



Contribution ID: 120

Type: TALK

Managing research data in plant sciences through the DataPLANT ontology service landscape

Tuesday 5 November 2024 13:00 (20 minutes)

The DataPLANT consortium, a German National Research Data Infrastructure (NFDI), aims to provide plant researchers a robust and sustainable infrastructure for managing research data. Since the complexity of research data continues to grow, effective methods for managing, annotating, and sharing this data becomes increasingly important. DataPLANT integrates different established concepts for FAIR research data management and ontologies to provide tools and services to aid plant researchers in their research data management (RDM).

At the core of the DataPLANT infrastructure is the Annotated Research Context (ARC), a data-centric approach to capturing and structuring the entire research cycle. By leveraging the ISA (Investigation-Study-Assay) standard, Research Object Crate, and Common Workflow Language, the ARC serves as a standardized and comprehensive method for researchers to document their experimental designs, protocols, workflows, and data in a structured format. By utilizing Git services, data provenance is tracked, facilitating collaboration between multiple researchers involved in a common project.

To assist researchers with the ARC creation and data annotation, the Swate tool, a spreadsheet-based software was developed, which allows researchers to annotate their data with standardized metadata. This process leverages selected ontologies relevant in plant research, which are stored in a database (SwateDB) and linked to the Swate tool via an API, allowing users to search for specific terms that fit their needs. In addition, DataPLANT manages the curation of the DataPLANT biology ontology (DPBO), a broker ontology that fills in gaps by providing missing terms not yet available in existing ontologies. SwateDB updates occur through the Swate OBO Updater (Swobup) via Git repository changes, ensuring that researchers have access to the most up-to-date ontologies. Making further use of Git's capabilities, users can easily request new terms during their annotation process and contribute to the SwateDB, either through opening new issues, or through direct contributions via pull requests. The request for the addition of a new term will then be reviewed by the DataPLANT team and incorporated into the DPBO to immediately provide the user with the option to add their term in their metadata spreadsheets. Each newly added term immediately gets a new persistent identifier to serve as an immutable link to this term. As a long-term solution for maintaining the new terms, each new addition will be evaluated individually and pushed to existing ontologies, which have a defined scope that should include this term. If a term is accepted by an external ontology, the original DPBO term will be deprecated and linked to the new term in the external ontology. In the future, this process will be improved by automating the term reading from the spreadsheets and creating new terms in DPBO for every metadata term that was not already taken from the SwateDB. Furthermore, ontologies from other research areas can be easily integrated into the current framework, making it a flexible resource for guiding scientist through their RDM processes.

With our approach, we show that standards such as ISA in combination with ontologies can be efficiently used across all life science domains for (meta)data annotation.

Please specify "other"

In addition, please add 3 to 5 keywords.

ontologies, RDM, DataPLANT, ARC

Please specify "other"

researchers and technicians in their day-to-day lab work, data professionals who provide and maintain infrastructure, data professionals and stewards

For whom will your contribution be of most interest?

other (please specify below)

Please assign yourself (presenting author) to one of the following groups.

Researchers

Primary author: DOERPHOLZ, Hannah (Forschungszentrum Jülich)

Presenter: DOERPHOLZ, Hannah (Forschungszentrum Jülich)

Session Classification: Session E1

Track Classification: Connecting research data: 4. Metadata annotation and management