

The ARC Club

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

Dominik Brilhaus – CEPLAS Data Science
June 28th, 2023

Annotated Research Context (ARC)



The ARC Club

Preparation

before August 16th, 2023

Dominik Brilhaus – 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](#).

The screenshot shows the first step of a two-step registration process. On the left, a white rectangular form is titled "SIGN UP" in blue capital letters. Below the title, the text "Get access to DataPLANT infrastructure and services" is displayed in bold black font. There are three input fields: a top field labeled "Email", a bottom-left field labeled "First name", and a bottom-right field labeled "Last name". To the right of these fields is a large blue rectangular area containing descriptive text. The text explains that the platform supports transforming results into FAIR Digital Objects and enabling seamless collaborations between lab-members and project partners. It highlights the traceability feature of Git and the Keycloak Single Sign-On solution.

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 Brilhaus – CEPLAS Data Science

August 16th, 2023

ARC Club Goals

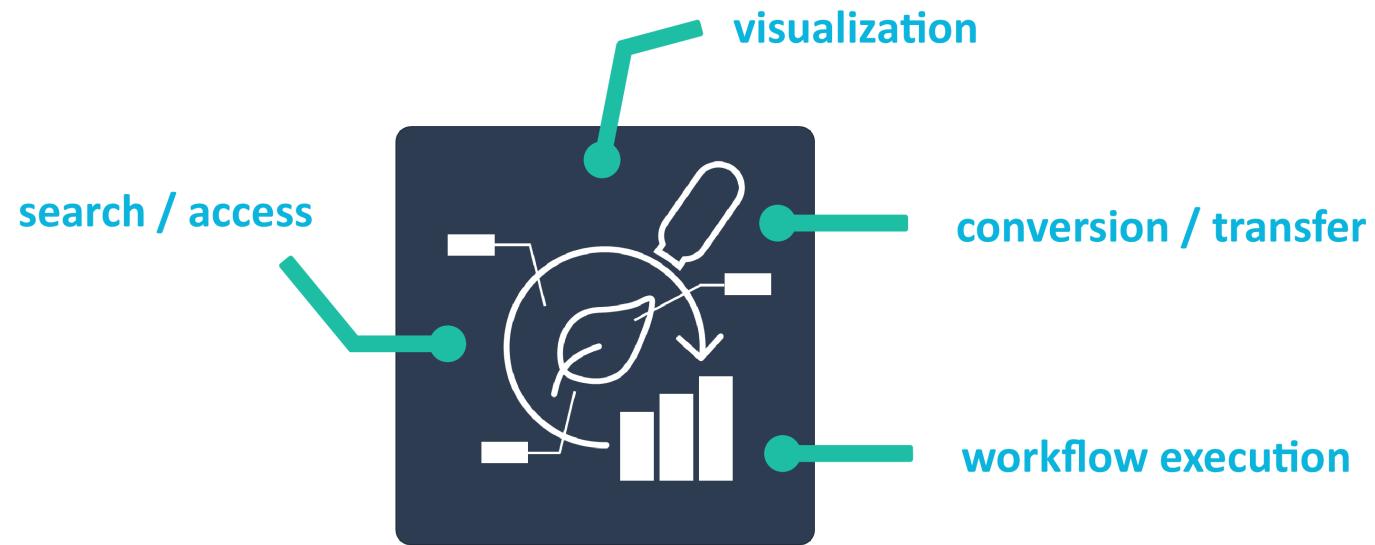
- Move existing datasets into ARCs
- Share them via the DataHUB
- First few steps into ARCs
- Data users can pick them up from there

Rules: perfect is the enemy of good

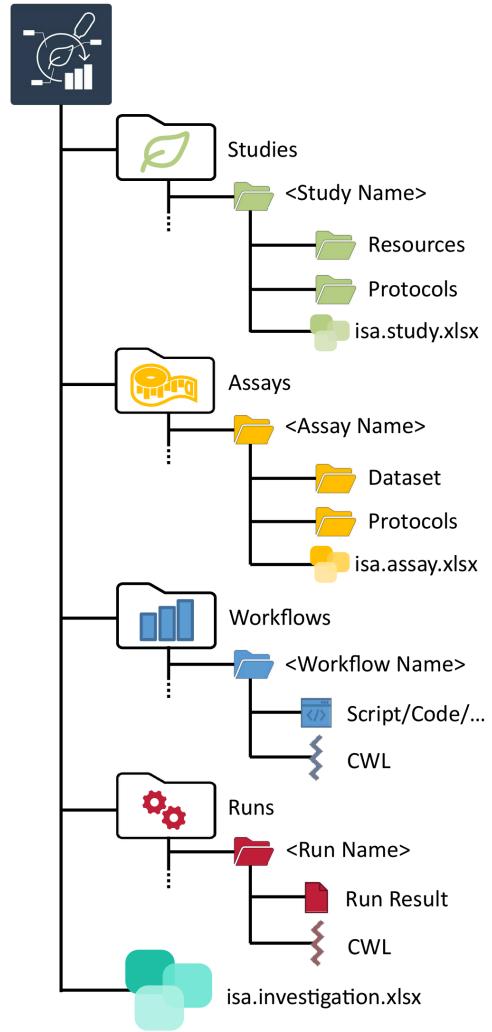
- There is no perfect ARC
- There is no complete ARC
- The only bad ARCs are those that don't exist yet.

🚀 Let's get started, the rest is easy 🚀

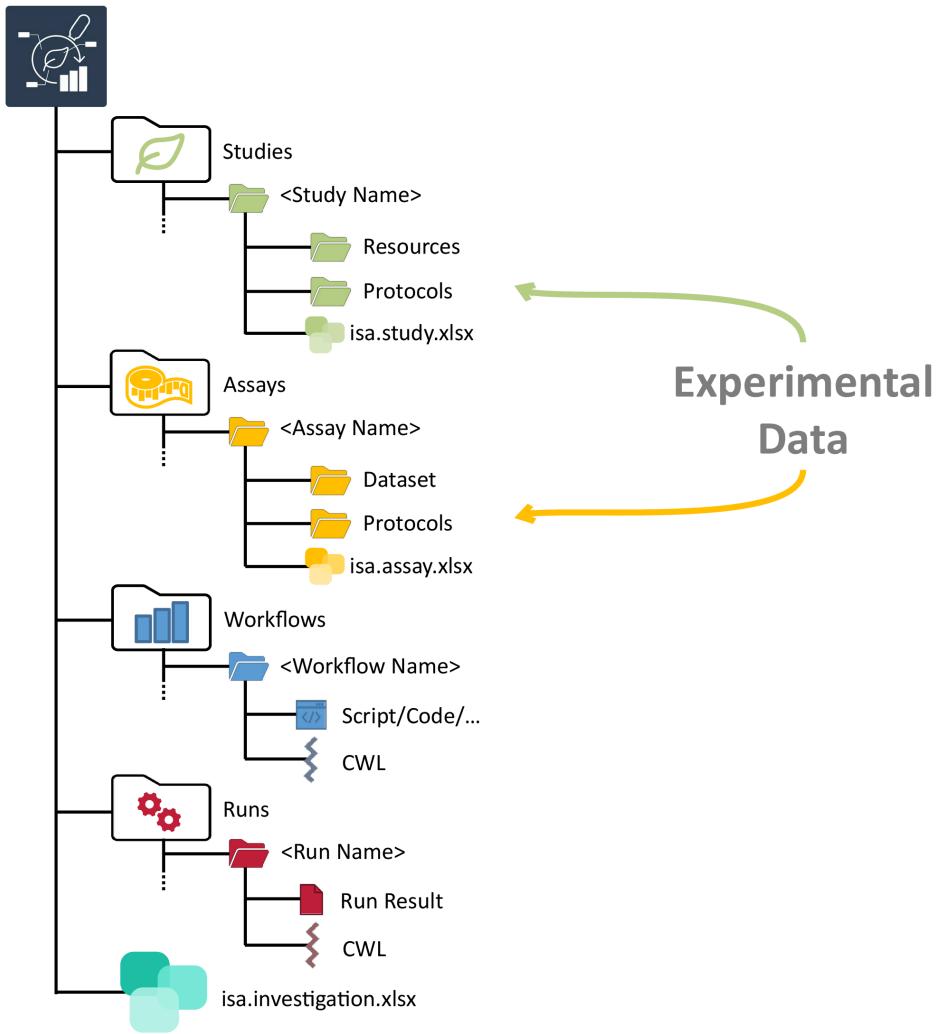
Annotated Research Context (ARC)



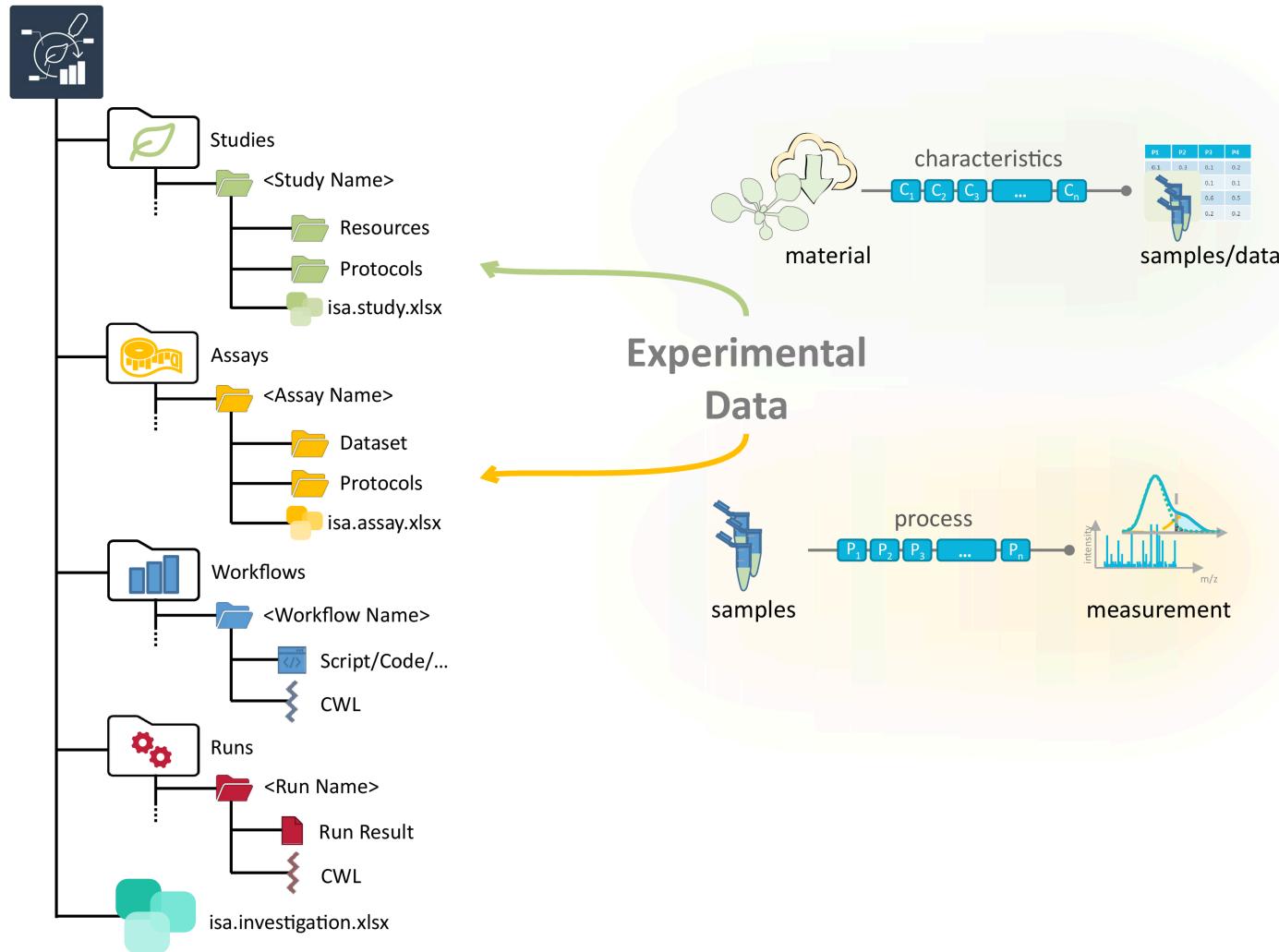
What does an ARC look like?



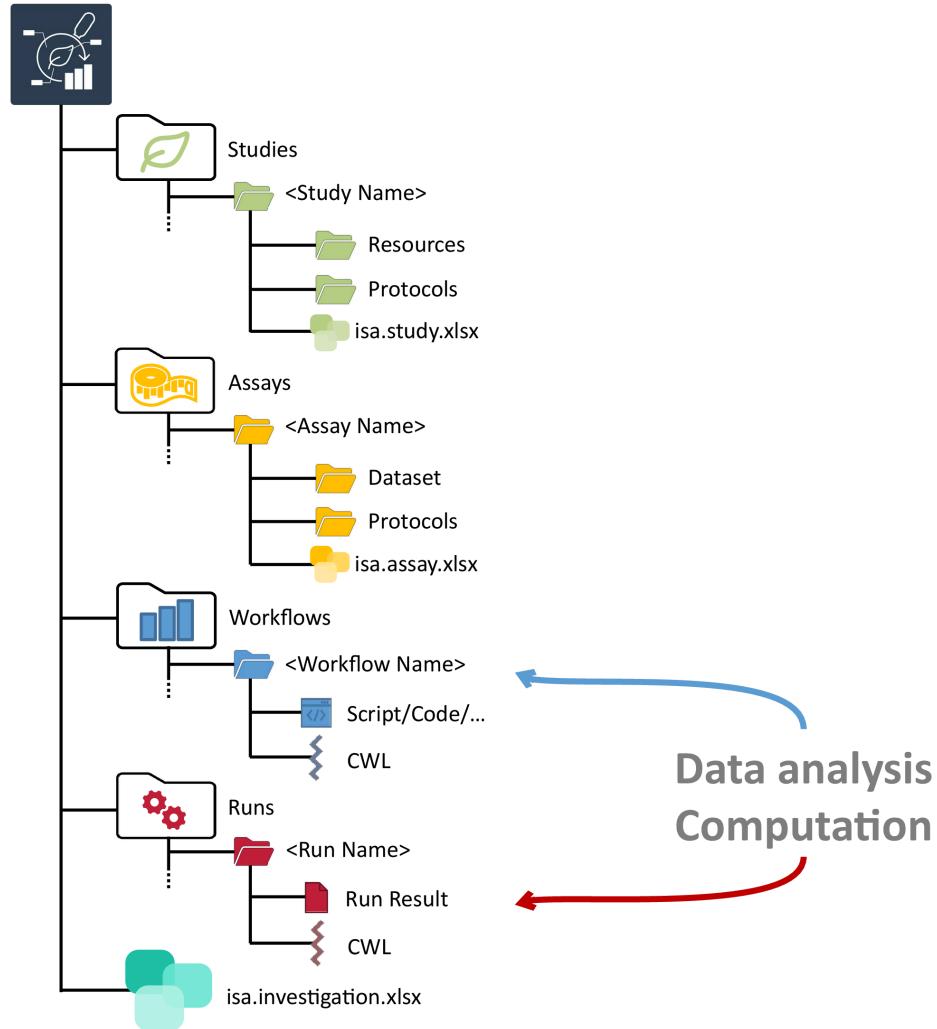
What does an ARC look like?



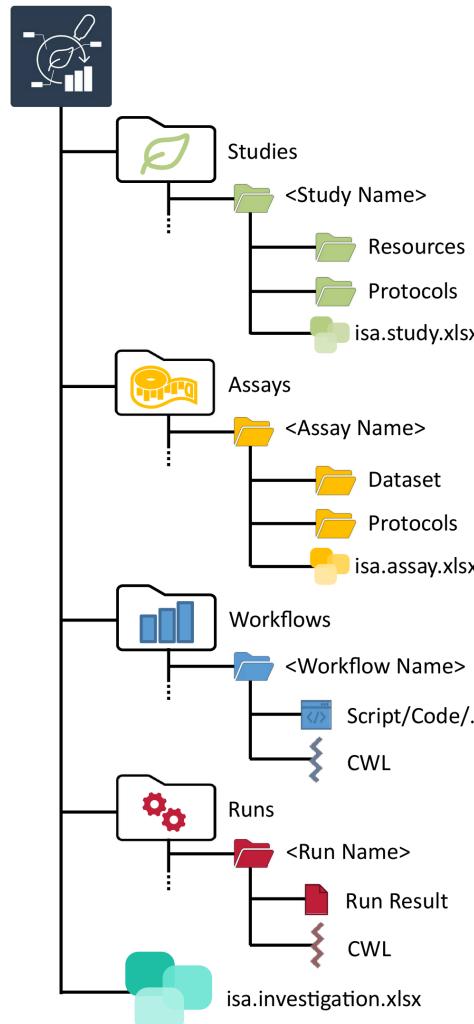
What does an ARC look like?



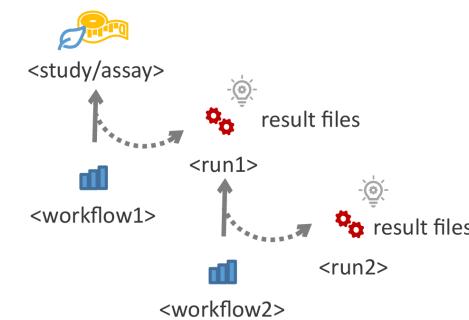
What does an ARC look like?



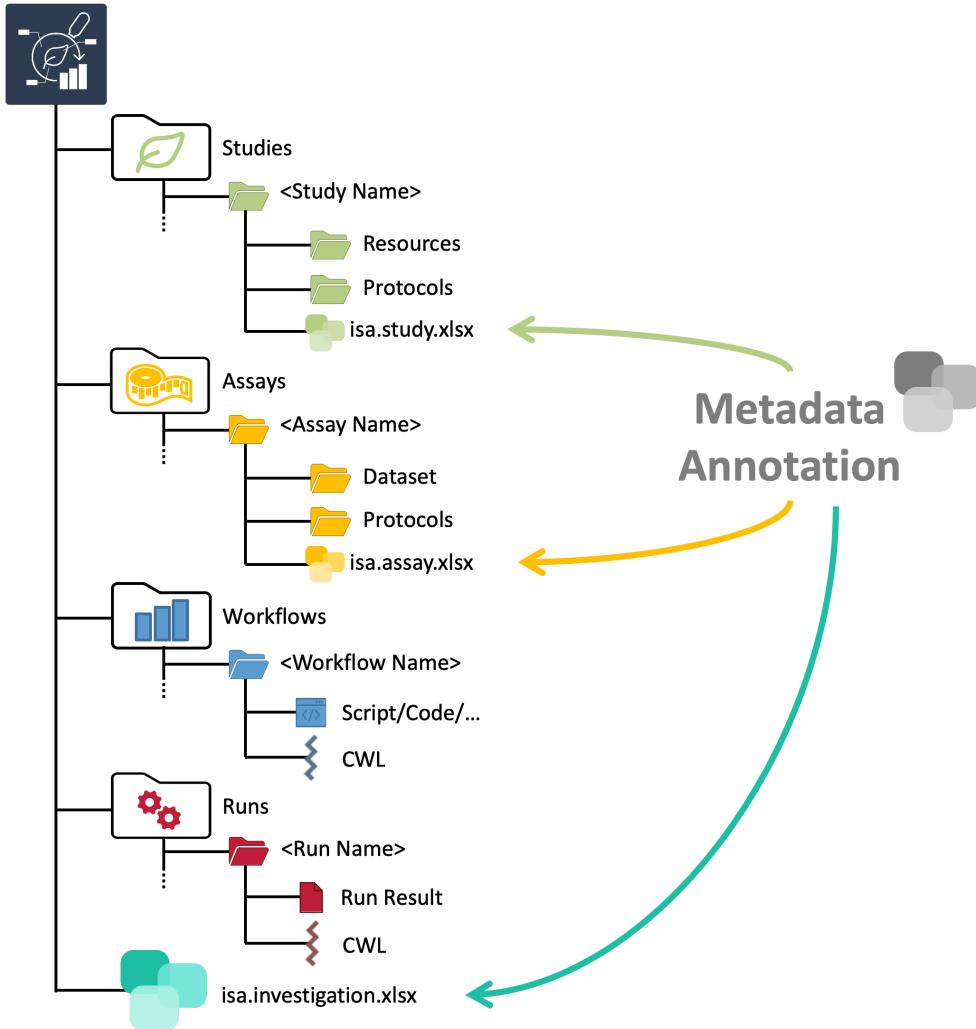
What does an ARC look like?



Data analysis
Computation



What does an ARC look like?





FINDABLE

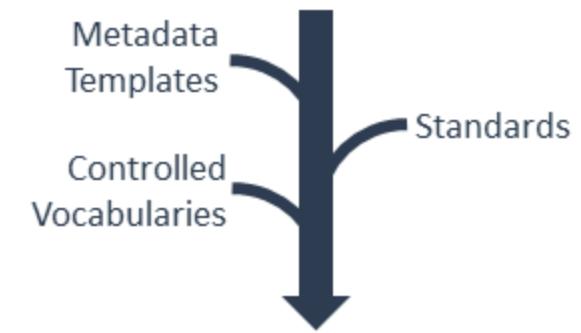
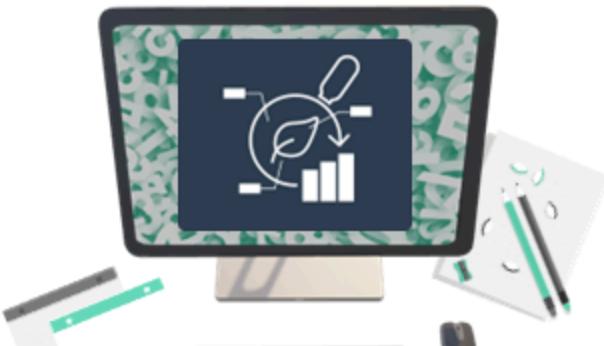
ACCESSIBLE

INTEROPERABLE

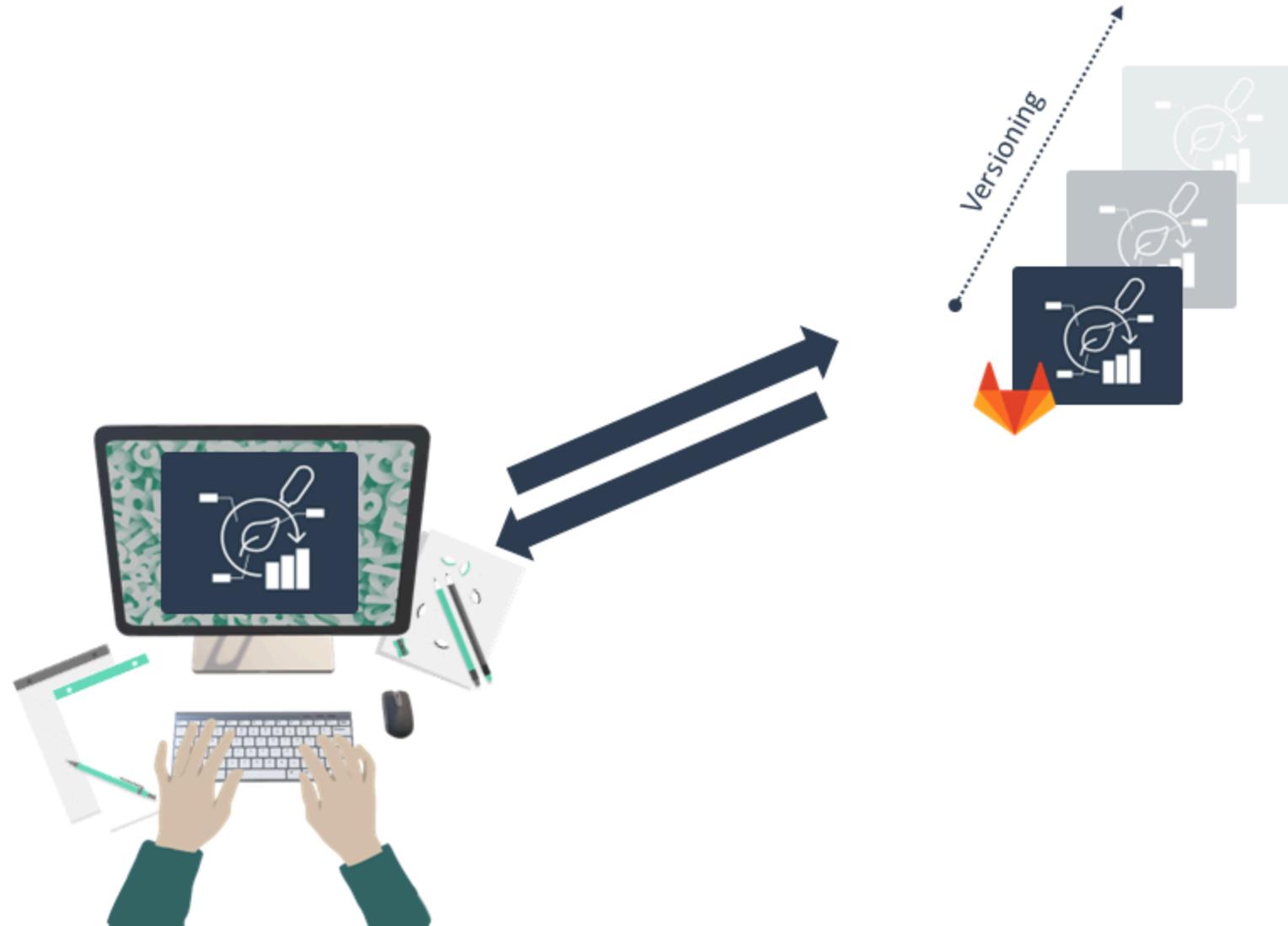
REUSABLE

Metadata
Templates

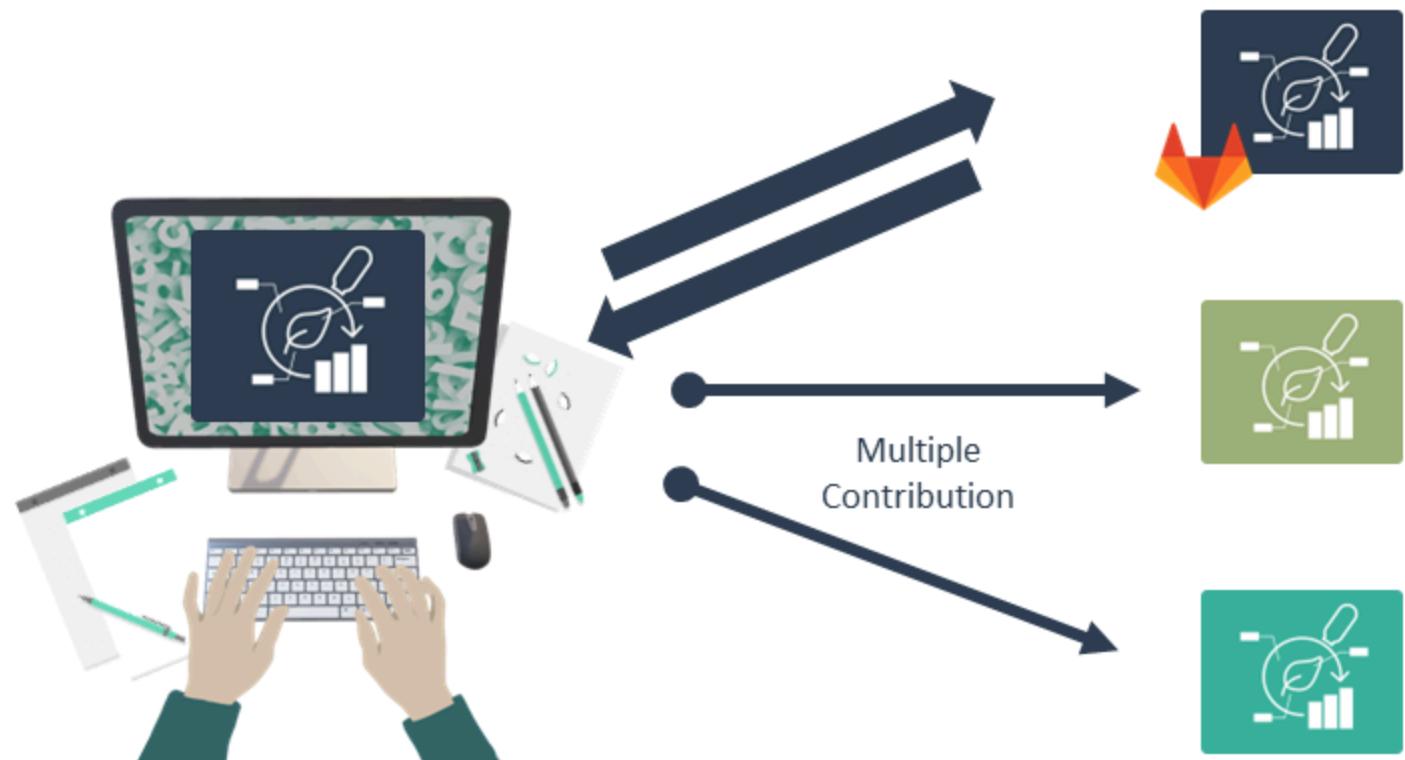
Controlled
Vocabularies



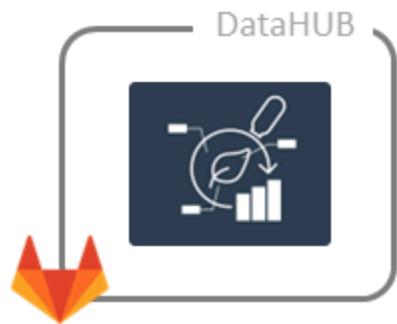






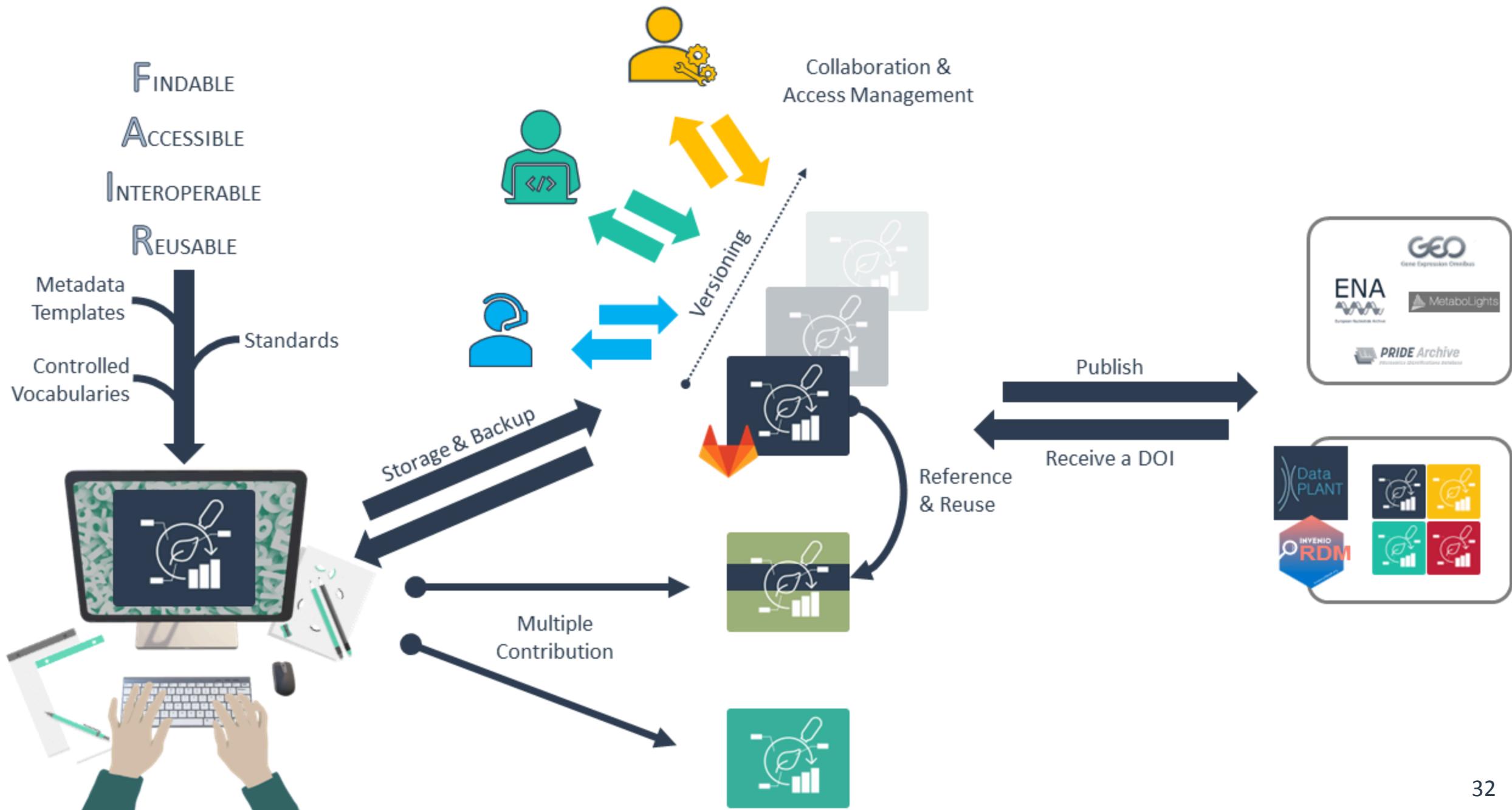






Publish
Receive a DOI





Contributors

Slides presented here include contributions by

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

The ARC Club

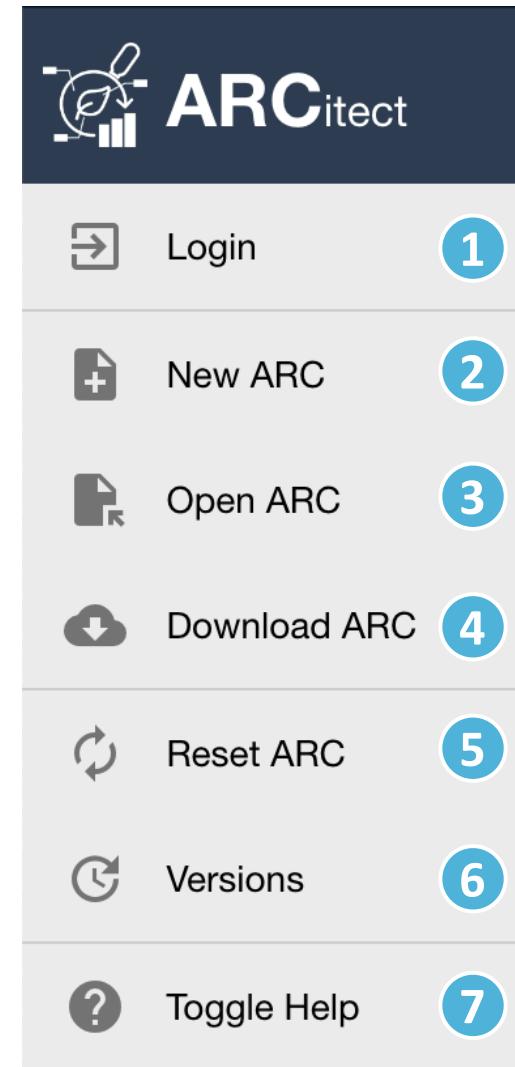
ARCitect QuickStart

Dominik Brilhaus – 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

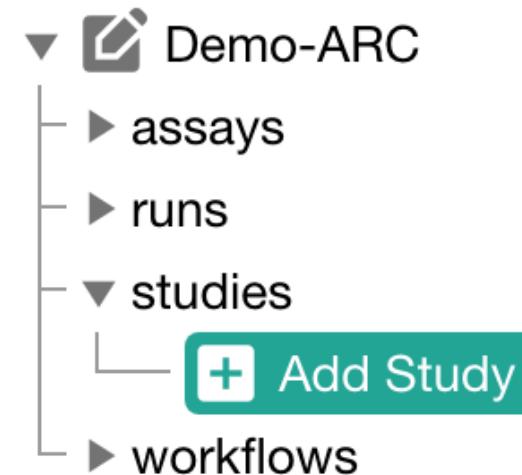
💡 Make sure that no ARC exists at

`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



Add information about your study

In the study panel you can add

- general metadata,
- people, and
- publications

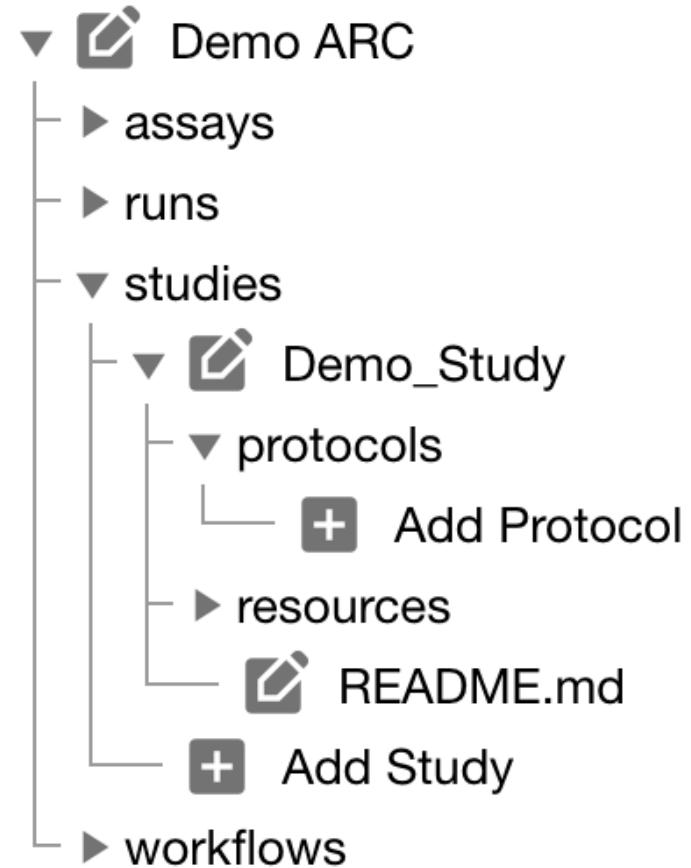
The screenshot shows a study panel interface. At the top left is a file tree:

```
/ Users / dominikbrilhaus / gitlab_dataplant  
/ demo-arcs / Demo-ARC  
  - assays  
  - runs  
  - studies  
    - Demo_Study (selected)  
      - protocols  
      - resources  
      - README.md 0.00 B  
      + Add Study  
    - workflows
```

The main area is titled "Study" and contains "General Meta Data of the Study". It includes fields for Identifier (Demo_Study) and Title (Demo_Study). Below this is a "Description" section with a placeholder "A textual description of the study". There are also "Submission Date" and "Public Release Date" fields, each with a calendar icon and a placeholder "The date the study was released publicly". At the bottom right are "UPDATE" and "RESET" buttons. A navigation bar at the bottom has tabs for "People", "Publications", and "Data".

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

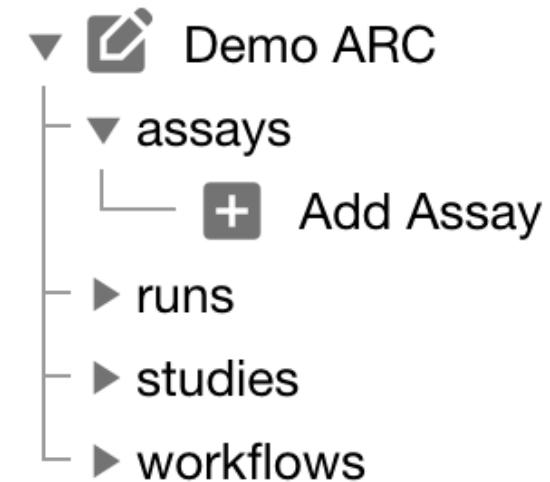


IMPORT PROTOCOL

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

Add Assay

Assay Identifier
Demo Assay

Studies
Create New Study

Demo_Study

 ADD ASSAY

CANCEL

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

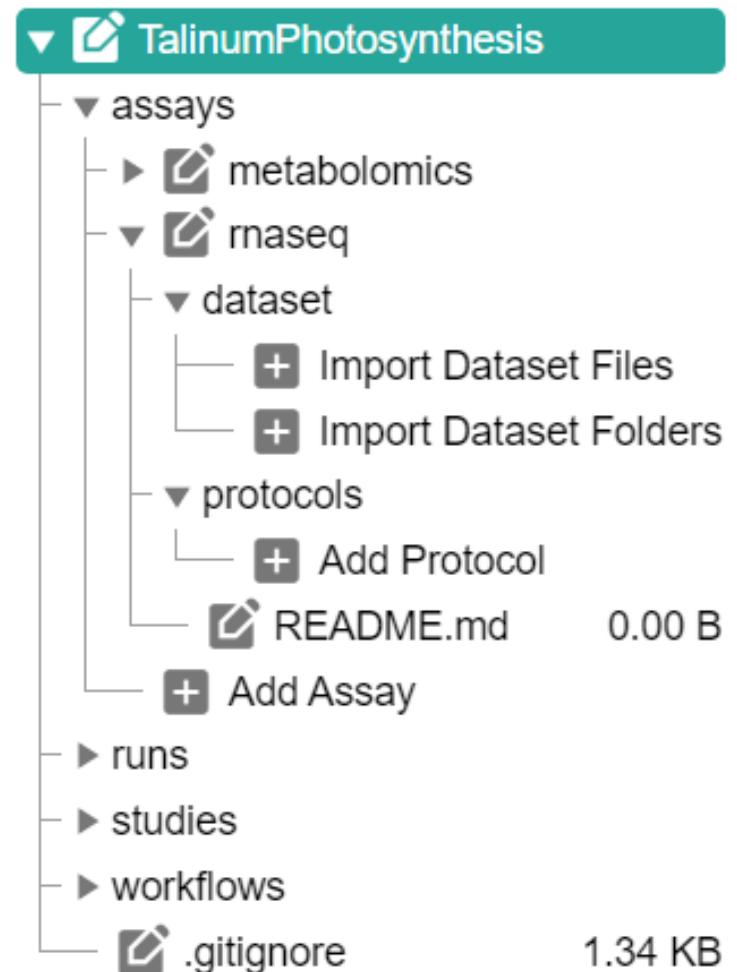
Add protocols and datasets

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.

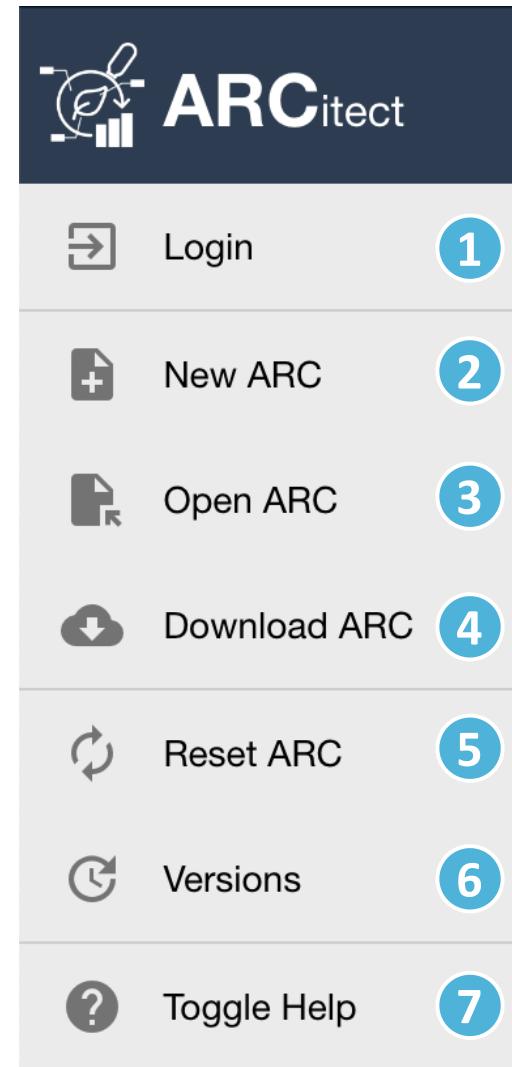
 Depending on the file size, this may take a while. Test this with a small batch of files first.



Login to the DataHUB

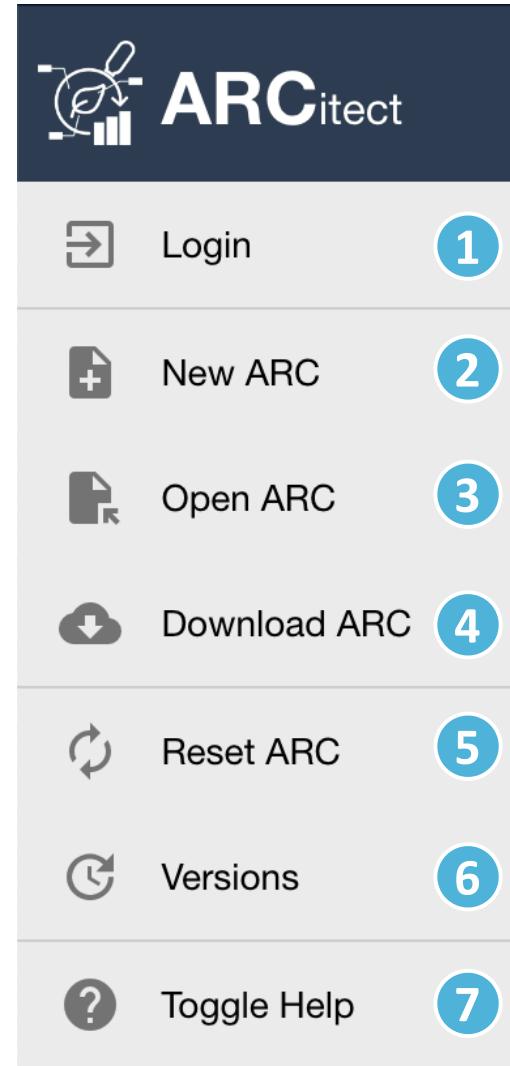
Click **Login** (1) in the sidebar to login to the DataHUB.

 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 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 large text area for 'Commit Message' is present with a placeholder 'A short description of the made changes'. Below it is a 'Changes' section listing modified files: '.arc/', 'assays/', 'isa.investigation.xlsx', 'runs/', 'studies/', and 'workflows/'. At the bottom are buttons for 'REFRESH', 'COMMIT', 'UPLOAD', and 'DOWNLOAD'.

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), 'eMail' (demo@nfdi4plants.org), and 'Remote' (https://git.nfdi4plants.org/demouser/Demo-ARC.git). There is a large 'Commit Message' input field 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

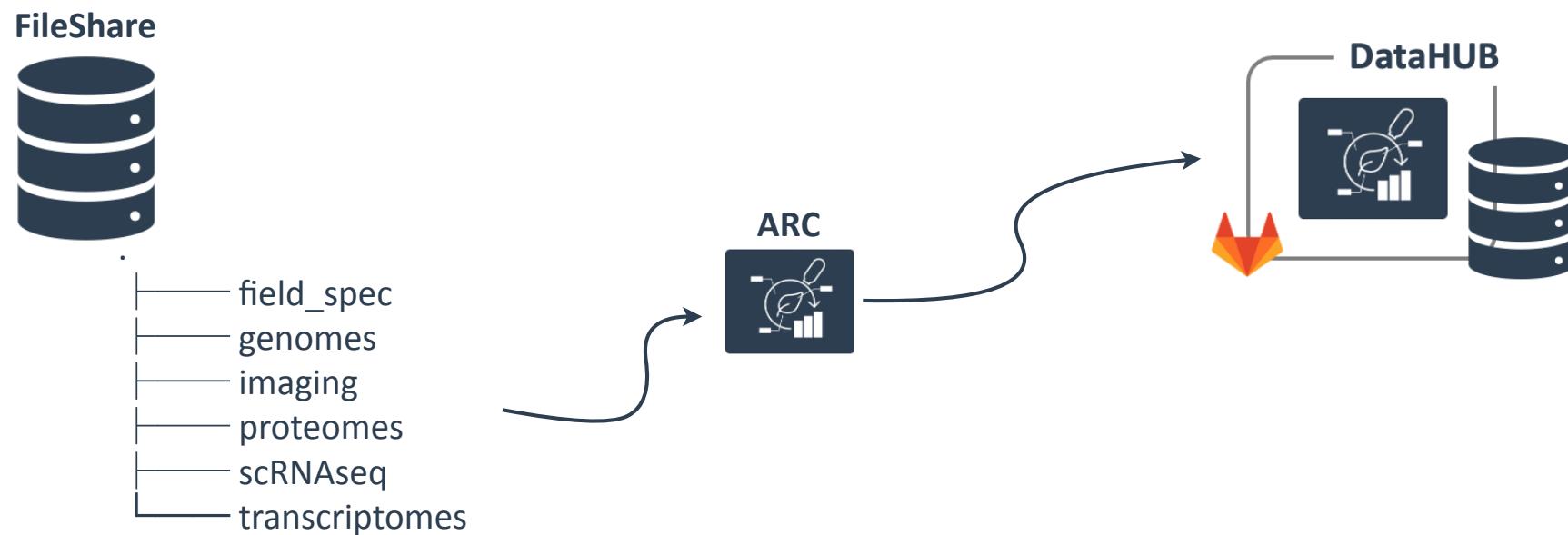
Changes

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

REFRESH COMMIT UPLOAD DOWNLOAD

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 haves

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>
```

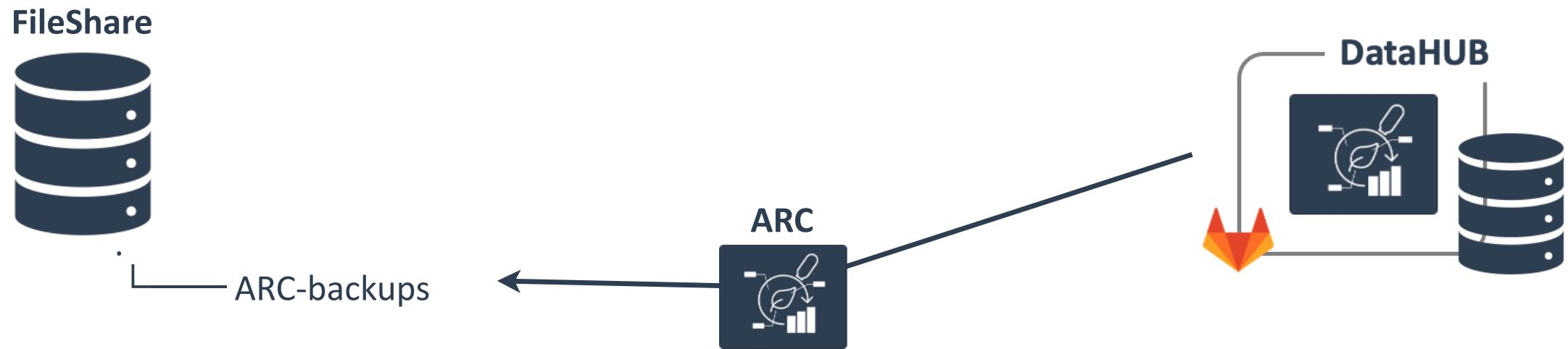
Copy data

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 Brilhaus – 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
- MacOSTextEdit

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

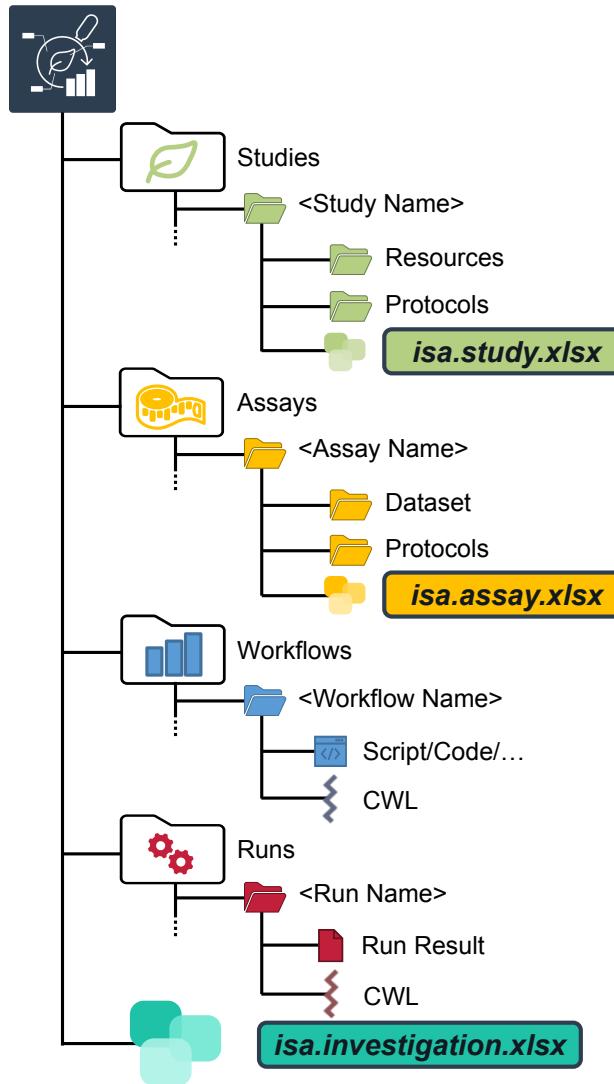
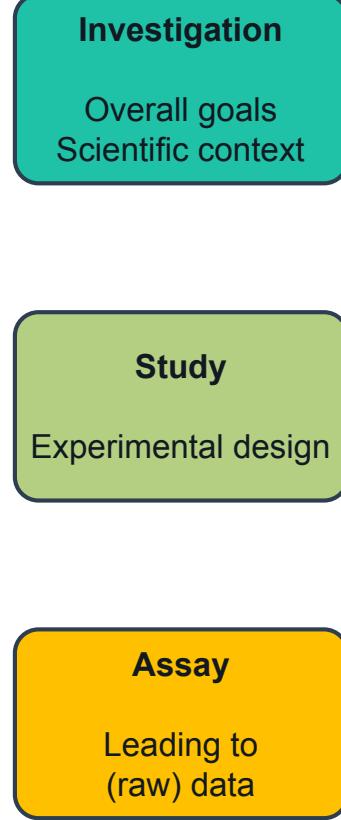
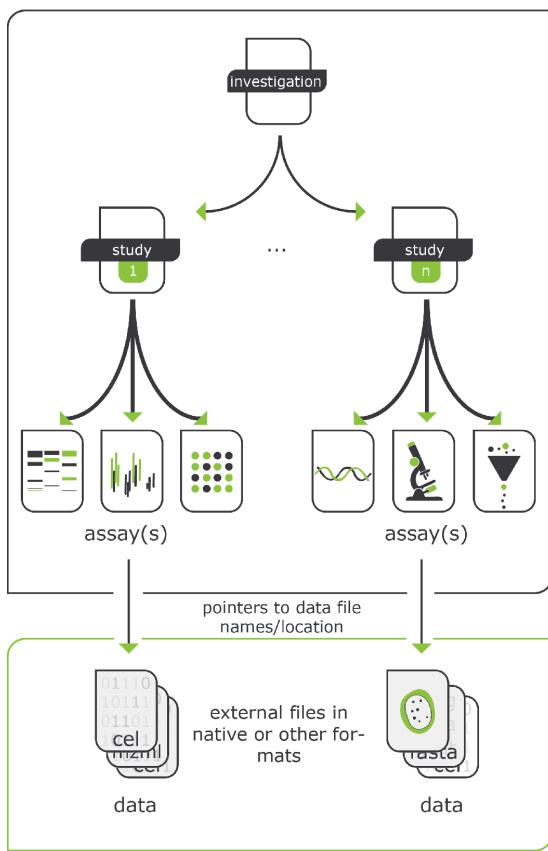
Today we want to

S+ ... annotate the experimental data



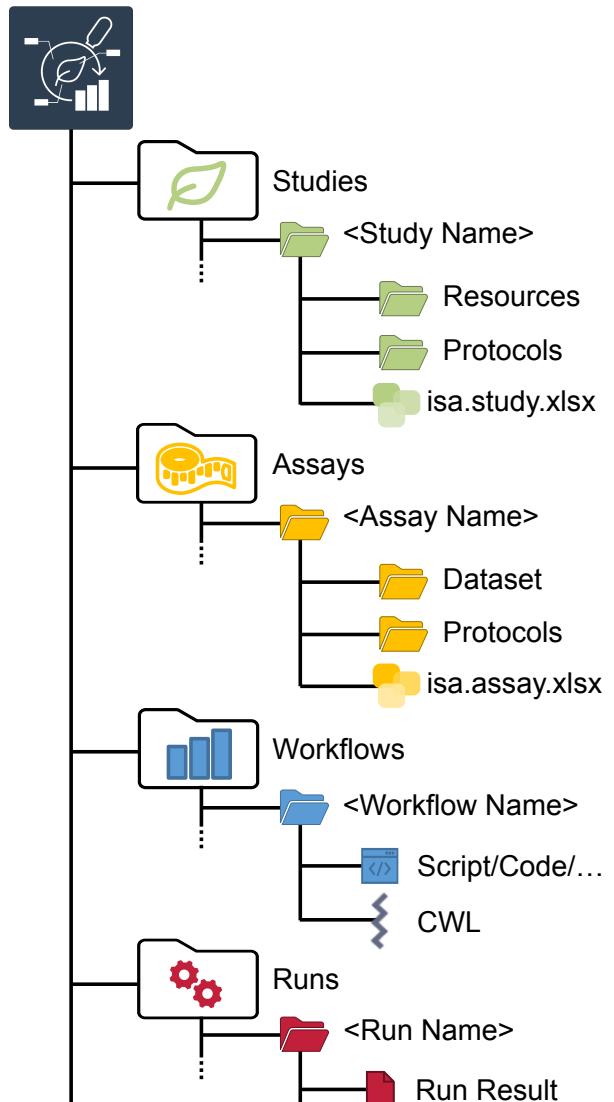
Intro ISA

ARC builds on ISA



ARC builds on ISA

Metadata Annotations



Source Name	Characteristics [Organism]	Sample Name
Ath001	<i>Arabidopsis thaliana</i>	Lre001
Ath002	<i>Arabidopsis thaliana</i>	Lre002
Ath003	<i>Arabidopsis thaliana</i>	Lre003
...

Source Name	Parameter [instrument model]	Raw Data File
Lre001	6130 Quadrupole LC/MS	Lre001.wiff
Lre002	6130 Quadrupole LC/MS	Lre002.wiff
Lre003	6130 Quadrupole LC/MS	Lre003.wiff
...

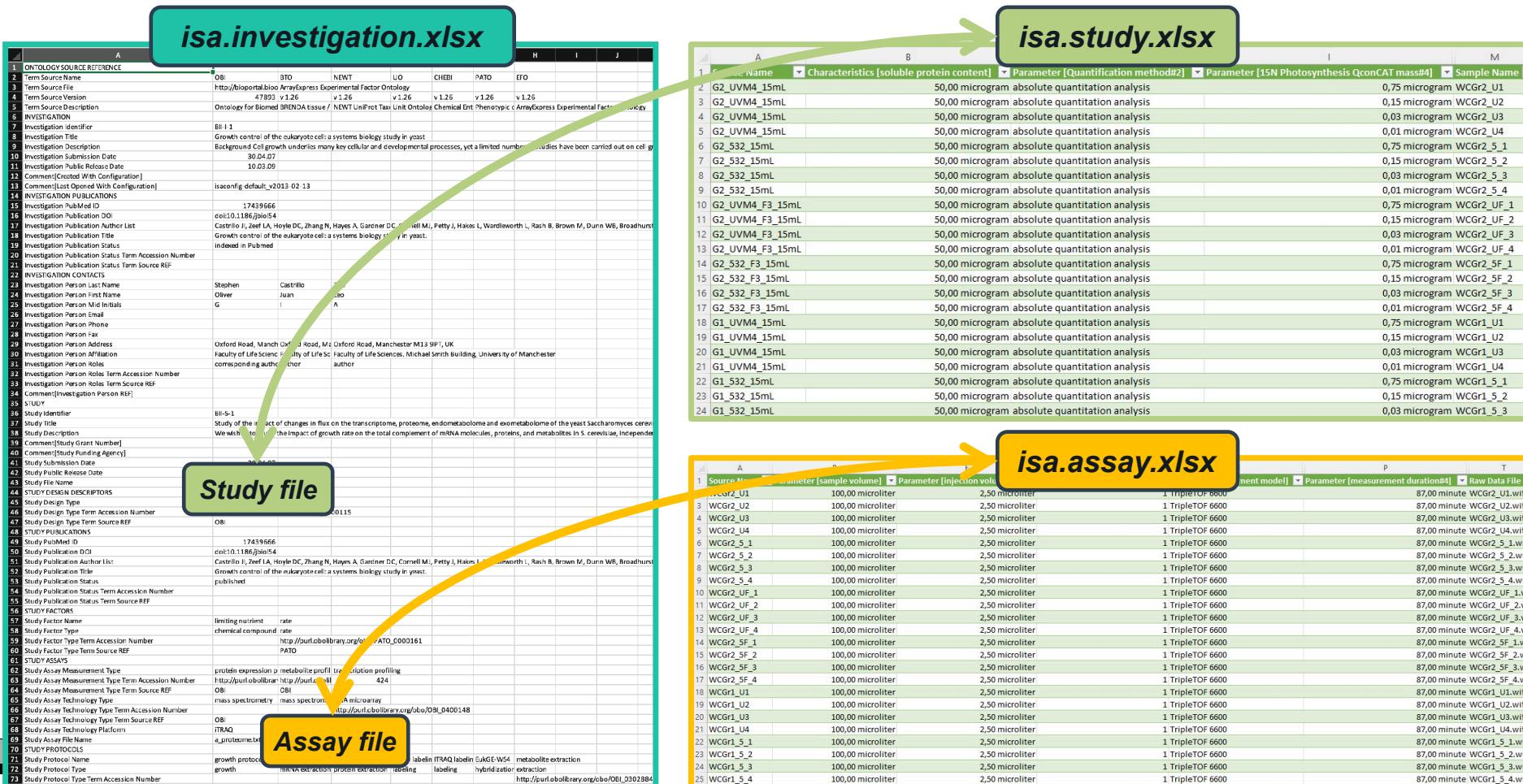
Investigation Title	
Investigation Description	
Investigation Person Last Name	

isa.<>.xlsx files within ARCs

isa.investigation.xlsx

A	B	C	D	E	F	G	H	I	J
ONTOLOGY SOURCE REFERENCE									
1 Term Source Name	OBI	BTO	NENT	LO	CHEBI	PATO	EFO		
2	http://biportal.bioArrayExpress Experimental Factor Ontology								
3 Term Source Version	47893	v1.26	v1.26	v1.26	v1.26	v1.26	v1.26		
4 Term Source Description	Ontology for Biomed BRENDa tissue / NEWT UniProt Tax Unit Ontology Chemical Ent Phenotypic c ArrayExpress Experimental Factor Ontology								
5 INVESTIGATION									
6 Investigation Identifier	BII-1								
7 Investigation Title	Growth control of the eukaryote cell: a systems biology study in yeast								
8 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.								
9 Investigation Start Date	30.04.07								
10 Investigation End Date	10.05.09								
11 Investigation Public Release Date									
12 Comment[Entered With Configuration]	isaconfig default_v2013_02_13								
13 Comment[Last Opened With Configuration]									
14 INVESTIGATION PUBLICATIONS									
15 Investigation PubMed ID	17439666								
16 Investigation Publication DOI	doi:10.1186/jbio54								
17 Investigation Publication Author List	Castroillo J, Zeeb LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell M, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst								
18 Investigation Publication Title	Growth control of the eukaryote 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 Publication Status Term Source URI									
23 Investigation Person Last Name	Stephen	Castroillo	Zeeb						
24 Investigation Person First Name	Oliver	Juan	Lao						
25 Investigation Person Mid Initials	G		A						
26 Investigation Person Email									
27 Investigation Person Phone									
28 Investigation Person Fax									
29 Investigation Person Address	Oxford Road, Manch Oxford Road, M2 Oxford Road, Manchester M13 9PT, UK								
30 Investigation Person Affiliation	Faculty of Life Sci	Faculty of Life Sci	Faculty of Life Sciences	Michael Smith Building	University of Manchester				
31 Investigation Person Roles	corresponding authc author		author						
32 Investigation Person Roles Term Accession Number									
33 Investigation Person Roles Term Source REF									
34 Comment[Investigation Person REF]									
35 STUDY									
36 Study Identifier	BII-S-1								
37 Study Title	Study of the impact of changes in flux on the transcriptome, proteome, endometabolome and exometabolome of the yeast <i>Saccharomyces cerevisiae</i> . We wished to study the impact of growth rate on the total complement of mRNA molecules, proteins, and metabolites in <i>S. cerevisiae</i> . Independent								
38 Study Description									
39 Comment[Study Grant Number]									
40 Comment[Study Funding Agency]									
41 Study Submission Date	30.04.07								
42 Study Public Release Date	10.05.09								
43 Study File Name	S_BII-S-1.txt								
44 STUDY DESIGN DESCRIPTORS									
45 Study Design Type	Intervention design								
46 Study Design Type Term Accession Number	http://purl.obolibrary.org/obo/OBI_0001115								
47 Study Design Type Term Source REF	OBI								
48 STUDY PUBLICATIONS									
49 Study PubMed ID	17439666								
50 Study Publication DOI	doi:10.1186/jbio54								
51 Study Publication Author List	Castroillo J, Zeeb LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell M, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst								
52 Study Publication Title	Growth control of the eukaryote 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	http://purl.obolibrary.org/obo/PATO_000161								
60 Study Factor Type Term Source REF	PATO								
61 STUDY ASSAYS									
62 Study Assay Measurement Type	protein expression	metabolite profile	transcription profiling						
63 Study Assay Measurement Type Term Accession Number	http://purl.obolibrary.org/obo/OBI_0400148								
64 Study Assay Measurement Type Term Source REF	OBI	OBI	OBI						
65 Study Assay Technology Type	mass spectrometry	mass spectrometry	RNA microarray						
66 Study Assay Technology Type Term Accession Number	http://purl.obolibrary.org/obo/OBI_0400148								
67 Study Assay Technology Type Term Source REF	OBI	OBI	OBI						
68 Study Assay Technology Platform	ITRAQ	LC-MS/MS	Affymetrix						
69 Study Assay File Name	a_proteome.txt	a_metabolome.txt	a_transcriptome.txt						
70 STUDY PROTOCOLS									
71 Study Protocol Name	growth	mRNA extraction	protein extraction	biotin labelin	ITRAQ labelin	EukGE-W54	metabolite extraction		
72 Study Protocol Type	growth	mRNA extraction	protein extraction	labeling	loading	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 tak. This was done using Enzo! For each target, a hybridisation cocktail was made using the								
76 Study Protocol URI									
77 Study Protocol									
78 Study Protocol Description									
79 Study Protocol									
80 Study Protocol									
81 Study Protocol									
82 Study Protocol									
83 Study Protocol									
84 Study Protocol									
85 Study Protocol									
86 Study Protocol									
87 Study Protocol									
88 Study Protocol									
89 Study Protocol									
90 Study Protocol									
91 Study Protocol									
92 Study Protocol									
93 Study Protocol									
94 Study Protocol									
95 Study Protocol									
96 Study Protocol									
97 Study Protocol									
98 Study Protocol									
99 Study Protocol									
100 Study Protocol									
101 Study Protocol									
102 Study Protocol									
103 Study Protocol									
104 Study Protocol									
105 Study Protocol									
106 Study Protocol									
107 Study Protocol									
108 Study Protocol									
109 Study Protocol									
110 Study Protocol									
111 Study Protocol									
112 Study Protocol									
113 Study Protocol									
114 Study Protocol									
115 Study Protocol									
116 Study Protocol									
117 Study Protocol									
118 Study Protocol									
119 Study Protocol									
120 Study Protocol									
121 Study Protocol									
122 Study Protocol									
123 Study Protocol									
124 Study Protocol									
125 Study Protocol									
126 Study Protocol									
127 Study Protocol									
128 Study Protocol									
129 Study Protocol									
130 Study Protocol									
131 Study Protocol									
132 Study Protocol									
133 Study Protocol									
134 Study Protocol									
135 Study Protocol									
136 Study Protocol									
137 Study Protocol									
138 Study Protocol									
139 Study Protocol									
140 Study Protocol									
141 Study Protocol									
142 Study Protocol									
143 Study Protocol									
144 Study Protocol									
145 Study Protocol									
146 Study Protocol									
147 Study Protocol									
148 Study Protocol									
149 Study Protocol									
150 Study Protocol									
151 Study Protocol									
152 Study Protocol									
153 Study Protocol									
154 Study Protocol									
155 Study Protocol									
156 Study Protocol									
157 Study Protocol									
158 Study Protocol									
159 Study Protocol									
160 Study Protocol									
161 Study Protocol									
162 Study Protocol									
163 Study Protocol									
164 Study Protocol									
165 Study Protocol									
166 Study Protocol									
167 Study Protocol									
168 Study Protocol									
169 Study Protocol									
170 Study Protocol									
171 Study Protocol									
172 Study Protocol									
173 Study Protocol									
174 Study Protocol									
175 Study Protocol									
176 Study Protocol									
177 Study Protocol									
178 Study Protocol									
179 Study Protocol									
180 Study Protocol									
181 Study Protocol									
182 Study Protocol									
183 Study Protocol									
184 Study Protocol									
185 Study Protocol									
186 Study Protocol									
187 Study Protocol									
188 Study Protocol									
189 Study Protocol									
190 Study Protocol									

Study and assay files are registered in the investigation file



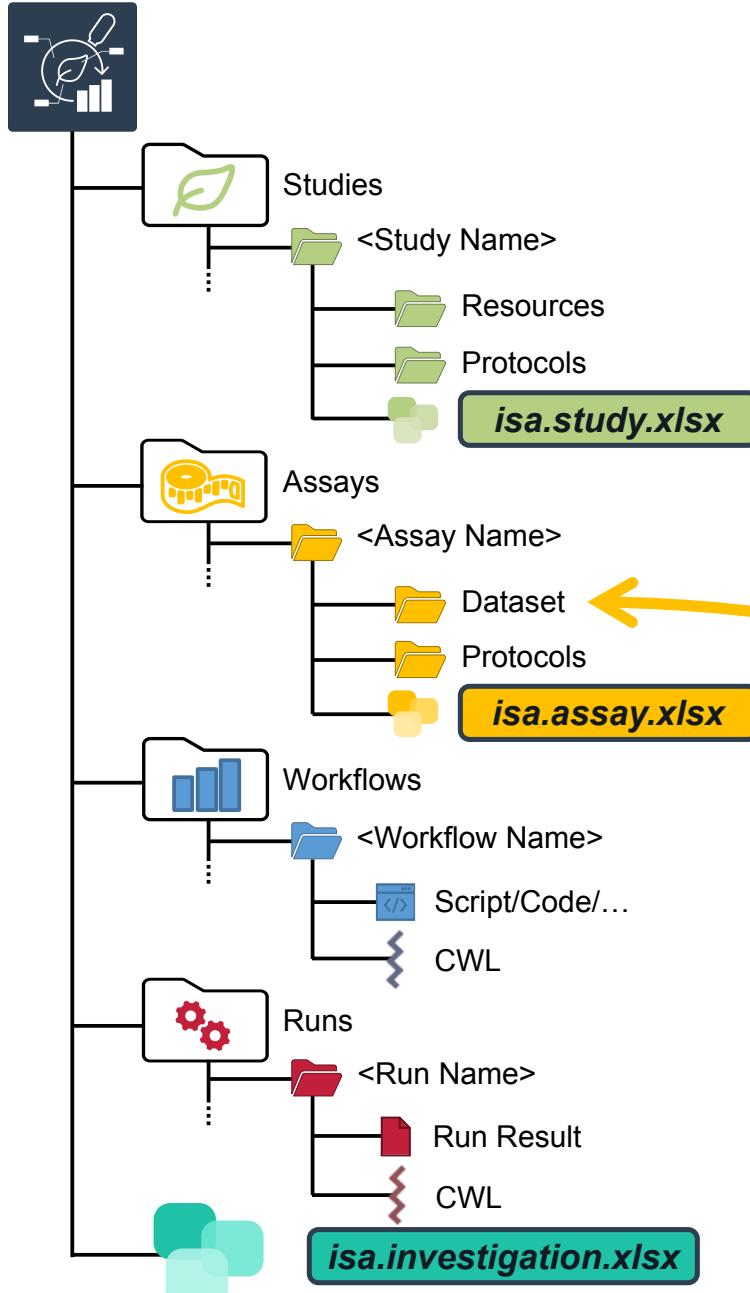
The output of a study or assay file can function as input for a new isa.assay.xlsx

Output building blocks:

- Sample Name
- Raw Data File
- Derived Data File

	A	B	C	D	E	F	G	H	I	J	K	L	M
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

	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								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
6	WCGr2_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_1.wiff
7	WCGr2_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_2.wiff
8	WCGr2_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_3.wiff
9	WCGr2_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr2_5_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_5_1	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_1.wiff
23	WCGr1_5_2	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_2.wiff
24	WCGr1_5_3	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_3.wiff
25	WCGr1_5_4	100,00 microliter	2,50 microliter	1	TripleTOF 6600								87,00 minute WCGr1_5_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



isa.investigation.xlsx

This screenshot shows the contents of the **isa.investigation.xlsx** spreadsheet, which includes three main tabs:

- Study Data (Top Tab)**: Shows a list of samples and their characteristics. The columns include Source Name, Characteristics [soluble protein content], Parameter [Quantification method#2], Parameter [15N Photosynthesis QconCAT mass#4], and Sample Name. The data is as follows:

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_532_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

- Assay Data (Middle Tab)**: Shows a list of samples and their parameters. The columns include Source Name, Parameter [sample volume], Parameter [injection vol.], and Raw Data File. The data is as follows:

Source Name	Parameter [sample volume]	Parameter [injection vol.]	Raw Data File
WCGr2_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_UF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_UF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_SF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_SF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_SF_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr2_SF_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_U1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_U2	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_U3	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_U4	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_5_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_5_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_5_3	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_5_4	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_UF_1	100,00 microliter	2,50 microliter	1 TripleTOF 6600
WCGr1_UF_2	100,00 microliter	2,50 microliter	1 TripleTOF 6600

- Investigation Data (Bottom Tab)**: Shows a list of raw data files. The columns include Source Name, Parameter [sample volume], Parameter [injection vol.], and Raw Data File. The data is as follows:

Source Name	Parameter [sample volume]	Parameter [injection vol.]	Raw Data File
WCGr2_5_1.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr2_5_2.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr2_5_3.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr2_5_4.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_U1.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_U2.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_U3.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_U4.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_5_1.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_5_2.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_5_3.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_5_4.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_UF_1.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_UF_2.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_UF_3.wiff	100,00 microliter	2,50 microliter	WIFF File
WCGr1_UF_4.wiff	100,00 microliter	2,50 microliter	WIFF File

Annotation Building Block types

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

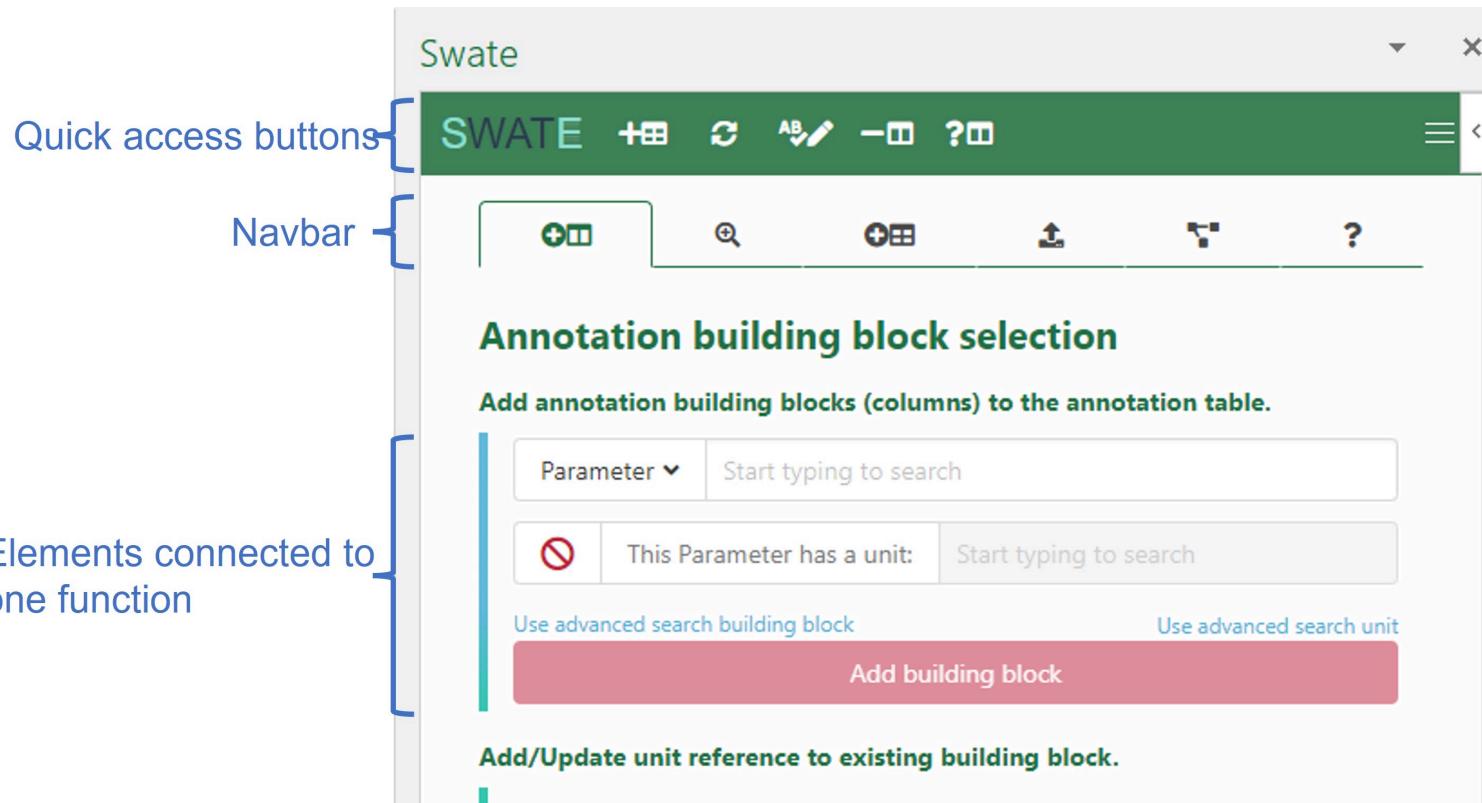
The screenshot shows a Microsoft Excel spreadsheet titled "6130 Quadrupole LC/MS". The table has columns labeled A through AB. Annotations highlight specific columns:

- Protocol Type/Protocol REF**: Points to the second column.
- Characteristic**: Points to the third column.
- Factor**: Points to the fourth column.
- Component**: Points to the fifth column.
- Sample Name/Raw Data File/Derived Data File**: Points to the sixth column.
- New Parameter**: Points to the seventh column.

A sidebar on the right is titled "Building Blocks" and lists categories like "Instrument model", "Instrument Model", "Instrument", and "Agilent instrument model". A note at the bottom right states: "Parameter columns describe steps in your experimental workflow, e.g. the centrifugation time or the temperature of a heating block. Note: 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."

Swate hands-on with demo data

Swate Overview



Major areas of the Swate user interface.

Let's annotate the plant samples first

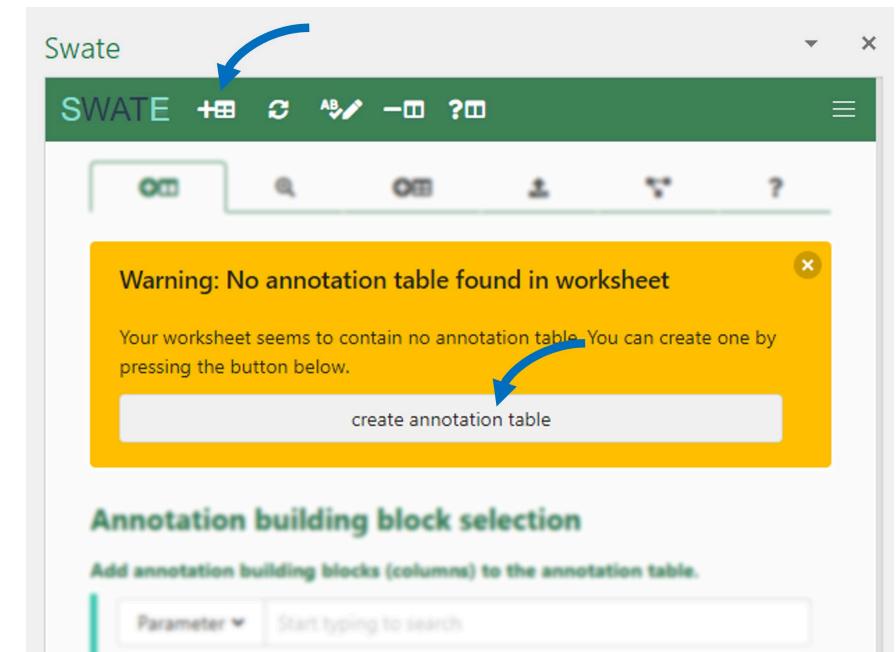
1. Navigate to the demo ARC.
2. Open the lab notes `studies/talinum_drought/protocols/plant_material.txt` in a text editor.
3. Open the empty `studies/talinum_drought/isa.study.xlsx` workbook in Excel.

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

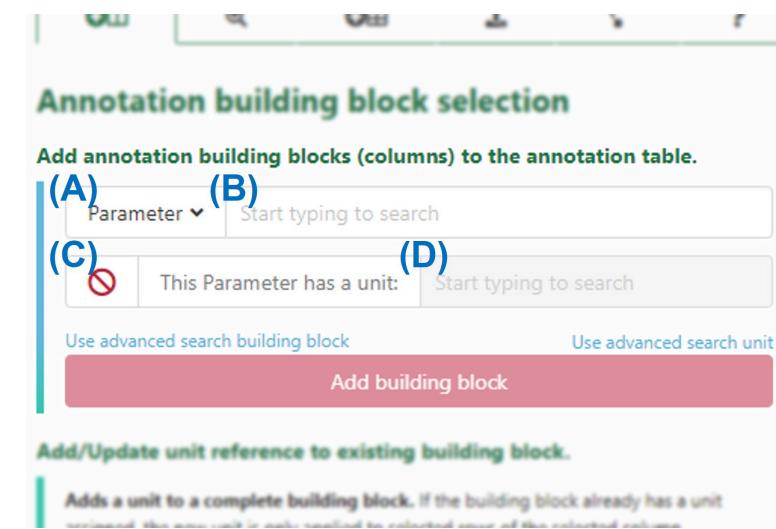
💡 Only one annotation table can be added per Excel sheet



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.

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:

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

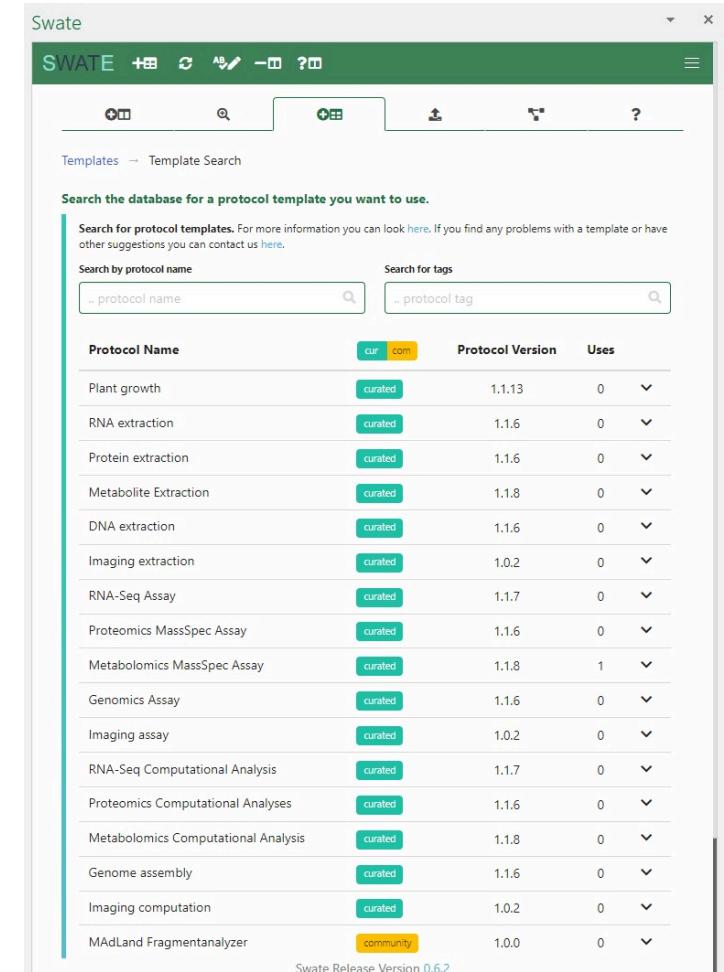
1. Navigate to the demo ARC.
2. Open the lab notes `assays/rnaseq/protocols/RNA_extraction.txt` in a text editor.
3. Open the empty `assays/rnaseq/isa.assay.xlsx` workbook in Excel.

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' screen in the Swate software. The interface includes a top navigation bar with icons for file operations like New, Open, Save, and Print. Below the navigation is a search bar with placeholder text 'Search the database for a protocol template you want to use.' and two search fields: 'Search by protocol name' and 'Search for tags'. A large table lists various protocol templates with columns for 'Protocol Name', 'Protocol Version', and 'Uses'. Each row includes a status indicator (curated or community) and a dropdown menu. The table lists protocols such as Plant growth, RNA extraction, Protein extraction, Metabolite Extraction, DNA extraction, Imaging extraction, RNA-Seq Assay, Proteomics MassSpec Assay, Metabolomics MassSpec Assay, Genomics Assay, Imaging assay, RNA-Seq Computational Analysis, Proteomics Computational Analyses, Metabolomics Computational Analysis, Genome assembly, Imaging computation, and MADLand Fragmentanalyzer. The bottom right corner of the window displays 'Swate Release Version 0.6.2'.

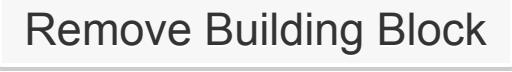
Protocol Name	Protocol Version	Uses
Plant growth	1.1.13	0
RNA extraction	1.1.6	0
Protein extraction	1.1.6	0
Metabolite Extraction	1.1.8	0
DNA extraction	1.1.6	0
Imaging extraction	1.0.2	0
RNA-Seq Assay	1.1.7	0
Proteomics MassSpec Assay	1.1.6	0
Metabolomics MassSpec Assay	1.1.8	1
Genomics Assay	1.1.6	0
Imaging assay	1.0.2	0
RNA-Seq Computational Analysis	1.1.7	0
Proteomics Computational Analyses	1.1.6	0
Metabolomics Computational Analysis	1.1.8	0
Genome assembly	1.1.6	0
Imaging computation	1.0.2	0
MADLand Fragmentanalyzer	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")
```

Contributors

Slides presented here include contributions by

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

