



**CEPLAS**

Cluster of Excellence on Plant Sciences

# Start Your ARC Workshop Series

(Preparation for) Session 03

Dominik Brillhaus - CEPLAS Data Science



# Checklist ARC data annotation hands-on



- ✓ Install Swate on your computer
- ✓ Prepare an ARC with your data

# Swate Installation

Please follow [these instructions](#) to install the latest version of Swate.

# Part 1: Prepare your ARC

## Your fresh ARC folder

1.  Create a new folder, which you want to initialize as an ARC.
2.  Open the command line inside the folder or navigate via command line to that folder.

For example:

```
mkdir -p ~/Desktop/workshop-arcs/my-arc  
cd ~/Desktop/workshop-arcs/my-arc
```

# ARC initialization

Initialize your ARC by executing

```
arc init
```

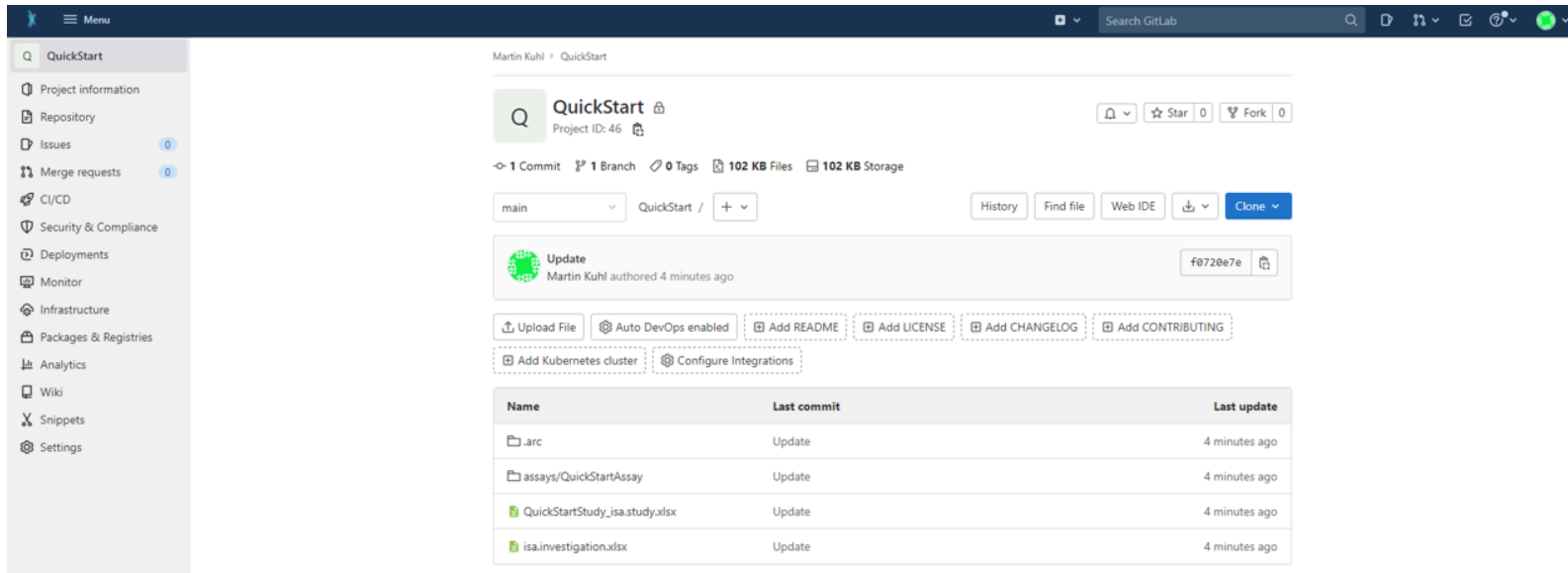
## Upload your local ARC to the DataHUB

In order to connect your local ARC with a remote ARC, you can specify the remote address with the flag `-r` followed by the URL of the remote ARC

```
arc sync -r https://git.nfdi4plants.org/<username>/my-arc
```

# Successful ARC synchronization

Check if the upload was successful by visiting your ARC at the respective URL in your browser.



The screenshot shows the GitLab interface for a project named "QuickStart". The left sidebar contains a menu with options like Project information, Repository, Issues, Merge requests, CI/CD, Security & Compliance, Deployments, Monitor, Infrastructure, Packages & Registries, Analytics, Wiki, Snippets, and Settings. The main content area displays the project details, including the repository name "QuickStart", Project ID "46", and statistics: 1 Commit, 1 Branch, 0 Tags, 102 KB Files, and 102 KB Storage. A green "Update" button is visible, indicating a successful commit. Below the update button, there are options to "Upload File", "Auto DevOps enabled", "Add README", "Add LICENSE", "Add CHANGELOG", "Add CONTRIBUTING", "Add Kubernetes cluster", and "Configure Integrations". A table lists the files in the repository, showing their names, last commit status, and last update time.

Name	Last commit	Last update
.arc	Update	4 minutes ago
assays/QuickStartAssay	Update	4 minutes ago
QuickStartStudy_isa.study.xlsx	Update	4 minutes ago
isa.investigation.xlsx	Update	4 minutes ago



## Part 2: Adding subfolders for your data

# Preparation

Before you can fill your ARC, let's revise your data.

- What is my **investigation**?
- What is my **study**?
- Which **assay** did I perform?
  - What is my (raw) **dataset**?
  - What **protocols** did I use?

💡 It might help to have this info ready in a text file. In particular, try to think about useful identifiers for investigation, study and assay.

## ISA investigation

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

```
arc investigation create --identifier <YourInvestigationID>
```

💡 Avoid using spaces in the identifier. Use underscores and capital letters instead.

## ISA studies and assays

The ISA study ( **s** ) and ISA assay ( **a** ) workbooks allow you to annotate your experimental data.

## Adding a study

Add a study folder with the isa.study.xlsx workbook including an identifier to your ARC with

```
arc s add --studyidentifier <YourStudyID>
```

## Adding an assay

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

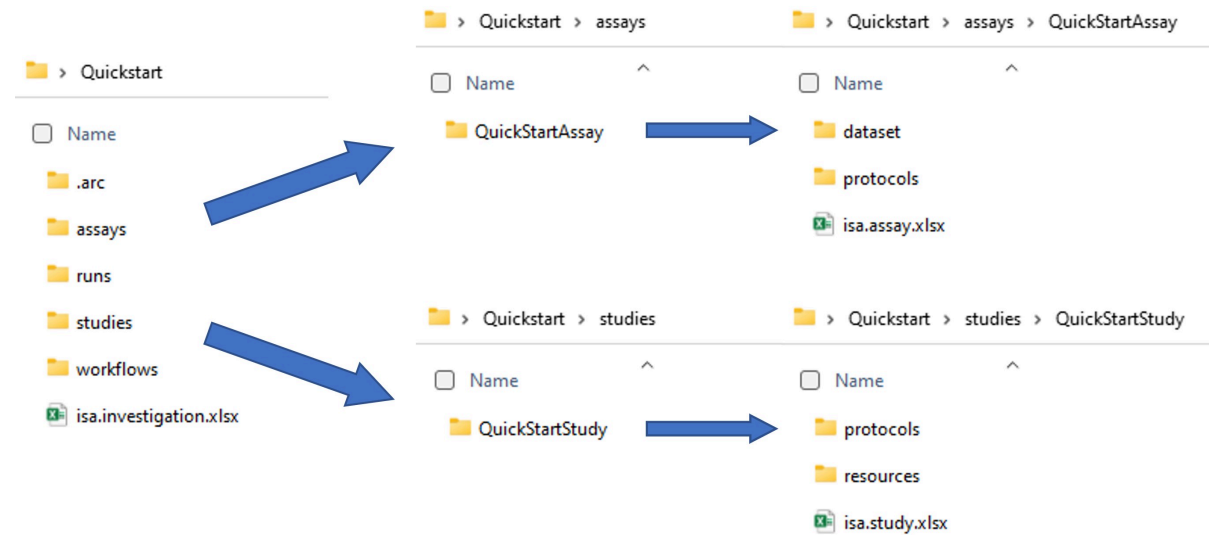
```
arc assay add --studyidentifier <YourStudyID> --assayidentifier <YourAssayID>
```

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

# ISA studies and assays

The ARC Commander adds subdirectories to the *studies* and *assays* folders. 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



## Sync your ARC to the DataHUB

To save the changes, sync the ARC to the DataHUB including a message ( `-m` ).

```
arc sync -m "initiate arc structure"
```

 Check your ARC in the DataHUB.



# Part 3: Structure your data

# Start adding your data to the ARC

Try to place your resources, datasets and protocols for each study and assay in the respective folders.

💡 This part will likely require some iteration. 🗣️ Talk to us.

- You might need to add new studies and assays or revise your overall structure during the process
- Sometimes you might not immediately find a suitable position for your data

## Part 4: Invite collaborators

- Unless changed, your ARC is set to private by default.
- To collaborate, you can invite lab colleagues or project partners to your ARC by following the steps on the subsequent slides.
- To get started [sign in](#) to the DataHUB and open the ARC you want to share.

# 1. Click on **Project Information** in the left navigation panel

Demo User / Demo\_ARC

[CQC pipelines](#) on DataHUB received major updates and are now opt-in. If you used the DataHUB publication service before, head [here](#) to learn how to opt-in.

## Demo\_ARC

main Demo\_ARC / +

History Find file Edit Code

arc init  
Demo User authored 1 week ago

Name	Last commit	Last update
assays	arc init	1 week ago
runs	arc init	1 week ago
studies	arc init	1 week ago
workflows	arc init	1 week ago
.gitignore	arc init	1 week ago
isa.investigation.xlsx	arc init	1 week ago

### Project information

pipeline passed Publish ARC

- 1 Commit
- 2 Branches
- 0 Tags
- 21 KiB Project Storage
- Auto DevOps enabled
- Wiki
- + Add README
- + Add LICENSE
- + Add CHANGELOG
- + Add CONTRIBUTING
- + Add Kubernetes cluster
- + Configure Integrations

**Created on**  
July 13, 2024

# 2. Click on Members

13

12

Search or go to...

Project

Demo\_ARC

Pinned

Manage

Activity

Members

Labels

Plan

Code

Build

Secure

Deploy

Operate

Monitor

Analyze

Settings

Demo User / Demo\_ARC / Members

CQC pipelines

 on DataHUB received major updates and are now opt-in. If you used the DataHUB publication service before, head [here](#) to learn how to opt-in.

Project members

Import from a project

Invite a group

Invite members

You can invite a new member to Demo\_ARC or invite another group.

Members 1

Filter members

Account

Account	Source	Max role	Expiration	Activity
<div><div></div><div>Demo User</div><div>@DemoUser</div></div> <div>It's you</div>	Direct member by <a href="#">Demo User</a>	Owner	<div>Expiration date</div> <div></div>	<div>8+ Sep 27, 2023</div> <div>✓ Jul 13, 2024</div> <div>⌂ Jul 21, 2024</div> <div></div>

### 3. Click on **Invite members**

13

12

Search or go to...

Project

Demo\_ARC

Pinned

1Manage

Activity

2Members

Labels

Plan

Code

Build

Secure

Deploy

Operate

Monitor

Analyze

Settings

Demo User / Demo\_ARC / Members

CQC pipelines on DataHUB received major updates and are now opt-in. If you used the DataHUB publication service before, head [here](#) to learn how to opt-in.

Project members

Import from a project

Invite a group

3Invite members

Members 1

Filter members

Account

Account	Source	Max role	Expiration	Activity
<div><div></div><div>Demo User</div><div>@DemoUser</div><div>It's you</div></div>	Direct member by <a href="#">Demo User</a>	Owner	<div>Expiration date</div>	<div>8+ Sep 27, 2023</div> <div>✓ Jul 13, 2024</div> <div>⌂ Jul 21, 2024</div>

# Invite members



You're inviting members to the **Demo\_ARC** project.

**Username, name or email address**

4

Select members or type email addresses

**Select a role**

Guest



[Read more](#) about role permissions

**Access expiration date (optional)**

YYYY-MM-DD



## Invite members



✓ Guest

Reporter

Developer

Maintainer

Owner

4

5

\_ARC project.

ses

Guest



[Read more](#) about role permissions

**Access expiration date (optional)**

YYYY-MM-DD





# Choosing the proper role

## Guests

Have the least rights. They will not be able to see the content of your ARC (only the wiki page).

## Reporters

Have **read access** to your ARC. 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.

*A detailed list of all permissions for the individual roles can be found [here](#)*

# Congratulations!

You have just shared your ARC with a collaborator.



# Contributors

Slides presented here include contributions by

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