This document is work in progress. If you experience any inconsistencies, have questions or would like to suggest additions, please feel free to contact us or open an issue in our Helpdesk.

We are very happy that you chose our tools and infrastructure to create and share your own ARCs. In this QuickStart we focus on explaining the ARC structure and its different components.

- Viola's ARC
- isa.investigation.xlsx
- Studies
  - isa.study.xlsx
- Assays
  - isa.assay.xlsx
- Workflows
- Runs
- Cheat sheet

Viola's ARC

Let's imagine a scenario where your project partner suggests at a conference to use this cool new Annotated Research Context (ARC) for your collaboration. Convinced by the versioning system and the single point of entry logic, you are motivated to set up your first own ARC after returning to the lab and fill it with your latest project results. Back home, however, you only remember the basic ARC structure and something about some isa.xlsx files. So how do you transfer your project into the empty ARC your project partner shared with you?

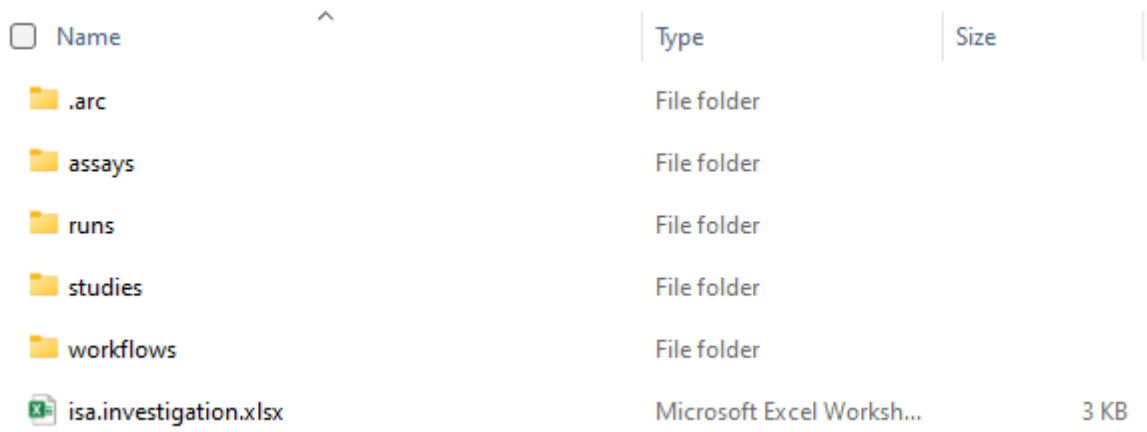


Figure 1: ARCStructure

To answer this question, we will first take a look back at Viola's metadata example:

> Viola investigates the effect of the plant circadian clock on sugar metabolism in *W. mirabilis*. For her PhD project, which is part of an EU-funded consortium in Prof. Beetroot's lab, she acquires seeds from a South-African Botanical Society. Viola grows the plants under different light regimes, harvests leaves from a two-day time series experiment, extracts polar metabolites as well as RNA and submits the samples to nearby core facilities for metabolomics and transcriptomics measurements, respectively. After a few weeks of iterative consultation with the facilities' heads as well as technicians and computational biologists involved, Viola receives back a wealth of raw and processed data. From the data she produces figures and wraps everything up to publish the results in the *Journal of Wonderful Plant Sciences*.

The entire information given in this example can be stored within an ARC. To illustrate the ARC specifications, we will highlight and explain every (sub)directory and ISA-file of the ARC with references to Viola's example.

isa.investigation.xlsx

The ISA investigation workbook allows you to record administrative metadata of your project. In Viola's example, the title of the project, the contact persons, and related publications correspond to such metadata. Besides that, the workbook can also contain a short description of your project, but also lists included studies with respective design types, assays, protocols, etc.. Although we recommend to use the ARC Commander for adding these metadata, you can of course fill the workbook (and also the isa.study.xlsx and isa.assay.xlsx) manually.

Studies

In the `studies` (sub)folders you can collect material and resources used within your studies. Corresponding information in Viola's project include the source of her seeds (South-African Botanical Society), how she grew the plants, and the design of the experiment (two-day time series, etc.).

In case your investigation contains more than one study, each of these studies is placed in an individual subdirectory. The "resources" directory allows you to store material samples or external data as virtual sample files. You can use the protocol subdirectory to store free-text protocols that describe how the samples or materials were created.

	A	B
1	ONTOLOGY SOURCE REFERENCE	
2	Term Source Name	
3	Term Source File	
4	Term Source Version	
5	Term Source Description	
6	INVESTIGATION	
7	Investigation Identifier	20220803
8	Investigation Title	The effect of the plant circadian clock on sugar metabolism in *W. mirabilis*.
9	Investigation Description	
10	Investigation Submission Date	2022-04-03
11	Investigation Public Release Date	2022-08-01
12	INVESTIGATION PUBLICATIONS	
13	Investigation Publication PubMed ID	32778xxx
14	Investigation Publication DOI	doi:10.1038/wxyz1170
15	Investigation Publication Author List	Viola Sugarcane, Herb Sequoia, Rosemary Foxglove, Oliver Sage, and Jasmine Beetroot
16	Investigation Publication Title	The effect of the plant circadian clock on sugar metabolism in *W. mirabilis*.
17	Investigation Publication Status	published
18	Investigation Publication Status Term Accession Number	
19	Investigation Publication Status Term Source REF	
20	INVESTIGATION CONTACTS	
21	Investigation Person Last Name	Beetroot
22	Investigation Person First Name	Edward
23	Investigation Person Mid Initials	
24	Investigation Person Email	edward.beetroot@nfdi4plants.org
25	Investigation Person Phone	+49 1234 5678

Figure 2: Investigation

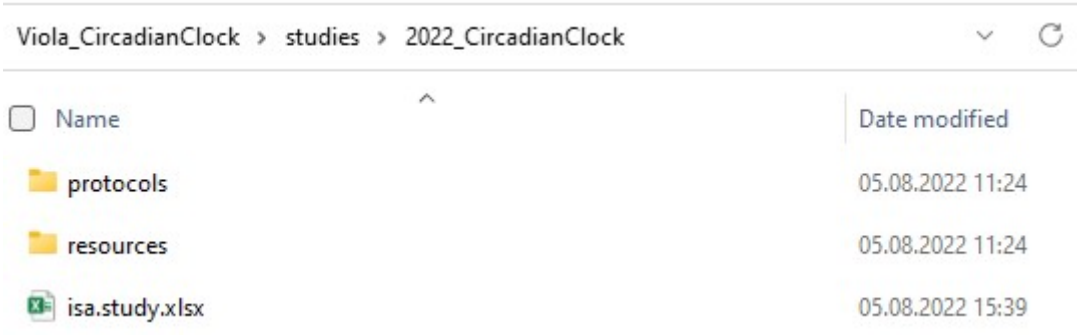


Figure 3: StuSubdirectories

isa.study.xlsx

Every study contains one `isa.study.xlsx` file to specify the characteristics of all material and resources. Resources described in a study file can be the input for one or multiple assays or workflows. The workbook contains (at least) two worksheets:

- "2022_CircadianClock": A worksheet with the name of your study to annotate the properties of your source material following the ISA model. While this can be done manually, we recommend using our ontology supported annotation tool Swate.
- "Study": Viola collected the administrative metadata of her study in this worksheet. This information can later be transferred into the `isa.investigation.xlsx` using the ARC Commander.

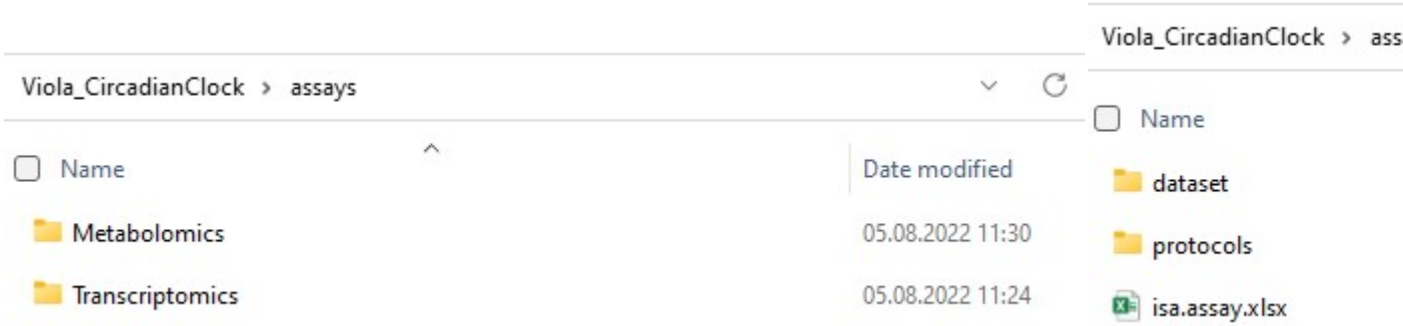
	A	B	E	H	K
1	Source Name	Characteristics [Organism]	Characteristics [Organism part]	Characteristics [Growth day length]	Characteristics [Sample type]
2	Wmir001	Welwitschia mirabilis	leaf	12 h	
3	Wmir002	Welwitschia mirabilis	leaf	12 h	
4	Wmir003	Welwitschia mirabilis	leaf	12 h	
5	Wmir004	Welwitschia mirabilis	leaf	12 h	
6	Wmir005	Welwitschia mirabilis	leaf	12 h	
7	Wmir006	Welwitschia mirabilis	leaf	12 h	
8	Wmir007	Welwitschia mirabilis	leaf	12 h	
9	Wmir008	Welwitschia mirabilis	leaf	12 h	
10	Wmir009	Welwitschia mirabilis	leaf	12 h	
11	Wmir010	Welwitschia mirabilis	leaf	14 h	
12	Wmir011	Welwitschia mirabilis	leaf	14 h	
13	Wmir012	Welwitschia mirabilis	leaf	14 h	
14	Wmir013	Welwitschia mirabilis	leaf	14 h	
15	Wmir014	Welwitschia mirabilis	leaf	14 h	
16	Wmir015	Welwitschia mirabilis	leaf	14 h	
17	Wmir016	Welwitschia mirabilis	leaf	14 h	
18	Wmir017	Welwitschia mirabilis	leaf	14 h	
19	Wmir018	Welwitschia mirabilis	leaf	14 h	
20	Wmir019	Welwitschia mirabilis	leaf	14 h	
21	Wmir020	Welwitschia mirabilis	leaf	14 h	
22	Wmir021	Welwitschia mirabilis	leaf	14 h	

	A	B	C
1	STUDY METADATA		
2	Study Identifier	2022_CircadianClock	
3	Study Title	The effect of the plant circadian clock on sugar metabolism in W. mirabilis.	
4	Study Description		
5	Study Submission Date	2022-04-03	
6	Study Public Release Date	2022-08-01	
7	Study File Name	2022_CircadianClock\isa.study.xlsx	
8	STUDY DESIGN DESCRIPTORS		
9	Study Design Type	time-series experiment	
10	Study Design Type Term Accession Number		
11	Study Design Type Term Source REF		
12	STUDY PUBLICATIONS		
13	Study Publication PubMed ID	32778xxx	
14	Study Publication DOI	doi:10.1038/wxyz1170	
15	Study Publication Author List	Viola Sugarcane, Herb Sequoia, Rosemary Foxglove, Oliver Sage, and Jasmine Beetroot	
16	Study Publication Title	The effect of the plant circadian clock on sugar metabolism in W. mirabilis.	
17	Study Publication Status	published	
18	Study Publication Status Term Accession Number		
19	Study Publication Status Term Source REF		
20	STUDY FACTORS		
21	Study Factor Name	Light exposure time	
22	Study Factor Type		

2022_CircadianClock | Study

Assays

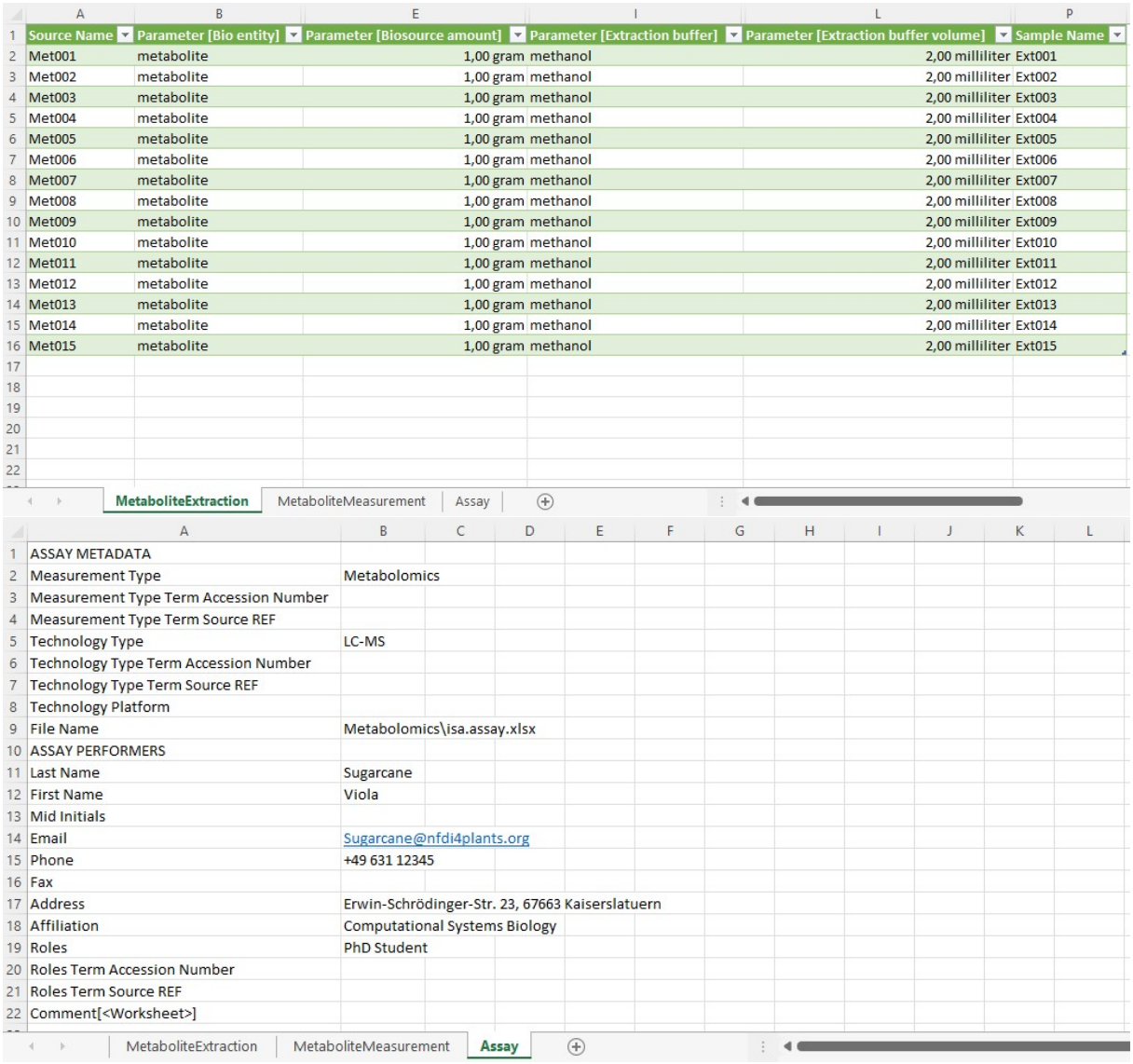
The `assays` folder allows you to store data and metadata from experimental processes or analytical measurements. Each assay is a collection of files stored in a single directory, including corresponding metadata files in form of an `isa.assay.xlsx`. Viola needs two subdirectories, one for her metabolomics and one for her transcriptomics dataset, respectively. Assay data files and free-text protocols are placed in individual subdirectories. Data files produced by an assay can be the input for one or multiple workflows.



isa.assay.xlsx

Viola can annotate her experimental workflows of the metabolomics and transcriptomics assays with process parameters in the `isa.assay.xlsx` file, which needs to be present for every assay. The workbook contains two or more worksheets, depending on the number of used protocols:

- “MetaboliteExtraction”: A worksheet with the name of the used protocol to annotate the experimental workflow, in this case for extraction of metabolites. While this can be done manually, we recommend using our ontology supported annotation tool Swate.
- “MetaboliteMeasurement”: A worksheet that describes the quantification of polar metabolites using gas-chromatography mass-spectrometry.
- “Assay”: Viola collected the administrative metadata of her assay in this worksheet. This information can later be transferred into the `isa.investigation.xlsx` using the ARC Commander.



Source Name	Parameter [Bio entity]	Parameter [Biosource amount]	Parameter [Extraction buffer]	Parameter [Extraction buffer volume]	Sample Name
Met001	metabolite	1,00 gram	methanol	2,00 milliliter	Ext001
Met002	metabolite	1,00 gram	methanol	2,00 milliliter	Ext002
Met003	metabolite	1,00 gram	methanol	2,00 milliliter	Ext003
Met004	metabolite	1,00 gram	methanol	2,00 milliliter	Ext004
Met005	metabolite	1,00 gram	methanol	2,00 milliliter	Ext005
Met006	metabolite	1,00 gram	methanol	2,00 milliliter	Ext006
Met007	metabolite	1,00 gram	methanol	2,00 milliliter	Ext007
Met008	metabolite	1,00 gram	methanol	2,00 milliliter	Ext008
Met009	metabolite	1,00 gram	methanol	2,00 milliliter	Ext009
Met010	metabolite	1,00 gram	methanol	2,00 milliliter	Ext010
Met011	metabolite	1,00 gram	methanol	2,00 milliliter	Ext011
Met012	metabolite	1,00 gram	methanol	2,00 milliliter	Ext012
Met013	metabolite	1,00 gram	methanol	2,00 milliliter	Ext013
Met014	metabolite	1,00 gram	methanol	2,00 milliliter	Ext014
Met015	metabolite	1,00 gram	methanol	2,00 milliliter	Ext015

MetaboliteExtraction | MetaboliteMeasurement | Assay

ASSAY METADATA	
Measurement Type	Metabolomics
Measurement Type Term Accession Number	
Measurement Type Term Source REF	
Technology Type	LC-MS
Technology Type Term Accession Number	
Technology Type Term Source REF	
Technology Platform	
File Name	Metabolomics\isa.assay.xlsx
ASSAY PERFORMERS	
Last Name	Sugarcane
First Name	Viola
Mid Initials	
Email	Sugarcane@nfdi4plants.org
Phone	+49 631 12345
Fax	
Address	Erwin-Schrödinger-Str. 23, 67663 Kaiserslatuern
Affiliation	Computational Systems Biology
Roles	PhD Student
Roles Term Accession Number	
Roles Term Source REF	
Comment[<Worksheet>]	

MetaboliteExtraction | MetaboliteMeasurement | Assay

Workflows

In an ARC `workflows` represent the processing steps used in computational analyses and other transformations of data originating from studies and assays. Typical examples include data cleaning and preprocessing, computational analysis, or visualization. The outcomes of these workflows ("run results") are stored in runs.

Viola received for her transcriptome and metabolome assays various processed data files, which she now can use to generate some nice plots. Additionally, the computational biologists sent her the code used for data processing, including an executable Common Workflow Language (CWL) file, which contains a standardized tool or workflow description. She stores these files in individual subdirectories for each workflow.

Runs

After Viola generated her plots, she placed them in individual subdirectories, specific to the run they were generated with. In general, you can use the runs folder to store plots, tables, or similar result files that derive from computations on study, assay, external data or other runs.

Cheat sheet

We hope that these examples nicely illustrated the ARC structure and that you are now ready to produce your own ARCs. Use the figure below as a cheat sheet to remember where to store which files. Or follow the ARC Commander QuickStart to try it out yourself.

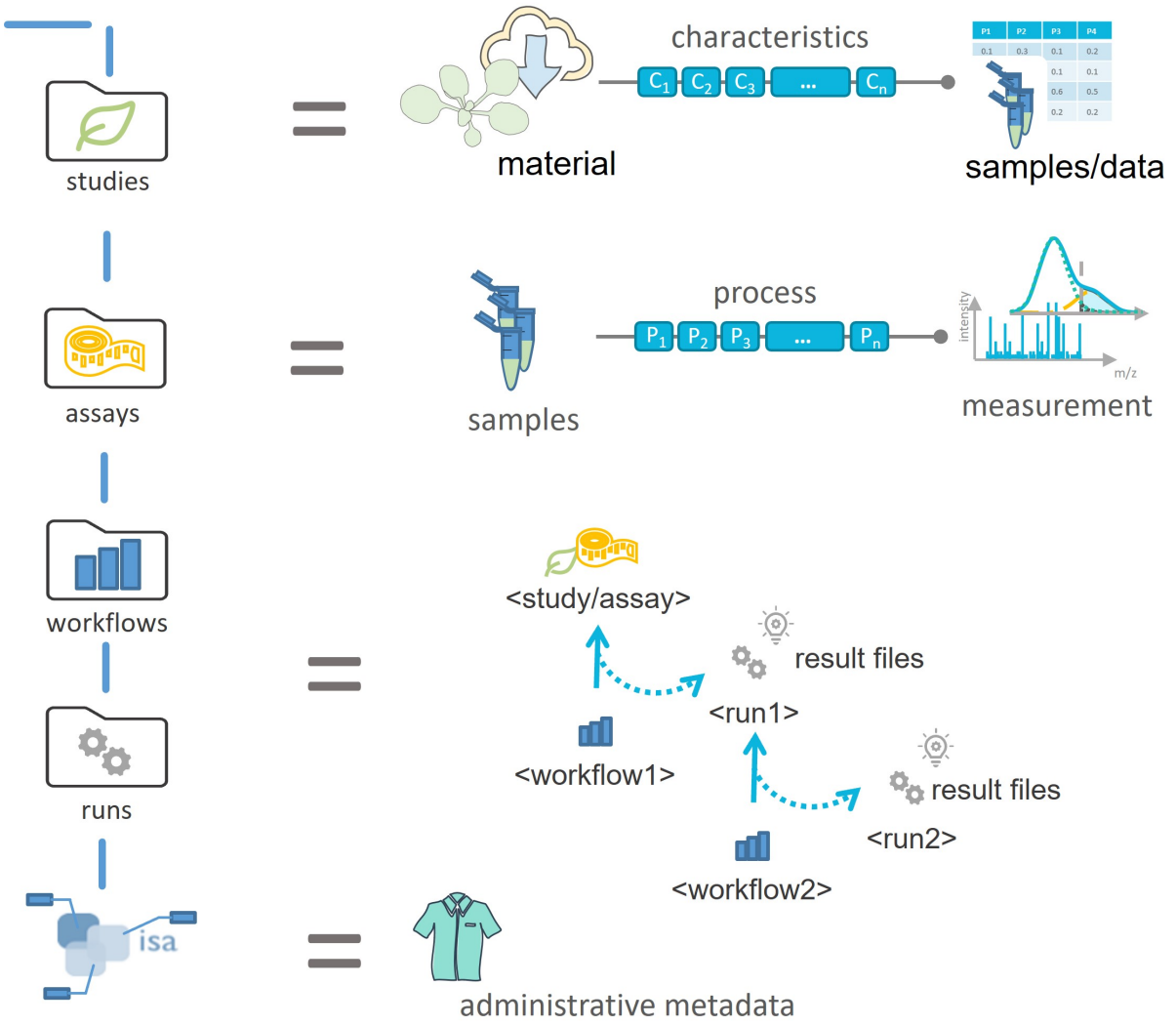


Figure 4: IllustratedARCStructure