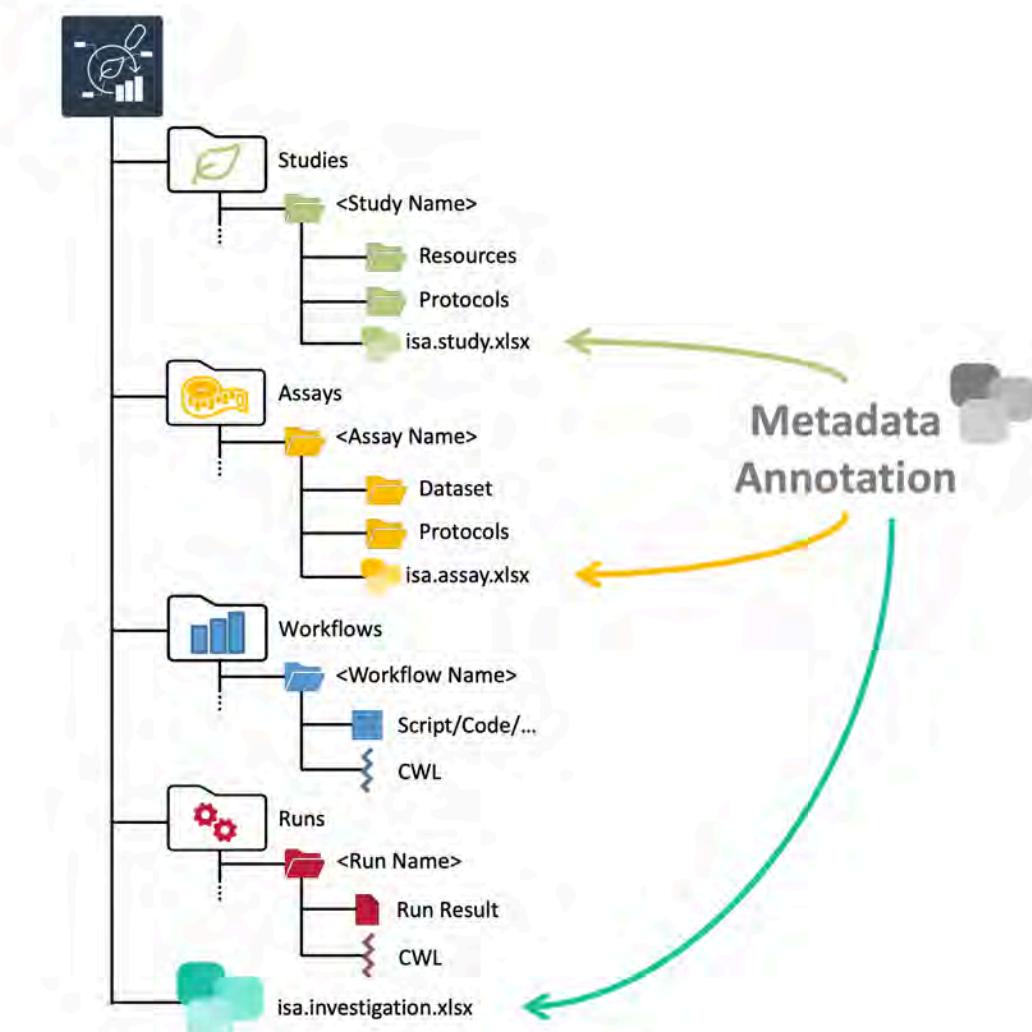
What does an ARC look like?



What does an ARC look like?



What does an ARC look like?



Resources



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...

Contributors

Slides presented here include contributions 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: Martin Kuhl
github: <https://github.com/Martin-Kuhl>
orcid: <https://orcid.org/0000-0002-8493-1077>

Block 3 – ARCitect Hands-on

October 5th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

Check-in

Registration

Did everyone [sign-up](#) at the DataHUB?

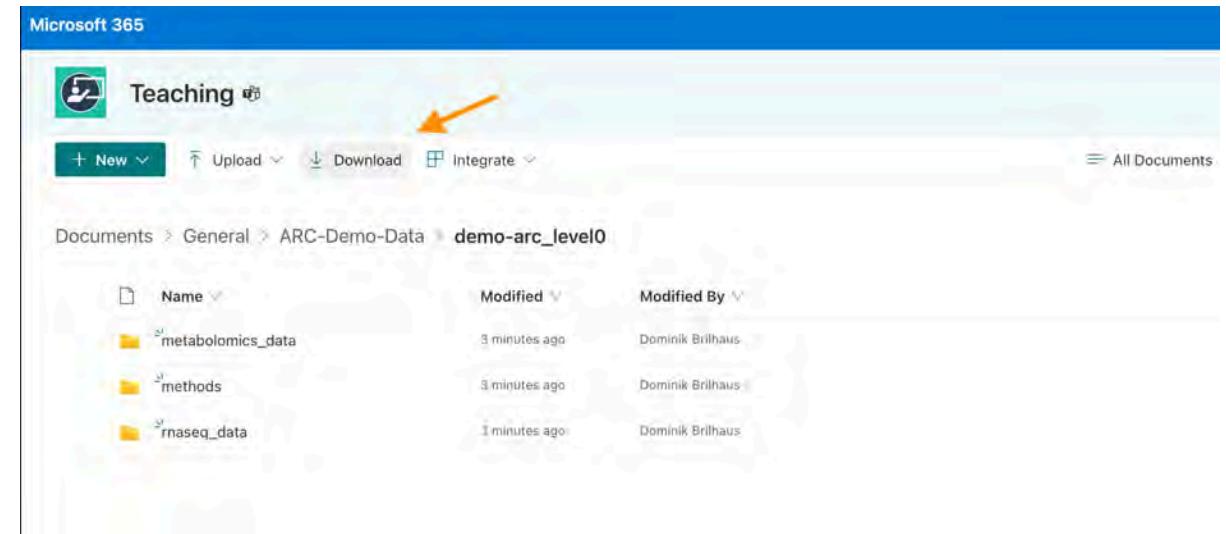
ARCitect installation

Please install the latest version of the ARCitect: <https://github.com/nfdi4plants/ARCitect>

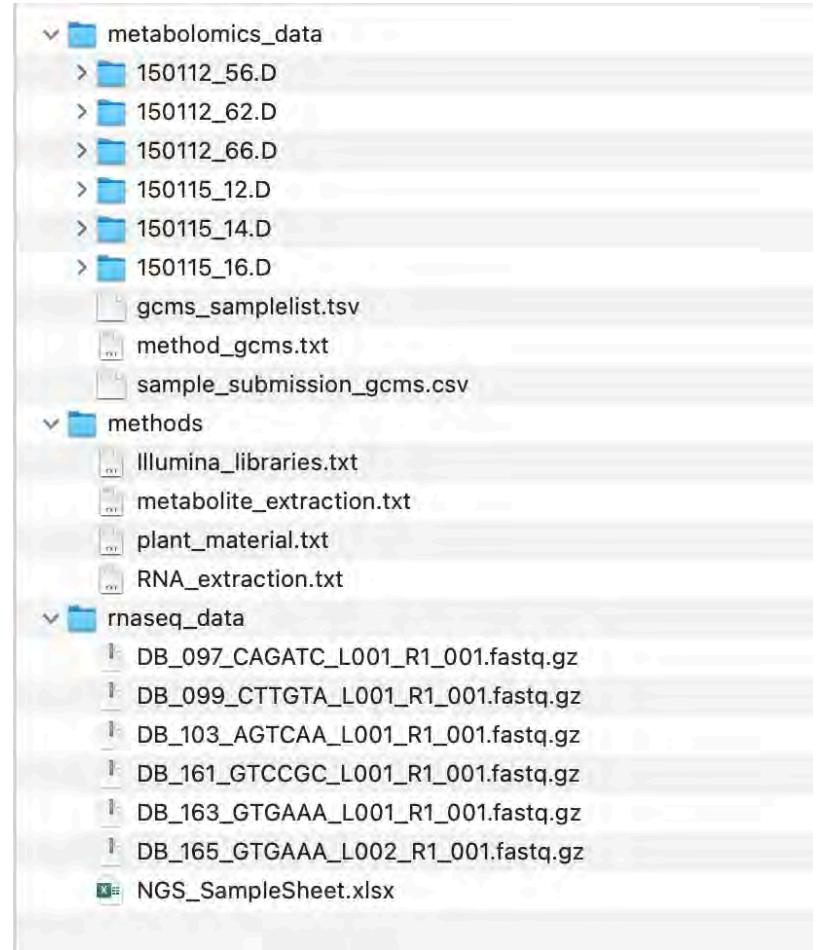
🔥 (released September 20th, 2023) 🔥

Download the demo data

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



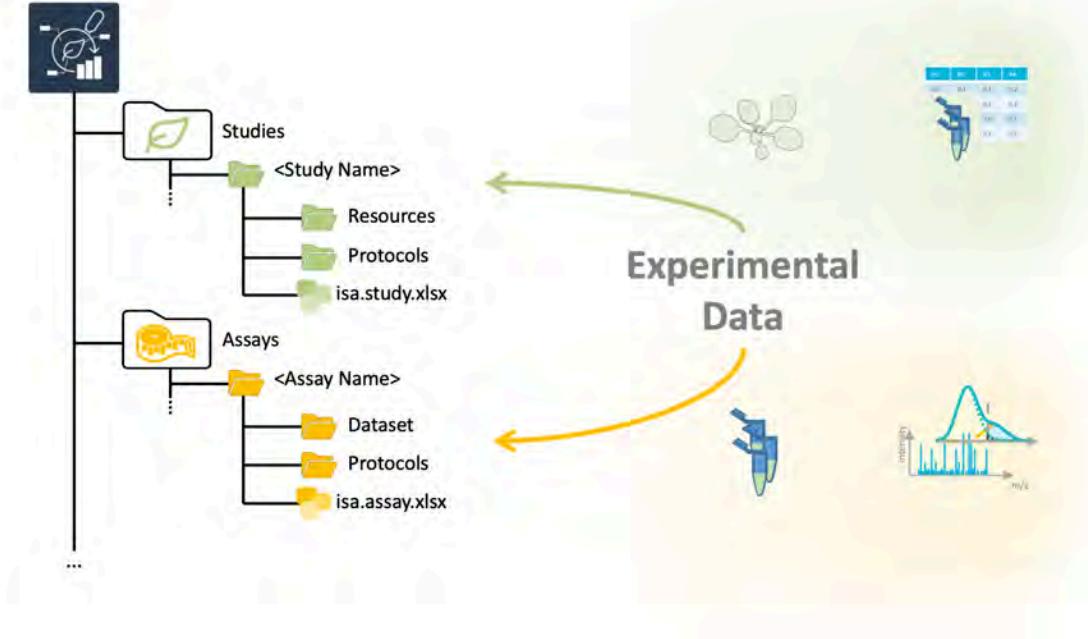
You just received your data



Goal

- Structure,
- (Annotate, and)
- Share your experimental data.

 We'll talk about data annotation
later

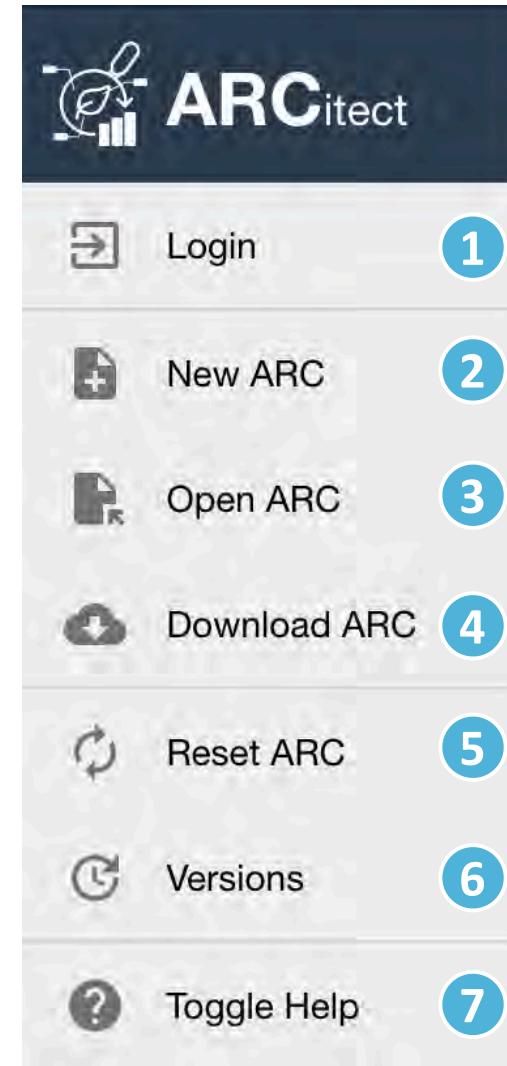


Open the ARCitect

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:

1. for the ARC folder on your machine
2. to create your ARC in the DataHUB at

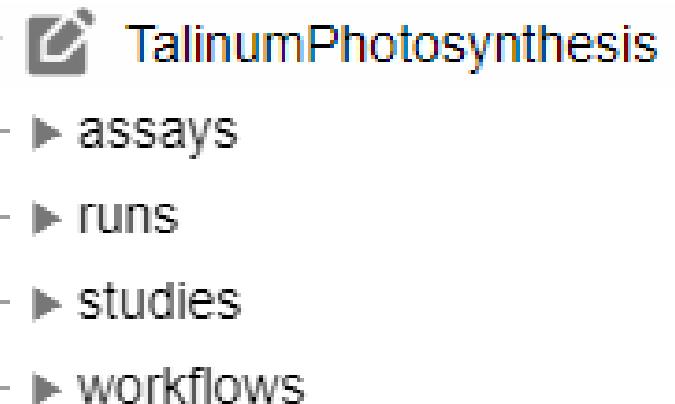
<https://git.nfdi4plants.org/<YourUserName>/<YourARC>> (see next steps)

3. as the identifier for your investigation

💡 Make sure that no ARC exists at

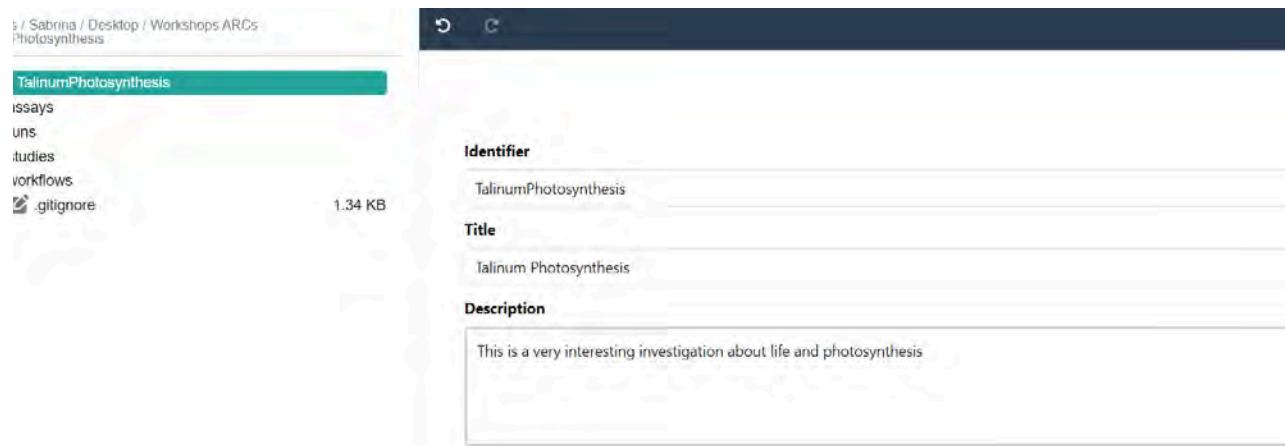
<https://git.nfdi4plants.org/<YourUserName>/<YourARC>> . Otherwise you will sync to that ARC.

💡 Avoid spaces in your ARC's name



Add a description and title to the investigation

1. Click on the ARC's name
2. Add a title (e.g. "Talinum Photosynthesis")
3. Add a description
4. Click "Update" to save your changes



Add contributors

In the section "People" click "ADD PERSON" to add at least one contributor

The screenshot shows a software interface with a sidebar and a main content area. The sidebar on the left displays a file tree structure:

- / Users / dominikbrilhaus / Desktop
- / TalinumPhotosynthesis
- └── ▶ **TalinumPhotosynthesis**
 - ▶ assays
 - ▶ runs
 - ▶ studies
 - ▶ workflows

The main content area contains several sections:

- Investigation**
General Meta Data of the Investigation
- People**
Authors and Collaborators
- Publications**
Papers, Books and Other Media

A prominent green button labeled **ADD PERSON** with a person icon is located in the lower right quadrant of the main content area.

Add contributor details

💡 For each person that you add, make sure to add

- First Name
- Last Name
- Email
- Affiliation

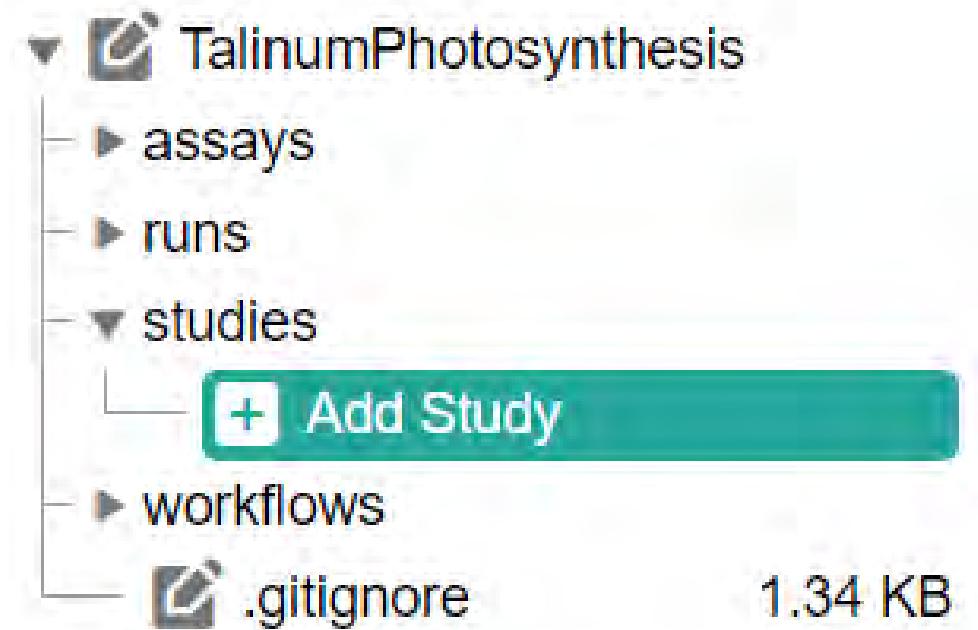
Contacts

Your First Name Your Last Name
Your ORCID 6/10

First Name Your First Name	Last Name Your Last Name	
Mid Initials <input type="text"/>	ORCID Your ORCID <input type="button" value="Search"/>	
Affiliation Your Affiliation	Address <input type="text"/>	
Email yourEmailAdress@uni.de	Phone <input type="text"/>	Fax <input type="text"/>
Roles 1. Author NCIT:NCIT:C42781 <input type="button" value="Delete"/> <input type="button" value="+"/>		

Add a study

by clicking "Add Study" and entering
talinum_drought as identifier for the
study



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 --:--

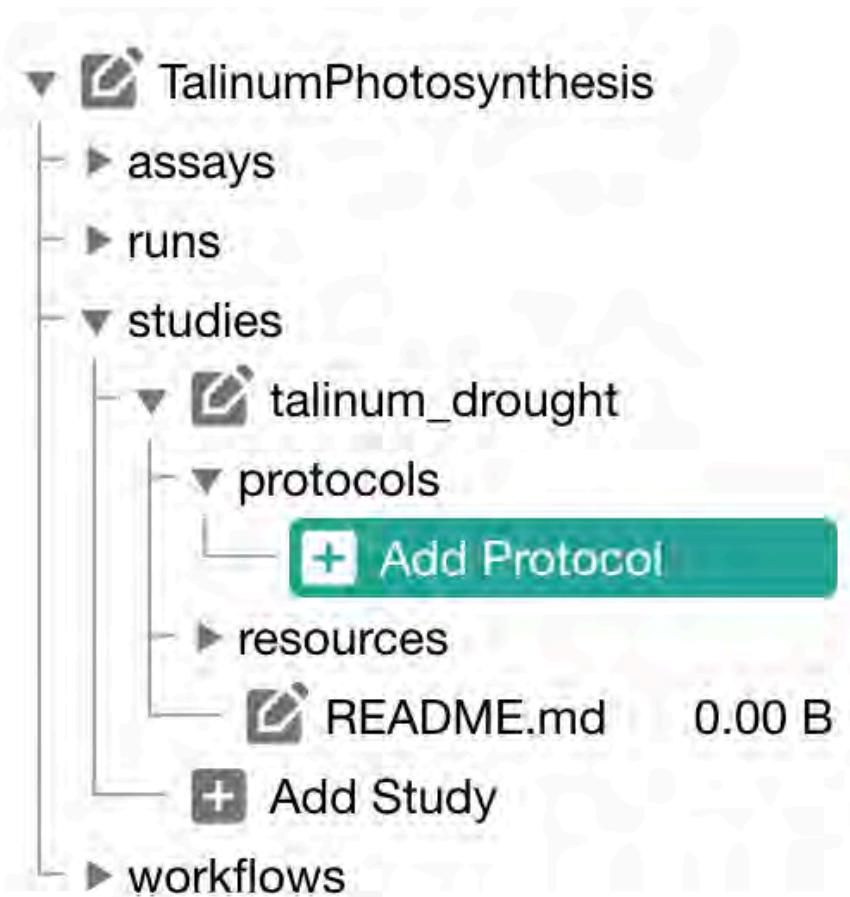
Public Release Date
tt.mm.jjjj --:--

Study Design Descriptors

+

Add a protocol to the study

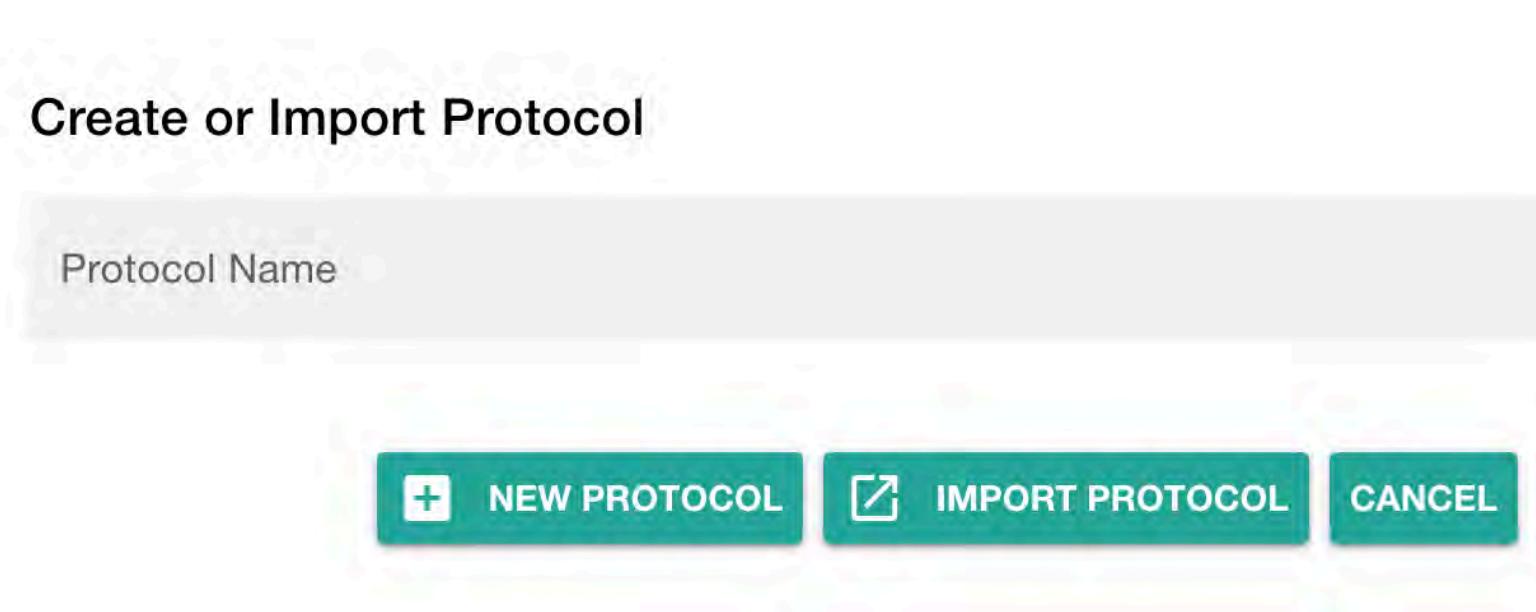
Click "Add Protocol" in the sidebar to add a protocol to the study



Adding protocols

You can either

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



Transfer the protocol information

From the demo data, transfer the lab notes stored in `plant_material.txt` to the `talinum_drought` study.

Add an assay to the ARC

Click "Add Assay" in the sidebar



Name and link the assay

1. Enter **rnaseq** as the identifier

2. Link the assay to the study

talinum_drought

Add Assay

Add Assay
rnaseq

Study Identifiers
talinum_drought 

talinum_drought

 ADD ASSAY CANCEL

Add information about the assay

In the assay panel you can

1. link or unlink the assay to studies, and
2. define the assay's
 - measurement type
 - technology type, and
 - technology platform.
3. add data process information

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 information about the assay

1. Add the following information:

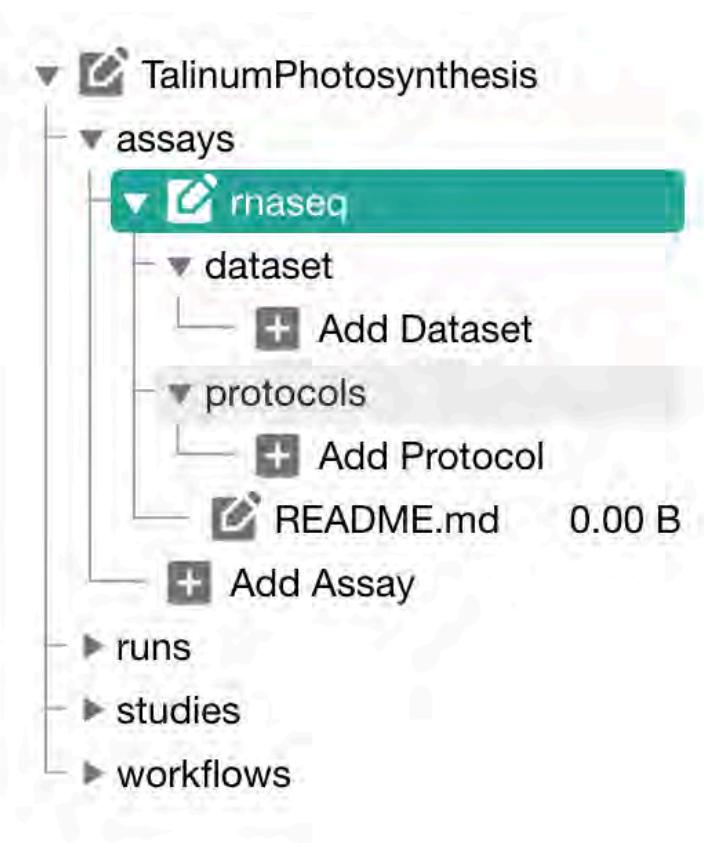
- Measurement Type: Gene Expression Analysis
- Technology type: Next Generation Sequencing
- Technology platform: Illumina HiSeq 2500

2. Click "Update" to save your changes

Add protocols and datasets

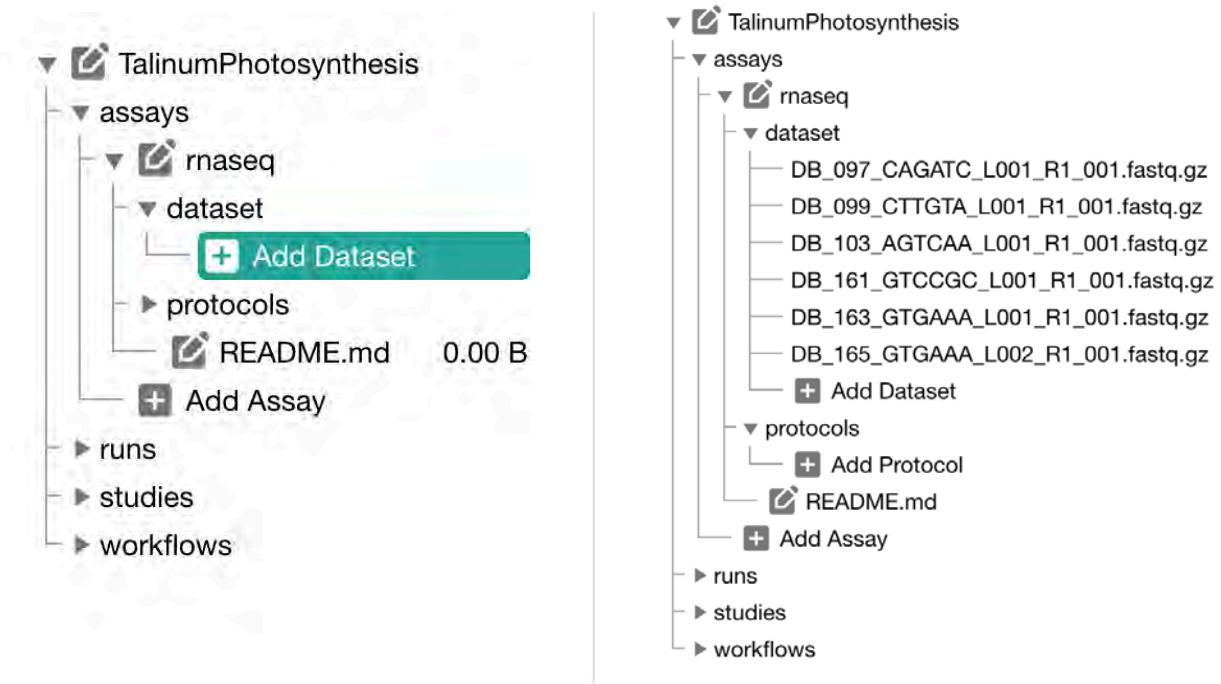
In the file tree you can

- **add a dataset** and
- **protocols** associated with that dataset



Import the demo dataset to the ARC

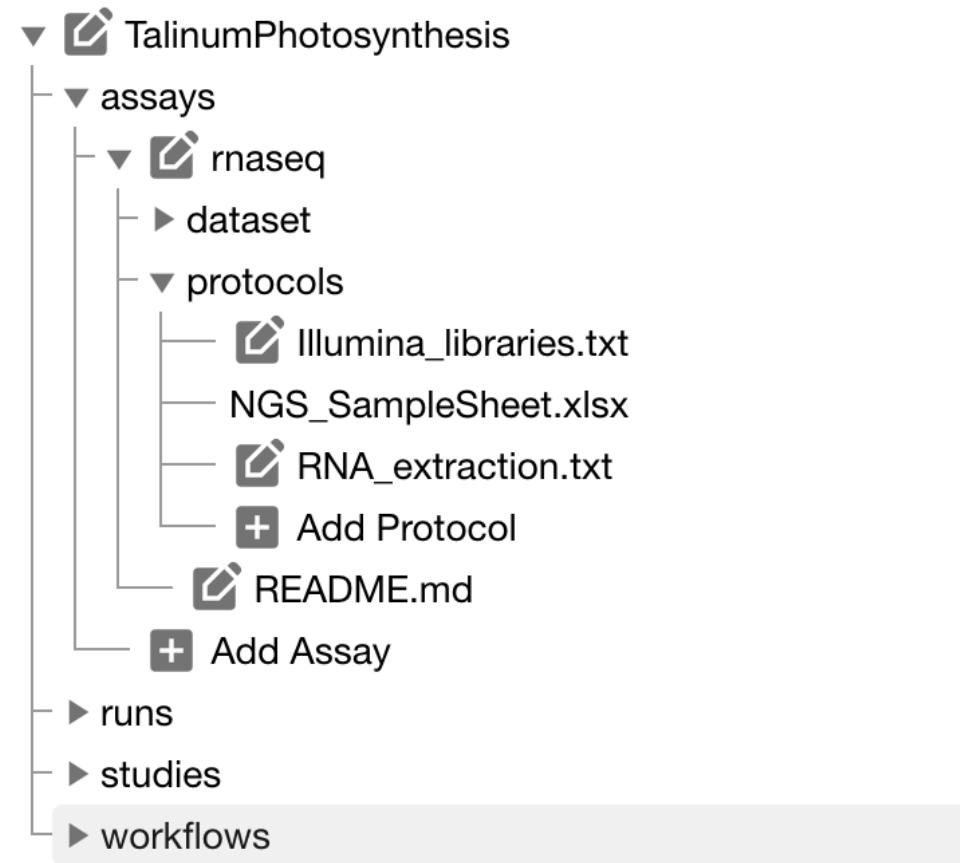
1. Click "Add Dataset"
2. Select the *.fastq.gz files from the demo data



Import the protocols

From the demo data, import the lab notes related to the `rnaseq` assay:

- `RNA_extraction.txt`
- `Illumina_libraries.txt`
- `NGS_SampleSheet.xlsx`



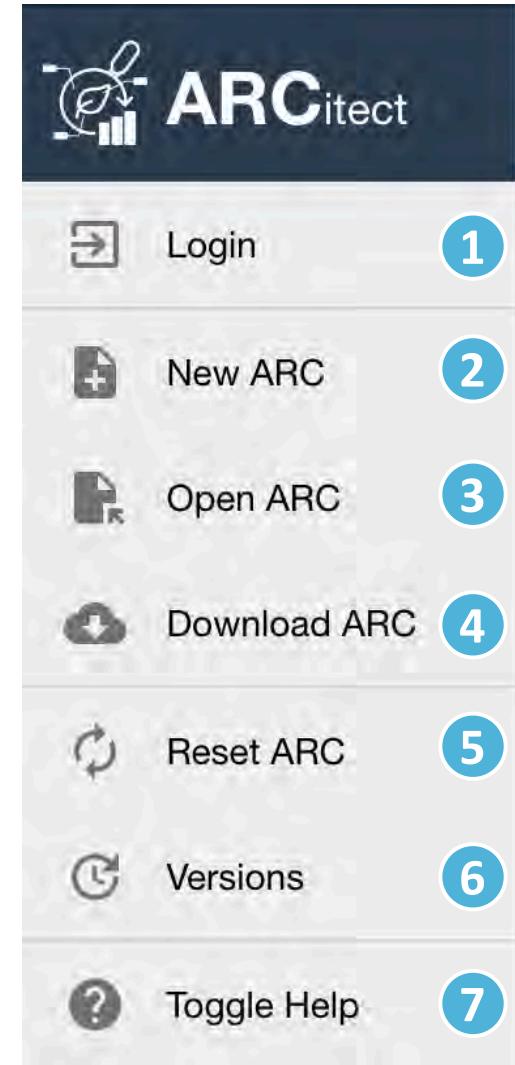
Collaborate and share



Login to the DataHUB

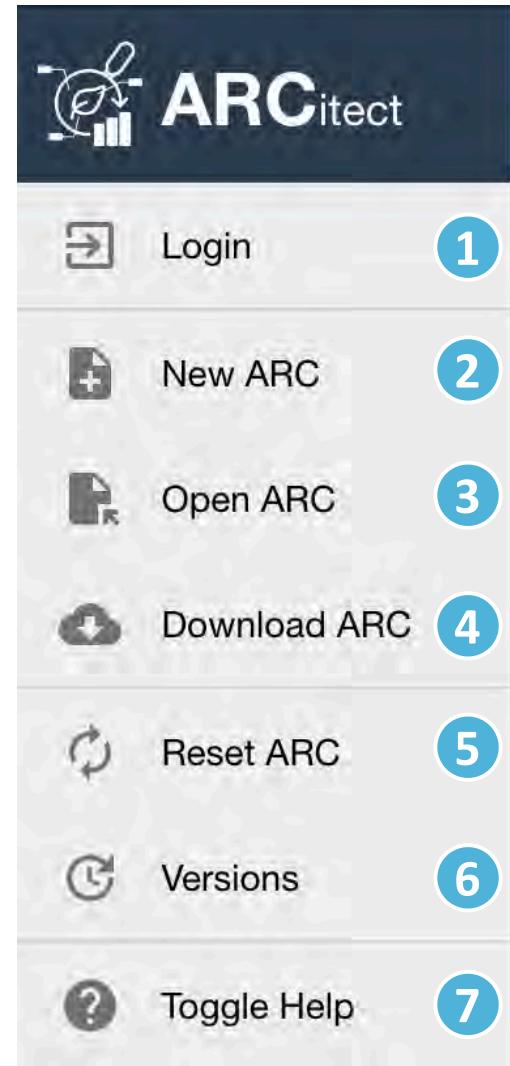
Click **Login** (1) in the sidebar to login to the DataHUB.

 This automatically opens your browser at the DataHUB (<https://git.nfdi4plants.org>) and asks you to login, if you are not already logged in.



Versions: Connection to the DataHUB

To communicate with the DataHUB, navigate to **Versions (6)**



Versions

The versions panel allows you to

- store the local changes to your ARC in form of "commits",
- sync the changes to the DataHUB, and
- check the history of your ARC

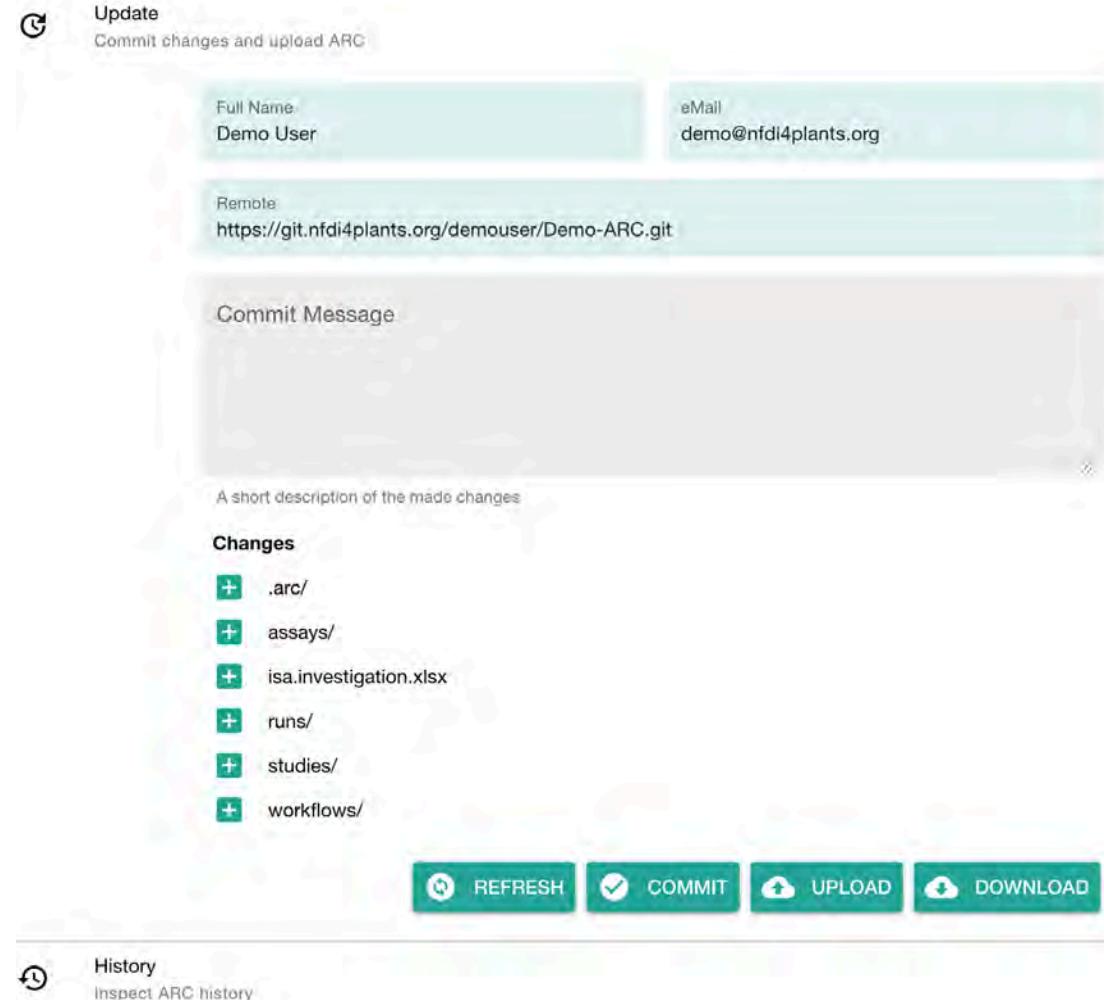
The screenshot shows the 'Update' section of the DataHUB interface. It includes fields for 'Full Name' (Demo User), 'eMail' (demo@nfdi4plants.org), and 'Remote' (https://git.nfdi4plants.org/demouser/Demo-ARC.git). A 'Commit Message' field is present with placeholder text: 'A short description of the made changes'. Below this is a 'Changes' section listing modified files: '.arc/', 'assays/', 'isa.investigation.xlsx', 'runs/', 'studies/', and 'workflows/'. At the bottom are buttons for 'REFRESH', 'COMMIT', 'UPLOAD', and 'DOWNLOAD'. A 'History' section at the bottom shows a timeline of recent commits.

Connection to the DataHUB

If you are logged in, the versions panel shows

- your DataHUB's *Full Name* and *eMail*
- the URL of the current ARC in the DataHUB

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



Upload your ARC to the DataHUB

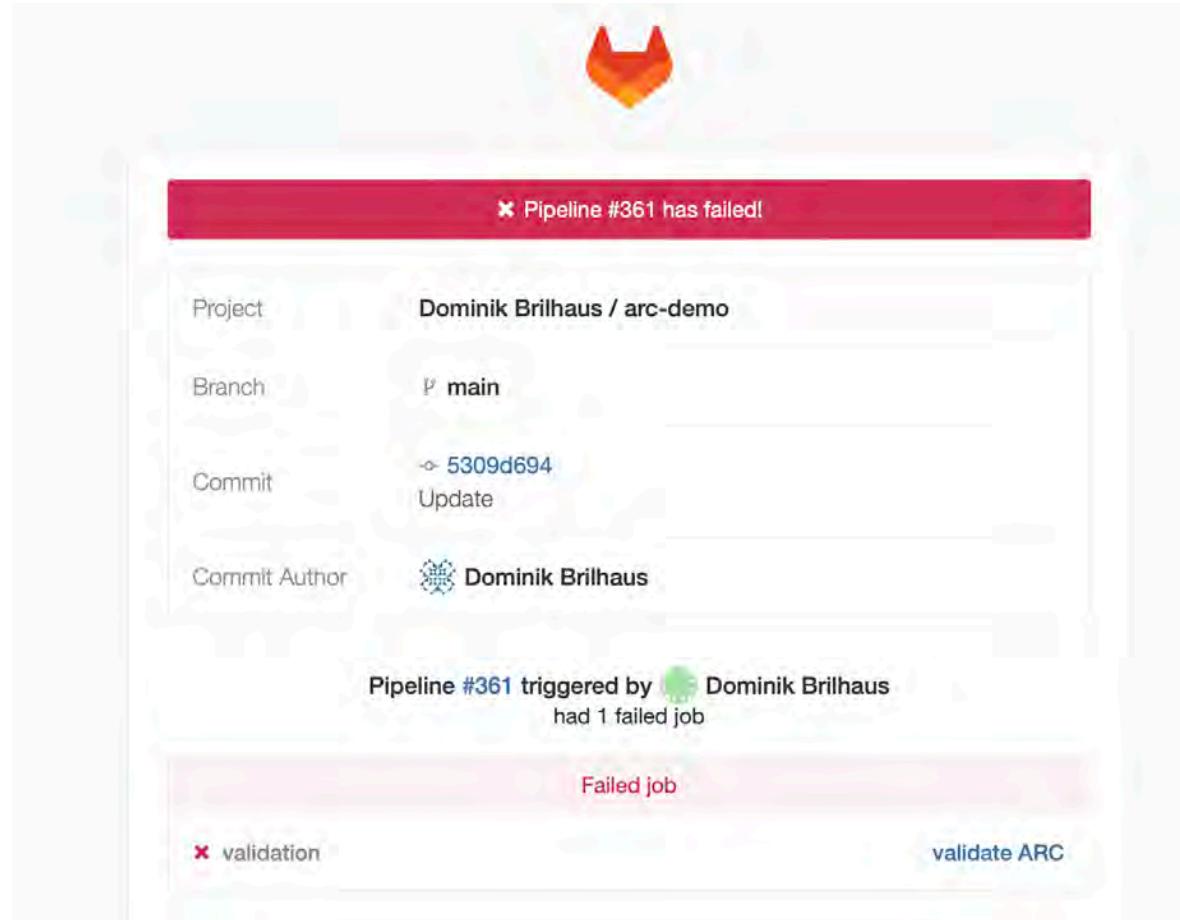
1. Enter a "commit message" to shortly describe the changes to your ARC
2. Click "COMMIT" to save your changes locally
3. Click "UPLOAD" to upload your ARC to the DataHUB

The screenshot shows a user interface for committing changes to an ARC. At the top, there's a header with the text "Update" and "Commit changes and upload ARC". Below this, there are fields for "Full Name" (set to "Demo User") and "eMail" (set to "demo@nfdi4plants.org"). A "Remote" field shows the URL "https://git.nfdi4plants.org/demouser/Demo-ARC.git". A "Commit Message" input field is present with placeholder text "A short description of the made changes". Under the "Changes" section, a list of modified files is shown: ".arc/", "assays/", "isa.investigation.xlsx", "runs/", "studies/", and "workflows/". At the bottom, there are four buttons: "REFRESH", "COMMIT" (which has a checkmark icon), "UPLOAD", and "DOWNLOAD". Below the buttons, a "History" section is visible with the text "Inspect ARC history".

Check whether your ARC was uploaded successfully

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

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

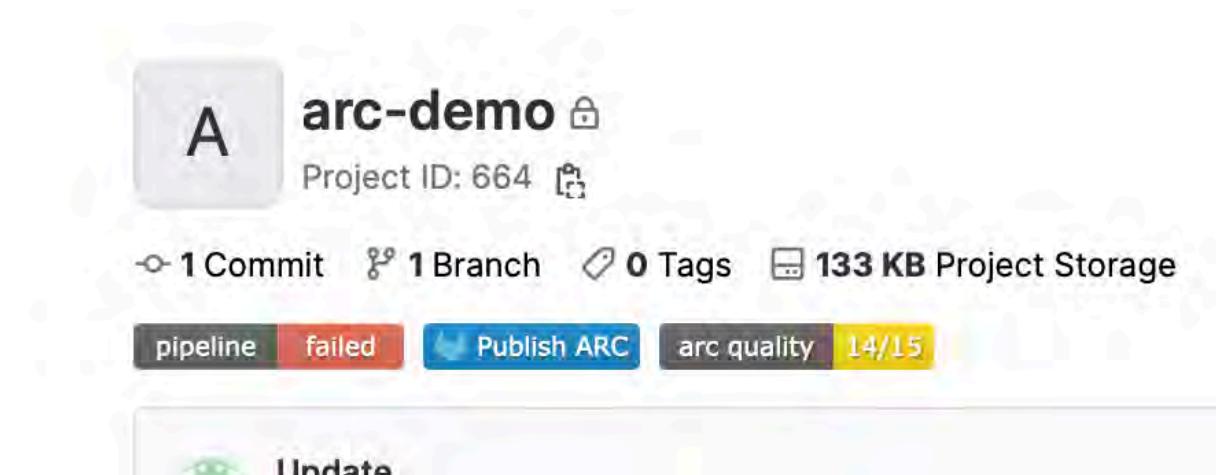


🔥 Don't worry 😊

Pipeline Failed

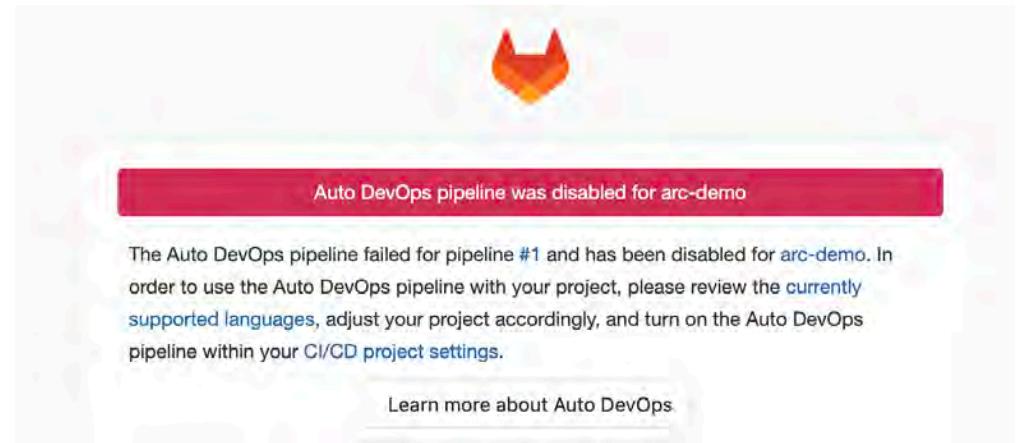
- a "continuous quality control" (CQC) pipeline validates your ARC
- This fails if one of the following metadata items is missing:

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



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 'CI/CD' settings page in GitLab. On the left, there is a sidebar with various project management and monitoring tools like Security & Compliance, Deployments, Packages and registries, Infrastructure, Monitor, Analytics, Wiki, Snippets, Settings (General, Integrations, Webhooks, Access Tokens, Repository, Merge requests), and CI/CD. The 'CI/CD' section is currently selected. On the right, the main content area is titled 'Auto DevOps'. It contains a sub-section titled 'Default to Auto DevOps pipeline' which is checked and labeled 'instance enabled'. Below this, there is a note about adding a Kubernetes cluster integration or creating an environment variable. Under 'Deployment strategy', the 'Continuous deployment to production' radio button is selected. At the bottom, there is a 'Save changes' button.

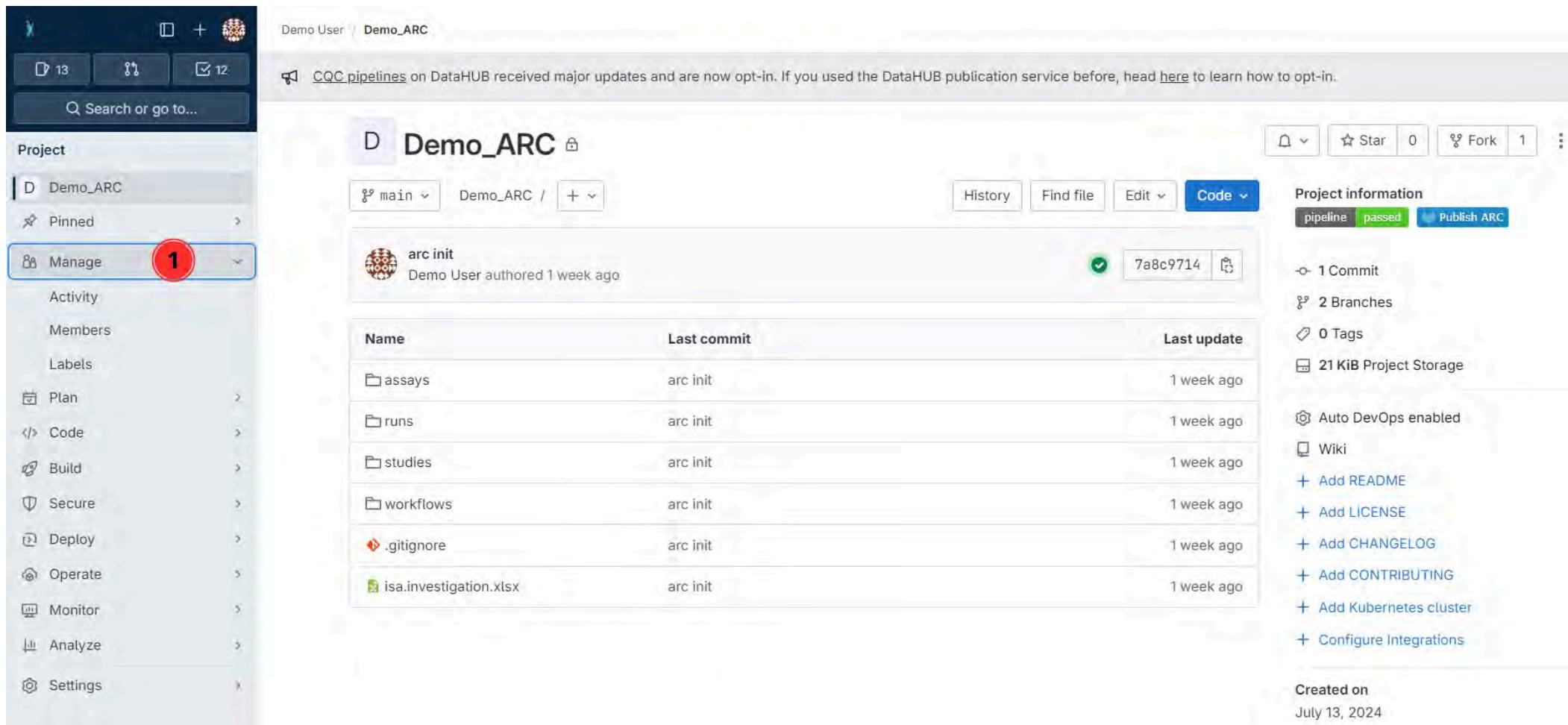
Collaborate and share



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

1. Click on Project Information in the left navigation panel



The screenshot shows the DataHUB project management interface for the 'Demo_ARC' project. The left sidebar has a red circle around the 'Manage' button, which is highlighted with a blue border. The main content area displays a list of files and their commit history. On the right, there is a 'Project information' sidebar with various metrics and configuration options.

Project Information Sidebar:

- pipeline: passed
- Published ARC
- 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

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

2. Click on Members

The screenshot shows the DataHUB interface with the following details:

- Project Sidebar:** On the left, there is a sidebar with various project management options: Demo ARC (selected), Pinned, Manage (circled with red number 1), Activity, Members (circled with red number 2), Labels, Plan, Code, Build, Secure, Deploy, Operate, Monitor, Analyze, and Settings.
- Breadcrumbs:** Demo User / Demo_ARC / Members
- Notification Bar:** 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 Table:** A table showing one member:

Account	Source	Max role	Expiration	Activity
Demo User @DemoUser	Direct member by Demo User	Owner	Expiration date 8+ Sep 27, 2023 ✓ Jul 13, 2024 ✗ Jul 21, 2024	⋮
- Buttons:** Import from a project, Invite a group, and Invite members.

3. Click on Invite members

The screenshot shows the DataHUB interface for a project named 'Demo_ARC'. The left sidebar has a 'Project' section with 'Demo_ARC' selected, and a 'Members' item highlighted with a red circle labeled '2'. The main area is titled 'Project members' and displays a table of members. The table has columns: Account, Source, Max role, Expiration, and Activity. One member is listed: 'Demo User @DemoUser' (highlighted with a red circle labeled '1'), who is a 'Direct member by Demo User' with a 'Owner' role. The expiration date is set to 'Sep 27, 2023' with an 'Expiration date' button. The activity shows 'Jul 13, 2024' and 'Jul 21, 2024'. At the top right of the main area, there are three buttons: 'Import from a project', 'Invite a group', and 'Invite members' (highlighted with a red circle labeled '3'). A banner at the top of the main area states: '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.'

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

4. Search for potential collaborators

Invite members X

You're inviting members to the **Demo_ARC** project.

Username, name or email address 4

Select members or type email addresses

Select a role

Guest ▼

[Read more about role permissions](#)

5. Select a role

The screenshot shows a user interface for selecting a role. On the left, there is a sidebar with five roles listed: Guest (selected), Reporter, Developer, Maintainer, and Owner. Below this is a dropdown menu also labeled 'Guest'. At the bottom, there is a link to 'Read more about role permissions'. A red circle with the number '4' is overlaid on the 'ARC project.' button, which is located to the right of the sidebar. A red circle with the number '5' is overlaid on the 'Guest' dropdown menu.

Guest

ARC project.

4

Reporter

Developer

Maintainer

Owner

Guest

5

Read more about role permissions

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.

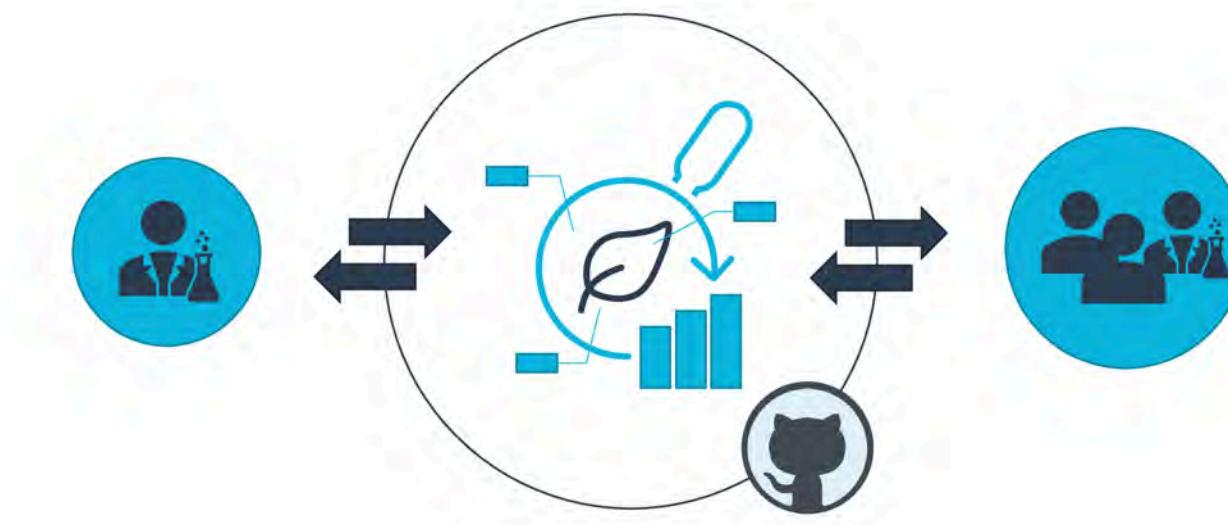
Maintainers

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](#)

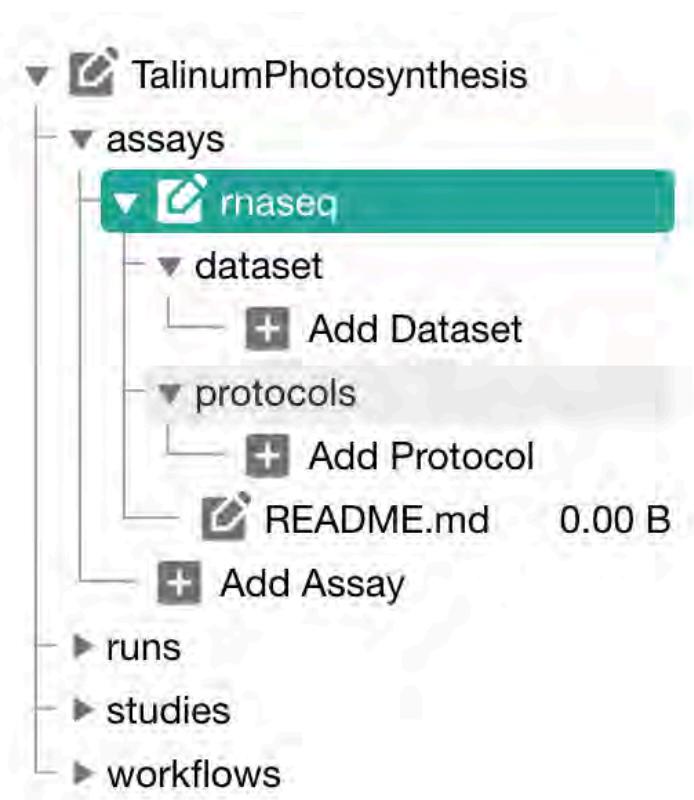
Congratulations!

You have just shared your ARC with a collaborator.



Add the remaining data

1. Add another assay (metabolomics)
 - i. Add the protocols
 - ii. Add the dataset
 2. Go to the Versions panel
 - i. Add a "commit message"
 - ii. Upload your changes to the DataHUB



Check the progress of your ARC

1. Navigate to Versions

2. Check the History panel at the bottom

The screenshot shows a file tree on the left and a history panel on the right.

File Tree:

- / Users / dominikbrilhaus / gitlab_dataplant
- / 2023-10-05_CSCS-Workshop
- / TalinumPhotosynthesis
- assays
 - ▶ metabolomics
 - ▶ rnaseq
 - + Add Assay
- ▶ runs
- studies
 - ▶ talinum_drought
 - + Add Study
- ▶ workflows

History Panel:

- 06.10.2023 11:02 - DOMINIK BRILHAUS (BRILHAUS@NFDI4PLANTS.ORG)
add assay metabolomics
- 05.10.2023 11:02 - DOMINIK BRILHAUS (BRILHAUS@NFDI4PLANTS.ORG)
add assay rnaseq
- 05.10.2023 11:01 - DOMINIK BRILHAUS (BRILHAUS@NFDI4PLANTS.ORG)
add study
- 05.10.2023 11:01 - DOMINIK BRILHAUS (BRILHAUS@NFDI4PLANTS.ORG)
Create empty ARC

Your ARC is ready

 Initiated an ARC

 Structured and ...

 ... annotated experimental data

 Shared with collaborators



DataHUB

Deleting an ARC

1. Click on *Settings* in the sidebar of your ARC
2. Navigate to the general (1) settings
3. In the advanced section (4) you can delete your ARC

The screenshot shows the 'General Settings' page for a project named 'Demo_ARC'. The sidebar on the left lists various project management options. Step 1 is highlighted over the 'General' settings option. Step 2 is highlighted over the 'Project name' field, which contains 'Demo_ARC'. Step 3 is highlighted over the 'Project avatar' section, where a placeholder image 'D' is shown and a file upload button is visible. Step 4 is highlighted over the 'Advanced' section at the bottom of the page.

Demo User / Demo_ARC / General Settings

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.

Search or go to...

Project

- Settings
- General** (1)
- Integrations
- Webhooks
- Access Tokens
- Repository
- Merge requests
- CI/CD
- Packages and registries
- Monitor
- Usage Quotas

Help

Naming, topics, avatar

Update your project name, topics, description, and avatar.

Project name Demo_ARC **Project ID** 1584

Topics

Search for topic

Topics are publicly visible even on private projects. Do not include sensitive information in topic names. [Learn more](#).

Project description (optional)

Project avatar

Choose file... No file chosen.

The ideal image size is 192 x 192 pixels. The maximum file size allowed is 200 KiB.

Visibility, project features, permissions

Choose visibility level, enable/disable project features and their permissions, disable email notifications, and show default emoji reactions.

Badges

Add badges to display information about this project. [What are badges?](#)

Service Desk

Enable and disable Service Desk. Some additional configuration might be required. [Learn more](#).

Advanced (4)

Housekeeping, export, archive, change path, transfer, and delete.

Follow your progress in the DataHUB

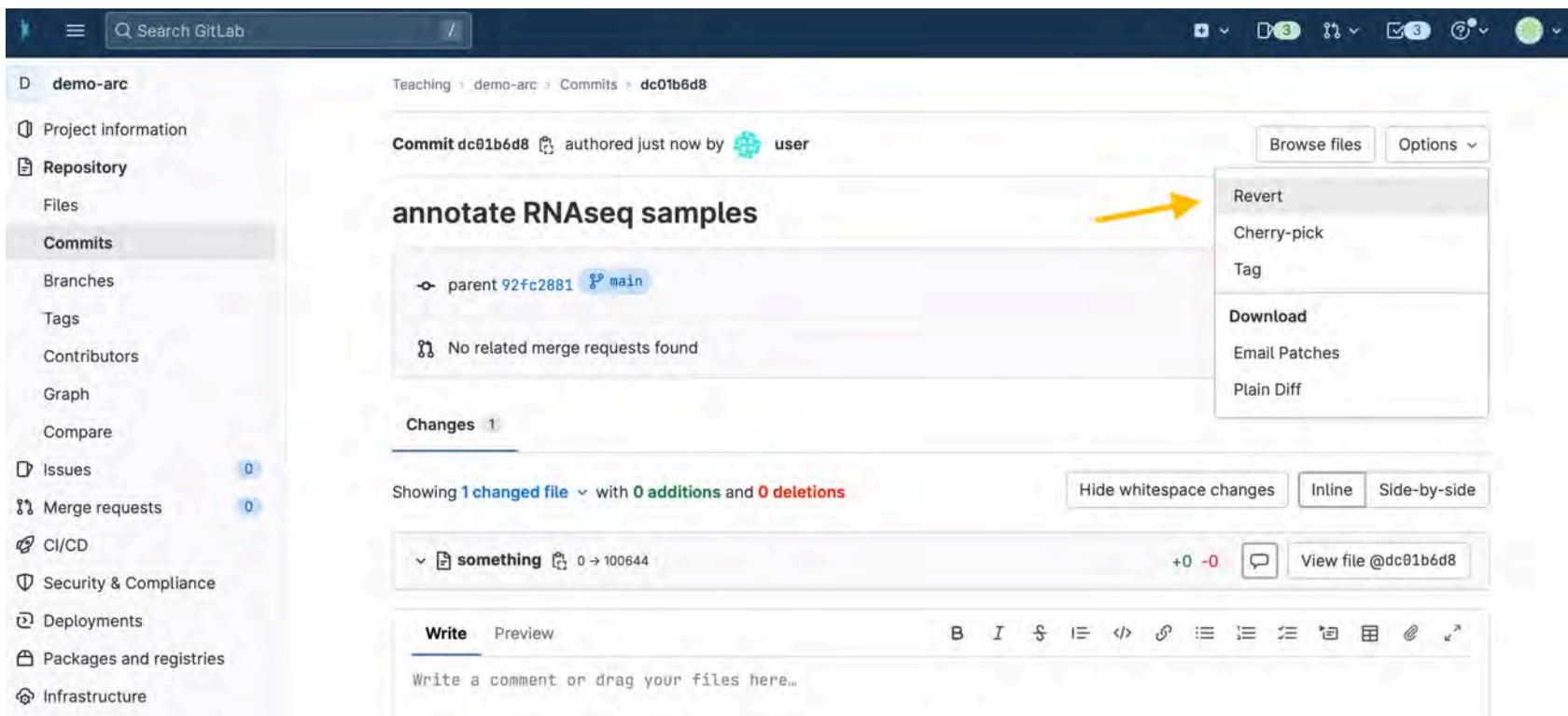
1. Open your ARC in the DataHUB
2. In the sidebar on the left, navigate to **Repository** → **Commits**
3. There you find a chronological list of syncing messages together with who synced and when

The screenshot shows a GitLab interface for a repository named 'demo-arc'. The sidebar on the left lists various repository sections: Project information, Repository, Files, Commits (which is highlighted with a yellow arrow), Branches, Tags, Contributors, Graph, Compare, Issues (0), Merge requests (0), CI/CD, Security & Compliance, and Deployments. The main content area displays a list of commits made on '29 Jun, 2023' at 5:00 AM. The commits are:

- annotate RNAseq samples (user authored just now)
- added RNAseq dataset (user authored 1 minute ago)
- added folder structure (user authored 2 hours ago)
- Update (user authored 3 hours ago)
- initialized ARC structure (user authored 4 hours ago)

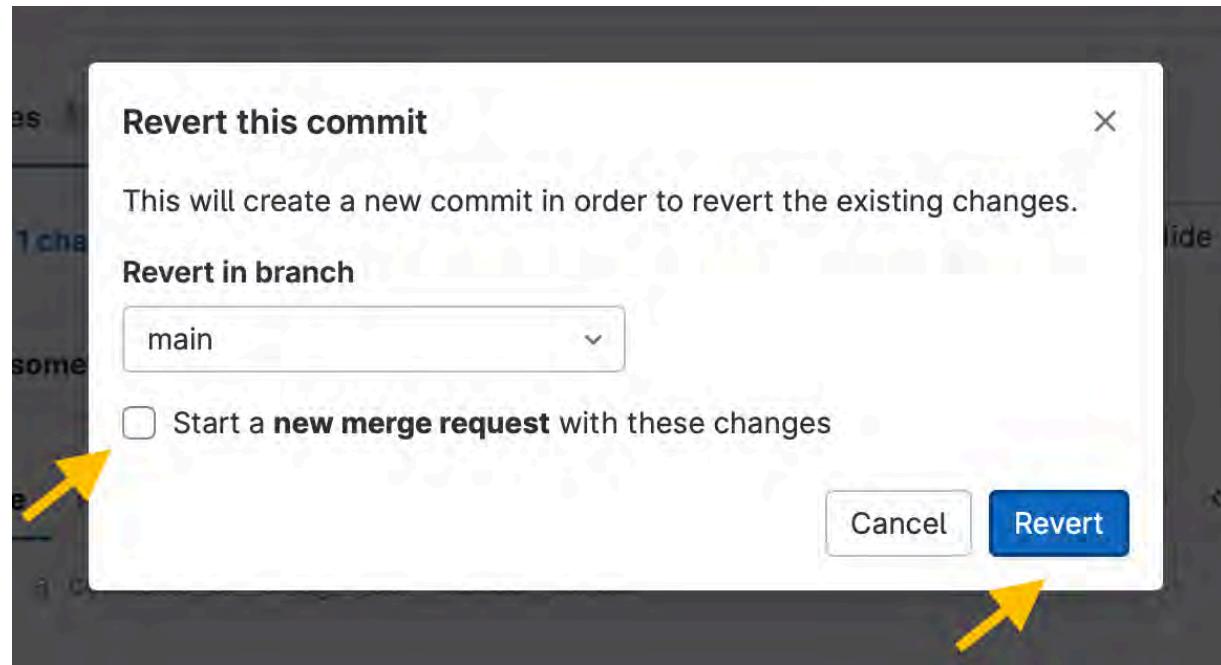
Undo latest changes

4. Click on the latest (i.e. uppermost) commit.
5. In the top-right corner select **Revert** from the drop-down menu **Options**.



Undo latest changes

6. Un-check the box "Start a new merge request with these changes".
7. Click "Revert"



Update your local ARC

If your ARC has changed in the DataHUB (by yourself or collaborators), you need to update your "local" version of the ARC.

1. Navigate to Versions
2. Click "Download"

Contributors

Slides presented here include contributions 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: Sabrina Zander
github: <https://github.com/SabrinaZander>
orcid: <https://orcid.org/0009-0000-4569-6126>

Q&A and Wrap-up

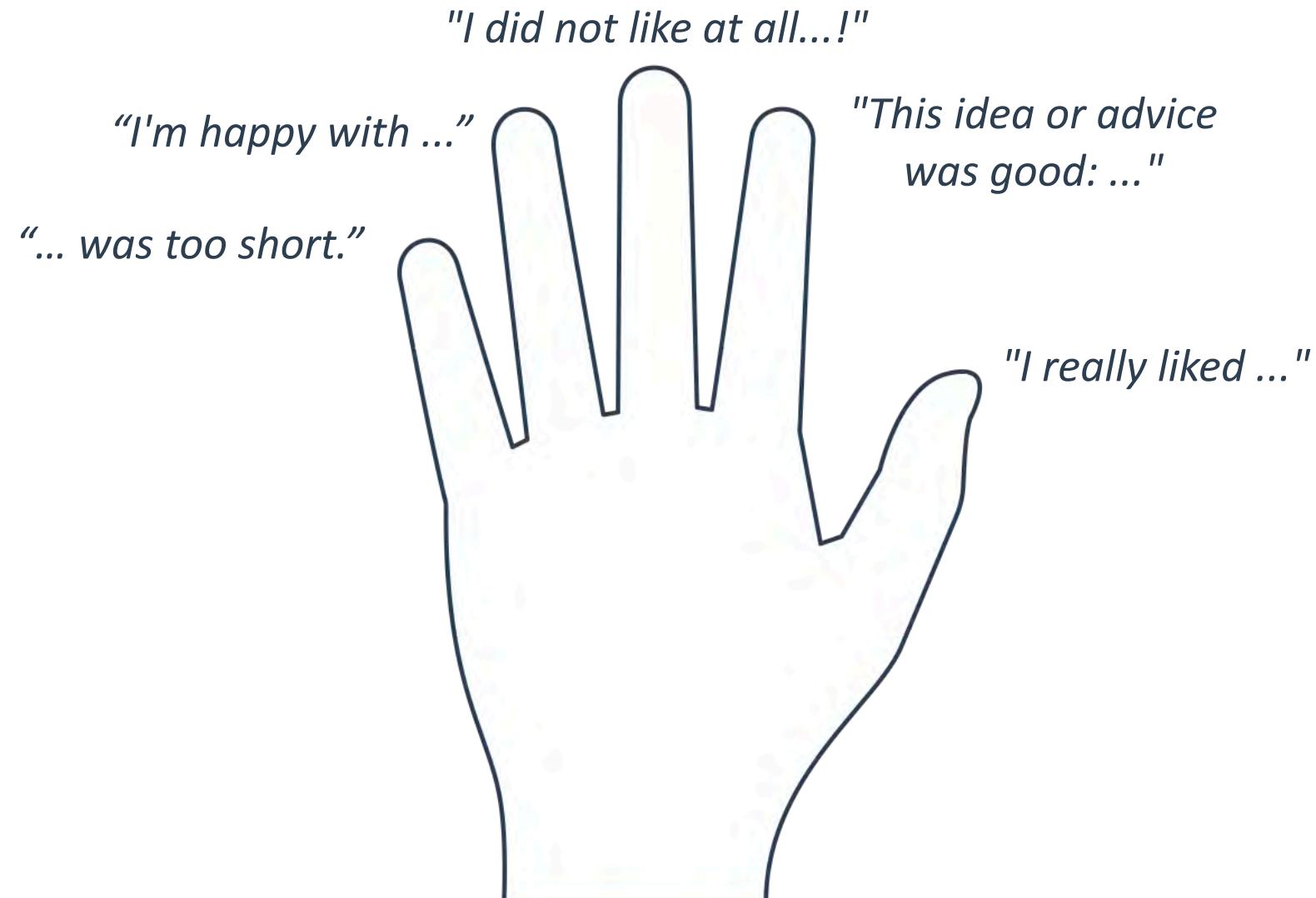
Tentative agenda

Time	Topics
13:00 - 14:00	Welcome and intro
14:00 - 14:15	<i>Short break</i>
14:15 - 16:00	ARC and ARCitect Hands-on
16:00 - 16:15	<i>Short break</i>
16:15 - 17:00	Q&A

Preparation for next week

- Please try to prepare your own ARC
- Please install SWATe

Five-Finger-Feedback



Resources



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...

Start your ARC Workshop

for CSCS

October 11th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

Welcome back

1. Welcome and feedback
2. Metadata and ISA
3. Swate Hands-On
4. ARC Ecosystem Summary
5. Q & A

Feedback

💡 Share your experience

Block 6 – Metadata and ISA

October 11th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

**What is
metadata?**

Viola's PhD Project

Exercise: Take 5 minutes to note down the metadata

Viola investigates the effect of the plant circadian clock on sugar metabolism in *W. mirabilis*. For her PhD project, which is part of an EU-funded consortium in Prof. Beetroot's lab, she acquires seeds from a South-African botanical society. Viola grows the plants under different light regimes, harvests leaves from a two-day time series experiment, extracts polar metabolites as well as RNA and submits the samples to nearby core facilities for metabolomics and transcriptomics measurements, respectively. After a few weeks of iterative consultation with the facilities' heads as well as technicians and computational biologists involved, Viola receives back a wealth of raw and processed data. From the data she produces figures and wraps everything up to publish the results in the Journal of Wonderful Plant Sciences.

Metadata everywhere

Viola investigates the effect of the plant circadian clock on sugar metabolism in *W. mirabilis*. For her PhD project, which is part of an EU-funded consortium in Prof. Beetroot's lab, she acquires seeds from a South-African botanical society. Viola grows the plants under different light regimes, harvests leaves from a two-day time series experiment, extracts polar metabolites as well as RNA and submits the samples to nearby core facilities for metabolomics and transcriptomics measurements, respectively. After a few weeks of iterative consultation with the facilities' heads as well as technicians and computational biologists involved, Viola receives back a wealth of raw and processed data. From the data she produces figures and wraps everything up to publish the results in the Journal of Wonderful Plant Sciences.

Project metadata

project design

- researcher
- institute and project
- biological context
- research question
- purpose of data collection
- ...

experimental processes

- origin and nature of the biological material
- lab protocols
- instrument model
- ...

data-analytical processes

- algorithms
- tools
- software versions and dependencies employed
- ...

Other types of metadata

bibliographic

- Title
- Publication date and title
- Description
- Author
- Contacts
- Keywords
- ...

legal or administrative

- data origin, ownership, provenance,
- licensing
- ethical aspects
- ...

technical

- expected data volume
- storage location
- file formats
- ...

Metadata from a FAIR perspective

Findable

- metadata names the content of the data
- basis for search engines
- makes it categorizable for people and machines

Interoperable

- metadata identifies software and file formats
- required conversions between file formats

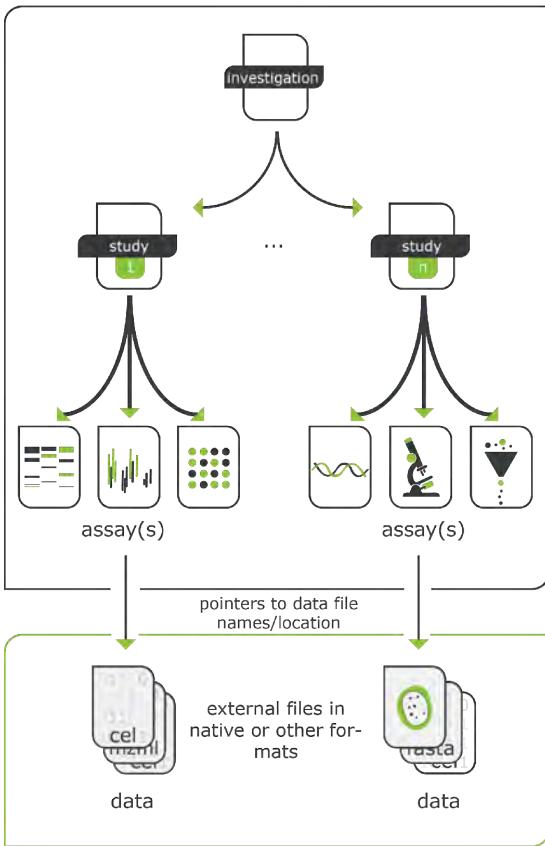
Reusable

- obtain and reuse research data according to clear rules described in licenses

Accessible

- information about origin
- location of storage
- access rights

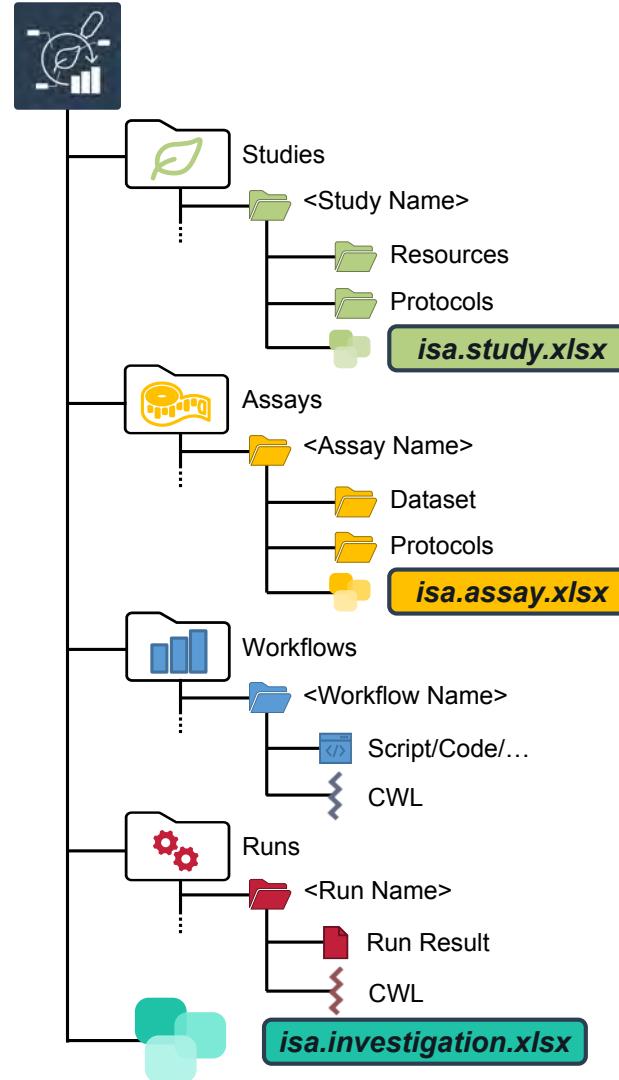
ARC builds on ISA



Investigation
Overall goals
Scientific context

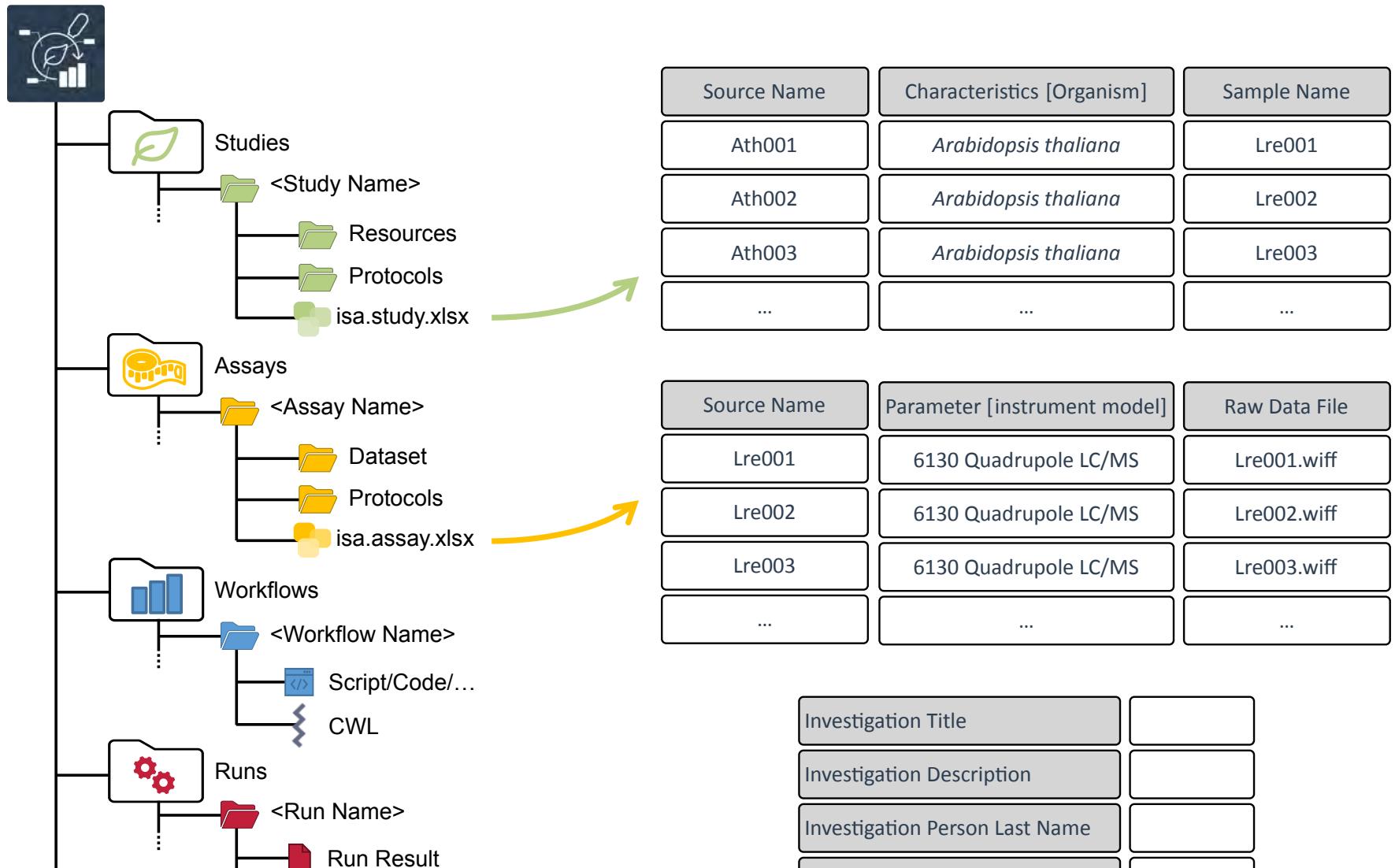
Study
Experimental design

Assay
Leading to (raw) data



ARC builds on ISA

Metadata Annotations



isa.<>.xlsx files within ARCs

isa.investigation.xlsx

DATACLOUD SOURCE REFERENCE	OB	STO	NENT	I_O	CIBIO	PATD	EFO
Term Source File	http://bioportal.bioontology.org/ontologies/Experimental Factor Ontology						
Term Source Version	47803_v126	v1.26	v1.26	v1.26	v1.26	v1.26	v1.26
Term Source Description	Ontology for Biomed BRINDA Issue / NWTF UniProt Tax Link Ontology Chemical Entity Phenotypic Array Exchange Diamerical factor Ontology						
INVESTIGATION							
Investigation Identifier	BII_1						
Investigation Title	Growth control of the yeast cell: a systems biology study in yeast						
Investigation Description	Background Cell growth underlies many key cellular and developmental processes, yet a limited number of studies have been carried out on cell growth control. This study aims to address this gap by using a systems biology approach to study cell growth control in yeast.						
Investigation Submission Date	30.04.07						
Investigation Public Release Date	19.03.07						
Investigation Publication Status	Published						
Investigation Publication Status Term Accession Number							
Investigation Publication Status Term Source REF							
INVESTIGATION PUBLICATIONS							
Investigation PubMed ID	17439666						
Investigation Publication DO	med101186/pmid54						
Investigation Publication Author List	Castroli J, Zeeb JA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hayes L, Wettieworth L, Rash B, Brown JV, Dunn WB, Broadhurst						
Investigation Publication Title	Growth control of the yeast cell: a systems biology study in yeast						
Investigation Publication Status	published						
Investigation Publication Status Term Accession Number							
Investigation Publication Status Term Source REF							
STUDY							
Study Identifier	BII_1						
Study Title	Study of the impact of changes in flux on the transcriptome, proteome, endometabolome and exometabolome of the yeast Saccharomyces cerevisiae						
Study Description	Study of the impact of growth rate on the total component of mRNA molecules, proteins, and metabolites in S. cerevisiae. Independent						
Comment[Study Grant Number]							
Comment[Study Funding Agency]							
Study Submission Date	30.04.07						
Study Release Date	10.03.09						
Study Identifier	BII_5_Lntr						
STUDY DESIGN DESCRIPTIONS							
Study Design Type	intervention design						
Study Design Type Term Accession Number	http://purl.bioontology.org/obo/ATO_0000015						
Study Design Type Term Source REF	OB						
STUDY PUBLICATIONS							
Study PubMed ID	17439666						
Study Publication DOI	med101186/pmid54						
Study Publication Author List	Castroli J, Zeeb JA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hayes L, Wettieworth L, Rash B, Brown JV, Dunn WB, Broadhurst						
Study Publication Title	Growth control of the yeast cell: a systems biology study in yeast						
Study Publication Status	published						
Study Publication Status Term Accession Number							
Study Publication Status Term Source REF							
STUDY FACTORS							
Study Factor Name	binding number	-rate					
Study Factor Type	physical response	-rate					
Study Factor Type Term Accession Number	http://caisafedlibrary.org/bio/ATO_0000161						
Study Factor Type Term Source REF	PATO						
STUDY ASSAYS							
Study Assay Measurement Type	protein expression profile	transcription profiling					
Study Assay Measurement Type Term Accession Number	http://purl.bioontology.org/obo/ATB_0400148						
Study Assay Measurement Type Term Source REF	OB	CBI	CBI				
Study Assay Technique Type	mass spectrometry	mass spectrometry	mass spectrometry	mass spectrometry			
Study Assay Technique Type Term Accession Number	http://purl.bioontology.org/obo/ATB_0400148						
Study Assay Technique Type Term Source REF	OB	CBI	CBI				
Study Assay Technique Platform	ITraq	LC-MS/MS	Affymetrix				
Study Assay File Name	W_00000001	W_00000001	W_00000001	W_00000001			
STUDY PROTOCOLS							
Study Protocol Name	Growth protocol	mRNA extraction protein extraction	protein labeling	labeling	hybridization	hybridization extraction	
Study Protocol Type	growth	mRNA extraction protein extraction	labeling	labeling	hybridization	hybridization extraction	
Study Protocol Type Term Accession Number							
Study Protocol Type Term Source REF							
Study Protocol Description	1. Biomass samples (1. Biomass samples (45 ml) were taken. This was done using Dexta. For each target, a hybridization cocktail was made using the						
Study Protocol JID							

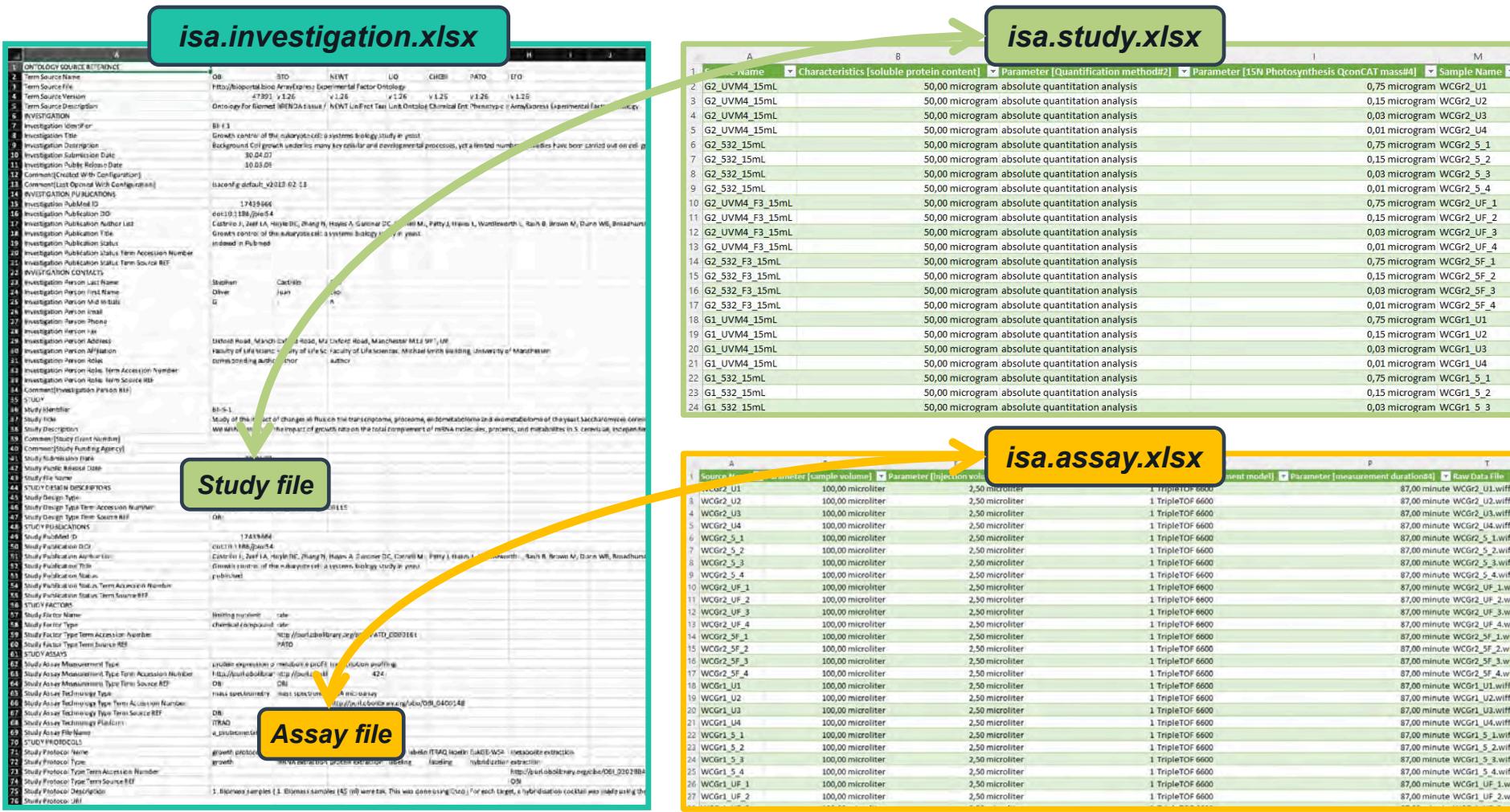
isa.study.xlsx

A	B	C	D	E	F
Source Name	Characteristics [soluble protein content]	Parameter [Quantification method#2]	Parameter [15N Photosynthesis QconCAT mass#4]	Sample Name	
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WGCr2_U1
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WGCr2_U2
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WGCr2_U3
G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WGCr2_U4
G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WGCr2_5_1
G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WGCr2_5_2
G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WGCr2_5_3
G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WGCr2_5_4
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WGCr2_UF_1
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WGCr2_UF_2
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WGCr2_UF_3
G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WGCr2_UF_4
G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WGCr2_UF_5_1
G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WGCr2_UF_5_2
G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WGCr2_UF_5_3
G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WGCr2_UF_5_4
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WGCr1_U1
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WGCr1_U2
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WGCr1_U3
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WGCr1_U4
G1_S32_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WGCr1_5_1
G1_S32_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WGCr1_5_2
G1_S32_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WGCr1_5_3

isa.assay.xlsx

A	B	C	D	E	F	G	H	I	J	K	L	M
Source Name	Parameter [sample volume]	Parameter [injection vol]	Parameter [measurement model]	Parameter [measurement duration#4]	Raw Data File							
WGCr2_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_U1.wiff							
WGCr2_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_U2.wiff							
WGCr2_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_U3.wiff							
WGCr2_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_U4.wiff							
WGCr2_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_5_1.wiff							
WGCr2_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_5_2.wiff							
WGCr2_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_5_3.wiff							
WGCr2_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_5_4.wiff							
WGCr2_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_UF_1.wiff							
WGCr2_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_UF_2.wiff							
WGCr2_UF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_UF_3.wiff							
WGCr2_UF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_UF_4.wiff							
WGCr2_SF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_SF_1.wiff							
WGCr2_SF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_SF_2.wiff							
WGCr2_SF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_SF_3.wiff							
WGCr2_SF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr2_SF_4.wiff							
WGCr1_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_U1.wiff							
WGCr1_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_U2.wiff							
WGCr1_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_U3.wiff							
WGCr1_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_U4.wiff							
WGCr1_S3_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_S3_1.wiff							
WGCr1_S3_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_S3_2.wiff							
WGCr1_S3_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_S3_3.wiff							
WGCr1_S3_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_S3_4.wiff							
WGCr1_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_UF_1.wiff							
WGCr1_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute	WGCr1_UF_2.wiff							

Study and assay files are registered in the investigation file



The output of a study or assay file can function as input for a new isa.assay.xlsx

Output building blocks:

- Sample Name
- Raw Data File
- Derived Data File

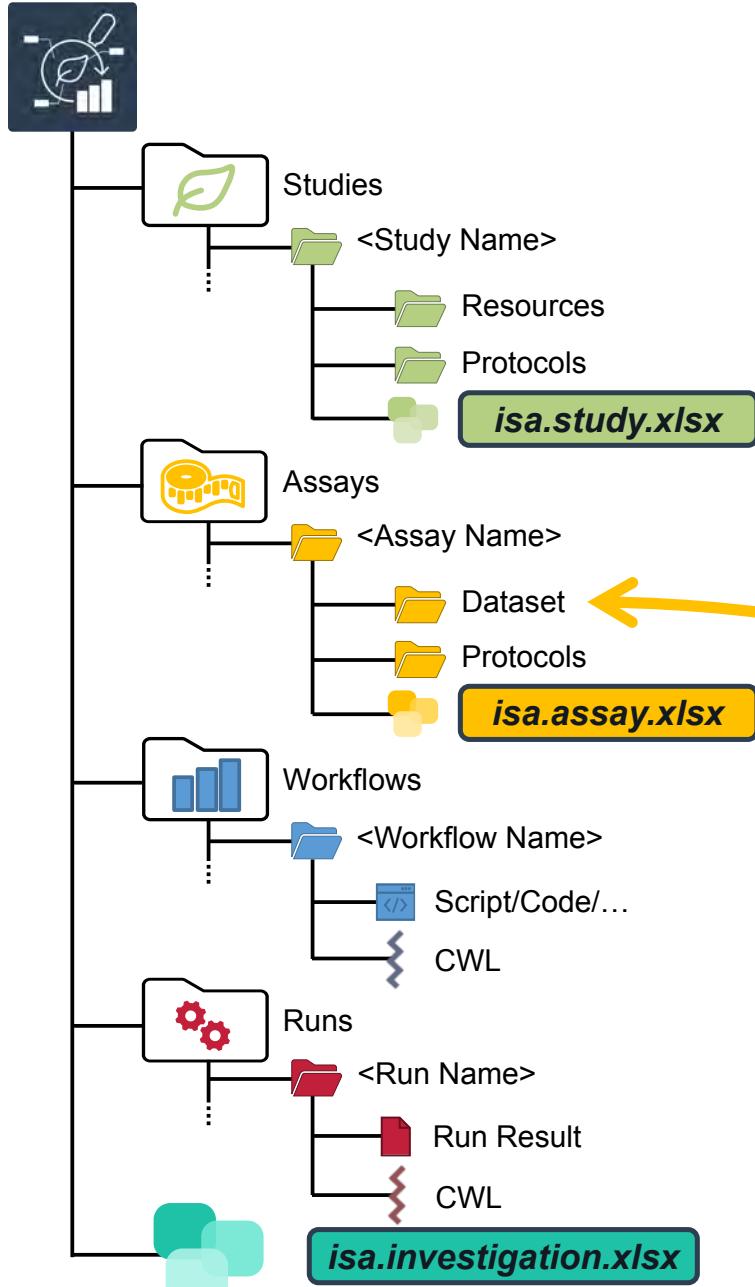
A	B	C	D	E	F	G	H	I	J	K	L	M
Source Name	Characteristics [soluble protein content]	Parameter [Quantification method#2]	Parameter [15N Photosynthesis QconCAT mass#4]	Sample Name								
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,75 microgram WCGr2_U1
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,15 microgram WCGr2_U2
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,03 microgram WCGr2_U3
G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,01 microgram WCGr2_U4
G2_532_15mL	50,00 microgram	absolute quantitation analysis										0,75 microgram WCGr2_5_1
G2_532_15mL	50,00 microgram	absolute quantitation analysis										0,15 microgram WCGr2_5_2
G2_532_15mL	50,00 microgram	absolute quantitation analysis										0,03 microgram WCGr2_5_3
G2_532_15mL	50,00 microgram	absolute quantitation analysis										0,01 microgram WCGr2_5_4
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis										0,75 microgram WCGr2_UF_1
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis										0,15 microgram WCGr2_UF_2
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis										0,03 microgram WCGr2_UF_3
G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis										0,01 microgram WCGr2_UF_4
G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis										0,75 microgram WCGr2_5F_1
G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis										0,15 microgram WCGr2_5F_2
G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis										0,03 microgram WCGr2_5F_3
G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis										0,01 microgram WCGr2_5F_4
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,75 microgram WCGr1_U1
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,15 microgram WCGr1_U2
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,03 microgram WCGr1_U3
G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis										0,01 microgram WCGr1_U4
G1_532_15mL	50,00 microgram	absolute quantitation analysis										0,75 microgram WCGr1_5_1
G1_532_15mL	50,00 microgram	absolute quantitation analysis										0,15 microgram WCGr1_5_2
G1_532_15mL	50,00 microgram	absolute quantitation analysis										0,03 microgram WCGr1_5_3

isa.study.xlsx

Samples

A	B	C	D	E	F	G	H	I	J	K	L	M
Source Name	Parameter [sample volume]	Parameter [injection volu										
WCGr2_U1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_U1.wiff
WCGr2_U2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_U2.wiff
WCGr2_U3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_U3.wiff
WCGr2_U4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_U4.wiff
WCGr2_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_1.wiff
WCGr2_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_2.wiff
WCGr2_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_3.wiff
WCGr2_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_4.wiff
WCGr2_UF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_UF_1.wiff
WCGr2_UF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_UF_2.wiff
WCGr2_UF_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_UF_3.wiff
WCGr2_UF_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_UF_4.wiff
WCGr2_SF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_SF_1.wiff
WCGr2_SF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_SF_2.wiff
WCGr2_SF_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_SF_3.wiff
WCGr2_SF_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_SF_4.wiff
WCGr1_U1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_U1.wiff
WCGr1_U2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_U2.wiff
WCGr1_U3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_U3.wiff
WCGr1_U4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_U4.wiff
WCGr1_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_1.wiff
WCGr1_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_2.wiff
WCGr1_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_3.wiff
WCGr1_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_4.wiff
WCGr1_UF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_UF_1.wiff
WCGr1_UF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_UF_2.wiff

isa.assay.xlsx



isa.study.xlsx

A	B	C	D	E	F	G	H	I	J	K	L	M
1	Source Name	Characteristics [soluble protein content]	Parameter [Quantification method#2]	Parameter [15N Photosynthesis QconCAT mass#4]	Sample Name							
2	G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr2_U1						
3	G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr2_U2						
4	G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr2_U3						
5	G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WCGr2_U4						
6	G2_532_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr2_5_1						
7	G2_532_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr2_5_2						
8	G2_532_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr2_5_3						
9	G2_532_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WCGr2_5_4						
10	G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr2_UF_1						
11	G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr2_UF_2						
12	G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr2_UF_3						
13	G2_UVM4_F3_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WCGr2_UF_4						
14	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr2_5F_1						
15	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr2_5F_2						
16	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr2_5F_3						
17	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WCGr2_5F_4						
18	G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr1_U1						
19	G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr1_U2						
20	G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr1_U3						
21	G1_UVM4_15mL	50,00 microgram	absolute quantitation analysis		0,01 microgram	WCGr1_U4						
22	G1_532_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr1_5_1						
23	G1_532_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr1_5_2						
24	G1_532_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr1_5_3						

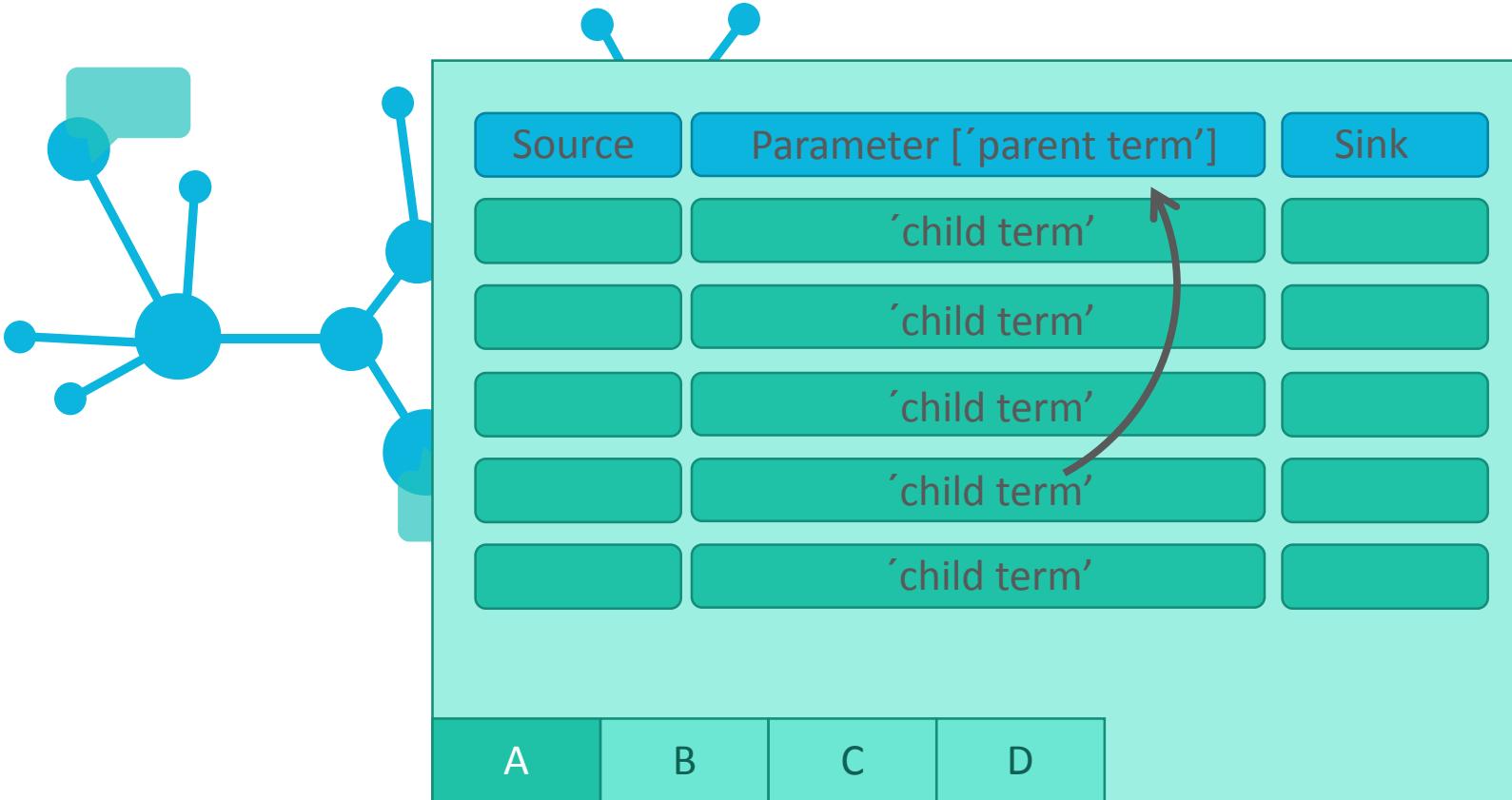
isa.assay.xlsx

A	B	C	D	E	F	G	H	I	J	K	L	M
1	Source Name	Parameter [sample volume]	Parameter [injection vol]	Parameter [measurement duration#1]	Raw Data File							
2	WCGr2_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
3	WCGr2_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
4	WCGr2_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
5	WCGr2_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
6	WCGr2_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
7	WCGr2_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
8	WCGr2_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
9	WCGr2_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
10	WCGr2_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
11	WCGr2_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
12	WCGr2_UF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
13	WCGr2_UF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
14	WCGr2_5F_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
15	WCGr2_5F_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
16	WCGr2_5F_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
17	WCGr2_5F_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
18	WCGr1_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
19	WCGr1_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
20	WCGr1_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
21	WCGr1_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
22	WCGr1_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
23	WCGr1_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
24	WCGr1_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
25	WCGr1_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
26	WCGr1_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							
27	WCGr1_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	WIFF File							

Raw data

Swate

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 (columns)

The screenshot shows a Microsoft Excel spreadsheet titled "Sheet1" with data from row 1 to 50. The columns include "Source Name", "Protocol type", "Characteristic [sample label]", "Factor [temperature]", "Parameter [instrument model]", "Component [software]", and "Sample Name". A callout bubble labeled "New Parameter" points to the "Parameter [instrument model]" column. To the right, a "Building Blocks" dialog box is open, listing various instrument models and instruments, with a note about adding annotation building blocks.

New Parameter

Building Blocks

Add annotation building blocks (columns) to the annotation table.

Parameter	instrument mod
instrument model	MS(100003)
Instrument Model	NCIT(C177410)
Ad	
instrument	MS(100004)
instrument	UFD(000054)
Agilent instrument model	MS(100049)

Can't find the term you are looking for? Try

Parameter columns describe steps in your experimental workflow, e.g. the centrifugation time or the temperature used for your assay. Multiple Parameter columns form a protocol. There is no limitation for the number of Parameter columns per table. You can find more information on our website.

Annotation Building Block types

- Source Name (Input)
- Protocol Columns
 - Protocol Type, Protocol Ref
- Characteristic
- Parameter
- Factor
- Component
- Output Columns
 - Sample Name, Raw Data File, Derived Data File

The screenshot shows a Microsoft Excel-like spreadsheet application titled "Swate" with the ribbon menu. The main window displays a table with columns labeled: Source Name, Protocol Ref, Characteristic (sample label), Factor (temperature), Parameter (instrument model), Component (software), and Sample Name. A "New Parameter" column is also present. Several cells in the table are highlighted with callout boxes and arrows pointing to specific terms:

- "Characteristic" points to the "Characteristics" column.
- "Protocol Type/Protocol REF" points to the "Protocol Ref" column.
- "Factor" points to the "Factor (temperature)" column.
- "Sample Name/Raw Data File Derived Data File" points to the "Sample Name" column.

A vertical sidebar on the right is titled "Building Blocks" and lists categories such as "Parameter", "Instrument model", "Instrument", and "Agilent instrument model". A tooltip at the bottom right of the sidebar says: "Parameter columns describe steps in your experimental workflow, e.g. the centrifugation time or the temperature used for your assay. Multiple parameter columns form a protocol. There is no limitation for the number of parameter columns in a table. You can find more information on our website".

Source Name	Protocol Ref	Characteristic (sample label)	Factor (temperature)	Parameter (instrument model)	Component (software)	Sample Name
G1_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G2_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G3_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G4_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4
G5_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_1
G6_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_2
G7_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_3
G8_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_4
G9_UVMA_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G10_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G11_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G12_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4
G13_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_1
G14_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_2
G15_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_3
G16_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_4
G17_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G18_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G19_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G20_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4
G21_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_1
G22_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_2
G23_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_3
G24_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_4
G25_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G26_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G27_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G28_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4
G29_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_1
G30_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_2
G31_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_3
G32_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_4
G33_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G34_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G35_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G36_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4
G37_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_1
G38_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_2
G39_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_3
G40_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_4
G41_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G42_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G43_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G44_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4
G45_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_1
G46_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_2
G47_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_3
G48_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_S2_4
G49_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U1
G50_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U2
G51_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U3
G52_UVMA_F1_15ml		30.00 degree Celsius 6130 Quadrupole LC/MS			Analyst	WCGRI_U4

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

Ontology term search

The screenshot shows a Microsoft Excel spreadsheet titled "Sheet1" with data in columns A through AB. The data consists of rows 1 through 52, each containing information such as Source Name, Protocol Type, Characteristic [sample label], Factor [temperature], Parameter [instrument model], Component [software], and Sample Name. The "Parameter [instrument model]" column contains values like "G2_UV4M_15mL", "G2_UV4M_F3_15mL", "G1_UV4M_15mL", etc., which are highlighted in green. The "Component [software]" column contains values like "Analyst", "WCGr2_U1", "WCGr2_U2", etc. The "Sample Name" column contains values like "WCGr2_U1", "WCGr2_U2", "WCGr2_U3", etc.

To the right of the spreadsheet, a "Swate" window is open. The title bar says "SWATE". The main area has tabs for "Ontology term search" and "Advanced search". In the "Ontology term search" tab, there is a search bar with the text "instrument n 6130" and a dropdown menu showing "6130 Quadrupole LC/MS". Below the search bar, it says "Can't find the Term you are looking for? Try Advanced Search!". At the bottom, it says "Still can't find what you need? Get in contact with us!"

Fill your table with ontology terms

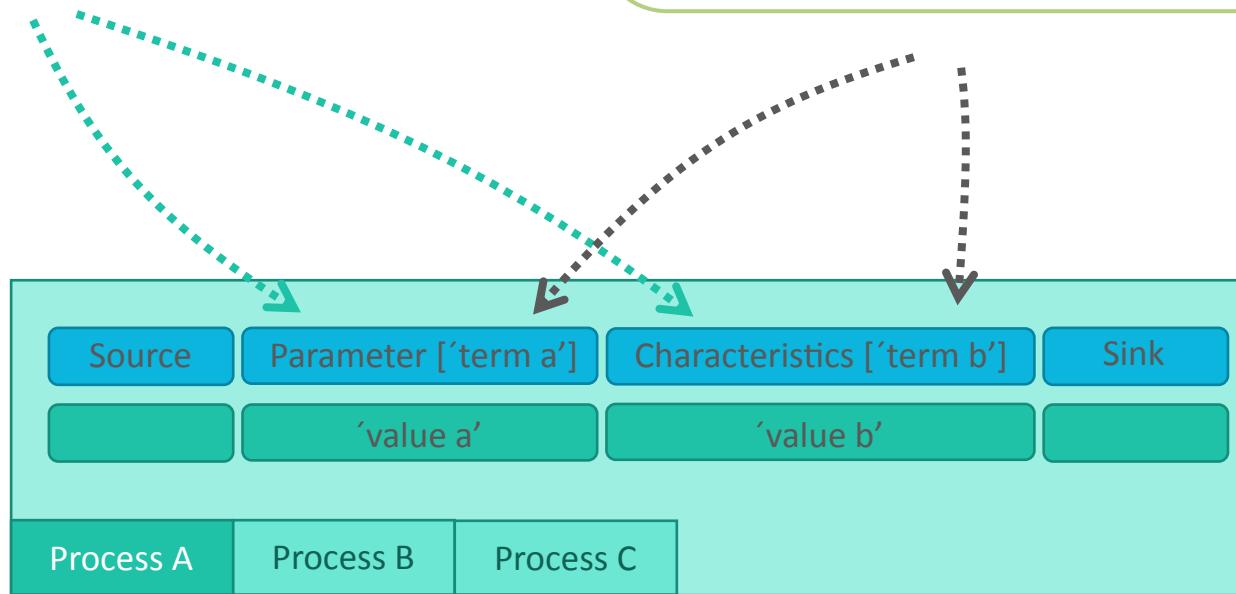
The screenshot shows a Microsoft Excel spreadsheet titled "Sheet1" with data in columns A through AB. The data consists of rows 1 through 50, each containing information such as Source Name, Protocol Type, Characteristic [sample label], Factor [temperature], Parameter [instrument model], Component [software], and Sample Name. Many cells in the table are highlighted in green, indicating they are being populated by the Swate add-in.

A floating window titled "Swate" is overlaid on the spreadsheet, showing an "Ontology term search" interface. The search bar contains the text "6130 Quadrupole LC/MS". Below the search bar are two buttons: "Use related term directed search." and "Fill selected cells with this term".

The Excel ribbon at the top includes tabs for File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Help, and Table Design. The Data tab is currently selected. The ribbon also features various icons for data management, such as Get & Transform Data, Queries & Connections, Sort & Filter, Data Tools, Forecast, and Outline. The "Table Design" tab is visible in the ribbon.

At the bottom of the screen, there are several icons: Data PLANT, CEPLAS, CC BY, Accessibility: Investigate, and a small logo with a person icon.

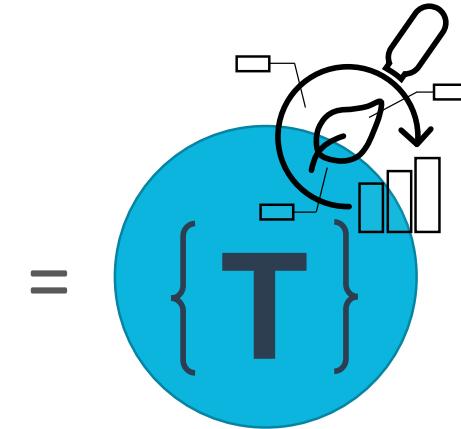
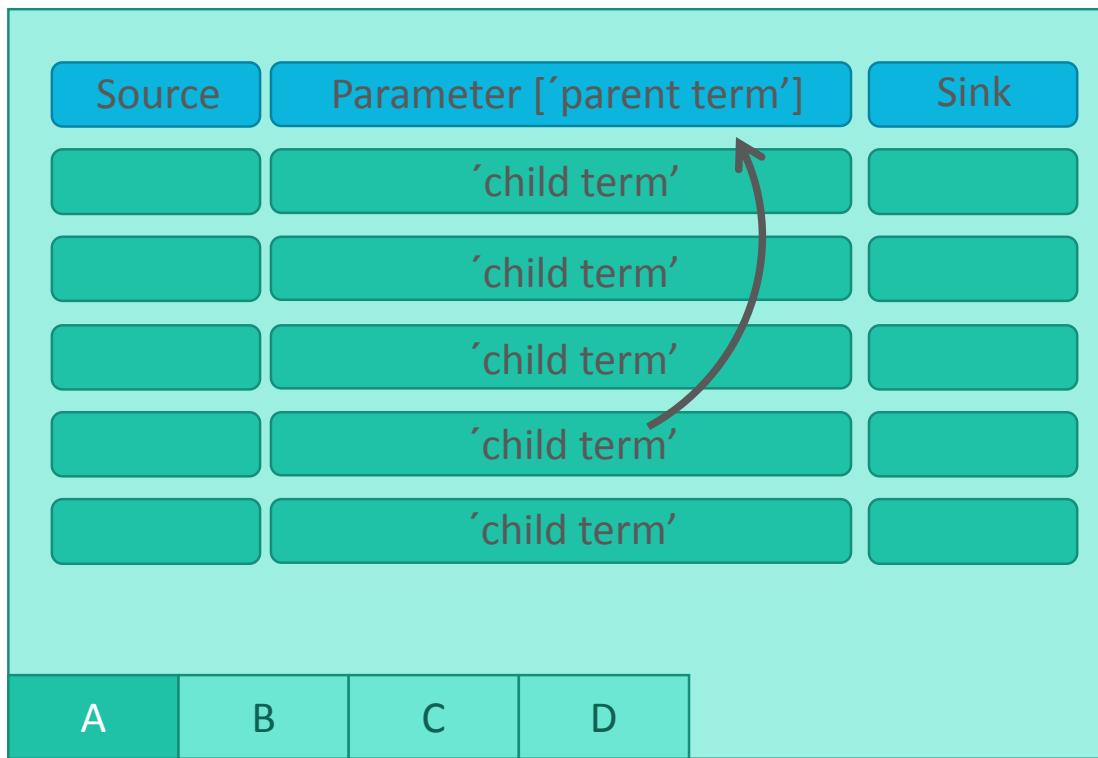
Hierarchical combination of ontologies



[isa.study.xlsx](#) or [isa.assay.xlsx](#)

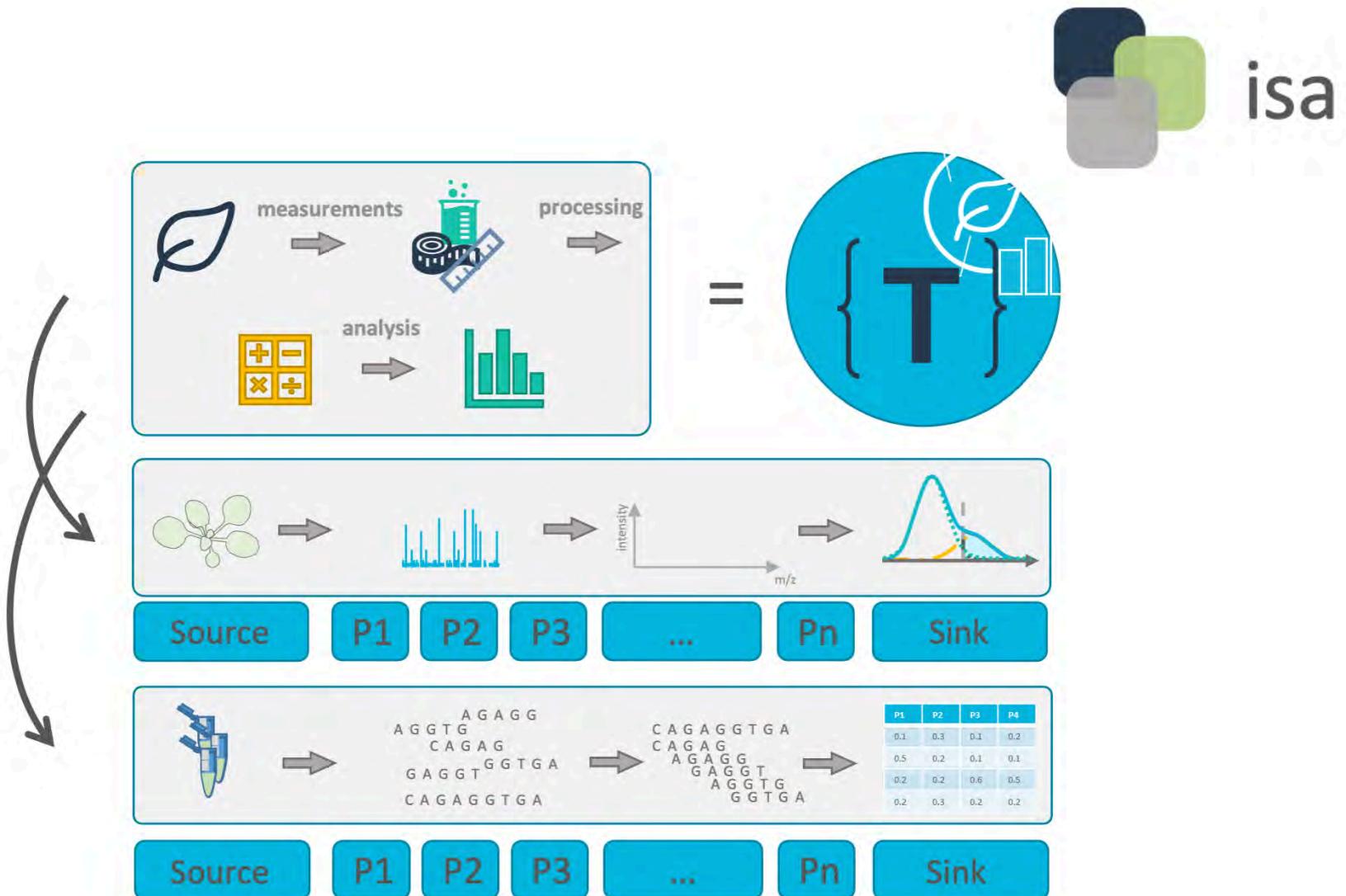
Swate templates

Checklists and Templates



Metadata standards or repository requirements can be represented as templates

Realization of lab-specific metadata templates



Directly import templates via Swate

- DataPLANT curated
- Community templates

The screenshot shows the Swate software interface. At the top, there's a toolbar with various icons. Below the toolbar, the title "Swate" is displayed, followed by the version "SWATE 1.1.13". The main area is titled "Templates → Template Search" and contains a search bar with placeholder text "... protocol name..." and "... protocol tag...". A sub-instruction says "Search for protocol templates. For more information you can look [here](#). If you find any problems with a template or have other suggestions you can contact us [here](#)." There are two search buttons: "Search by protocol name" and "Search for tags". The main content area is a table listing 20 protocol templates. The columns are "Protocol Name", "cur com", "Protocol Version", and "Uses". Each row includes a small green "curated" button and a dropdown arrow. The listed protocols include Plant growth, RNA extraction, Protein extraction, Metabolite Extraction, DNA extraction, Imaging extraction, RNA-Seq Assay, Proteomics MassSpec Assay, Metabolomics MassSpec Assay, Genomics Assay, Imaging assay, RNA-Seq Computational Analysis, Proteomics Computational Analyses, Metabolomics Computational Analysis, Genome assembly, Imaging computation, and MAdLand Fragmentanalyzer. The "MAdLand Fragmentanalyzer" entry has a yellow "community" button instead of a "curated" button.

Protocol Name	cur com	Protocol Version	Uses
Plant growth	curated	1.1.13	0
RNA extraction	curated	1.1.6	0
Protein extraction	curated	1.1.6	0
Metabolite Extraction	curated	1.1.8	0
DNA extraction	curated	1.1.6	0
Imaging extraction	curated	1.0.2	0
RNA-Seq Assay	curated	1.1.7	0
Proteomics MassSpec Assay	curated	1.1.6	0
Metabolomics MassSpec Assay	curated	1.1.8	1
Genomics Assay	curated	1.1.6	0
Imaging assay	curated	1.0.2	0
RNA-Seq Computational Analysis	curated	1.1.7	0
Proteomics Computational Analyses	curated	1.1.6	0
Metabolomics Computational Analysis	curated	1.1.8	0
Genome assembly	curated	1.1.6	0
Imaging computation	curated	1.0.2	0
MAdLand Fragmentanalyzer	community	1.0.0	0

Swate Release Version 0.6.2

Contributors

Slides presented here include contributions by

- name: Dominik Brilhaus
github: <https://github.com/brilator>
orcid: <https://orcid.org/0000-0001-9021-3197>
- name: Martin Kuhl
github: <https://github.com/Martin-Kuhl>
orcid: <https://orcid.org/0000-0002-8493-1077>
- name: Sabrina Zander
orcid: <https://orcid.org/0009-0000-4569-6126>

Block 7 – Swate hands-on

October 11th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

Goals

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

Check Swate installation

 Make sure [Swate is installed](#):

1. Open Excel (online or Desktop)
2. Go to the [Insert](#) tab: Click the arrow next to "My Add-ins". There you should be able to select Swate.
3. Go to the [Data](#) tab: you should see the Swate (Core) add-in.

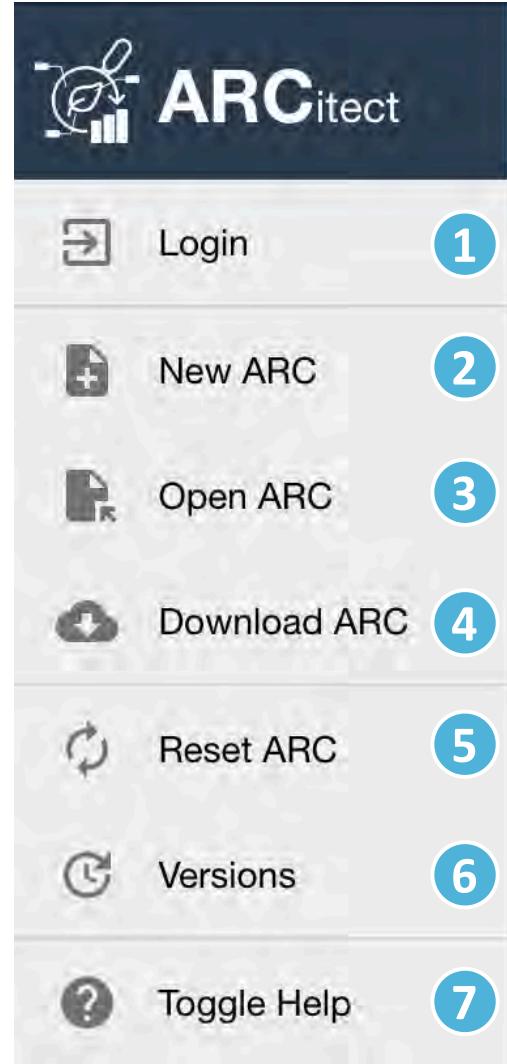
 Alternatively, you can use [Swate standalone](#)

( this is however *work in progress* and likely to change)

Have a simple text editor ready

- Windows Notepad
- MacOS TextEdit

Recommended text editor with code highlighting, git support, terminal, etc: [Visual Studio Code](#)



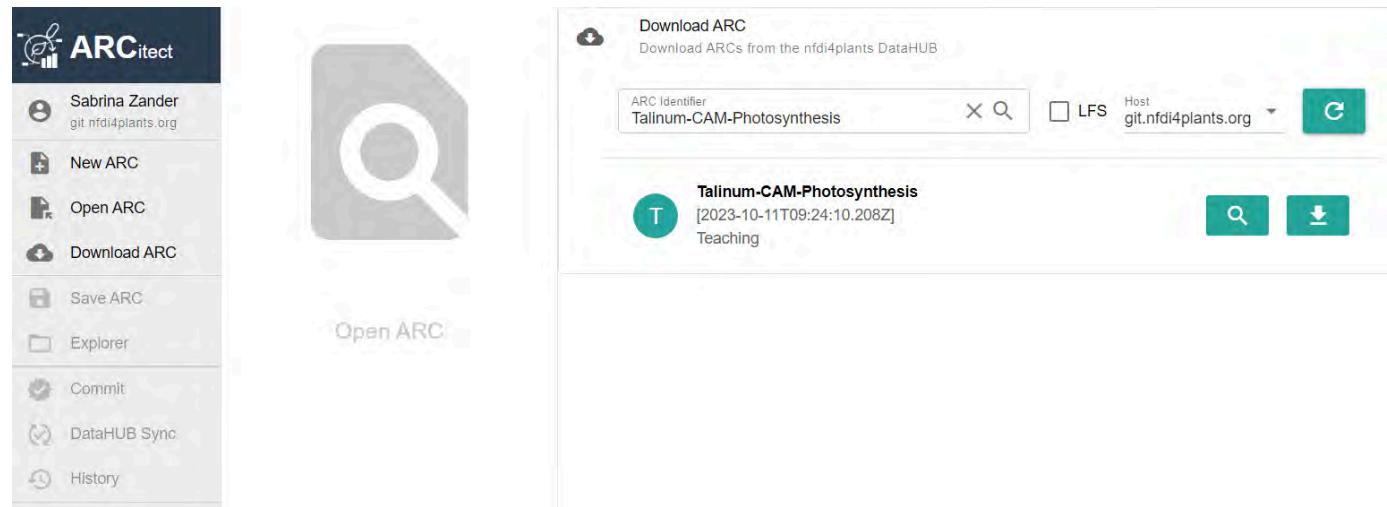
Download the demo data

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.



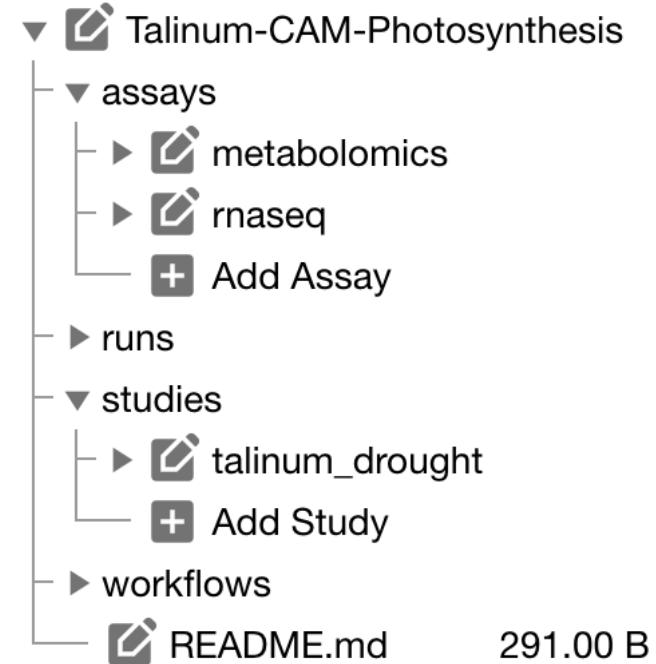
 This is basically the ARC we created last session.

Where we left off last time

 Initiated an ARC

 Structured and ...

 Shared with collaborators



Today we want to

 ... annotate the experimental data

Swate hands-on with demo data

Swate Overview

The screenshot shows a Microsoft Excel spreadsheet titled "tabelle1.xlsx" with a table containing 37 rows of data. The columns are labeled A through K. Column A contains "Source Name", column B contains "Characteristics (sample label)", column C contains "Factor [temperature unit]", and column D contains "Data File Name". The data includes various temperature measurements and corresponding data files. A green box highlights the first few rows under the "source" label. A blue box highlights the "factor" column under the "characteristic" label. A red box highlights the "datafile / sample" column under the "characteristic" label.

The Swate software window is open to the right, showing a search bar and a list of annotations. A green box highlights the "new parameter" section. A blue box highlights the "annotation building block selection" section. A red box highlights the "datafile / sample" section. The Swate interface includes sections for "Add annotation building blocks (columns) to the annotation table.", "Parameter", "instrument", "instrument model", "instrument vendor", "medical instrument", and "Mascot:Instrument". It also features a search bar, an "Advanced Search" link, and a "More about Parameter:" section with a "Use parameters to annotate your experimental workflow..." note.

Source Name	Characteristics (sample label)	Factor [temperature unit]	Data File Name
Heat_15A_OD_R1	15N	32.00 degree Celsius	Heat_15A_OD_R1.wiff
Heat_15A_OD_R2	15N	32.00 degree Celsius	Heat_15A_OD_R2.wiff
Heat_180A_OD_R1	15N	32.00 degree Celsius	Heat_180A_OD_R1.wiff
Heat_180A_OD_R2	15N	32.00 degree Celsius	Heat_180A_OD_R2.wiff
Heat_2880A_OD_R1	15N	32.00 degree Celsius	Heat_2880A_OD_R1.wiff
Heat_2880A_OD_R2	15N	32.00 degree Celsius	Heat_2880A_OD_R2.wiff
Heat_5760A_OD_R1	15N	32.00 degree Celsius	Heat_5760A_OD_R1.wiff
Heat_5760A_OD_R2	15N	32.00 degree Celsius	Heat_5760A_OD_R2.wiff
Heat_5760A_15D_R1	15N	32.00 degree Celsius	Heat_5760A_15D_R1.wiff
Heat_5760A_15D_R2	15N	32.00 degree Celsius	Heat_5760A_15D_R2.wiff
Heat_5760A_180D_R1	15N	32.00 degree Celsius	Heat_5760A_180D_R1.wiff
Heat_5760A_180D_R2	15N	32.00 degree Celsius	Heat_5760A_180D_R2.wiff
Heat_5760A_2880D_R1	15N	32.00 degree Celsius	Heat_5760A_2880D_R1.wiff
Heat_5760A_2880D_R2	15N	32.00 degree Celsius	Heat_5760A_2880D_R2.wiff
Heat_5760A_5760D_R1	15N	32.00 degree Celsius	Heat_5760A_5760D_R1.wiff
Heat_5760A_5760D_R2	15N	32.00 degree Celsius	Heat_5760A_5760D_R2.wiff
Cold_15A_OD_R1	15N	4.00 degree Celsius	Cold_15A_OD_R1.wiff
Cold_15A_OD_R2	15N	4.00 degree Celsius	Cold_15A_OD_R2.wiff
Cold_180A_OD_R1	15N	4.00 degree Celsius	Cold_180A_OD_R1.wiff
Cold_180A_OD_R2	15N	4.00 degree Celsius	Cold_180A_OD_R2.wiff
Cold_2880A_OD_R1	15N	4.00 degree Celsius	Cold_2880A_OD_R1.wiff
Cold_2880A_OD_R2	15N	4.00 degree Celsius	Cold_2880A_OD_R2.wiff
Cold_5760A_OD_R1	15N	4.00 degree Celsius	Cold_5760A_OD_R1.wiff
Cold_5760A_OD_R2	15N	4.00 degree Celsius	Cold_5760A_OD_R2.wiff
Cold_5760A_15D_R1	15N	4.00 degree Celsius	Cold_5760A_15D_R1.wiff
Cold_5760A_15D_R2	15N	4.00 degree Celsius	Cold_5760A_15D_R2.wiff
Cold_5760A_180D_R1	15N	4.00 degree Celsius	Cold_5760A_180D_R1.wiff
Cold_5760A_180D_R2	15N	4.00 degree Celsius	Cold_5760A_180D_R2.wiff
Cold_5760A_2880D_R1	15N	4.00 degree Celsius	Cold_5760A_2880D_R1.wiff
Cold_5760A_2880D_R2	15N	4.00 degree Celsius	Cold_5760A_2880D_R2.wiff
Cold_5760A_5760D_R1	15N	4.00 degree Celsius	Cold_5760A_5760D_R1.wiff
Cold_5760A_5760D_R2	15N	4.00 degree Celsius	Cold_5760A_5760D_R2.wiff
Highlight_15A_OD_R1	15N	22.00 degree Celsius	Highlight_15A_OD_R1.wiff
Highlight_15A_OD_R2	15N	22.00 degree Celsius	Highlight_15A_OD_R2.wiff
Highlight_180A_OD_R1	15N	22.00 degree Celsius	Highlight_180A_OD_R1.wiff
Highlight_180A_OD_R2	15N	22.00 degree Celsius	Highlight_180A_OD_R2.wiff

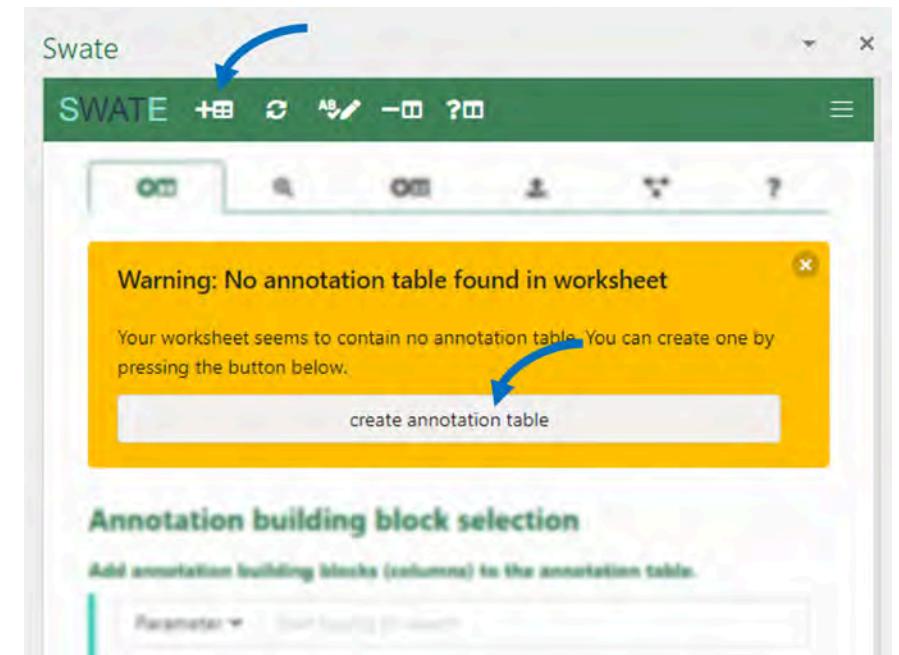
Let's annotate the plant samples first

1. Navigate to the demo ARC.
2. Open the lab notes `studies/talinum_drought/protocols/plant_material.txt` in a text editor.
3. Open the empty `studies/talinum_drought/isa.study.xlsx` workbook in Excel.

Create an annotation table

Create a Swate annotation table via the
create annotation table button in the yellow pop-up box
OR click the Create Annotation Table quick access
button.

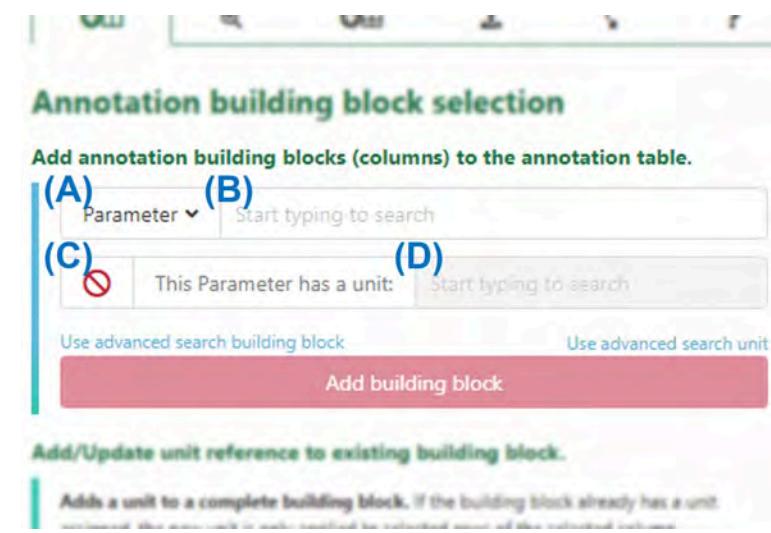
- 💡 Each table is by default created with one input (Source Name) and one output (Sample Name) column
- 💡 Only one annotation table can be added per Excel sheet



Add a building block

1. Navigate to the *Building Blocks* tab via the navbar. Here you can add *Building Blocks* to the table.
2. Instead of *Parameter* select *Characteristic* from the drop-down menu (A)
3. Search for **organism** in the search bar (B). This search looks for suitable *Terms* in our *Ontology* database.
4. Select the Term with the id **OBI:0100026** and,
5. Click **Add building block**.

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



Annotation building block selection

Add annotation building blocks (columns) to the annotation table.

(A) Parameter (B) Start typing to search

(C)  This Parameter has a unit: (D) Start typing to search

Use advanced search building block Use advanced search unit

Add building block

Add/Update unit reference to existing building block.

Add a unit to a complete building block. If the building block already has a unit

Insert values to annotate your data

1. Navigate to the *Terms* tab in the Navbar
2. In the annotation table, select any number of cells below **Characteristic**
[organism]

3. Click into the search field in Swate.

|  You should see **organism** showing in a field in front of the search field
 The search will now yield results related to **organism**

4. In the search field, search for "Talinum fruticosum"

5. Select the first hit and click **Fill selected cells with this term**

Add a building block with a unit

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



This 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. 425.00 microeinstein per square meter per second

Showing ontology reference columns

Hold **Ctrl** and click the *Autoformat Table* quick access button to adjust column widths and un-hide all hidden columns.

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

 This feature is currently not supported on MacOS

Update ontology reference columns

Click the **Update Ontology Terms** quick access buttons.

 This updates all reference columns according to the main column. In this case the reference columns for **Parameter [light intensity exposure]** are updated with the id and source ontology of the **microeinsteин per square meter per second** unit.

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

Hiding ontology reference columns

Click the  quick access button without holding  to hide all reference columns.

Exercise



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

Let's annotate the RNA Seq data

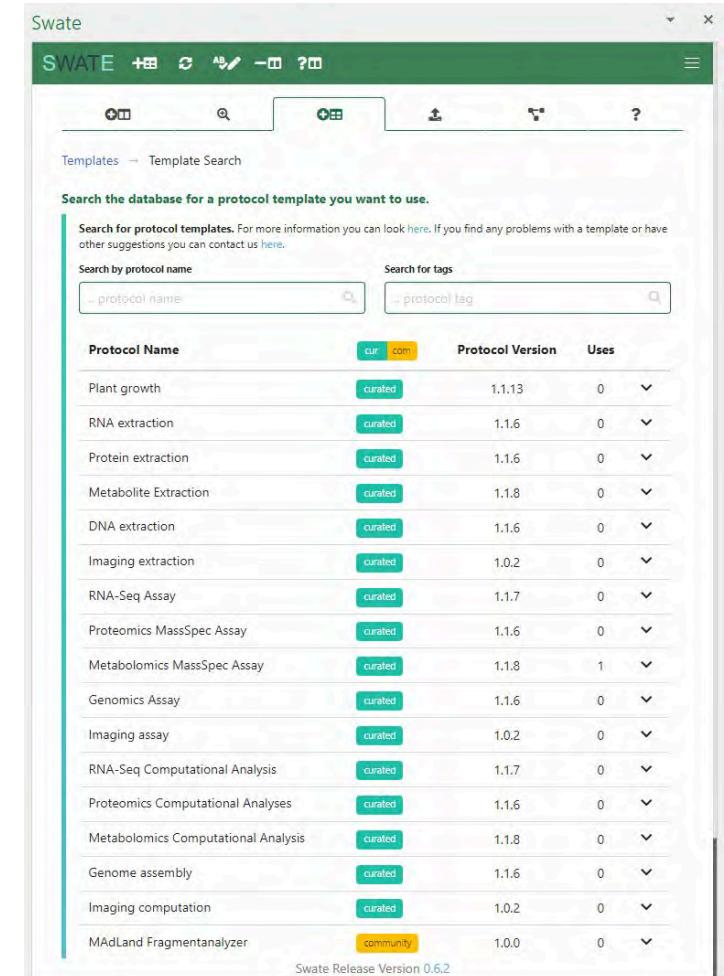
1. Navigate to the demo ARC.
2. Open the lab notes `assays/rnaseq/protocols/RNA_extraction.txt` in a text editor.
3. Open the empty `assays/rnaseq/isa.assay.xlsx` workbook in Excel.

Use a template

1. Navigate to *Templates* in the Navbar and click *Browse database* in the first function block.

 Here you can find 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.

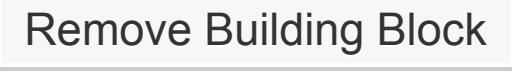


Adding / Updating unit references

Sometimes you need to add or update the unit of an existing building block.

1. Select any number of rows of the **Parameter [biosource amount]** building block to mark it for the next steps.
2. Open the *Building Blocks* tab
3. In the bottom panel "Add/Update unit reference to existing building block", search for the unit "milligram". Select the unit term and click **Update unit for cells**.
 If you already had values in the main column they will be updated automatically.
4. Click the *Update Ontology Terms* quick access button, to update the reference columns.

Remove building blocks

If there are any Building Blocks which do not fit your experiment you can use the  quick access button to remove it including all related (hidden) reference columns.

 Due to the hidden reference columns, we recommend not to delete table columns via usual Excel functions.

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

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.

A small detour on "Excel Tables"

Swate uses Excel's "table" feature to annotate workflows. Each table represents one *process* from input (e.g. plant leaf material) to output (e.g. leaf extract).

Example workflows with three *processes* each:

- Plant growth → sampling → extraction
- Measured data files → statistical analysis → result files

 Excel tables allow to group data that belongs together inside one sheet. This is not to be confused with a (work)sheet or workbook.

```
workbook          (e.g. "isa.assay.xlsx")
  └── worksheet    (e.g. "plant_growth")
    └── table       (e.g. "annotationTable")
```

Annotation with ARCitect

 Is not yet available.

Contributors

Slides presented here include contributions by

- name: Dominik Brilhaus
github: <https://github.com/brilator>
orcid: <https://orcid.org/0000-0001-9021-3197>
- name: Kevin Frey
github: <https://github.com/Freymaurer>
orcid: <https://orcid.org/0000-0002-8493-1077>
- name: Martin Kuhl
github: <https://github.com/Martin-Kuhl>
orcid: <https://orcid.org/0000-0002-8493-1077>
- name: Sabrina Zander



ARC Ecosystem Demo

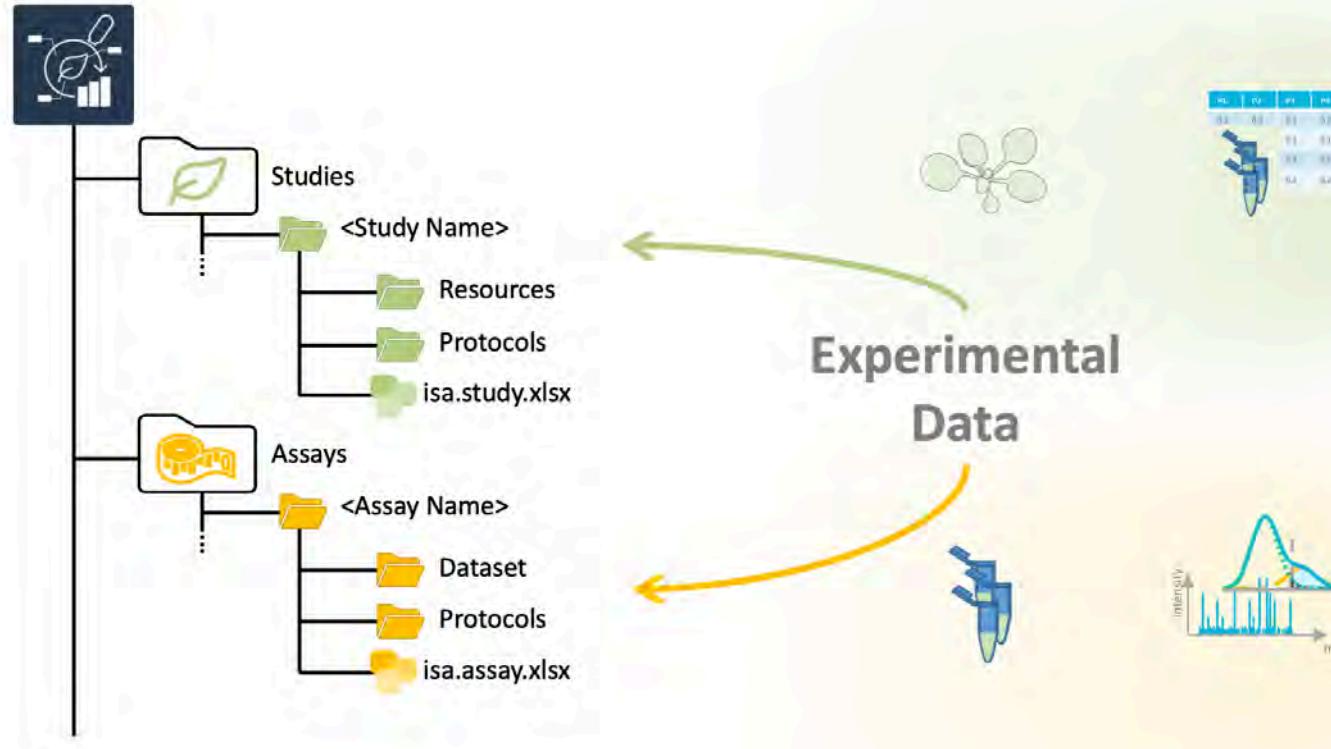
"A FAIR RDM journey along a (mutable) data life cycle"

October 11th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

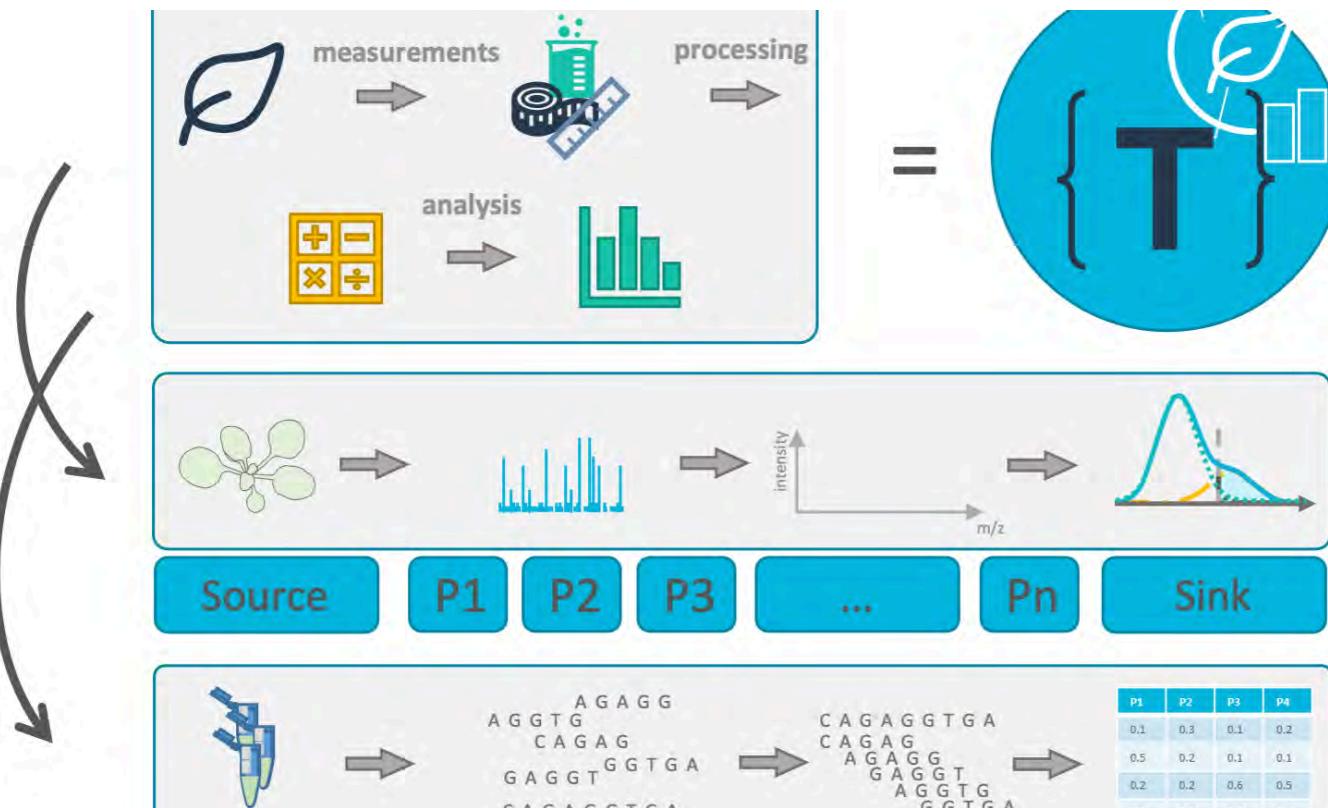


Collect



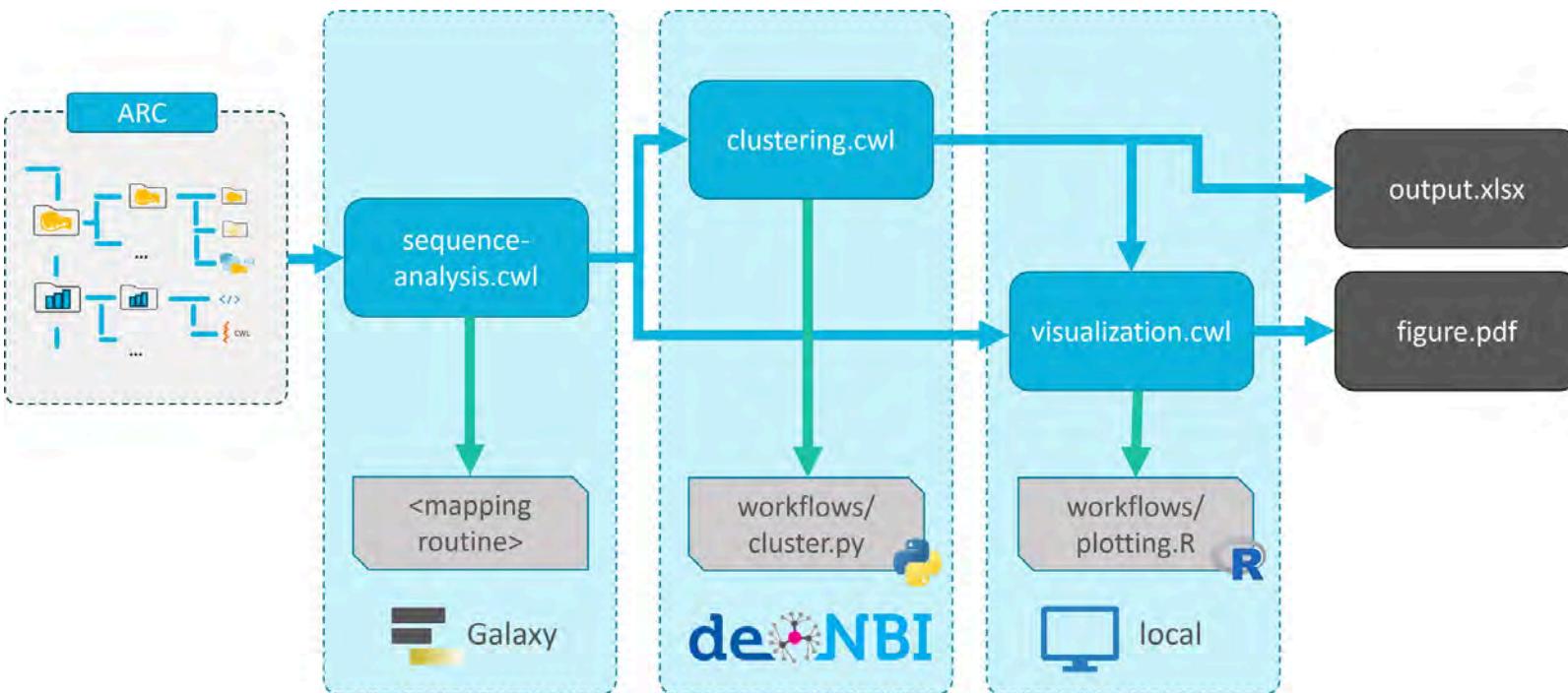


Process (e.g. annotate)



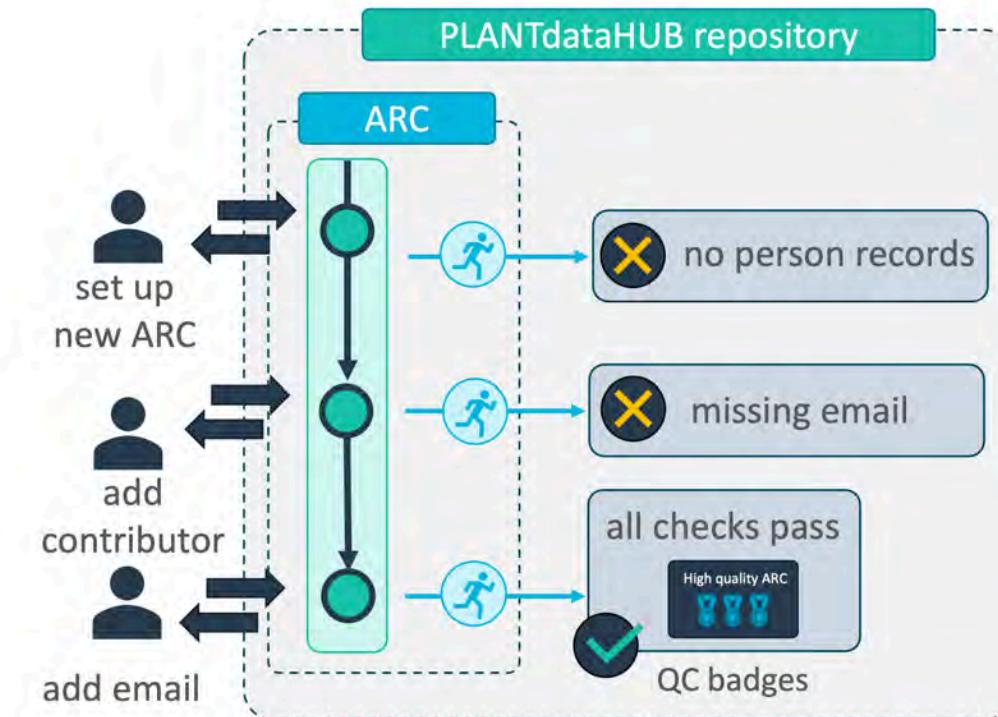


Analyse





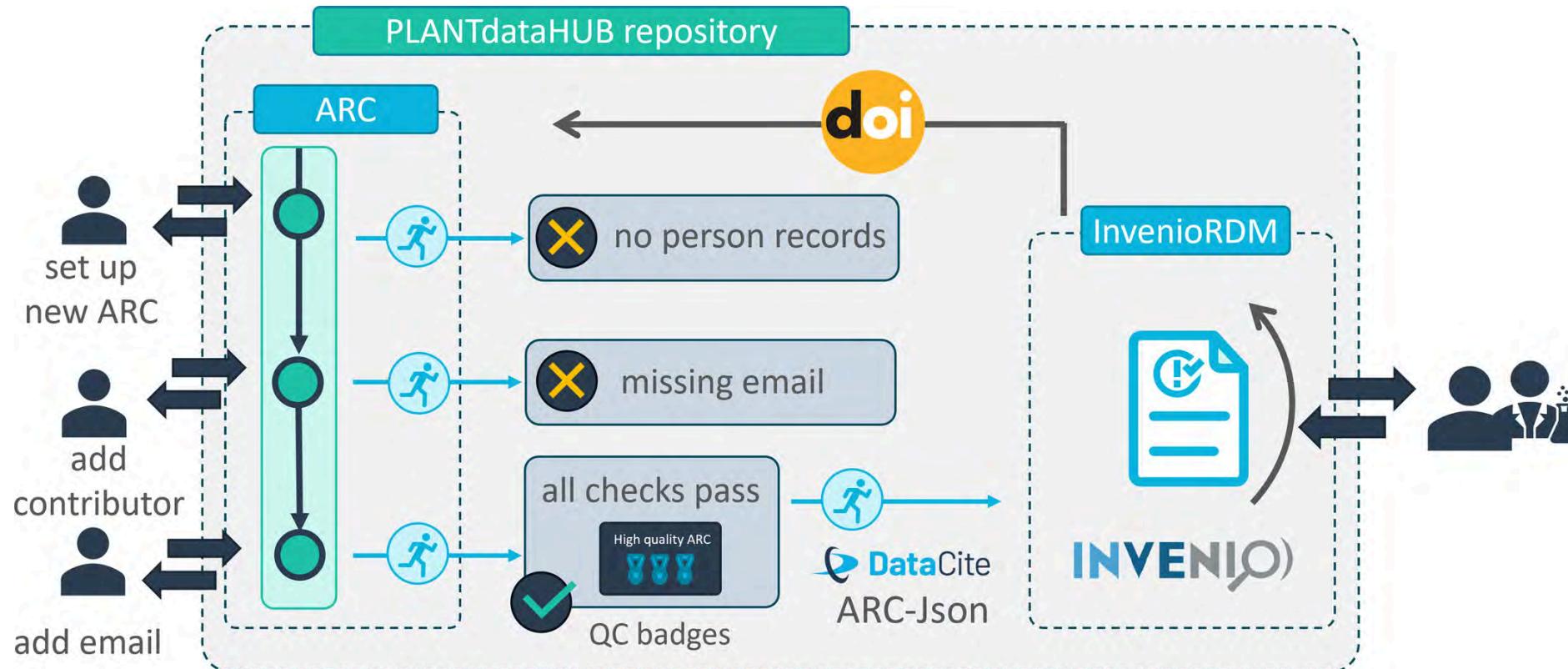
Preserve



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>



Preserve and publish



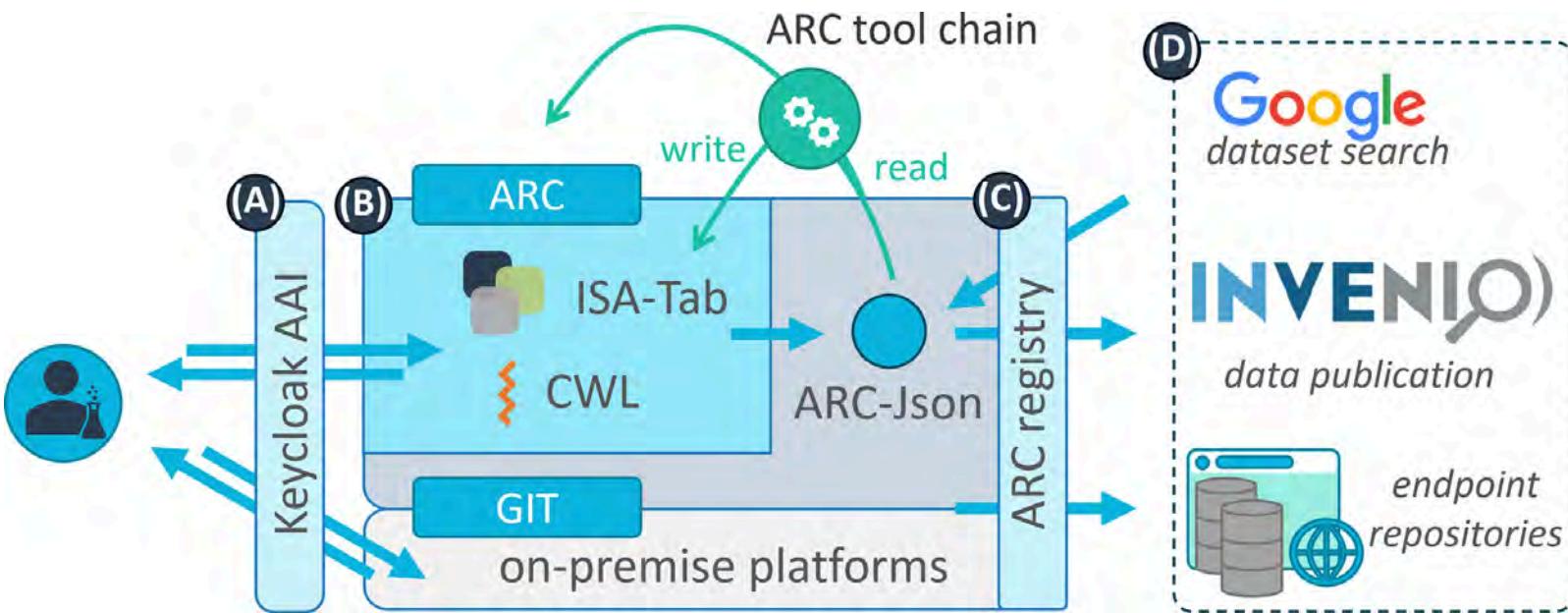


Share and collaborate

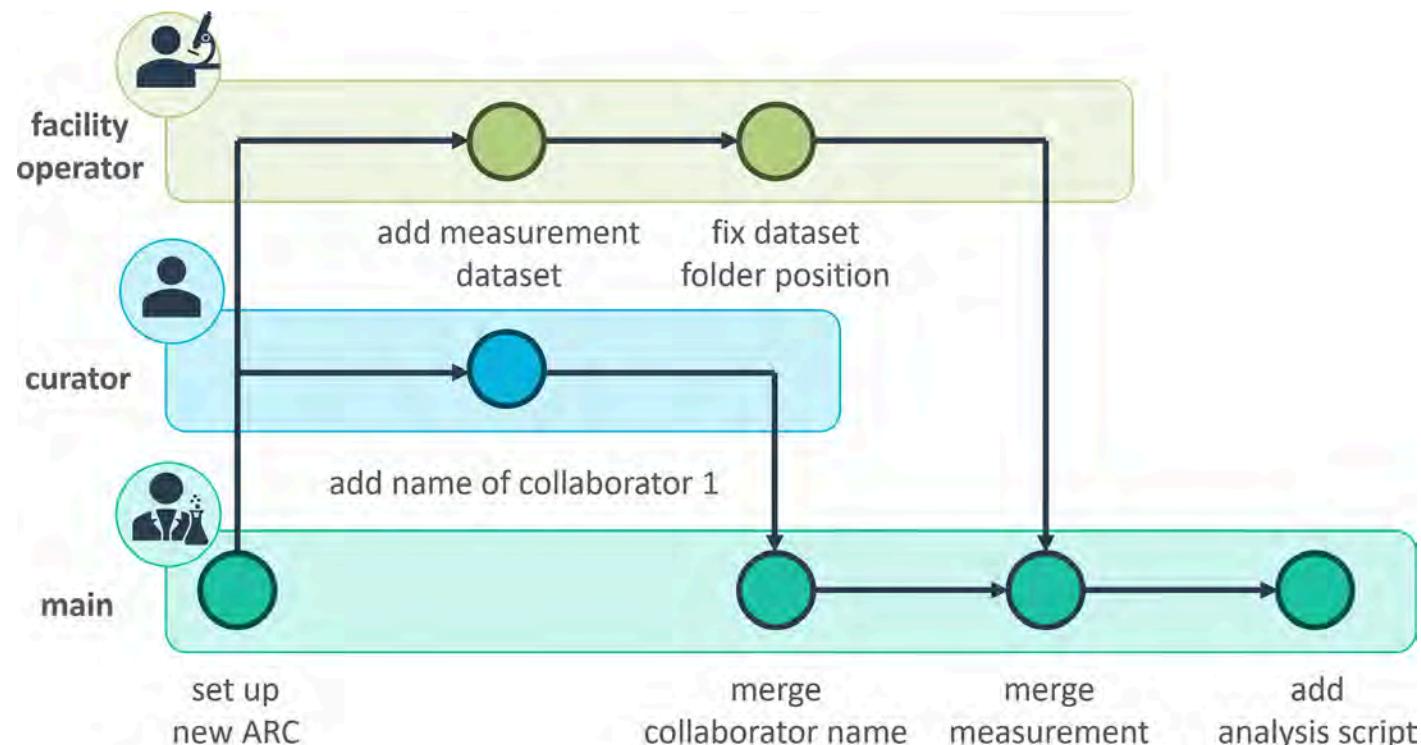




Reuse

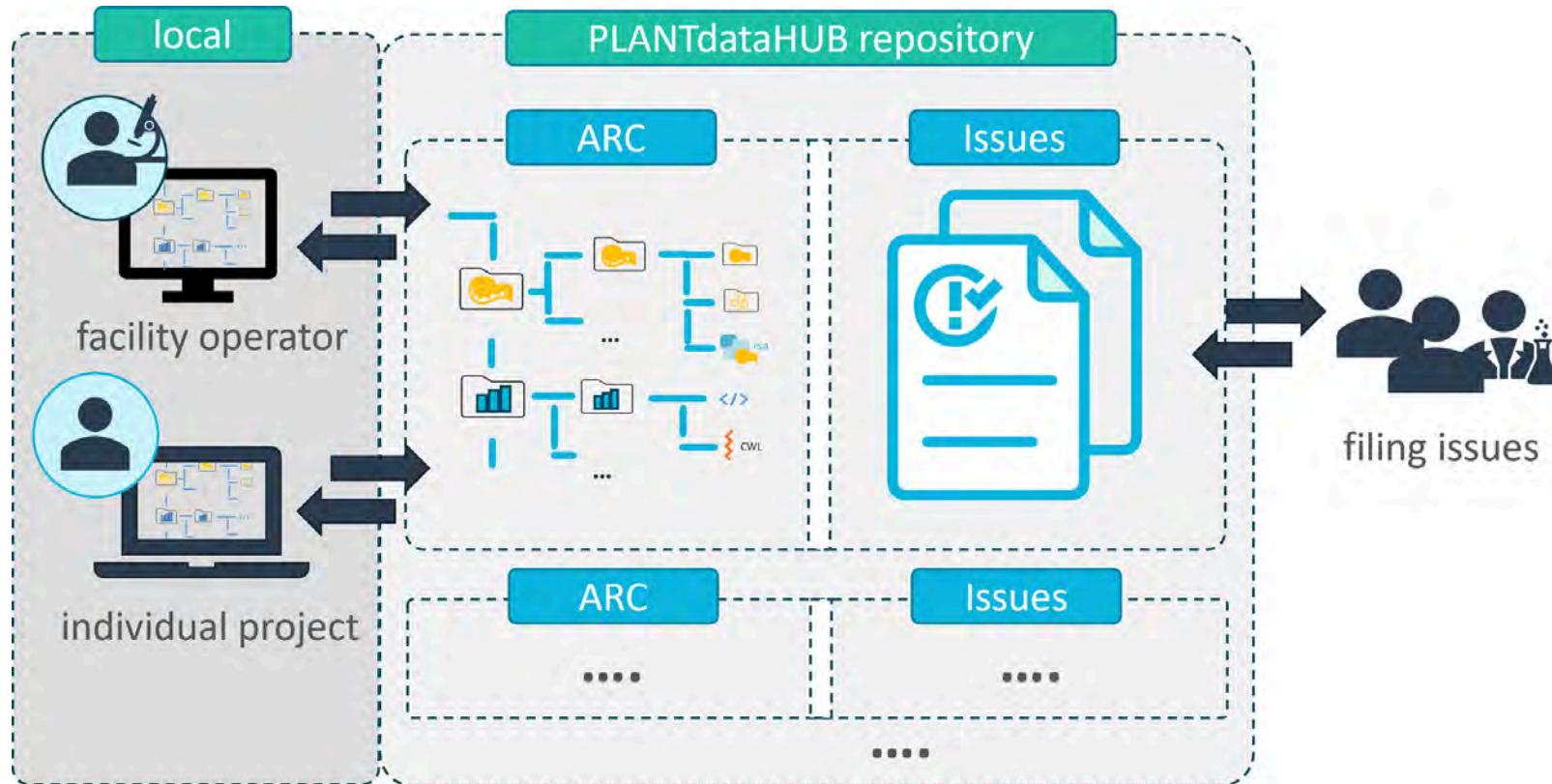


Mutable data life cycle





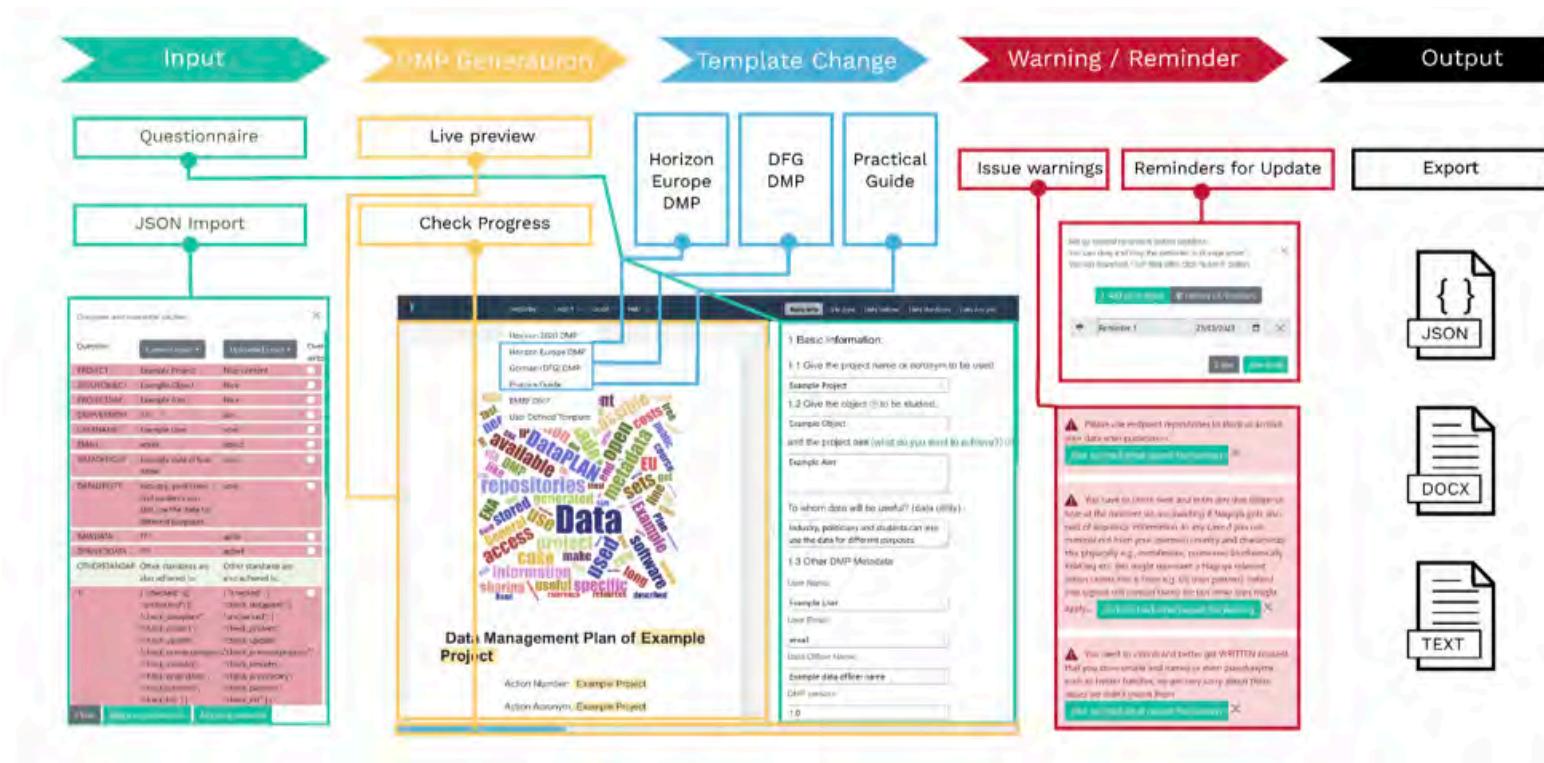
Plan (ARC scale)





Plan (proposal scale)

<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>

Contributors

Slides presented here include contributions by

- name: Dominik Brilhaus
github: <https://github.com/brilator>
orcid: <https://orcid.org/0000-0001-9021-3197>

