



ARC WORKSHOP

Angela Kranz, Stella Eggels
6th of February 2025

Time schedule

09:00 - 10:00 Intro RDM, NFDI, DataPLANT & DataPLANT MetadataToolchain

10:00 - 12:15 Intro ARCitect and demonstration

12:15 - 13:30 Lunch break

13:30 - 14:00 Intro Metadata Annotation and demonstration

14:00 - 14:45 Create your own ARC

14:45 - 15:00 Short break

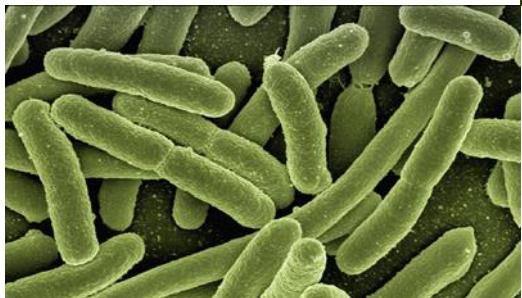
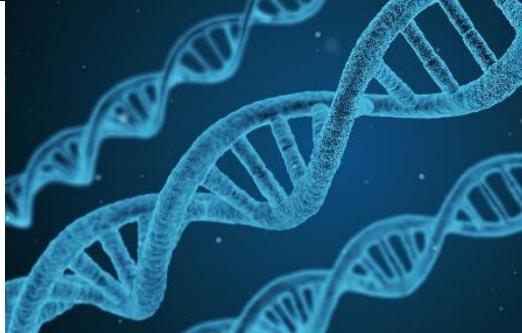
15:00 - 15:45 Intro DataHUB and demonstration

15:45 - 16:45 Continue with your own ARC

16:45 - 17:00 Q & A and wrap-up

Intro to research data management and DataPLANT

What kind of research data do you work with?



Research Data Management

Research data management (RDM) includes all activities associated with

- processing,
- storage,
- preserving and
- publication

of research data.

RDM accompanies the research process from the initial planning to the archiving, re-use or deletion of the data

Research Data Management

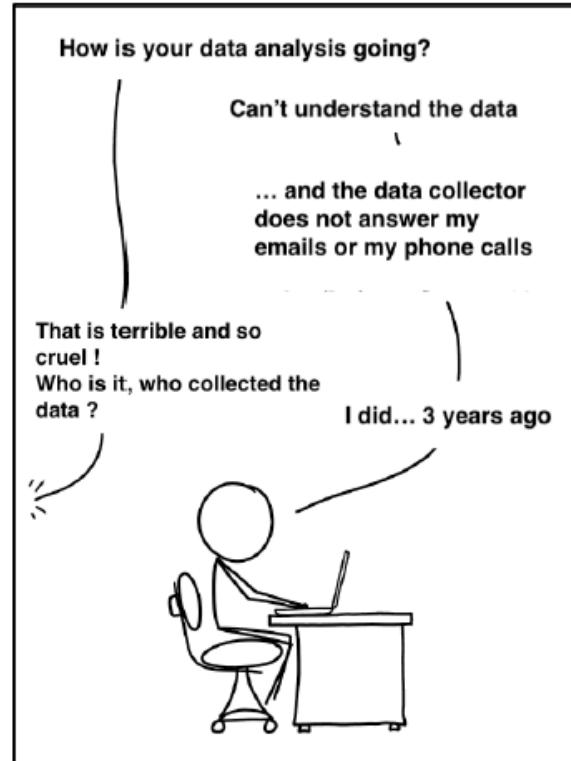
Take good care of your data!

Why should I?



Why RDM?

- Increase transparency
- Make data accessible
- Save time (writing, reusing)
- Reduce the risk of data loss
- Optimize the costs
- Facilitate future reuse and sharing
- Improve citations

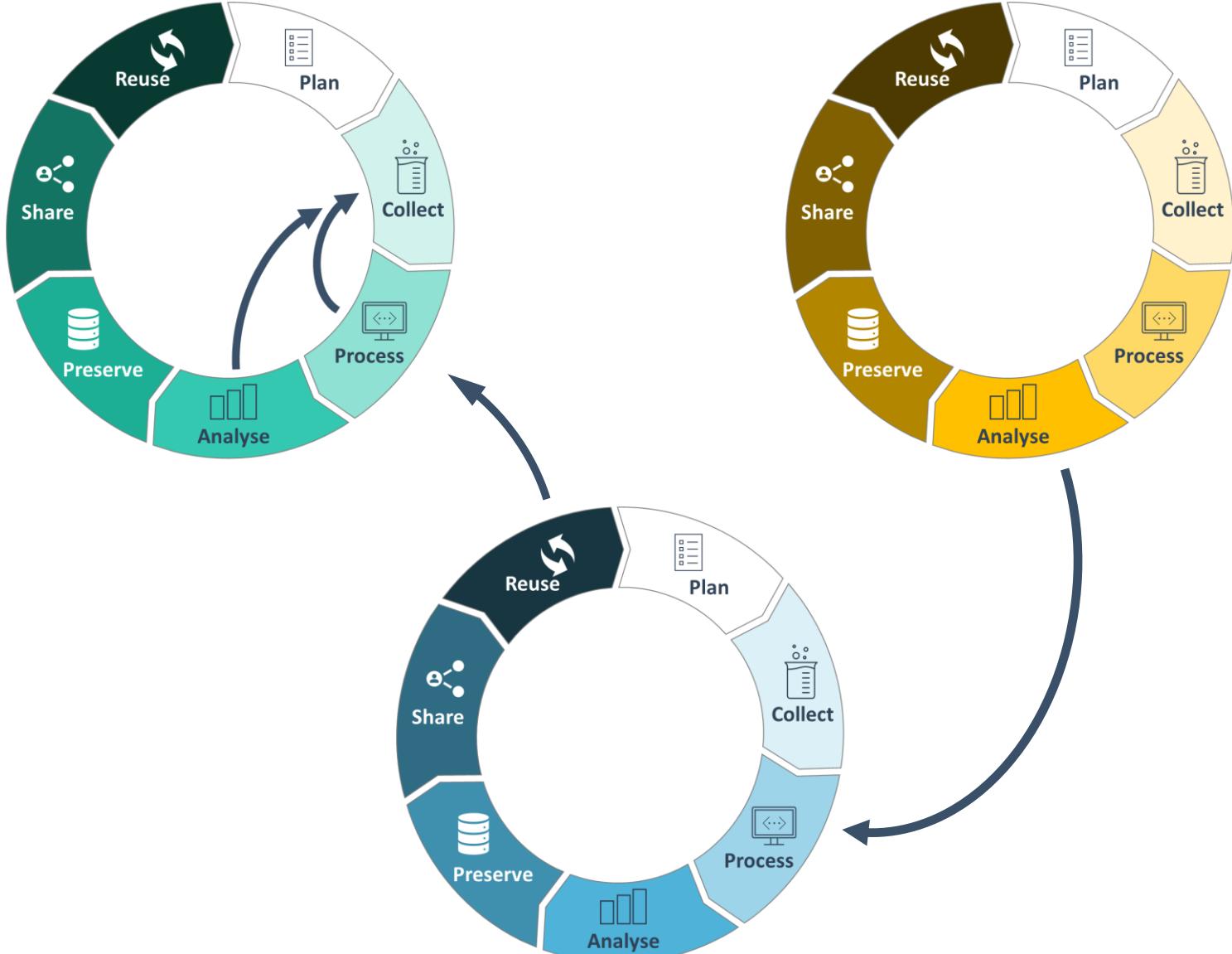


Your first collaborators
are your future selves,
be nice to them !

your future self, by Julian Colomb, CC-BY-NC, derived from NDPA Normal File Format, CC-BY-NC, by Randall Munroe

derived by Julian Colomb from <https://xkcd.com/1459/>





The FAIR principles

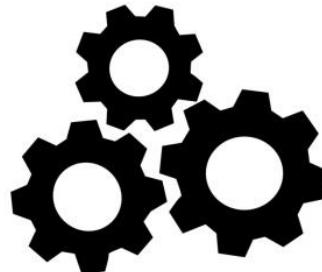
F
indable



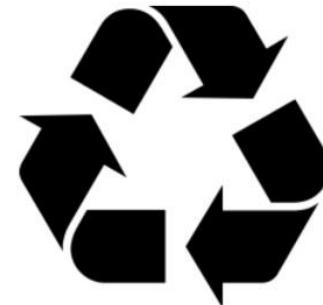
A
ccessible



I
nteroperable



R
eusable



Source: Pundir, Sangya. https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg [last access: 27.09.2018], CC-BY-SA-4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.en>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand J. Aalberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.

Is your data FAIR?

Findable | Accessible | Interoperable | Reusable

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



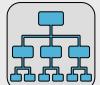
Benefits of being FAIR



Findable



Accessible

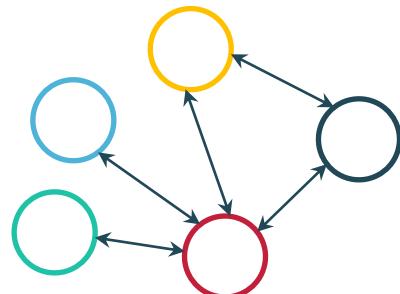


Interoperable

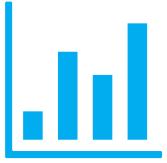


Reusable

Easier collaboration & sharing



Increased findability and visibility



Reproducibility



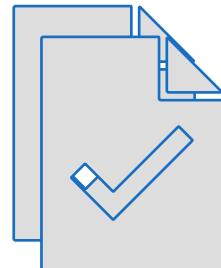
Added-value to the research community



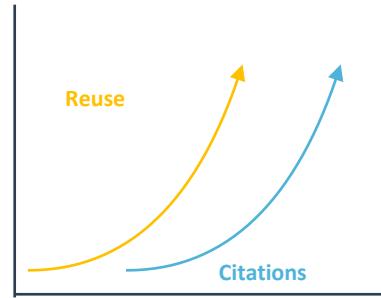
EMBL-EBI



Compliance with funding policies

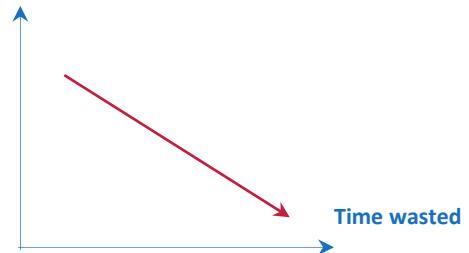


Receive due credit

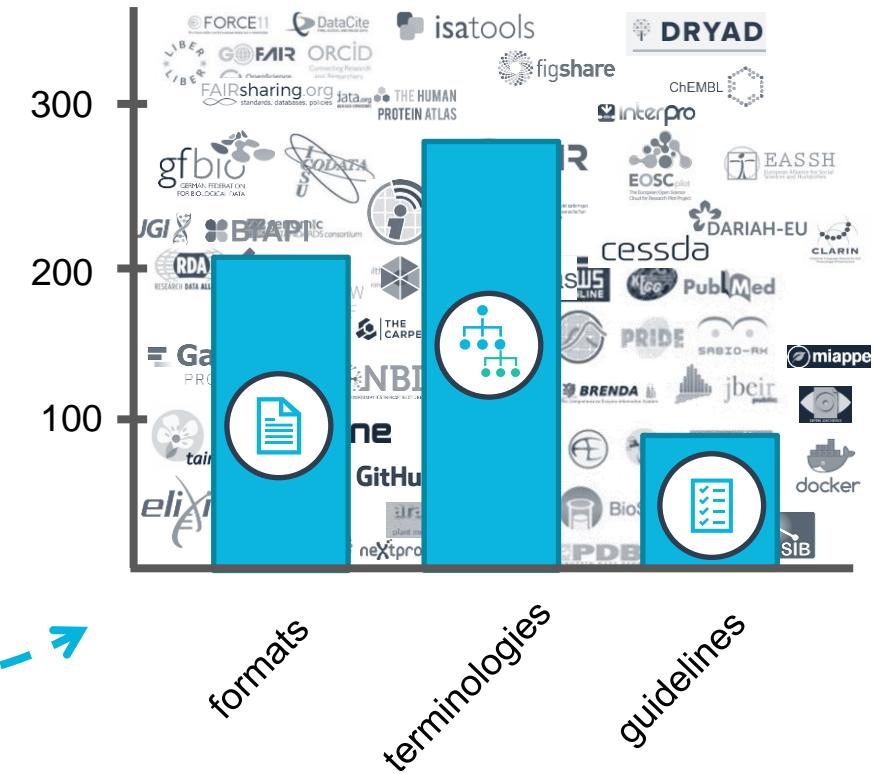
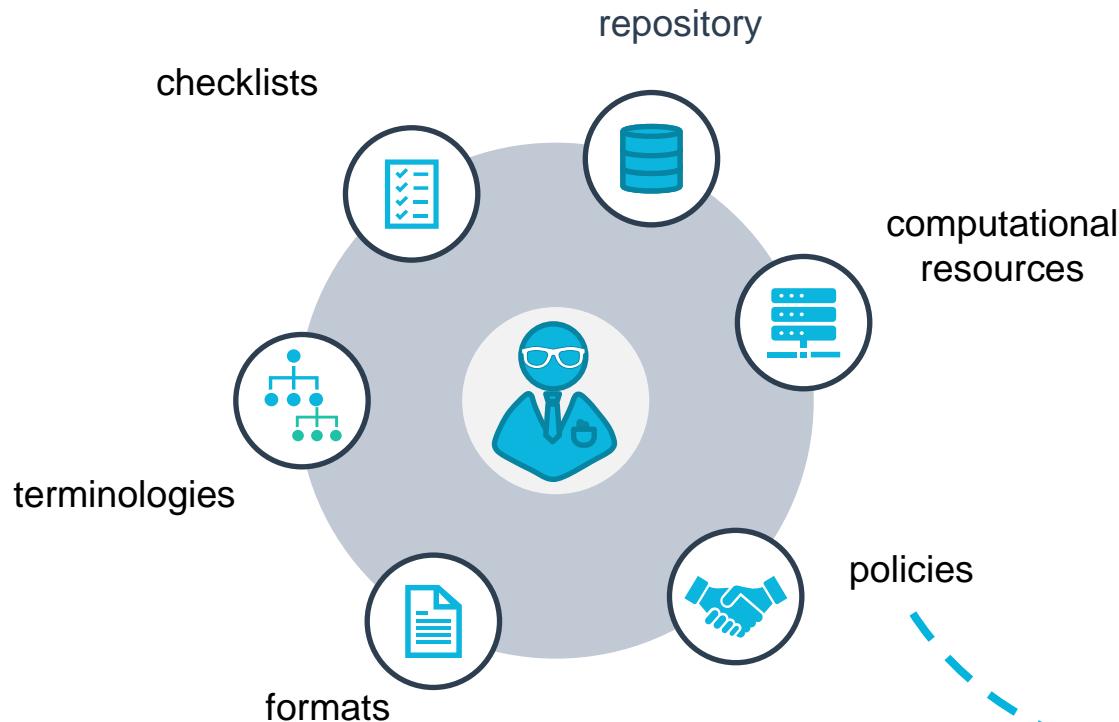


Saves time & workload

FAIR

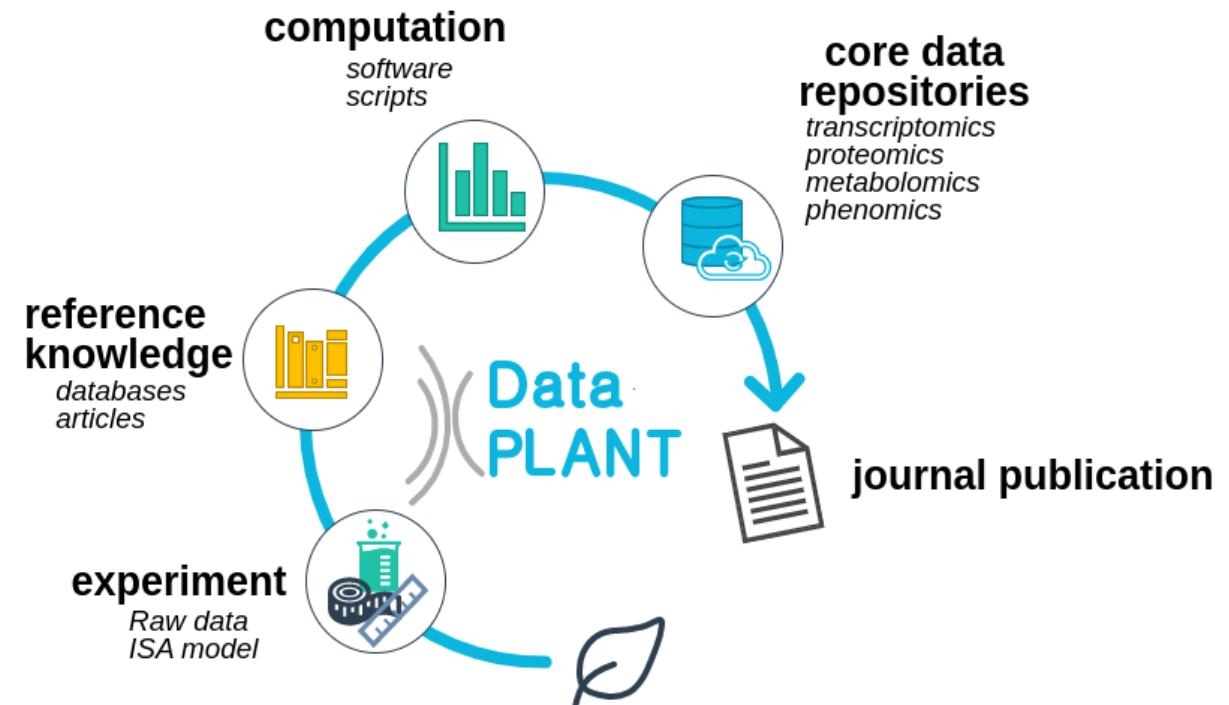


New responsibilities for researchers



Currently, there are more than 600 content 'standards'

- DataPLANT is a “Nationale Forschungsdateninfrastruktur” (NFDI)
- Effectively manage own research data and metadata annotation
- Provide **flexible** tools, standards and services to structure research data according to FAIR principles

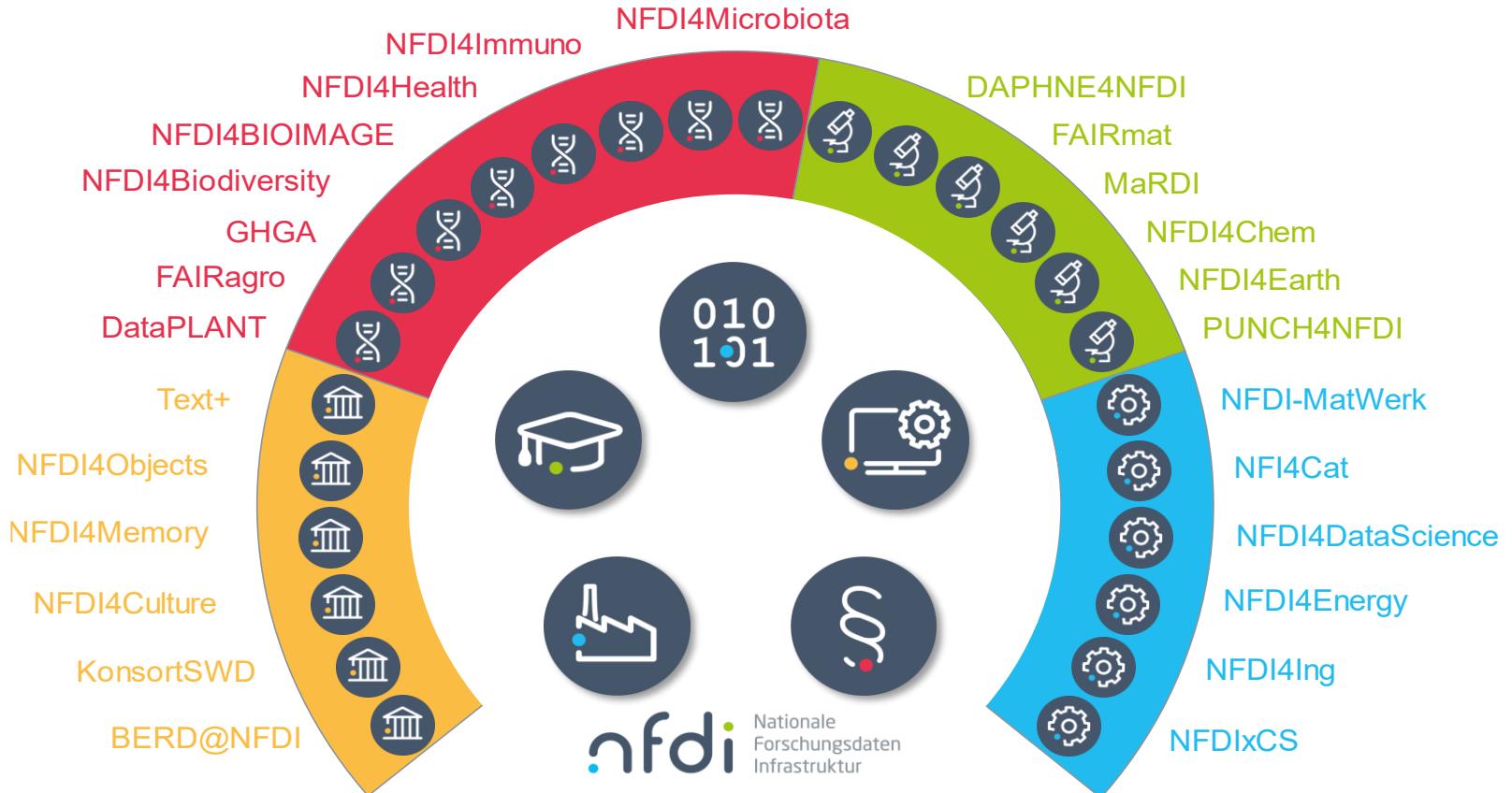


DataPLANT is a (German) NFDI



The aim of the national research data infrastructure (NFDI) is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally.

NFDI Consortia



Order and structure

Question

When have I ever missed order and/or structure of my data?

Advantages of good structure

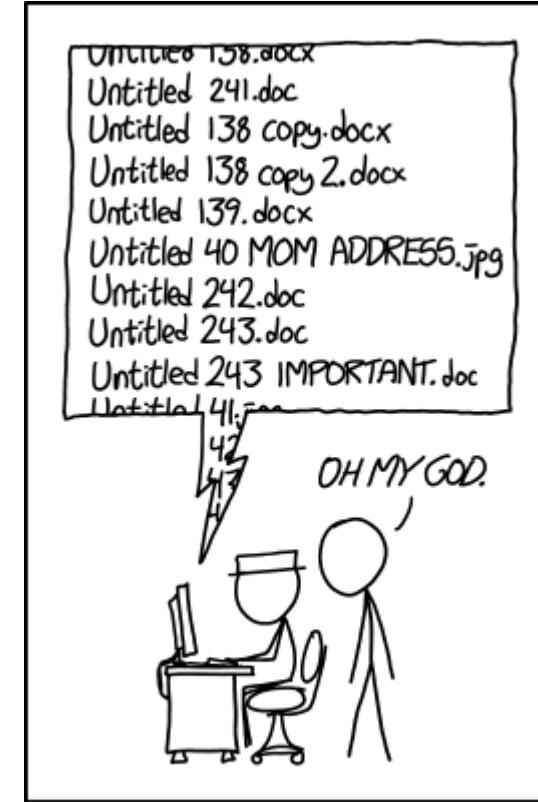
- What, how and why it was done remains comprehensible even after years
 - Others can work with the data
 - Double work is avoided
 - Data loss is avoided
- > it can be worked more efficiently



<https://www.ezcomputersolutions.com/wp-content/uploads/2017/07/grocery-store-comparison-1024x600.jpg>

File naming

- Primary identifier of a file
- Good and meaningful names
 - Hint towards contents of file
 - Help in discovery, classification, sorting, versioning
- Consider (in future)
 - Searching
 - Sorting
 - Uniqueness



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Adding dates to file names

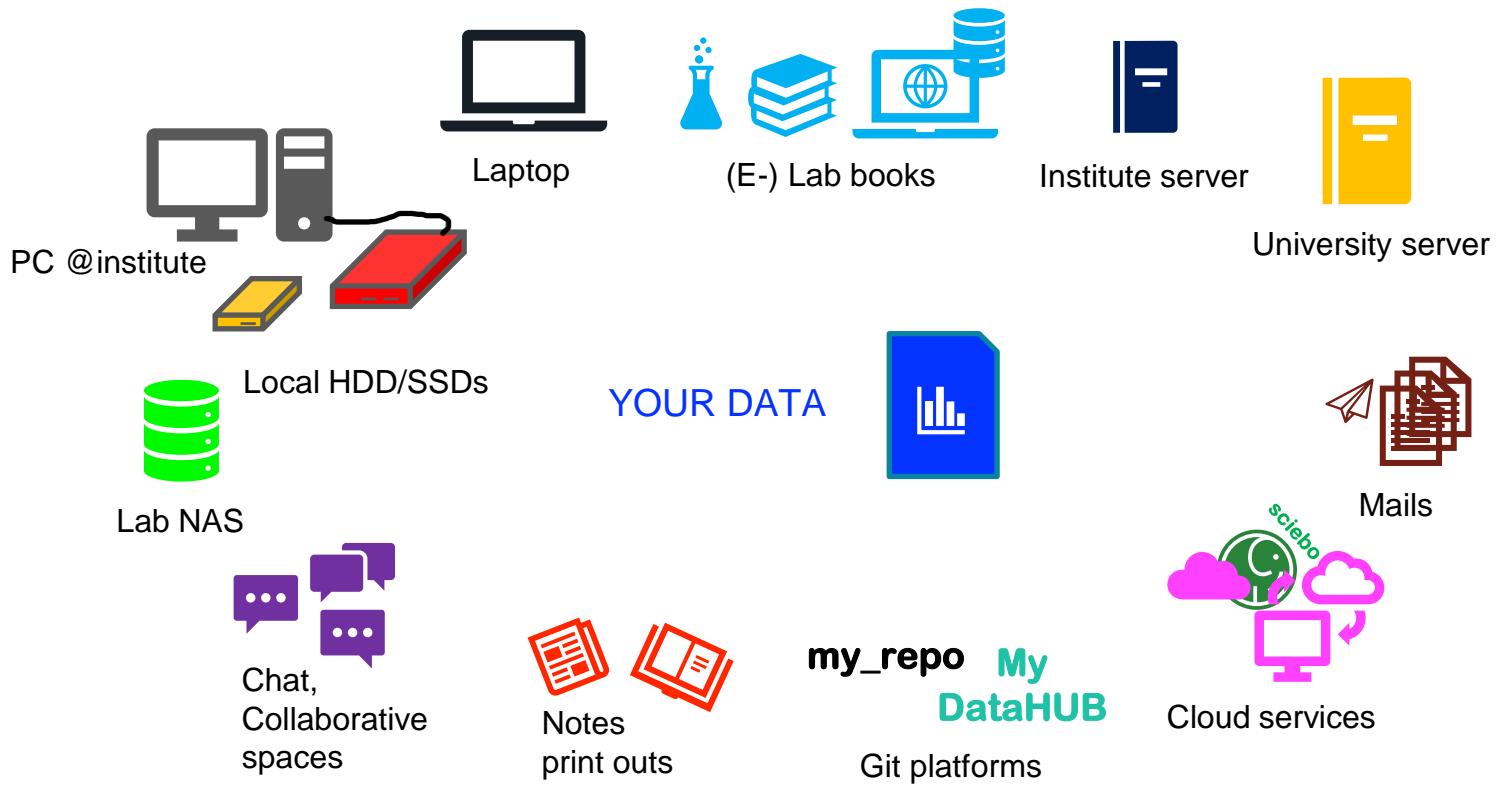
- Dates in the ISO8601 format (YYYY-MM-DD) are recommended as prefix to file names of time-stamped files
- Automatically sorted (independent of the operating system)
- Intuitively human-readable
20221111_presentation vs. 2022-11-11_presentation
- Easy to parse by scripts
 - 2022 // 11 // 11
 - 2022 // Nov // 11



https://www.reddit.com/r/ProgrammerHumor/comments/upffyq/dating_a_programmer/

Data Storage and Backup

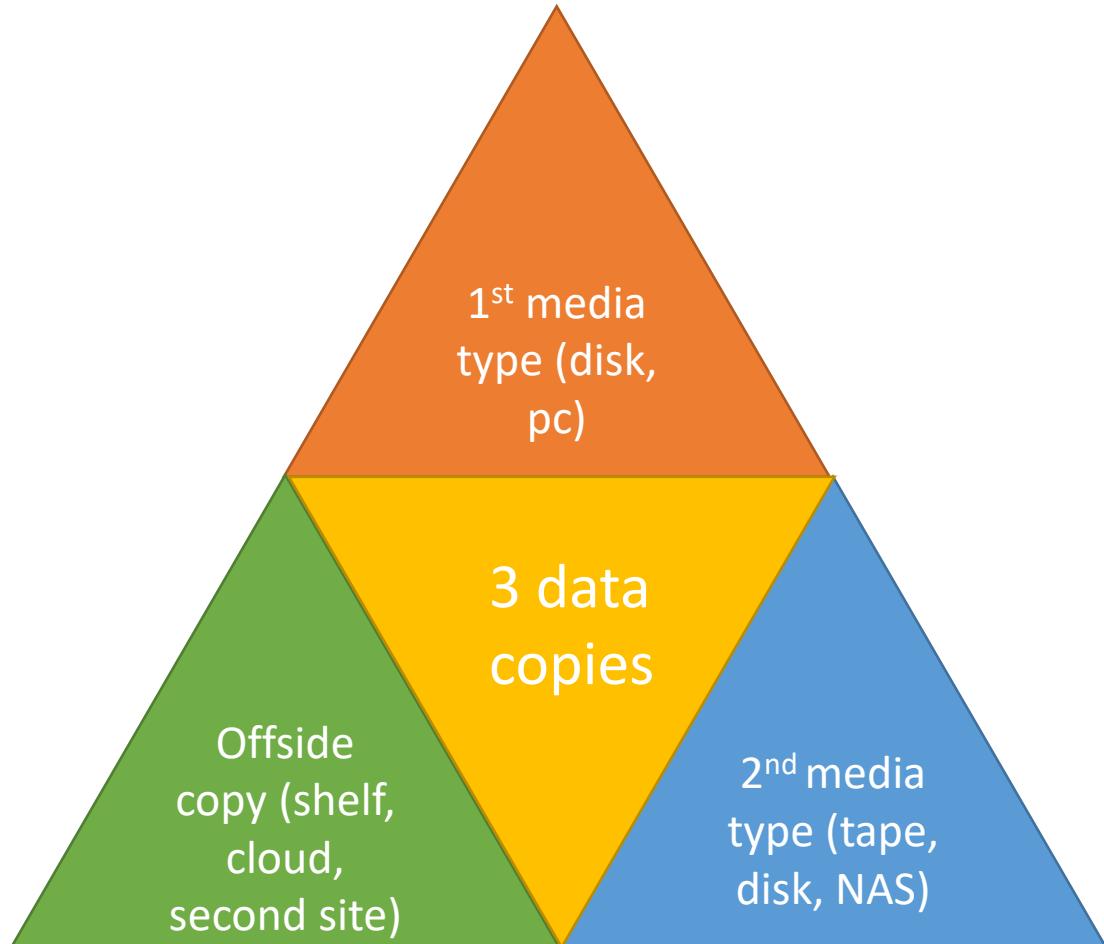
Scattered data



Backup rule

3-2-1 backup rule

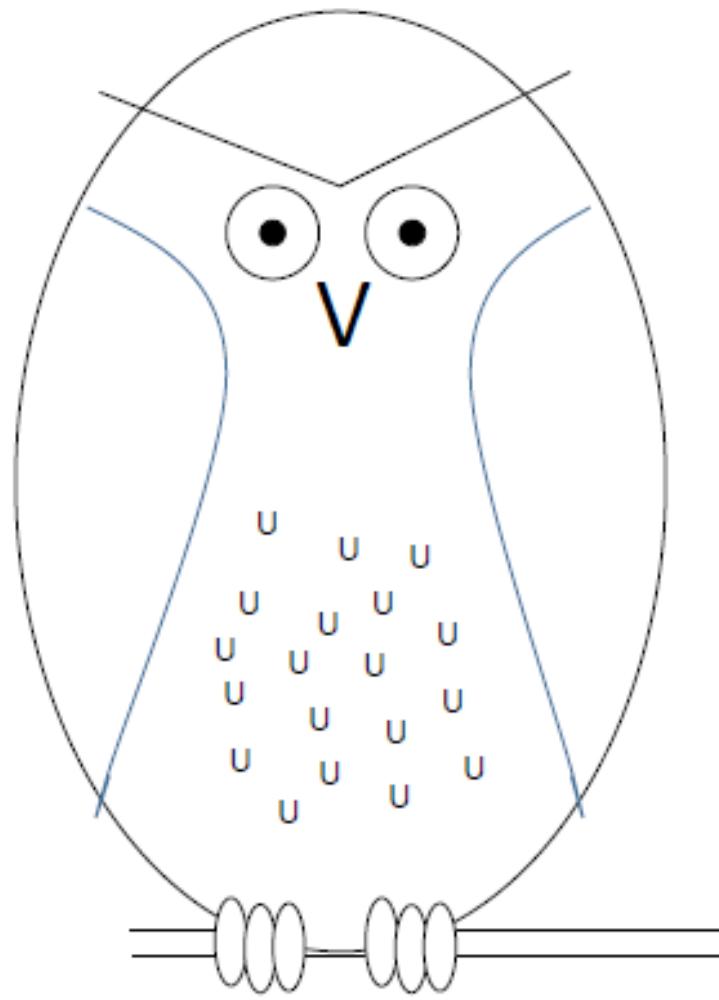
- Three different copies
- Two storage media
- One copy offsite



Documentation

Exercise

1. Draw a large oval that takes up about 2/3 of the side height
2. Draw two circles in the top part of the oval, about 1/5 the height from the top. Draw a smaller circle in each of the two circles and color it
3. Draw a very wide V that extends outside of both sides of the oval, with the point of the V crossing vertically over the center of the two larger circles from step 2
4. Draw curves from the top left and right, curving inward for about 1/4 of the oval, then curving outward to the bottom
5. Draw stretched ovals at the bottom of the large oval, and then two horizontal lines
6. Draw dots or small U's in the area between the curves from step 4
7. Add a single V just below the circles from step 2



Content of documentation

- Description of the research project
- Project goals
- Hypotheses
- Information on the collection of data (methods, units, periods, places, devices used)
- Measures for data cleansing
- Structure of the data and relationship of files to each other
- Explanation of variables, labels and codes
- Differences between different versions
- Information on access and terms of use

Documentation

Which forms of documentation are you using in your lab?



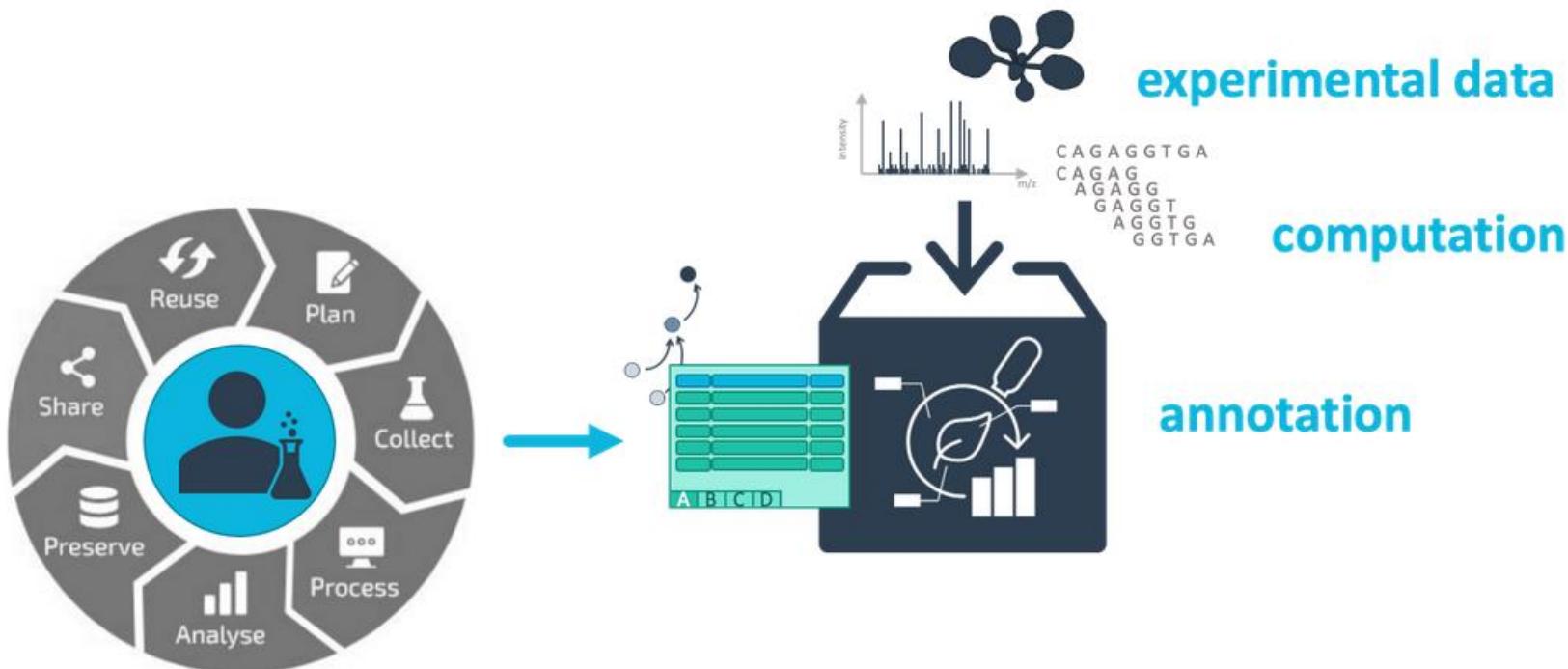
Form of documentation

- Documentation files e.g.
 - README Files
 - Data Dictionaries
 - Codebooks
- Documentation software
 - e.g. Electronic Lab Notebooks
- Articles in data journal
- Wiki

DataPLANT – ARC, Metadata, Data Sharing

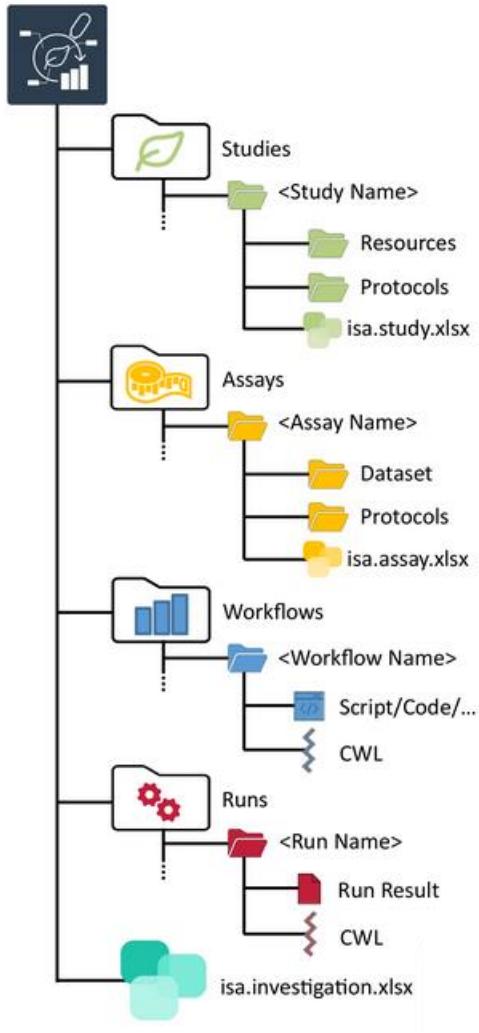
Annotated research context (ARC)

Simple folder and file structure that becomes FAIR by following a few minor principles

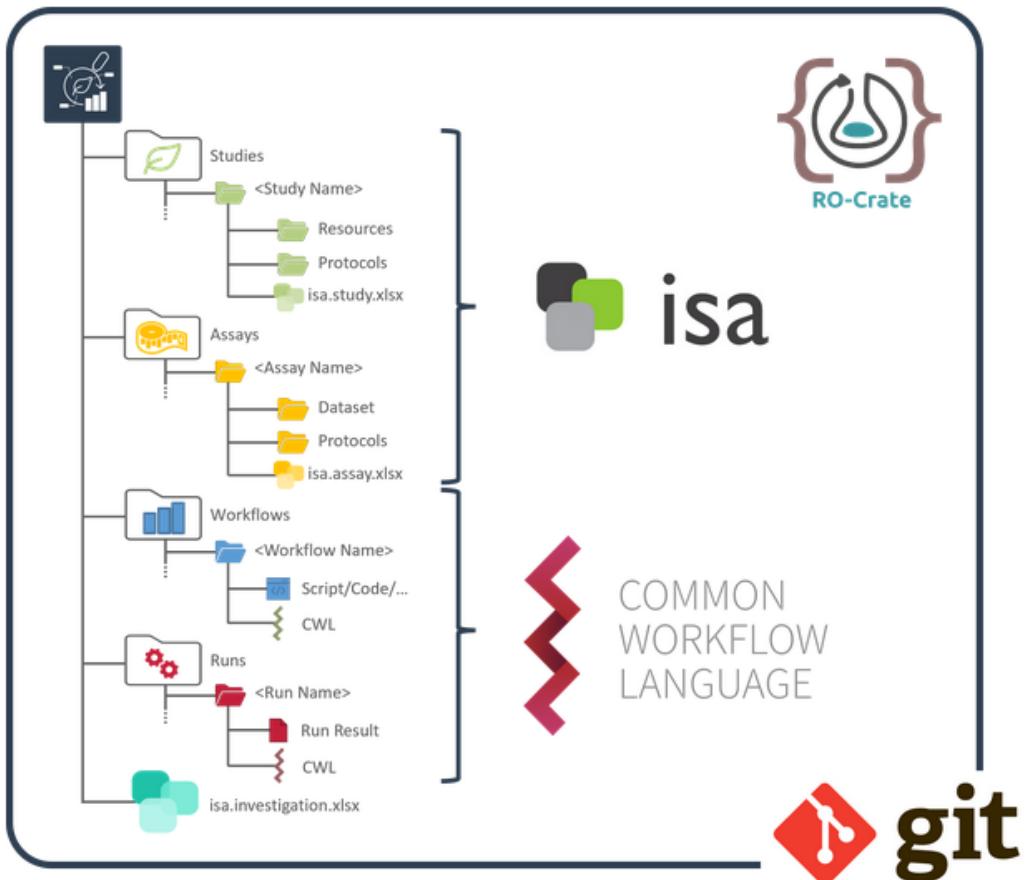


Your entire investigation in a single unified bag

What does an ARC look like?

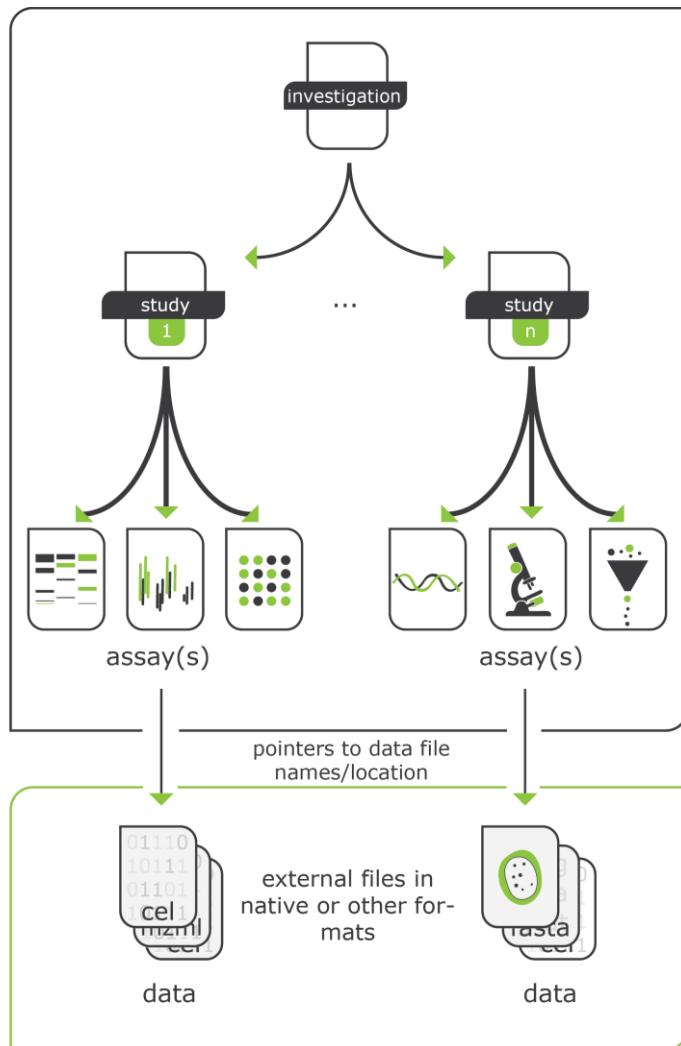


ARC builds on standards



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

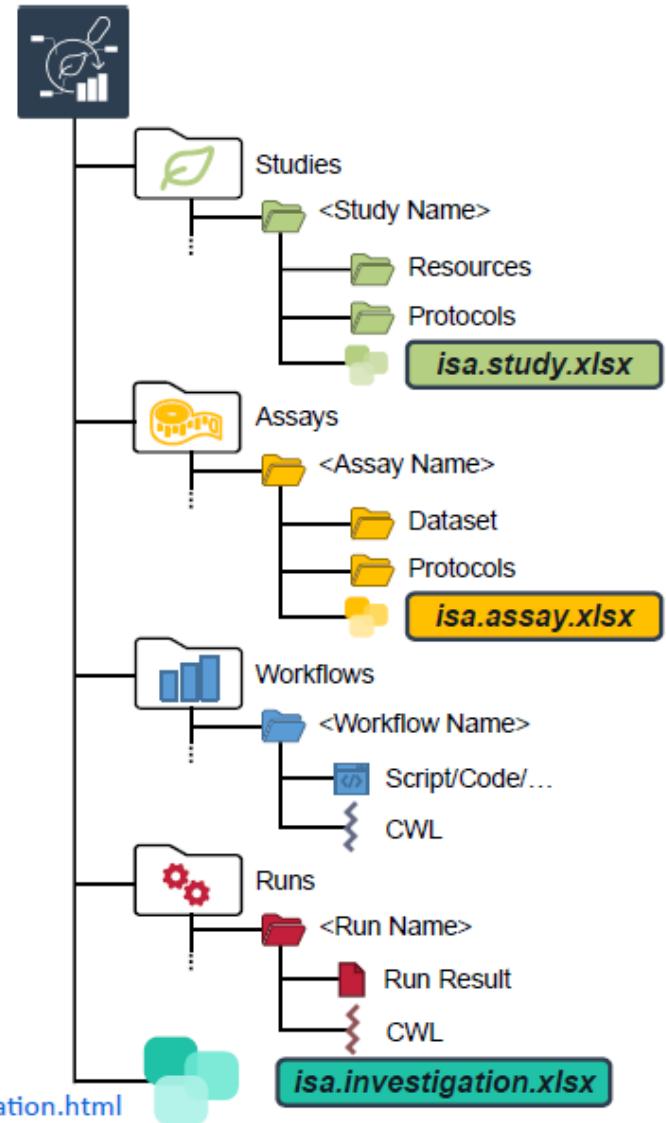
Structure your data – ISA standard



Investigation
Overall goals
Scientific context

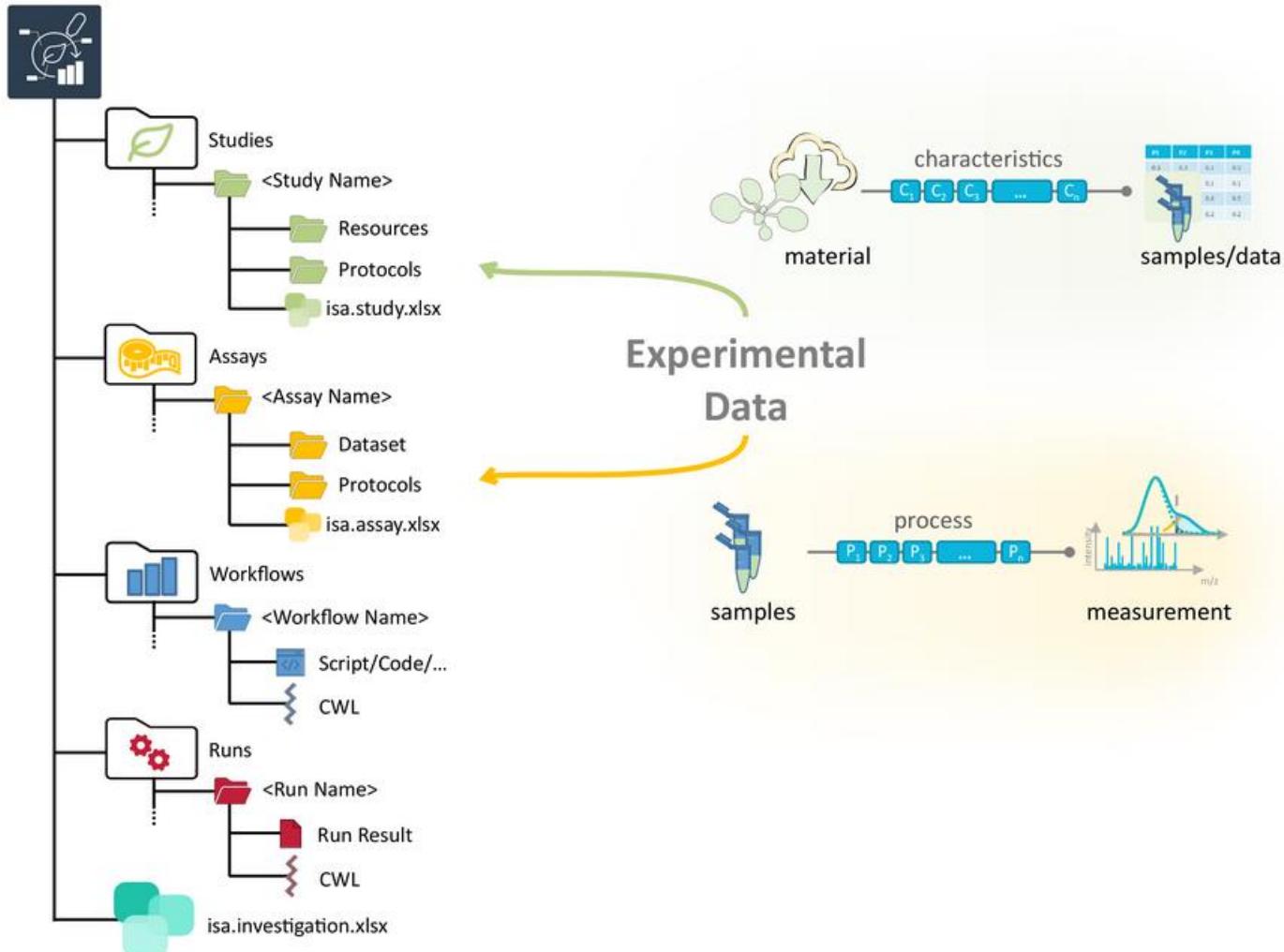
Study
Experimental design

Assay
Leading to (raw) data

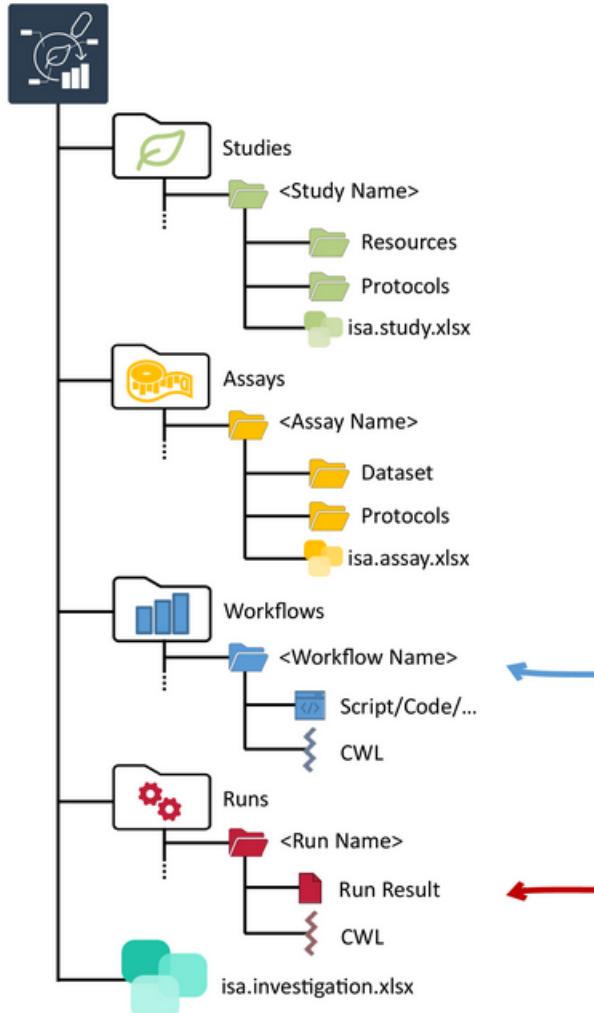


ISA Tools: <https://isa-tools.org/format/specification.html>

ARCs store experimental data

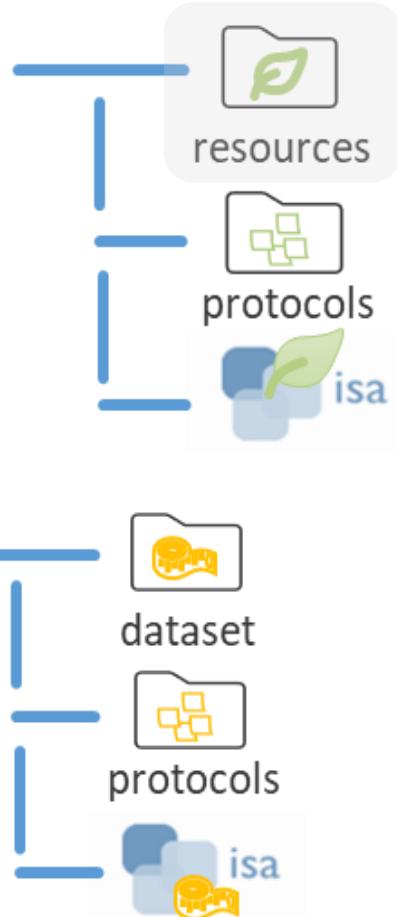


Computations can be run inside ARCs



How does a real-world ARC look like?

Study > Resources



Alignments

AterDL6	MSFRDDN-EEARNDLRRPFIHTGSWYRMGSRQSSMMGSSQVIRDSSISVLACVLIVALGP	59
AterDL4	MSFRDDNTEGRNDLRRFLHTGSWYRMGSRQSSMLESSQVIRDSSISVLACVLIVALGP	60
	***** * .*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	IQFGFTCGYSSPTQAAITKDLGLTVSEYSVFGSLSNVGAMVGAIASCGIAEYIGRKGSIM	119
AterDL4	IQFGFTCGYSSPTQAAITKDLGLTVSEYSVFGSLSNVGAMVGAIASCGIAEYVGRKGSIM	120
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	IAAIPNIIGWLSPAKTSPFLYMGRLLEGPGVGIISYYFPVYIAEIAPQNMRRGLGSVN	179
AterDL4	IAAIPNIIGWLSTSFAKDTSPFLYMGRLLEGPGVGIISYYFPVYIAEIAPQTMRGALGSVN	180
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	QLSVTIGIMLAYLLGLFVFWRILAVLGILPCTLLIPGLFFIPESPRNLAKMGMDEFETS	239
AterDL4	QLSVTIGIMLAYLLGLFVFWRILAVLGILPCTLLIPGLFFIPESPRNLAKMGLTDDETS	240
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	LQVLRGFTEDITVEVNEIKRSVASSTKRNTYRFVDLKRRRYFFLMVGIGLLVLQQGGI	299
AterDL4	LQVLRGFTEDITVEVNEIKRSVASSSKRSAVRFVDLKRRRYFFLMVGIGLLALQQLGGI	300
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	NGVLFYSSSTIFESAGVTSSNAATFGVGAIQVVATAISTWLVDKAGRLLLTISSVGMTIS	359
AterDL4	NGVLFYSSSTIFESAGVTSSNVATFGVGVVQVVAATGIAITWLVDKAGRLLLMISSIGMTIS	360
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	LIVIVAAAFYLKEFVSPDSDMYSNLISLISVVGVAMMVFFSLGMGPPIPWLIMSEILPVNIK	419
AterDL4	LIVIVAVAFYLKEFVSPDSNMINILSMVSVVGVAMVISCLGMGPPIPWLIMSEILPVNIK	420
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	GLAGSIATLANWFWSWLITMTANLLAWS:	
AterDL4	GLAGSIATLNNWFWSWLITMTANLLAWS:	
	*****:*****:*****:*****:*****:*****:*****:*****:*****:*****:	
AterDL6	EELQSLFR 467	
AterDL4	EELQALFR 468	
	::**	

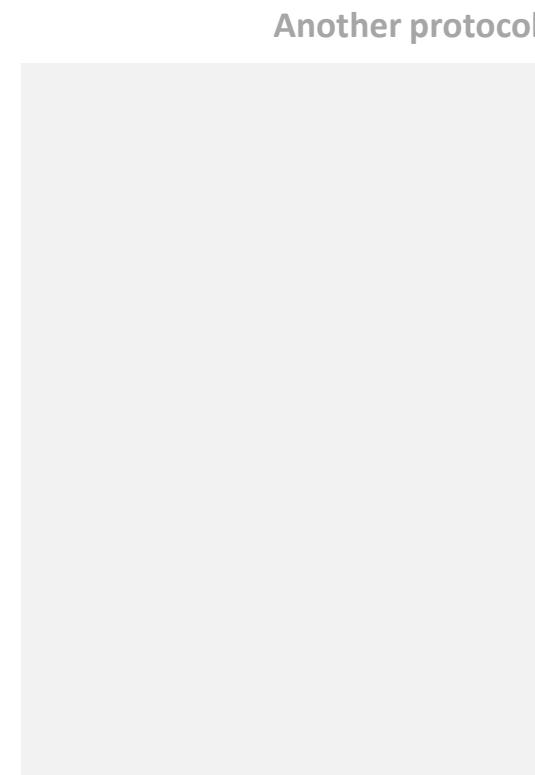
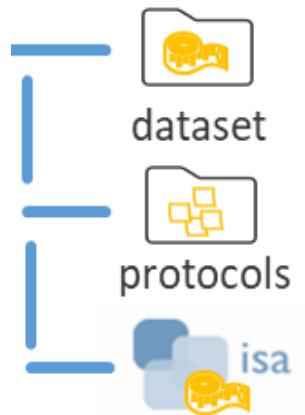
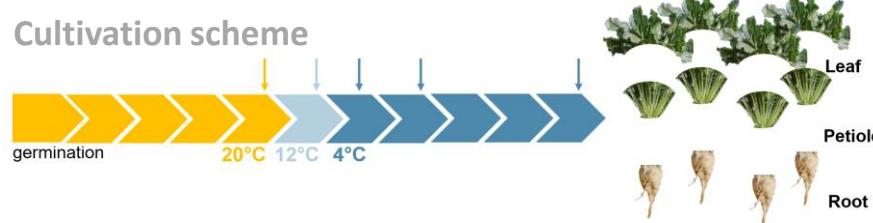
Another Resource

Sequence alignment results from another resource, showing sequence similarity between AterDL6 and AterDL4 across various genes. The sequences are aligned with gaps indicated by dashes (-) and highlighted differences in green boxes. The alignment shows high conservation of the sequence across the genes.

Fasta-Files

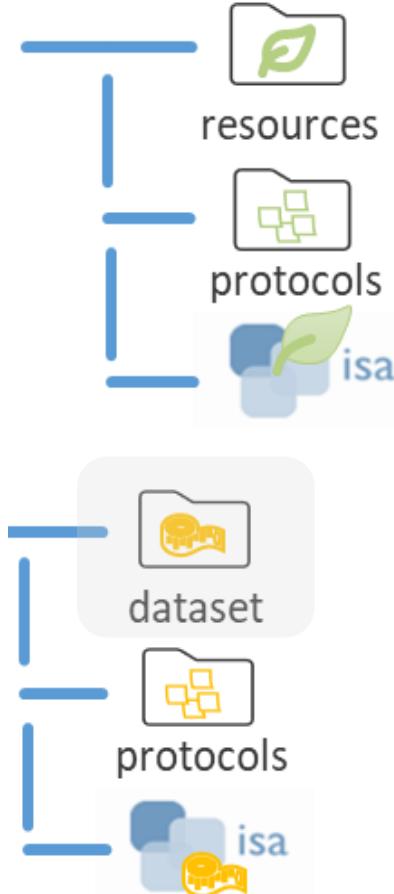
How does a real-world ARC look like?

Study > Protocols



How does a real-world ARC look like?

Assay > Dataset

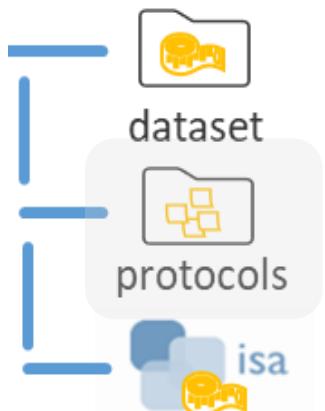


Data set directly from the machine

Acc ID	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6
NM_007818	67540.89	70924.09	80243.76	3501.2	5697.47	2426.72
NM_001105160	811.93	801.36	740.71	128.67	104.42	101.33
NM_028089	190.41	211.06	236.19	9.05	23.33	8.44
NM_016696	66.77	57.56	101.09	750.9	659.84	491.89
NM_013459	3.3	11.29	1.89	735.82	816.46	118.22
NM_007809	45.34	36.12	51.02	245.27	372.13	335.67
NM_009999	103.04	370.21	200.29	17.09	13.33	8.44
NM_133960	7708.78	6976.38	6569.04	1731	1641.81	1853.55
NM_027881	31.32	10.16	24.56	268.39	186.62	135.11
NM_054053	31.32	24.83	19.84	323.68	428.78	116.11
NM_007377	47.81	89.17	70.86	370.93	378.79	279.72
NM_028064	703.95	689.62	662.29	214.11	168.85	144.61
NM_008182	222.56	339.73	226.75	30.16	63.32	26.39
NM_013661	12.36	11.29	8.5	97.51	77.76	71.78
NM_007815	20613.09	25218.13	31540.46	5209.07	7680.3	6312.2

How does a real-world ARC look like?

Assay > Protocols



Freely written protocols

RNA-Preparation (Trizol)

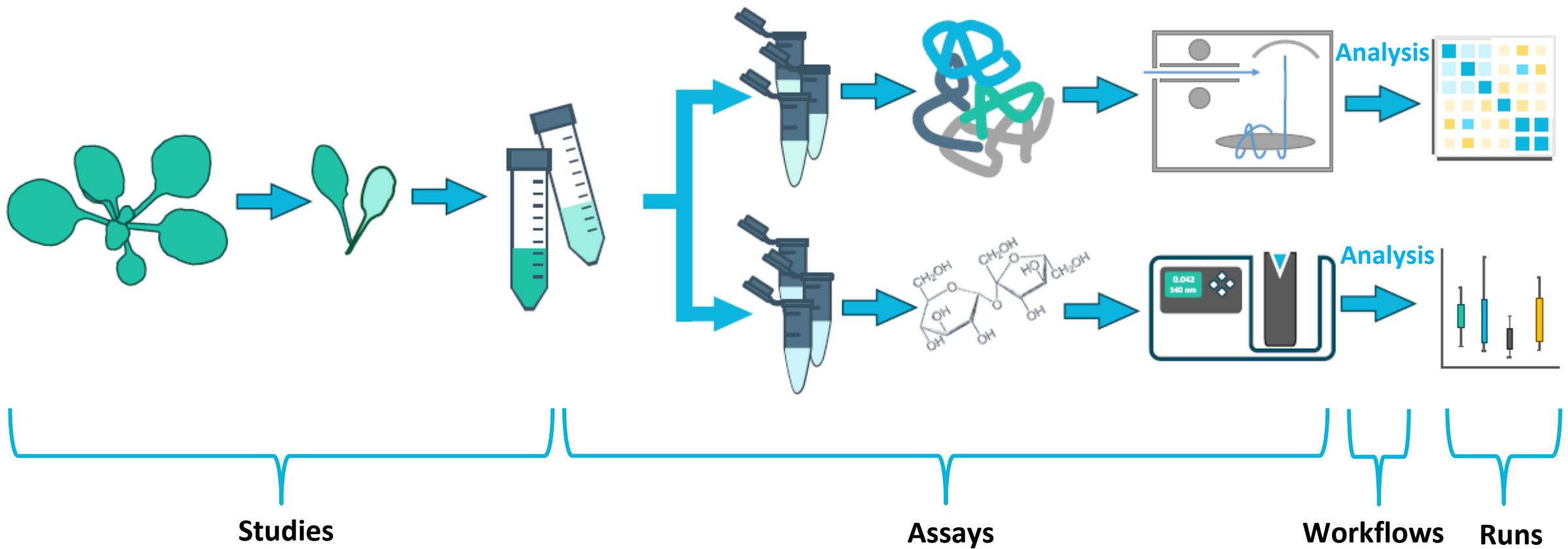
Reagents

70%	Ethanol
	Trizol
	Chloroform
	Isopropanol

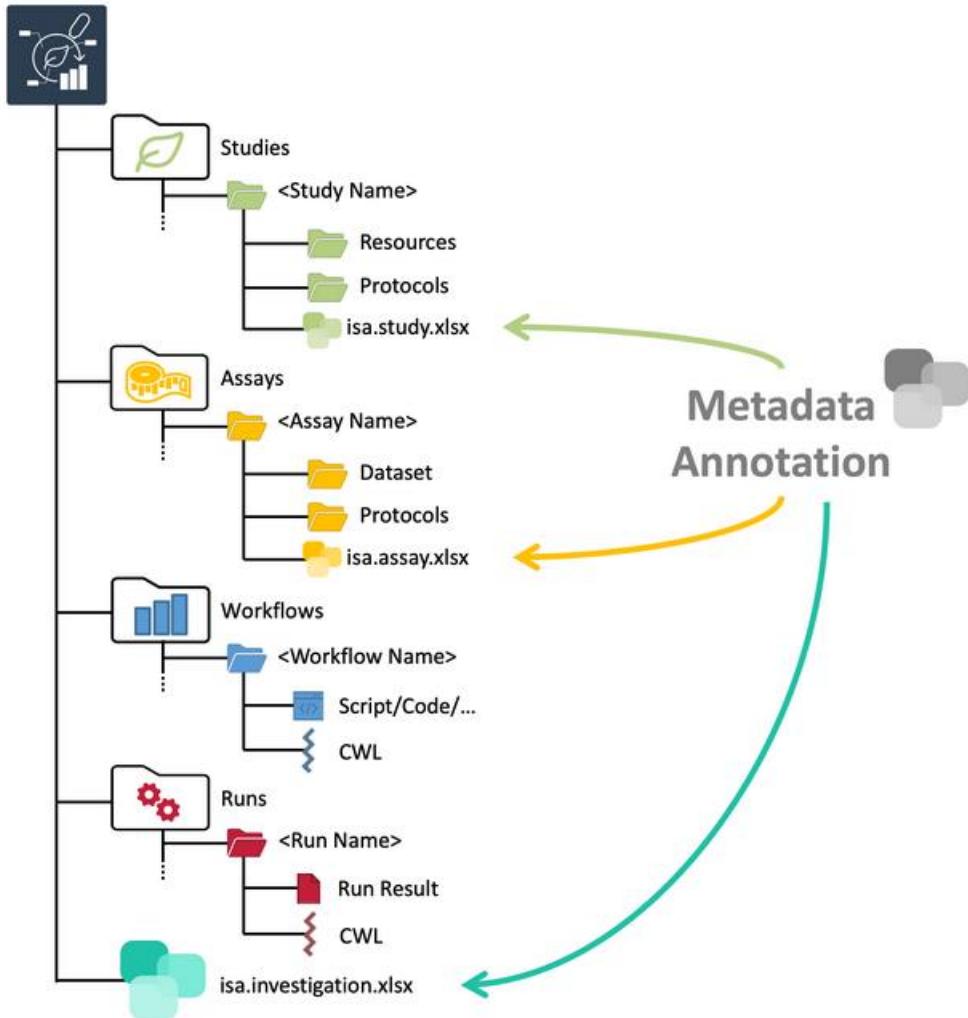
Protocol:

- add about 100mg Tissue to 1ml Trizol
- vortex 3x 30sec.
- incubate 15min. RT
- add 500µl Chloroform
- vortex 3x 30sec
- centrifugation 10min. 4°C 14000rpm
- transfer supernatant to a new tube (ca. 600µl)
- add 1Vol. of Isopropanol
- centrifugation 10min. 4°C 14000rpm
- throw away supernatant
- add 500µl 70% Ethanol
- centrifugation 5min. 4°C 14000rpm
- throw away supernatant
- dry the pellet under sterile bench
- add 50-70µl RNase free H₂O
- estimate RNA-content with the Nanodrop-machine

A small prototypic project

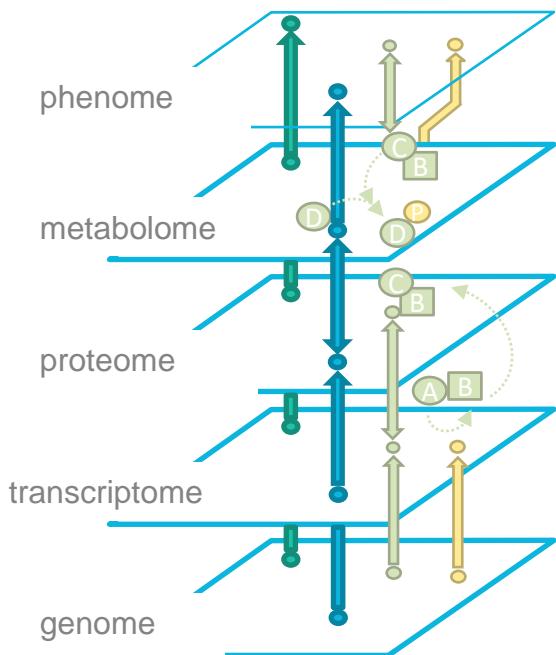


ARCs come with comprehensive metadata



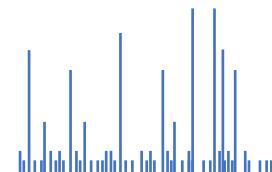
What is Metadata?

Data that describes your data

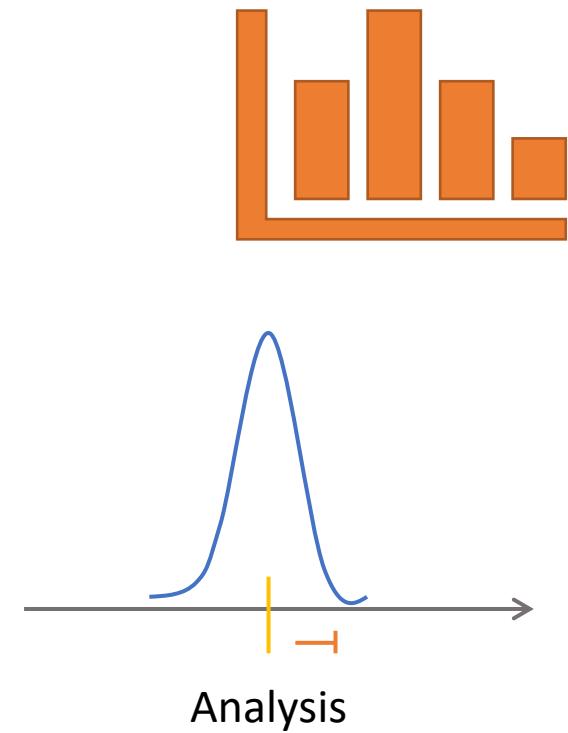


A G A G G
A G G T G
C A G A G
G G T G A
C A G A G G T G A

Sequencing



Measurements



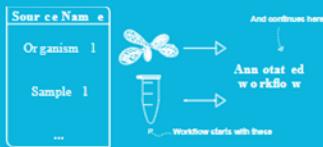
Analysis

Metadata: Simplify the annotation grammar

Input Columns

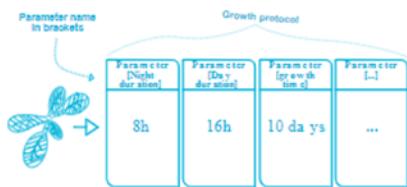
Every annotation table must start with an Input column, which defines the input of your table. This input value must be a unique identifier for an organism or a sample.

The number of Input columns per table is limited to one.



Parameter

Parameter columns describe steps in your experimental workflow, e.g., the temperature or extraction buffer used for your assay. Multiple Parameter columns form a protocol.



Characteristics

Characteristic columns describe inherent properties of the source material, e.g., a certain strain or ecotype, but also the temperature an organism was exposed to. You can use any number of Characteristic columns.



Output Columns

Per table only one output column is allowed, which can either be a Sample Name, a Raw Data File or a Derived Data File. Data files can be sources for computational workflows. The value of this column must be a unique identifier.



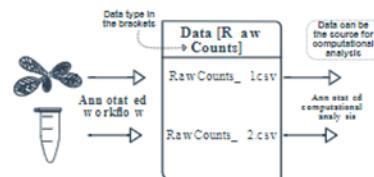
Factor

Use Factor columns to describe independent variables that determine the specific output of your experiment when process and analysis were identical. Factors are the most important building blocks for downstream computational analysis.

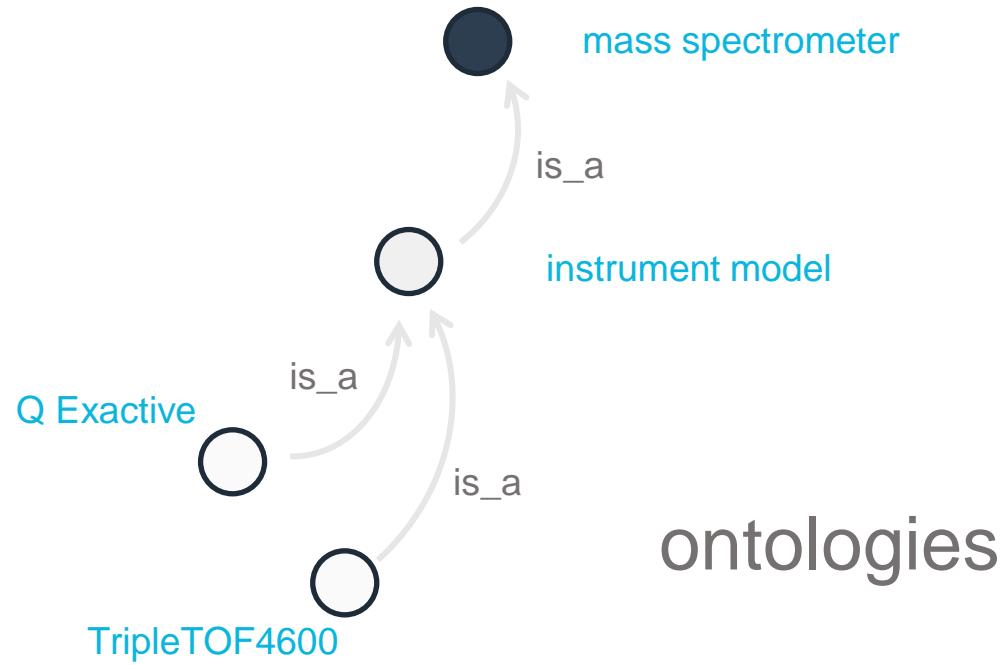
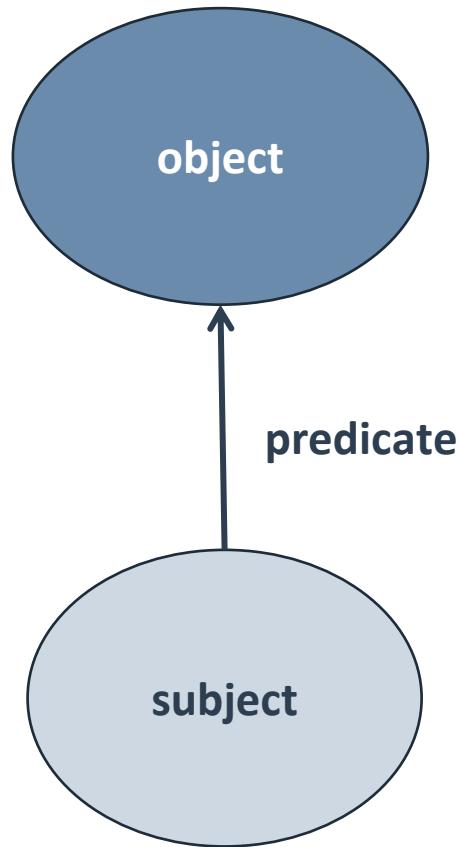


Component

Use these columns to list anything physical or a protocol that can be consumed, e.g. instrument names, software names, or reagents names.

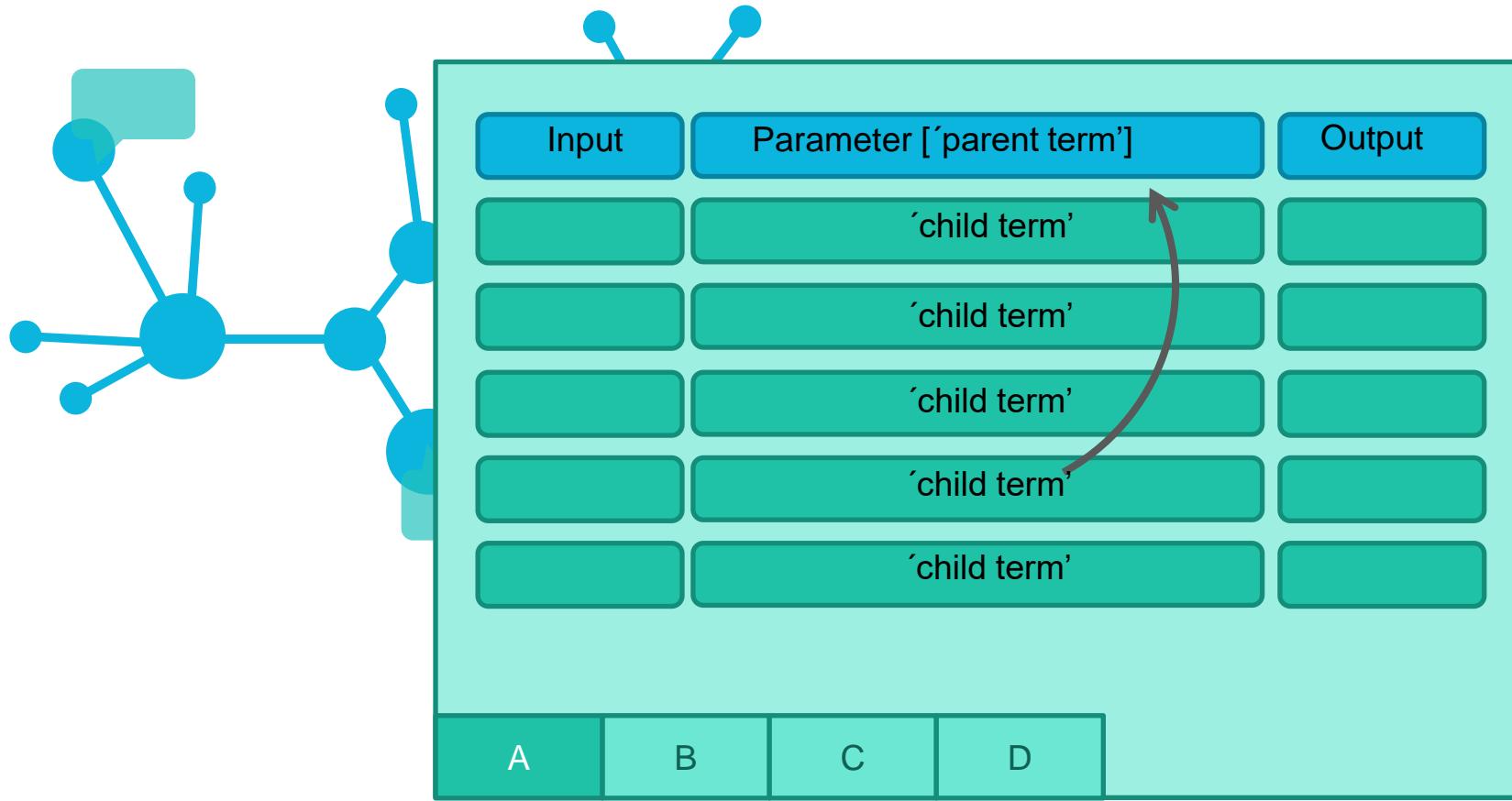


Using ontologies for annotation



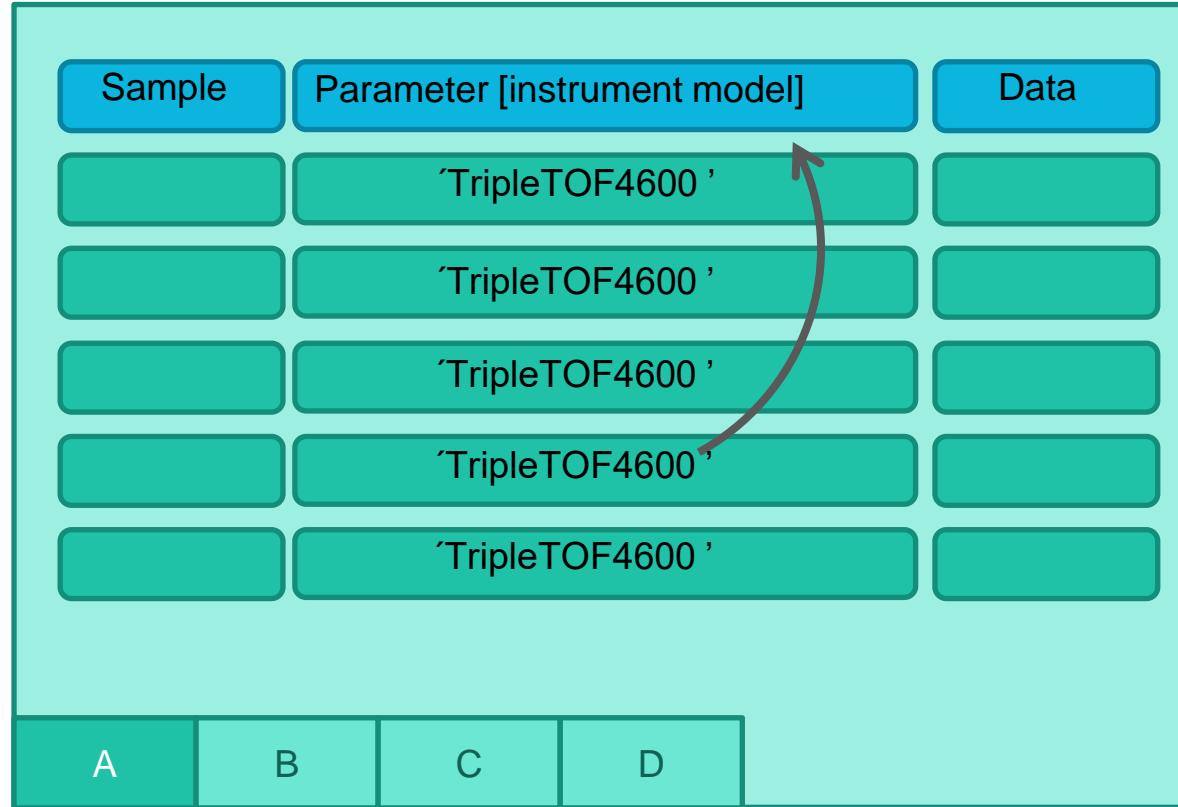
A graph of information: Terms (with definitions) as nodes & relationships as links between terms

Annotation by flattening the knowledge graph



- Low-friction metadata annotation
- Metadata annotation via ISA-Tab (spreadsheet-friendly ISA)
- Familiar Spreadsheet, row/column-based environment

Annotation principle



- Low-friction metadata annotation
- Metadata annotation via ISA-Tab (spreadsheet-friendly ISA)
- Familiar Spreadsheet, row/column-based environment

Swate – Metadata annotation tool

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

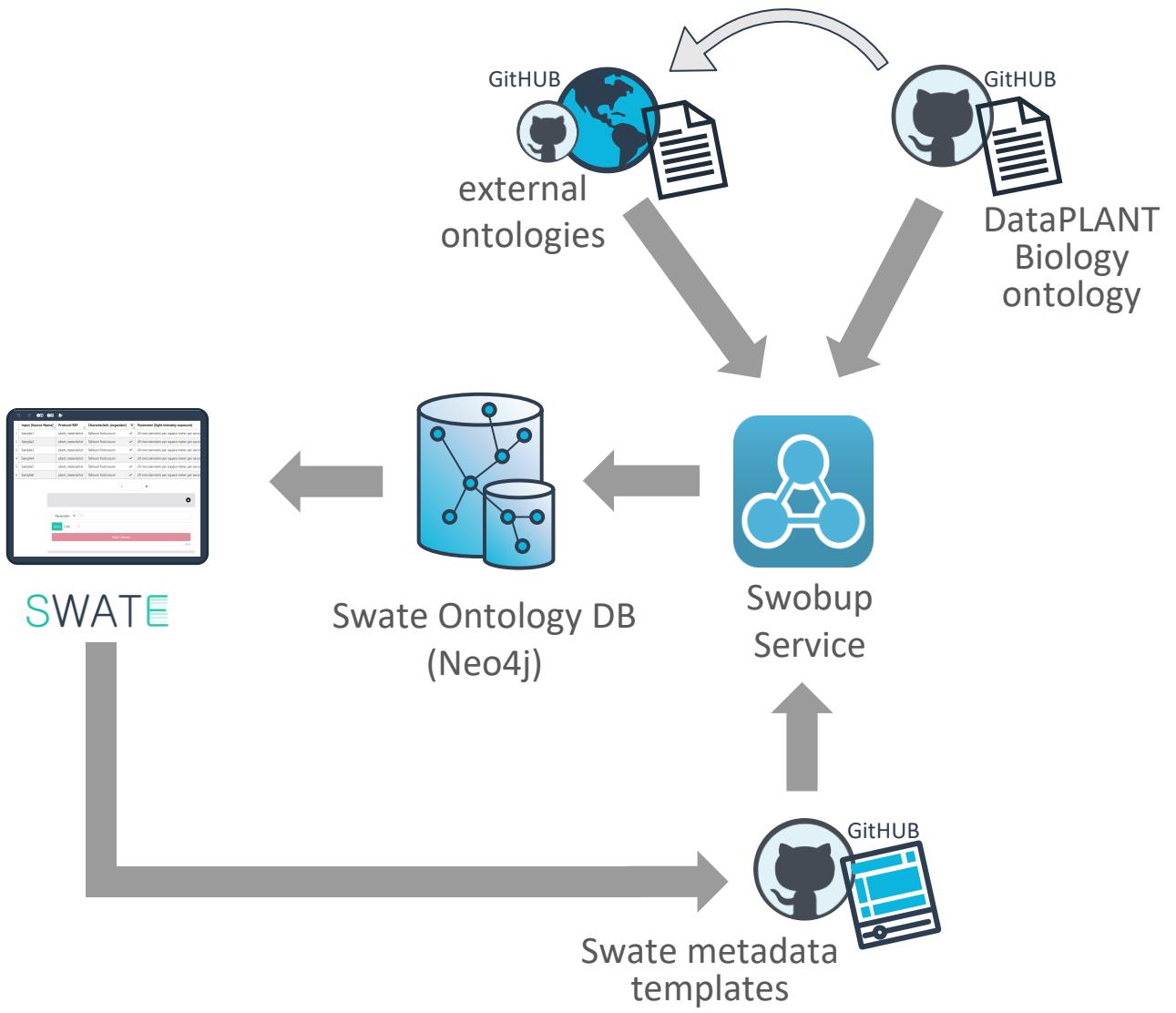
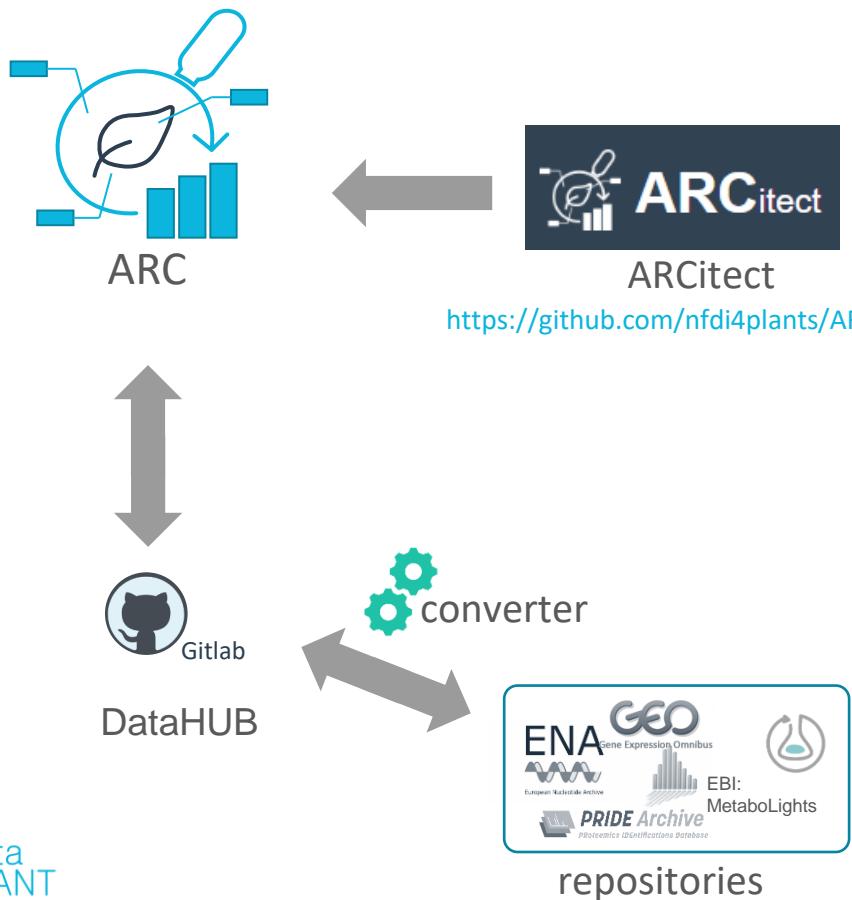
- Quick and efficient way to annotate study and assay metadata sheets
- Employs ontologies to standardize both the headers and respective values tables
- Repository checklists and requirements can be displayed as templates

The screenshot shows the Swate interface. At the top, there's a toolbar with various icons and a "Widgets" button. To the right is a "Save & delete" button. Below the toolbar is a table titled "Annotations" with columns: Input [Source Name], Characteristic [Organism], Factor [watering exposure], and Output [Sample Name]. The table contains six rows of data. A callout box labeled "Elements connected to one function" points to the "Characteristic [Organism]" column. At the bottom left, there's a diagram showing a "Metadata sheet" pointing to "Metadata" and an "Annotation table" pointing to "plant_material". On the right side, there's a sidebar titled "Ontology term search" with a search bar containing "Talinum fruticosum". A callout box labeled "Sidebar" points to this sidebar area. At the bottom right, it says "Swate Release Version v1.0.0-beta.03 Host Browser".

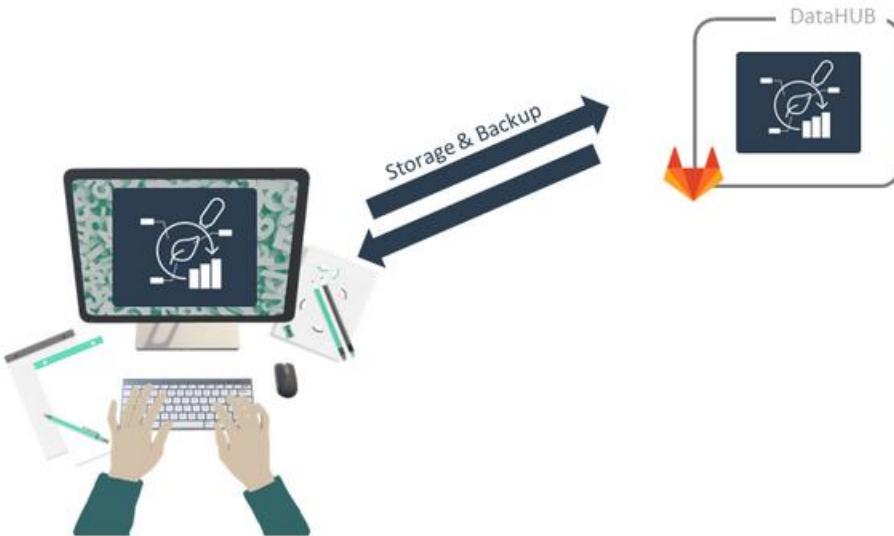
	Input [Source Name]	Characteristic [Organism]	Factor [watering exposure]	Output [Sample Name]
1	DB_097	Talinum fruticosum	✓ 12 days drought	CAM_01
2	DB_099	Talinum fruticosum	✓ 12 days drought	CAM_02
3	DB_103	Talinum fruticosum	✓ 12 days drought	CAM_03
4	DB_161	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_01
5	DB_163	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_02
6	DB_165	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_03

Metadata sheet → Metadata
Annotation table → plant_material

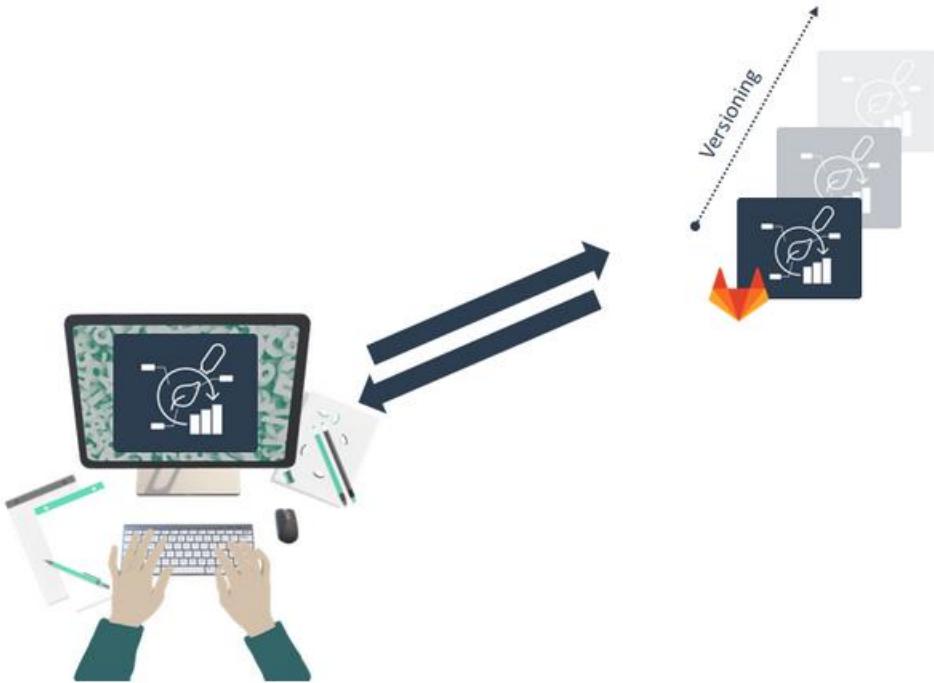
DataPLANT tool chain



You can store your ARC in the DataHUB



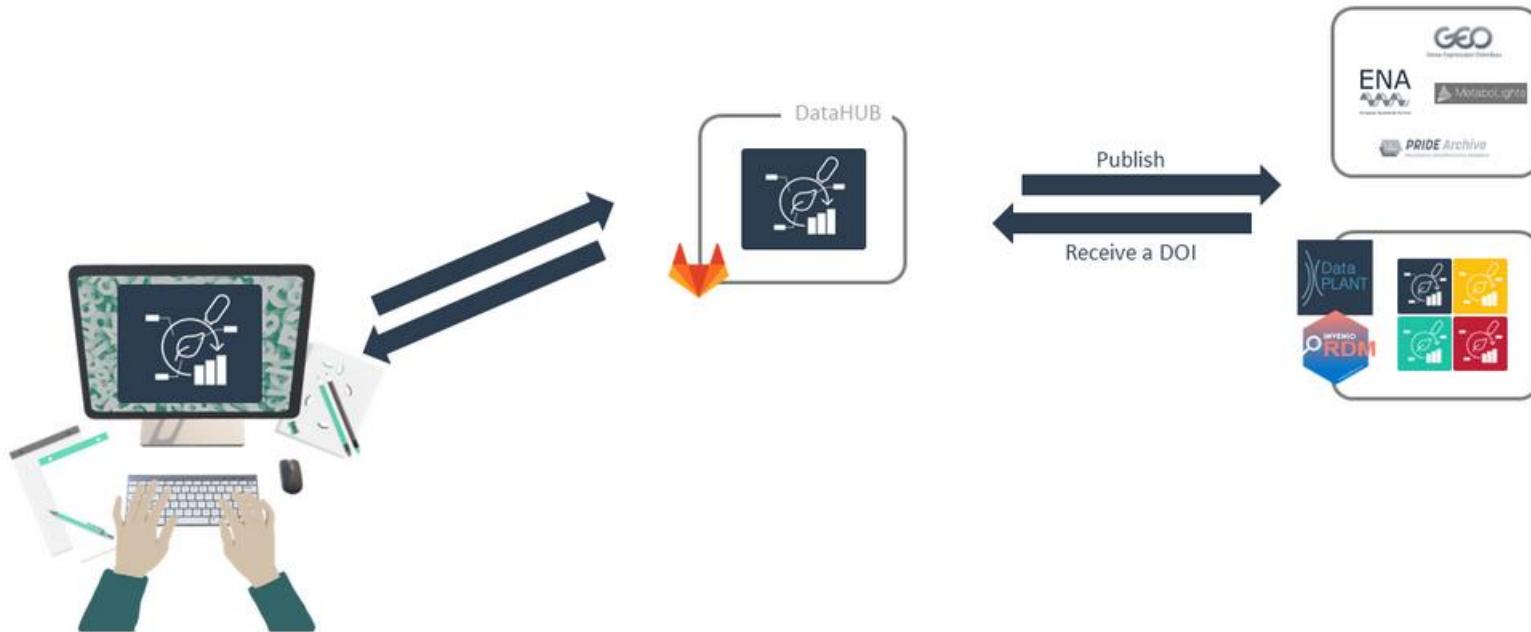
ARCs are versioned



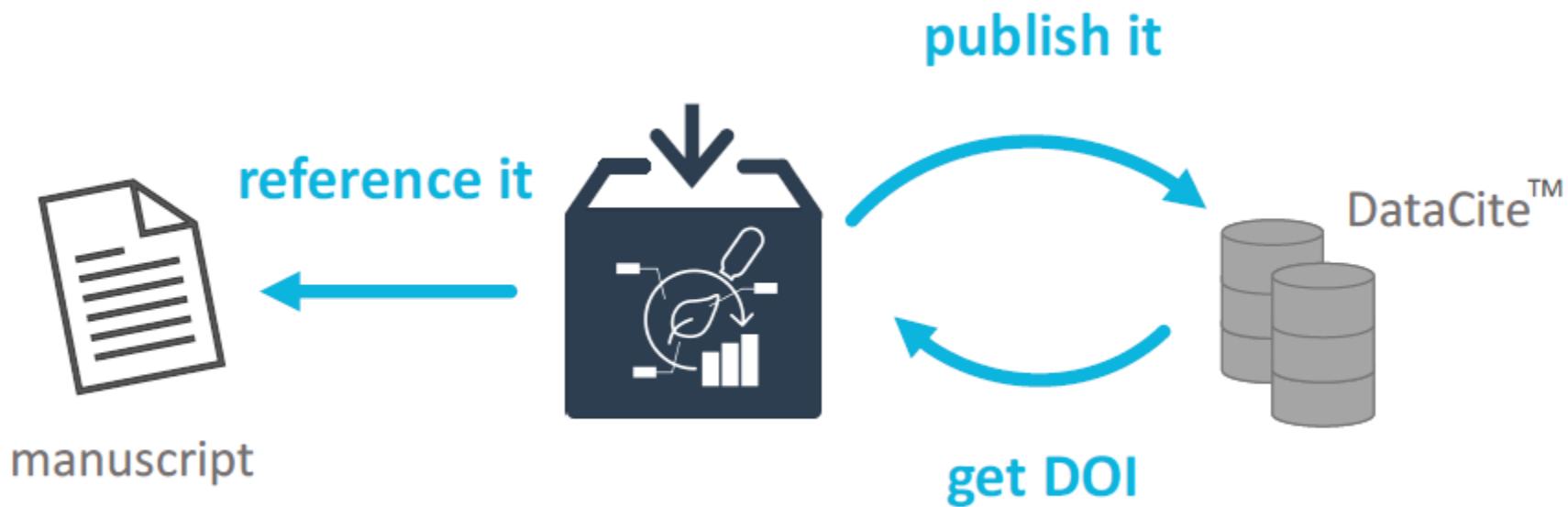
You can invite collaborators



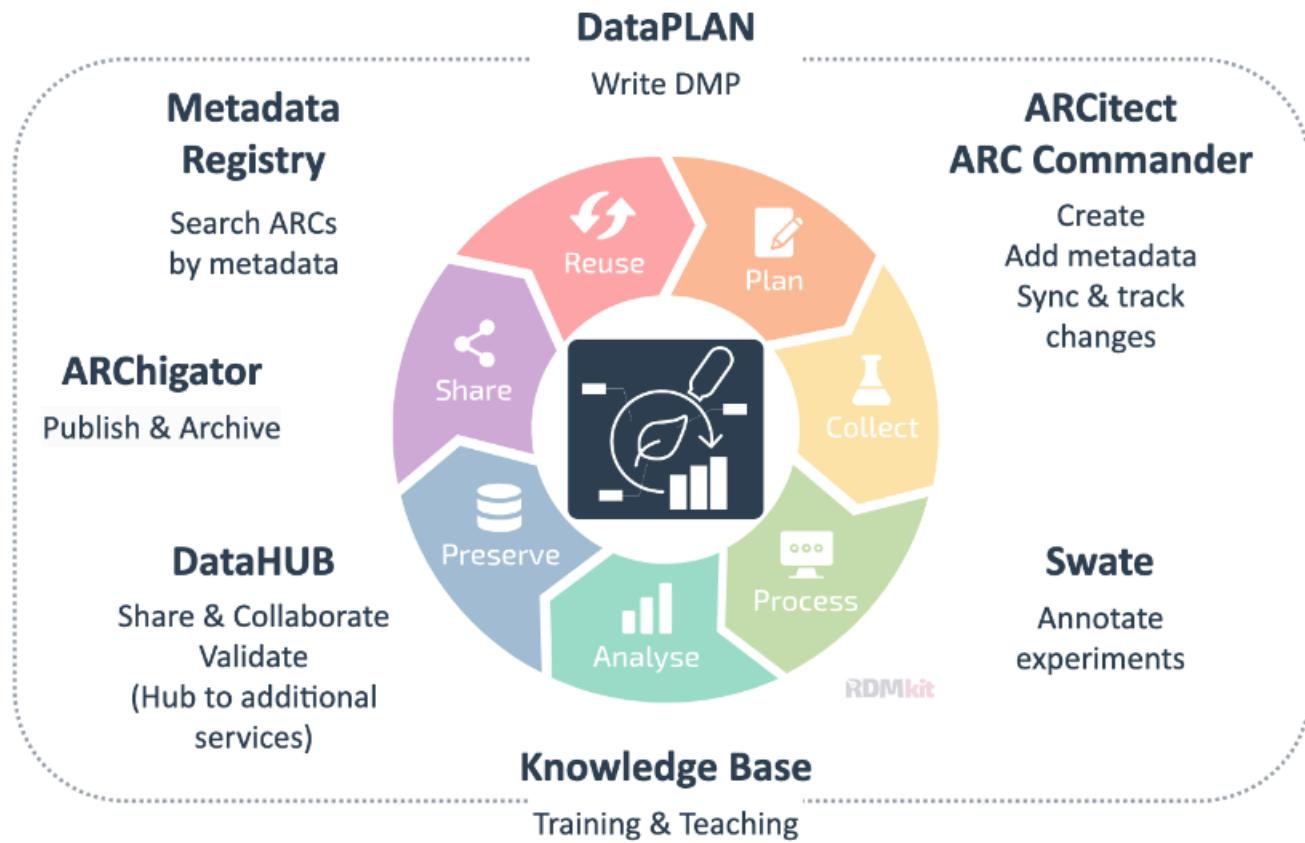
Publish your ARC



Publish your ARC, get a DOI



The ARC ecosystem



ARCitect walkthrough: create an ARC

ARCitect

The screenshot illustrates the ARCitect application interface, which integrates local file management and DataHUB synchronization.

Left Panel: The main menu includes:

- Create / open / download ARCs
- Commit, and view changes and files
- Sync changes with the DataHUB

Middle Panel: The **ARC folder structure** shows the directory tree:

```
C:/Users/Stella.Eggels/Desktop/ARCs/Maize_Water_use_efficiency_24
```

- Maize_Water_use_efficiency_24
 - studies
 - assays
 - gasexchangemeasurements
 - dataset
 - protocols
 - README.md
 - RNAExtraction
 - workflows
 - runs
 - .gitignore

Right Panel: The **Metadata** entry screen for **gasexchanger** contains the following fields:

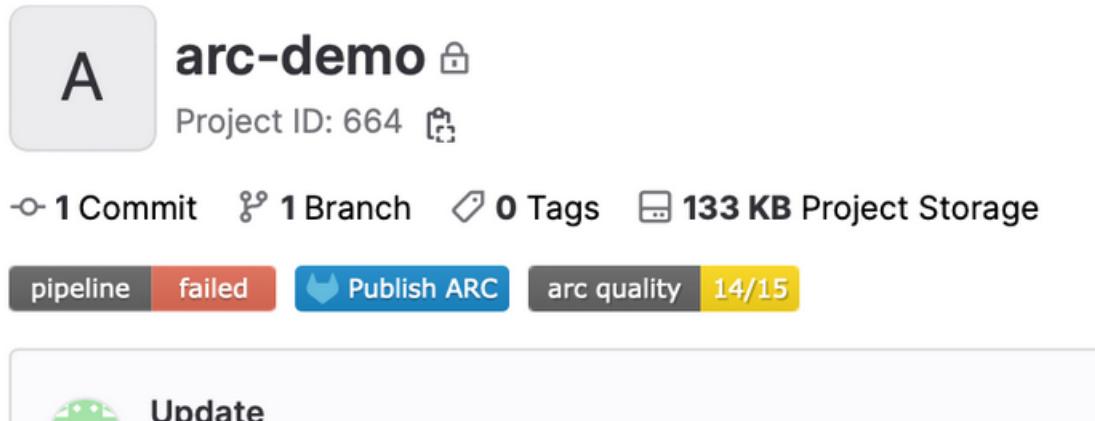
- Assay Metadata**
- Identifier**: gasexchangemeasurements
- Measurement Type**: leaf gas exchange
- Technology Type**: infrared gas analyzer
- Technology Platform**: licor 6800

A blue arrow labeled **Save metadata!** points to the save icon in the top right corner of the metadata form.

Pipeline failed

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

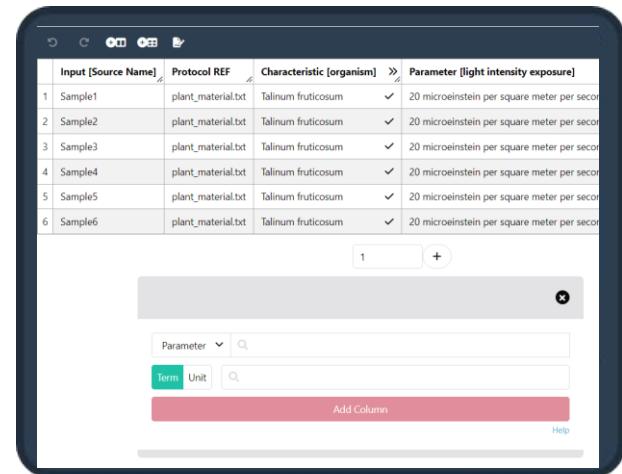
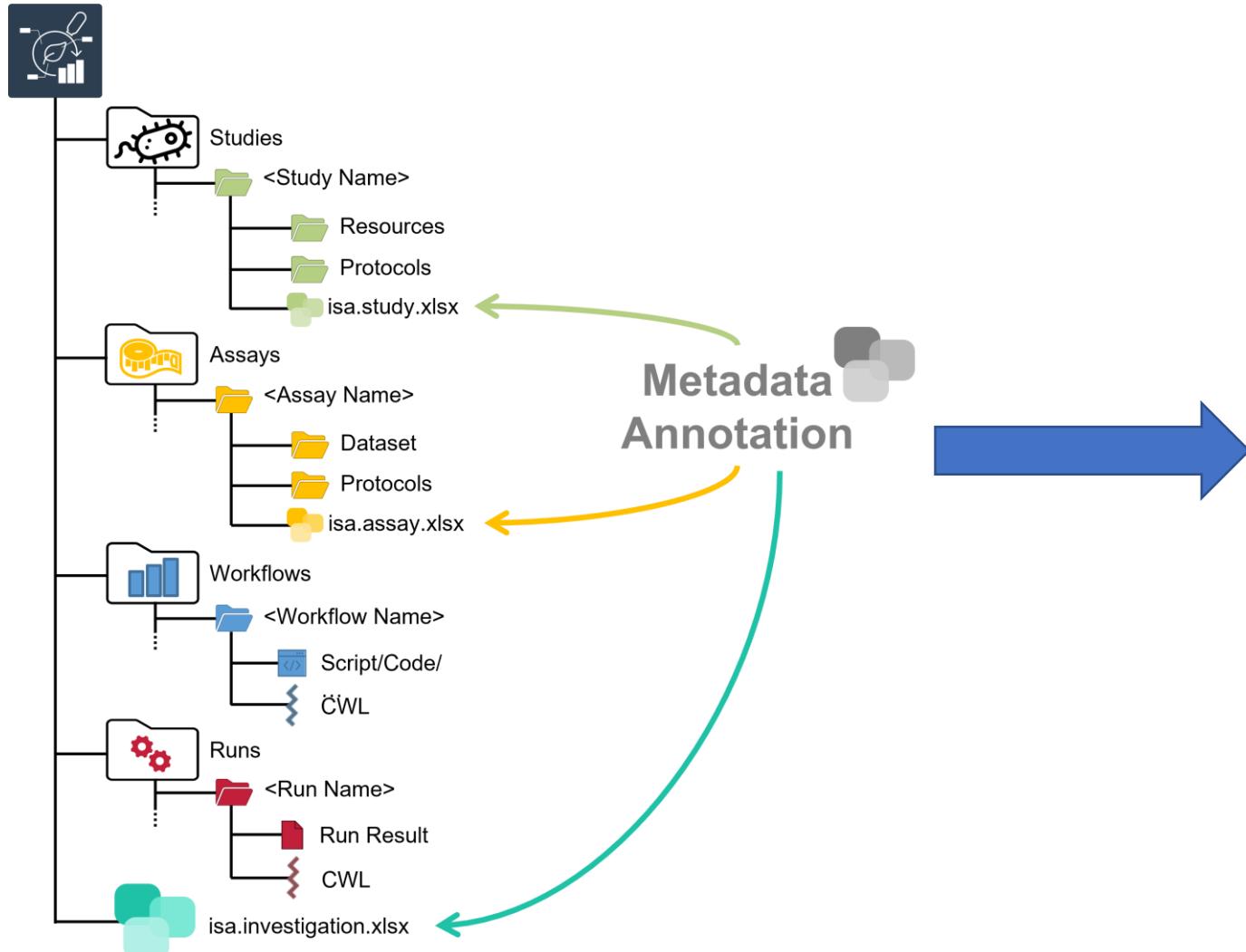
```
Investigation Identifier
Investigation Title
Investigation Description
Investigation Person Last Name
Investigation Person First Name
Investigation Person Email
Investigation Person Affiliation
```



If this happens, you will receive an email from GitLab

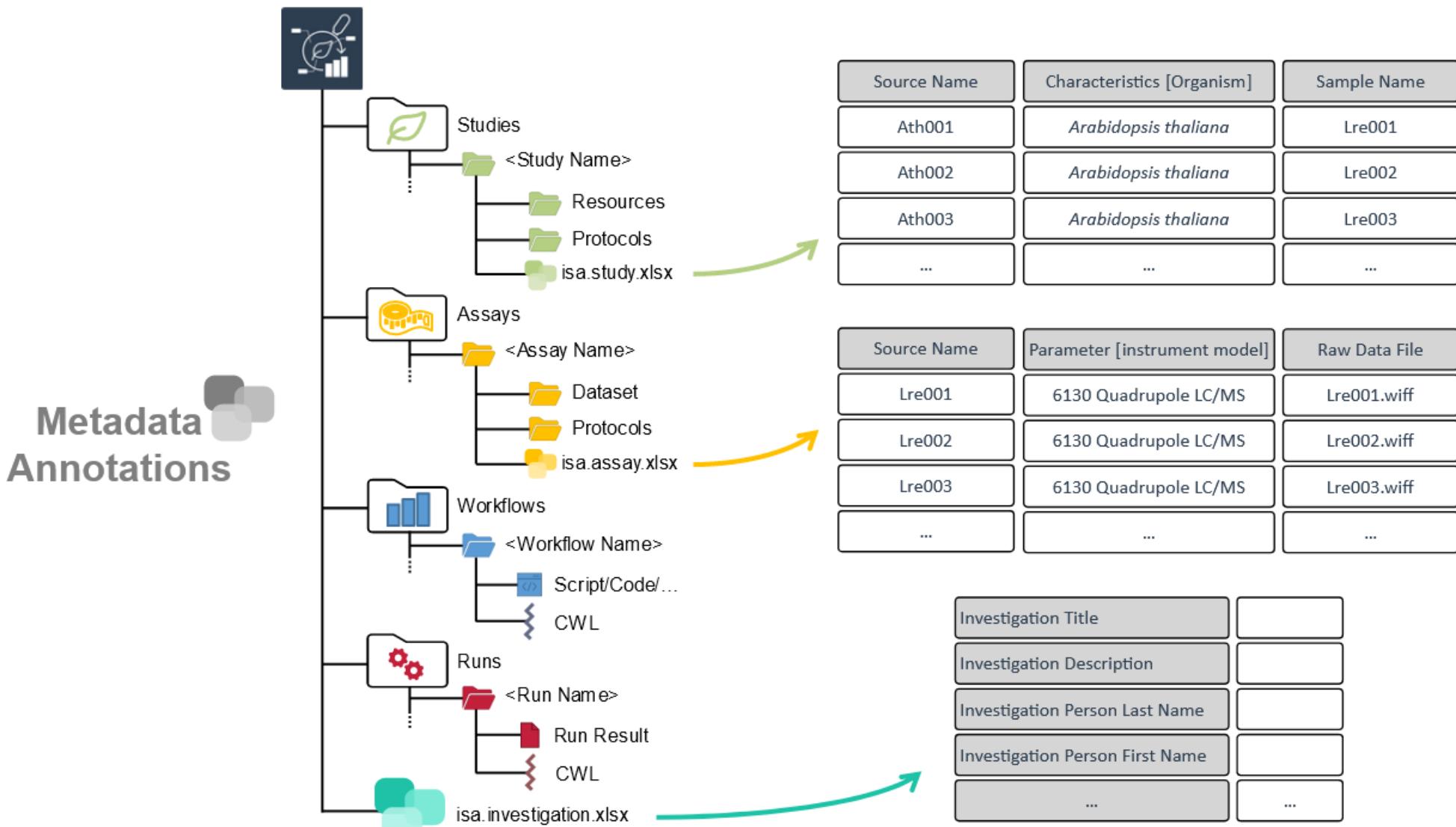
Metadata annotation

Metadata Annotation



SWATE

ARC builds on ISA to connect data



Swate as a standalone or ARCitect build-in

- The tool Swate is integrated into ARCitect and will automatically open once a new metadata sheet is added
- Alternatively Swate can be used in the browser: <https://swate-alpha.nfdi4plants.org/>

The image shows two screenshots of the ARCitect software interface. The left screenshot displays a file tree in a 'TestARC' folder on the desktop. A file named 'TestAssay' is highlighted with a green bar at the top, indicating it is the active or selected item. The right screenshot shows a detailed metadata entry form for the 'TestAssay' file. The form includes sections for Identifier (containing 'TestAssay'), Measurement Type (with fields for Term Name, TSR, and TAN), Technology Type (with fields for Term Name, TSR, and TAN), Technology Platform (with fields for Term Name, TSR, and TAN), and Performers (with a '+' button). A red box highlights the '+' button under the Performers section.

Swate overview

The screenshot illustrates the Swate interface, which includes the following components:

- Widgets**: A header bar containing various icons for file operations.
- Save & delete**: A button for saving and deleting data.
- Elements connected to one function**: A callout pointing to the central workspace where a table is displayed.
- Sidebar**: A vertical panel on the right side of the interface.
- Metadata sheet** and **Annotation table**: Labels indicating the types of data sheets available.
- Annotations**: A section at the bottom left showing a tree structure with "Metadata" and "plant_material" under it, with a plus sign indicating expandability.
- SWATE**: The main title of the application.
- Ontology term search**: A search bar for finding ontology terms.
- Search for an ontology term to fill into the selected field(s)**: A placeholder text for the search bar.
- Talinum fruticosum**: A search result entry.
- Parent: Organism, MIAPPE:0041**: The category and identifier for the search result.
- Use advanced search**: A link to more search options.
- Fill selected cells with this term**: A button to apply the search result to the workspace.
- Swate Release Version v1.0.0-beta.03 Host Browser**: The software version information.

Input [Source Name]	Characteristic [Organism]	Factor [watering exposure]	Output [Sample Name]
1 DB_097	Talinum fruticosum ✓	12 days drought	CAM_01
2 DB_099	Talinum fruticosum ✓	12 days drought	CAM_02
3 DB_103	Talinum fruticosum ✓	12 days drought	CAM_03
4 DB_161	Talinum fruticosum ✓	12 days drought + 2 days rewatered	reC3_01
5 DB_163	Talinum fruticosum ✓	12 days drought + 2 days rewatered	reC3_02
6 DB_165	Talinum fruticosum ✓	12 days drought + 2 days rewatered	reC3_03

Adding new building blocks

The screenshot shows the Swate software interface. On the left, there is a table with six rows of experimental data:

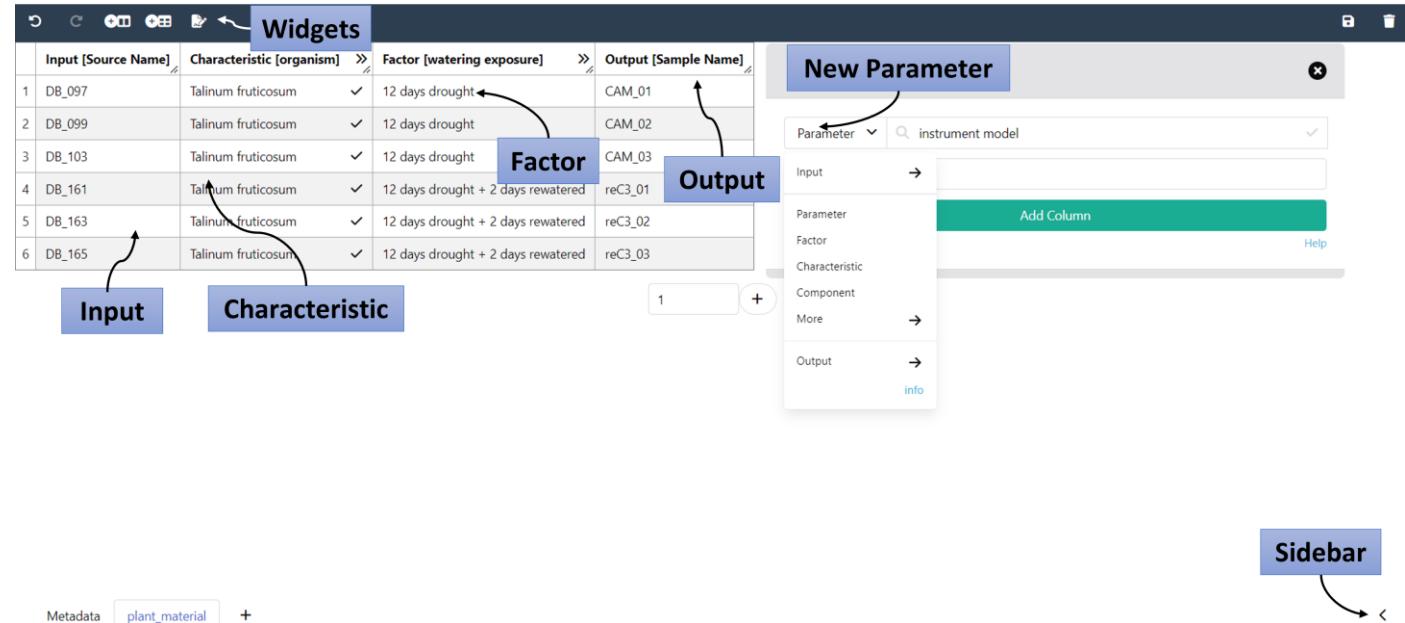
	Input [Source Name]	Characteristic [organism]	Factor [watering exposure]	Output [Sample Name]
1	DB_097	Talinum fruticosum	✓ 12 days drought	CAM_01
2	DB_099	Talinum fruticosum	✓ 12 days drought	CAM_02
3	DB_103	Talinum fruticosum	✓ 12 days drought	CAM_03
4	DB_161	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_01
5	DB_163	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_02
6	DB_165	Talinum fruticosum	✓ 12 days drought + 2 days rewatered	reC3_03

Below the table is a toolbar with various icons. A blue arrow points downwards from the toolbar towards the 'New building blocks' dialog box. The dialog box is titled 'New building blocks' and contains a search bar with the placeholder 'instrument model'. It also includes sections for 'Input', 'Parameter', 'Factor', 'Characteristic', 'Component', 'More', and 'Output'. A large green button labeled 'Add Column' is prominently displayed. At the bottom right of the dialog box is a 'Help' link.

Swate can be used for the annotation of **isa.study.xlsx** and **isa.assay.xlsx** files

Annotation building block types

- Input (e.g. Source Name, Sample Name, Data)
- Protocol columns
- Characteristic// Parameter// Component// Factor
- Output (e.g. Sample Name, Data)



Fill your table with ontology terms

New Assay + + ↻ ↺

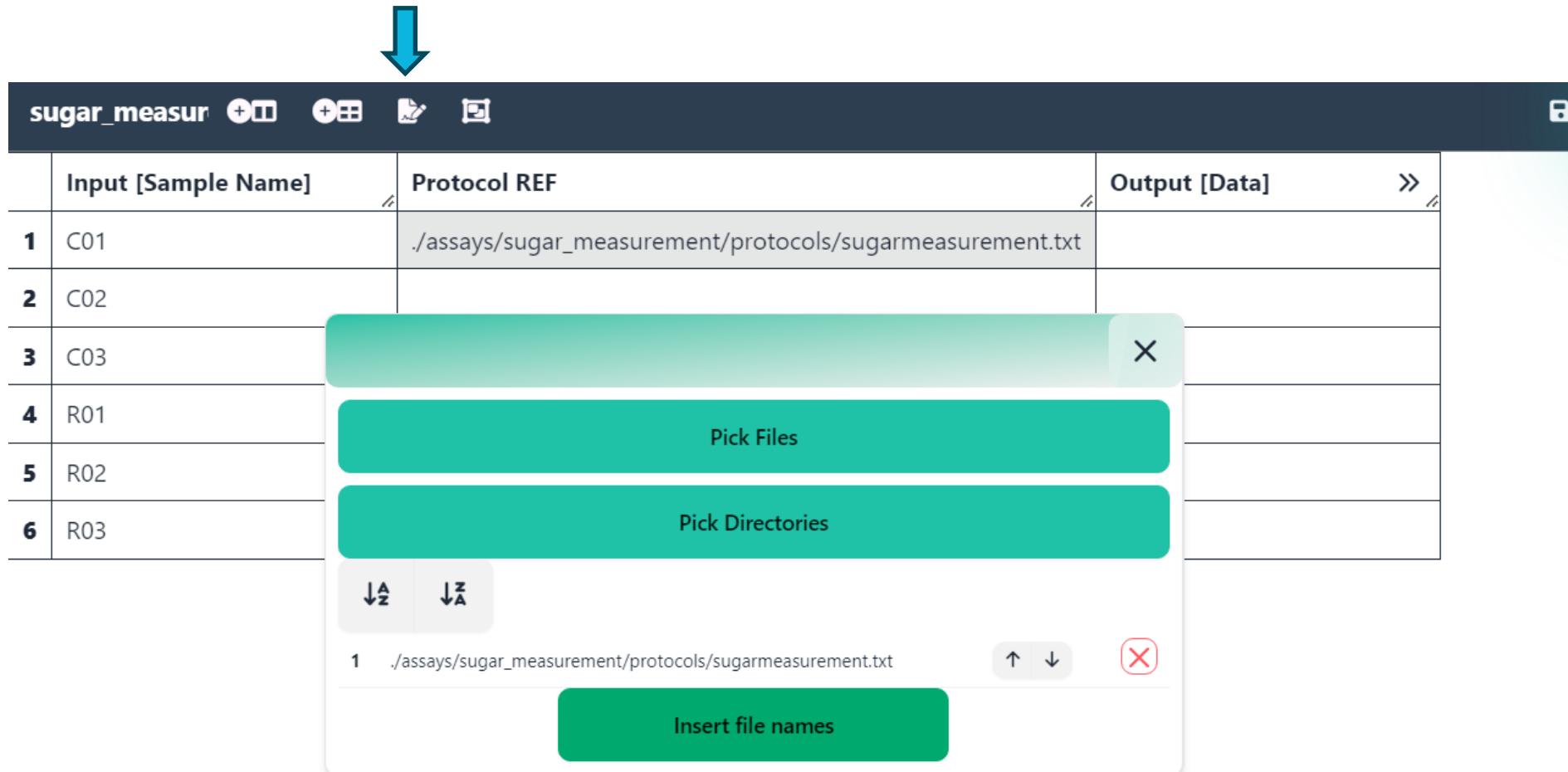
	Input [Source Name]	Characteristic [organism]	»	Parameter [watering exposure]	»	Output [Sample Name]
1	DB1	Zea		12 days drought		C01
2	DB2	Zea	▼	ght		C02
3	DB3	Zea mays	▼	ght		C03
4	DB4	Zea mays	▼	ght + 2 days rewatered		R01
5	DB5	Zea perennis	▼	ght + 2 days rewatered		R02
6	DB6	Zea perennis	▼	ght + 2 days rewatered		R03
		Zea mexicana	▼			
		Zea luxurians	▼			
		Helicoverpa zea	▼			
		Helicoverpa zea	▼			
		Zea mays Pollen	▼			
		Zea nicaraguensis	▼			

• Automatic ontology term search in cells
• Related terms are marked by green symbol

1 ↻ +

Assay New Assay Table +

Picking file names for protocol or data reference



- Easily copy file and directory names for protocol or data files

Refer to columns/rows/cells of a data file



sugar_measur		+ II	+ III	✎	☒
	Input [Sample Name]	Protocol REF			Output [Data] »
1	C01	./assays/sugar_measurement/protocols/sugarmeasurement.txt			sugar_result.csv#col=2
2	C02				

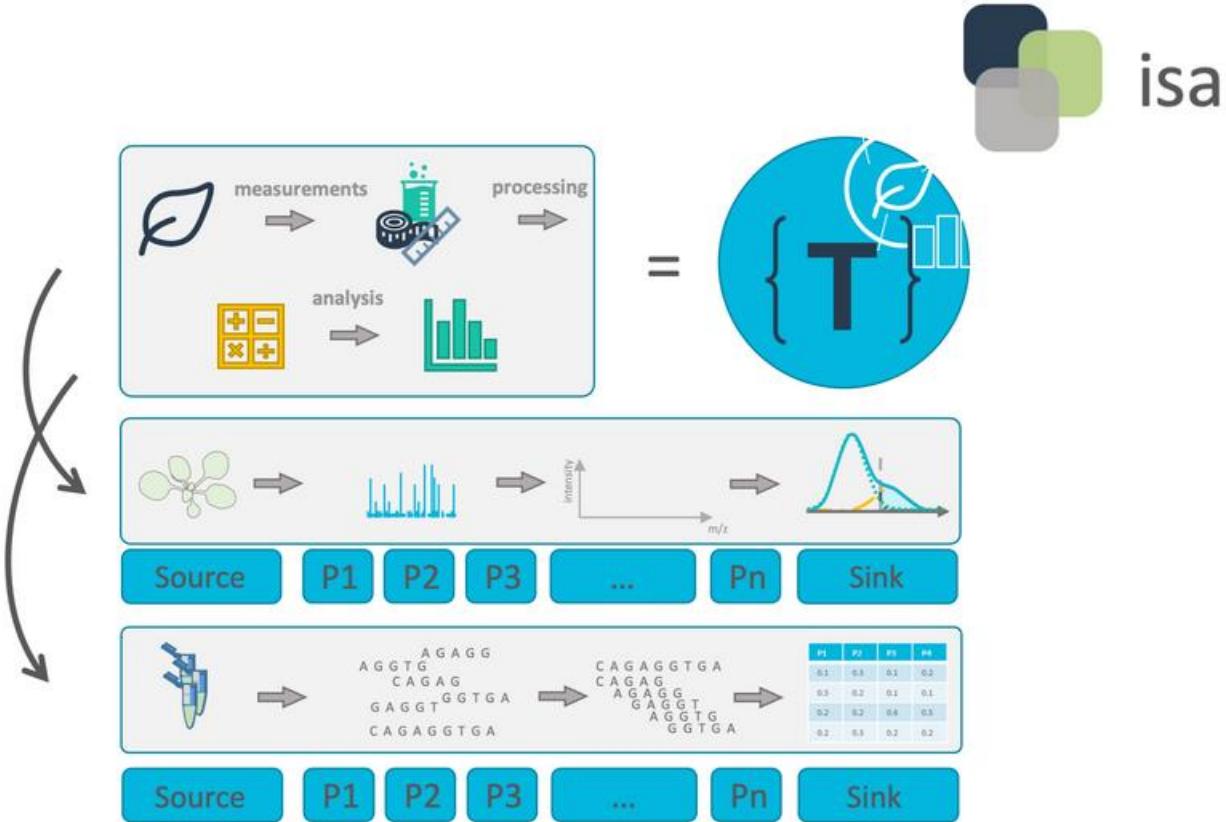
Data Annotator

.. update separator Update Has Header Output ⓘ Reset

sugar_result.csv - text/csv

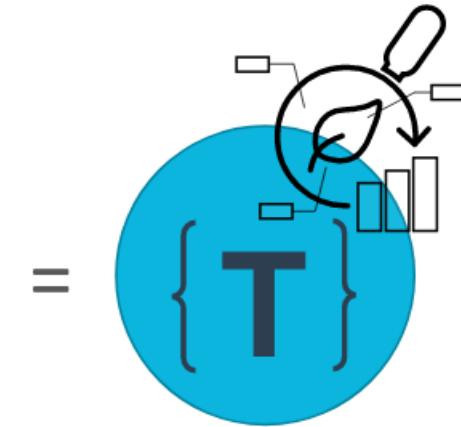
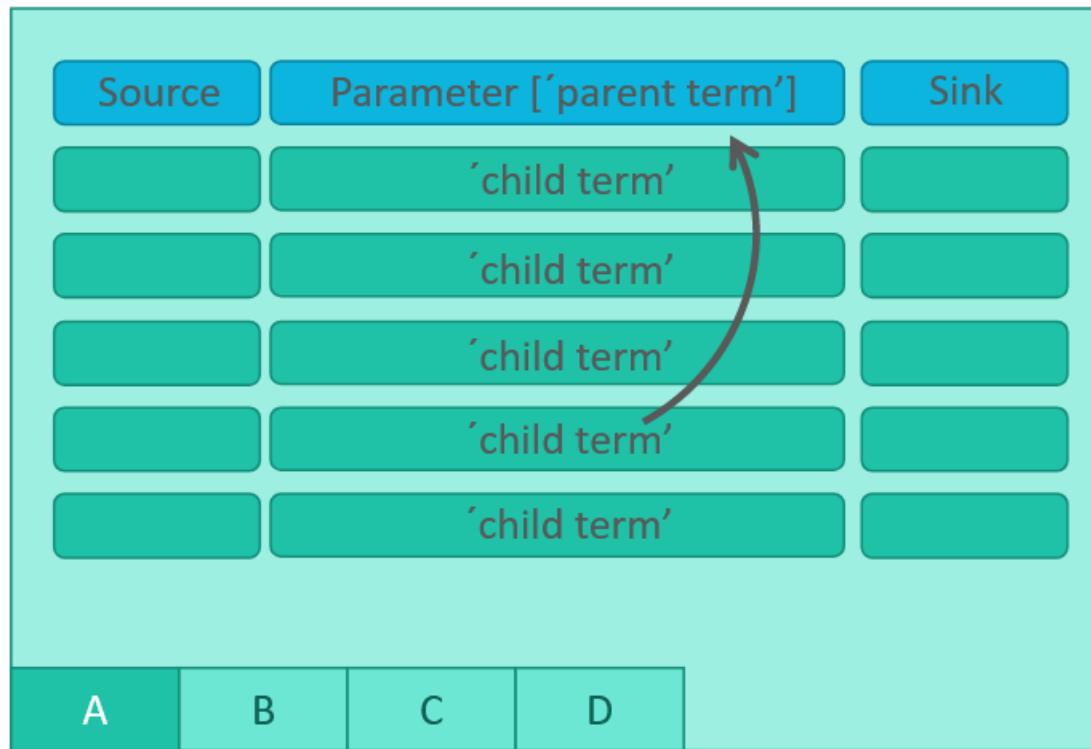
	MeasuredSugar	C01	C02	C03	R01	R02	R03
0	Glucose	27.671	24.881	22.634	24.665	32.118	29.067
1	Fructose	9.283	7.913	7.945	8.752	9.679	9.671
2	Sucrose	9.294	7.272	8.236	7.612	9.495	9.344

Realization of lab-specific metadata templates



Facilities can define their most common workflows as templates

Checklists and templates



Metadata standards or repository requirements can be represented as templates

Directly import templates via Swate

- DataPLANT curated
- Community templates

Search by template name

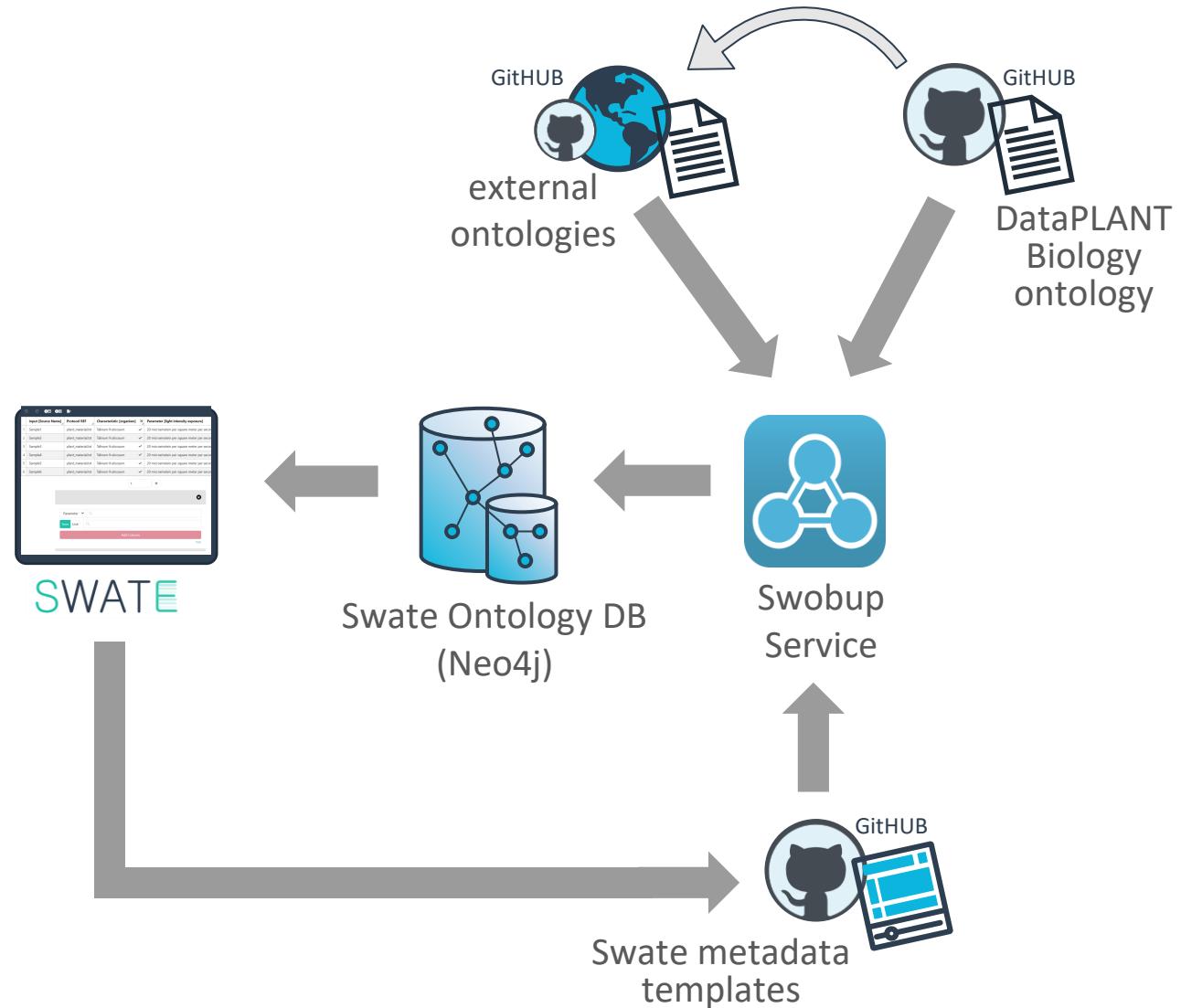
Search for tags

Select community

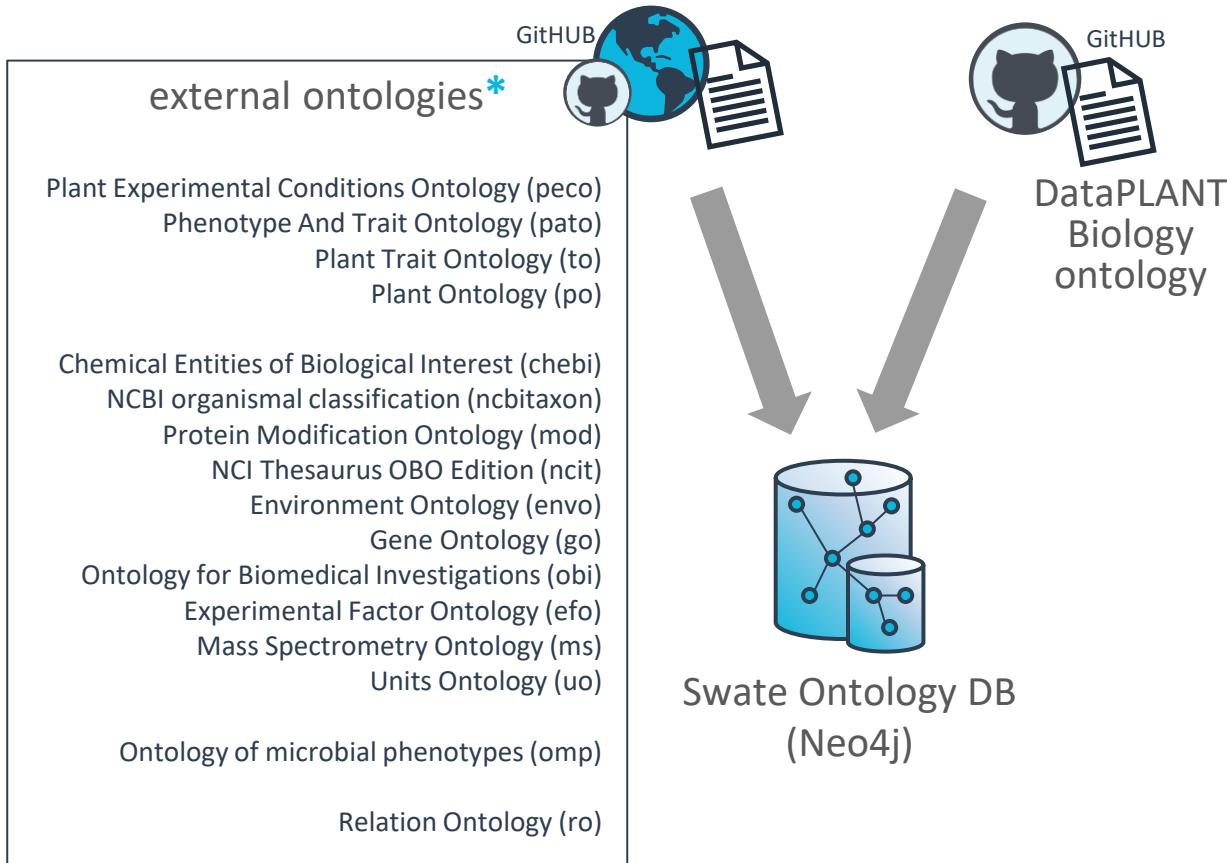
Template Name	Community	Template Version	Actions
ArrayExpress - Plant sample	curated	1.0.2	▼
ArrayExpress - Sequencing library	curated	1.0.2	▼
DNA extraction	curated	1.1.7	▼
ENA - Gene promoter annotated sequence	curated	1.0.1	▼
ENA - Plant sample	curated	1.0.1	▼
ENA - Plant sample - extension	curated	1.0.1	▼
ENA - Raw sequencing reads	curated	1.0.1	▼
ENA - Single CDS annotated sequence	curated	1.0.1	▼

[Help](#)

Ontology Service



Ontologies within SwateDB



*Collection can be easily adjusted to fit the needs of other research domains

- External ontologies covering a range of domains
 - plant sciences
 - biology and chemistry
 - experimental
 - upper level
- DataPLANT biology ontology (dpbo)
 - missing vocabulary needed for metadata annotation

DataPLANT Biology Ontology (dpbo)



Import ontology
Request the import of an already existing ontology into the SwateDB if you need >60% of terms from the ontology to annotate your data.

Import term
Import term from a registered ontology into dpbo

Add synonym
New synonym suggestion for a dpbo term.

Add new term
Suggest a new term to be added to dpbo

Custom issue template
Please use this template for all other issues or concerns

Fix/add xref
Suggest a new ref or report issue about an incorrect xref.

Get started

Get started

Get started

Get started

Get started

Get started

Edit templates

Don't see your issue here? [Open a blank issue.](#)

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



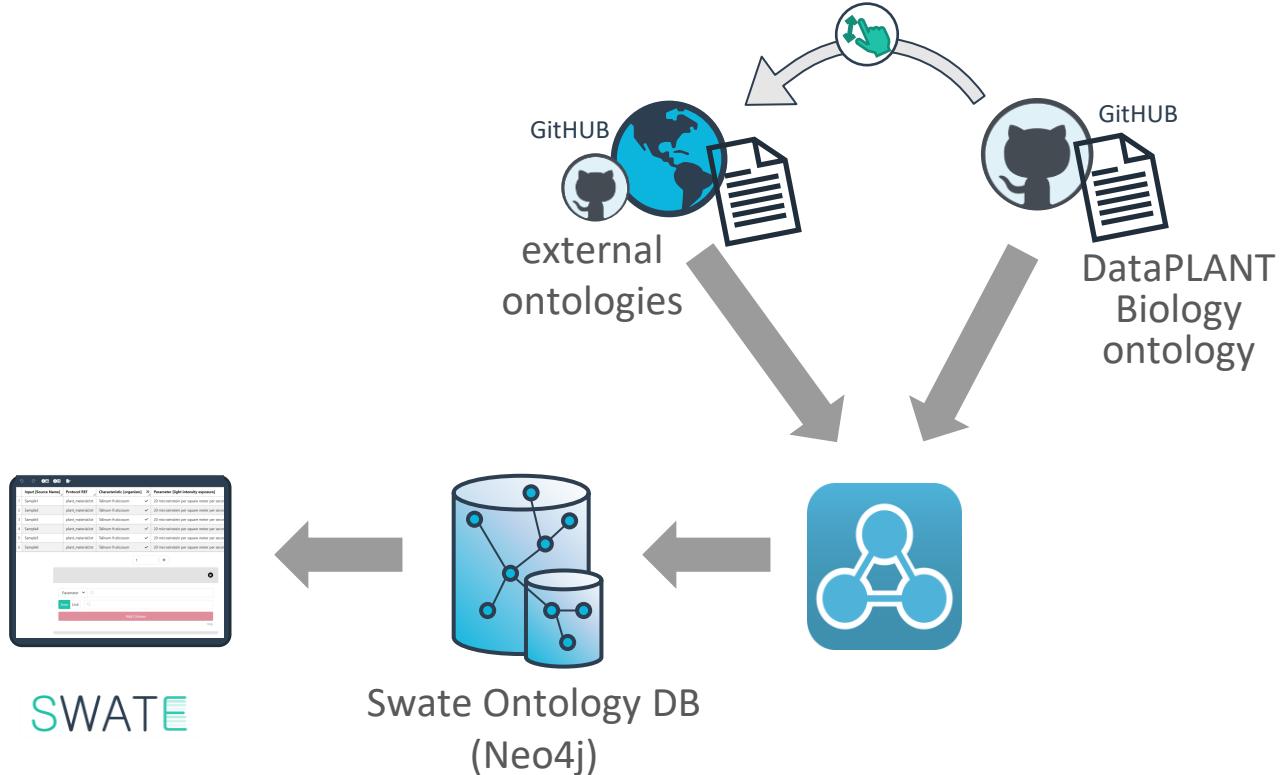
Submit a ticket

Please fill out this form to submit a new request to DataPLANT

- Users can submit issues to
 - suggest new terms
 - import terms or
 - improve already existing terms
- Users can also suggest other external ontologies to add to the SwateDB

https://github.com/nfdi4plants/nfdi4plants_ontology

Bridging the ontology gap



The DataPLANT biology ontology facilitates the exchange of knowledge between the researcher and the main ontology provider.

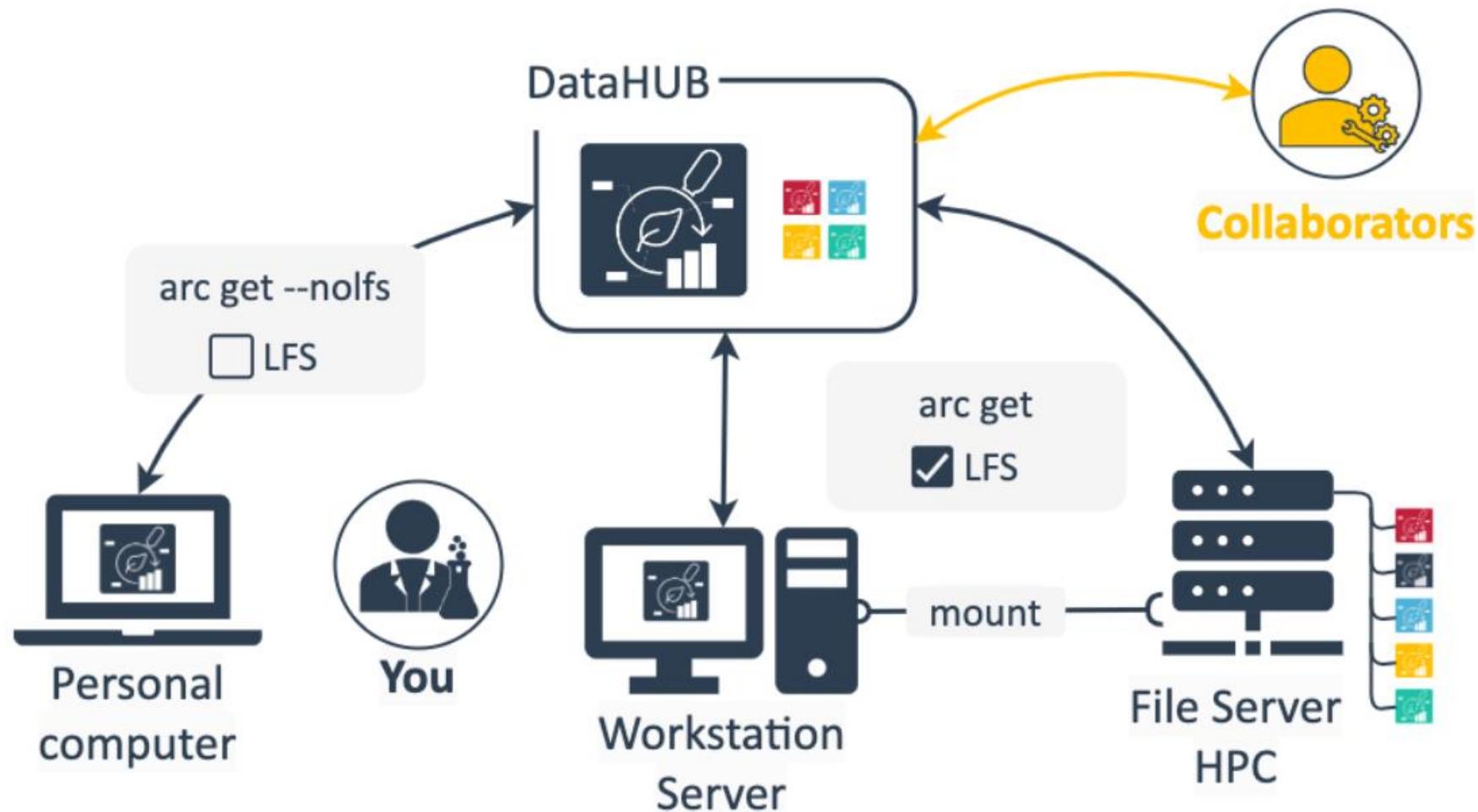
DataPLANT will act as a broker by suggesting terms back to relevant ontologies, thereby improving these external ontologies.

Contributions are always welcomed!

https://github.com/nfdi4plants/nfdi4plants_ontology

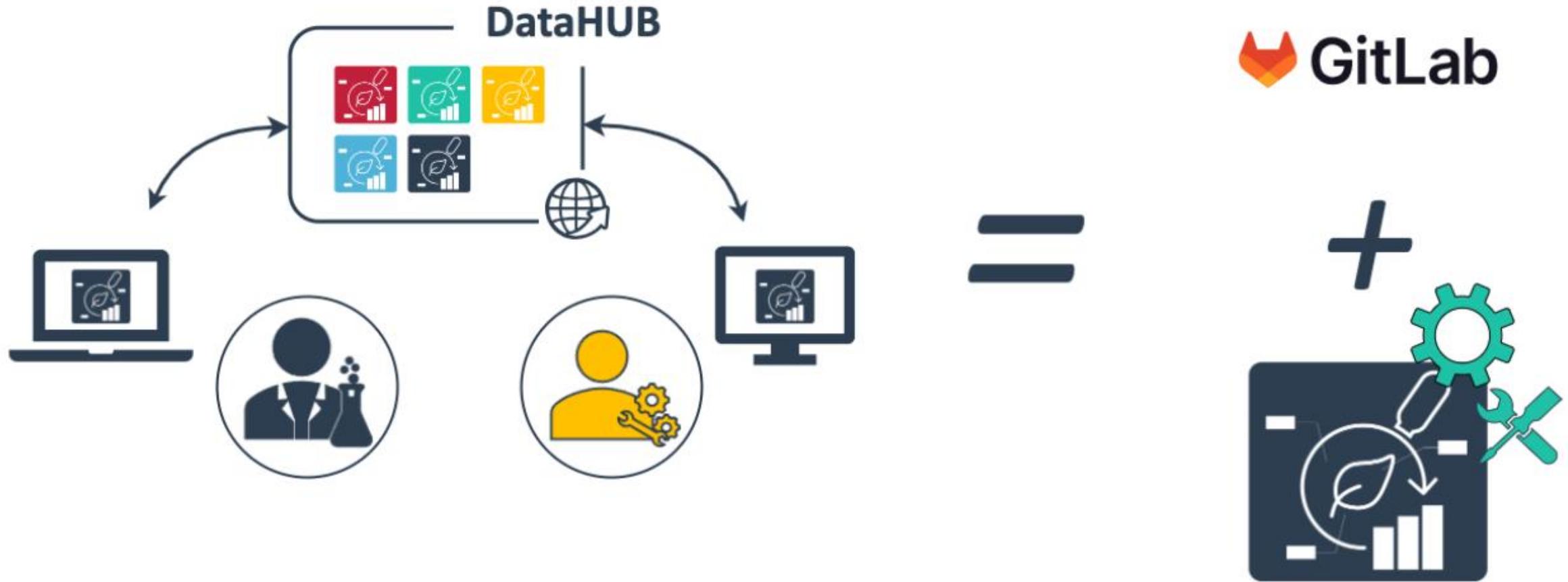
DataHUB

Where do I store my ARC?

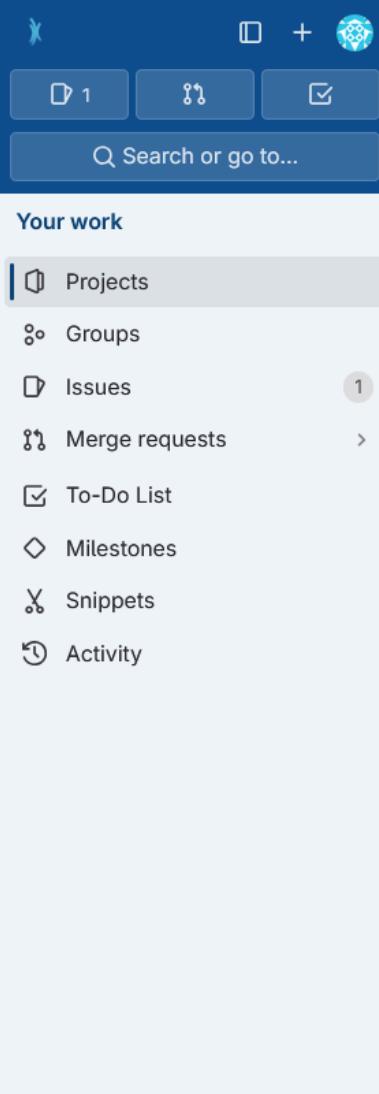


💡 ARCitect and ARC commander provide options to avoid syncing large files (LFS = Large file storage)

The DataPLANT DataHUB – a GitLab *Plus*



DataPLANT DataHUB



The sidebar on the left contains the following navigation items:

- Projects (selected)
- Groups
- Issues (1)
- Merge requests
- To-Do List
- Milestones
- Snippets
- Activity

Your work / Projects

Projects

[Explore projects](#) [New project](#)

Yours 60 Starred 0 Personal Inactive



Search or filter results...



Updated

	Usadellab / Barvista_ARC  Owner	 ★ 1 ⚡ 0 📈 0 📂 0	Updated 25 minutes ago
	Usadellab / szymanskilab / arab_env_2024i  Owner	 ★ 0 ⚡ 0 📈 0 📂 0	Updated 26 minutes ago
	Usadellab / ARCYeast  Owner	 ★ 1 ⚡ 0 📈 0 📂 0	Updated 29 minutes ago
	Usadellab / Lupins  Owner	 ★ 0 ⚡ 0 📈 0 📂 0	Updated 3 hours ago
	Usadellab / szymanskilab / genRE  Owner	 ★ 0 ⚡ 0 📈 0 📂 0	Updated 5 days ago
	Usadellab / szymanskilab / gene_characterisation_dCRE  Owner	 ★ 0 ⚡ 0 📈 0 📂 0	Updated 6 days ago

Project = ARC

- In the DataHUB, ARCs are called "projects"; they are the same.
- An ARC can be shared with individual users (invited as "members") or a group.

DataHUB groups

- A "Group" is a group of users with specific [permissions](#)
- A group can share ARCs
- A group can be invited to an ARC
- Groups can have subgroups

Namespaces

- Every user has a personal namespace, where they can upload or create new ARCs
- Every group and subgroup has its own namespace

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

 **Personal** is not the same as **private**

Visibility

The visibility of ARCs and groups can be managed individually for each ARC or group

Visibility

The visibility of each ARC can be managed in the settings of the ARC



Private – ARC access must be granted explicitly to each user or group.



Internal – ARC can be accessed by any logged in user.



Public – ARC can be accessed without authentication.



By default every ARC and every group is set to **private**.

Permissions and Roles

When inviting new members to an ARC or group, you can choose between different levels.

Permissions & Roles

Roles are assigned when adding a user to an ARC or to a group. This is a very simplified summary.

- | | |
|-------------------|---|
| Guest | – Can only see the ARC's wiki |
| Reporter | – Can read, but not add or edit data |
| Developer | – Reporter permissions + can read, add, and edit data |
| Maintainer | – Developer permissions + can add new members |
| Owner | – Maintainer + can delete ARC, manage memberships and permissions |



By default you are **Owner** of an ARC you create or upload to the DataHUB.

ARC DataHUB members // ARC investigation contacts

The screenshot shows the 'Members' section of a project named 'Samuilov-2018-BOU-PSP'. It lists 18 members and 1 group. The table includes columns for Account, Source, Max role, Expiration, and Activity. The activity column displays the creation date, grant date, and last activity for each member.

Account	Source	Max role	Expiration	Activity
Adriano Nunes-Nesi @nunesnesi	HHU Plant Biochemistry by Sebastian Triesch	Maintainer	Expiration date	User created: Jul 05, 2023 Access granted: Jul 10, 2023 Last activity: Aug 09, 2023
Andreas Weber @andreas.weber	HHU Plant Biochemistry by Sebastian Triesch	Maintainer	Expiration date	User created: Mar 10, 2023 Access granted: Jul 31, 2023 Last activity: Sep 11, 2023
Dominik Brilhaus It's you @brilator	Direct member by Dominik Brilhaus	Owner	Expiration date	User created: Feb 21, 2022 Access granted: Dec 06, 2023 Last activity: Mar 26, 2024
Franziska Fichtner @franziska.fichtner	HHU Plant Biochemistry by Sebastian Triesch	Maintainer	Expiration date	User created: Aug 11, 2023 Access granted: Aug 11, 2023 Last activity: Aug 11, 2023

DataHUB: ARC members

https://git.nfdi4plants.org/hhu-plant-biochemistry/Samuilov-2018-BOU-PSP/-/project_members

The screenshot shows the 'Investigation Contacts' section of the ARCIctect interface. It lists several contacts with their names, ORCID IDs, and scores. The contacts listed are Sladjana Samuilov, Nadine Rademacher, Samantha Flachbart, Leila Arab, Saleh Alfarraj, Franziska Kuhnert, Stanislav Kopriva, Andreas P. M. Weber, and Tabea Mettler-Altmann.

Contact	Score
Sladjana Samuilov <orcid>	4/10
Nadine Rademacher <orcid>	3/10
Samantha Flachbart <orcid>	3/10
Leila Arab <orcid>	3/10
Saleh Alfarraj <orcid>	3/10
Franziska Kuhnert <orcid>	3/10
Stanislav Kopriva <orcid>	3/10
Andreas P. M. Weber <orcid>	4/10
Tabea Mettler-Altmann <orcid>	3/10

ARCIctect: Investigation Contacts

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

Explore public ARCs

Explore / Projects

Explore ARCs

Most starred Trending Active Inactive All

Search or filter results...

Rank	User	Project Name	Description	Metrics	Last Updated
A	Timo Mühlhaus	ArcPrototype	A prototypic ARC that implements all specification standards accordingly	3 ⚡ 2 🔍 1 📂	Updated 2 months ago
B	HHU Plant Biochemistry	CEPLAS RNASeq Workshop 2022	Creative Commons Attribution Non Commercial Share Alike 4.0 International	2 ⚡ 0 🔍 0 📂	Updated 5 months ago
C	Benedikt Venn	Ru_ChlamyHeatstress	Algae cultures were grown mixotrophically (TAP). After 24h of 35°C/40°C the cells were shifted back to room temperature for 48h. 'omics samples were taken.	2 ⚡ 1 🔍 0 📂	Updated 1 month ago
D	Felix Jung	deepSTABp	MIT License	2 ⚡ 1 🔍 0 📂	Updated 5 months ago
E	Usadelab	Barvista_ARC	Early patterning of organ primordia during barley meristem development uncovered by imputation of gene expression at single cell level	1 ⚡ 0 🔍 0 📂	Updated 28 minutes ago
F	CEPLAS	Sphingomonas	This ARC contains the data, analysis and results of the pangenomic analysis of plant-associated bacteria with an emphasis on what separates plant-associated Sphingomonadaceae genomes from other plant-associated bacteria.	1 ⚡ 0 🔍 0 📂	Updated 1 month ago
G	Natural-Variation-and-Evolution	Microscopy_Collection	map-by-seq_CLSM-stacks	1 ⚡ 0 🔍 0 📂	Updated 6 months ago

Projects Groups CI/CD Catalog Topics Snippets Help

New project

Find published ARCs in the ARChive

The screenshot shows a dataset record on the DataPLANT platform. The top navigation bar includes the DataPLANT logo, a search bar, 'Communities' and 'My dashboard' links, and a 'Log in' button. Below the header, a sidebar on the left features a 'DataPLANT' icon and a 'DataPLANT' link. The main content area displays the following information:

Published May 27, 2024 | Version v1

In depth exploration of the genomic and metabolic diversity in tea varieties based on a newly constructed pan-genome of *Camellia sinensis*

Tariq, Arslan¹; Meng, Minghui²; Jiang, Xiaohui²; Bolger, Anthony³; Beier, Sebastian³; Buchmann, Jan¹; Fernie, Alisdair⁴; Wen, Weiwei²; Usadel, Björn³

[Show affiliations](#)

Citation

Tariq, A., Meng, M., Jiang, X., Bolger, A., Beier, S., Buchmann, J., Fernie, A., Wen, W., & Usadel, B. (2024). In depth exploration of the genomic and metabolic diversity in tea varieties based on a newly constructed pan-genome of *Camellia sinensis* [Data set]. DataPLANT. <https://doi.org/10.60534/30cpm-z1c67>

Style [APA](#)

Versions

Version v1	May 27, 2024
10.60534/30cpm-z1c67	

Details

DOI [DOI 10.60534/30cpm-z1c67](#)

Resource type Dataset

Publisher DataPLANT

Export

Looking for help? Use...

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

ARC user support chat: arc-user-support.slack.com
or open an issue on github

Find more information on...

<https://nfdi4plants.org/>

<https://git.nfdi4plants.org>

Acknowledgements



Björn Usadel
Kathryn Dumschott
Xiaoran Zhou
Hannah Dörpholz
Dominik Brilhaus
Sabrina Zander

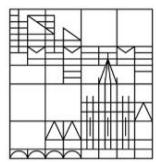
Timo Mühlhaus
Jens Krüger
Dirk von Suchodoletz



& the whole team at DataPLANT



Universität Konstanz



RUHR
UNIVERSITÄT
BOCHUM



Heinrich Heine
Universität
Düsseldorf



UNIVERSITÄT
HOHENHEIM



HelmholtzZentrum münchen
German Research Center for Environmental Health

