

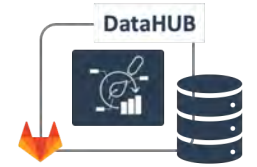
The ARC Club

a two-day adventure to prepare your lab for the
ARC universe

Dominik Brillhaus – CEPLAS Data Science

June 28th, 2023

Annotated Research Context (ARC)



The ARC Club

Preparation

before August 16th, 2023

Dominik Brillhaus – CEPLAS Data Science

Checklist hands-on sessions

💡 Please prepare the following before the workshop:

Required:

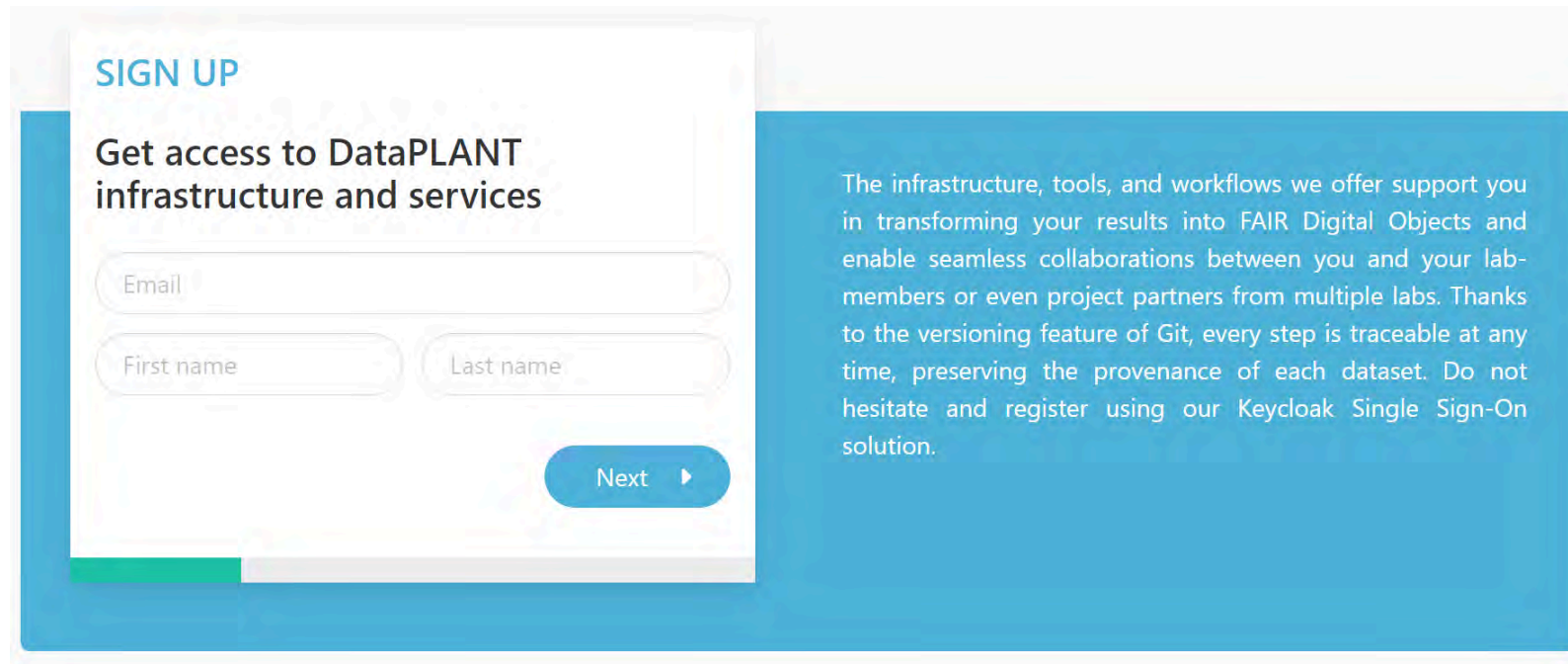
- ✓ Register at DataPLANT
- ✓ Install ARCitect on your computer
- ✓ Install Swate on your computer

Recommended (for trouble-shooting):

- ✓ Find your command line
- ✓ Install ARC Commander and dependencies on your computer
- ✓ Install VS Code

DataPLANT Registration

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

A screenshot of the DataPLANT registration form. The form is titled "SIGN UP" in blue. Below the title, it says "Get access to DataPLANT infrastructure and services". There are three input fields: "Email", "First name", and "Last name". A blue "Next" button with a right arrow is at the bottom right of the form. To the right of the form, there is a blue box with white text describing the services offered.

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

Join the group

Once signed-up and logged in, please join the [HHU Plant Biochemistry group](#).

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.

Recommended for trouble-shooting

💡 We will likely not use the tools on the next few slides.

However, as of now (early August 2023), it's probably better to have them ready for trouble-shooting and to show some inner workings of the ARC.

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.

ARC Commander Installation

Please install the latest version of the ARC Commander and dependencies for your operating system according to the [manual's setup instructions](#).

Check if the ARC Commander is functional by displaying the ARC Commander version and help menu:

```
arc --version
```

Setup ▾

- Installing Dependencies
- Configure Git
- Installing the ARC Commander
 - Windows
 - MacOS
 - Linux
- DataHUB Access
- Before we start

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>

The ARC Club

Day 1 – Into the ARC

Dominik Brillhaus – CEPLAS Data Science

August 16th, 2023

ARC Club Goals

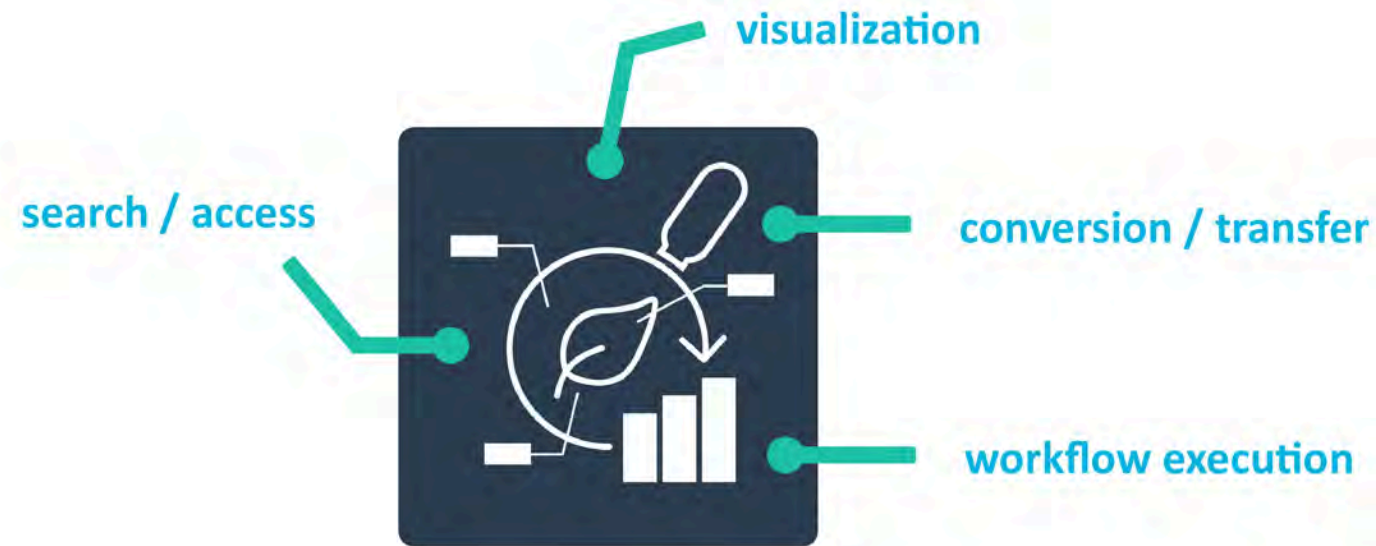
- Move existing datasets into ARCs
- Share them via the DataHUB
- First few steps into ARCs
- Data users can pick them up from there

Rules: perfect is the enemy of good

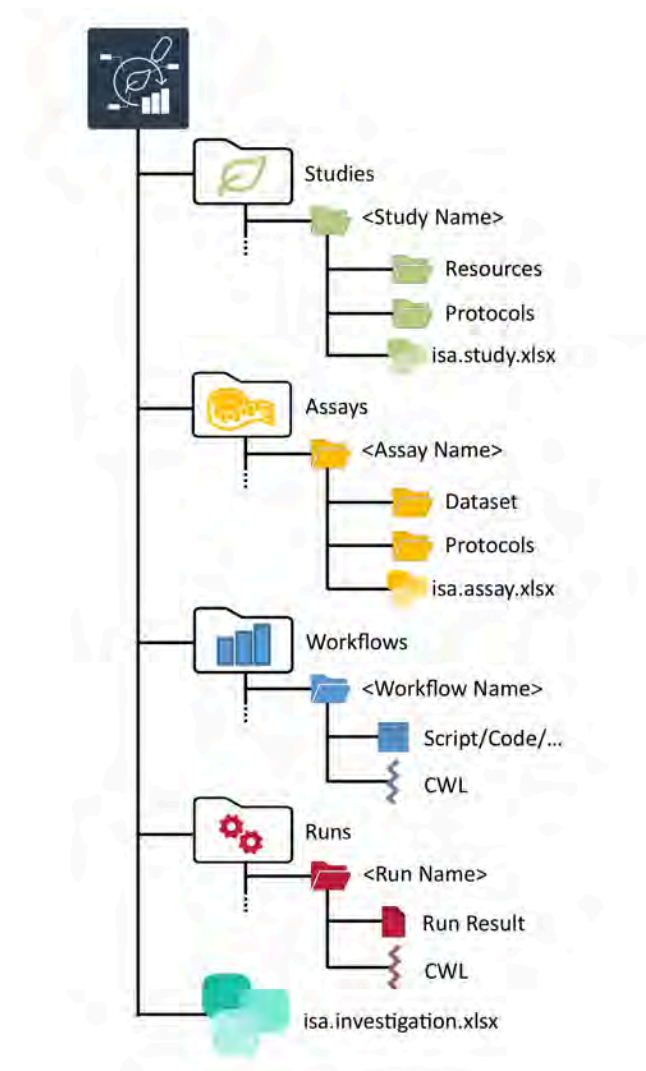
- There is no perfect ARC
- There is no complete ARC
- The only bad ARCs are those that don't exist yet.

 Let's get started, the rest is easy 

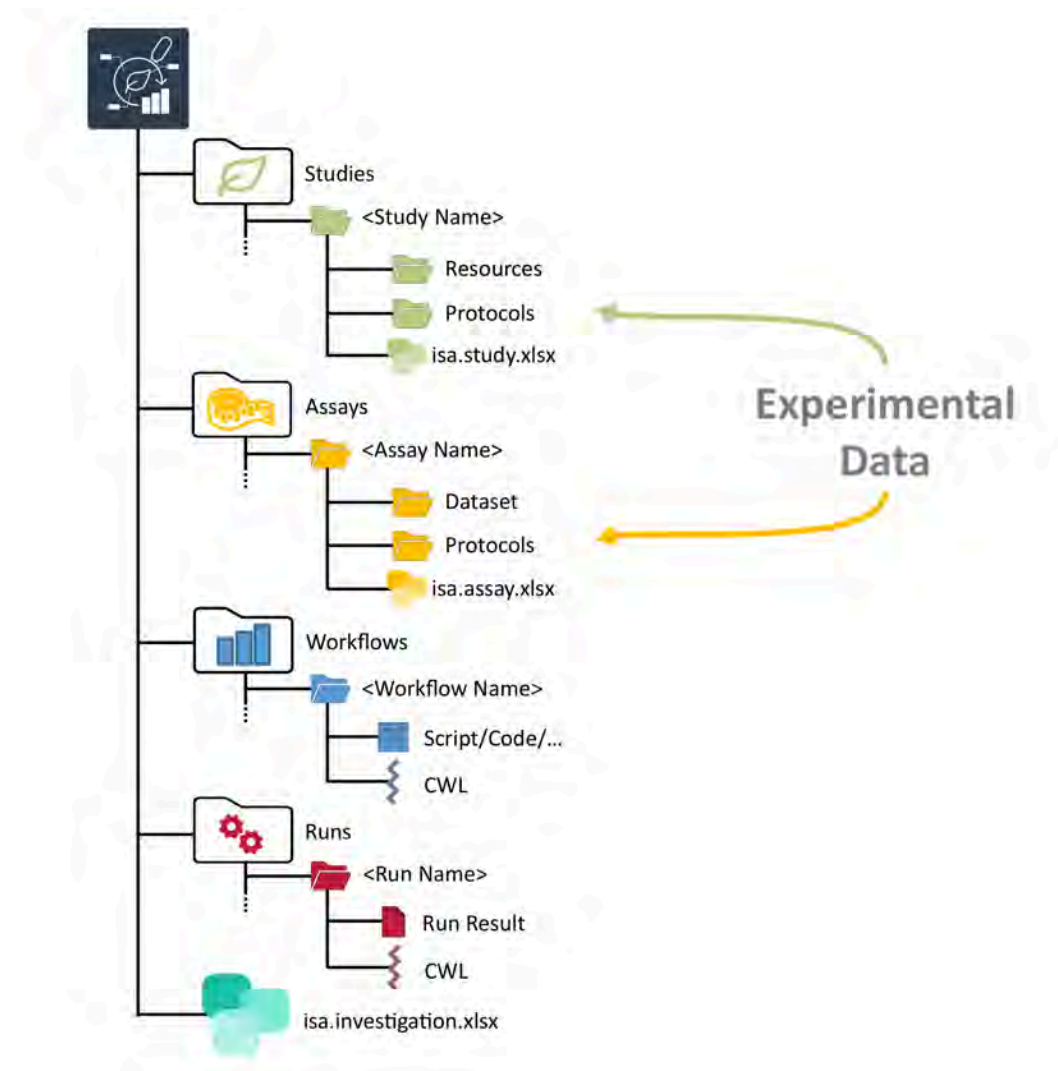
Annotated Research Context (ARC)



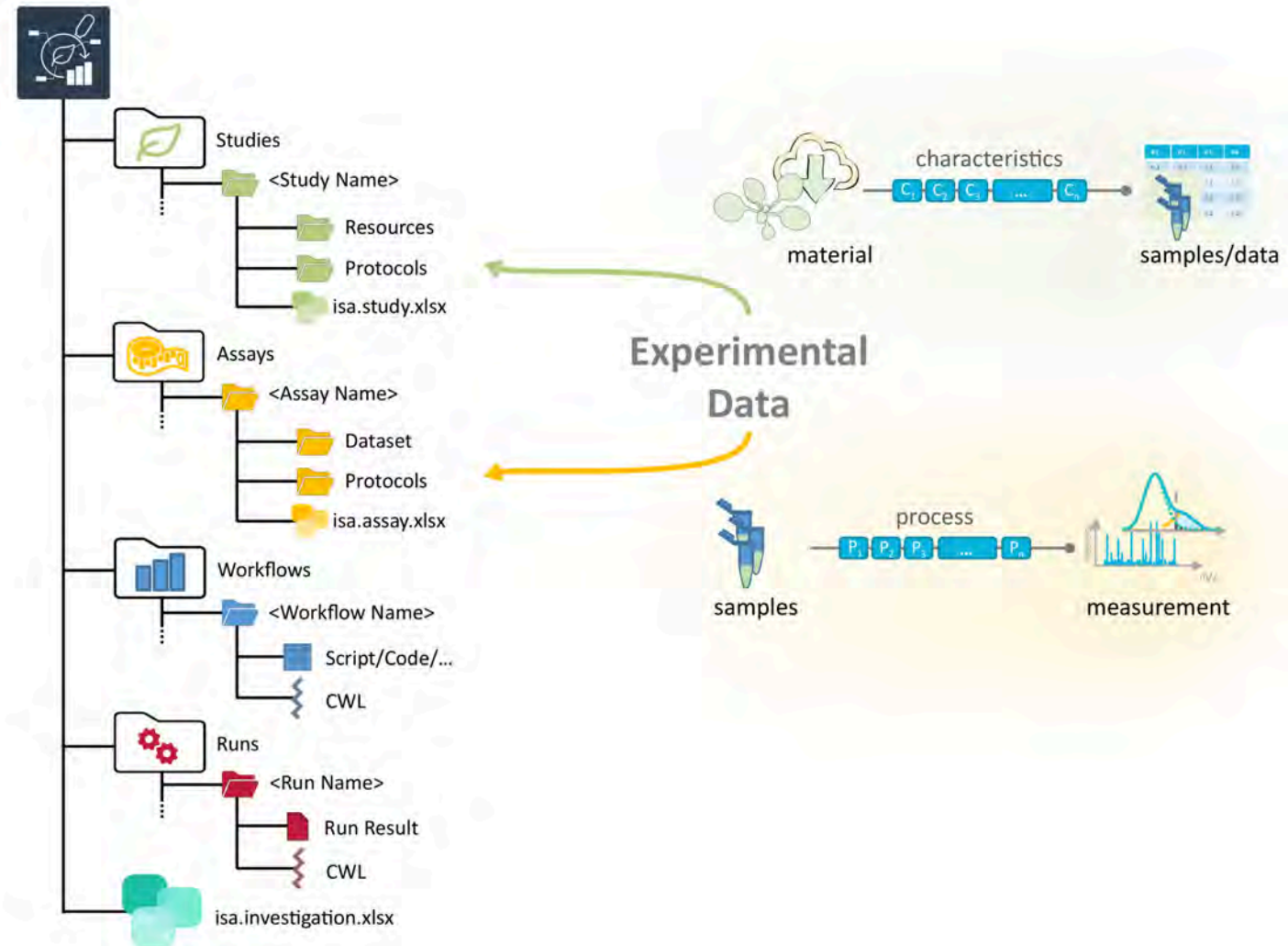
What does an ARC look like?



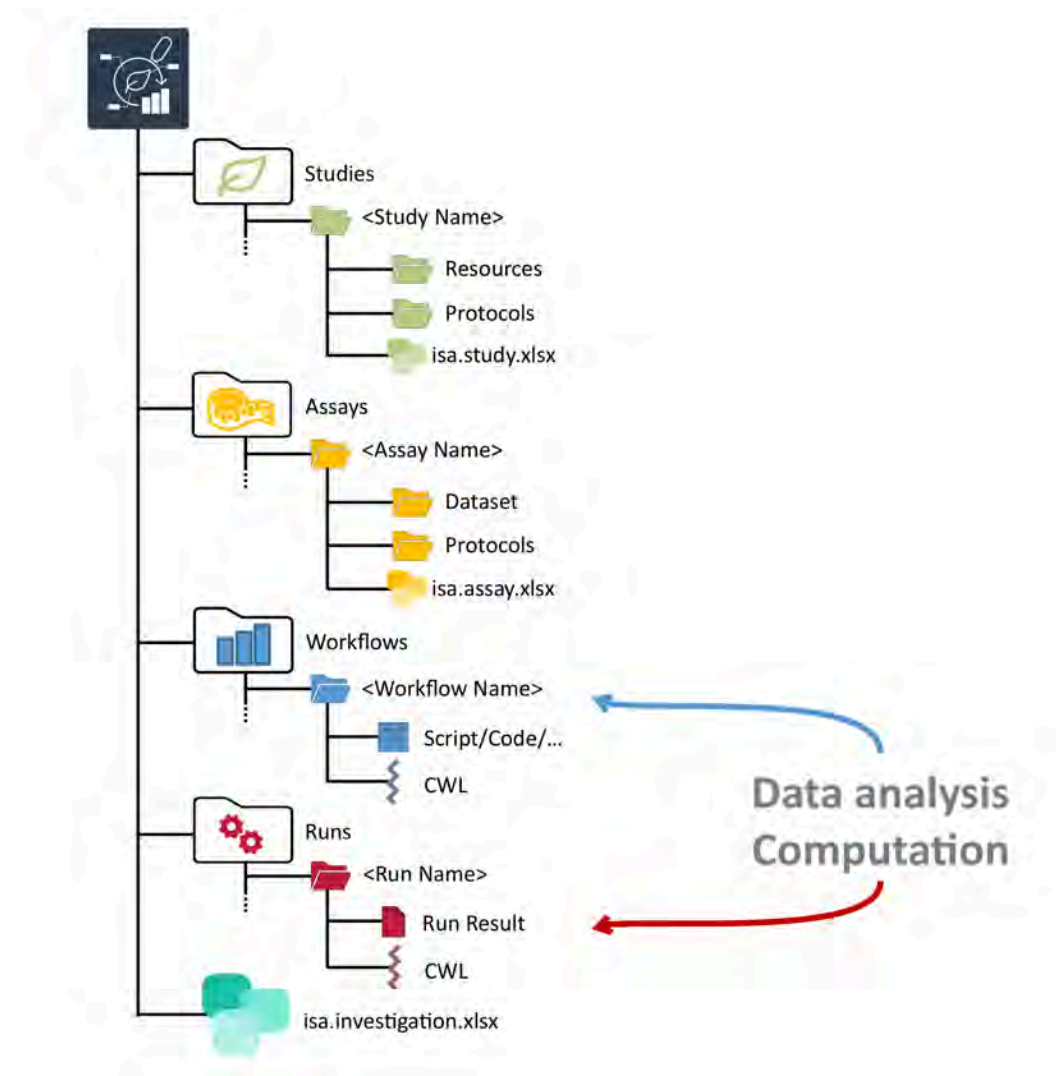
What does an ARC look like?



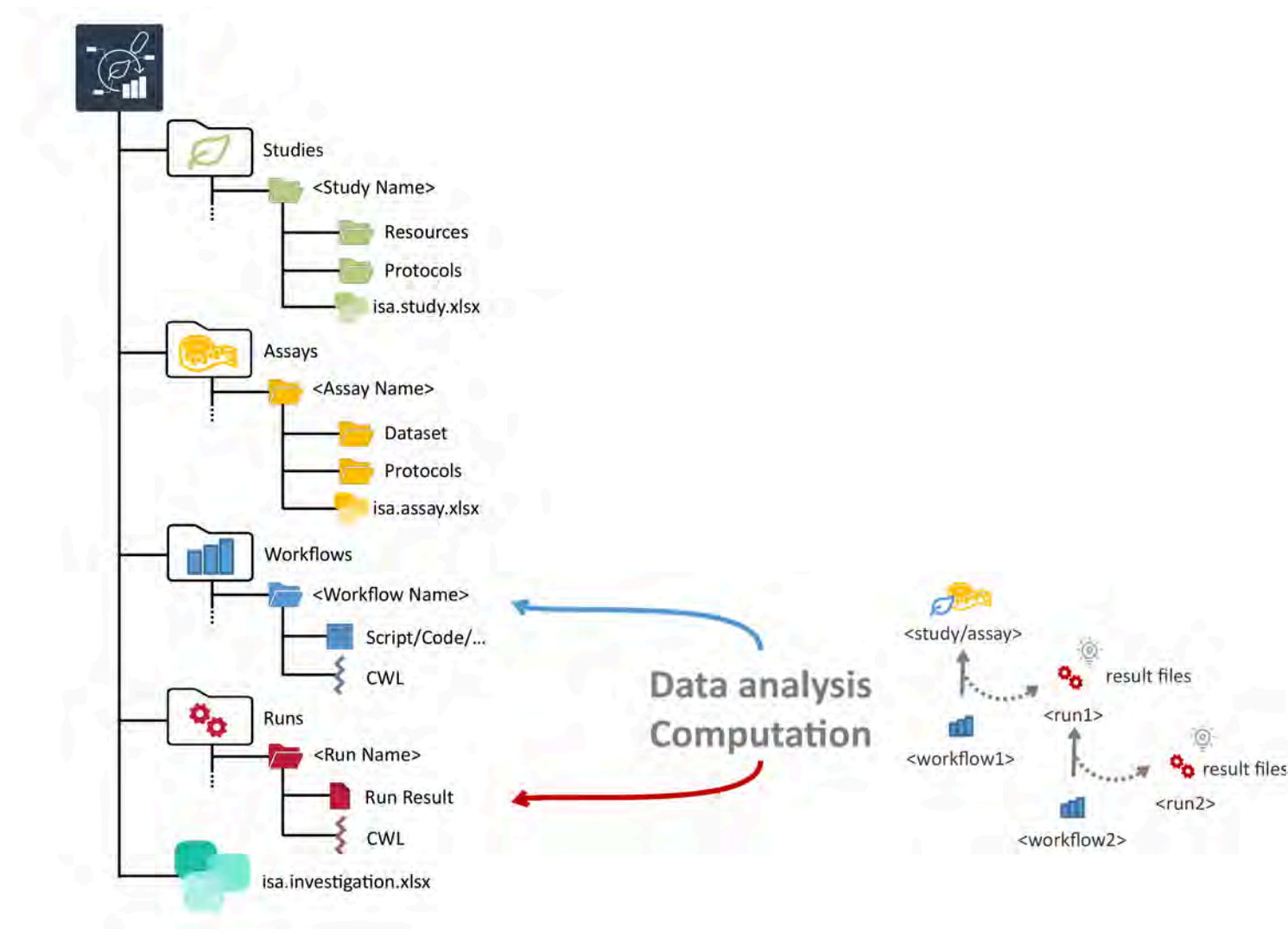
What does an ARC look like?



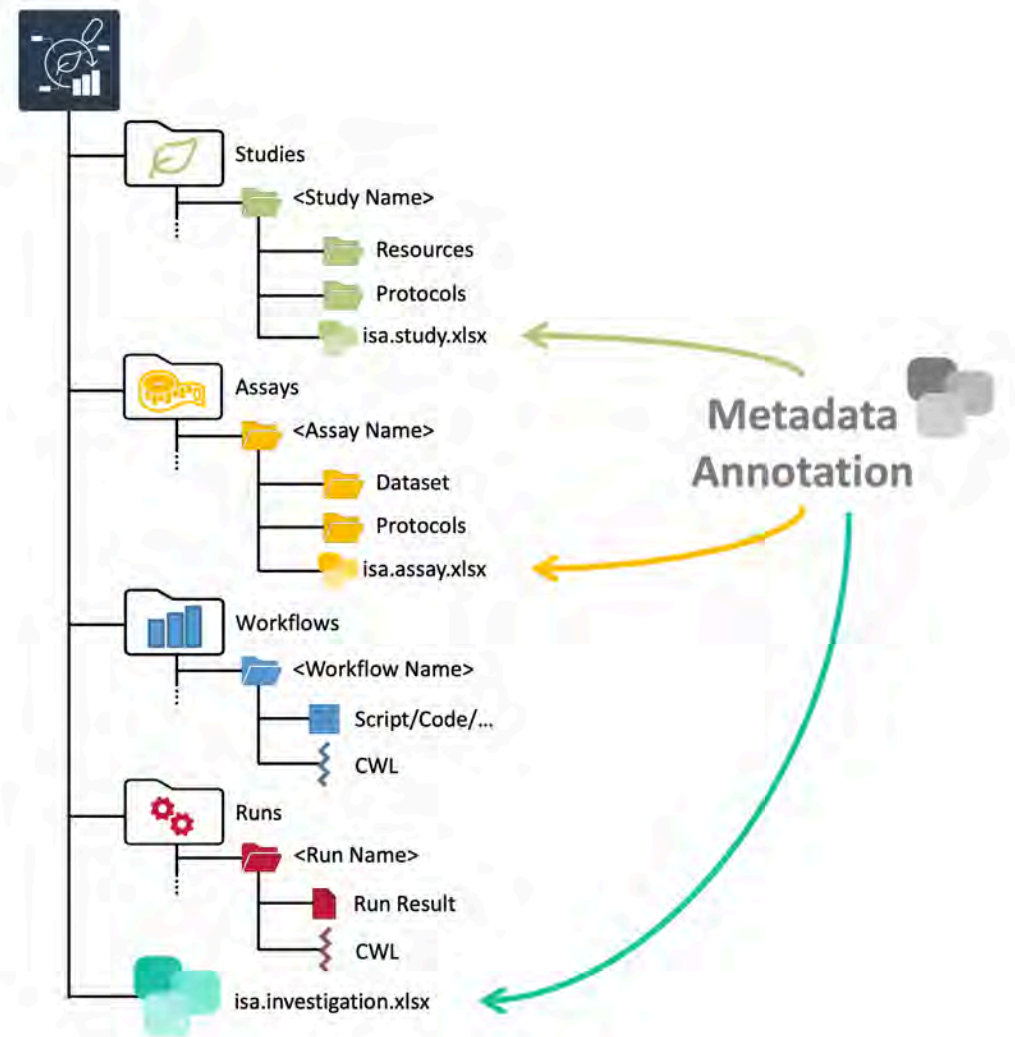
What does an ARC look like?



What does an ARC look like?



What does an ARC look like?





FINDABLE

ACCESSIBLE

INTEROPERABLE

REUSABLE

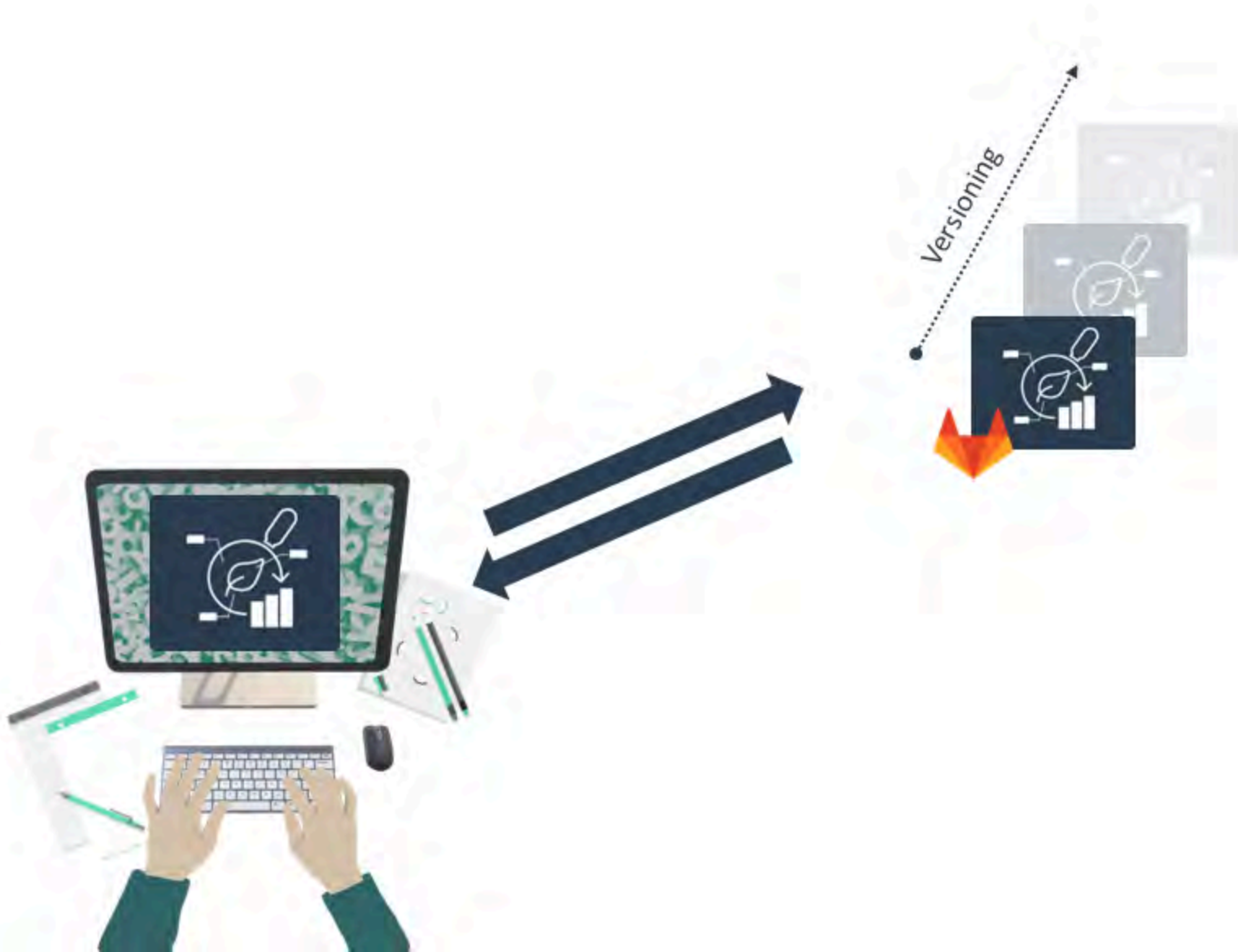
Metadata
Templates

Controlled
Vocabularies

Standards



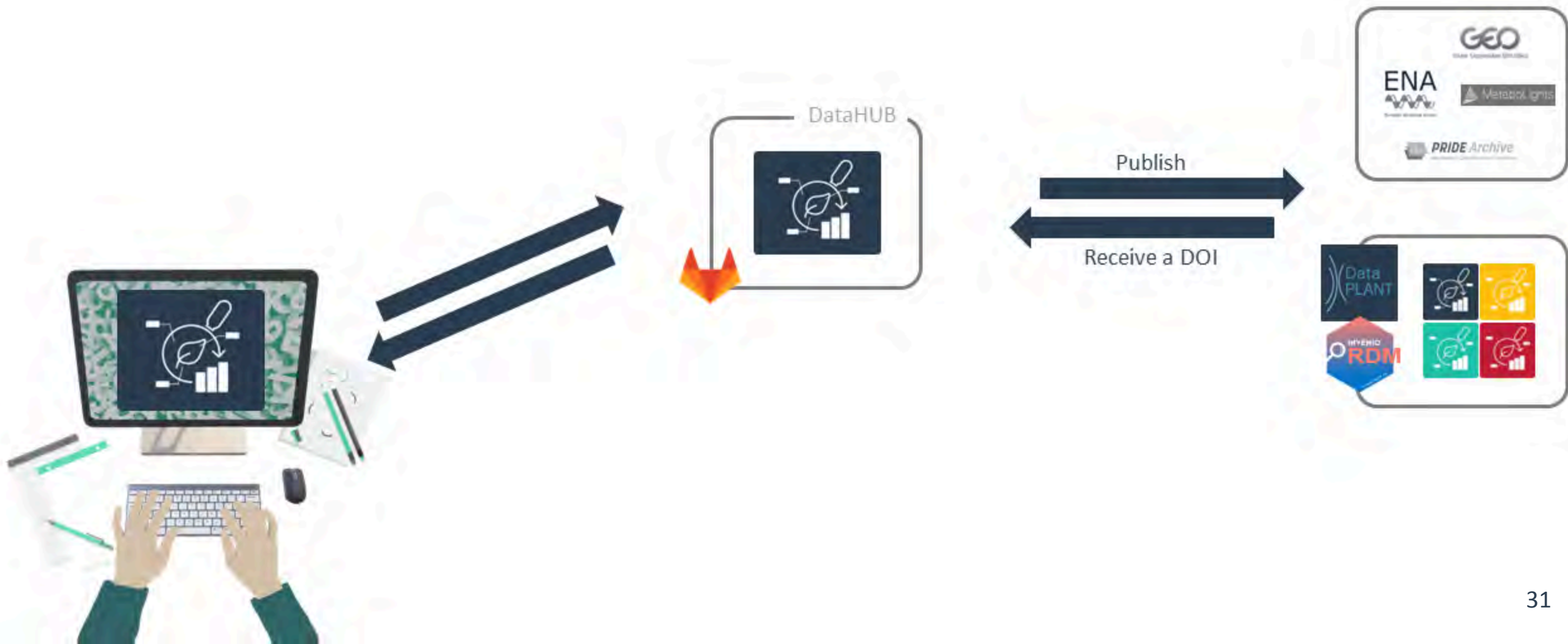


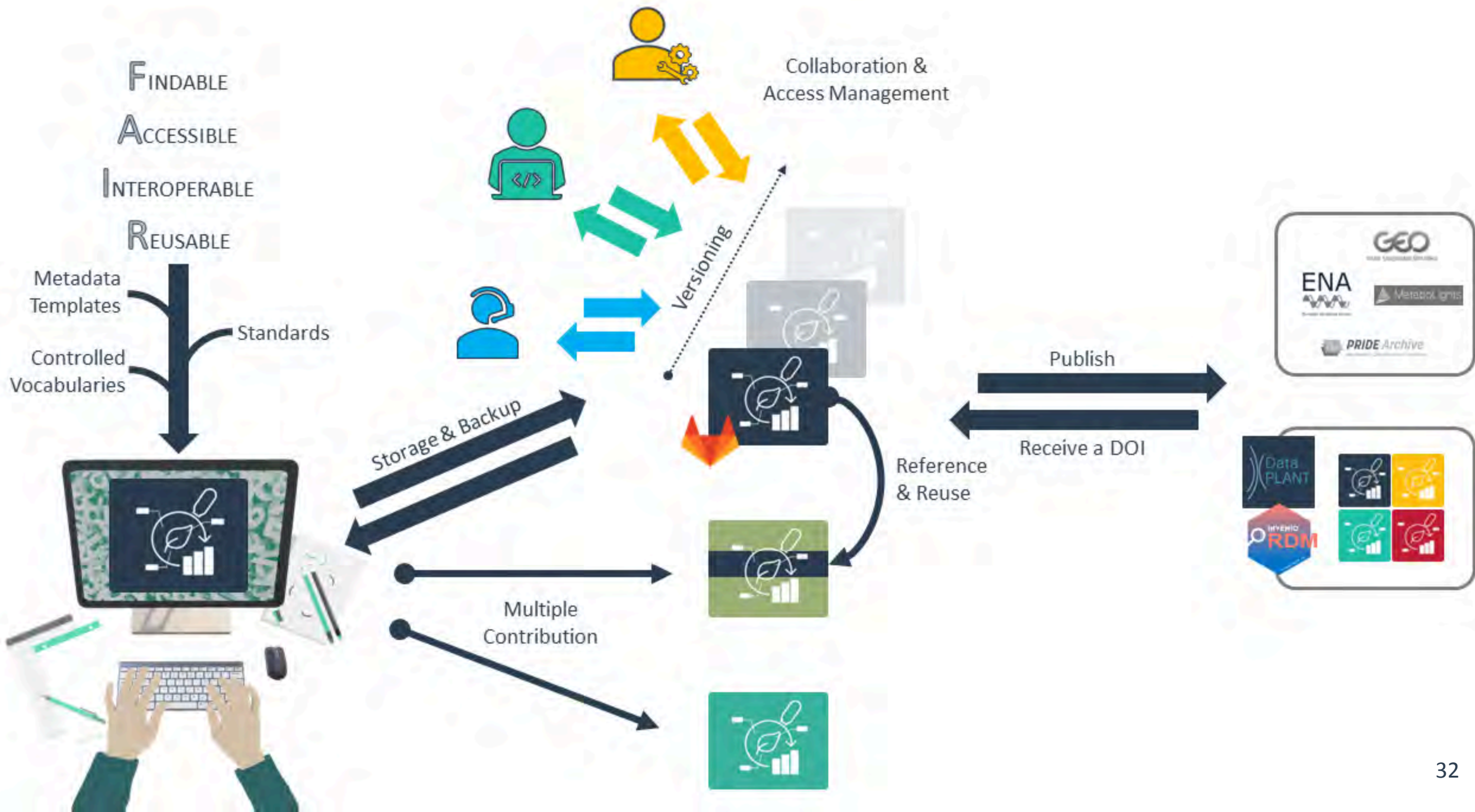












Contributors

Slides presented here include contributions by

- Dominik Brilhaus | [GitHub](#) | [ORCID](#)
- Cristina Martins Rodrigues | [GitHub](#) | [ORCID](#)
- Martin Kuhl | [GitHub](#) | [ORCID](#)

The ARC Club

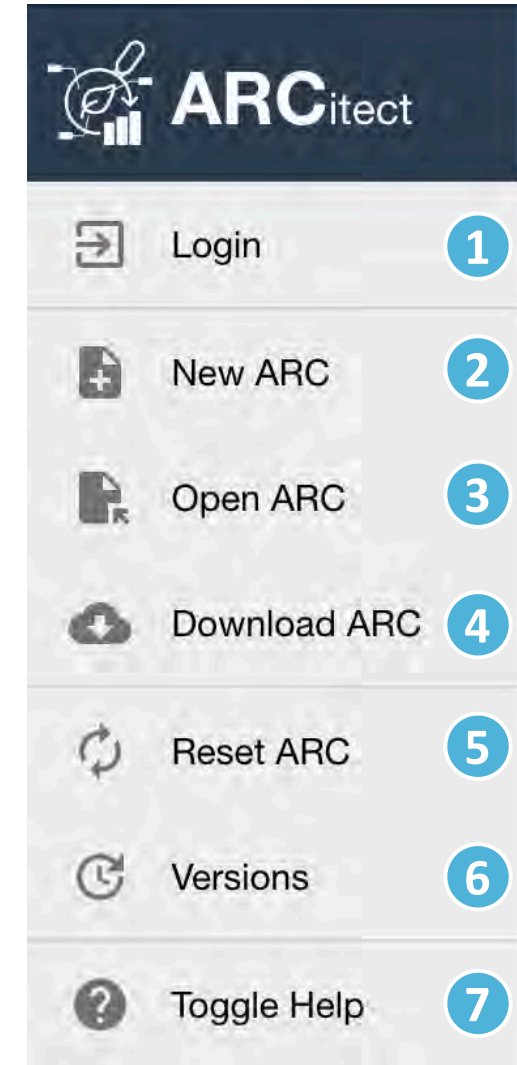
ARCitect QuickStart

Dominik Brillhaus – CEPLAS Data Science

August 16th, 2023

Initiate the ARC folder structure

1. Create a **New ARC** (2)
2. Select a location and name for your ARC



Your ARC's name

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

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


💡 Make sure that no ARC exists at

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



Add a study


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

▼  Demo-ARC

▶ assays

▶ runs

▼ studies

 Add Study

▶ workflows

Add information about your study

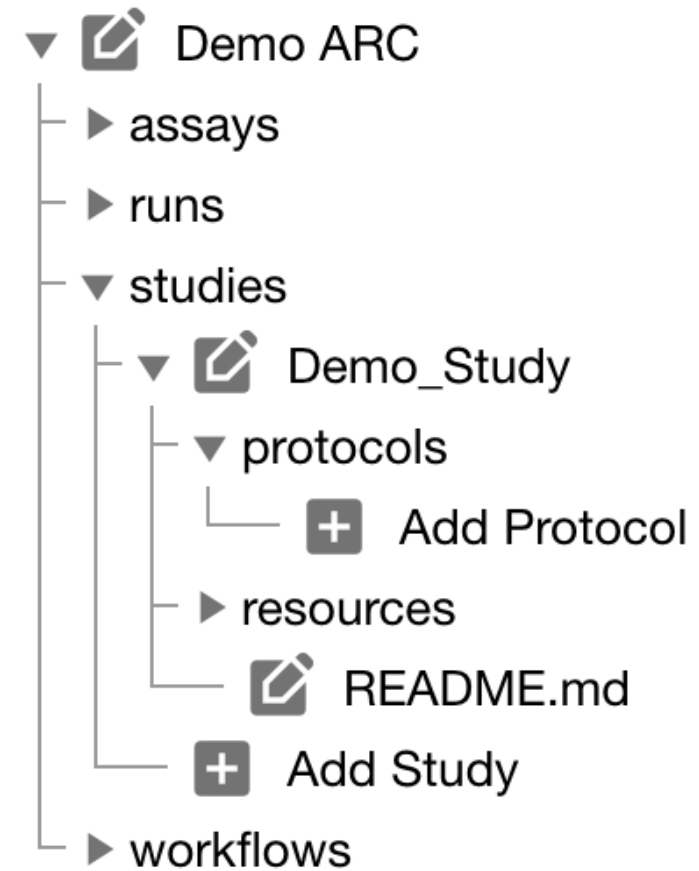
In the study panel you can add

- general metadata,
- people, and
- publications

The screenshot shows a web application interface for managing study data. On the left, a file explorer displays a directory structure under 'Demo-ARC', including folders for 'assays', 'runs', 'studies', and 'workflows'. The 'studies' folder is expanded, showing 'Demo_Study' (selected), 'protocols', 'resources', and a 'README.md' file (0.00 B). An 'Add Study' button is visible below the file list. On the right, the 'Study' panel is titled 'General Meta Data of the Study'. It contains a form with the following fields: 'Identifier' (set to 'Demo_Study'), 'Title' (set to 'Demo_Study'), and a large 'Description' text area. Below the description is a placeholder text: 'A textual description of the study'. Further down, there are two date fields: 'Submission Date' and 'Public Release Date', both with calendar icons. Below these are two buttons: 'UPDATE' (with a checkmark icon) and 'RESET' (with a circular arrow icon). At the bottom of the panel, there are three expandable sections: 'People' (Authors and Collaborators), 'Publications' (Papers, Books and Other Media), and 'Data' (Process Information).

Add protocols to your study

In the file tree you can **add protocols**



Add protocols

You can either

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

Create or Import Protocol

Protocol Name



NEW PROTOCOL

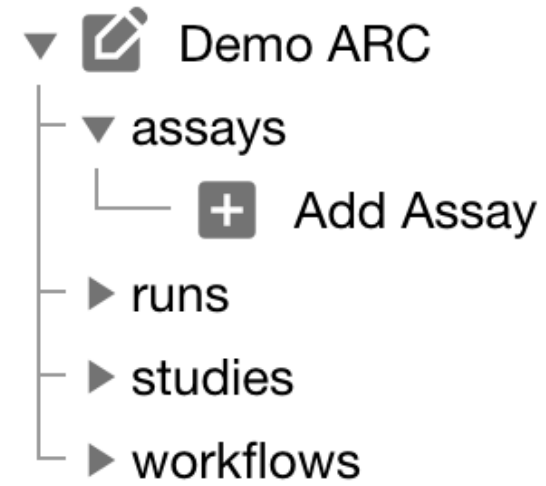


IMPORT PROTOCOL

CANCEL

Add an assay

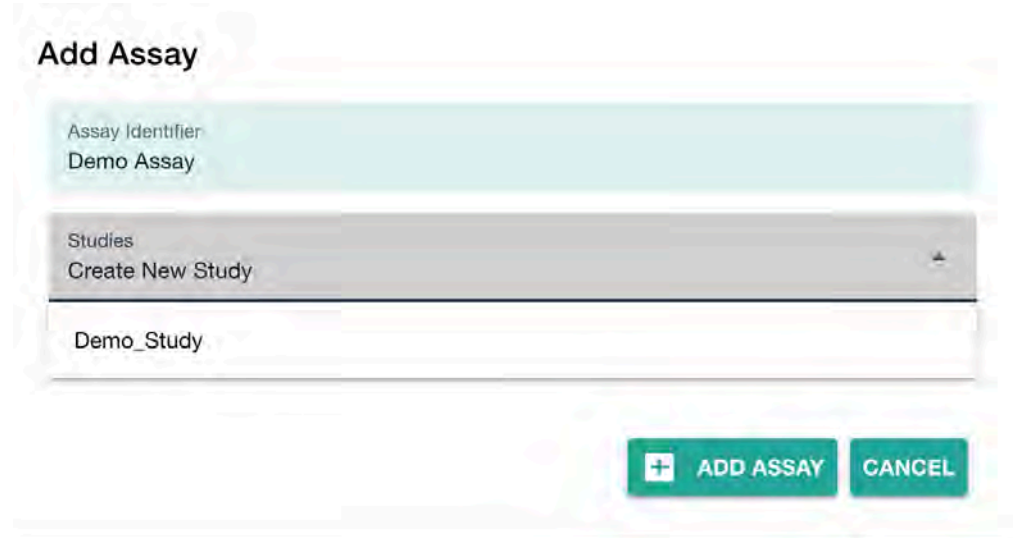
by clicking "Add Assay" and entering an identifier for your assay



Link your assay to a study

You can either

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



The screenshot shows a modal dialog box titled "Add Assay". It contains three input fields: "Assay Identifier" with the text "Demo Assay", "Studies" with a dropdown menu showing "Create New Study", and a text field containing "Demo_Study". At the bottom right, there are two buttons: "ADD ASSAY" (green with a plus icon) and "CANCEL" (grey).



Add information about your 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.

Assay
General Meta Data of the Assay

Assay Identifier Demo_Assay	Studies Demo_Assay
Measurement Type	
A term to qualify the endpoint, or what is being measured, e.g., gene expression profiling or protein identification.	
Technology Type	
Term to identify the technology used to perform the measurement, e.g., DNA microarray, mass spectrometry.	
Technology Platform	
Manufacturer and platform name, e.g., Bruker AVANCE.	

 UPDATE  RESET

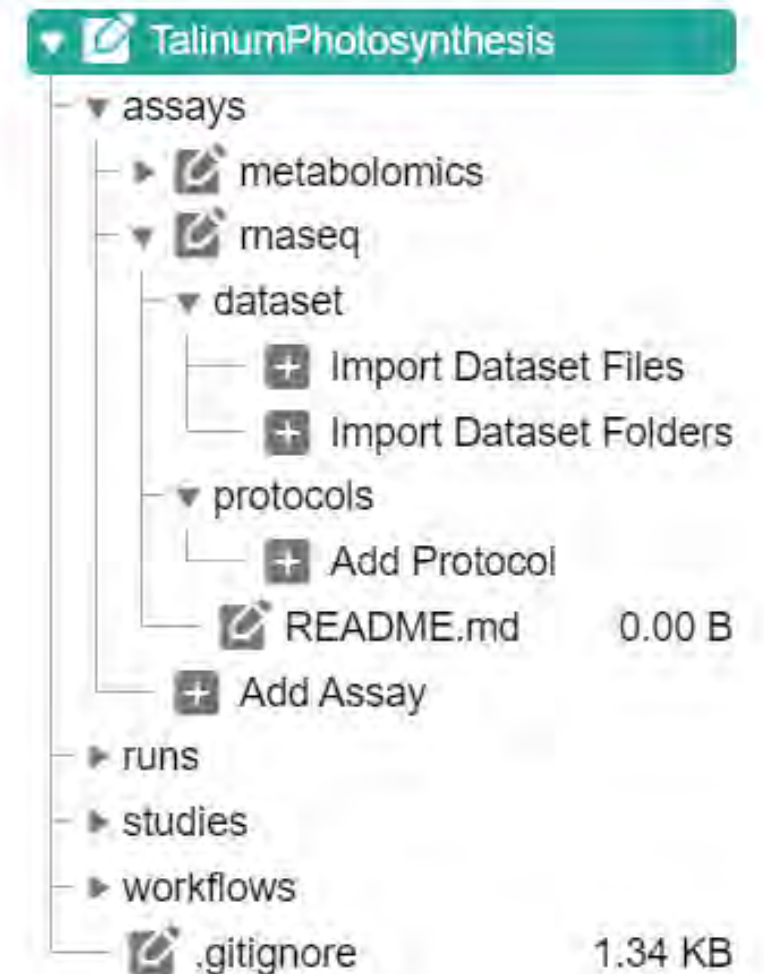
Add protocols and datasets

In the file tree you can

- **add a dataset** and
- **protocols** associated to that dataset.

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

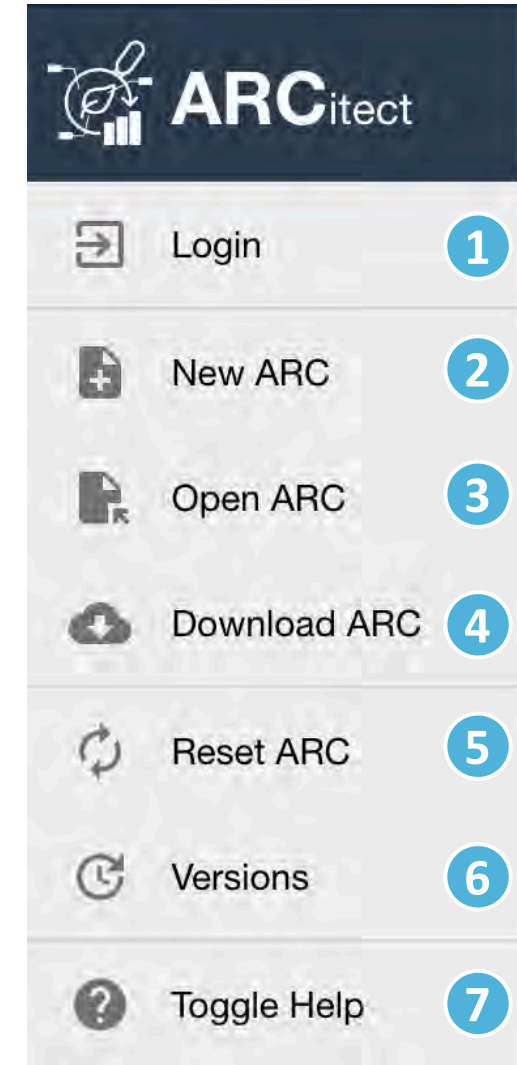
⚠️ Depending on the file size, this may take a while. Test this with a small batch of files first.



Login to the DataHUB

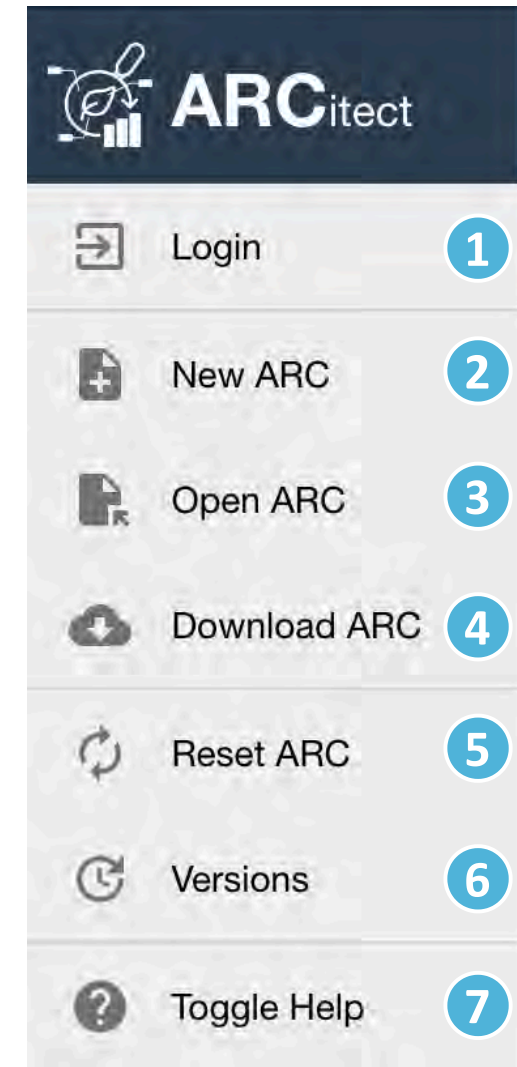
Click **Login** (1) in the sidebar to login to the DataHUB.

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



Upload your local ARC to the DataHUB

From the sidebar, 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 displays the 'Update' panel, which is used to commit changes and upload the ARC. The panel includes the following elements:

- Update** (with a refresh icon): 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 large text area for entering a message.
- Changes**: A list of files to be committed, each preceded by a green plus icon in a square:
 - .arc/
 - assays/
 - isa.investigation.xlsx
 - runs/
 - studies/
 - workflows/
- Buttons**: REFRESH, COMMIT, UPLOAD, and DOWNLOAD.
- History** (with a refresh icon): Inspect ARC history

Connection to the DataHUB

If you are logged in, the versions panel shows

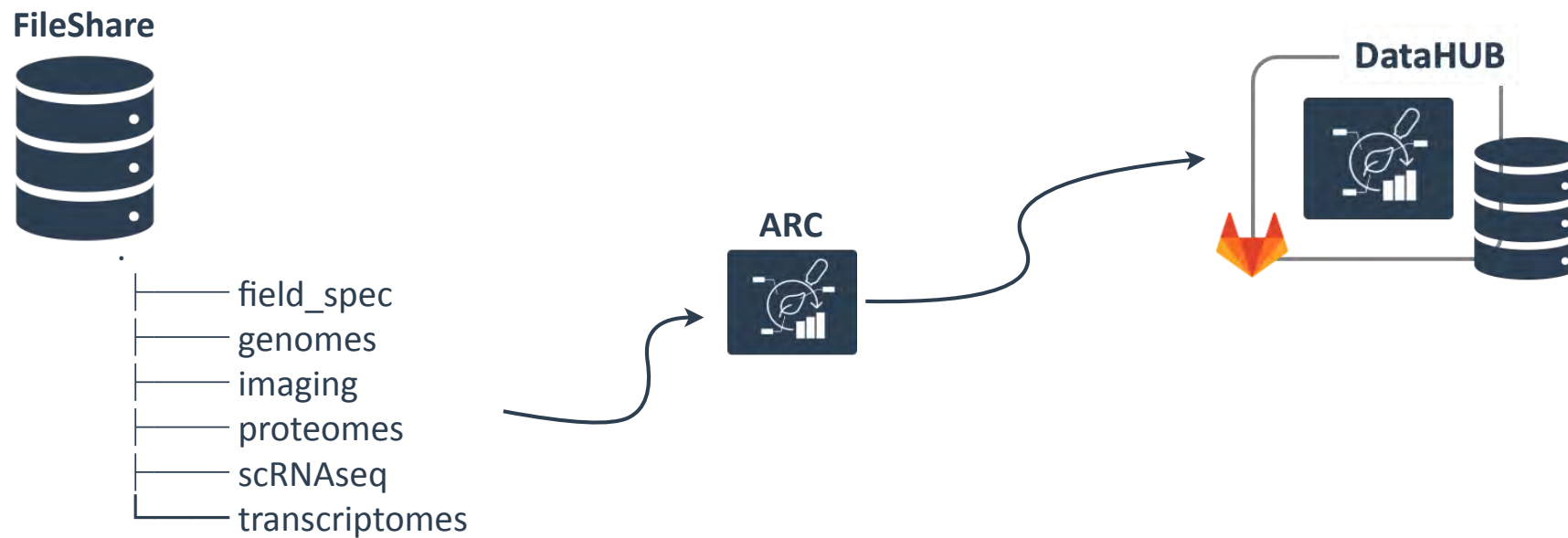
- your DataHUB's *Full Name* and *eMail*
- the URL of the current ARC in the DataHUB

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

The screenshot shows the 'Update' form in the DataHUB interface. At the top, there's a title 'Update' with a subtext 'Commit changes and upload ARC'. Below this, there are two input fields: 'Full Name' with the value 'Demo User' and 'eMail' with the value 'demo@nfdi4plants.org'. A 'Remote' field contains the URL 'https://git.nfdi4plants.org/demouser/Demo-ARC.git'. A large text area for 'Commit Message' is present, with a placeholder text 'A short description of the made changes'. Below the message area, a 'Changes' section lists files with green plus icons: '.arc/', 'assays/', 'isa.investigation.xlsx', 'runs/', 'studies/', and 'workflows/'. At the bottom of the form, there are four buttons: 'REFRESH', 'COMMIT', 'UPLOAD', and 'DOWNLOAD'. Below the form, there is a 'History' section with the subtext 'Inspect ARC history'.

The ARC Club

Moving from FileShare to DataHUB – *via* ARCs



Assign projects

Rough routine for each project

1. Identify the available data and resources
2. Create the ARC
3. Add metadata and data
4. Share via DataHUB group <https://git.nfdi4plants.org/hhu-plant-biochemistry/>

Low(er) hanging fruits: published projects

1. Add the authors
2. Add the publication(s)
 - i. Add citation and DOI
 - ii. Add supplemental
 - iii. Convert M&M to *protocols*
3. Reference data in public repositories
4. Add large data (e.g. from file share)
5. Set ARC to **public!**

More challenging ARCs

- (unpublished) left-overs of colleagues who have since moved
- unclear

Collect / derive as much info about the **investigation** as possible

MUST have

```
Investigation Identifier  
Investigation Title  
...  
Investigation Publication Status  
...  
Investigation Person Last Name  
Investigation Person First Name
```

💡 This and more investigation-level info can be collected in the ARC's

```
isa.investigation.xlsx
```

Create and share the ARC

```
arc init  
arc sync -f -r https://git.nfdi4plants.org/hhu-plant-biochemistry/<InvestigationID>
```

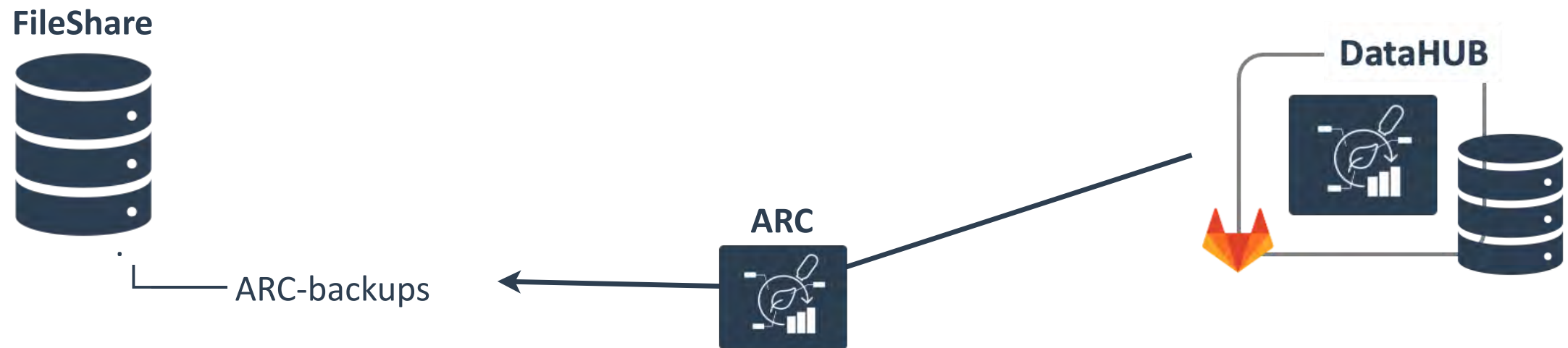
Copy data

1. **Copy** data to the ARC, do not **move** data from original source (we'll take care of that later)
2. Ideally use `rsync` rather than copying manually
3. Ideally use `md5` or `md5sum` to check for correct file transfer

💡 Ask the coders for help!

Perspective and administration in the future

Administration / Backup



The ARC Club

Intro and Hands-on Swate

Dominik Brillhaus – CEPLAS Data Science

August 17th, 2023

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 editors with code highlighting:

- Visual Studio Code <https://code.visualstudio.com/>

Download the demo data

```
git clone "https://demo-user:5ehDYeHcqP2MqVXsNNPu@git.nfdi4plants.org/teaching/demo-arc_level1.git"
```

Where we left off last time

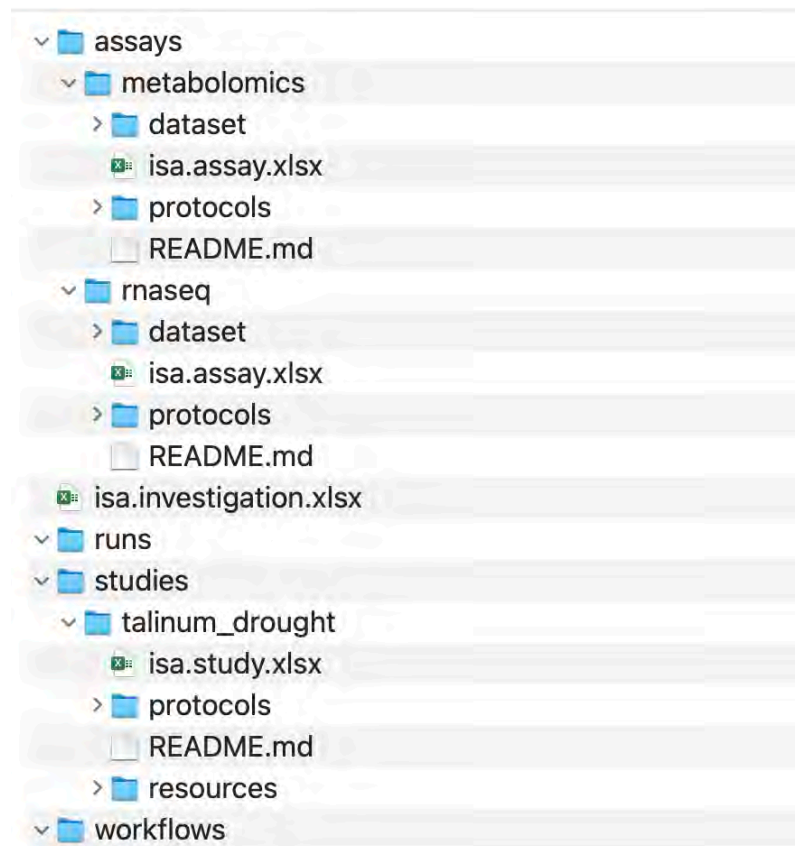
 Initiated an ARC

 Structured and ...

 Shared with collaborators

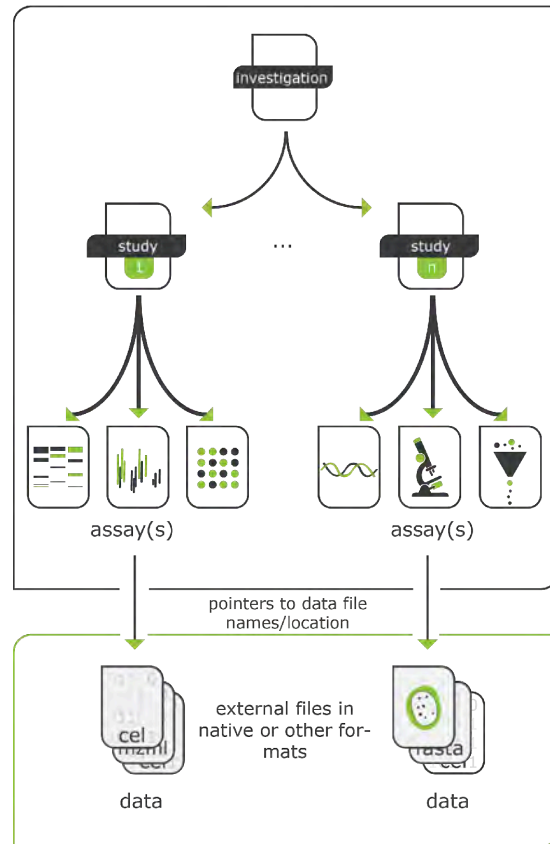
Today we want to

 ... **annotate the experimental data**



Intro ISA

ARC builds on ISA



Investigation

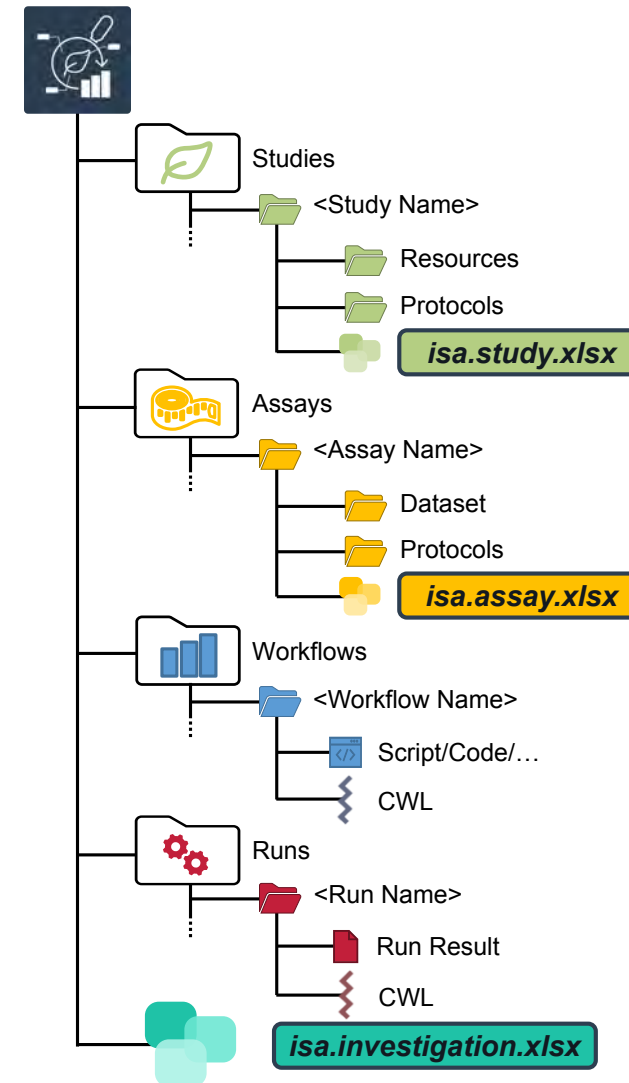
Overall goals
Scientific context

Study

Experimental design

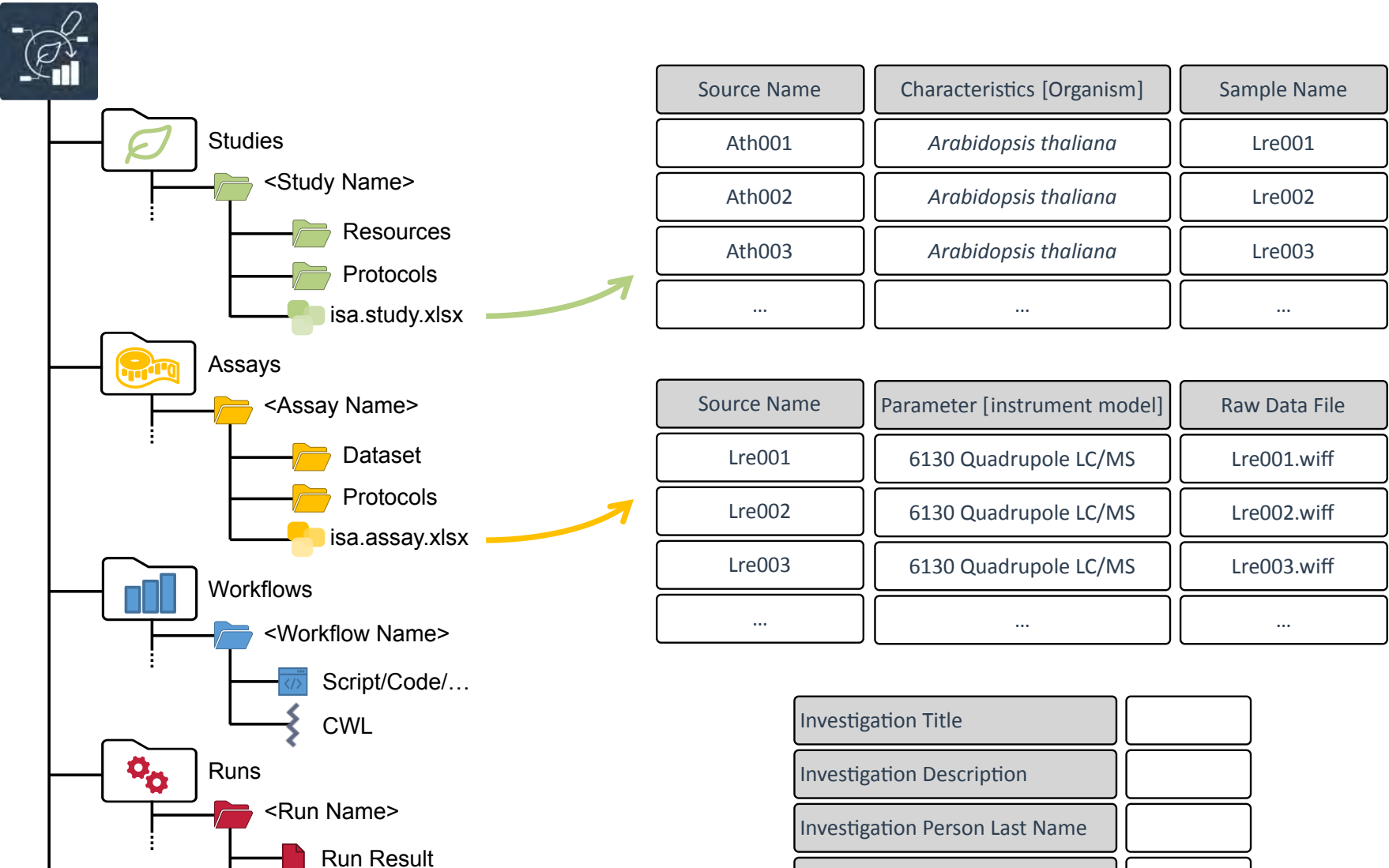
Assay

Leading to
(raw) data



ARC builds on ISA

Metadata
Annotations



isa.investigation.xlsx

isa.study.xlsx

isa.assay.xlsx



Study and assay files are registered in the investigation file

[illegible]

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

Output building blocks:

- Sample Name
- Raw Data File
- Derived Data File

isa.study.xlsx

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

Samples

isa.assay.xlsx

Source Name	Parameter [sample volume]	Parameter [injection volume]	Parameter [measurement duration#4]	Raw Data File
WCGr2_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_U1.wiff
WCGr2_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_U2.wiff
WCGr2_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_U3.wiff
WCGr2_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_U4.wiff
WCGr2_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5_1.wiff
WCGr2_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5_2.wiff
WCGr2_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5_3.wiff
WCGr2_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5_4.wiff
WCGr2_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_UF_1.wiff
WCGr2_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_UF_2.wiff
WCGr2_UF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_UF_3.wiff
WCGr2_UF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_UF_4.wiff
WCGr2_5F_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5F_1.wiff
WCGr2_5F_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5F_2.wiff
WCGr2_5F_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5F_3.wiff
WCGr2_5F_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr2_5F_4.wiff
WCGr1_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_U1.wiff
WCGr1_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_U2.wiff
WCGr1_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_U3.wiff
WCGr1_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_U4.wiff
WCGr1_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_5_1.wiff
WCGr1_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_5_2.wiff
WCGr1_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_5_3.wiff
WCGr1_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_5_4.wiff
WCGr1_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_UF_1.wiff
WCGr1_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600	87,00 minute WCGr1_UF_2.wiff

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

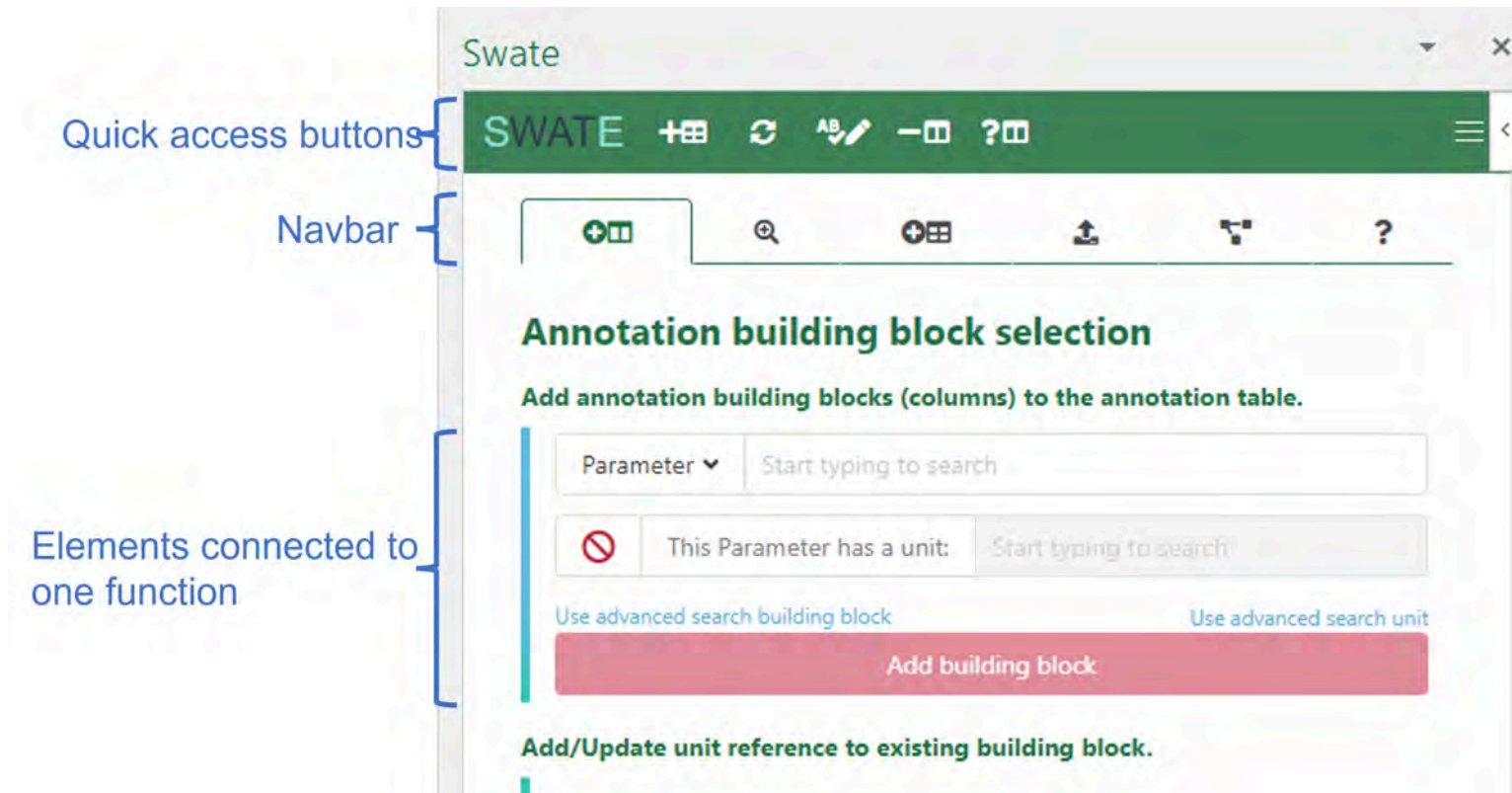
The screenshot displays a Microsoft Excel spreadsheet with a data table. The table has five main columns identified by annotations:

- Source:** Points to the first column, which contains values like "62_UMH_25m", "62_UMH_30m", etc.
- Characteristic:** Points to the second column, which contains values like "data extraction protocol", "data extraction protocol", etc.
- Component:** Points to the third column, which contains values like "30.00 degree Celsius", "30.00 degree Celsius", etc.
- Factor:** Points to the fourth column, which contains values like "4.00 degree Celsius", "4.00 degree Celsius", etc.
- Sample Name/ Raw Data File:** Points to the fifth column, which contains values like "WCS01_1", "WCS01_2", etc.

A "New Parameter" label points to a column on the right side of the table, which contains values like "WCS01_1", "WCS01_2", etc.

Swate hands-on with demo data

Swate Overview



Major areas of the Swate user interface.

Let's annotate the plant samples first

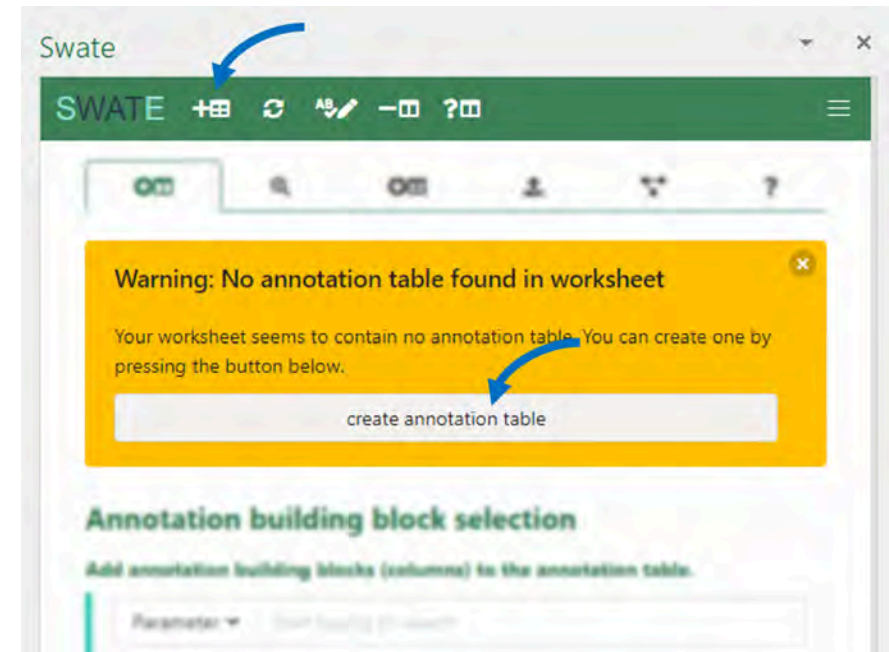
1. Navigate to the demo ARC.
2. Open the lab notes `studies/talinum_drought/protocols/plant_material.txt` in a text editor.
3. Open the empty `studies/talinum_drought/isa.study.xlsx` workbook in Excel.

Create an annotation table

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

💡 Each table is by default created with one input (Source Name) and one output (Sample Name) column

💡 Only one annotation table can be added per Excel sheet



Add a building block

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

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

Annotation building block selection

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

(A) Parameter (B) Start typing to search

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

Use advanced search building block Use advanced search unit

Add building block

Add/Update unit reference to existing building block.

Add a unit to a complete building block, if the building block already has a unit.

Insert values to annotate your data

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

💡 You should see **organism** showing in a field in front of the search field

💡 The search will now yield results related to **organism**

4. In the search field, search for "Talinum fruticosum"
5. Select the first hit and click **Fill selected cells with this term**

Add a building block with a unit

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


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


Insert unit-values to annotate your data

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

💡 You can see the numbers being complemented with the chosen unit, e.g. `425.00`
`microeinstein per square meter per second`

Showing ontology reference columns


Hold  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  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 **microeinstein per square meter per second** unit.

Your ISA table is growing

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

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

1

+

Hiding ontology reference columns

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

Exercise

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

Let's annotate the RNA Seq data

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

Use a template

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

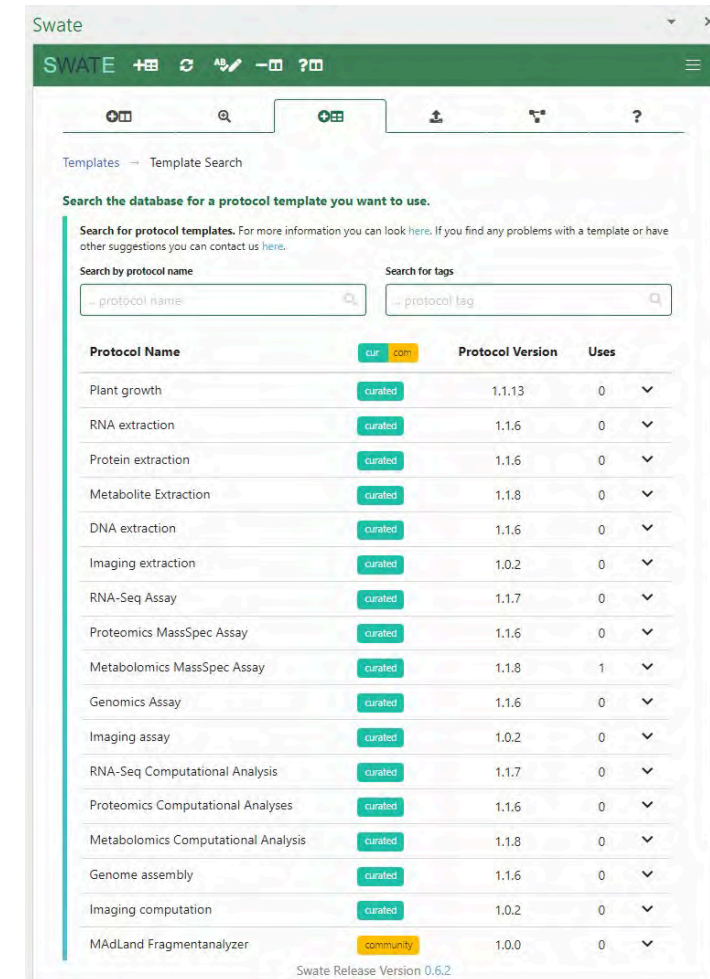


Here you can find community created workflow annotation templates

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


- You will see a preview of all building blocks which are part of this template.

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

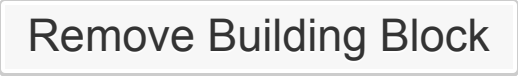



Adding / Updating unit references

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

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

Remove building blocks

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

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

New process, new worksheet

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

Exercise

Try to fill the two sheets with the protocol details:

- `assays/rnaseq/protocols/RNA_extraction.txt` and
- `assays/rnaseq/protocols/Illumina_libraries.txt`

Your ISA table is ready 🎉

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

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

A small detour on "Excel Tables"

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

Example workflows with three *processes* each:

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

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

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

Contributors

Slides presented here include contributions by

- Dominik Brilhaus | [GitHub](#) | [ORCID](#)
- Kevin Frey | [GitHub](#) | [ORCID](#)
- Martin Kuhl | [GitHub](#) | [ORCID](#)

