



DataPLANT's QuickStart on ARCs

V1.1

April 2022

We are very happy that you chose our tools and infrastructure to create and share your own ARCs. In this QuickStart we focus on how to use the "ARC Commander" to store your data and "SWATE" to enrich it with metadata.

This document is work in progress. If you experience any inconsistencies, have questions or would like to suggest additions, please feel free to send a message to: [info\[at\]nfdi4plants.org](mailto:info[at]nfdi4plants.org) or open an issue at <https://github.com/nfdi4plants/quickstart>.

- [Setup and environment](#)
 - [Required softwares](#)
 - [The command line](#)
- [ARC initialization](#)
- [Adding metadata](#)
 - [ISA investigation](#)
 - [ISA studies and assays](#)
- [Sharing your ARC](#)
 - [DataPLANT registration and access](#)
 - [ARC synchronization](#)
 - [Invite collaborators](#)
- [Data annotation](#)
 - [SWATE](#)
- [DataPLANT Support](#)
- [The Minimalist's ARC-QuickStart](#)

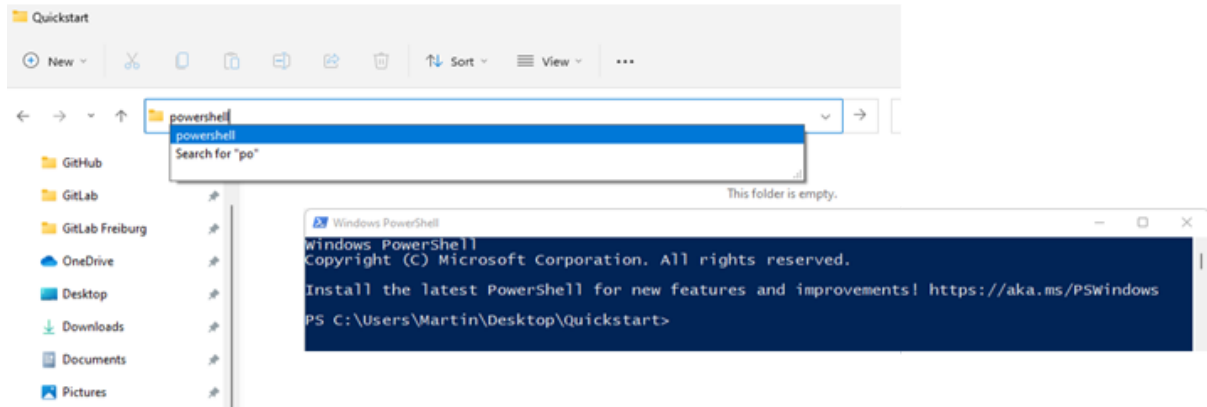
Setup and environment

Required softwares

- ☐ Please download the latest version of the [ARC Commander](#) for your operating system and install it according to [these instructions](#).
- ☐ Prerequisites for using the ARC Commander are [git](#) and [git LFS](#)

The command line

- Most of this quickstart (especially the section [ARC initialization](#)) is based on the command line (Windows: cmd, powershell; Linux and Mac: terminal). The following picture shows exemplarily how to open a powershell on windows by entering *powershell* into the explorer path:



- Text formatted as code blocks represents commands to copy/paste into the command line:

```
echo "hello - I am a code block"
```

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

```
arc --version
arc --help
```

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Martin\Desktop\Quickstart> arc
no valid subcommand has been specified.

SUBCOMMANDS:

  init <init args>      Initialize basic folder structure
  export <export args>  Exports the full arc to a json object
  sync <sync args>      Synchronize the ARC with its upstream repository. Commits changes made in the ARC. If a
                        remote is set or is given, also pulls from there and pushes all previously made commits.
  get <get args>        Download an ARC from a remote repository (e.g. from gitlab)
  update               Update items in the arc against each other
  --version            Get the ArcCommander's current version
  investigation, i <verb and args> Investigation file functions
  study, s <verb and args> Study functions
  assay, a <verb and args> Assay functions
  configuration, config <verb and args> Configuration editing
  remoteaccess, remote <verb and args> Subcommands for handling access functionality to remote repositories

  Use 'arc.exe <subcommand> --help' for additional information.

OPTIONS:
  --workingdir, -p <working directory> Set the base directory of your ARC
  --verbosity, -v <verbosity>         Set the amount of additional printed information: 0 -> No information, 1 (Default) -> Basic
                                      Information, 2 -> Additional information
  --help                               display this list of options.

PS C:\Users\Martin\Desktop\Quickstart>
```

ARC initialization

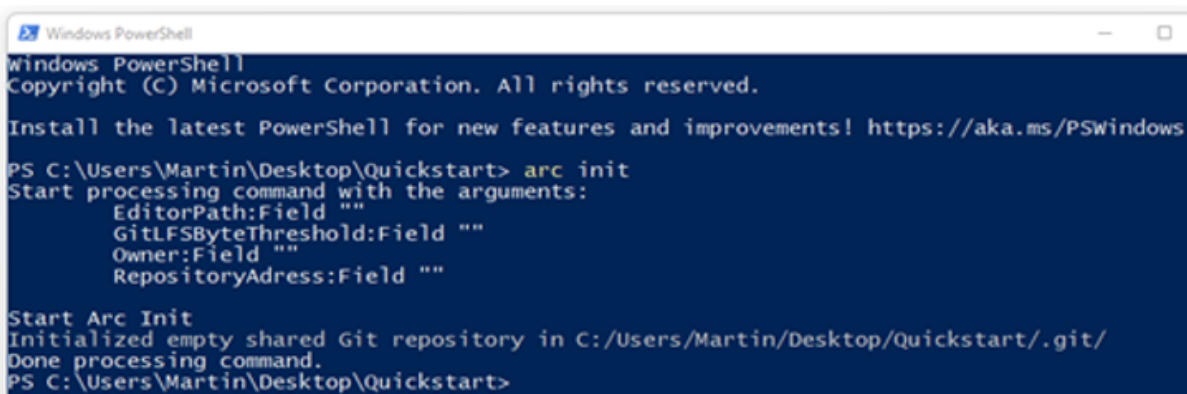
1. Create and navigate to a local folder, which you want to initialize as an ARC.

```
mkdir ~/QuickStartARC;  
cd ~/QuickStartARC
```






2. Initialize your ARC by executing

```
arc init
```

3. This will create the general ARC folder structure:



```
Windows PowerShell  
Copyright (C) Microsoft Corporation. All rights reserved.  
  
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows  
  
PS C:\Users\Martin\Desktop\Quickstart> arc init  
Start processing command with the arguments:  
  EditorPath:Field ""  
  GitLFSByteThreshold:Field ""  
  Owner:Field ""  
  RepositoryAddress:Field ""  
  
Start Arc Init  
Initialized empty shared Git repository in C:/Users/Martin/Desktop/Quickstart/.git/  
Done processing command.  
PS C:\Users\Martin\Desktop\Quickstart>
```

<input type="checkbox"/> Name	Date modified	Type
 .arc	22.03.2022 08:27	File folder
 assays	22.03.2022 08:29	File folder
 runs	22.03.2022 08:27	File folder
 studies	22.03.2022 08:29	File folder
 workflows	22.03.2022 08:36	File folder

Adding metadata

ISA investigation

The ISA investigation workbook allows you to record administrative metadata of your project. Add the isa.investigation.xlsx workbook including an identifier to your ARC with

```
arc i create -i "QuickStartInvestigation"
```

ISA studies and assays

The ISA study and ISA assay workbooks allow you to annotate your experimental data.

1. Add an isa.study.xlsx workbook including an identifier to your ARC with

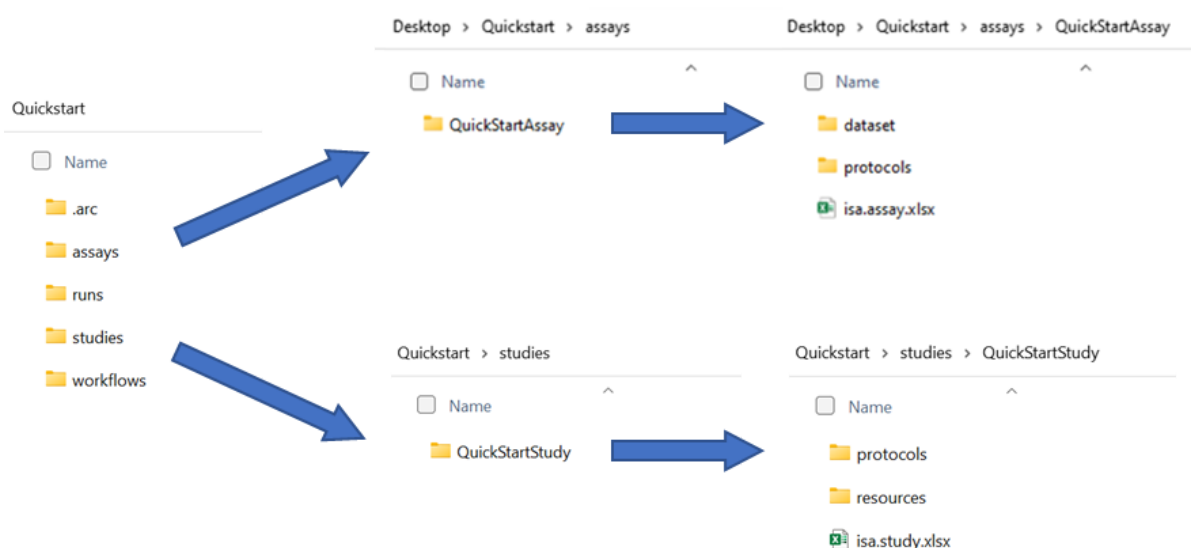
```
arc s add -s "QuickStartStudy"
```

2. Add an isa.assay.xlsx workbook including an identifier to your ARC with

```
arc a add -s "QuickStartStudy" -a "QuickStartAssay"
```

Note: An assay must be linked to a study. If a study does not exist, it will be created automatically in this step.

- The ARC Commander will add a subdirectories to the *studies* and *assays* folder. Your ARC should look similar to this now:



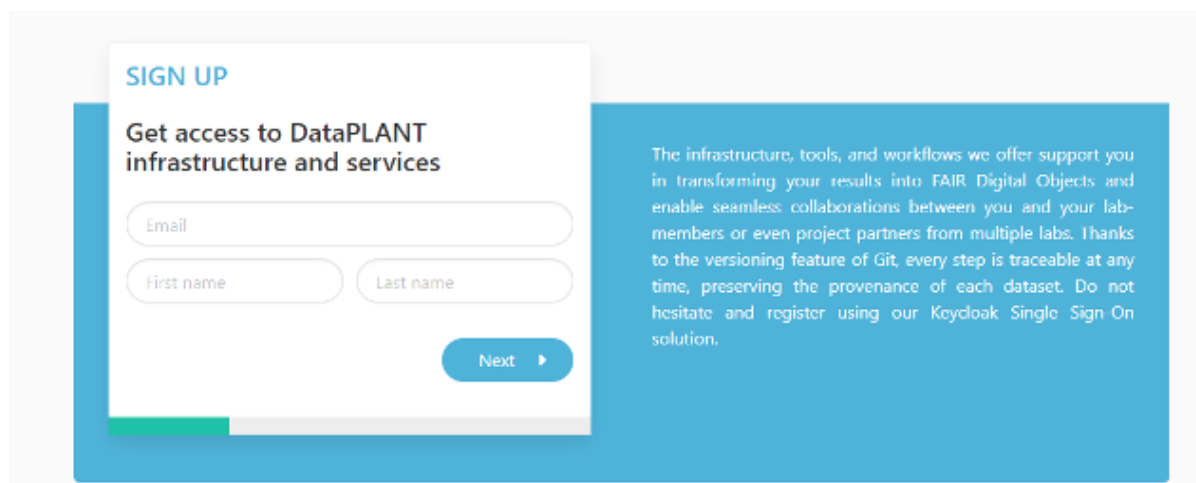
- These steps can be repeated to add as many studies and assays as needed. Accordingly, more subdirectories will be added. Multiple assays can be grouped in a study when the same StudyIdentifier in the text editor window is used.

3. Place the data for each assay in the respective dataset folder.

Sharing your ARC

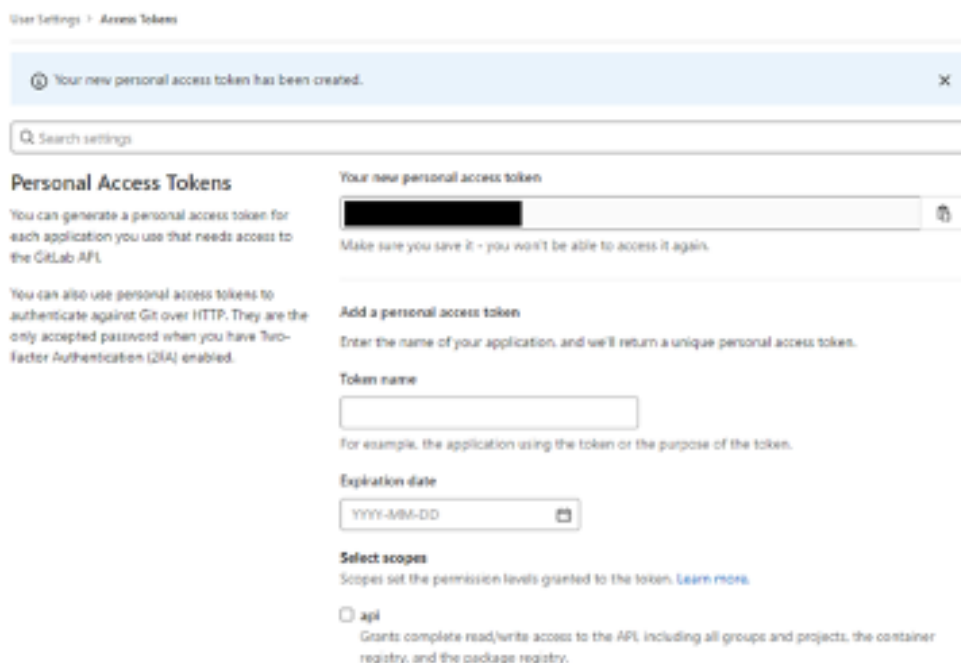
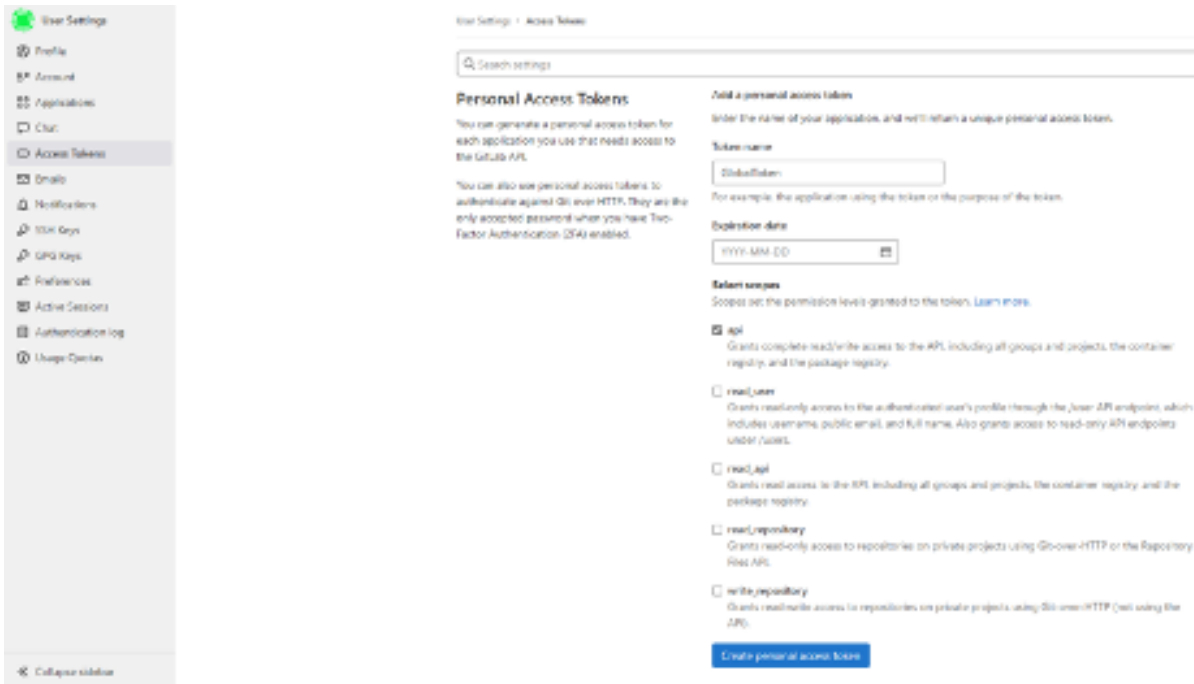
DataPLANT registration and access

In case you are not a member of DataPLANT yet, please visit <https://register.nfdi4plants.org> to register. Afterwards, you will be granted access to DataPLANT's DataHUB, available under <https://git.nfdi4plants.org>. The DataHUB allows you to share your ARCs with registered lab or project partners.

A screenshot of the DataPLANT registration page. The page has a light blue header with the text "SIGN UP" in white. Below the header, the main heading is "Get access to DataPLANT infrastructure and services". There are three input fields: "Email", "First name", and "Last name". A blue "Next" button with a right arrow is positioned below the "Last name" field. To the right of the form, there is a blue box containing white text that reads: "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."

After successful registration, please visit the [DataHUB](#) to set an access token for ARC Commander synchronization:

1. Sign-in in the top right corner. Click on your profile picture in the top right corner and go to *Preferences -> Access Tokens*.
2. Create an api access token with a name of your choice. These tokens grant read and write access to all of your groups and projects. Make sure you save your access token upon successful creation, as this is the only time you will have access to the token (in case you lose the token, you can simply create a new one).



ARC synchronization

Create and connect a remote repository

1. In the [DataHUB](#), create a new blank repository by clicking "New project/repository" in the plus drop down menu of the navigation bar on top.
2. Connect your local ARC with the remote specifying the remote address with the flag `-r` in combination with the URL of your remote repository

```
# Note: This command needs to be adapted with the respective URL of your
DataHUB ARC
# arc sync -r git@git.nfdi4plants.org:<YourUserName>/<YourRepository.git>
```

3. Synchronize your ARC with the DataHUB using the command `arc sync`. The ARC Commander will ask for your credentials, where you need to enter your DataHUB handle (displayed on the DataHUB when clicking on your profile picture) and the newly generated access token.

```

PS C:\Users\Martin\Desktop\Quickstart> arc sync -r https://git.nfdi4plants.org/martinkuhl/QuickStart
Start processing command with the arguments:
  Branch:Field ""
  CommitMessage:Field ""
  RepositoryAddress:Field "https://git.nfdi4plants.org/martinkuhl/QuickStart"

Start Arc Sync
GIT: Martin Kuhl
GIT: martinkuhl137@gmail.com
-----
GIT: Tracking "**/dataset/**"
-----
[master (root-commit) f0720e7] Update
5 files changed, 2 insertions(+)
create mode 100644 .arc/ArcCommander.log
create mode 100644 QuickStartStudy_isa.study.xlsx
create mode 100644 assays/QuickStartAssay/README.md
create mode 100644 assays/QuickStartAssay/isa.assay.xlsx
create mode 100644 isa.investigation.xlsx
[master (root-commit) f0720e7] Update
  
```

Windows Security

Git Credential Manager

Enter credentials for 'https://git.nfdi4plants.org/'

Benutzername

Kennwort

OK Cancel

4. If no repository exists under the given URL, the ARC Commander will produce an error ensuring that you spelled the URL correctly. Use `arc sync -f` to force synchronization to the specified URL.

```

[master (root-commit) f0720e7] Update
ERROR: Remote repo was set, but does not exist.
Check whether it was spelled correctly. If not, you can run "arc sync" again using the --repositoryAddress argument.
If you want to create a new remote repository instead, you can run "arc sync -f" to force push the local repository to a new upstream.
Done processing command.
PS C:\Users\Martin\Desktop\Quickstart> arc sync -f
Start processing command with the arguments:
  Branch:Field ""
  CommitMessage:Field ""
  Force:Flag
  RepositoryAddress:Field ""

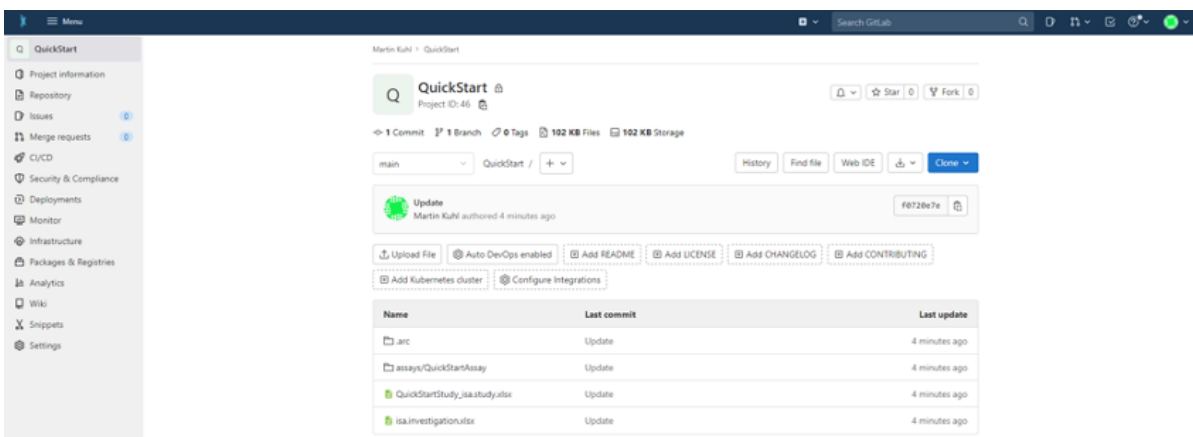
Start Arc Sync
GIT: Martin Kuhl
GIT: martinkuhl137@gmail.com
-----
GIT: Tracking "**/dataset/**"
-----
On branch main
Untracked Files:
(use "git add <files>..." to include in what will be committed)
.gitattributes
  
```

Note: In case you did not set your git user name and email address you might get a warning to do so. These are needed for displaying them on the git commits. You can update the settings with

```

git config --global user.name <your_name>
git config --global user.email <your_email>
  
```

5. Check if the upload was successful by visiting the respective URL.



Invite collaborators

You can invite lab-colleagues or project partners to join your ARC for collaboration. While inside your ARC on the DataHUB, click on *Project information* -> *Members* in the left navigation panel. Search for registered researchers and select a role for each individually. These roles come along with different rights.

Briefly:

- **Guests:** Have the least rights. This is recommended for people you ask for consultancy.
- **Developers:** The choice for most people you want to invite to your ARC. Developers have read and write access, but cannot maintain the project on the DataHUB, e.g. inviting others.
- **Maintainers:** Gives the person the same rights as you have (except of removing you from your own project). This is recommended for inviting PIs or group leaders allowing them to add their group members for data upload or analysis to the project as well.

The screenshot shows the 'Project members' page in the GitLab QuickStart interface. The left sidebar contains a navigation menu with options like Project information, Activity, Labels, Members (selected), Repository, Issues, Merge requests, CI/CD, Security & Compliance, Deployments, Monitor, Infrastructure, Packages & Registries, Analytics, Wiki, Snippets, and Settings.

The main content area is titled 'Project members' and includes a sub-header 'You can invite a new member to QuickStart or invite another group.' Below this are two tabs: 'Invite member' (active) and 'Invite group'.

The 'Invite member' form contains the following fields:

- GitLab member or Email address:** A search bar with the placeholder text 'Search for members to update or invite'.
- Select a role:** A dropdown menu currently set to 'Guest'. A link 'Learn more about roles.' is provided below the dropdown.
- Access expiration date:** A date input field with the placeholder text 'Expiration date'.
- Buttons:** 'Invite' and 'Import' buttons.

Below the form is a table titled 'Members' showing the current project members:

Account	Source	Access granted	Max role	Expiration
Martin Kuhl (xym) @martinkuhl	Direct member	6 days ago by Martin Kuhl	Maintainer	Expiration date

Note: A detailed usage instruction for the ARC Commander can be found [here](#).

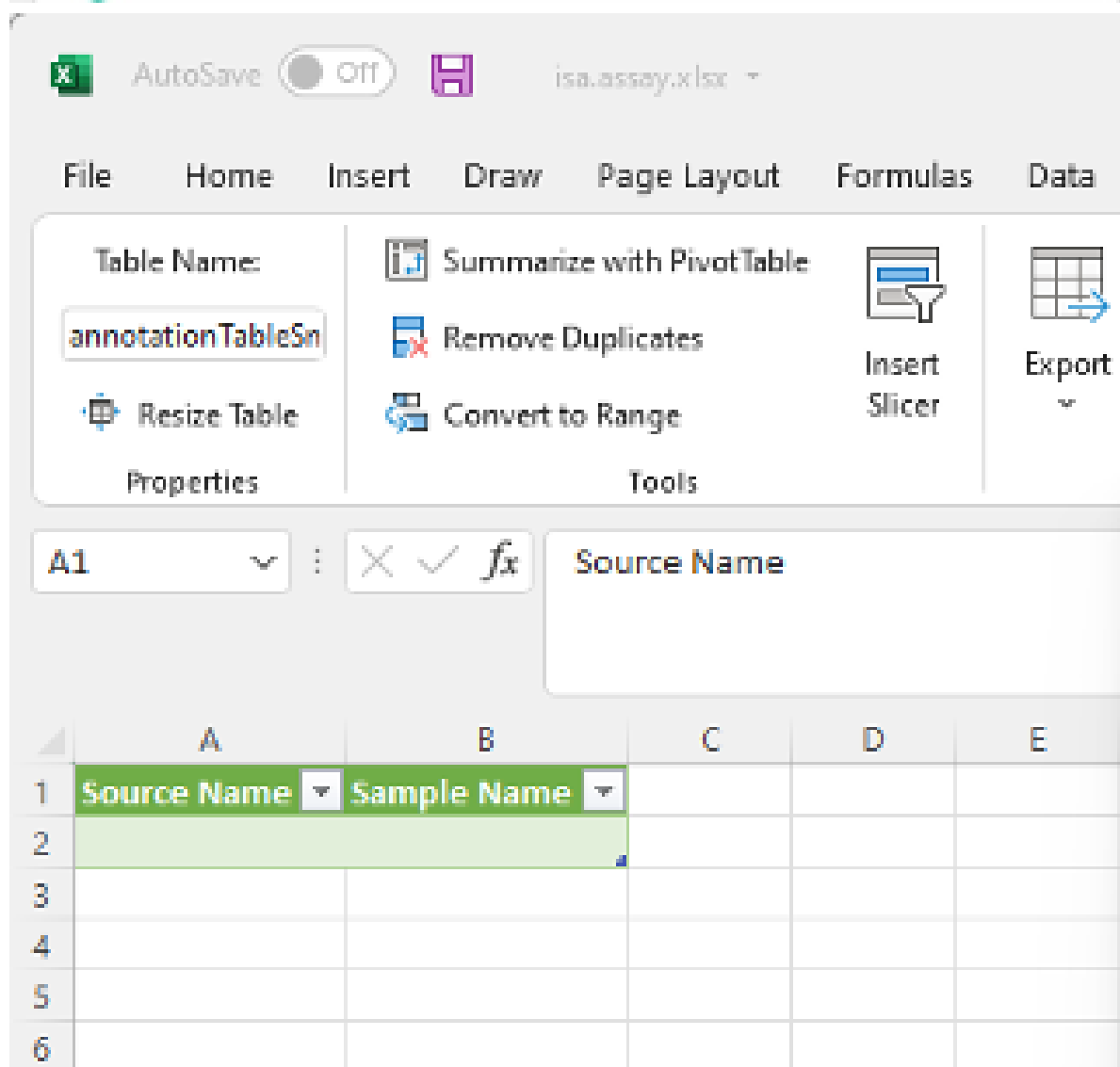
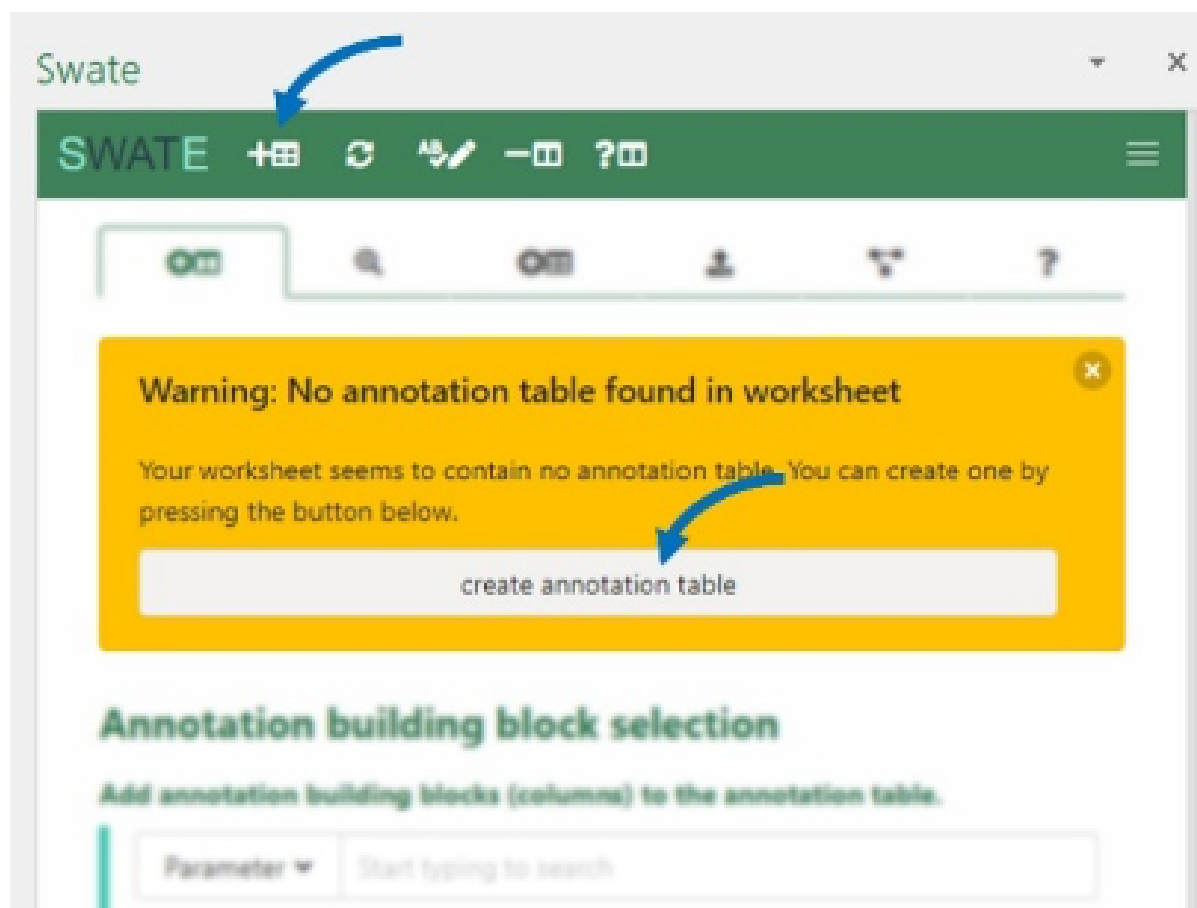
Data annotation

Your ARC should now contain one `isa.investigation.xlsx` and one or more `isa.study.xlsx` and `isa.assay.xlsx` file(s), respectively. Use the `isa.study.xlsx` to describe the characteristics of your samples, e.g. how you grew your plant, and `isa.assay.xlsx` to annotate the experimental analyses.

SWATE

DataPLANT provides the Excel Add-In SWATE to support you in data annotation.

- ☐ Download and install the newest SWATE version according to [these instructions](#).
- ☐ In case you use an Excel version older than Excel 2019, please install [SWATE for Excel online](#).
- Use the *create annotation table* button in the yellow pop-up box (this only appears if you start SWATE on an Excel worksheet without an existing annotation table). An annotation table with the building blocks *Source Name* and *Sample Name* will be generated.



- Annotate your table with help of the [annotation principles](#).

Briefly:

- *Characteristics* are used for study descriptions and describe inherent properties of the source material (e.g. a certain strain).
- *Parameters* describe steps in your experimental workflow (e.g. an instrument model or a growth chamber), and
- *Factors* describe independent variables that result in a specific output (e.g. the light intensity).
- The combination of ISA (Characteristics, Parameter, Factor) and a biological or technological ontology (e.g. temperature, strain, instrument model) gives the flexibility to display an ontology term, e.g. temperature, as a regular process parameter or as the factor your study is based on (Parameter [temperature] or Factor [temperature]).

Customize your table by adding building blocks

1. Choose the type of building block you want to add (A).
2. If you chose a descriptive building block type (building blocks besides Sample Name, Source Name, and Data File Name), use search field (B) to search for an Ontology Term. SWATE accesses the SwateDB with a list of established external ontologies designated suitable for use in plant science. In addition, we feature our own ontology NFDI4PSO to extend the DB with missing, but necessary terms.
3. If you want to add a building block with a unit, check box (C) and use search field (D) to look for a fitting unit term, e.g. degree Celsius as unit for Parameter [temperature].
4. If you could not find a fitting term, you can use the Advanced Term Search with the blue links above the *Add building block* button. If you still could not find a fitting term, use free text input.

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](#)

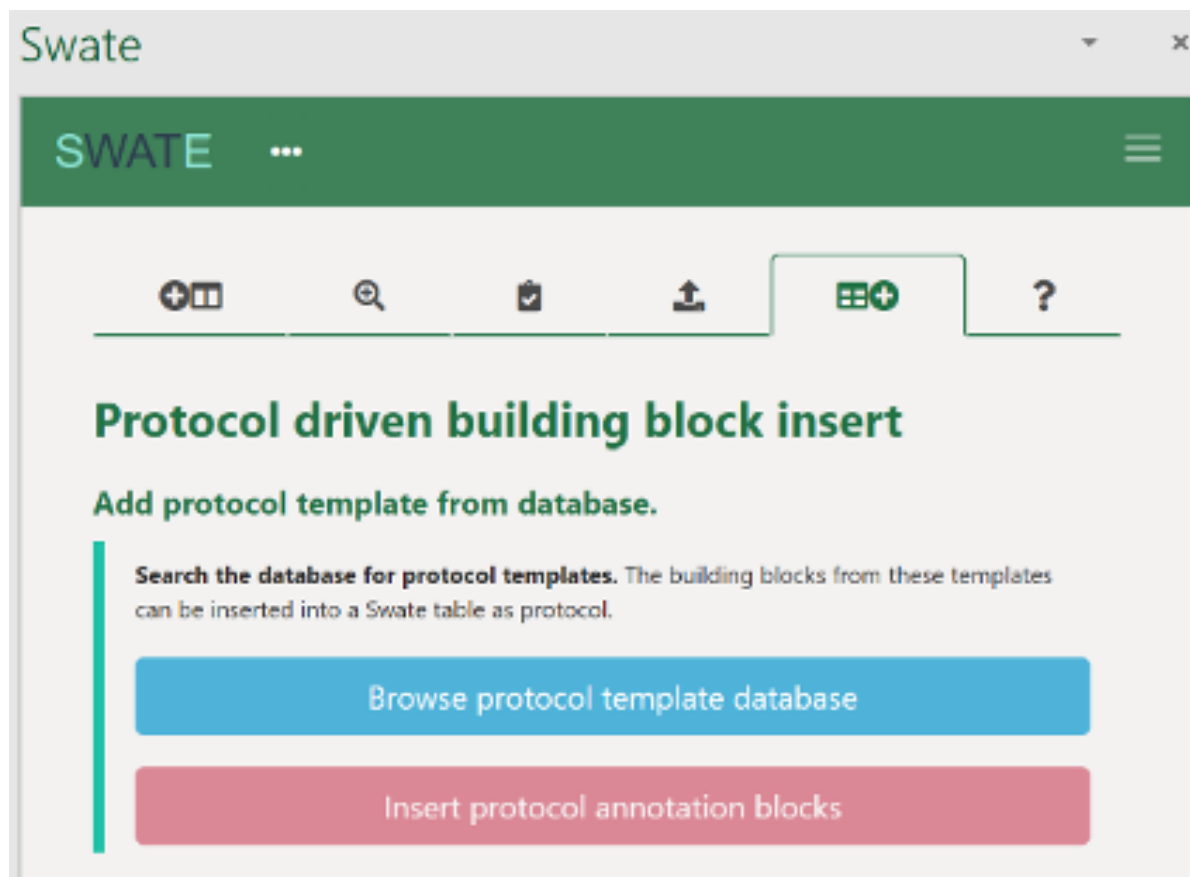
Add building block

Add/Update unit reference to existing building block.

5. For more information on customizing your annotation table click [here](#).

Use templates

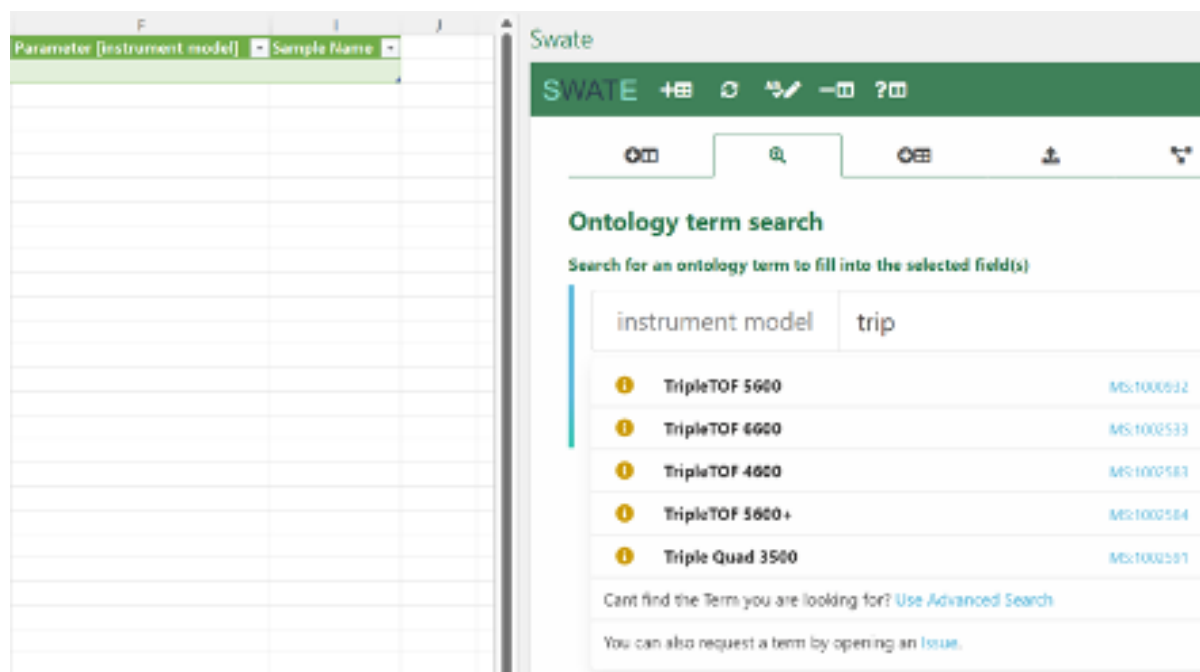
Alternatively, you can also use one of DataPLANT's [SWATE templates](#). You can find them under the *Protocol Insert* tab in SWATE.



Annotate your samples and data

Fill the cells beneath each building block with ontology terms to note the respective *Characteristics*, *Parameter*, and *Factor* values of your experiment. Using the ontology term search function, you can fill multiple cells at once.

1. When *Use related term directed search* (A) is enabled, SWATE will suggest a selection of suitable terms within the ontology for the column header, e.g. *TripleTOF 5600* for *instrument model*.
2. When term directed search (A) is disabled, SWATE will still suggest ontology terms, but without relation to the column header.
3. If you could not find a fitting term, use free text input.



Note: More information on how to use SWATE can be found [here](#).

DataPLANT Support

For further assistance, feel free to reach out via our [helpdesk](#) or by contacting us [directly](#).

The Minimalist's ARC-QuickStart

- ☒ You know how to use a command line
- ☒ You have created an ARC before
- ☒ The latest version of the [ARC Commander](#) as well as [git](#) and [git LFS](#) are installed on your computer
- ☒ You have a [DataPLANT](#) account
- ☒ Your computer is linked to the [DataHUB](#) via an ssh key or a personal access token.

Voila! You are ready to follow these few steps to create a minimal ARC sharable via DataPLANT's DataHUB:

1. Visit the [DataHUB](#), create a new repository and copy the URL to your clipboard.
2. Replace the `<variables>` in the following code block with your information and execute it in your command line.

```
# Clone your repository locally
arc get -r git@git.nfdi4plants.org:<YourUserName>/<YourRepository.git>

# navigate to the cloned repository
cd <YourRepository>

# Setup the ARC structure with one study and one assay
arc init
arc i create -i "<YourInvestigation>"
arc a add -s "<YourStudy>" -a "<YourAssay>"
arc sync -m "initialize ARC structure"
```