Concept for an exchangeable metadata structure for electronic labbooks based on Mediawiki

Thomas Gruber[1], David Pape[1], Martin Voigt[1],
Oliver Knodel[1], Guido Juckeland[1]

[1] Department of Information Services and Computing

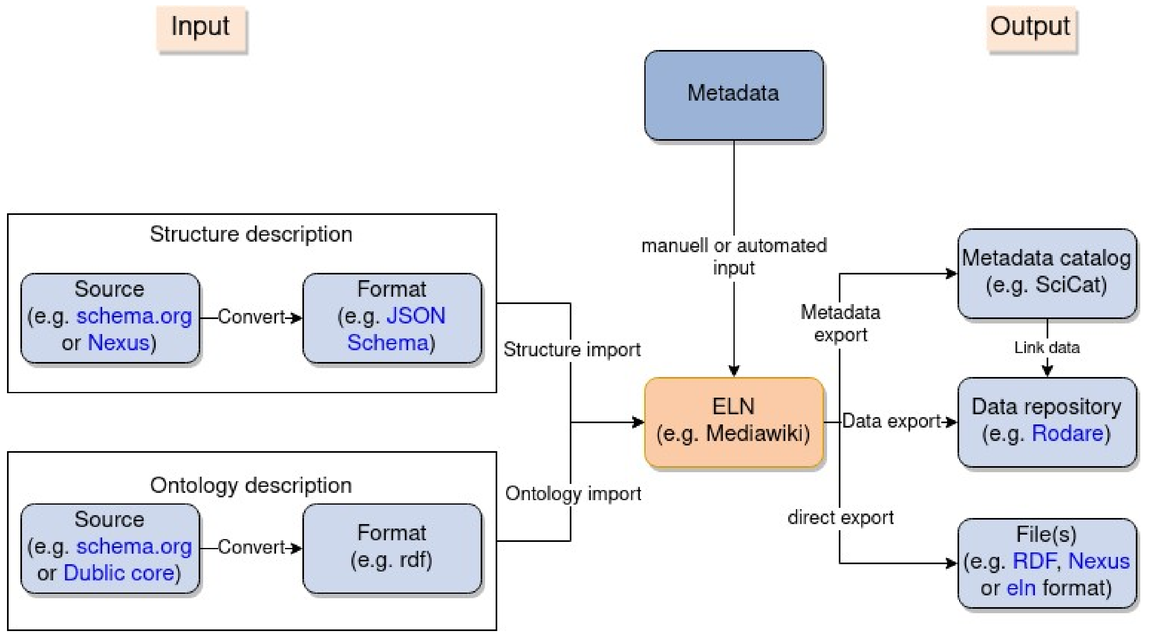
HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Goal - Export functionality with schema origin notation

An essential challenge when creating FAIR datasets is the often underestimated I, which stands for interoperability. Especially for a dataset that is meant to be exported from its ecosystem, it is important to store the metadata and data in an appropriate exchange format based on standards. One possible source for metadata is an electronic lab notebook that stores it in a structured manner. In many cases the internal structure does not match any established metadata schema and a mapping is required for a meaningful export.

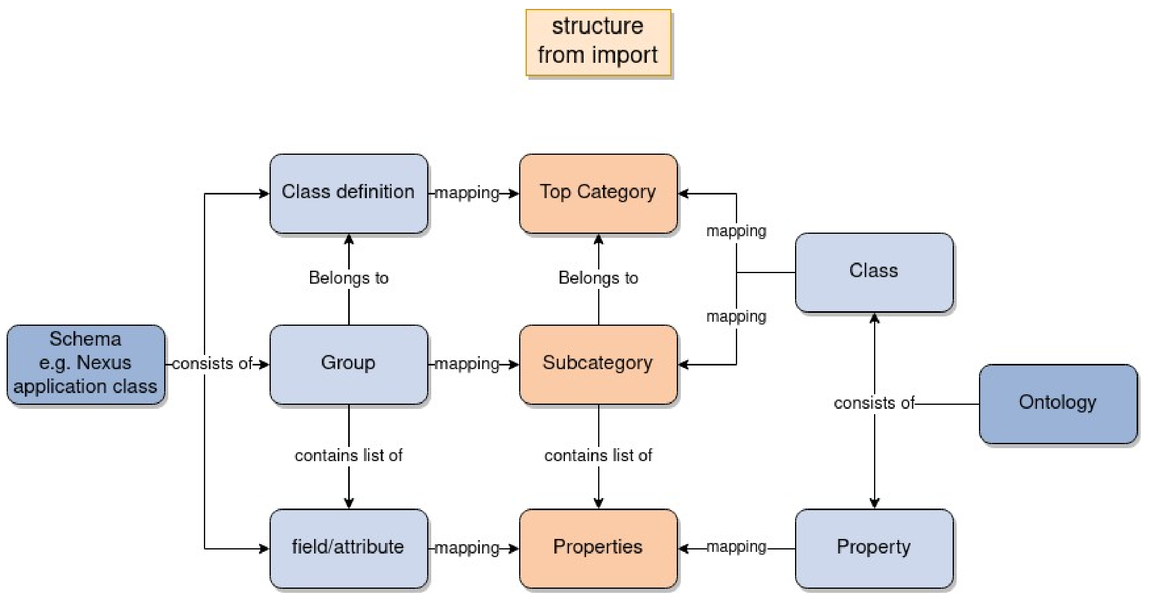
This poster presents the necessary steps for a generic export from an electronic lab notebook (ELN) based on semantic Mediawiki, e.g. for ingestion into SciCat or interoperable Nexus files. It is a flexible and iterative procedure, which makes it perfectly usable for existing documentation where metadata schema are applied at a later stage or need to be updated. In addition, the reference to the original metadata schema is known in the whole pipeline and could be included in the export.

Interactions with the Electronic Lab Notebook (ELN)



1. The source of metadata schema and ontology needs to be imported
2. Local documentation needs to be linked to predefined source
3. The ELN is filled with (meta)data
4. Select the export type for a specific measurement
5. Generic export creates for a pre-defined metadata schema the cor-rect output following the standard and contains the references to the origin

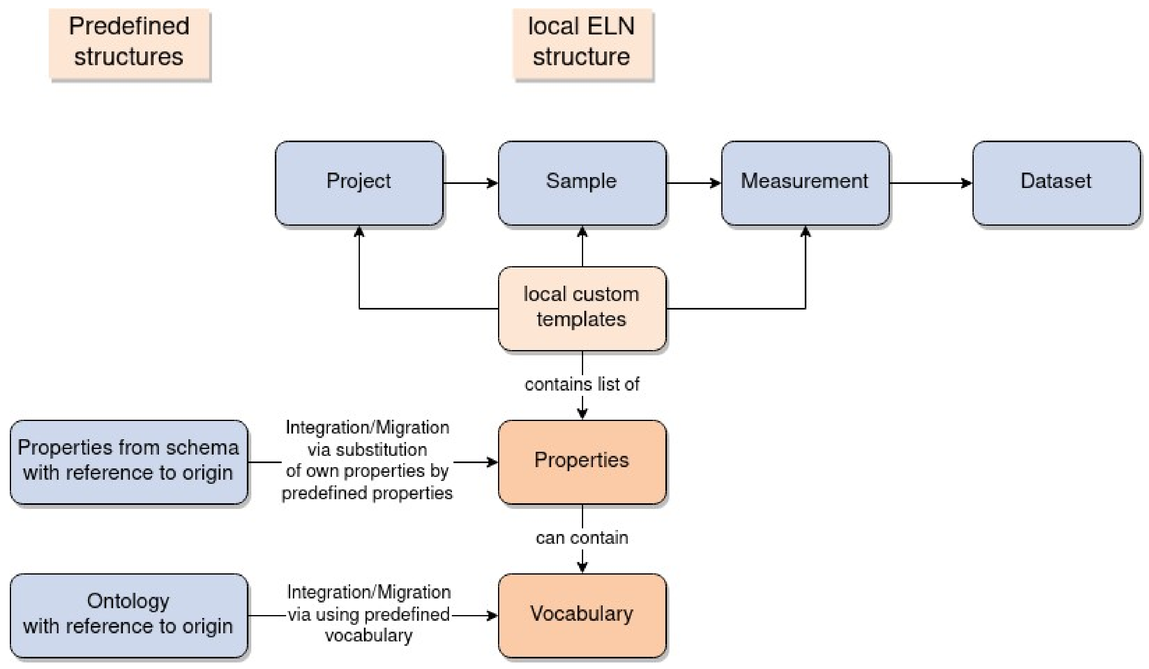
Representing metadata schema within Mediawiki-based Labbooks



- Metadata Schema contains a structure of groups and properties
- Ontology contains classes and terms
- Both needs to be mapped / assigned to ELN specific equivalent entities to represent the same structures and contains references to the origin
- Structure definitions are the bases for structured generic export
- If the reference to origin and structure is on property level, then the existing local structure can be individually defined

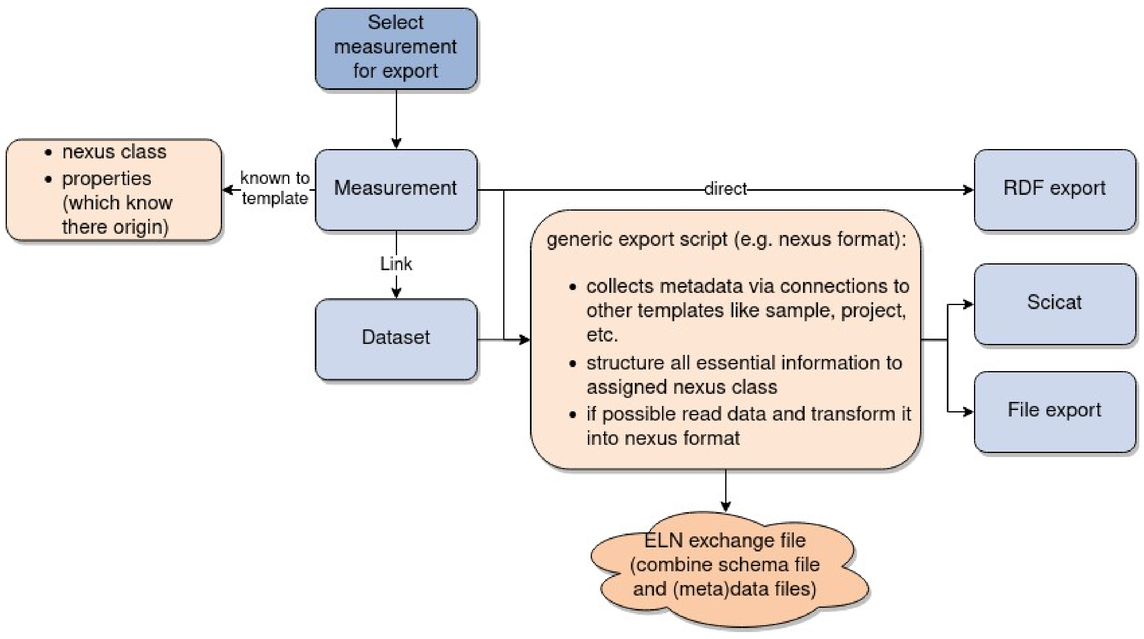
Integration of predefined schema and ontology into existing documentation

- Often the documentation already exists in a unique structure containing e.g. templates to represent different entity type
- Each template contains a list of properties
- Predefined vocabulary lists are assigned on property level
- Integration means substitution of properties and vocabularies with predefined (imported) structures
- Integration process can be done iteratively
- Local ELN and schema development are independent and can be updated any time



Generic export into different formats

- Generic export of an entity collects all properties including its references to certain metadata schema
- The more properties containing references, the richer on linked data is the export
- Once the schema is established for a local measurement it could serve as an “eln” exchange format containing the schema and data file



Key aspects to be considered

- Define metadata schema exchange format: e.g. JSON Schema or JSON-LD
- Conversion from metadata schema source (e.g. Nexus) into exchange format
- Input parser: import exchange file into ELN
- Output parser: export ELN (meta)data content following exchange format
- Property and class definitions within ELN should contain origin of source to provide it for generic export

Outlook

- Define a common metadata schema exchange format
- Test Link-ML to combine metadata schema with ELN specific form definitions
- Test concept with Nexus format in own Mediawiki instance reusing existing tools and concepts
- Metadata schema exchange format could be prototype to extend the “eln” exchange format from the ELN consortium and other initiative (e.g. SciMesh) together with the data file
- It supports the 2 step approach: having individual existing local documentation as first step and the public available exchange format with applied standard

Department of Information Services and Computing · Computational Science Department
Thomas Gruber · t.gruber@hzdr.de · https://www.hzdr.de/fwcc-d

hzdr.de/elabbook

DRESDEN concept
SCIENCE AND INNOVATION CAMPUS