

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](#).

The screenshot shows a registration form titled "SIGN UP" with the sub-instruction "Get access to DataPLANT infrastructure and services". It features three input fields: "Email", "First name", and "Last name". Below these is a blue "Next" button with a right-pointing arrow. To the right of the form is a large blue text block explaining the benefits of registration, mentioning FAIR Digital Objects, seamless collaborations, and Git versioning.

SIGN UP

Get access to DataPLANT infrastructure and services

Email

First name

Last name

Next ▶

The infrastructure, tools, and workflows we offer support you in transforming your results into FAIR Digital Objects and enable seamless collaborations between you and your lab-members or even project partners from multiple labs. Thanks to the versioning feature of Git, every step is traceable at any time, preserving the provenance of each dataset. Do not hesitate and register using our Keycloak Single Sign-On solution.

Role and consortium

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

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

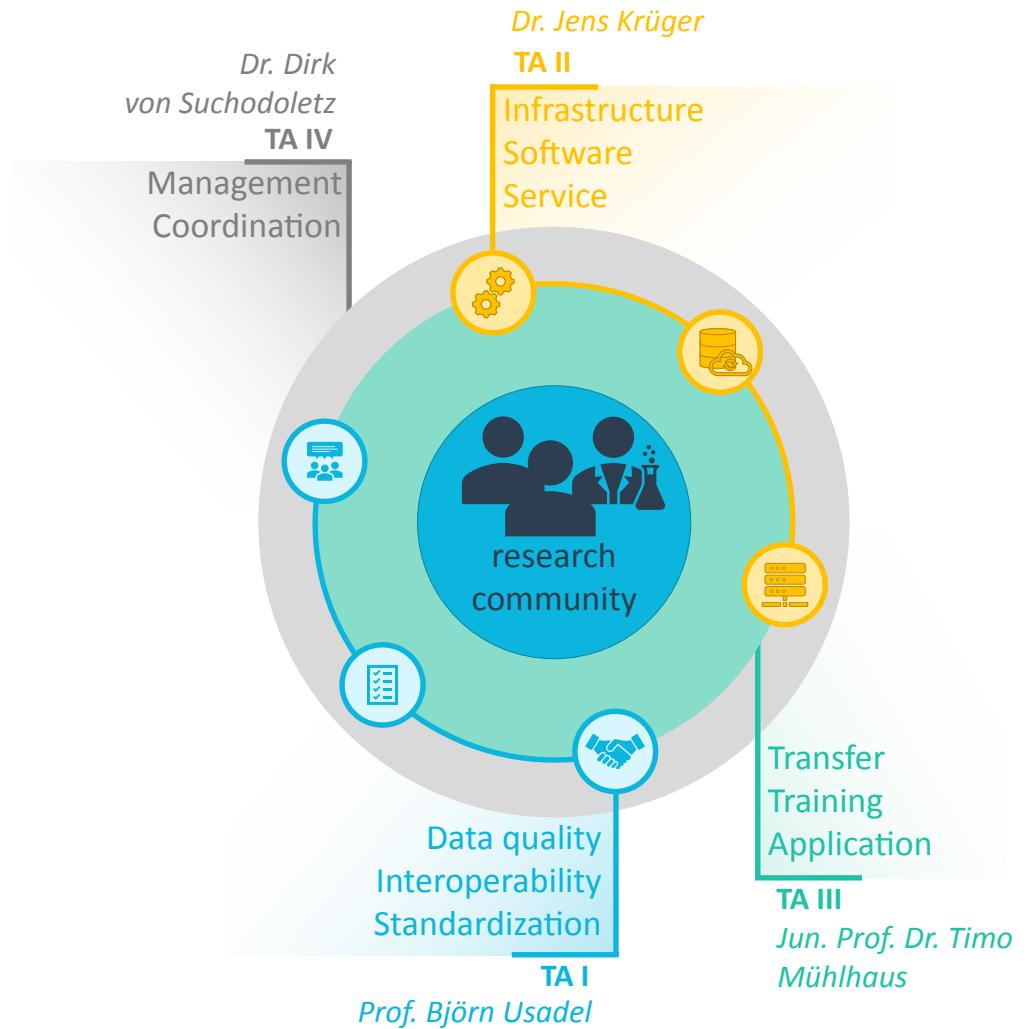
Block 2 – Intro to DataPLANT and ARC

October 5th, 2023

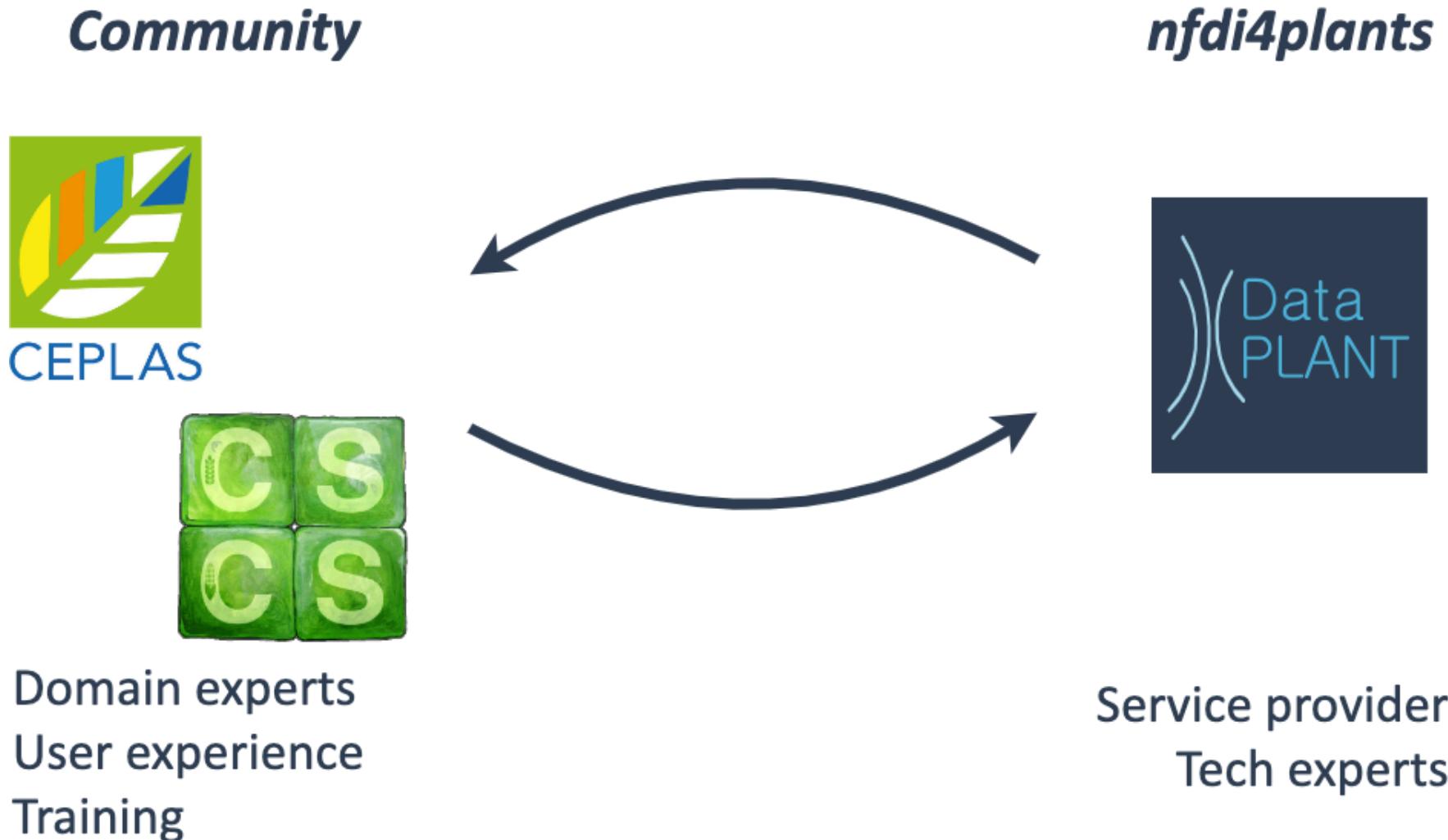
Dominik Brilhaus, [CEPLAS Data Science](#)

DataPLANT – The NFDI4Plants

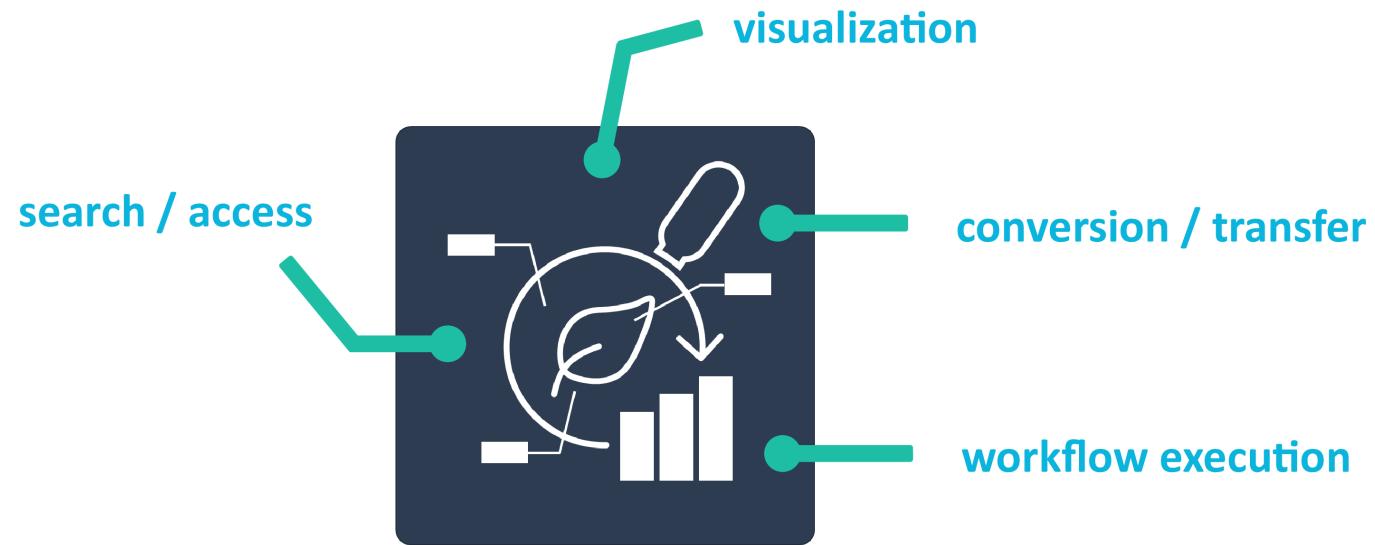
- NFDI: "Nationale Forschungsdaten Infrastruktur" – www.nfdi.de
- Funded since end of 2020



Data Stewardship between DataPLANT and the community



Annotated Research Context (ARC)





FINDABLE

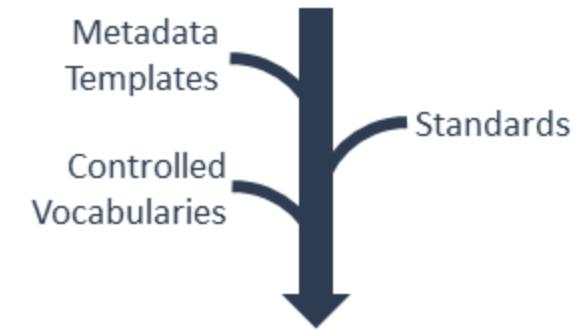
ACCESSIBLE

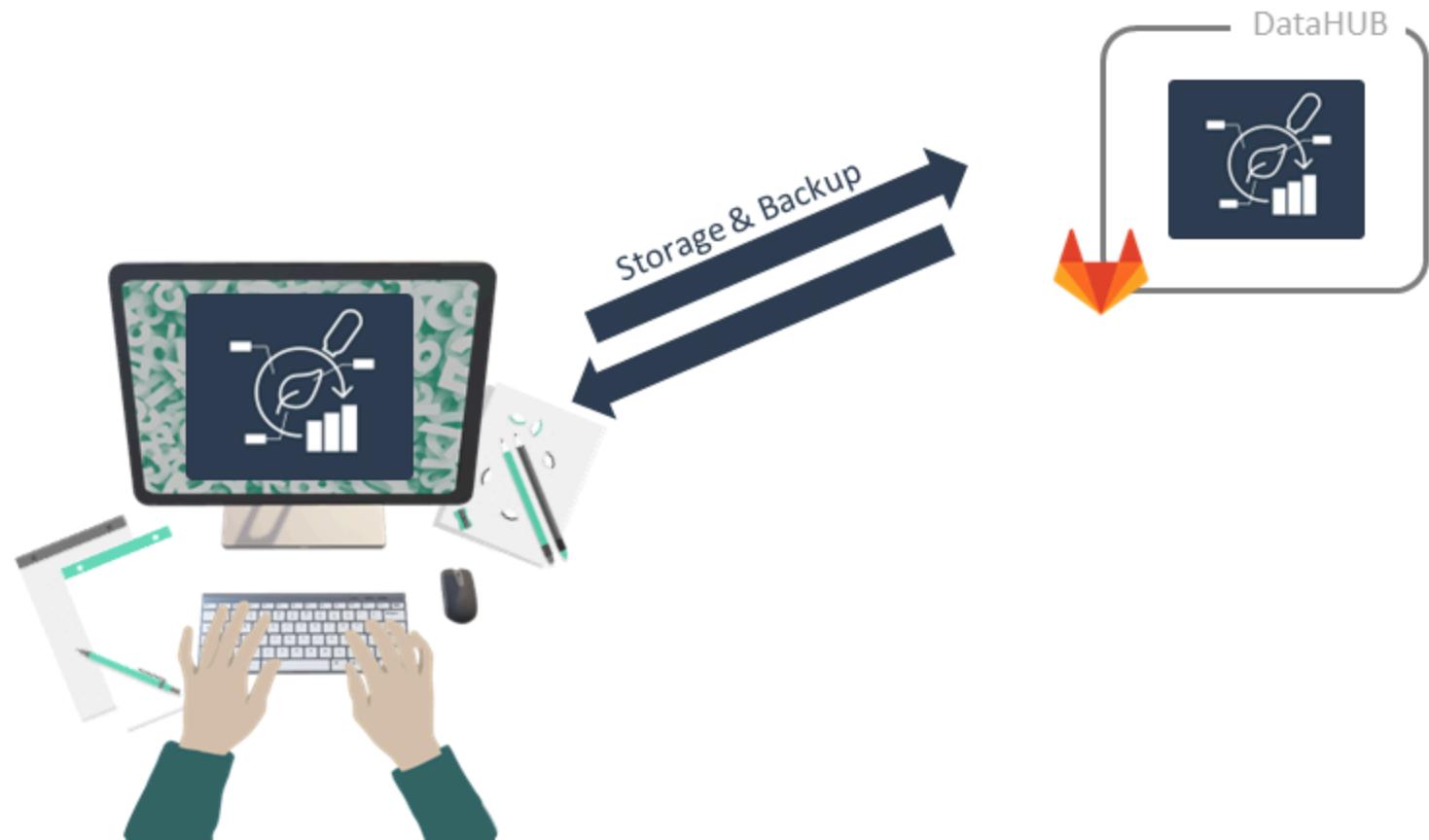
INTEROPERABLE

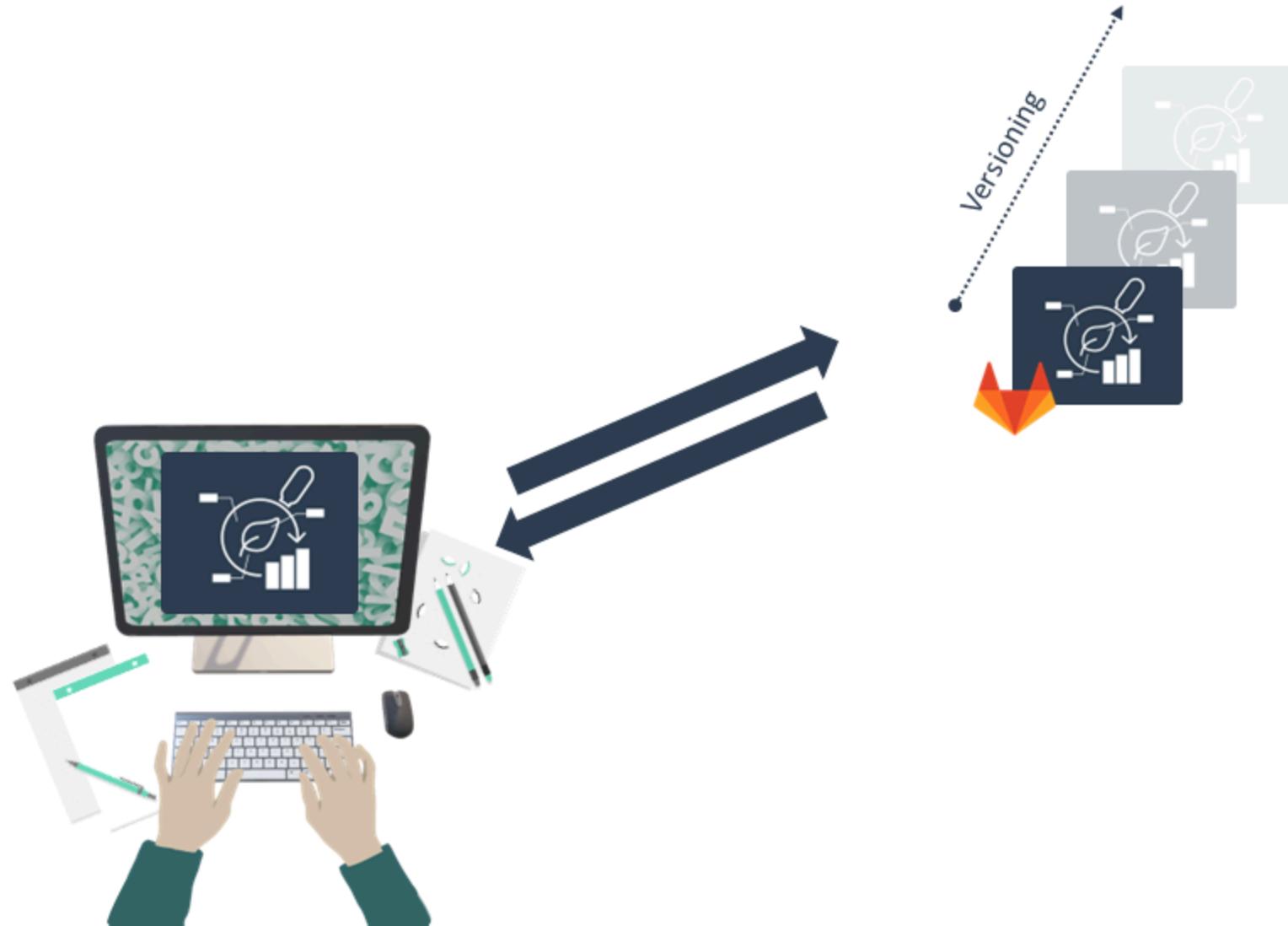
REUSABLE

Metadata
Templates

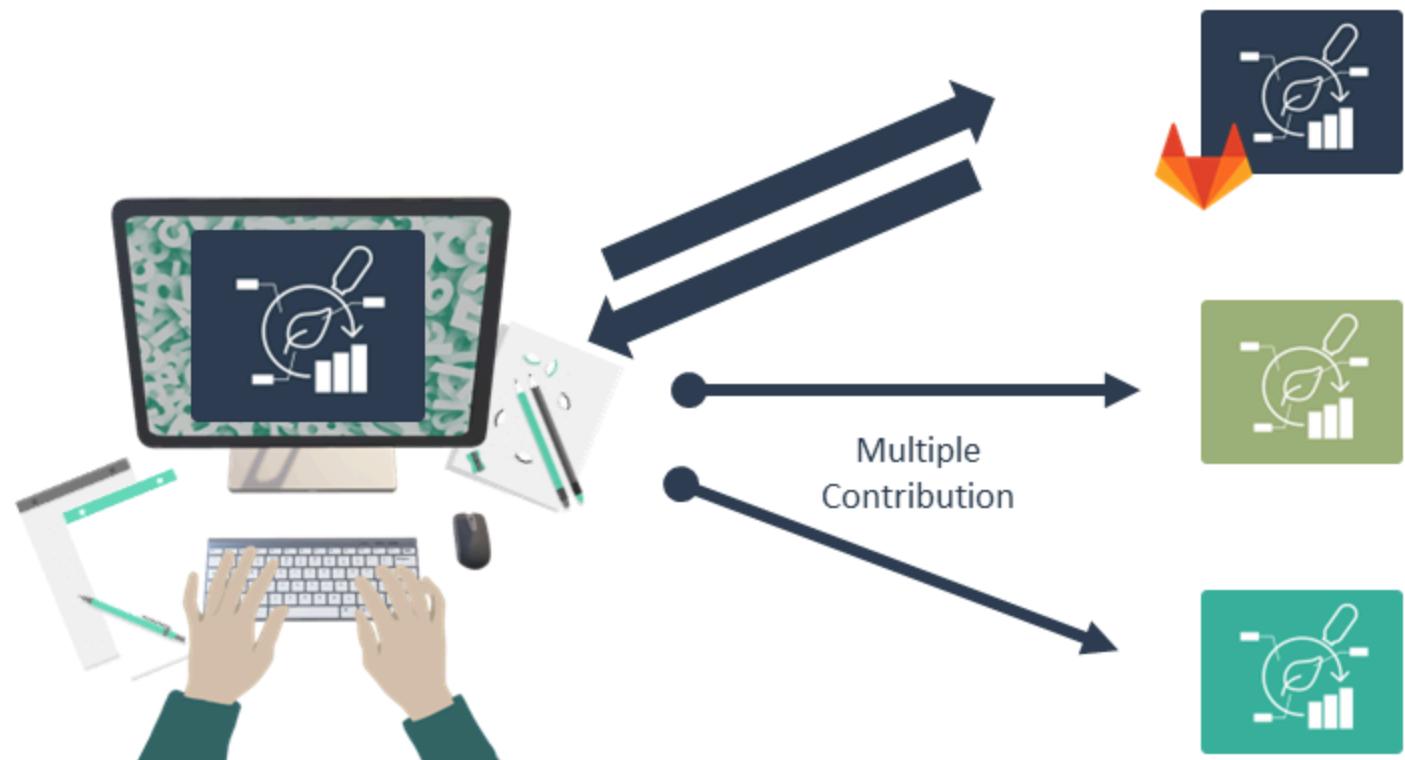
Controlled
Vocabularies



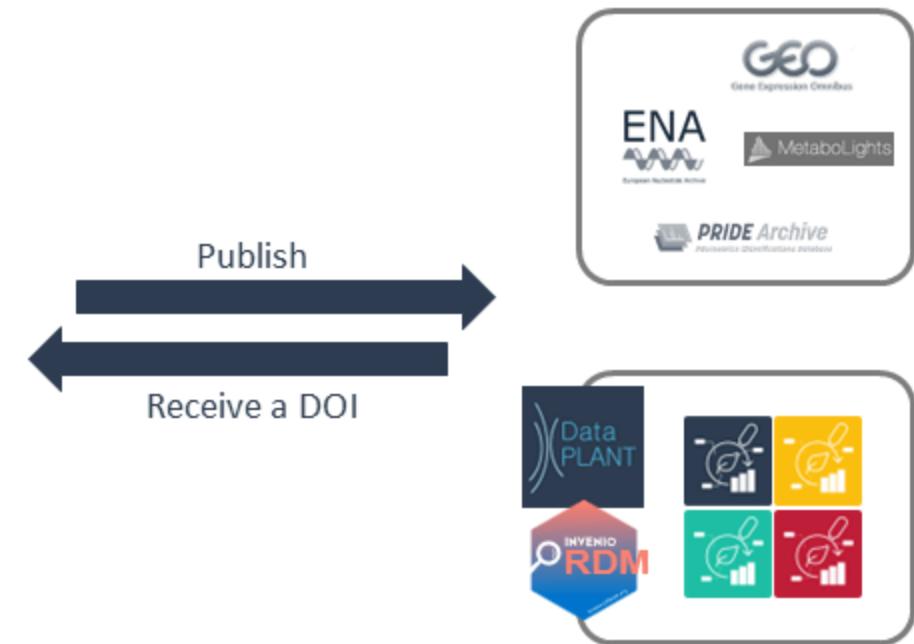


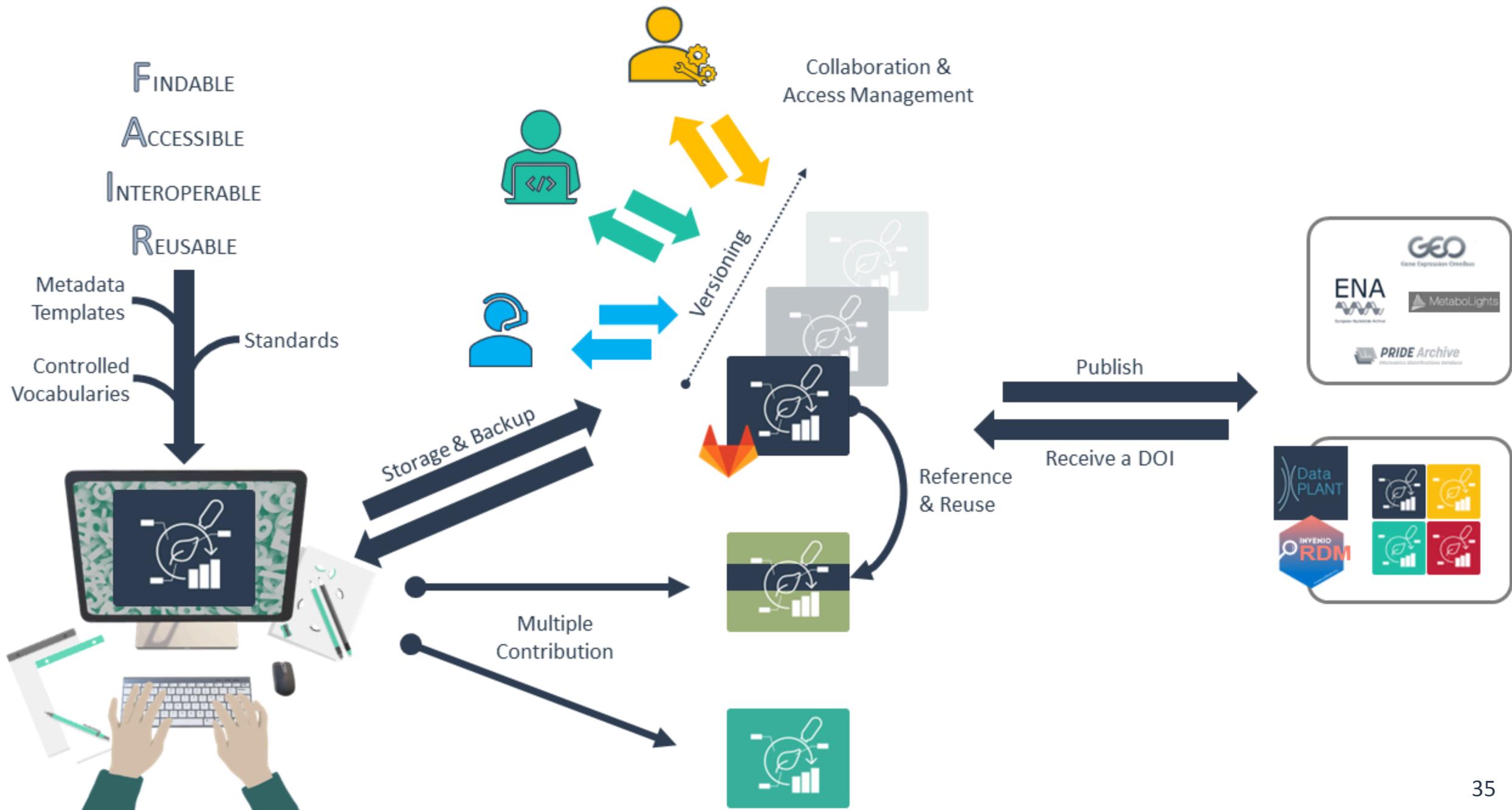




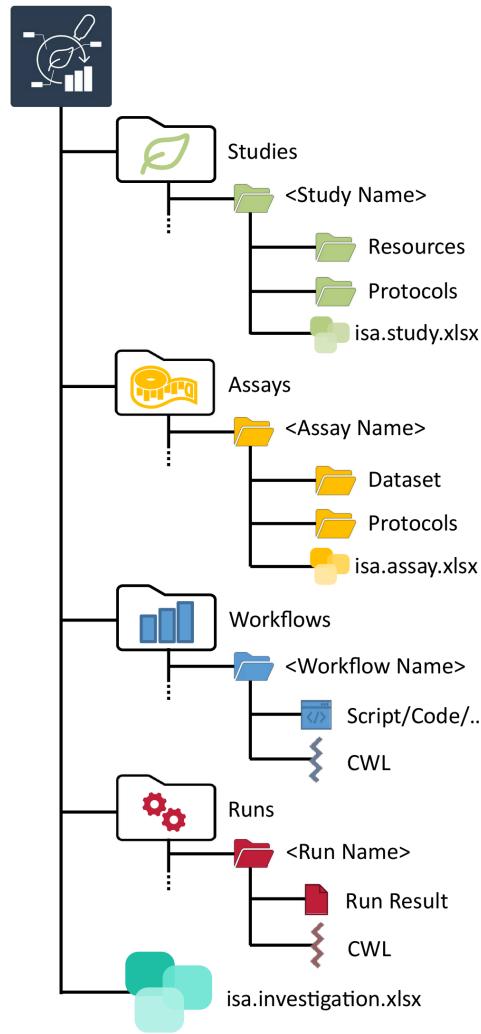




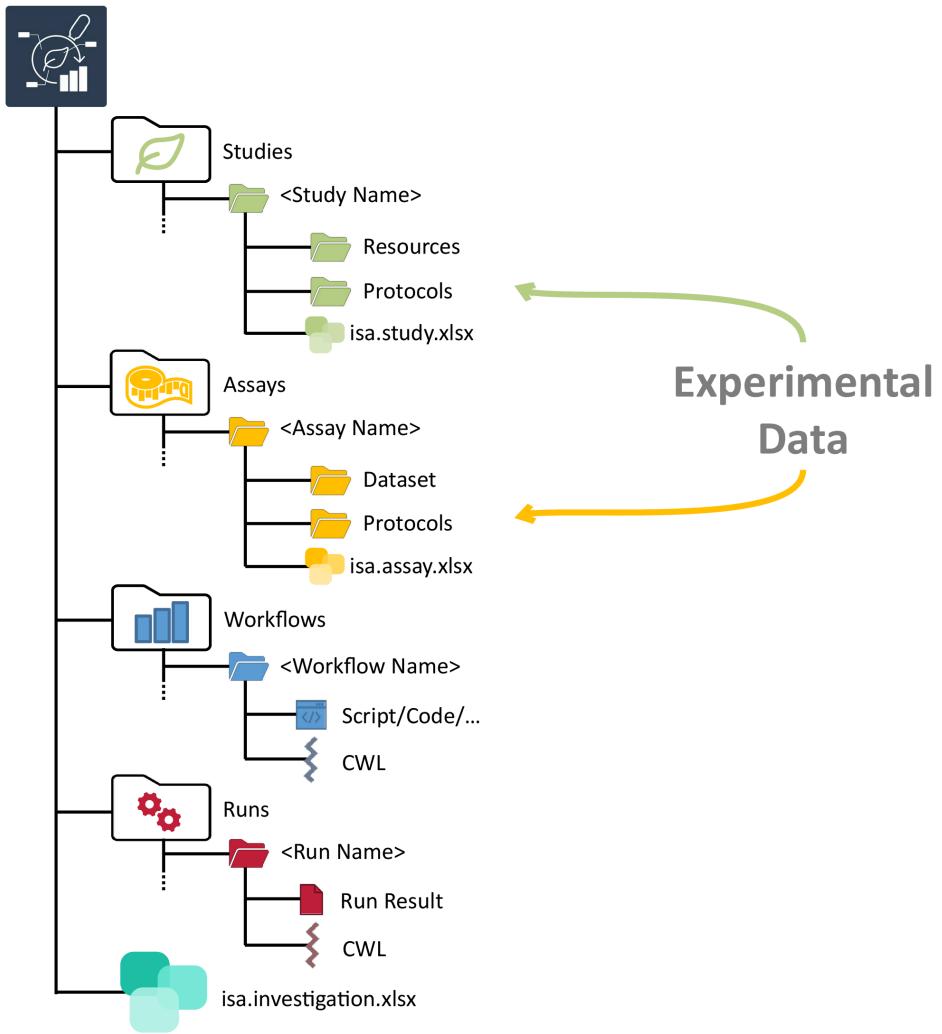




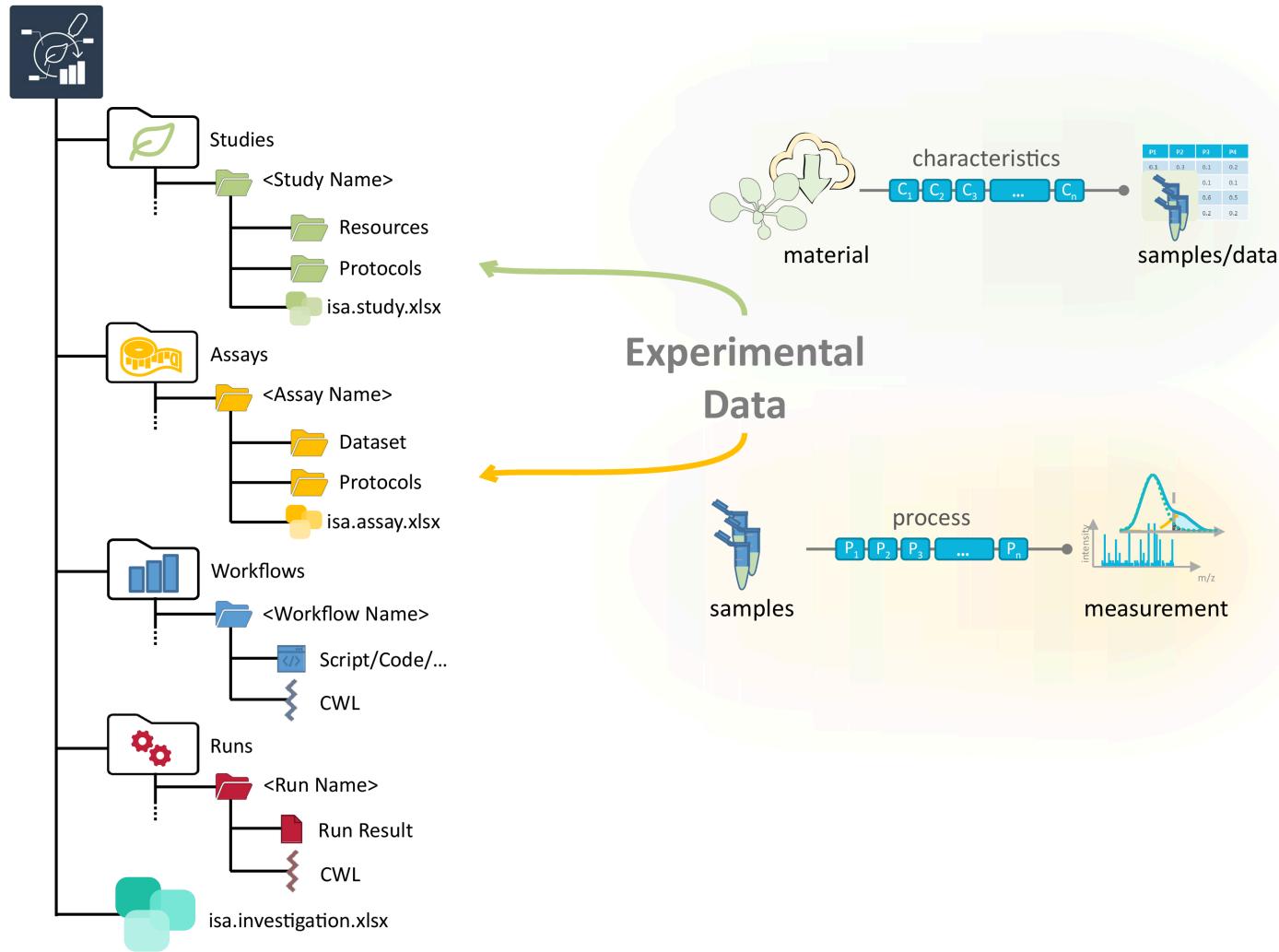
What does an ARC look like?



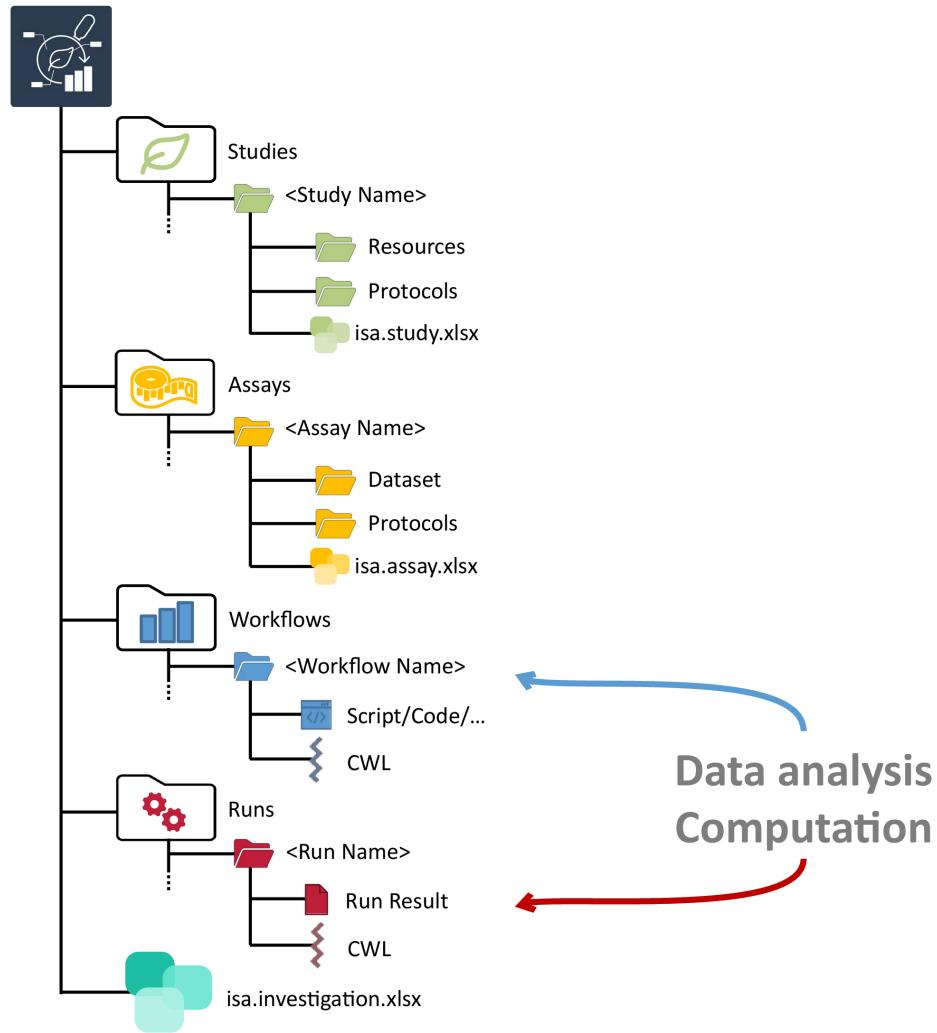
What does an ARC look like?



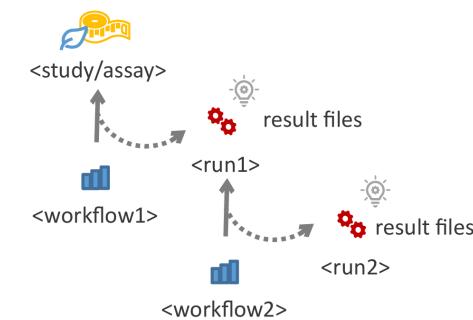
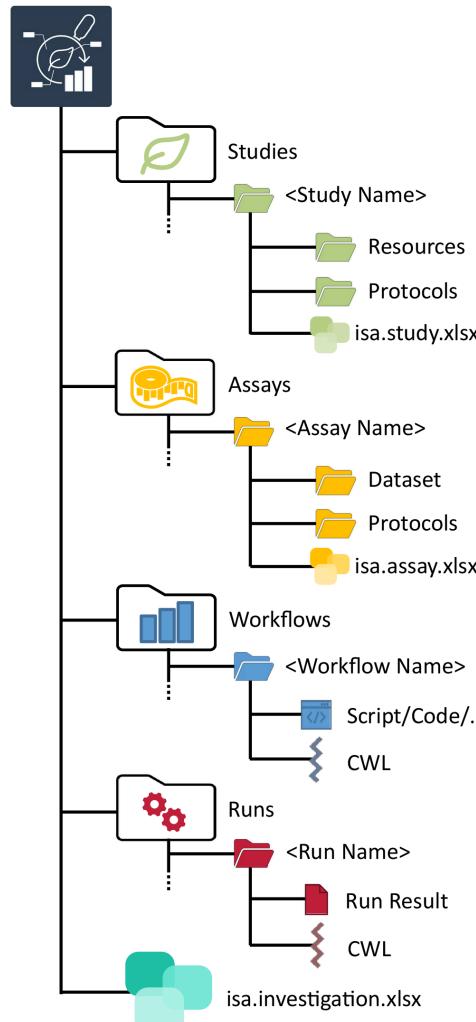
What does an ARC look like?



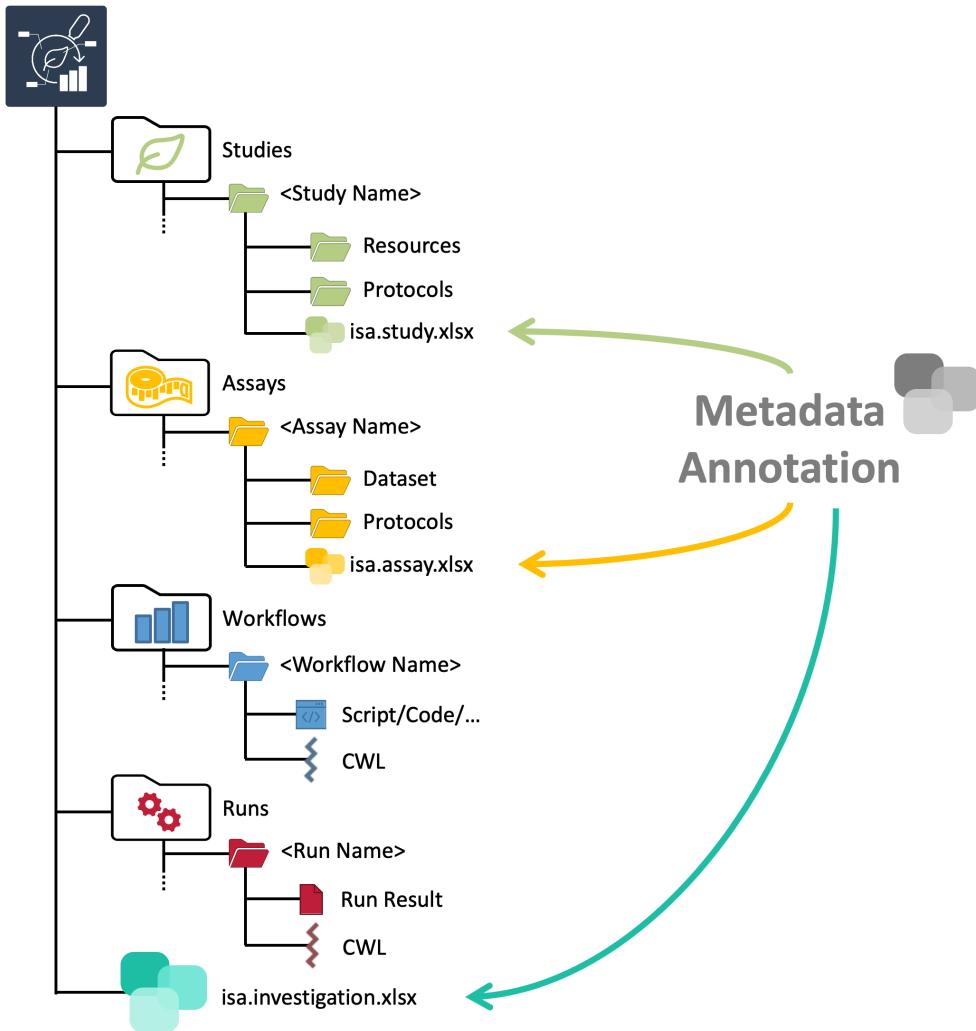
What does an ARC look like?



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

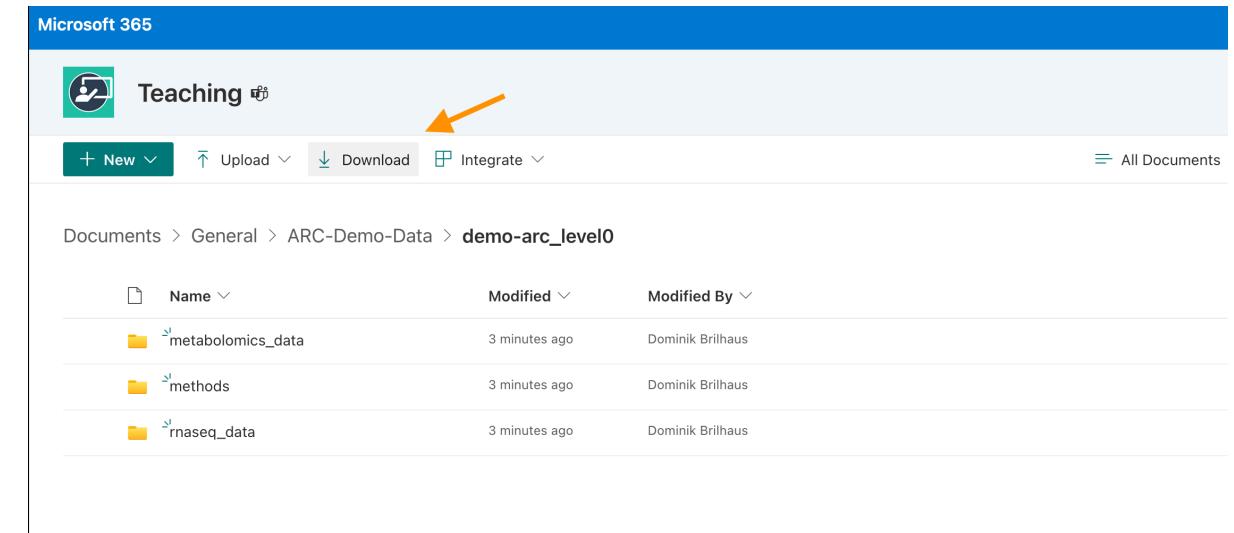
Did everyone [sign-up](#) at the DataHUB?

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



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

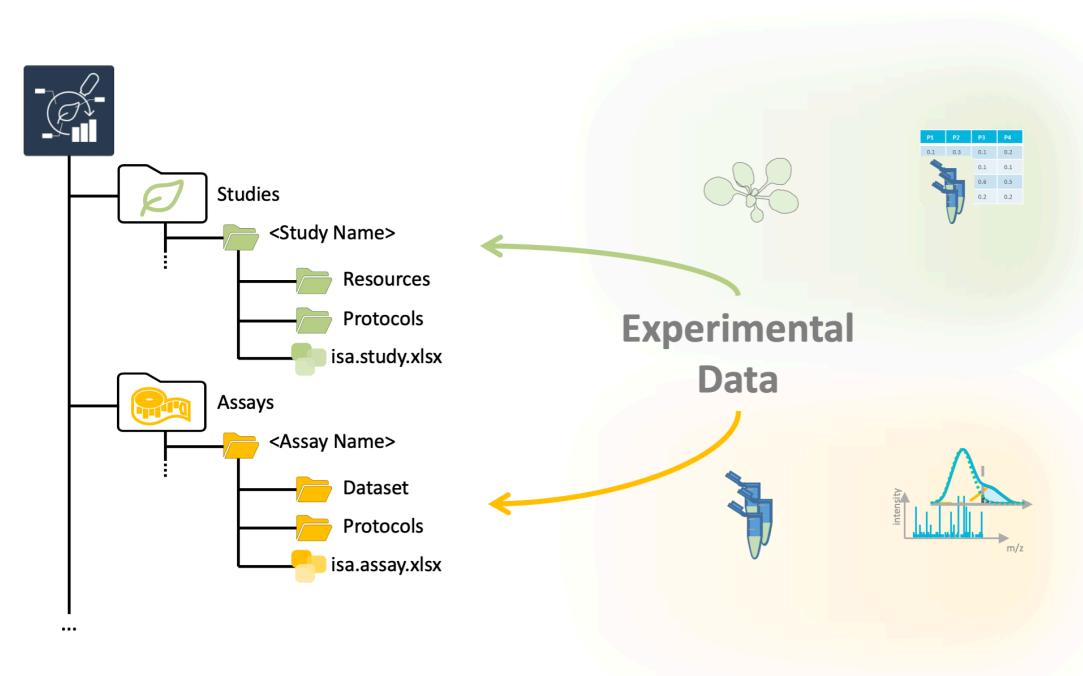
You just received your data

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

Goal

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

 We'll talk about data annotation
later

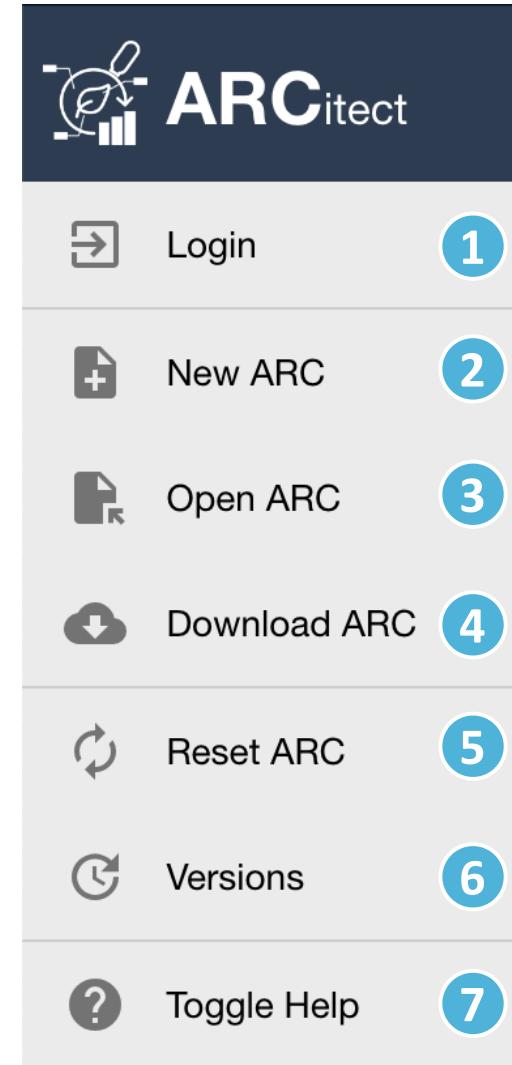


Open the ARCitect

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

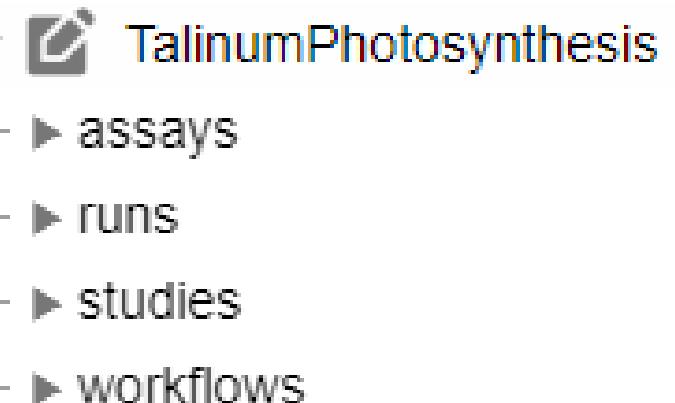
<https://git.nfdi4plants.org/<YourUserName>/<YourARC>> (see next steps)

3. as the identifier for your investigation

💡 Make sure that no ARC exists at

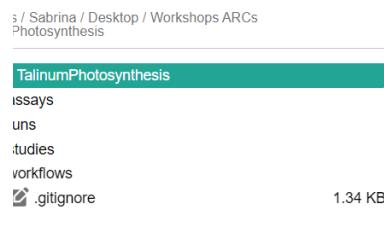
<https://git.nfdi4plants.org/<YourUserName>/<YourARC>> . Otherwise you will sync to that ARC.

💡 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

A screenshot of an 'Update' dialog box. It has fields for 'Identifier' (set to 'TalinumPhotosynthesis'), 'Title' (set to 'Talinum Photosynthesis'), and 'Description' (containing the text 'This is a very interesting investigation about life and photosynthesis'). There is also a 'Cancel' button at the top left and an 'Update' button at the bottom right.

Add contributors

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

/ Users / dominikbrilhaus / Desktop
/ TalinumPhotosynthesis

▼  **TalinumPhotosynthesis**

- ▶ assays
- ▶ runs
- ▶ studies
- ▶ workflows



Investigation

General Meta Data of the Investigation



People

Authors and Collaborators



ADD PERSON



Publications

Papers, Books and Other Media



Add contributor details

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

- First Name
- Last Name
- Email
- Affiliation

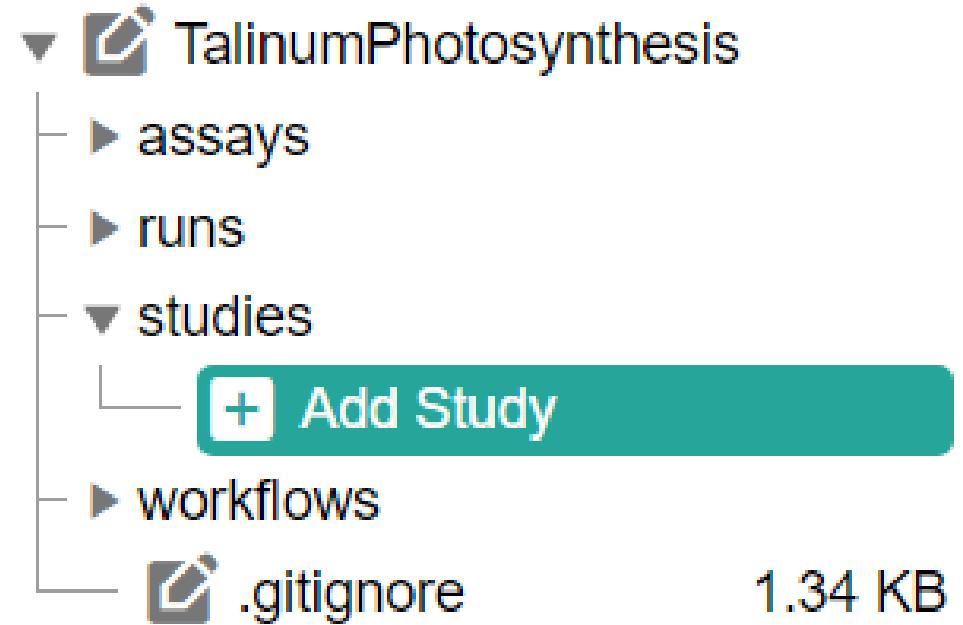
Contacts

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

Description

Contacts

Publications

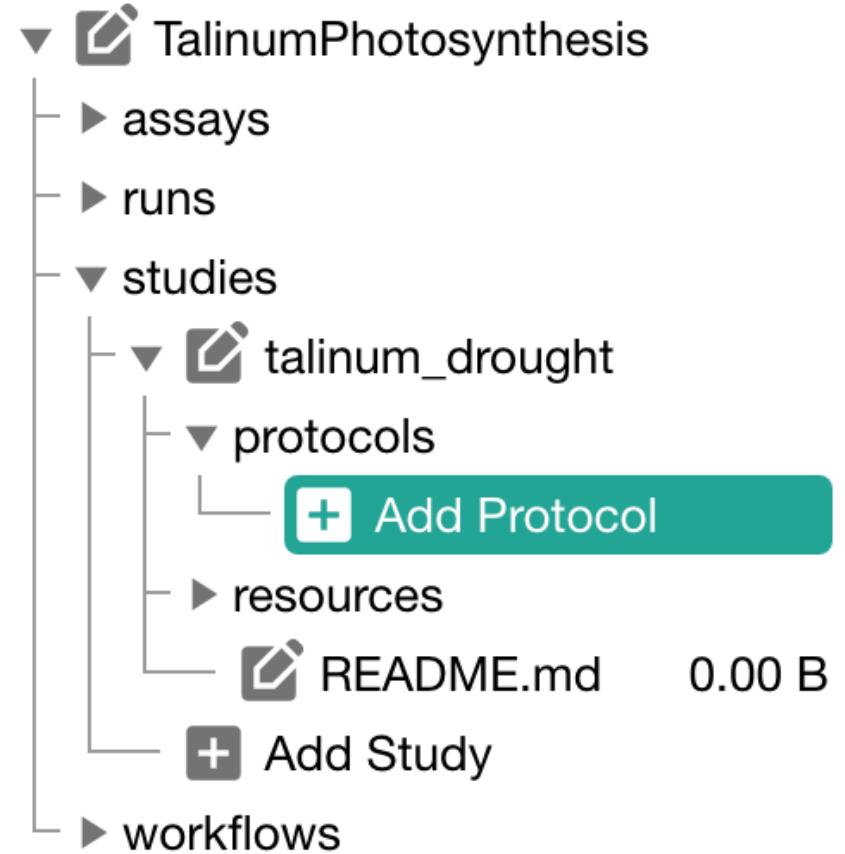
Submission Date

Public Release Date

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

Create or Import Protocol

Protocol Name



NEW PROTOCOL



IMPORT PROTOCOL

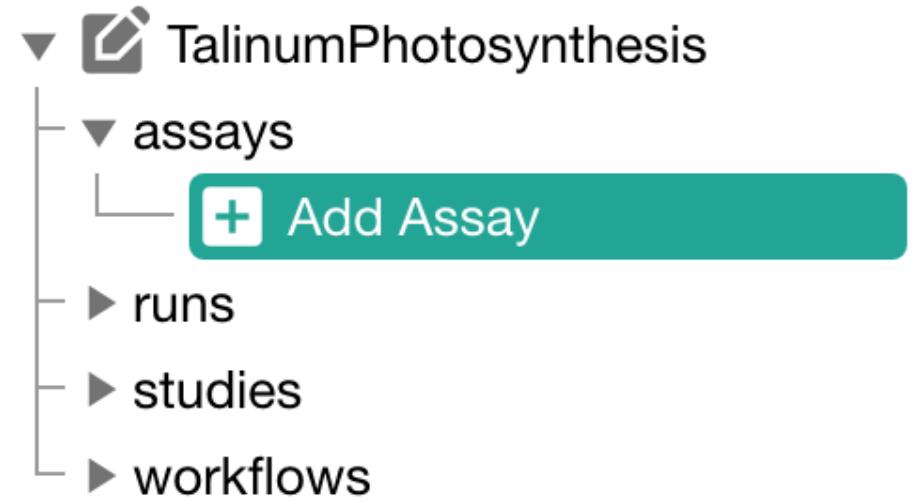
CANCEL

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		
<input type="text" value="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:

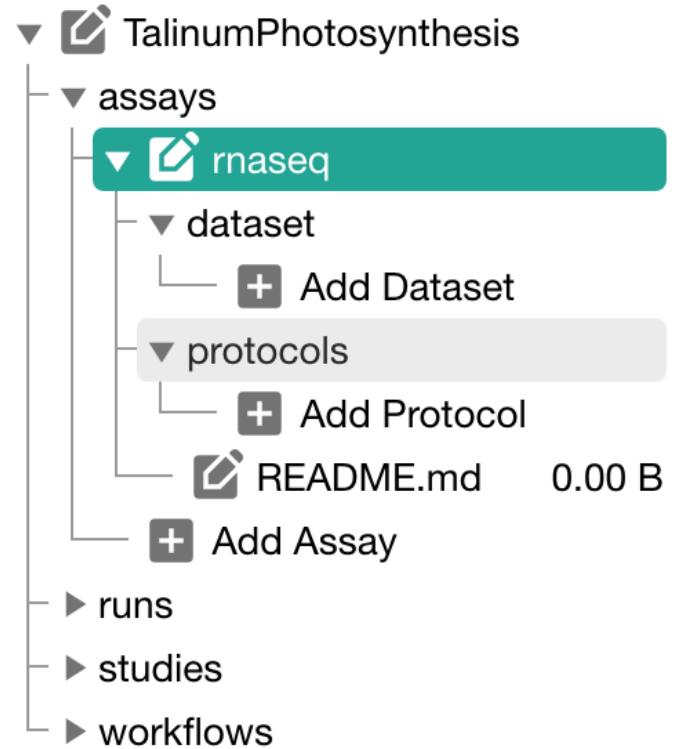
- Measurement Type: Gene Expression Analysis
- Technology type: Next Generation Sequencing
- Technology platform: Illumina HiSeq 2500

2. Click "Update" to save your changes

Add protocols and datasets

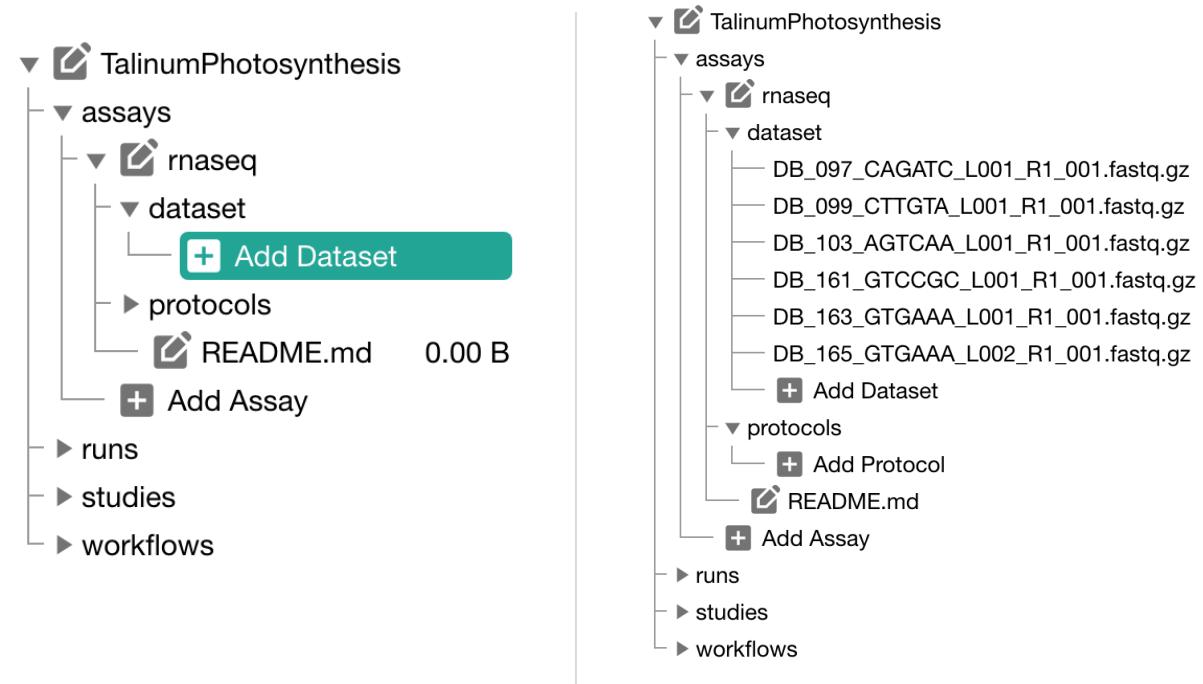
In the file tree you can

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



Import the demo dataset to the ARC

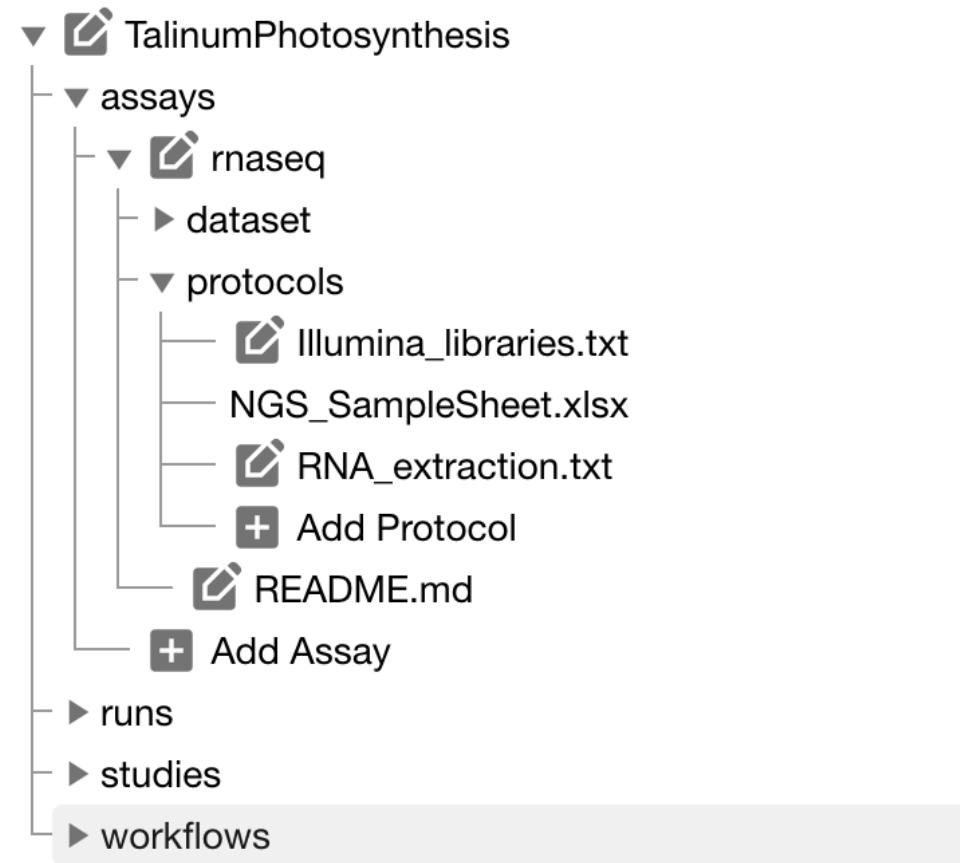
1. Click "Add Dataset"
2. Select the *.fastq.gz files from the demo data



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`



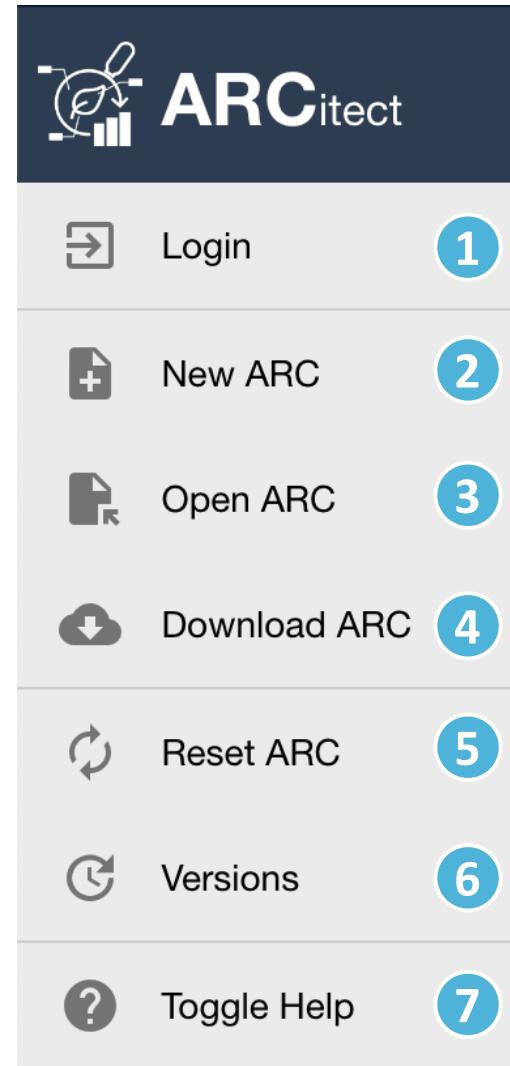
Collaborate and share



Login to the DataHUB

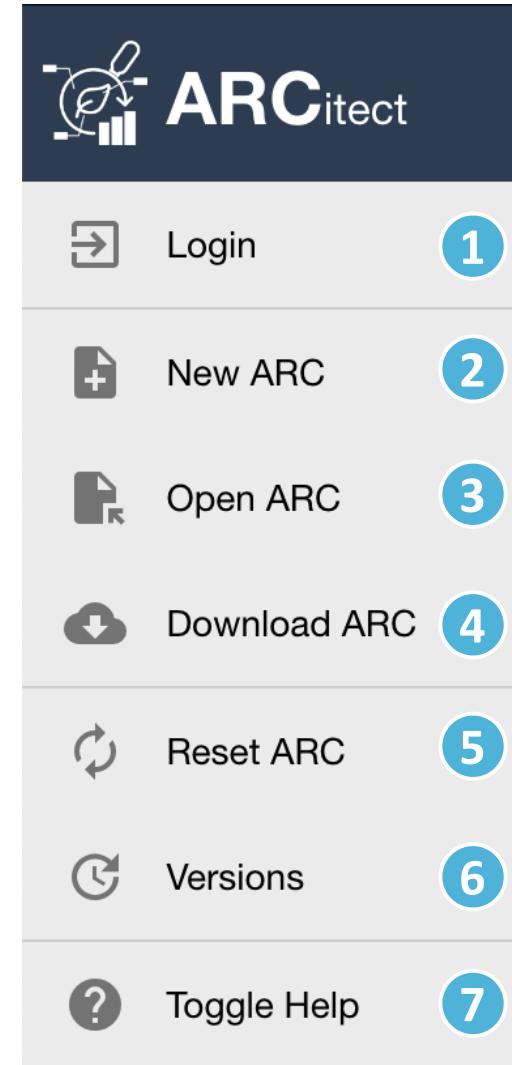
Click **Login** (1) in the sidebar to login to the DataHUB.

 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' input field is present with a placeholder 'A short description of the made changes'. Below it is a 'Changes' list with items like '.arc/', 'assays/', 'isa.investigation.xlsx', 'runs/', 'studies/', and 'workflows/'. At the bottom are buttons for 'REFRESH', 'COMMIT', 'UPLOAD', and 'DOWNLOAD'.

Update
Commit changes and upload ARC

Full Name
Demo User

eMail
demo@nfdi4plants.org

Remote
https://git.nfdi4plants.org/demouser/Demo-ARC.git

Commit Message

A short description of the made changes

Changes

- + .arc/
- + assays/
- + isa.investigation.xlsx
- + runs/
- + studies/
- + workflows/

REFRESH COMMIT UPLOAD DOWNLOAD

History
Inspect ARC history

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>>

The screenshot shows the 'Update' section of the DataHUB interface. It includes fields for 'Full Name' (Demo User) and 'eMail' (demo@nfdi4plants.org), a 'Remote' URL (https://git.nfdi4plants.org/demouser/Demo-ARC.git), a large 'Commit Message' input field, and a 'Changes' list. The changes listed are: .arc/, assays/, isa.investigation.xlsx, runs/, studies/, and workflows/. At the bottom are buttons for 'REFRESH', 'COMMIT', 'UPLOAD', and 'DOWNLOAD'.

Update
Commit changes and upload ARC

Full Name
Demo User

eMail
demo@nfdi4plants.org

Remote
https://git.nfdi4plants.org/demouser/Demo-ARC.git

Commit Message

A short description of the made changes

Changes

- + .arc/
- + assays/
- + isa.investigation.xlsx
- + runs/
- + studies/
- + workflows/

REFRESH COMMIT UPLOAD DOWNLOAD

History
Inspect ARC history

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

Update
Commit changes and upload ARC

Full Name
Demo User

eMail
demo@nfdi4plants.org

Remote
<https://git.nfdi4plants.org/demouser/Demo-ARC.git>

Commit Message

A short description of the made changes

Changes

- + .arc/
- + assays/
- + isa.investigation.xlsx
- + runs/
- + studies/
- + workflows/

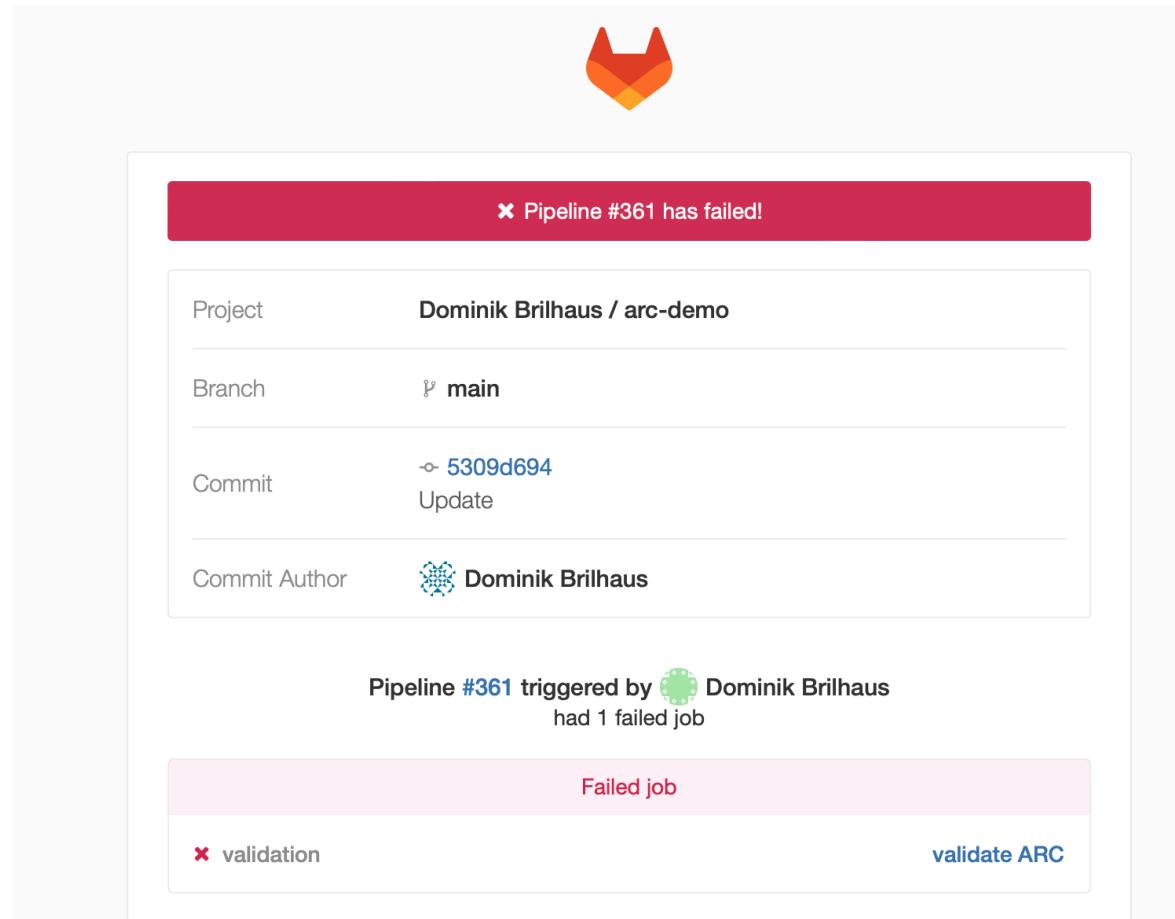
 REFRESH  COMMIT  UPLOAD  DOWNLOAD

History
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?



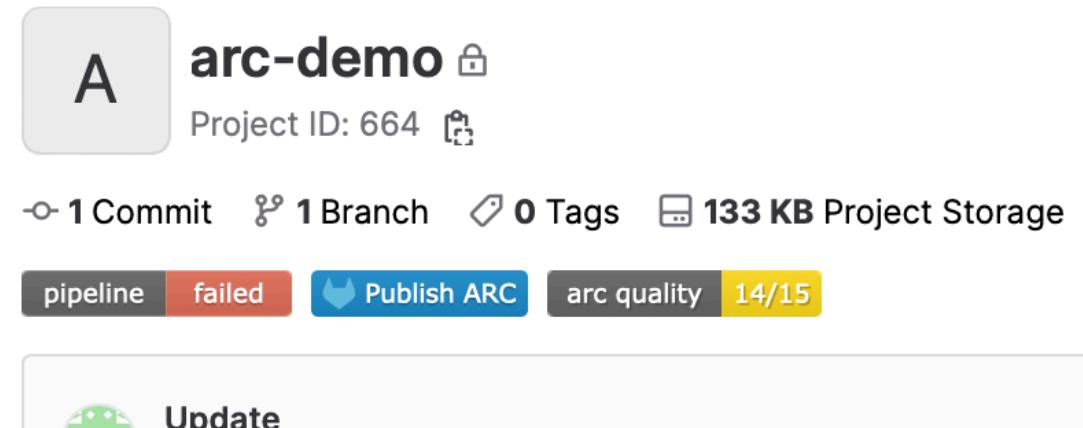
Don't worry 😊



Pipeline Failed

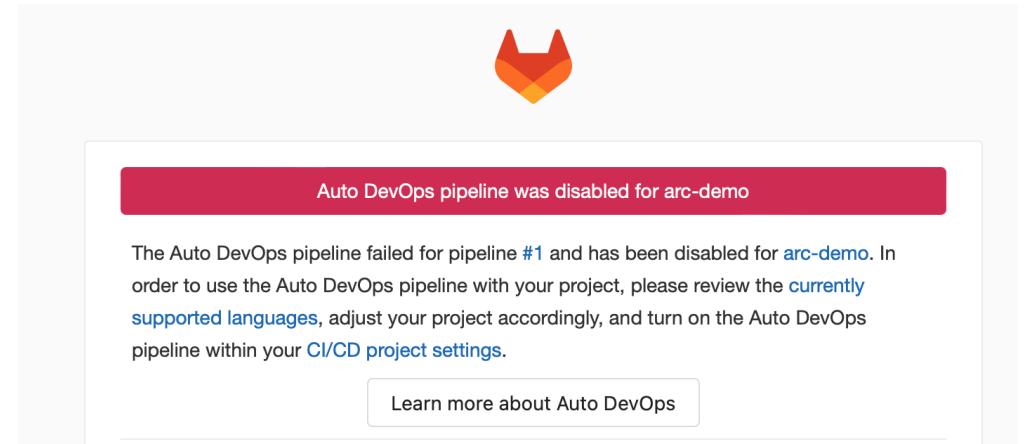
- a "continuous quality control" (CQC) pipeline validates your ARC
- This fails if one of the following metadata items is missing:

```
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



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

Reactivate the CQC pipeline

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

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

The screenshot shows the GitLab CI/CD settings interface. On the left, there is a sidebar with various project management and development tools listed: Security & Compliance, Deployments, Packages and registries, Infrastructure, Monitor, Analytics, Wiki, Snippets, Settings (which is currently selected), General, Integrations, Webhooks, Access Tokens, Repository, Merge requests, CI/CD (which is highlighted in blue), Packages and registries, Monitor, and Usage Quotas. The main content area is titled "Auto DevOps". It contains a sub-section "Deployment strategy" with three radio button options: "Continuous deployment to production" (selected), "Continuous deployment to production using timed incremental rollout", and "Automatic deployment to staging, manual deployment to production". There is also a note about adding a Kubernetes cluster integration or creating an AUTO_DEVOPS_PLATFORM_TARGET CI variable. A "Save changes" button is located at the bottom of this section. Below this, there are sections for "Runners" and "Artifacts", each with an "Expand" button.

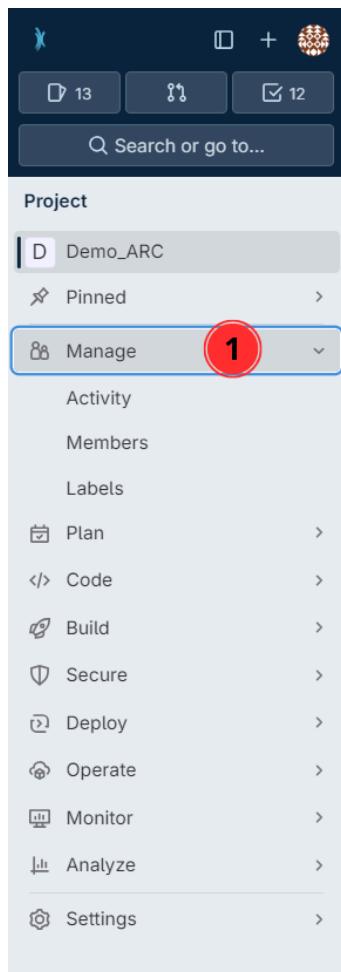
Collaborate and share



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. 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 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 recent commit 'arc init' by 'Demo User' from 1 week ago. A table lists files and their last commits. To the right, a 'Project information' section provides metrics like 1 Commit, 2 Branches, 0 Tags, and 21 KiB Project Storage, along with links to enable Auto DevOps and add various project files.

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 'Members' item highlighted with a red circle labeled '2'. The main area shows the 'Project members' section with one member listed. The member 'Demo User' (@DemoUser) is shown as a direct member by 'Demo User'. The member has an expiration date of Sep 27, 2023, and activity dates from Jul 13, 2024, to Jul 21, 2024. There are buttons for 'Import from a project', 'Invite a group', and 'Invite members'.

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 Sep 27, 2023 ✓ Jul 13, 2024 ✗ Jul 21, 2024	⋮

3. Click on Invite members

The screenshot shows the 'Members' page for the 'Demo_ARC' project. The left sidebar has a 'Members' item highlighted with a red circle labeled '2'. The main area shows a table of members with one entry. The 'Invite members' button in the top right is also highlighted with a red circle labeled '3'.

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"/>	8+ Sep 27, 2023 ✓ Jul 13, 2024 ✗ Jul 21, 2024

4. Search for potential collaborators

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](#)

5. Select a role

A screenshot of a user interface showing a list of roles for a project. The 'Guest' role is selected and highlighted with a blue border. A red circle with the number '4' is positioned near the top right of the list area. A red circle with the number '5' is positioned near the bottom right of the list area. The list includes:

- Guest (selected)
- Reporter
- Developer
- Maintainer
- Owner

Below the list is a dropdown menu set to 'Guest'. At the bottom of the screen, there is a link to 'Read more about role permissions'.

ARC project.

ses

4

5

Guest

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.

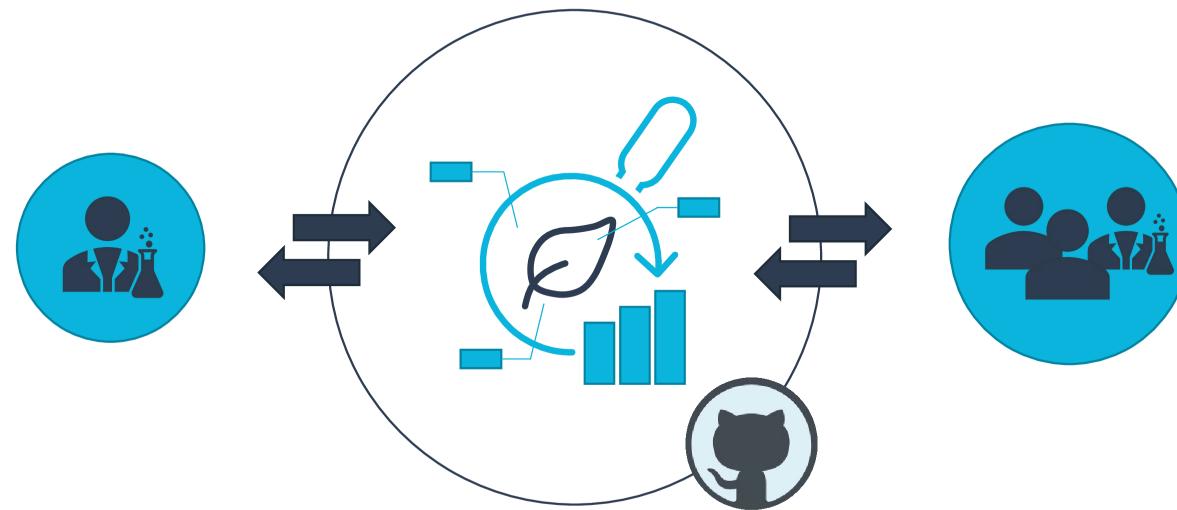
Maintainers

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](#)

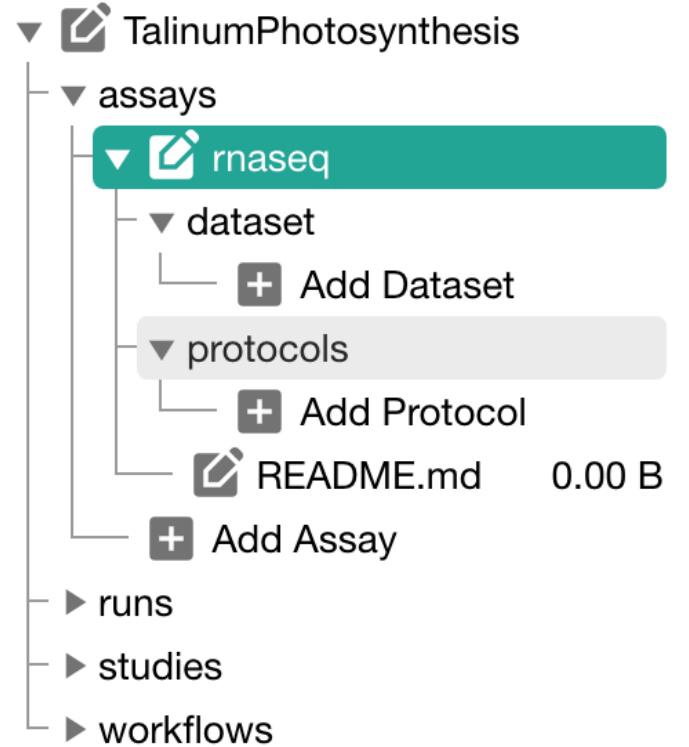
Congratulations!

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 GitLab interface with a sidebar and a main content area.

File Tree (Left):

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

History Panel (Right):

Update
Commit changes and upload ARC

History
Inspect ARC history

- 05.10.2023 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

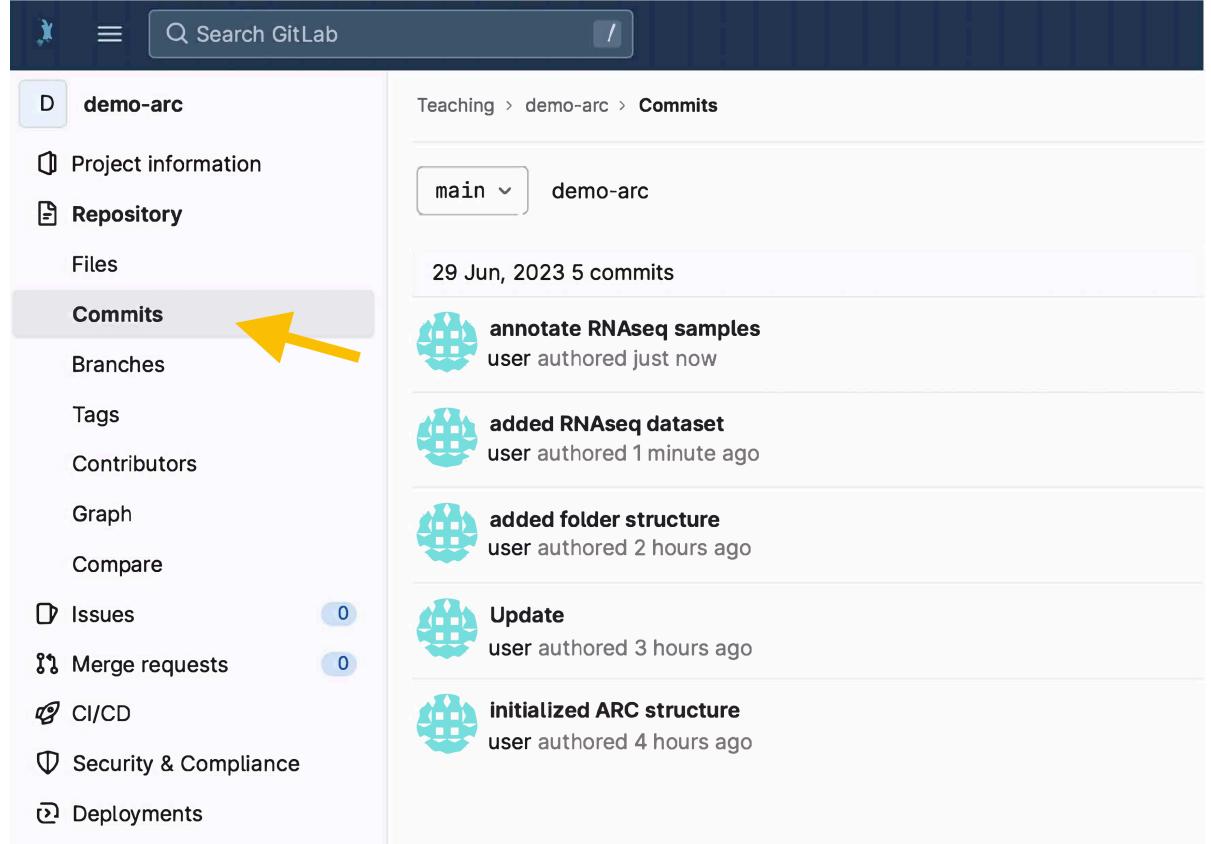
Deleting an ARC

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 settings sections, with 'General' highlighted by a yellow circle labeled '1'. The main content area is titled 'Naming, topics, avatar'. It includes fields for 'Project name' (set to 'Demo_ARC'), 'Project ID' (set to '1584'), and 'Topics' (a search bar). Below these are sections for 'Project description (optional)' and 'Project avatar', which currently shows a placeholder image 'D' and a file upload button. A 'Save changes' button is located at the bottom of this section. Further down, there are sections for 'Visibility, project features, permissions', 'Badges', and 'Service Desk', each with an 'Expand' button. At the very bottom of the page, a section is labeled '4 Advanced' with a yellow circle labeled '4', followed by the text '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

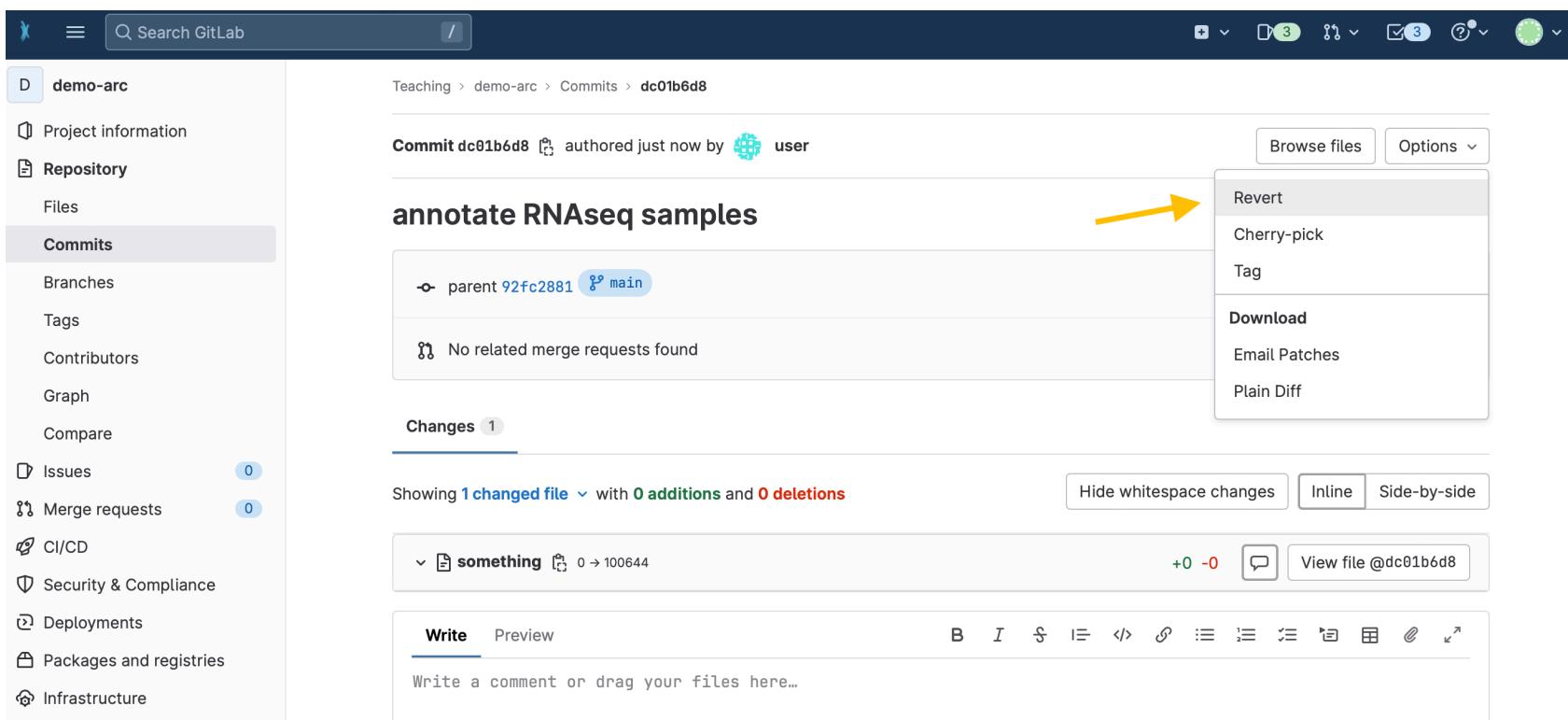


The screenshot shows the GitLab interface for the project 'demo-arc'. The sidebar on the left lists various repository management options: Project information, Repository, Files, **Commits** (which is highlighted with a yellow arrow), Branches, Tags, Contributors, Graph, Compare, Issues (0), Merge requests (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. Each commit is represented by a teal circular icon with a white globe symbol, followed by the commit message, the author's name, and the time it was authored.

Commit Details	Author	Time
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

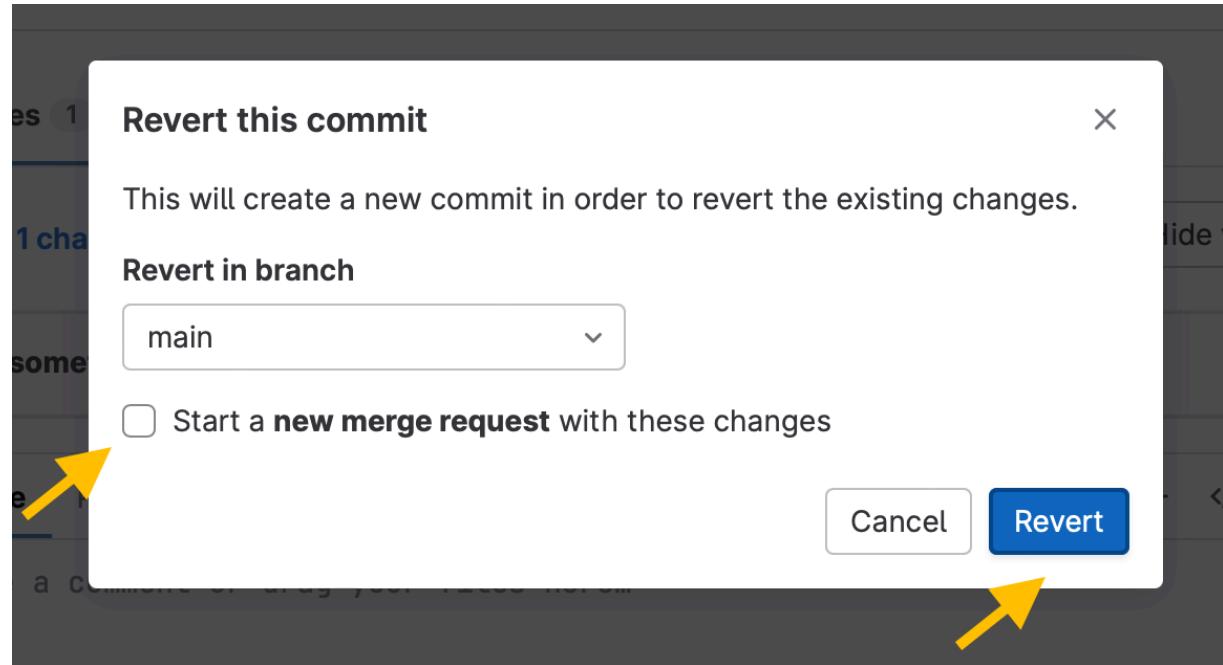
Undo latest changes

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

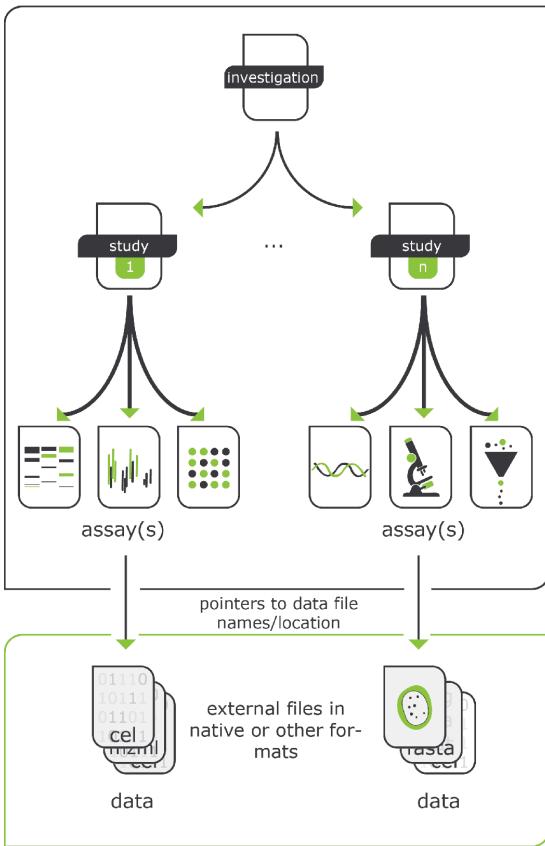
Reusable

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

Accessible

- information about origin
- location of storage
- access rights

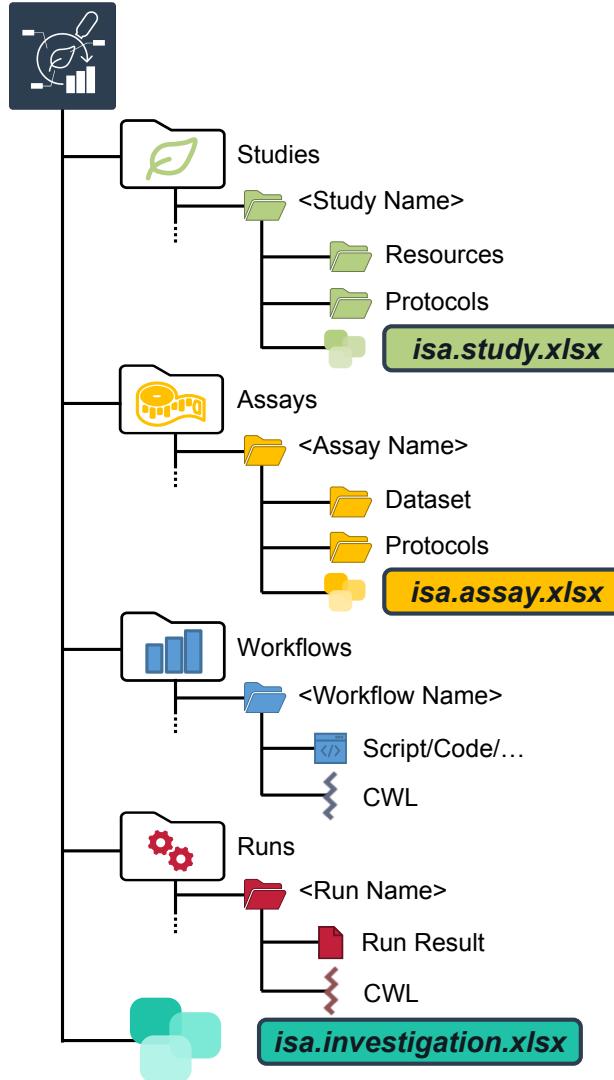
ARC builds on ISA



Investigation
Overall goals
Scientific context

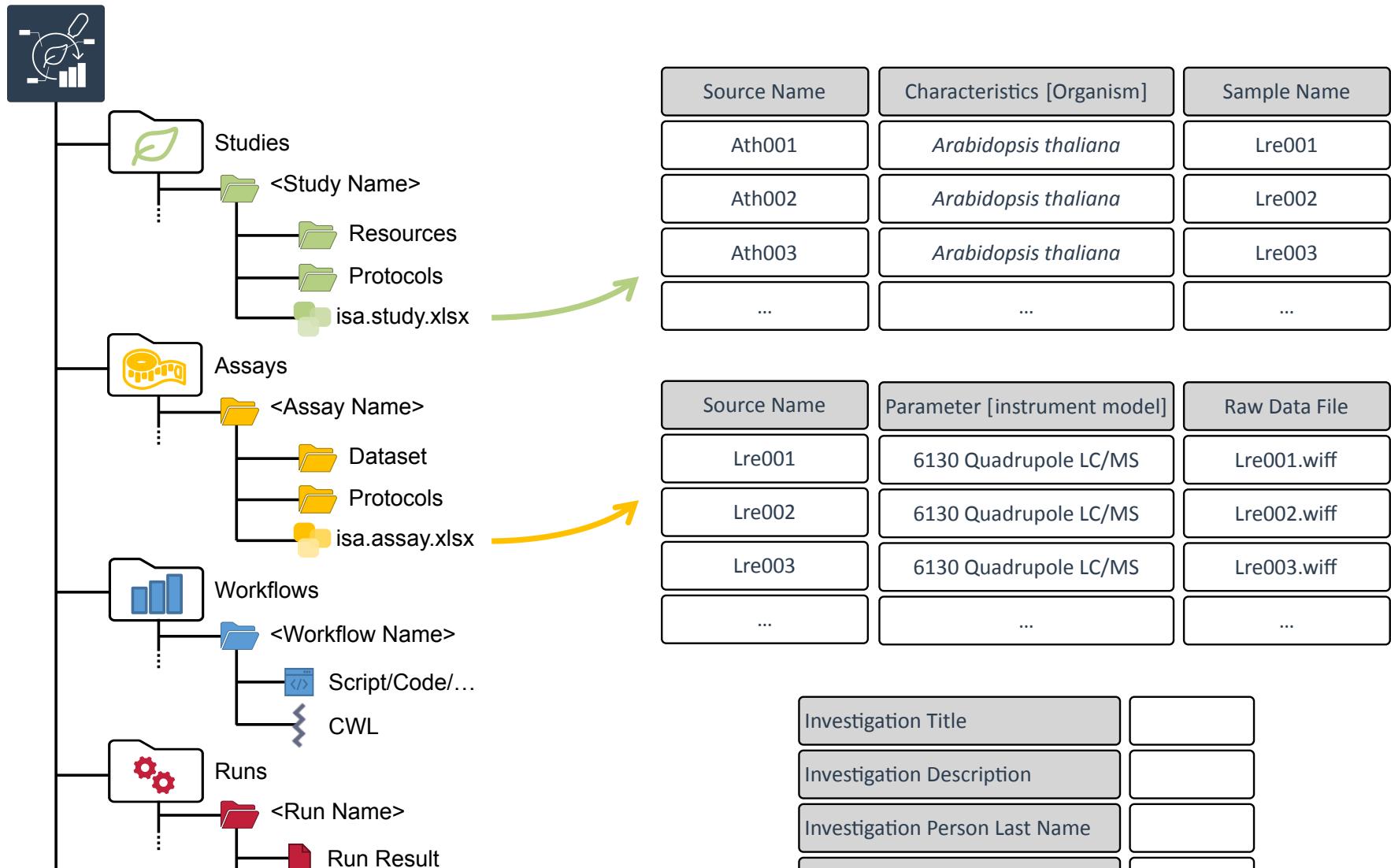
Study
Experimental design

Assay
Leading to (raw) data



ARC builds on ISA

Metadata Annotations



isa.<>.xlsx files within ARCs

ISA INVESTIGATION.XLSX									
1	ONTOLOGY SOURCE REFERENCE	OB	BTO	NEWT	UO	CHEBI	PATO	EFO	
2	Term Source Name								
3	Term Source File	http://biportal.ncbi.nlm.nih.gov/ArrayExpress/Experimental_Factor_Ontology							
4	Term Source Version	47993_v1.26	v.1.26	v.1.26	v.1.26	v.1.26	v.1.26	v.1.26	
5	Term Source Description	Ontology for Biomed BRENDA/tissue / NEWT/UniProt Tax Unit Ontology Chemical Ent Pheno/phenotypic c ArrayExpress Experimental Factor Ontology							
6	INVESTIGATION								
7	Investigation Identifier	BI-1							
8	Investigation Title	Growth control of the yeast cell: a systems biology study in yeast							
9	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.							
10	Investigation Submission Date	30.04.07							
11	Investigation Public Release Date	10.03.09							
12	Comment[:Created With Configuration]								
13	Comment[:Last Opened With Configuration]	isaconfig default_v2013_02_13							
14	INVESTIGATION PUBLICATIONS								
15	Investigation PubMed ID	17439666							
16	Investigation Publication DOI	doi:10.1186/jlife54							
17	Investigation Publication Author List	Castro I, Zee LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst							
18	Investigation Publication Title	Growth control of the yeast cell: a systems biology study in yeast.							
19	Investigation Publication Status	Published							
20	Investigation Publication Status Term Accession Number								
21	Investigation Publication Status Term Source REF								
22	INVESTIGATION CONTACTS								
23	Investigation Person Last Name	Stephen	Castillo	Zeeb					
24	Investigation Person First Name	Oliver	Juan	Lao					
25	Investigation Person Mid Initials	G	I	A					
26	Investigation Person Email								
27	Investigation Person Home Phone								
28	Investigation Person Work Phone								
29	Investigation Person Address	Oxford Road, Manchester Oxford Road, M6 Oxford Road, Manchester M13 9PT, UK							
30	Investigation Person Affiliation	Faculty of Life Sciences Faculty of Life Sciences Michael Smith Building, University of Manchester corresponding auth. author							
31	Investigation Person Roles								
32	Investigation Person Roles Term Accession Number								
33	Investigation Person Roles Term Source REF								
34	Comment[:Investigation Person Ref:]								
35	STUDY								
36	Study Identifier	BI-1							
37	Study Title	Study of the impact of changes in flux on the transcriptome, proteome, endometabolome and exometabolome of the yeast Saccharomyces cerevisiae							
38	Study Description	We wished to study the impact of growth rate on the total complement of mRNA molecules, proteins, and metabolites in S. cerevisiae, independent of growth rate.							
39	Comment[:Study Grant Number:]								
40	Comment[:Study Funding Agency:]								
41	Study Submission Date	30.04.07							
42	Study Public Release Date	10.03.09							
43	Study File Name	s_BI-1.txt							
44	STUDY DESIGN DESCRIPTORS								
45	Study Design Type	Intervention design							
46	Study Design Type Term Accession Number								
47	Study Design Type Term Source REF								
48	STUDY CONDITIONS								
49	Study Published Date								
50	Study Publication DOI	17439666							
51	Study Publication Author List	Castro I, Zee LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst							
52	Study Publication Title	Growth control of the yeast cell: a systems biology study in yeast.							
53	Study Publication Status	published							
54	Study Publication Status Term Accession Number								
55	Study Publication Status Term Source REF								
56	STUDY FACTORS								
57	Study Factor Name	limiting nutrient	rate						
58	Study Factor Type	chemical compound	rate						
59	Study Factor Type Term Accession Number								
60	Study Factor Type Term Source REF								
61	STUDY ASSAYS								
62	Study Assay Measurement Type	protein expression	a metabolic profile	transcription profiling					
63	Study Assay Measurement Type Term Accession Number								
64	Study Assay Measurement Type Term Source REF								
65	Study Assay Technology Type	mass spectrometry	mass spectrometry	DNA microarray					
66	Study Assay Technology Type Term Accession Number								
67	Study Assay Technology Type Term Source REF								
68	Study Assay Technology Platform	OB	OB	OB					
69	Study Assay Platform File Name	iTRAQ	LC-MS/MS	Affymetrix					
70	Study Assay Platform ID	a_proteome.txt	a_metabolome.txt	a_transcriptome.txt					
71	Study Protocol Name	growth protocol	mRNA extraction	protein extraction	biotin label	iTRAQ labeling	EukGE WS4	metabolite extraction	
72	Study Protocol Type	growth	mRNA extraction	protein extraction	labeling	labeling	hybridization	extraction	
73	Study Protocol Type Term Accession Number								
74	Study Protocol Type Term Source REF								
75	Study Protocol Description	1. Biomass samples (1. Biomass samples (45 ml) were taken. This was done using Enzo							
76	Study Protocol URI								

A	B	C	D	E	F	G
1	Source Name	Characteristics [soluble protein content]	Parameter [Quantification method#2]	Parameter [1S1N 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_S32_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr2_5_1
7	G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr2_5_2
8	G2_S32_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr2_5_3
9	G2_S32_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_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr2_5F_1
15	G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr2_5F_2
16	G2_S32_F3_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr2_5F_3
17	G2_S32_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_S32_15mL	50,00 microgram	absolute quantitation analysis		0,75 microgram	WCGr1_5_1
23	G1_S32_15mL	50,00 microgram	absolute quantitation analysis		0,15 microgram	WCGr1_5_2
24	G1_S32_15mL	50,00 microgram	absolute quantitation analysis		0,03 microgram	WCGr1_5_3

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Source Name	Parameter [sample volume]	Parameter [injection vol]	Parameter [measurement model]	Parameter [measurement duration#4]	Raw Data File														
2	WCGr2_U1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_U1.wiff
3	WCGr2_U2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_U2.wiff
4	WCGr2_U3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_U3.wiff
5	WCGr2_U4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_U4.wiff
7	WCGr2_S_1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_S_1.wiff
8	WCGr2_S_2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_S_2.wiff
8	WCGr2_S_3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_S_3.wiff
9	WCGr2_S_4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_S_4.wiff
10	WCGr2_UF_1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_UF_1.wiff
11	WCGr2_UF_2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_UF_2.wiff
12	WCGr2_UF_3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_UF_3.wiff
13	WCGr2_UF_4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_UF_4.wiff
14	WCGr2_SF_1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_SF_1.wiff
15	WCGr2_SF_2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_SF_2.wiff
16	WCGr2_SF_3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_SF_3.wiff
17	WCGr2_SF_4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr2_SF_4.wiff
18	WCGr1_U1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_U1.wiff
19	WCGr1_U2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_U2.wiff
20	WCGr1_U3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_U3.wiff
21	WCGr1_U4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_U4.wiff
22	WCGr1_S_1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_S_1.wiff
23	WCGr1_S_2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_S_2.wiff
24	WCGr1_S_3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_S_3.wiff
25	WCGr1_S_4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_S_4.wiff
26	WCGr1_UF_1	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_UF_1.wiff
27	WCGr1_UF_2	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_UF_2.wiff
28	WCGr1_UF_3	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_UF_3.wiff
29	WCGr1_UF_4	100,00	microliter	2,50	microliter	1	TripleTOF	6600									87,00	minute	WCGr1_UF_4.wiff

Study and assay files are registered in the investigation file

isa.investigation.xlsx

Study file

Assay file

isa.study.xlsx

isa.assay.xlsx

Column	Characteristic [soluble protein content]	Parameter [Quantification method#2]	Parameter [15N Photosynthesis QconCAT mass#4]	Sample Name
1	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,75 microgram WGr2_U1	
2	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,15 microgram WGr2_U2	
3	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,03 microgram WGr2_U3	
4	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,01 microgram WGr2_U4	
5	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,75 microgram WGr2_5_1	
6	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,15 microgram WGr2_5_2	
7	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,03 microgram WGr2_5_3	
8	G2_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,01 microgram WGr2_5_4	
9	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,75 microgram WGr2_UF_1	
10	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,15 microgram WGr2_UF_2	
11	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,03 microgram WGr2_UF_3	
12	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,01 microgram WGr2_UF_4	
13	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,75 microgram WGr2_5_F_1	
14	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,15 microgram WGr2_5_F_2	
15	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,03 microgram WGr2_5_F_3	
16	G2_UVM4_F3_15mL	50,00 microgram absolute quantitation analysis	0,01 microgram WGr2_5_F_4	
17	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,75 microgram WGr1_U1	
18	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,15 microgram WGr1_U2	
19	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,03 microgram WGr1_U3	
20	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,01 microgram WGr1_U4	
21	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,75 microgram WGr1_5_1	
22	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,15 microgram WGr1_5_2	
23	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis	0,03 microgram WGr1_5_3	
24	G1_UVM4_15mL	50,00 microgram absolute quantitation analysis		

Column	Parameter [sample volume]	Parameter [injection volume]	Parameter [measurement model#4]	Raw Data File
1	WGr2_U1	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_U1.wiff
2	WGr2_U2	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_U2.wiff
3	WGr2_U3	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_U3.wiff
4	WGr2_U4	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_U4.wiff
5	WGr2_S_1	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_S_1.wiff
6	WGr2_S_2	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_S_2.wiff
7	WGr2_S_3	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_S_3.wiff
8	WGr2_S_4	100,00 microliter	1 TripleTOF 6000	87,00 minute WGr2_S_4.wiff
9	WGr2_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6000
10	WGr2_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6000
11	WGr2_UF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6000
12	WGr2_UF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6000
13	WGr2_SF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6000
14	WGr2_SF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6000
15	WGr2_SF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6000
16	WGr2_SF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6000
17	WGr1_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6000
18	WGr1_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6000
19	WGr1_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6000
20	WGr1_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6000
21	WGr1_U5	100,00 microliter	2,50 microliter	1 TripleTOF 6000
22	WGr1_U6	100,00 microliter	2,50 microliter	1 TripleTOF 6000
23	WGr1_U7	100,00 microliter	2,50 microliter	1 TripleTOF 6000
24	WGr1_U8	100,00 microliter	2,50 microliter	1 TripleTOF 6000
25	WGr1_U9	100,00 microliter	2,50 microliter	1 TripleTOF 6000
26	WGr1_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6000
27	WGr1_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6000

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

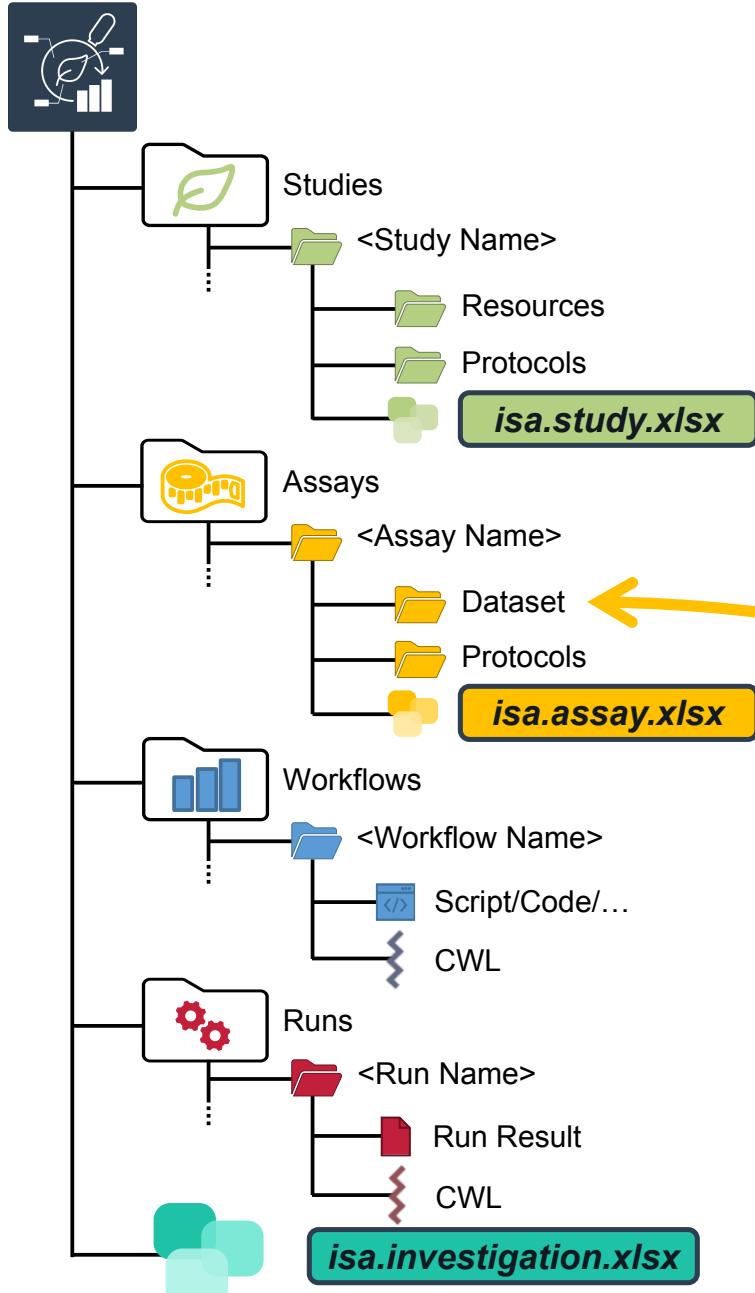
A	B	C	D	E	F	G	H	I	J	K	L	M
1	Source Name	Characteristics [soluble protein content]	Parameter [Quantification method#2]	Parameter [15N Photosynthesis QconCAT mass#4]	Sample Name							
2	G2_UVM4_15mL	50,00 microgram	absolute quantitation analysis			0,75 microgram	WCGr2_U1					
3	G2_UVM4_15mL	50,00 microgram	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_SF_1					
15	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis			0,15 microgram	WCGr2_SF_2					
16	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis			0,03 microgram	WCGr2_SF_3					
17	G2_532_F3_15mL	50,00 microgram	absolute quantitation analysis			0,01 microgram	WCGr2_SF_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.study.xlsx

Samples

A	B	C	D	E	F	G	H	I	J	K	L	M
1	Source Name	Parameter [sample volume]	Parameter [injection volu									
2	WCGr2_U1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
3	WCGr2_U2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
4	WCGr2_U3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
5	WCGr2_U4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
6	WCGr2_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
7	WCGr2_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
8	WCGr2_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
9	WCGr2_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
10	WCGr2_UF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
11	WCGr2_UF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
12	WCGr2_UF_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
13	WCGr2_UF_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
14	WCGr2_SF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
15	WCGr2_SF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
16	WCGr2_SF_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
17	WCGr2_SF_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
18	WCGr1_U1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
19	WCGr1_U2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
20	WCGr1_U3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
21	WCGr1_U4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
22	WCGr1_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
23	WCGr1_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
24	WCGr1_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
25	WCGr1_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
26	WCGr1_UF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
27	WCGr1_UF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							

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

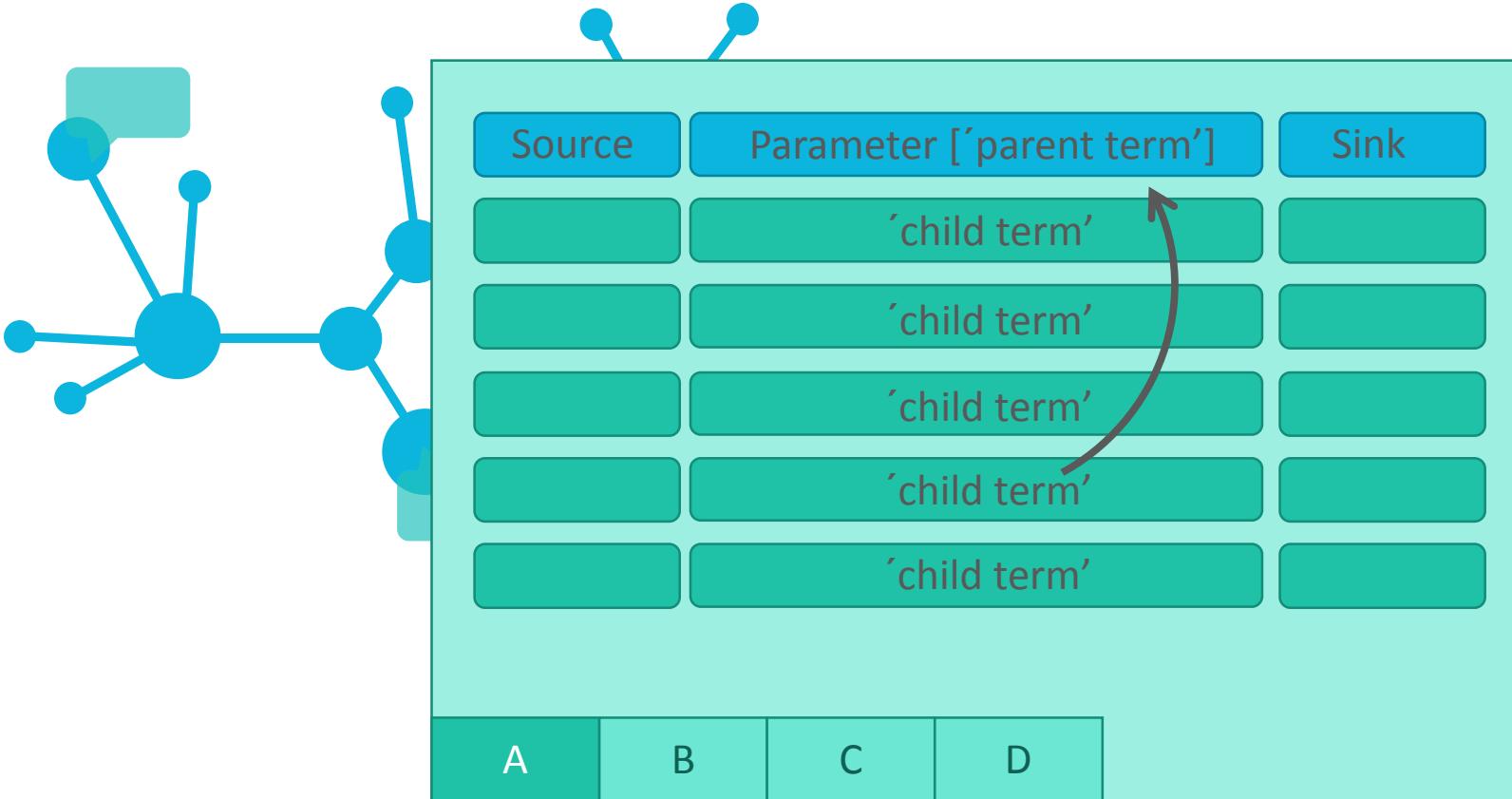
A	B	C	D	E	F	G	H	I	J	K	L	M
1	Source Name	Parameter [sample volume]	Parameter [injection vol.									
2	WCGr2_U1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
3	WCGr2_U2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
4	WCGr2_U3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
5	WCGr2_U4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
6	WCGr2_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
7	WCGr2_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
8	WCGr2_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
9	WCGr2_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
10	WCGr2_UF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
11	WCGr2_UF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
12	WCGr2_UF_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
13	WCGr2_UF_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
14	WCGr2_SF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
15	WCGr2_SF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
16	WCGr2_SF_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
17	WCGr2_SF_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
18	WCGr1_U1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
19	WCGr1_U2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
20	WCGr1_U3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
21	WCGr1_U4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
22	WCGr1_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
23	WCGr1_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
24	WCGr1_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
25	WCGr1_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
26	WCGr1_UF_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600							
27	WCGr1_UF_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600							

Raw data

The raw data consists of WIFF files for each run, corresponding to the entries in the assay table. The files are named WCGr2_5_1.wiff, WCGr2_5_2.wiff, WCGr2_5_3.wiff, WCGr2_5_4.wiff, WCGr1_U1.wiff, WCGr1_U2.wiff, WCGr1_U3.wiff, WCGr1_U4.wiff, WCGr1_5_1.wiff, WCGr1_5_2.wiff, WCGr1_5_3.wiff, WCGr1_5_4.wiff, WCGr1_UF_1.wiff, WCGr1_UF_2.wiff, WCGr1_UF_3.wiff, and WCGr1_UF_4.wiff.

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 "isa.study (1).xlsx" with a data table containing approximately 50 rows of experimental data. The columns include "Source Name", "Protocol Type", "Characteristic [sample label]", "Factor [temperature]", "Parameter [Instrument model]", "Component [Software]", and "Sample Name". A callout bubble points to the "Parameter [Instrument model]" column with the text "New Parameter".

To the right of the Excel window, the "Swate" add-in interface is open. It displays a "Building Blocks" section with a list of instrument models and instruments. A search bar at the bottom says "...Can't find the term you are looking for? Try". Below the search bar, a note states: "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."

Parameter	instrument mod
instrument model	MS:1000031
Instrument Model	NCIT:C177610
instrument	MS:1000463
instrument	EFO:0000548
Agilent instrument model	MS:1000490

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 an Excel spreadsheet titled 'isa.study (1).xlsx' with a data table containing approximately 50 rows of experimental data. The columns include 'Source Name', 'Protocol Type', 'Characteristic [sample label]', 'Factor [temperature]', 'Parameter [instrument model]', 'Component [software]', and 'Sample Name'. The 'Characteristic' column is highlighted with a green background for several rows. The 'Protocol Type' column is also highlighted with a green background. The 'Factor' column is highlighted with a green background. The 'Sample Name' column is highlighted with a green background. A 'Building Blocks' pane is open on the right side of the screen, listing various annotations such as 'Instrument Model', 'Instrument', and 'Agilent instrument model'. A callout labeled 'New Parameter' points to the 'Parameter' column header in the Building Blocks pane. Other callouts point to the 'Characteristic', 'Protocol Type/Protocol Ref', 'Factor', and 'Sample Name/Raw Data File/Derived Data File' columns in the main data table.

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_UV4_15mL", "G2_UV4_15mL", "G2_UV4_15mL", etc., which are used for the search.

A modal window titled "Swate" is open on the right side of the screen, specifically the "Ontology term search" tab. The search bar contains the text "instrument n 6130". Below the search bar, it says "6130 Quadrupole MS:1000470 LC/MS". A message at the bottom of the modal says "Cant find the Term you are looking for? Try Advanced Search!". Another message below that says "Still can't find what you need? Get in contact with us!".

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 selected, showing various data analysis tools like Sort & Filter, Text to Columns, Flash Fill, Remove Duplicates, Consolidate, Relationships, Manage Data Model, What-If Analysis, Forecast Sheet, Group, Ungroup, Subtotal, Outline, and Swate.

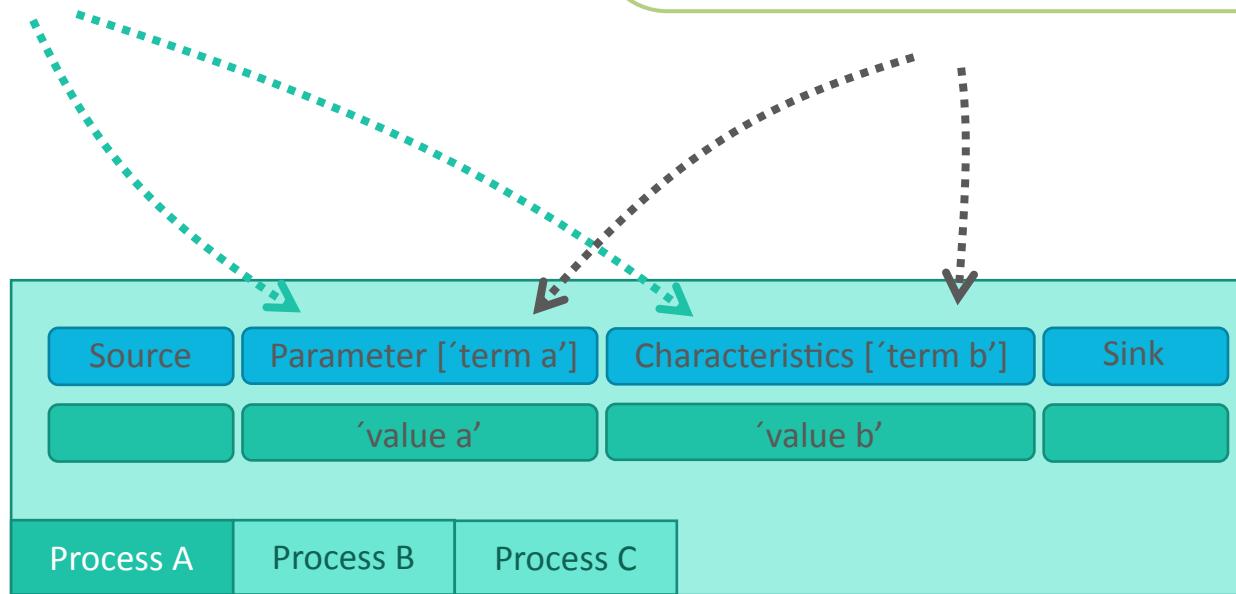
Fill your table with ontology terms

The screenshot shows a Microsoft Excel spreadsheet titled "Sheet1". The table has columns labeled A through AB. The first few rows of data are as follows:

Source Name	Protocol Type	Characteristic [sample label]	Factor [temperature]	Parameter [instrument model]	Component [software]	Sample Name
G2_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_U1
G2_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_U2
G2_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_U3
G2_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_U4
G2_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_S_1
G2_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_S_2
G2_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_S_3
G2_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_S_4
G2_UVMA_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_UF_1
G2_UVMA_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_UF_2
G2_UVMA_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_UF_3
G2_UVMA_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_UF_4
G2_532_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_SF_1
G2_532_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_SF_2
G2_532_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_SF_3
G2_532_F3_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr2_SF_4
G1_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_U1
G1_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_U2
G1_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_U3
G1_UVMA_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_U4
G1_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_S_1
G1_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_S_2
G1_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_S_3
G1_532_15mL	data extraction protocol	15N	30,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_S_4
G1_UVMA_F7_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_UF_1
G1_UVMA_F7_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_UF_2
G1_UVMA_F7_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_UF_3
G1_UVMA_F7_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_UF_4
G1_532_F10_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_SF_1
G1_532_F10_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_SF_2
G1_532_F10_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_SF_3
G1_532_F10_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr1_SF_4
G3_UVMA_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_U1
G3_UVMA_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_U2
G3_UVMA_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_U3
G3_UVMA_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_U4
G3_532_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_S_1
G3_532_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_S_2
G3_532_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_S_3
G3_532_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_S_4
G3_UVMA_F1_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_UF_1
G3_UVMA_F1_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_UF_2
G3_UVMA_F1_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_UF_3
G3_UVMA_F1_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_UF_4
G3_532_F2_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_SF_1
G3_532_F2_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_SF_2
G3_532_F2_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_SF_3
G3_532_F2_15mL	data extraction protocol	15N	4,00 degree Celsius	6130 Quadrupole LC/MS	Analyst	WCGr3_SF_4

A floating window titled "Swate" is open, showing an "Ontology term search" interface. It contains a search bar with the text "6130 Quadrupole LC/MS" and a button "Fill selected cells with this term".

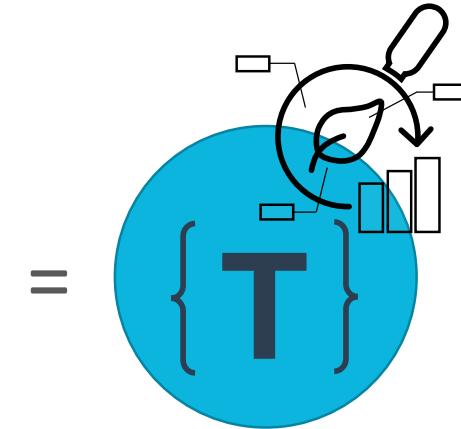
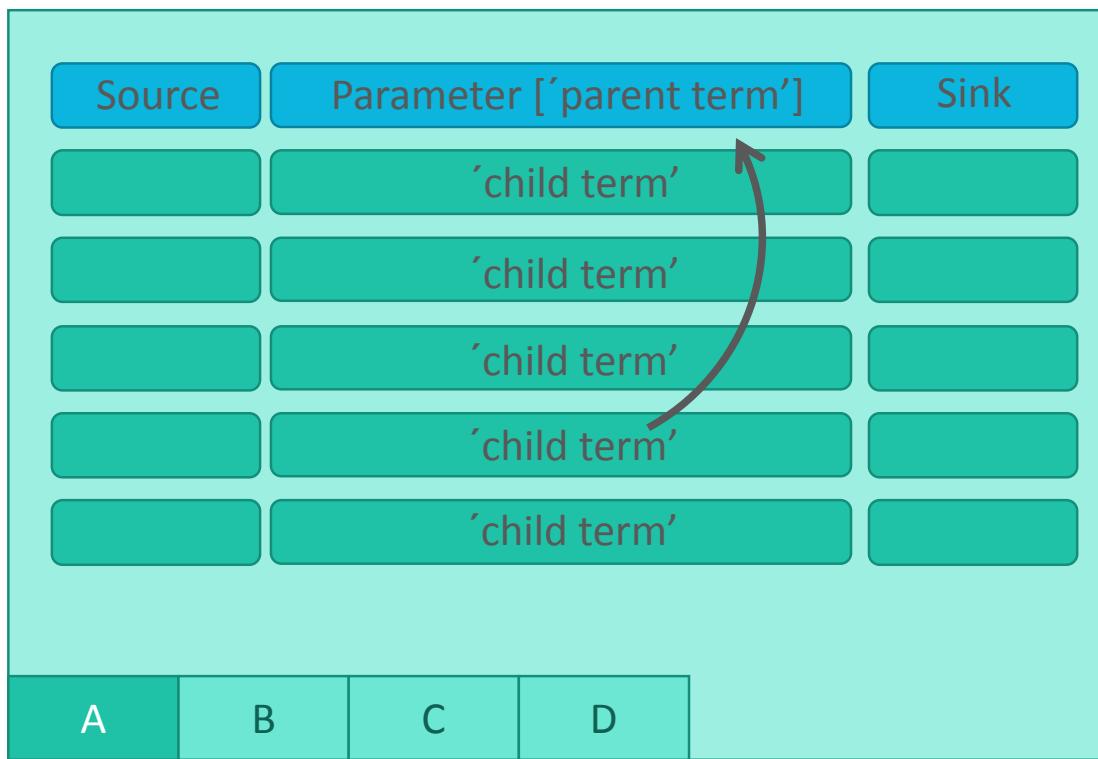
Hierarchical combination of ontologies



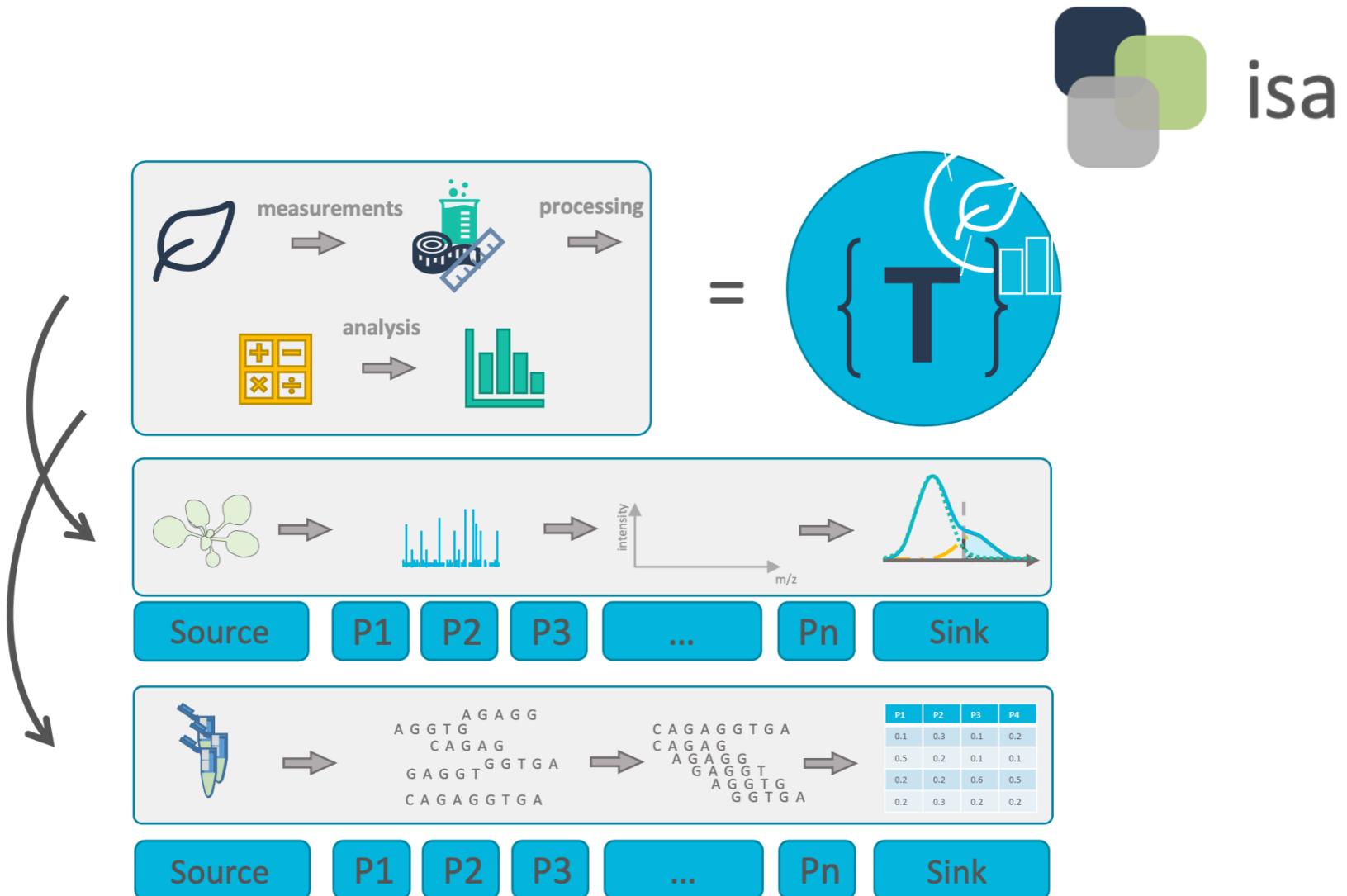
[isa.study.xlsx or isa.assay.xlsx](#)

Swate templates

Checklists and Templates

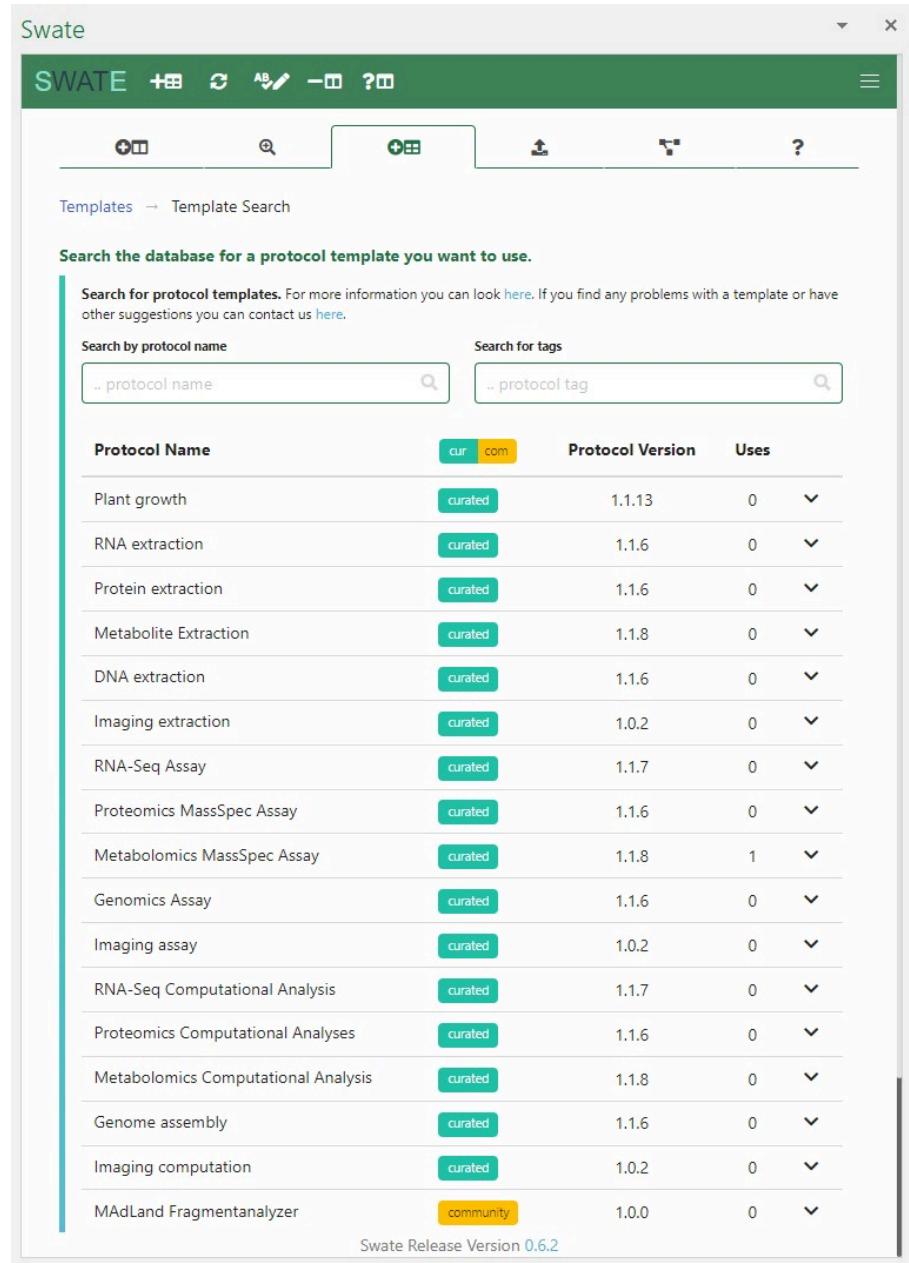


Realization of lab-specific metadata templates



Directly import templates via Swate

- DataPLANT curated
- Community templates



The screenshot shows the Swate software interface. At the top, there's a green header bar with the word "Swate" and various icons. Below the header is a toolbar with buttons for creating new templates, searching, and other functions. The main area is titled "Templates → Template Search" and contains a search bar and a "Search by protocol name" input field. To the right of these fields is a "Search for tags" input field. The main content area displays a table of protocol templates:

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

At the bottom right of the main area, it says "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.

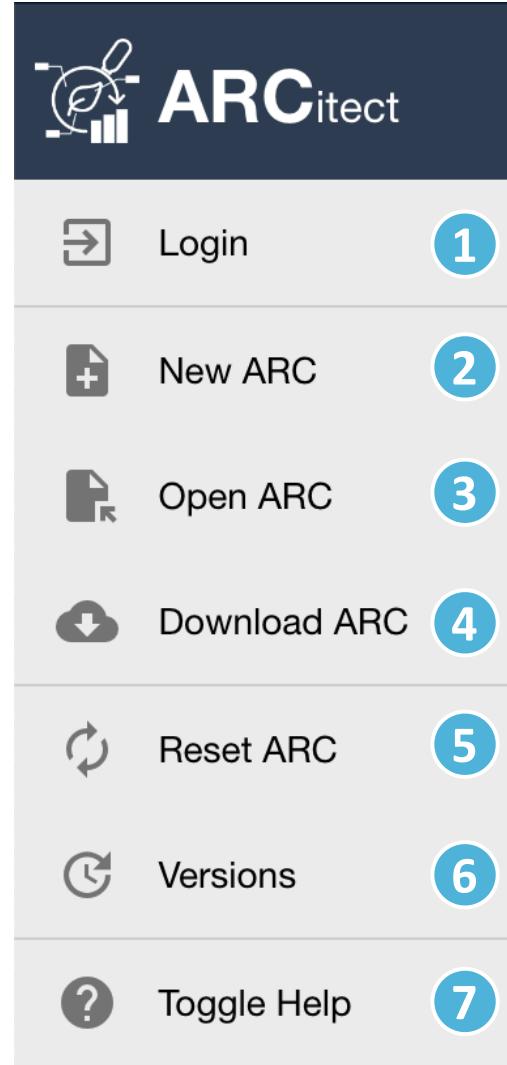
 Alternatively, you can use [Swate standalone](#)

(⚠️ 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](#)



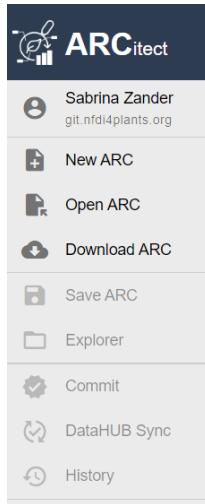
Download the demo data

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.



Open ARC

A screenshot of the "Download ARC" interface. The title bar says "Download ARC" and "Download ARCs from the nfdi4plants DataHUB". The search bar contains "Talinum-CAM-Photosynthesis" with a search icon. Below it, a dropdown shows "Host git.nfdi4plants.org" and a refresh icon. A list item "Talinum-CAM-Photosynthesis [2023-10-11T09:24:10.208Z] Teaching" is shown with a green circular icon containing a white letter "T", a search icon, and a download icon.

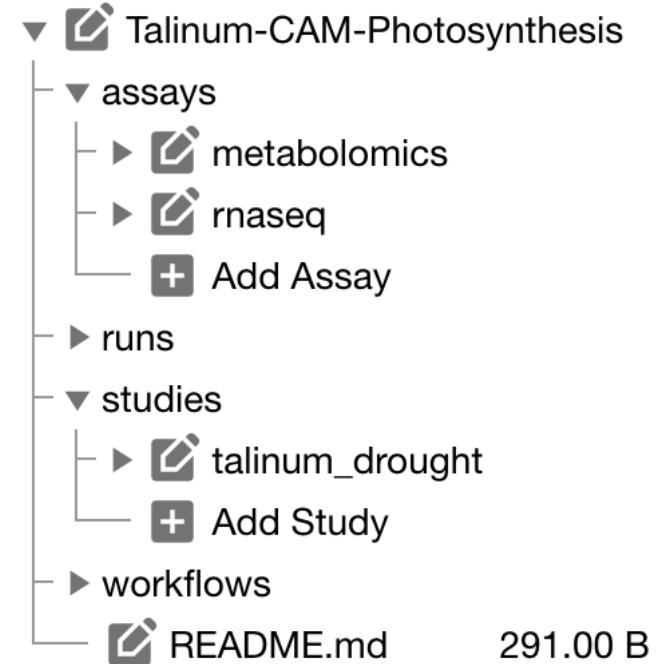
This is basically the ARC we created last session.

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

source

	Characteristics [sample label]	Factor [temperature unit]	Data File Name
1	Source Name	15N	32.00 degree Celsius
2	Heat_15A_OD_R1	15N	32.00 degree Celsius
3	Heat_15A_OD_R2	15N	32.00 degree Celsius
4	Heat_180A_OD_R1	15N	32.00 degree Celsius
5	Heat_180A_OD_R2	15N	32.00 degree Celsius
6	Heat_2880A_OD_R1	15N	32.00 degree Celsius
7	Heat_2880A_OD_R2	15N	32.00 degree Celsius
8	Heat_5760A_OD_R1	15N	32.00 degree Celsius
9	Heat_5760A_OD_R2	15N	32.00 degree Celsius
10	Heat_5760A_15D_R1	15N	32.00 degree Celsius
11	Heat_5760A_15D_R2	15N	32.00 degree Celsius
12	Heat_5760A_180D_R1	15N	32.00 degree Celsius
13	Heat_5760A_180D_R2	15N	32.00 degree Celsius
14	Heat_5760A_2880D_R1	15N	32.00 degree Celsius
15	Heat_5760A_2880D_R2	15N	32.00 degree Celsius
16	Heat_5760A_5760D_R1	15N	32.00 degree Celsius
17	Heat_5760A_5760D_R2	15N	32.00 degree Celsius
18	Cold_15A_OD_R1	15N	4.00 degree Celsius
19	Cold_15A_OD_R2	15N	4.00 degree Celsius
20	Cold_180A_OD_R1	15N	4.00 degree Celsius
21	Cold_180A_OD_R2	15N	4.00 degree Celsius
22	Cold_2880A_OD_R1	15N	4.00 degree Celsius
23	Cold_2880A_OD_R2	15N	4.00 degree Celsius
24	Cold_5760A_OD_R1	15N	4.00 degree Celsius
25	Cold_5760A_OD_R2	15N	4.00 degree Celsius
26	Cold_5760A_15D_R1	15N	4.00 degree Celsius
27	Cold_5760A_15D_R2	15N	4.00 degree Celsius
28	Cold_5760A_180D_R1	15N	4.00 degree Celsius
29	Cold_5760A_180D_R2	15N	4.00 degree Celsius
30	Cold_5760A_2880D_R1	15N	4.00 degree Celsius
31	Cold_5760A_2880D_R2	15N	4.00 degree Celsius
32	Cold_5760A_5760D_R1	15N	4.00 degree Celsius
33	Cold_5760A_5760D_R2	15N	4.00 degree Celsius
34	Highlight_15A_OD_R1	15N	22.00 degree Celsius
35	Highlight_15A_OD_R2	15N	22.00 degree Celsius
36	Highlight_180A_OD_R1	15N	22.00 degree Celsius
37	Highlight_180A_OD_R2	15N	22.00 degree Celsius

new parameter

otation building block selection

parameter

instrument

instrument model

instrument vendor

medical instrument

Mascot:Instrument

datafile / sample

characteristic

factor

More about Parameter:

Use parameters to annotate your experimental workflow. You can group parameters to create a protocol. You can find more information on our website.

Swate Release Version 0.4.7

Let's annotate the plant samples first

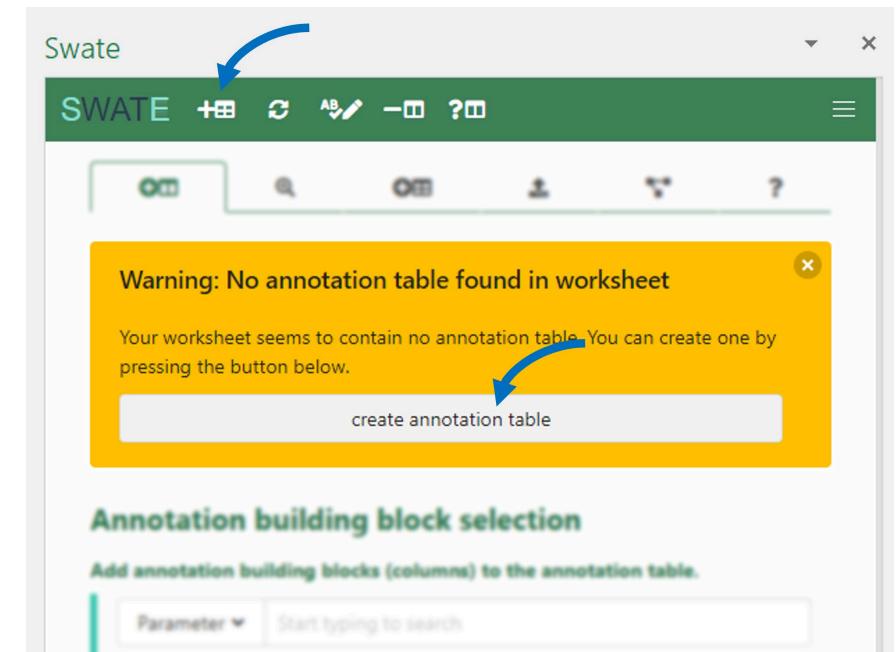
1. Navigate to the demo ARC.
2. Open the lab notes `studies/talinum_drought/protocols/plant_material.txt` in a text editor.
3. Open the empty `studies/talinum_drought/isa.study.xlsx` workbook in Excel.

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

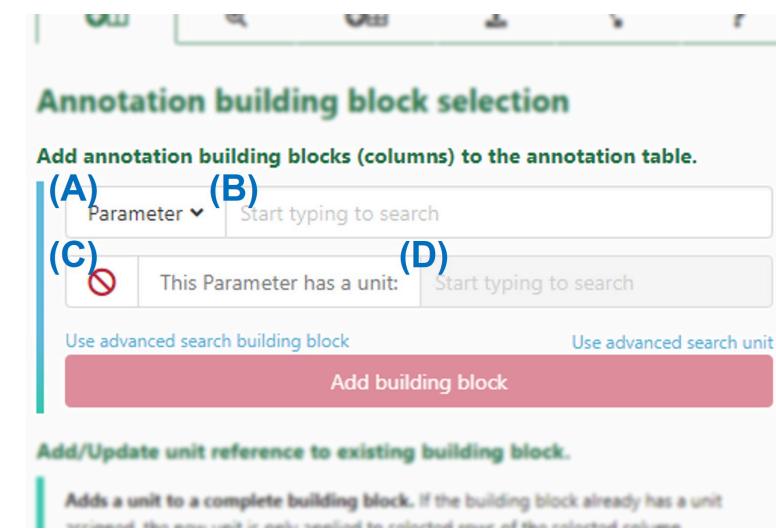
💡 Only one annotation table can be added per Excel sheet



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.

Adds 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 microeinsteins 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:

ISA Table Overview				
	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

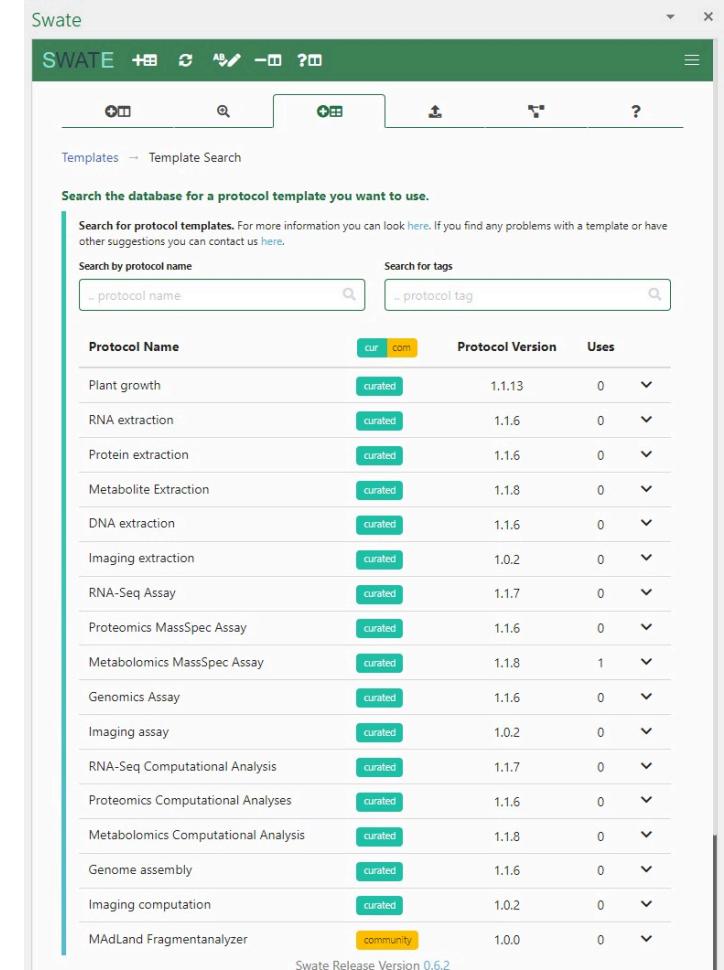
1. Navigate to the demo ARC.
2. Open the lab notes `assays/rnaseq/protocols/RNA_extraction.txt` in a text editor.
3. Open the empty `assays/rnaseq/isa.assay.xlsx` workbook in Excel.

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.



The screenshot shows the 'Template Search' page in the Swate software. The top navigation bar includes 'SWATE' and various icons. Below the header, there are search fields for 'Search by protocol name' and 'Search for tags'. A main table lists protocol templates with columns for 'Protocol Name', 'Protocol Version', and 'Uses'. The table includes entries for 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. A note at the bottom of the table area states: '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](#)'.

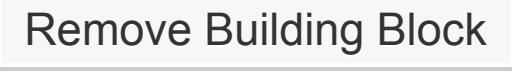
Protocol Name	cur com	Protocol Version	Uses
Plant growth	curated	1.1.13	0
RNA extraction	curated	1.1.6	0
Protein extraction	curated	1.1.6	0
Metabolite Extraction	curated	1.1.8	0
DNA extraction	curated	1.1.6	0
Imaging extraction	curated	1.0.2	0
RNA-Seq Assay	curated	1.1.7	0
Proteomics MassSpec Assay	curated	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

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")
```

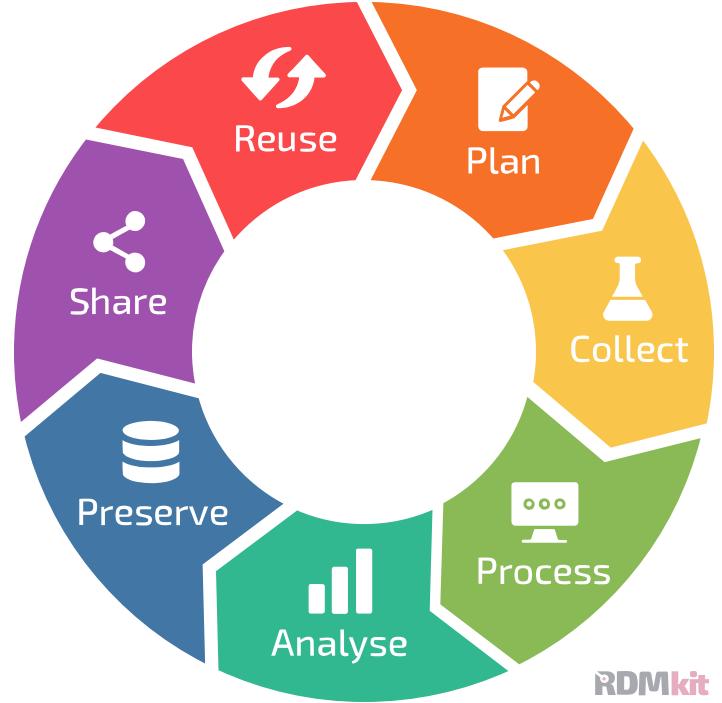
Annotation with ARCitect

 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

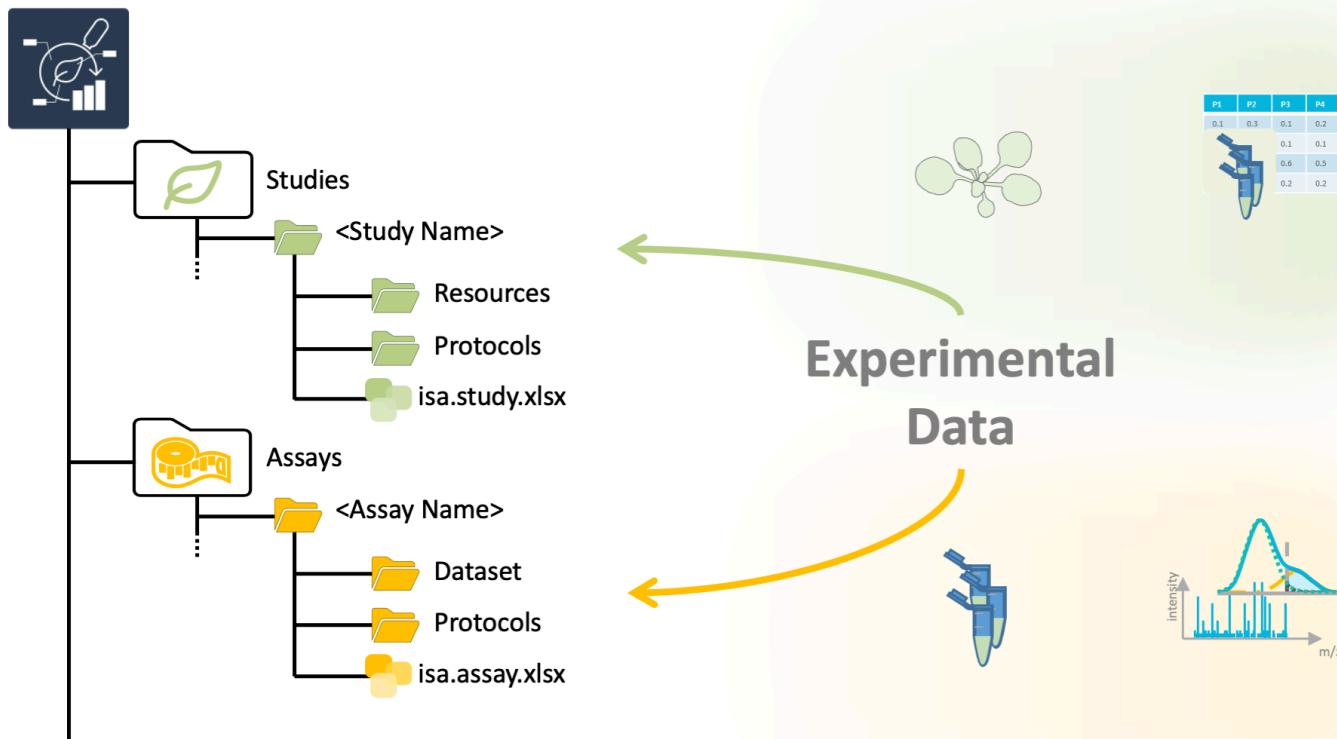
"A FAIR RDM journey along a (mutable) data life cycle"

October 11th, 2023

Dominik Brilhaus, [CEPLAS Data Science](#)

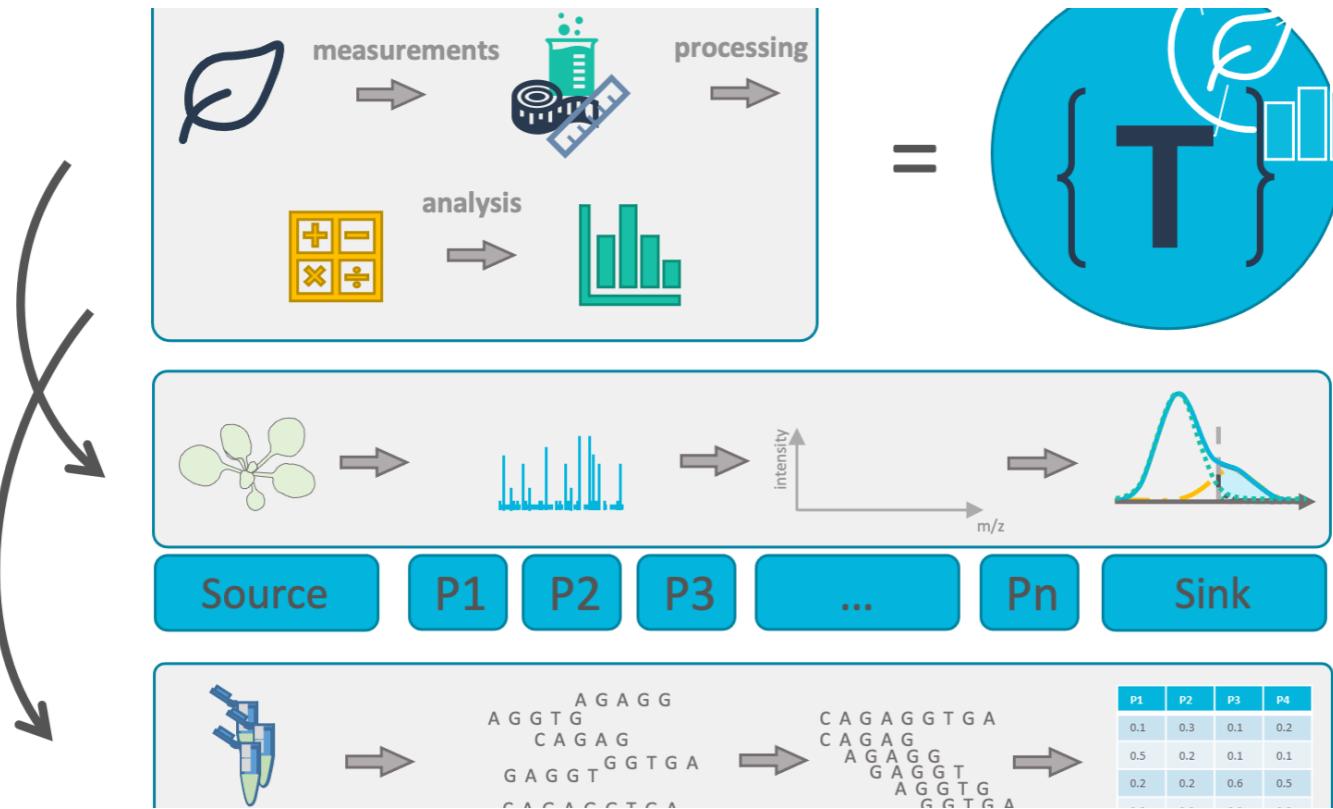


Collect



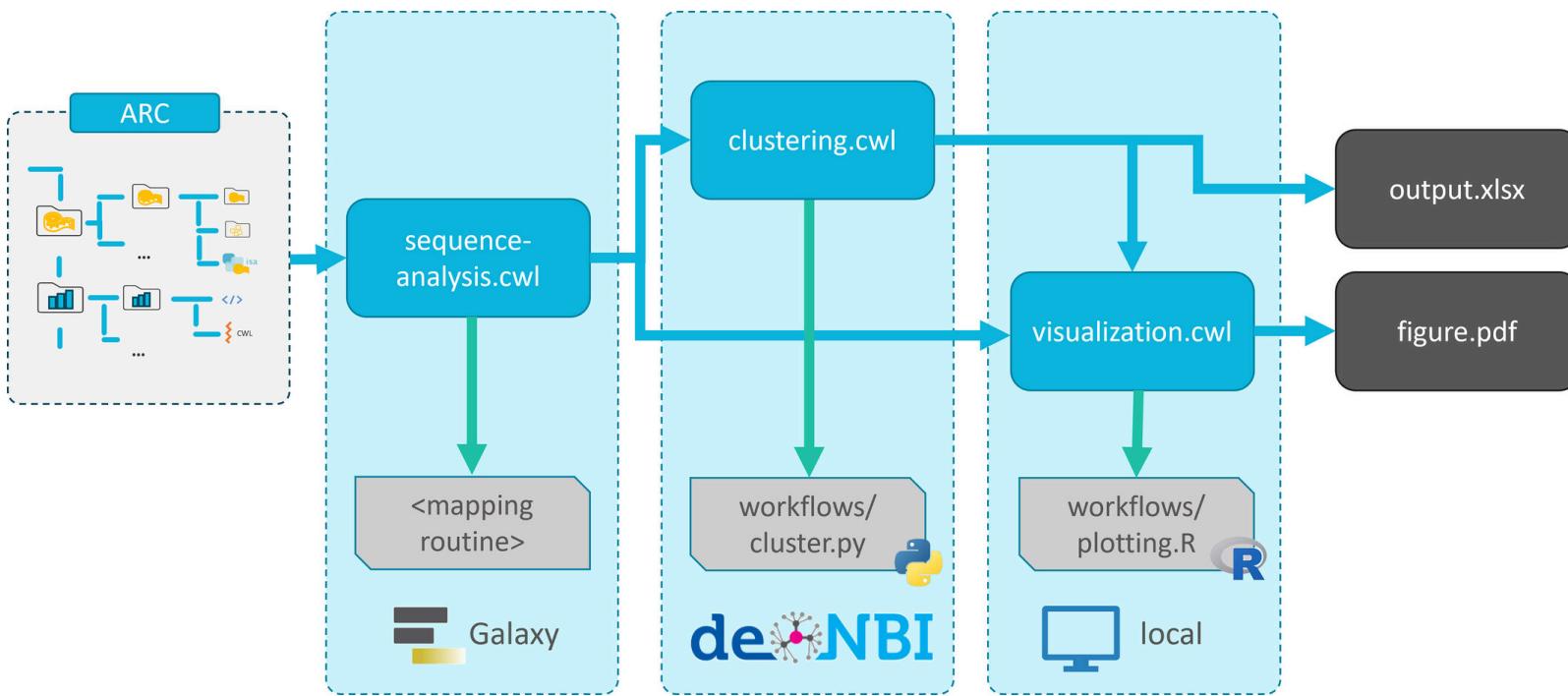


Process (e.g. annotate)



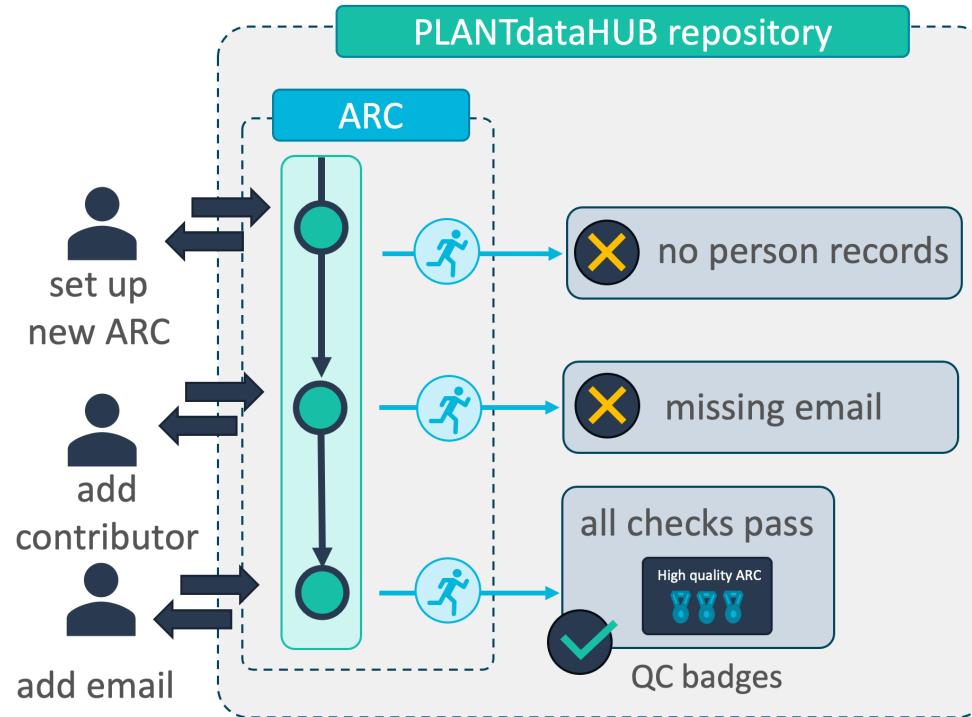


Analyse





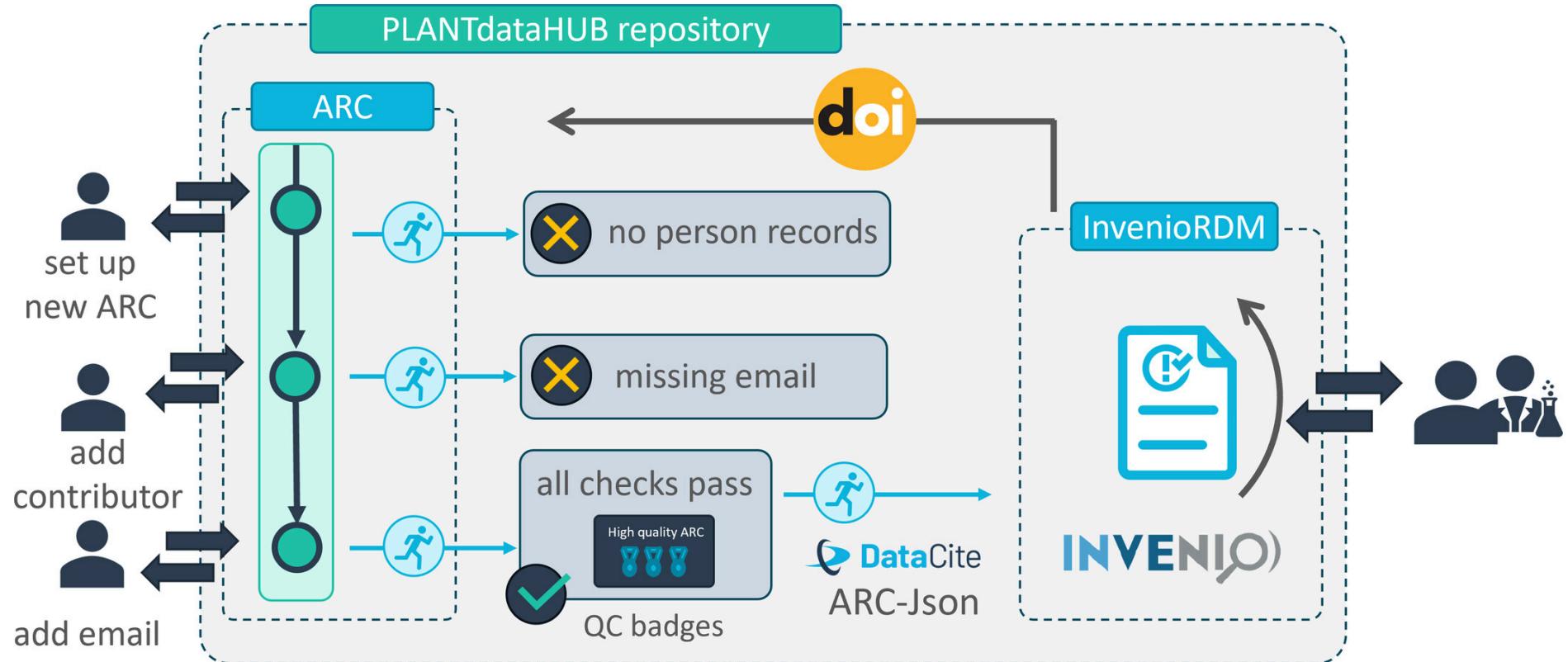
Preserve



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



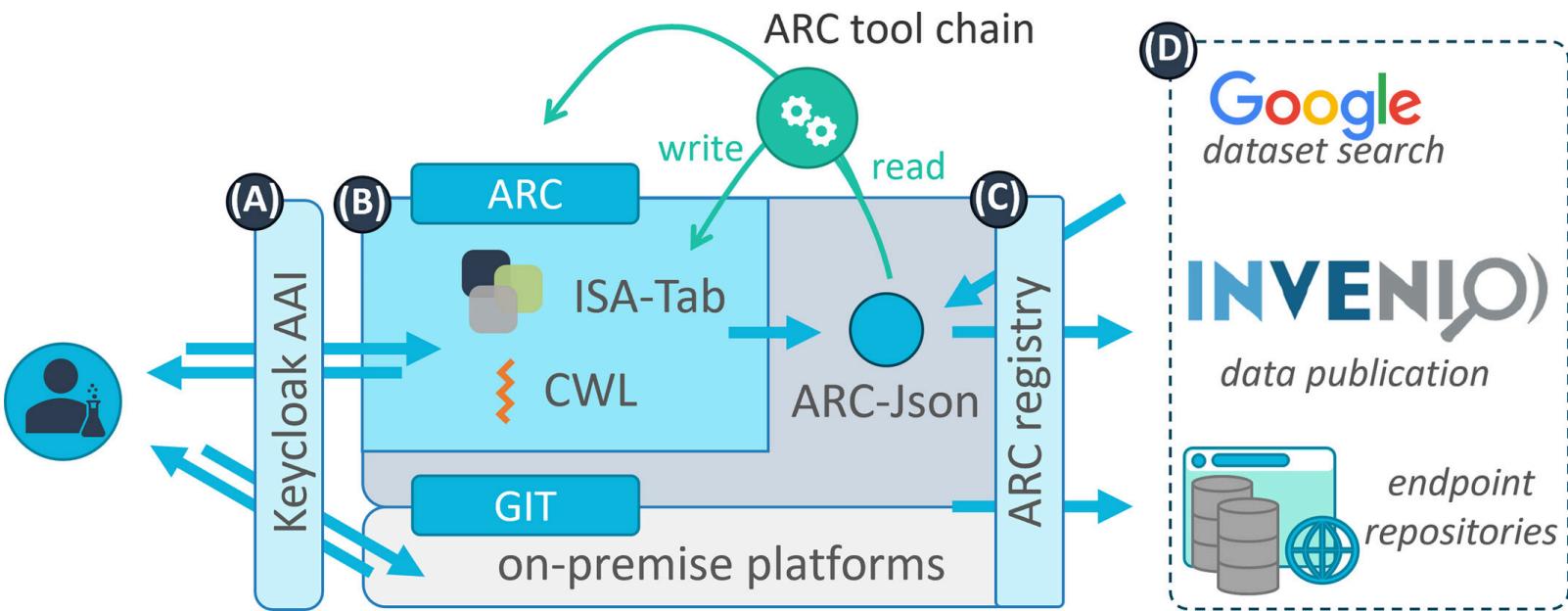


Share and collaborate

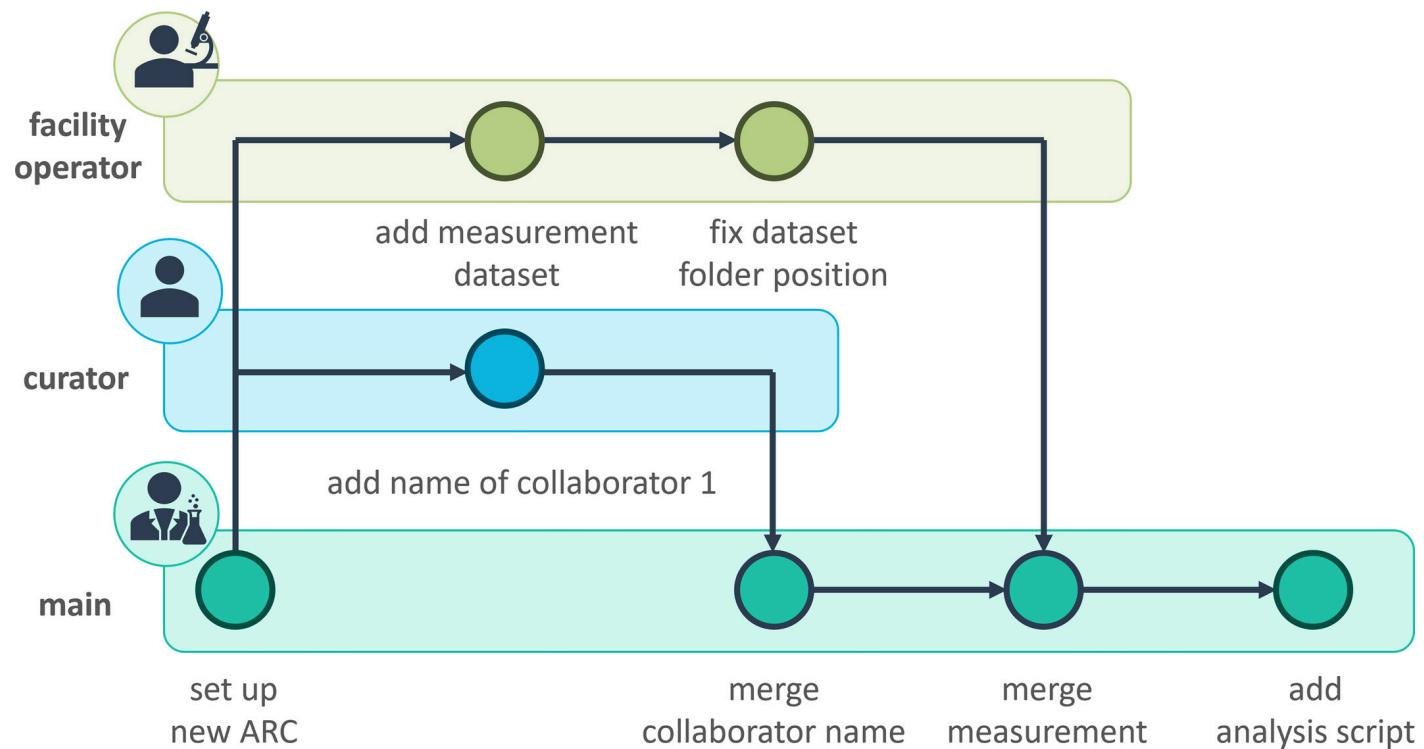




Reuse

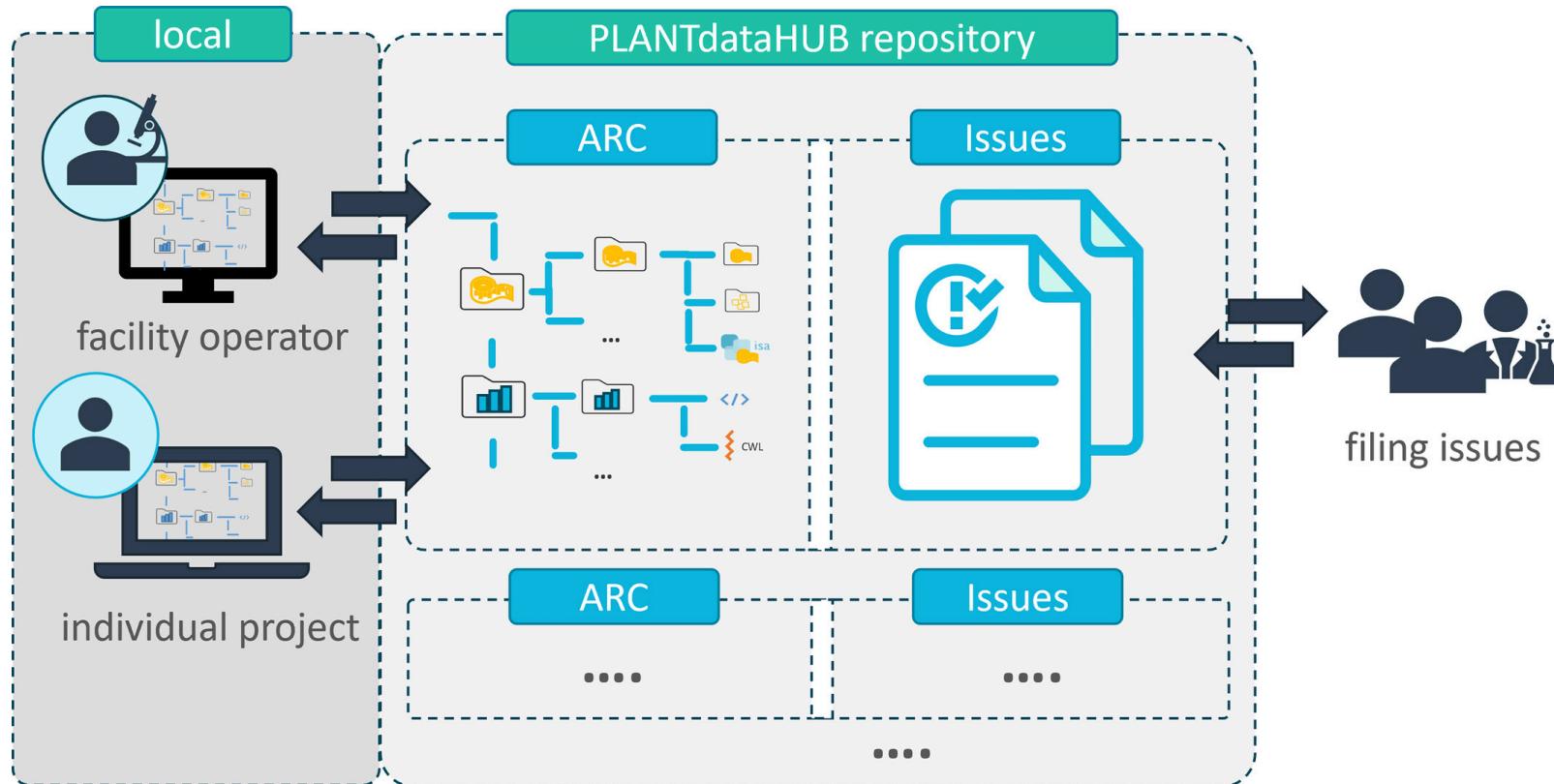


Mutable data life cycle





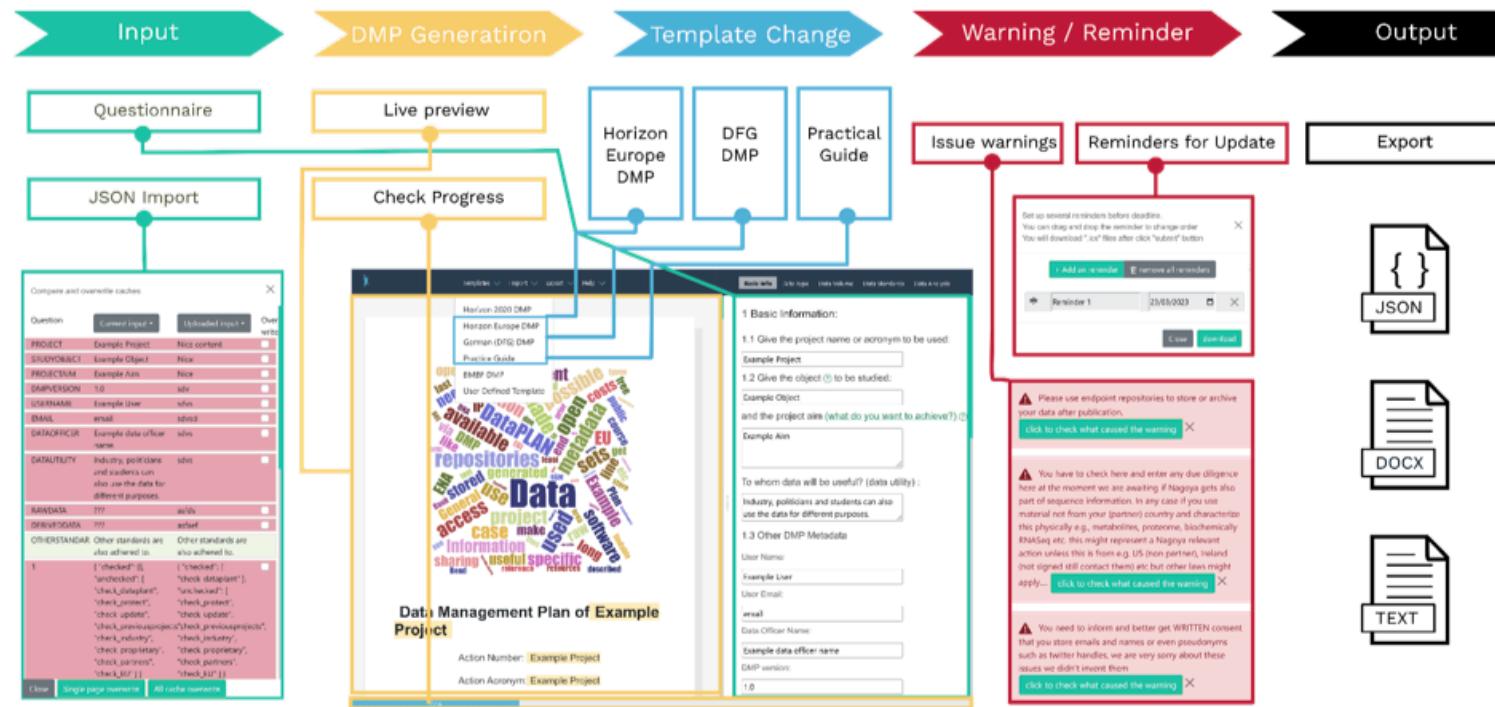
Plan (ARC scale)





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>

