ARC and elabFTW Introduction IMET

February, 2025

Sabrina Zander (MibiNet)

Stella Eggels (DataPLANT)

Tentative agenda

Morning

Time	Topics
09:30 - 09:45	Welcome, Intro RDM and NFDI
09:45 - 10:15	Overview DataPLANT, and DataPLANT Tool chain
10:15 - 10:45	Hands-on Start Here Guide Part I
10:45 - 11:00	*Short break* ☕
11:00 - 11:30	Metadata Intro
11:30 - 12:15	Hands-on Start Here Guide Part II

Afternoon

Time	Topics
12:15 - 13:15	*Lunch break* 🍕
13:15 - 14:15	Hands-on Start Here Guide Part III
14:15 - 14:30	DataHUB Intro
14:30 - 15:00	Hands-on Start Here Guide Part IV
15:00 - 15:30	Q & A ARC
15:30 - 15:45	*Short break* ☕
15:45 - 16:15	elabFTW Intro
16:15 - 16:45	elab2ARC tool
16:45 - 17:30	Q & A and wrap-up

House-keeping

- Use the pad to raise questions and feedback
- Copy / paste links (hands-on)

Training Materials

Slides are shared via DataPLANT Training Material

Resources – join the open source movement

DataPLANT (nfdi4plants)

DataPLANT Website: https://nfdi4plants.org/

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

DataHUB: https://git.nfdi4plants.org

GitHub: https://github.com/nfdi4plants

HelpDesk: https://helpdesk.nfdi4plants.org

💡 You can help us by raising issues, bugs, ideas...

🚀 NEW! ARC website: https://arc-rdm.org

Continuous support

Data managers in Düsseldorf, Cologne, Jülich and close by (CEPLAS, MibiNet, TRR341) offer support.

1. Slack Workspace for ad hoc support
2. Monthly user meeting (2nd Friday of the month)

→ Details.

CEPLAS
Cluster of Excellence on Plant Sciences

trr_341 plant ecological genetics

MibiNet

Introduce yourself

- Lab
- MibiNet / CEPLAS / TRR / ?
- Used code / programming language before
- Experience with Git / GitLab / GitHub?
- Have an ORCID
- My motivation / expectation
- My favorite lab assay

Let's draw a typical lab workflow 📝

BYOD – Goals

- First few steps into ARC ecosystem
- Move existing datasets into ARCs
- Share them via the DataHUB
- Annotate with metadata
- Introduction into elabFTW
- elab2ARC tool

FAIR data stewardship

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

https://doi.org/10.1038/sdata.2016.18

nature > scientific data > comment > article

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons

— Show fewer authors

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

The FAIR principles

Is your data FAIR?

Findable | **A**ccessible | **I**nteroperable | **R**eusable

- 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?

MibiNet – One SFB, six locations

Data silos impede collaboration

Missing interfaces impede collaboration

NFDI – Nationale Forschungsdaten Infrastruktur

nfdi Nationale Forschungsdaten Infrastruktur

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.de

NFDI – A community-targeted approach for RDM

nfdi.de

MibiNet connection to NFDIs

DataPLANT – *NFDI4plants*

- https://nfdi4plants.org
- https://arc-rdm.org

Data Stewardship between DataPLANT and the community

Community

nfdi4plants

Domain experts
User experience
Training

Service provider
Developers
Tech experts

Annotated Research Context (ARC)

Your entire investigation in a single unified bag

You can store your ARC in the DataHUB

ARCs are versioned

You can invite collaborators

Collaborate and contribute

Reuse data in ARCs

Publish your ARC

Publish your ARC, get a DOI

From ARC to repositories

Moving from paper to data publications

Project management

Weil, H.L., Schneider, K., et al. (2023), PLANTdataHUB: a collaborative platform for continuous FAIR data sharing in plant research. Plant J. https://doi.org/10.1111/tpj.16474

The ARC ecosystem

What does an ARC look like?

ARCs store experimental data

Computations can be run inside ARCs

ARCs come with comprehensive metadata

ARC builds on standards

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

https://isa-tools.org/ | https://www.commonwl.org/
https://www.researchobject.org/ro-crate/ | https://git-scm.com

Everything is a file

DataPLANT employs a **data-centric** approach to RDM

No technical lock-in

(Meta)data transparency with tool assistance but **no technical lock-in**

Hands-on part 1: Start Here guide

Start Here guide in the DataPLANT knowledge base.

- follow the fictional researcher Viola on her journey to create her first ARC

A small prototypic project

Hands-on part 1: Setup and Investigation

Follow the **Start Here** guide in the DataPLANT knowledge base.

📝 Stop after step **Describing the Investigation**

❓ Please ask if you have any question/problems ❓

Metadata

What is metadata?

ARCs come with comprehensive metadata

ARC builds on standards

https://isa-tools.org/ | https://www.commonwl.org/
https://www.researchobject.org/ro-crate/ | https://git-scm.com

ARC builds on ISA

Investigation

Overall goals
Scientific context

Study

Experimental design

Assay

Leading to
(raw) data

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

ISA abstract model in a nutshell

isa

Investigation
administrative (meta)data

- Summary
  - Titel
  - Description
- Person
- Organisation
- Publication reference

Study
descriptive (meta)data
information on the subject

- Characteristics
- Parameters
- Components
- Factors

Assay
descriptive (meta)data
information on the measurement

- Characteristics
- Parameters
- Components
- Factors

ARC builds on ISA to connect data

ARC builds on ISA to link data

- Samples are linked study-to-assay, assay-to-assay
- Raw data is linked to assays
- Protocols can be referenced
- …

Annotation by flattening the knowledge graph

- Low-friction metadata annotation
- Familiar spreadsheet, row/column-based environment

Annotation principle

- Low-friction metadata annotation
- Familiar spreadsheet, row/column-based environment

Adding new building blocks

Swate (now integrated in ARCitect) 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)
- Protocol columns
- Characteristic // Parameter // Factor // Component
- Output (e.g. Sample Name, Raw Data File, Derived Data File)

Let's take a detour on Annotation Principles

Ontology term search

Enable **related term directed search** to directly fill cells with child terms

Fill your table with ontology terms

Realization of lab-specific metadata templates

Facilities can define their most common workflows as templates

Directly import templates via Swate

- DataPLANT curated
- Community templates

Build on existing well-established standards

Integration of established (meta)data standards ensures compatibility by design

ISA and CWL – Connected by similarity

isa

COMMON
WORKFLOW
LANGUAGE

Hands-on part 2: Start Here guide, study & assay part

Start Here guide in the DataPLANT knowledge base.

Divide and conquer for reproducibility

Identifying the ‘study’ part

Protocol REF

growth_protocol.txt

A table-based organization schema

Referencing a protocol

This allows you to reference the free-text, human-readable protocol.

- It is recommended that the protocol is in an open format (.md|.txt|.docx|…)
- But everything is possible also an URI to an electronic lab notebook

Parameterizing the ‘study’

Finding the right metadata vocabulary

Parameters []

- Light intensity 200 µEinstein
- Temperature 6°C / 25°C
- Growing 4d

Characteristics []

- Arabidopsis thaliana
- Leaf
- Hydroponic culture
- Columbia

OLS: Finding the right metadata vocabulary

Temperature Dependence Temperature:Dependence_Annotation

http://purl.uniprot.org/core/Temperature_Dependence_Annotation

Indicates the optimum temperature for enzyme activity and/or the variation of enzyme activity with temperature variation; the thermostability/thermolability of the enzyme is also mentioned when it is known.

Ontology: UNIPROTRDFS

temperature AFO:/result#AFR_0001584

http://purl.allotrope.org/ontologies/result#AFR_0001584

A temperature (datum) is a quantity facet that quantifies some temperature. [Allotrope]

Ontology: AFO

temperature FBcv:0000466

http://purl.obolibrary.org/obo/FBcv_0000466

Mutation caused by exposure to a temperature that is higher or lower than 25 degrees Celsius.

Ontology: FBCV

temperature PATO:0000146

http://purl.obolibrary.org/obo/PATO_0000146

A physical quality of the thermal energy of a system.

Ontology: PATO

Also appears in: NGBO HTN CAO ZP AGRO OMIABIS OBIB MONDO TXPO MCO +

Ontology Lookup Sevice (OLS): https://www.ebi.ac.uk/ols4/

Finding the metadata vocabulary and descriptors

Finding the metadata vocabulary and descriptors

Parameters []

-
  - Light intensity 200 µEinstein
-
  - Temperature 6°C / 25°C
-
  - Growing 4d

Characteristics []

-
  - Arabidopsis thaliana
-
  - Leaf
-
  - Hydroponic culture
-
  - Columbia

Finding the metadata vocabulary and descriptors

Parameters []

- Light intensity
  - 200 µEinstein
- Temperature
  - 6°C / 25°C
- Growth time
  - 4d

Characteristics []

- Organism
  - Arabidopsis thaliana
- Tissue
  - Leaf
- Growth medium
  - Hydroponic culture
- Ecotype
  - Columbia

Identifying assays

Assay for sugar measurement

Separating different assay elements

Isolating the lab processes in an assay

Parameteterization: sugar extraction

- Vortex Mixer
  - 3 seconds
- Temperature
  - 95 degree celsius

Parameteterization: sugar measurement

- technical replicate
  - 1,2,3,...
- sample volume
  - 10 `microliter`
- buffer volume
  - 190 `microliter`
- cycle count
  - 5

Save time using standard methods and SOPs

Parameter []

- Protein Precipitation
  - acetone
- cleavage agent name
  - Trypsin
- sample preparation
  - reversed-phase solid-phase extraction
- ...

Component []

- chromatography instrument model
  - nanoElute2
- chromatography column model
  - PepSep C18 1.9u, 25cm x 75um
  - ...

Applying standard procedures to sample record

Hands-on part 2: ARCitect (and Swate)

Continue the **Start Here** guide in the DataPLANT knowledge base.

📝 Stop after step Add assay data

Hands-on part 3: Data

Options to annotate the data analysis

A computational workflow is like a protocol

Option I: Create a virtual assay

python Copy code

```python
import pandas as pd
import plotly.express as px

# Read the CSV file
data = pd.read_csv('result.csv')

# If the CSV doesn't have a clear index or row/column names, you may need to manually set
# For example, if the first column should be the index, you can use:
# data = pd.read_csv('result.csv', index_col=0)

# Create a heatmap using Plotly
fig = px.imshow(data,
                labels=dict(x="Columns", y="Rows", color="Value"),
                x=data.columns,
                y=data.index)

# Show the heatmap
fig.show()
```

heatmap.py

Option II: Create a workflow and run

```python
import pandas as pd
import plotly.express as px

# Read the CSV file
data = pd.read_csv('result.csv')

# If the CSV doesn't have a clear index or row/column names, you may need to manually set
# For example, if the first column should be the index, you can use:
# data = pd.read_csv('result.csv', index_col=0)

# Create a heatmap using Plotly
fig = px.imshow(data,
                labels=dict(x="Columns", y="Rows", color="Value"),
                x=data.columns,
                y=data.index)

# Show the heatmap
fig.show()
```

heatmap.py

Use CWL to wrap your workflow

CWL workflow document (*.cwl)

CWL job parameter (*.yaml)

CWL
tool descriptors (*.cwl)

Hands-on part 3: Data

Continue the **Start Here** guide in the DataPLANT knowledge base.

📝 We will skip the **Add a datamap** step

📝 Do the **Data analysis** step and stop

DataHUB

DataHUB overview

ARC builds on standards + Git

git

The DataPLANT DataHUB – a GitLab *Plus*

=

GitLab

+

On-premise DataHUBs

ARC services are available as on-premise option

Using the DataHUB to collaborate

Project management

Weil, H.L., Schneider, K., et al. (2023), PLANTdataHUB: a collaborative platform for continuous FAIR data sharing in plant research. Plant J. https://doi.org/10.1111/tpj.16474

ARCs come with their own wiki space

- directly associated to your ARC
- same access rights as your ARC
- share meeting minutes or ideas with collaboration partners
- keep ARC clean of files that are not considered "research data"

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 & 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.

Hands-on part 4: DataHUB

Continue the **Start Here** guide in the DataPLANT knowledge base.

📝 Until the final **complete** step

DataHUB concepts and terminology

Where do I store my ARC?

ARC storage and sharing

- DataHUB as "ground truth" / original clone
- You can sync and communicate all changes to your ARC via the DataHUB

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

Example setup to store and use ARCs

Personal computer

- work on small files
- annotate metadata
- add scripts, protocols

Workstation / Server

- work on large files
- run computations

FileShare

- mount to local machine, sync ARC from there

HPC

- direct connection HPC to DataHUB (depends on security settings)
- or mount to local machine and sync, ARC from there

Projects and Groups are not the same

- "Project" = ARC
- "Groups" = Group of users

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

Options to share an ARC via the DataHUB

ARC DataHUB members // ARC Investigation contacts

DataHUB: ARC members

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

ARCitect: Investigation Contacts

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

elabFTW Introduction

Login into eLabFTW HHU

eLab-FTW - Dashboard

eLabFTW - Experiment panel

Default "working area" for protocols of daily lab work

eLabFTW - Resource panel

Place to categorize content

eLabFTW - Experiment information

eLabFTW - Tag usage

Very individual, some examples:

- Investigation ID
- Subgroup
- Organism /part (e.g. A.thaliana, leaf, root, mitochondria)
- Goal (strain generation, plasmid generation)
- Methode (e.g. transformation, western blot, extraction, biolector)
- Genes/ Proteins (e.g. mKate2, eGfp, Rrm4)

Use keywords you would search/filter for

📝 Make rules for tag usage

eLabFTW - Extra fields

eLabFTW - Extra fields examples

Dropdown	URL	Number	Date

eLabFTW - Experiment categories examples

- Study
- Assay
- Discussion
- Teaching
- Demo
- R&D
- Production

eLabFTW - Categories in Resources

Use Resources to categorize content & manage lab.

Use unique identifier & naming conventions (e.g. Methode MTH0001, Media MED0001, Plasmid PL0001).

- List of Investigation
- Investigation
- Plant sheet
- Plasmid sheet
- Equipment
- Collections
- Methods
- Manuals
- Media & Solution
- Formblatt Z

eLabFTW - Scheduler

eLabFTW - Booking parameter

You can modify booking parameter for reach Resource

eLabFTW - Templates

Create templates for experiments you often use

- „Skeleton“ of a real experiment
- Select „Templates“ from User menu
- You can have different permissions for the template itself and for Experiment that will created from that template
- You can add tags, steps, links to it (e.g. link to project)
- Can also be imported directly in your text (allows merging several templates in one experiment)

eLabFTW - Tipps and tricks

- Use # to link experiments/resources in the text
- Use .gb files for plasmid cards (can be opend in a plasmid viewer in elabFTW)
- Use versions, possibility to upload new versions of files/methods

eLabFTW - Support

General

https://github.com/elabftw/elabftw

https://www.elabftw.net/ (Newsletter)

HHU specific

elabftw@hhu.de

https://rocketchat.hhu.de/channel/elb_eln

https://wiki.hhu.de/display/ZIM/eLabFTW

elab2ARC Documentation

https://nfdi4plants.github.io/nfdi4plants.knowledgebase/resources/elab2arc/

Working Scenario

1. Use ARCs to

- Organize and store (raw) data locally
- Share and communicate data with collaboration partners
- Annotate and prepare data for publication

2. Use elabFTW to

- Document daily lab work „on the fly“
- Organize lab routines with colleagues

eLabFTW & ARC Connection

Aim

- Reducing duplication/ additional workload
- Transformation of data from elabFTW into ARC
- User-friendly, easy implementation

Challanges

- ELNs are unstructed; automated translation and interoperation into well-structured ARC challenging

Solution

- Start with automated transition of data without
  - metadata mapping,
  - structured metadata extraction,
  - re-structuring of elabFTW content

Conversion

1 Content Preparation

ARC

eLabFTW entry

2 eLab2ARC Conversion

3 Updated ARC content shown in ARCitect

Conversion

The elab2ARC tool will automatically convert your elabFTW experiments into ARC format

- create a new assay folder with eLabFTW experiment name as assay name
- create the assay folder structure (dataset/protocols/isa.assay)
- convert experiment main text into a .md file and store it in the protocol folder
- add all attachments of the eLabFTW experiment into the dataset folder
- enter name/email/affiliation of the eLabFTW experiment metadata into the isa.assay sheet

elab2ARC Tool

Tool: https://nfdi4plants.github.io/elab2arc/

Documentation:

https://nfdi4plants.github.io/nfdi4plants.knowledgebase/resources/elab2arc/

Issues/Problems:

https://github.com/nfdi4plants/elab2arc

Contributors

If not referenced otherwise, figures and slides presented here were created by members of DataPLANT (https://nfdi4plants.org). Additional slides were contributed by

- name: Dominik Brilhaus
  github: https://github.com/brilator
  orcid: https://orcid.org/0000-0001-9021-3197
- name: Cristina Martins Rodrigues
  github: https://github.com/CMR248
  orcid: https://orcid.org/0000-0002-4849-1537
- name: Hajira Jabeen
  github: https://github.com/HajiraJabeen
  orcid: https://orcid.org/0000-0003-1476-2121
- name: Kevin Frey
  github: https://github.com/Freymaurer
  orcid: https://orcid.org/0000-0002-8493-1077
- name: Sabrina Zander
  orcid: https://orcid.org/0009-0000-4569-6126
- name: Martin Kuhl