



CEPLAS

Cluster of Excellence on Plant Sciences

Start Your ARC Workshop Series

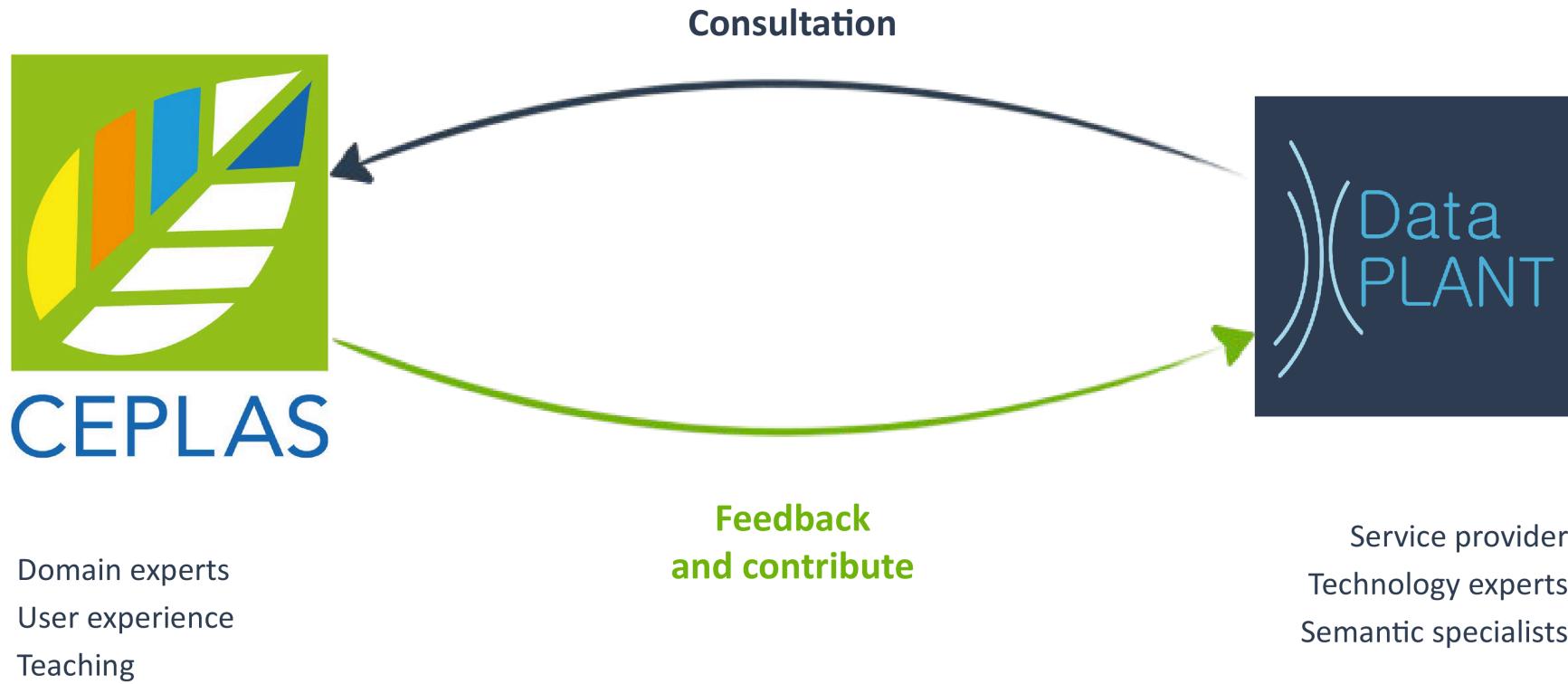
Session 01 - Introduction to the ARC (demo)

May 9th, 2023

Dominik Brilhaus - CEPLAS Data Science



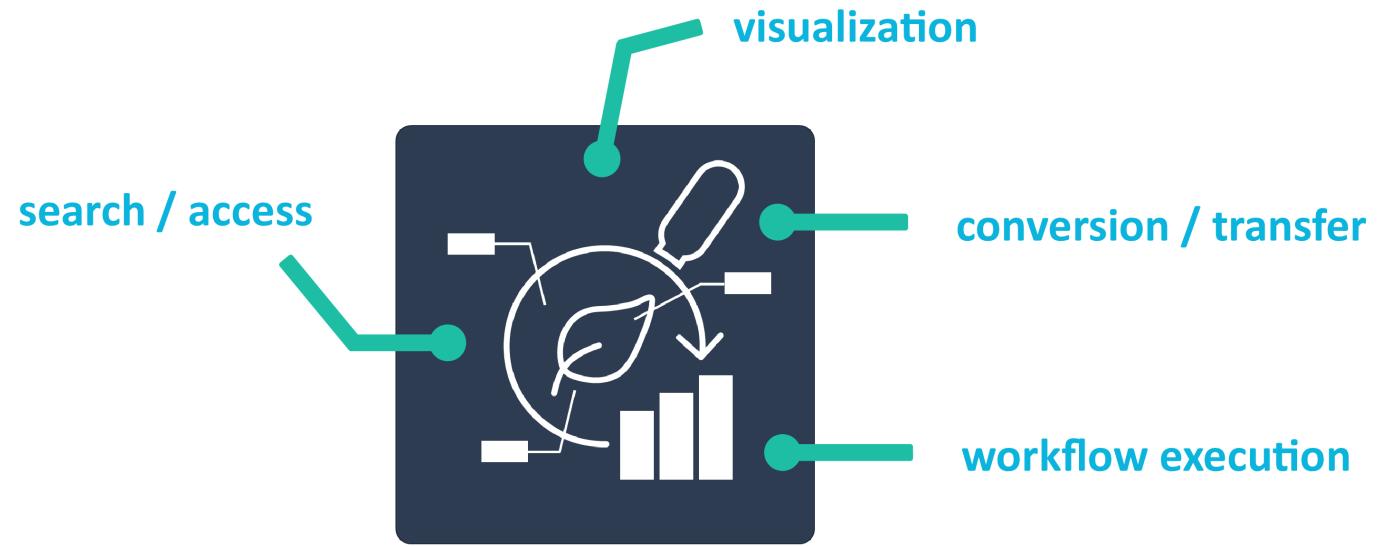
CEPLAS collaborates with DataPLANT



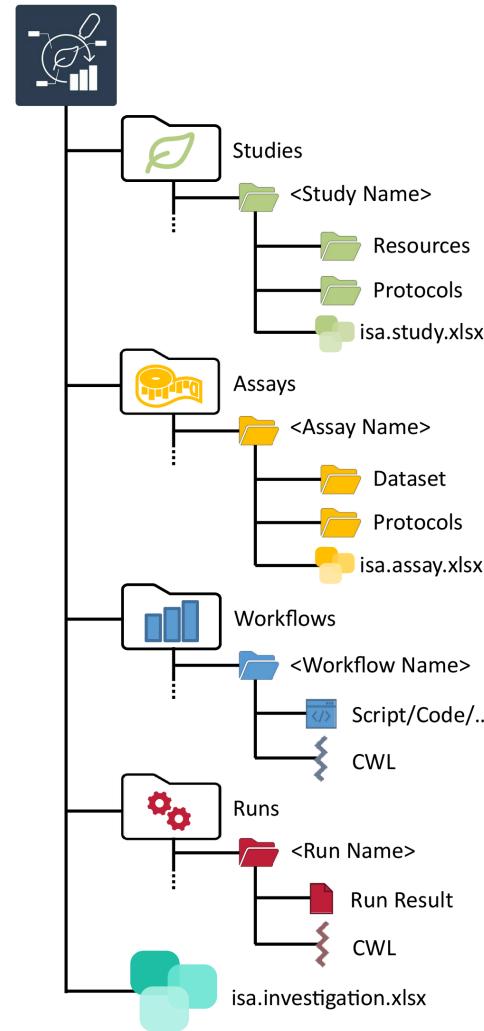
CEPLAS collaborates with DataPLANT



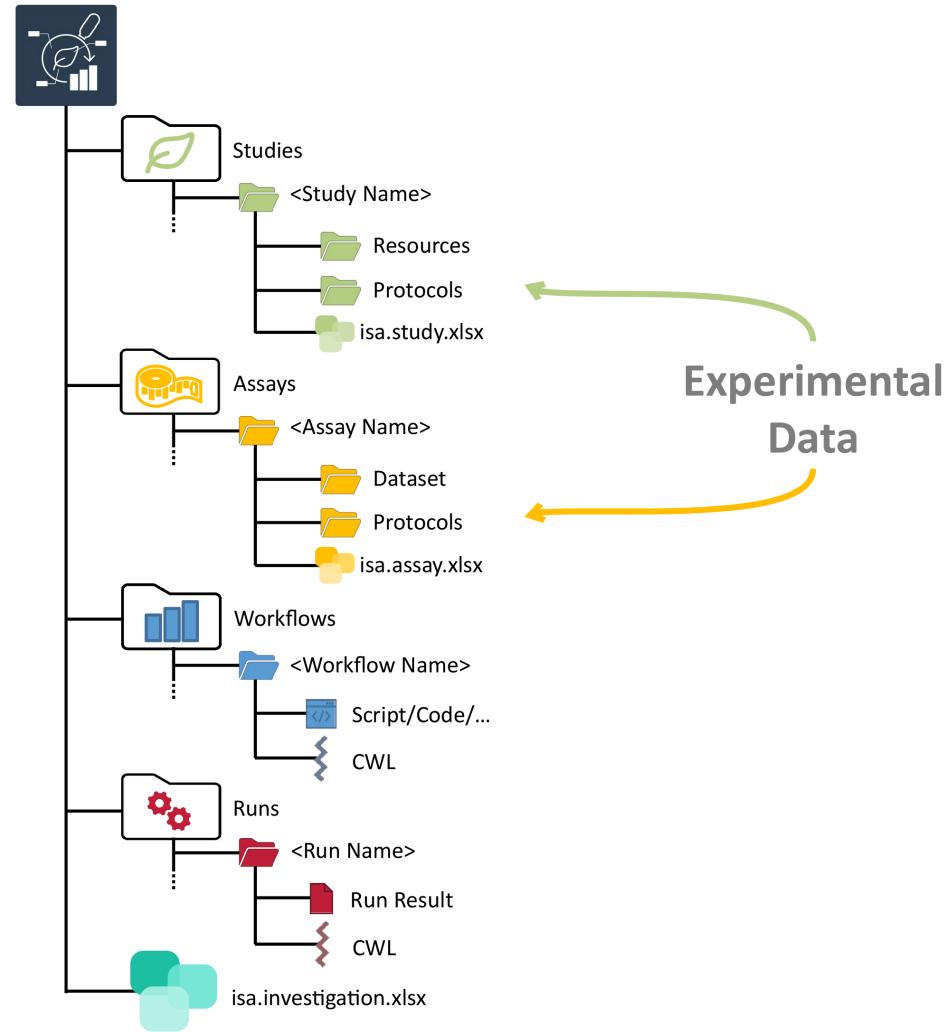
Annotated Research Context (ARC)



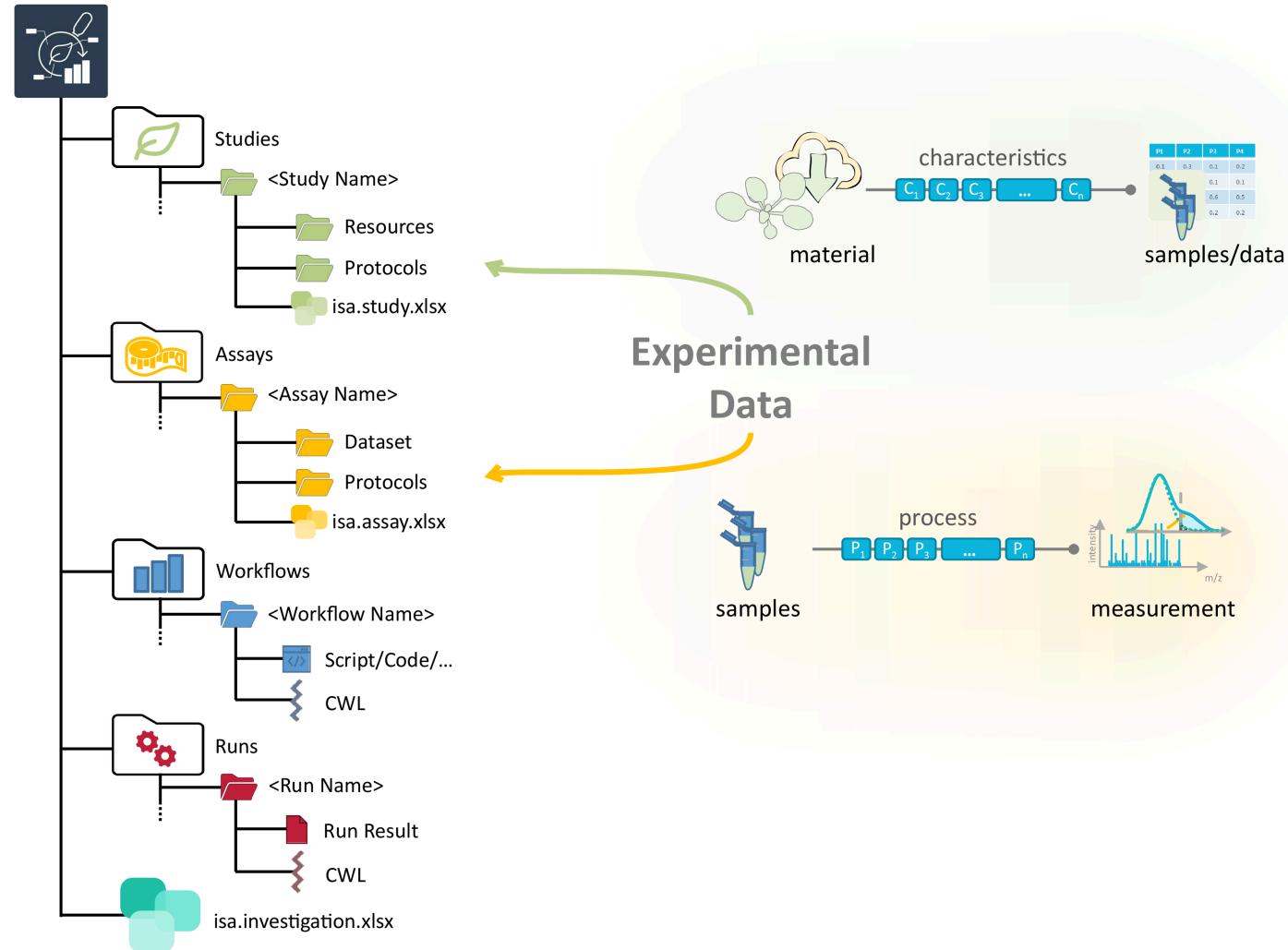
What does an ARC look like?



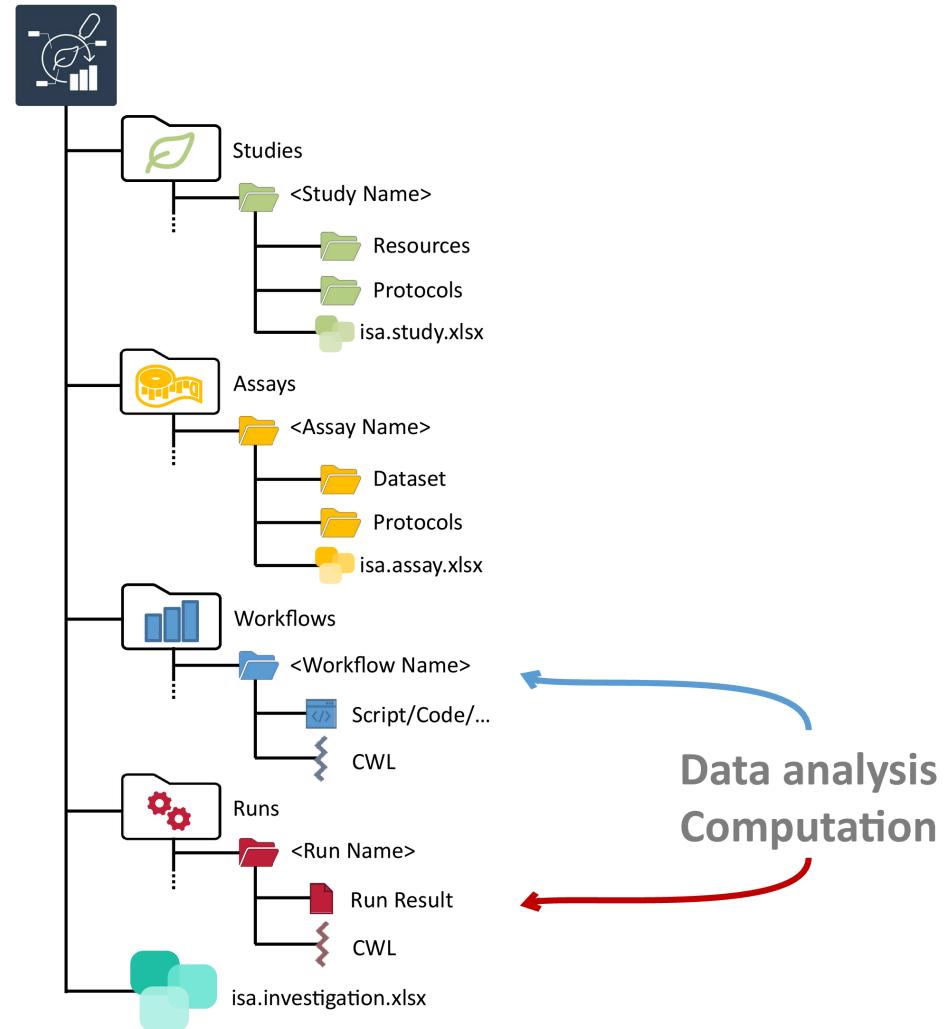
What does an ARC look like?



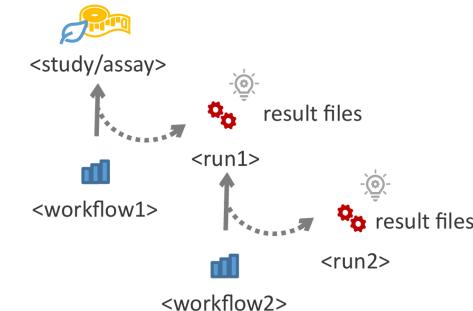
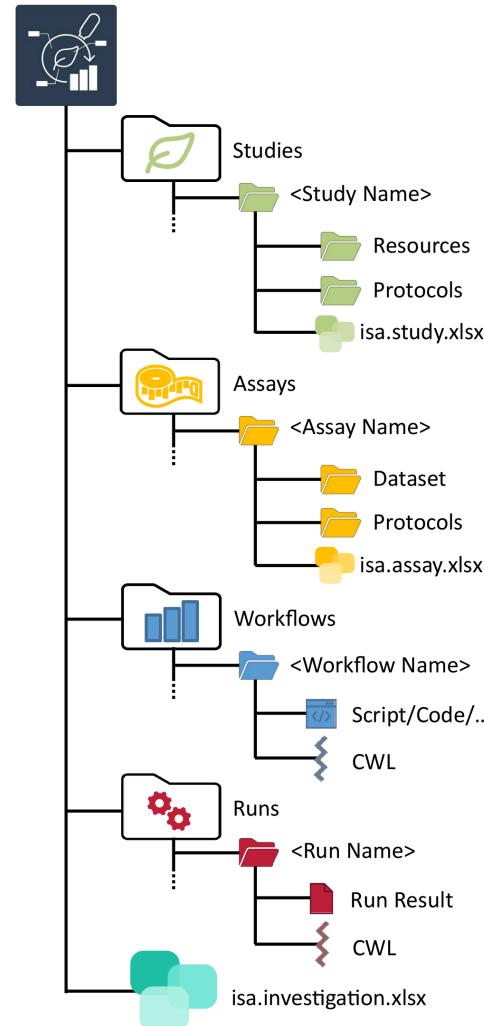
What does an ARC look like?



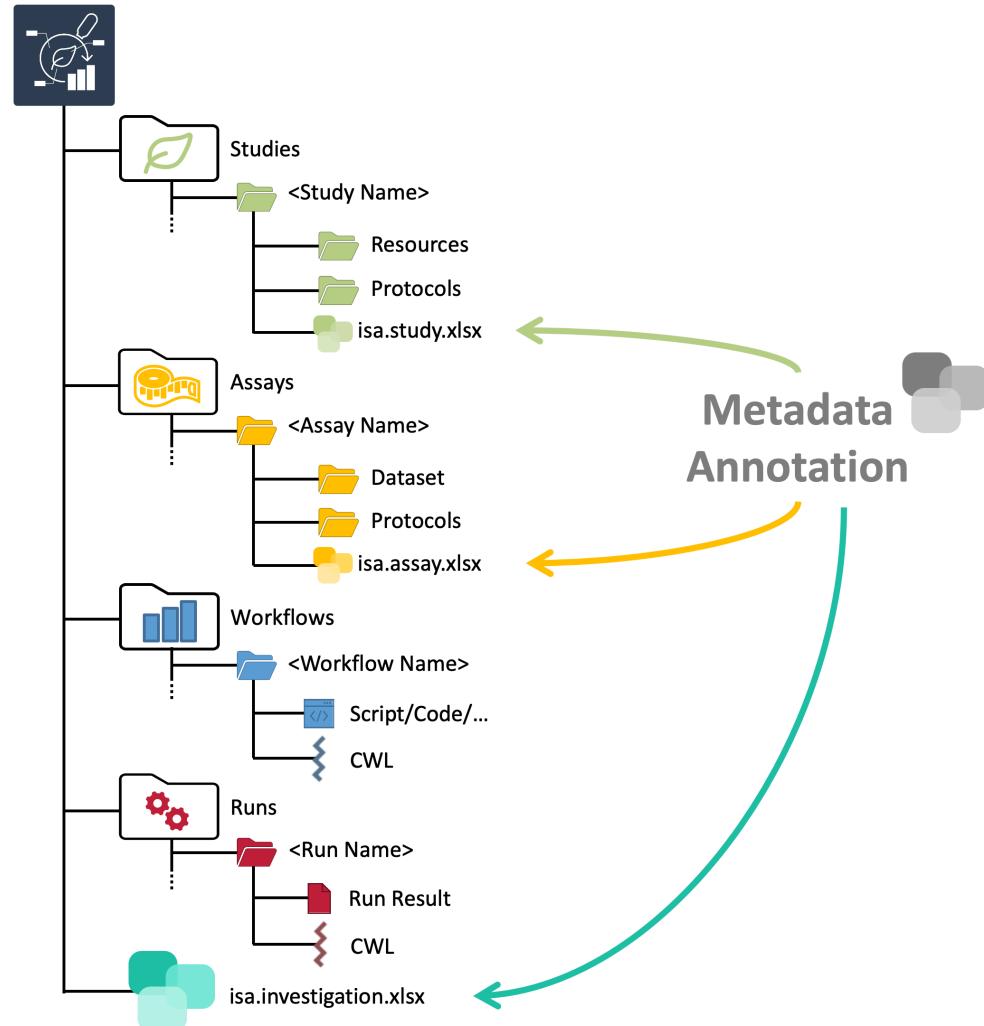
What does an ARC look like?



What does an ARC look like?



What does an ARC look like?





FINDABLE

ACCESSIBLE

INTEROPERABLE

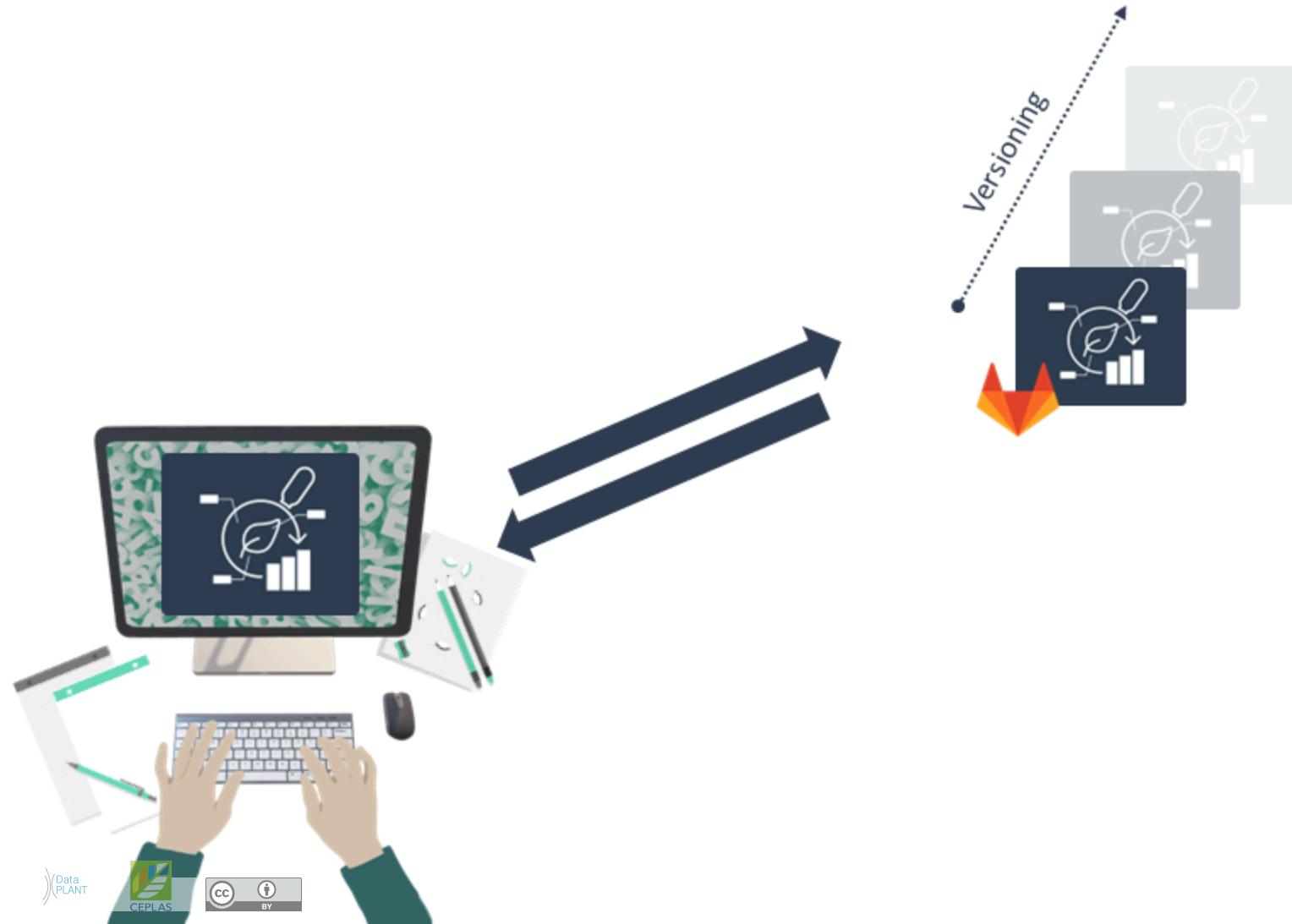
REUSABLE

Metadata
Templates

Controlled
Vocabularies

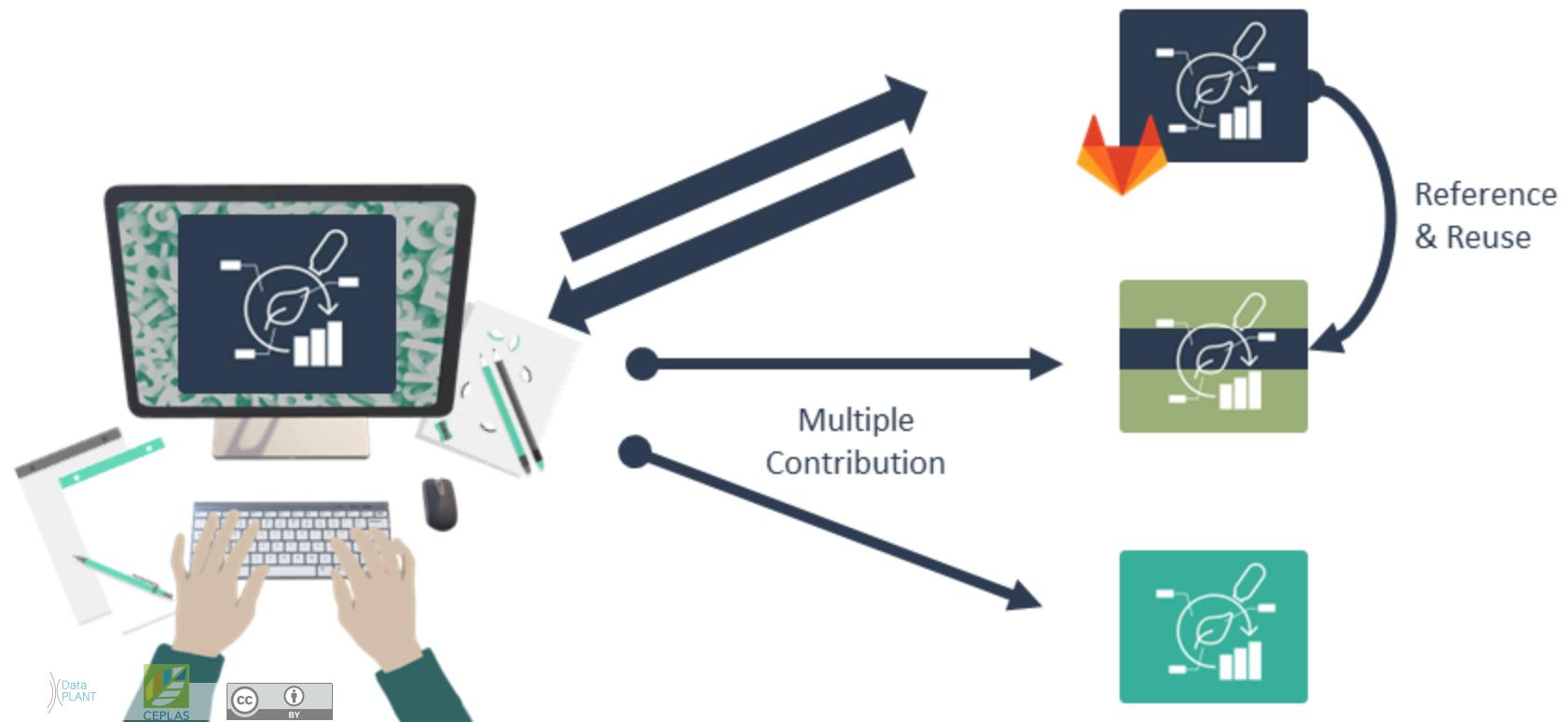


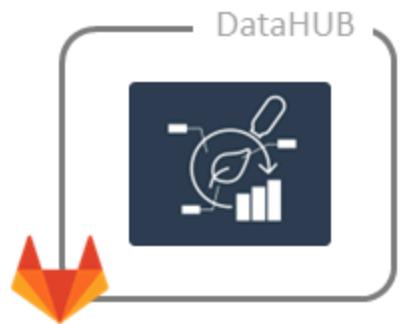
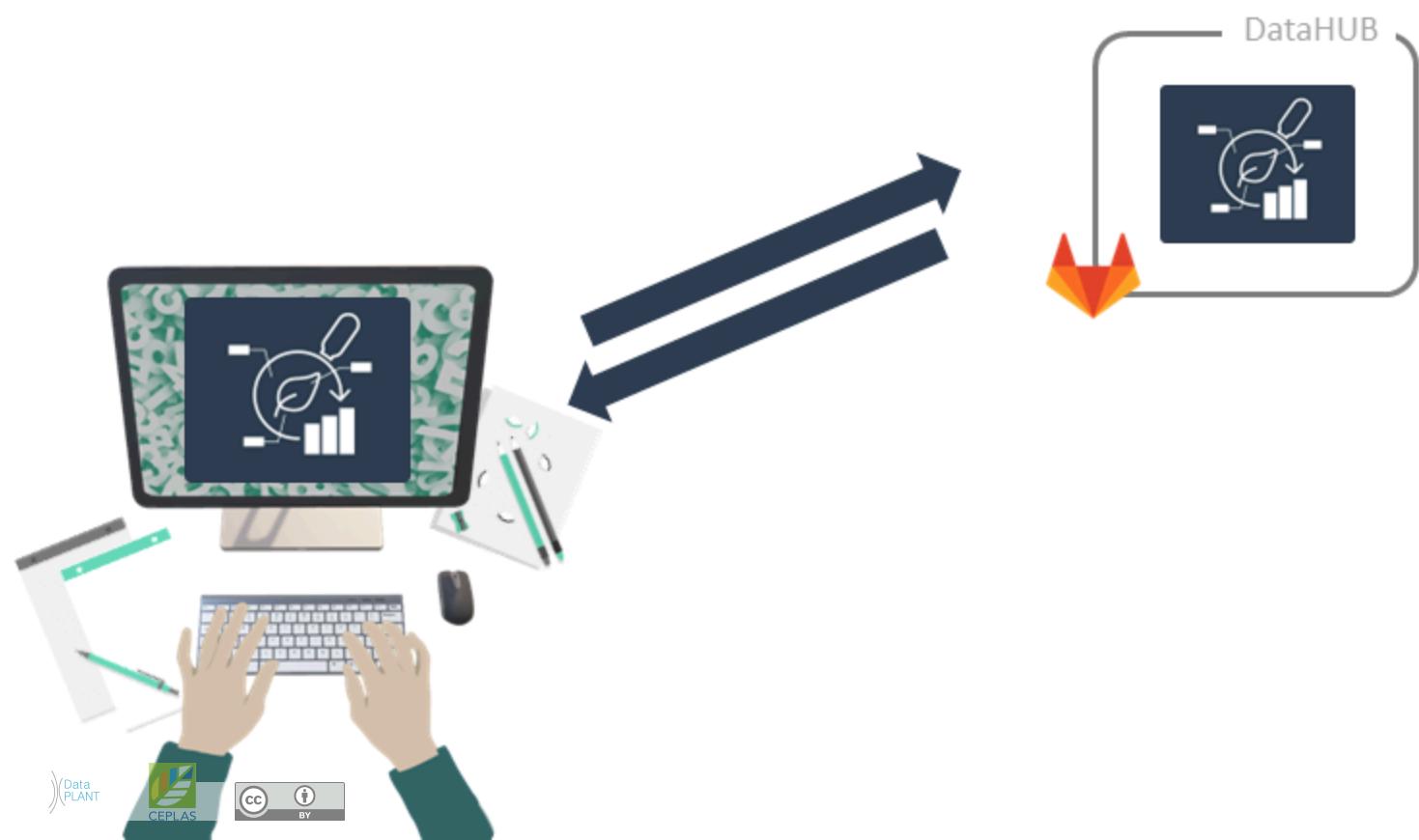






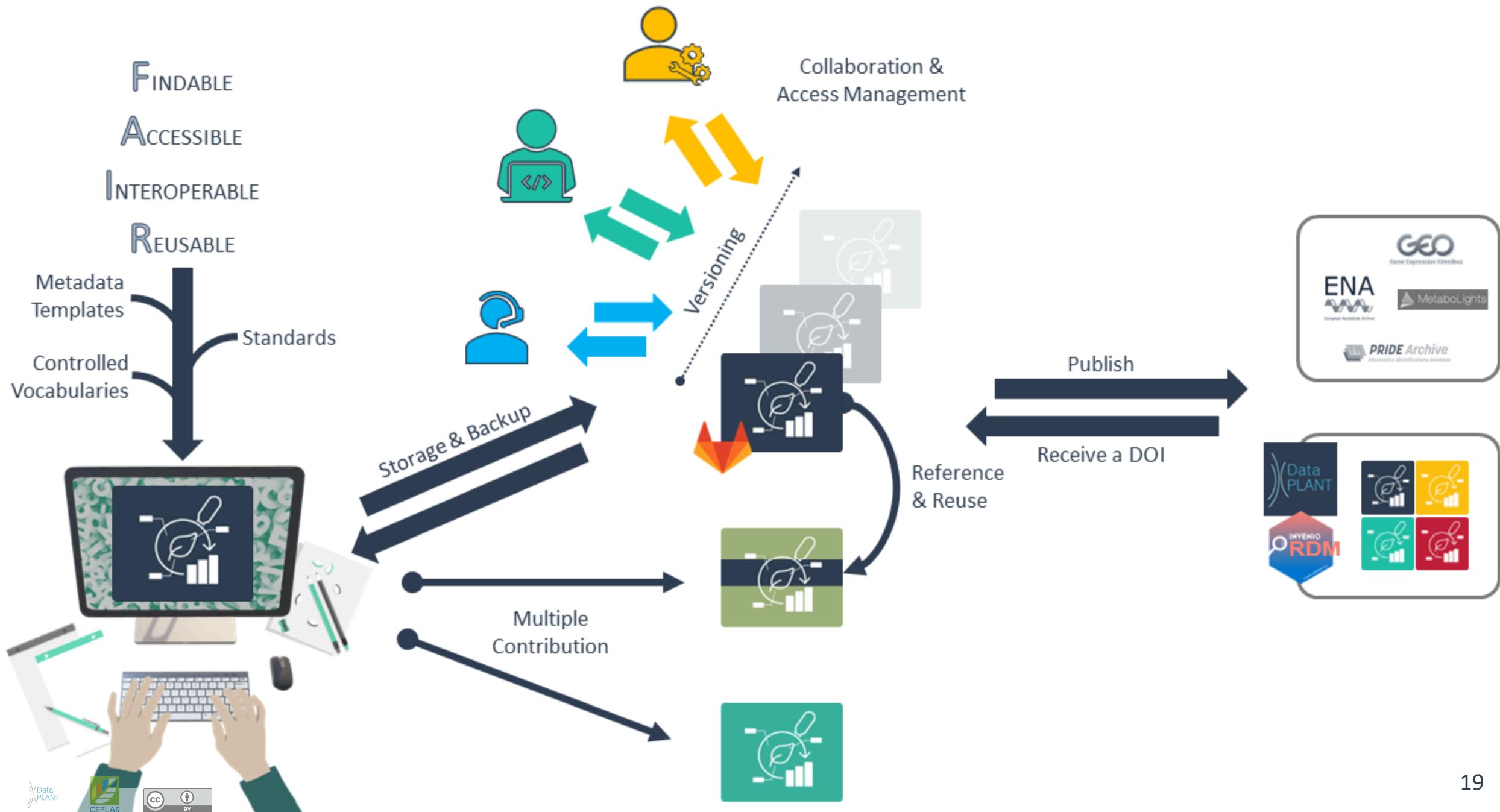






Publish
Receive a DOI





First steps towards your ARC

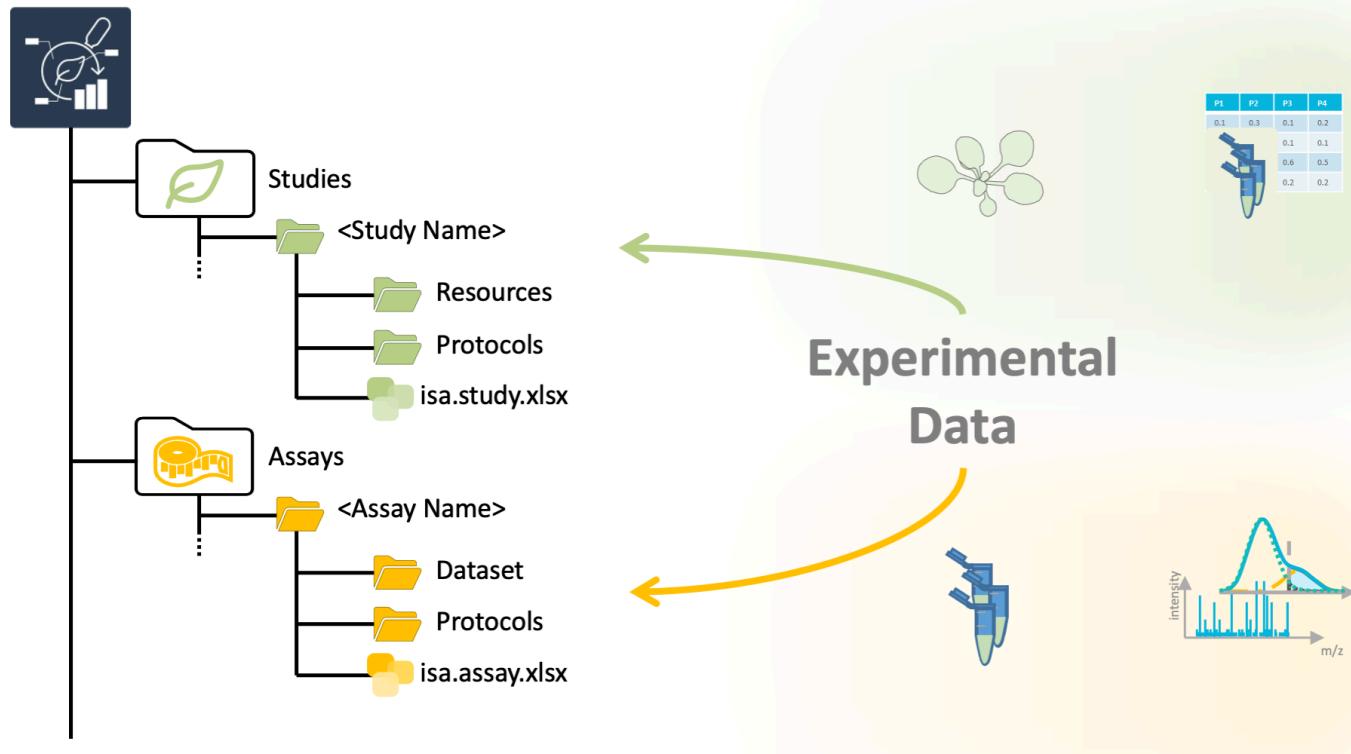
You just received your data

metabolomics_data
150112_56.D
150112_62.D
150112_66.D
150115_12.D
150115_14.D
150115_16.D
gcms_samplelist.tsv
method_gcms.txt
sample_submission_gcms.csv
methods
Illumina_libraries.txt
metabolite_extraction.txt
plant_material.txt
RNA_extraction.txt
rnaseq_data
DB_097_CAGATC_L001_R1_001.fastq.gz
DB_099_CTTGTA_L001_R1_001.fastq.gz
DB_103_AGTCAA_L001_R1_001.fastq.gz
DB_161_GTCCGC_L001_R1_001.fastq.gz
DB_163_GTGAAA_L001_R1_001.fastq.gz
DB_165_GTGAAA_L002_R1_001.fastq.gz
NGS_SampleSheet.xlsx

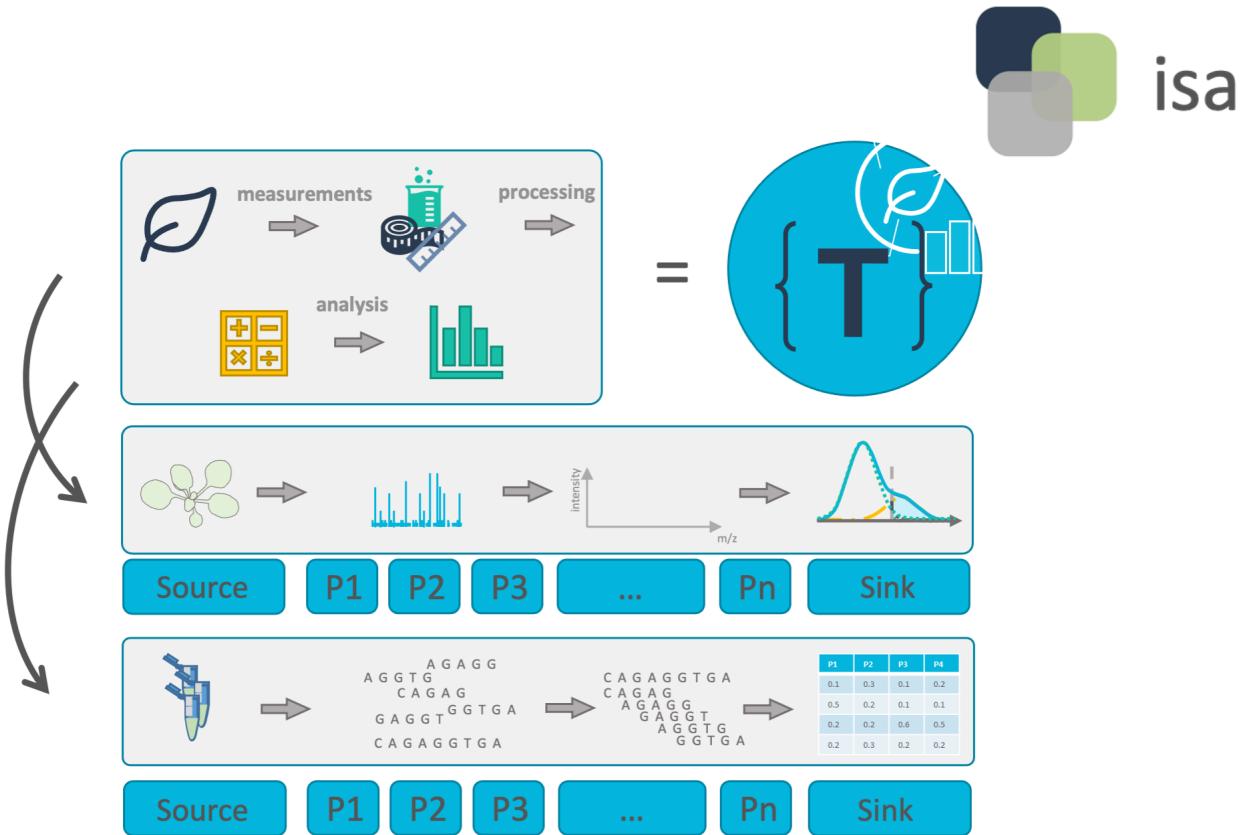
Goal

- Structure,
- Annotate, and
- Share **your experimental data.**

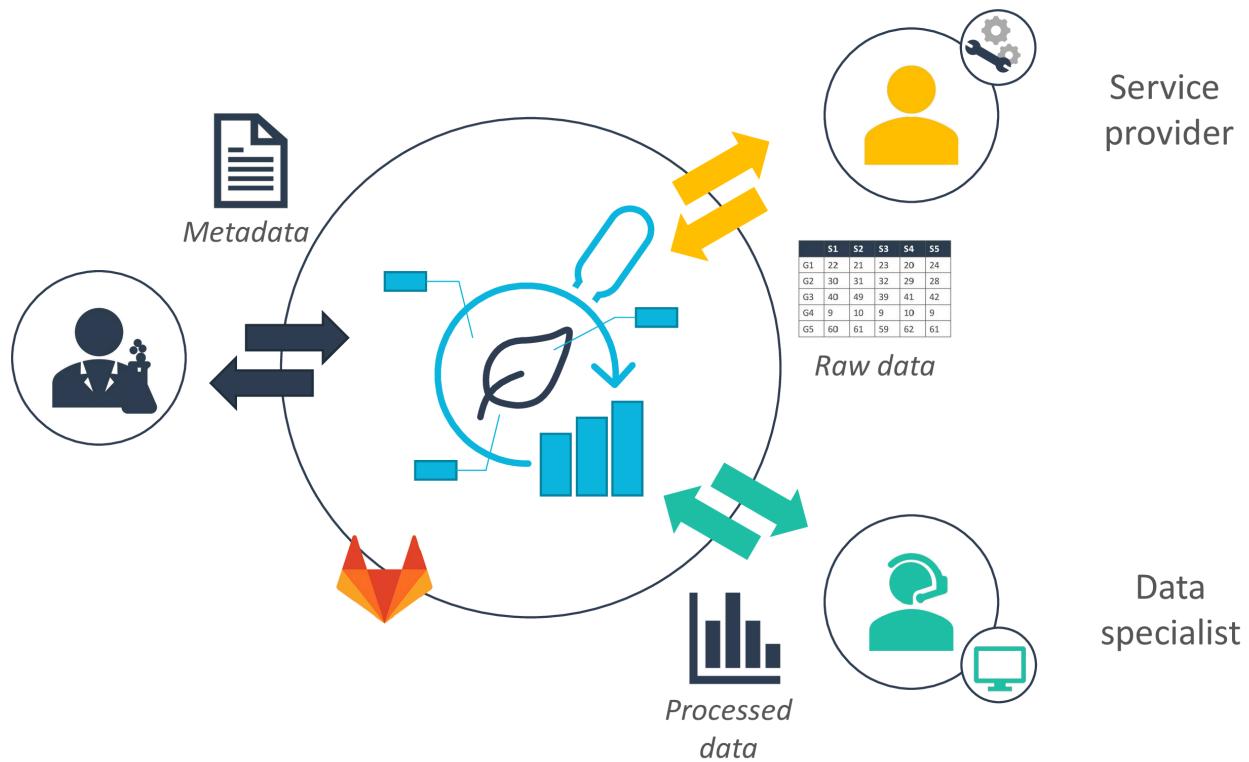
Structure your data



Annotate your data



Collaborate and share



Sync your ARC to the DataHUB

```
arc sync -f -r https://git.nfdi4plants.org/brilator/demo-arc
```

Invite collaborators to your ARC

<https://git.nfdi4plants.org/brilator/demo-arc>

Using the ARC Commander to setup an ARC

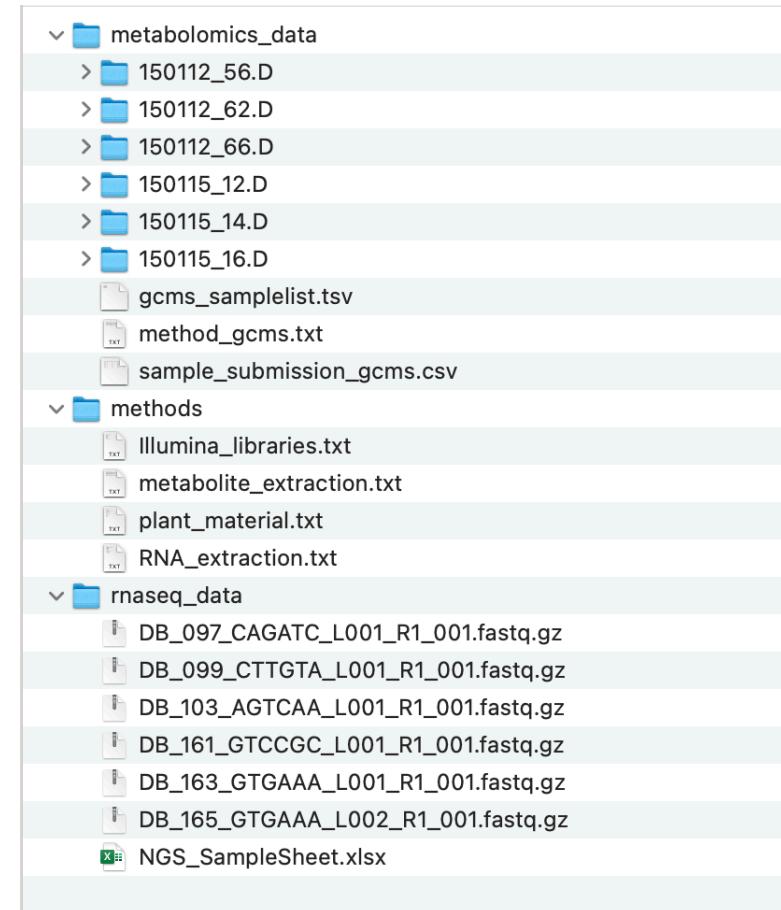
Initiate the ARC folder structure

```
arc init
```

Add metadata structure

```
arc investigation create -i TalinumPhotosynthesis  
arc study add -s talinum_drought  
arc assay add -s talinum_drought -a rnaseq  
arc assay add -s talinum_drought -a metabolomics
```

Your ARC is ready

 <p>The screenshot shows a file explorer interface with two main sections. The left section displays a directory structure for 'metabolomics_data' and 'rnaseq_data'. The 'metabolomics_data' section contains subfolders for runs like '150112_56.D', '150112_62.D', etc., and files like 'gcms_samplelist.tsv' and 'method_gcms.txt'. The 'rnaseq_data' section contains fastq.gz files for samples like 'DB_097_CAGATC_L001_R1_001.fastq.gz' and an 'NGS_SampleSheet.xlsx' file. The right section displays a directory structure for 'assays'. It includes subfolders for 'metabolomics' (containing 'dataset', 'isa.assay.xlsx', 'protocols', and 'README.md'), 'rnaseq' (containing 'dataset', 'isa.assay.xlsx', 'protocols', and 'README.md'), 'studies' (containing 'talinum_drought' with 'isa.study.xlsx', 'protocols', 'README.md', and 'resources'), and 'workflows'.</p> <table border="1"><tr><td>metabolomics_data</td><td>150112_56.D</td></tr><tr><td>metabolomics_data</td><td>150112_62.D</td></tr><tr><td>metabolomics_data</td><td>150112_66.D</td></tr><tr><td>metabolomics_data</td><td>150115_12.D</td></tr><tr><td>metabolomics_data</td><td>150115_14.D</td></tr><tr><td>metabolomics_data</td><td>150115_16.D</td></tr><tr><td>metabolomics_data</td><td>gcms_samplelist.tsv</td></tr><tr><td>metabolomics_data</td><td>method_gcms.txt</td></tr><tr><td>metabolomics_data</td><td>sample_submission_gcms.csv</td></tr><tr><td>methods</td><td>Illumina_libraries.txt</td></tr><tr><td>methods</td><td>metabolite_extraction.txt</td></tr><tr><td>methods</td><td>plant_material.txt</td></tr><tr><td>methods</td><td>RNA_extraction.txt</td></tr><tr><td>rnaseq_data</td><td>DB_097_CAGATC_L001_R1_001.fastq.gz</td></tr><tr><td>rnaseq_data</td><td>DB_099_CTTGTA_L001_R1_001.fastq.gz</td></tr><tr><td>rnaseq_data</td><td>DB_103_AGTCAA_L001_R1_001.fastq.gz</td></tr><tr><td>rnaseq_data</td><td>DB_161_GTCCGC_L001_R1_001.fastq.gz</td></tr><tr><td>rnaseq_data</td><td>DB_163_GTGAAA_L001_R1_001.fastq.gz</td></tr><tr><td>rnaseq_data</td><td>DB_165_GTGAAA_L002_R1_001.fastq.gz</td></tr><tr><td>rnaseq_data</td><td>NGS_SampleSheet.xlsx</td></tr><tr><td>assays</td><td>metabolomics</td></tr><tr><td>assays/metabolomics</td><td>dataset</td></tr><tr><td>assays/metabolomics</td><td>isa.assay.xlsx</td></tr><tr><td>assays/metabolomics</td><td>protocols</td></tr><tr><td>assays/metabolomics</td><td>README.md</td></tr><tr><td>assays</td><td>rnaseq</td></tr><tr><td>assays/rnaseq</td><td>dataset</td></tr><tr><td>assays/rnaseq</td><td>isa.assay.xlsx</td></tr><tr><td>assays/rnaseq</td><td>protocols</td></tr><tr><td>assays/rnaseq</td><td>README.md</td></tr><tr><td>assays</td><td>isa.investigation.xlsx</td></tr><tr><td>assays</td><td>runs</td></tr><tr><td>assays</td><td>studies</td></tr><tr><td>assays/studies</td><td>talinum_drought</td></tr><tr><td>assays/studies/talinum_drought</td><td>isa.study.xlsx</td></tr><tr><td>assays/studies/talinum_drought</td><td>protocols</td></tr><tr><td>assays/studies/talinum_drought</td><td>README.md</td></tr><tr><td>assays/studies/talinum_drought</td><td>resources</td></tr><tr><td>assays</td><td>workflows</td></tr></table>	metabolomics_data	150112_56.D	metabolomics_data	150112_62.D	metabolomics_data	150112_66.D	metabolomics_data	150115_12.D	metabolomics_data	150115_14.D	metabolomics_data	150115_16.D	metabolomics_data	gcms_samplelist.tsv	metabolomics_data	method_gcms.txt	metabolomics_data	sample_submission_gcms.csv	methods	Illumina_libraries.txt	methods	metabolite_extraction.txt	methods	plant_material.txt	methods	RNA_extraction.txt	rnaseq_data	DB_097_CAGATC_L001_R1_001.fastq.gz	rnaseq_data	DB_099_CTTGTA_L001_R1_001.fastq.gz	rnaseq_data	DB_103_AGTCAA_L001_R1_001.fastq.gz	rnaseq_data	DB_161_GTCCGC_L001_R1_001.fastq.gz	rnaseq_data	DB_163_GTGAAA_L001_R1_001.fastq.gz	rnaseq_data	DB_165_GTGAAA_L002_R1_001.fastq.gz	rnaseq_data	NGS_SampleSheet.xlsx	assays	metabolomics	assays/metabolomics	dataset	assays/metabolomics	isa.assay.xlsx	assays/metabolomics	protocols	assays/metabolomics	README.md	assays	rnaseq	assays/rnaseq	dataset	assays/rnaseq	isa.assay.xlsx	assays/rnaseq	protocols	assays/rnaseq	README.md	assays	isa.investigation.xlsx	assays	runs	assays	studies	assays/studies	talinum_drought	assays/studies/talinum_drought	isa.study.xlsx	assays/studies/talinum_drought	protocols	assays/studies/talinum_drought	README.md	assays/studies/talinum_drought	resources	assays	workflows	
metabolomics_data	150112_56.D																																																																														
metabolomics_data	150112_62.D																																																																														
metabolomics_data	150112_66.D																																																																														
metabolomics_data	150115_12.D																																																																														
metabolomics_data	150115_14.D																																																																														
metabolomics_data	150115_16.D																																																																														
metabolomics_data	gcms_samplelist.tsv																																																																														
metabolomics_data	method_gcms.txt																																																																														
metabolomics_data	sample_submission_gcms.csv																																																																														
methods	Illumina_libraries.txt																																																																														
methods	metabolite_extraction.txt																																																																														
methods	plant_material.txt																																																																														
methods	RNA_extraction.txt																																																																														
rnaseq_data	DB_097_CAGATC_L001_R1_001.fastq.gz																																																																														
rnaseq_data	DB_099_CTTGTA_L001_R1_001.fastq.gz																																																																														
rnaseq_data	DB_103_AGTCAA_L001_R1_001.fastq.gz																																																																														
rnaseq_data	DB_161_GTCCGC_L001_R1_001.fastq.gz																																																																														
rnaseq_data	DB_163_GTGAAA_L001_R1_001.fastq.gz																																																																														
rnaseq_data	DB_165_GTGAAA_L002_R1_001.fastq.gz																																																																														
rnaseq_data	NGS_SampleSheet.xlsx																																																																														
assays	metabolomics																																																																														
assays/metabolomics	dataset																																																																														
assays/metabolomics	isa.assay.xlsx																																																																														
assays/metabolomics	protocols																																																																														
assays/metabolomics	README.md																																																																														
assays	rnaseq																																																																														
assays/rnaseq	dataset																																																																														
assays/rnaseq	isa.assay.xlsx																																																																														
assays/rnaseq	protocols																																																																														
assays/rnaseq	README.md																																																																														
assays	isa.investigation.xlsx																																																																														
assays	runs																																																																														
assays	studies																																																																														
assays/studies	talinum_drought																																																																														
assays/studies/talinum_drought	isa.study.xlsx																																																																														
assays/studies/talinum_drought	protocols																																																																														
assays/studies/talinum_drought	README.md																																																																														
assays/studies/talinum_drought	resources																																																																														
assays	workflows																																																																														

Your ARC is ready

 Initiated an ARC

 Structured and ...

 ... annotated experimental data

 Shared with collaborators



Resources

CEPLAS Data Science

Website: <https://www.ceplas.eu/en/research/data-science-and-data-management/>

DataPLANT (nfdi4plants)

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

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

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

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

Acknowledgements

CEPLAS

Björn Usadel, Andrea Schrader, Hajira Jabeen (until 2021)

DataPLANT

Kaiserslautern: Timo Mühlhaus, Christoph Garth

- Kevin Frey, Lukas Weil, Oliver Maus, Kevin Schneider
(ARC Commander , Swate)
- Martin Kuhl (Knowledge Base)
- Jonas Lukasczyk (ARCitect)

Freiburg: Dirk von Suchodeletz

- Jonathan Bauer, Marcel Tschöpe (DataHUB , Swobup)
- Cristina Martins Rodrigues (Knowledge Base)

Five-Finger-Feedback



Contributors

Slides presented here include contributions by

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