The DataPLANT Ontology Service Landscape

Angela Kranz[1], Stella Eggels[1], Kathryn Dumschott[1], Marcel Tschöpe[2], Kevin Frey[3], Timo Mühlhaus[3], Dirk von Suchodoletz[2], Björn Usadel[1,4]

[1]IBG-4 Bioinformatics, BioSC, CEPLAS, Forschungszentrum Jülich, Jülich, Germany
[2]Computer Center, University of Freiburg, Freiburg im Breisgau, Germany
[3]Computational Systems Biology, RPTU University of Kaiserslautern, Germany
[4]Institute for Biological Data Science, Faculty of Mathematics and Natural Sciences, CEPLAS, Heinrich Heine University, Düsseldorf, Germany

Introduction

As part of the National Research Data Infrastructure (NFDI), the DataPLANT consortium aims to establish a strong and sustainable data infrastructure for advancing plant sciences [1]. At its core, DataPLANT features the Annotated Research Context (ARC), a novel container framework based on the ISA model, Research Object Crate, and Common Workflow Language (Fig. 1) [2]. The ARC offers researchers a standardized way to document and store experimental designs, protocols, workflows, and data.

Here, we present an overview of the DataPLANT Ontology Landscape, demonstrating its utility in easily annotating research data in plant and life sciences. Through the DataPLANT Ontology Landscape, researchers can adeptly handle data annotation challenges, ensuring their research remains accessible, discoverable, and interoperable.

Annotated Research Context

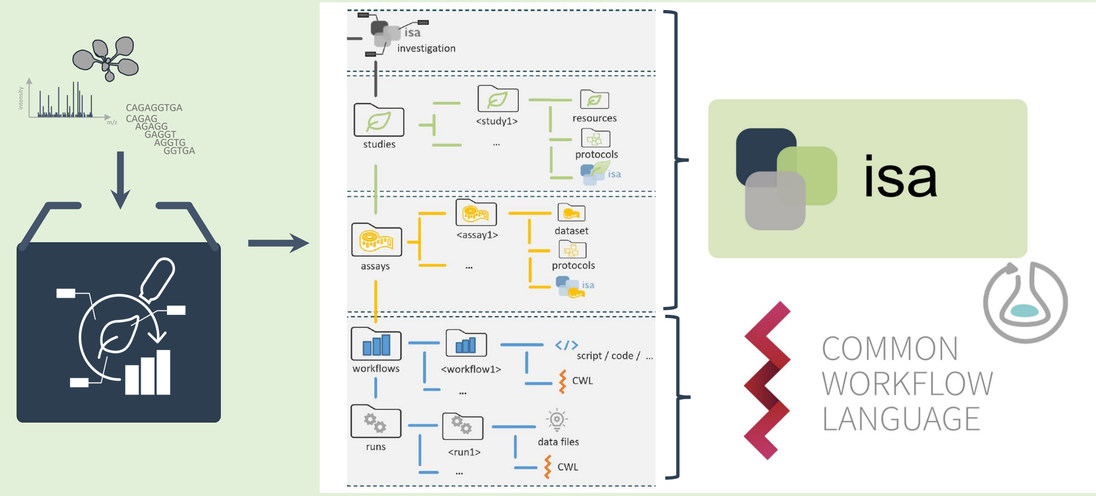

Figure 1: Overview of the ARC folder structure and its relationships to ISA and CWL.

The DataPLANT Toolchain for Data Annotation

The DataPLANT Metadata Toolchain

DataPLANT provides tools and ontologies for efficient metadata annotation in ARCs (Fig. 2):

- The spreadsheet-based tool Swate simplifies metadata annotation [3]
- DataPLANT stores diverse ontologies for plant sciences in the Swate Database (SwateDB) for easy access and updates
- The Swate OBO Updater (Swobup) automates the updating process by fetching ontology source files from a Git repository and integrating changes into SwateDB [4]

The DataPLANT Ontology Collection

- DataPLANT integrates external ontologies for the description of plant-related experimental conditions, traits, chemicals, organisms, protocols, materials, instruments, and data types [5]
- The **DataPLANT Biology Ontology (DPBO)** serves as a broker ontology by providing missing vocabulary that is needed by Swate users for custom metadata annotation [6]
- The ontologies used by DataPLANT can be browsed at the TIB terminology service [7]

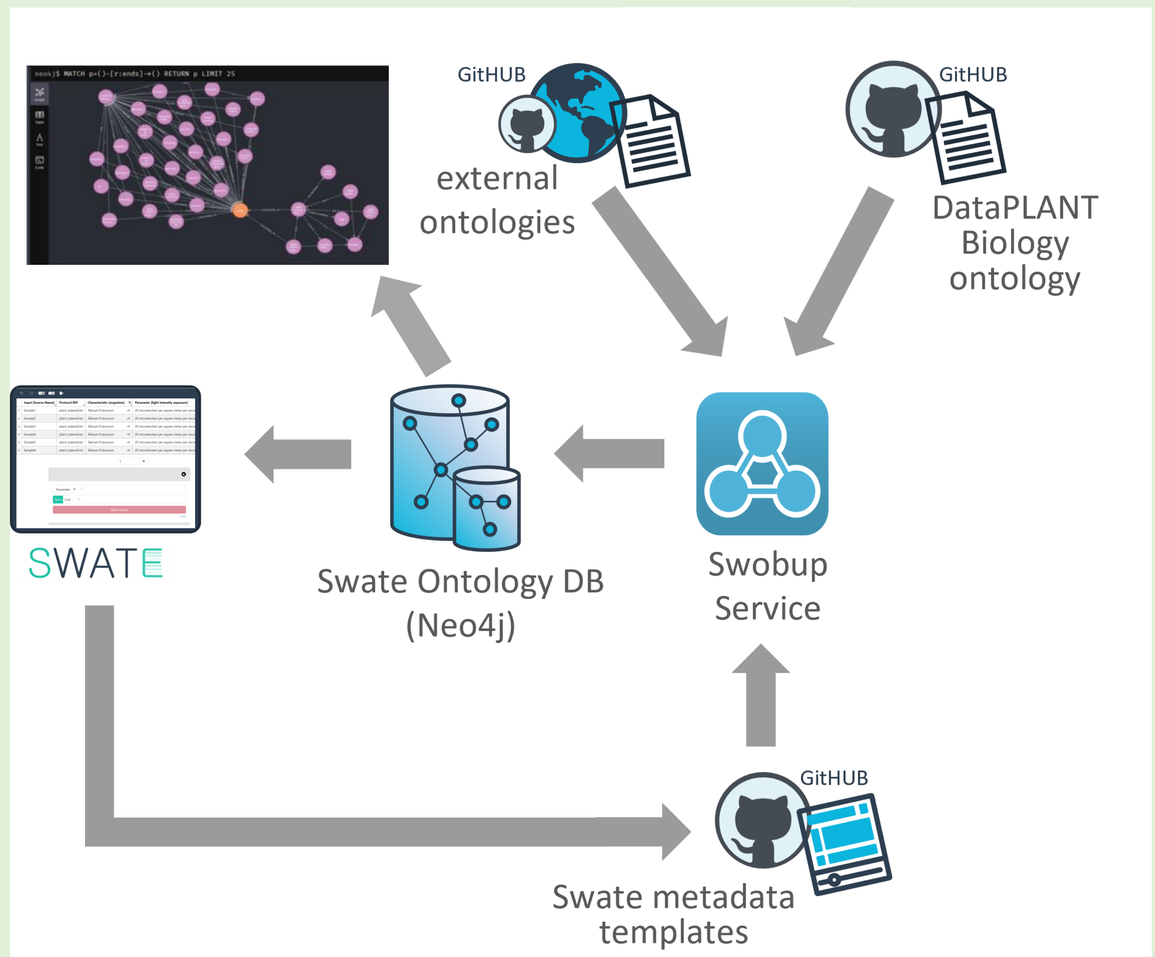

Figure 2: The DataPLANT Metadata Toolchain.

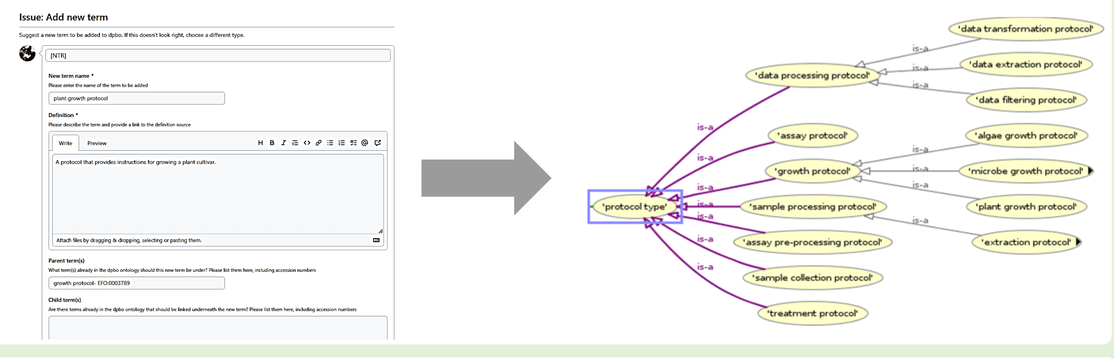

Figure 3: The DPBO curation workflow.

Community Involvement

The DataPLANT Ontology Team encourages community involvement by providing an easy way for users to contribute to ontologies (Fig. 3):

1. Users can report missing terms/ontologies via “Issues” in GitHub
2. The DataPLANT ontology team evaluates and incorporates them into DPBO
3. DataPLANT suggests missing terms back to main ontology providers

Summary

Ontologies are vital for organizing and integrating data seamlessly. DataPLANT uses the tool Swate and external ontologies in a graph database to annotate data. In addition, the DataPLANT Biology Ontology (DPBO) addresses vocabulary gaps during metadata annotation. Here, we emphasize the benefits of ontologies for data integration in plant and life sciences and the importance of collaborative ontology development.

References

[1] Weil et al. (2023), The Plant Journal 116 (4), 974-988. DOI: 10.1111/tpj.16474
[2] https://github.com/nfdi4plants/ARC-specification
[3] https://github.com/nfdi4plants/Swate
[4] https://github.com/nfdi4plants/Swobup
[5] Dumschott et al. (2023), Front Plant Sci. 14, 1279694. DOI: 10.3389/fpls.1279694
[6] https://github.com/nfdi4plants/nfdi4plants_ontology
[7] https://terminology.tib.eu/ts/collections?col=DataPLANT

Funded by
Deutsche Forschungsgemeinschaft
German Research Foundation
Project no. 442077441