Metadata in the Research Workflow: Tools for Enrichment and Validation of Structured Metadata

Improving research data management practices is both an organizational and a technical challenge: even in the same research field, (meta)data is often created, stored and processed in an ad-hoc manner. This results in a lack of a clear structure and standardization and makes the metadata "unFAIR". We present two tools that assist scientists in their research workflows to enrich, structure and validate their data and metadata. This increases machine interpretability and reusability, e.g. to ease (automatic) data analysis or metadata harvesting pipelines.

Metador is a web-based structured submission interface for uploading research data and linking it to prede-fined metadata in a structured form. Metadata is supplied by completing a form for each uploaded file. The form is configurable by JSON Schemas and can adjusted by the user depending on the type of the uploaded file. This ensures that captured metadata is specific to the uploaded file type and appropriate to the scientific domain. It is intended for deployment in research groups and designed for quick and easy integration into existing scientific workflows.
Currently we are extending Metador architecture and functionality into a versatile RDM platform focused on metadata standardization and harmonization. It will be designed as an open and extensible ecosystem of reusable generic building blocks and ready-to-deploy services. Combined, they will cover aspects from initial collection of metadata up to improved search, data extraction and data visualization.

DirSchema is a specification and validation tool that enforces requirements concerning the directory structure and metadata provided in datasets. It is intended to be used by researchers and research groups during dataset generation or preparation to harmonize metadata in datasets across users and groups. DirSchema can be used by individual researchers or research groups to validate their dataset directory structures against an agreed-upon JSON Schema based specification that is provided as a YAML file. Further it can be deployed as a building block in other local or web-based scenarios to perform the validation automatically.

Please assign your poster to one of the following keywords.

Tools

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

other (please specify)

Please specify "other" (stakeholder)

HMC Staff (FAIR Data Commons)

In addition please add keywords.

metadata, validation, JSON Schema, YAML

Primary authors: PIROGOV, Anton (Forschungszentrum Jülich GmbH); Ms D'MELLO, Fiona (Forschungszentrum Jülich GmbH); Mr SANDFELD, Stefan (Forschungszentrum Jülich GmbH); Mr HOFMANN, Volker (Forschungszentrum Jülich GmbH)

Presenter: PIROGOV, Anton (Forschungszentrum Jülich GmbH)

Session Classification: Postersession II

Track Classification: Postersession