

OPEN
ARTICLE

An ontology-based description of nano computed tomography measurements in electronic laboratory notebooks

F. Kirchner^{1,8}✉, D.C.F. Wieland^{1,8}, S. Irvine², S. Schimek¹, J. Reimers^{1,3}, R. Aversa⁴, A. Boubnov⁵, C. Lucas⁶, S. Flenner², I. Greving², A. Lopes Marinho¹, T. M. Wong^{1,2}, R. Willumeit-Römer¹, C. Eschke¹✉ & B. Zeller-Plumhoff^{1,7}

Scientific communities have recognized the importance of well-documented metadata generated during research. However, ensuring that metadata is findable, accessible, interoperable, and reusable (FAIR) remains a significant challenge. To address this, scientific communities are working towards making metadata available in semantically annotated knowledge graphs using semantic web technologies. In our proposed solution, the creation of a schema is initiated at the very beginning of the scientific process. This is transformed into a data collection platform using the electronic laboratory notebook framework, Herbie, which facilitates the automatic validation and semantic annotation of metadata. Using the example of synchrotron-radiation-based nano-computed-tomography measurements at a beamline, we demonstrate this approach. It effectively captures the complex metadata of such research instruments along with their various configurations, providing a user-friendly experience. We illustrate how Herbie converts all semantic documents into a user interface, ensuring that the entered data automatically meets all FAIR requirements. Additionally, we show how data can be extracted from the knowledge graph using SPARQL queries.

Introduction

When performing experimental measurements, specifically in situ three-dimensional (3D) imaging, a large amount of research data is being generated. Firstly, the actual raw data is produced, which is of main interest to scientists. Secondly, metadata is produced, explaining the measurement's surroundings, such as information about how the devices were set up, or at what point different stages of the experiment took place. It is clear that either part is not meaningful without the other, and it is general consensus in research communities, that all data should be sustainably stored in a FAIR way, *i.e.* findable, accessible, interoperable, and reusable¹. Nevertheless, in particular metadata is typically stored in an unstructured way, often ad hoc using spreadsheets or text documents, or inside a classical notebook. Furthermore, the terms used and level of detail of the recorded data may vary between different experiments.

Electronic laboratory notebooks (ELNs) offer the opportunity to store (meta-)data in a more structured manner and were broadly introduced and studied in the last decade^{2,3}. Higgins *et al.* provided a comprehensive comparison recently⁴, which highlighted that the median lifetime of ELN software packages was only 7 years. The authors suggested that the development of ELNs along user needs, understanding of laboratory culture

¹Institute of Metallic Biomaterials, Helmholtz-Zentrum Hereon, Max-Planck-Straße 1, 21502, Geesthacht, Germany.

²Institute of Materials Physics, Helmholtz-Zentrum Hereon, Max-Planck-Straße 1, 21502, Geesthacht, Germany.

³Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons, Forschungszentrum Jülich GmbH, 52425, Jülich, Germany. ⁴Karlsruhe Institute of Technology, Scientific Computing Center (SCC), Kaiserstraße 12, 76131, Karlsruhe, Germany.

⁵Karlsruhe Institute of Technology, Institute of Nanotechnology (INT), Kaiserstraße 12, 76131, Karlsruhe, Germany. ⁶Bruker Daltonics SPR, Falkenried 88, 20251, Hamburg, Germany. ⁷Data-driven Analysis and Design of Materials, Faculty of Mechanical Engineering and Marine Technologies, University of Rostock, 18059, Rostock, Germany. ⁸These authors contributed equally: F. Kirchner, D.C.F. Wieland. ✉e-mail: fabian.kirchner@hereon.de

fabian.kirchner@hereon.de; catriona.eschke@hereon.de

and ongoing commitment of institutional support were keys to the long-term successful adoption. For material science research specifically, eCAT⁵ and eLabFTW⁶ were two of the most popular frameworks, which were both data-centric ELNs for resource management (*i.e.* text, files and images, etc.). However, as Kanza *et al.* investigated in a recent user study⁷, semantic technologies, such as tagging, advanced semantic search, storing metadata, and linking to ontologies can help the interoperability and usefulness of ELNs. Later in its extended user study⁸, Kanza *et al.* found that simple semantic technology such as tags searching of documents could not enhance scientists' performance due to the personalized behavior of annotation and searching among different areas of scientific research. In the herein described specific scientific use case, the ELN should offer a rigid validation of very specialized data schema, and on the other hand, the stored data should be (at least partially) compatible with other solutions. Existing ELNs were considered, like e.g. FAIRDOME-SEEK, which is a web-platform for collecting heterogeneous research data⁹. Its data model is based on the "Investigation-Study-Assay" framework (ISA)¹⁰ which is suitable for a wide range of use cases by providing just a small set of common data fields. This increases compatibility but forces each use case to adhere to this exact schema, which might not always be fully possible or practical. On the other hand, such a small schema reduces the specificity of the data records. Namely, fields which are just meaningful for a specific beamline setup are not governed by any schema and therefore lack rigid validation and semantic annotation. This hinders possible compatibility with similar setups unless substantial post processing work is done and ultimately does not align with the requirements of the approach.

Hence, building a knowledge-centric framework using well-structured ontologies could be beneficial for ELNs, while there were only limited offers of ELNs using modern ontology technology for knowledge management in scientific research⁸. To the authors' best knowledge, the semantic electronic lab book *Herbie*¹¹ is the first ontology-based ELN framework directly working on a knowledge graph, which see increasing usage in fields where heterogeneous and interdisciplinary data are ever-present¹². It was developed for interdisciplinary material science research in a laboratory environment based on scientific user experience.

In this work, the research question of how to establish FAIR synchrotron radiation-based nano computed tomography (SRnCT) data is addressed. Therefore, the scientific objectives of developing a workflow to define a metadata schema and to store metadata in a semantically annotated manner—while considering the specific constraints of a synchrotron beamline—are discussed. SRnCT is a 3D X-ray imaging technique used in particular to study material and biological systems at resolutions well below 100 nm^{13,14}. *In situ* SRnCT measurements are conducted to investigate material functionality and dynamic behavior. Clearly, FAIR (meta)data management is essential if the data is to be used for follow-up analyses such as in *in silico* model validation. Unlike laboratory devices, synchrotron beamlines are reconfigured every 3–4 days, as this is the typical duration of an experiment granted by the peer-review process of large-scale facilities. Afterwards, new users from laboratories around the world arrive with specific requirements regarding available techniques, such as transmission X-ray microscopy, holotomography, or sample environments integrated into the beamline, all of which must be accommodated. From this perspective, an ELN is required that can capture the complexity of such a beamline and can be extended when new techniques or sample environments are introduced. Furthermore, the metadata schema and ELN must be strictly defined and validated, as users from different laboratories may be present, creating the risk of inconsistent definitions of similar terms if each group establishes its own metadata schema. In¹⁵ a novel experimental flow cell was developed and full-field nanotomographic imaging techniques were applied. The ELN implementation was tested during this beamtime to record the metadata with the goal of finding correlations between experimental results and the setup. Furthermore, this step also improves the repeatability of such experiments as the resulting metadata assets can be readily reused in similar experimental setups.

The workflow is designed as follows: Firstly, a specialized metadata schema for SRnCT experiments was created. This schema covers key metadata required to describe and reproduce the performed experiment. The procedure to establish the schema followed that of a metadata schema for scanning electron microscopy (SEM)¹⁶. The SRnCT metadata schema was then transformed into an application level ontology, which was aligned with the mid-level **PR**ovenance **I**nformation for **MA**terials science ontology (PRIMA) inside the research area (<https://github.com/Materials-Data-Science-and-Informatics/MDMC-NEP-top-level-ontology>). The actual collection of the metadata during the experiment was performed in the ELN *Herbie*. To this end, the ontology was extended by shapes constraint language (SHACL) documents. The SHACL documents were uploaded to *Herbie*, thus automatically creating user friendly web forms ready to be filled out by the performing scientists. This approach ensured that the metadata was captured, validated, and in particular FAIRly stored inside a semantically annotated resource description framework (RDF)-based knowledge graph. This data strategy is in line with well-established community best practices¹⁷. We show how the data entries within the knowledge graph can be transformed into an XML document which adheres to the metadata schema using a SPARQL query. Furthermore the usefulness of the ontology is tested by creating SPARQL queries from scientifically relevant competency questions and running these against the knowledge graph. The applicability of the presented approach to similar experimental setups is discussed.

Methods

Design of metadata schema. The SRnCT schema was developed alongside a number of other schemas for characterization techniques, by the scientists and technicians performing these experiments. The first represented technique schema, for SEM, is well documented within reference¹⁶ and registered in MetaRepo (<https://metarepo.nffa.eu/api/v1/schemas/sem>), the metadata repository of the Nanoscience Foundries and Fine Analysis (NFFA)-EUROPE Pilot (NEP). Other techniques similarly represented in the form of metadata schemas include magnetic resonance imaging (https://metarepo.nffa.eu/api/v1/schemas/mri_schema) and transmission electron microscopy (<https://metarepo.nffa.eu/api/v1/schemas/tem>). Whilst originally intended to serve as a schema for all X-ray nanoCT experiments, it was quickly established that the metadata requirements for measurements

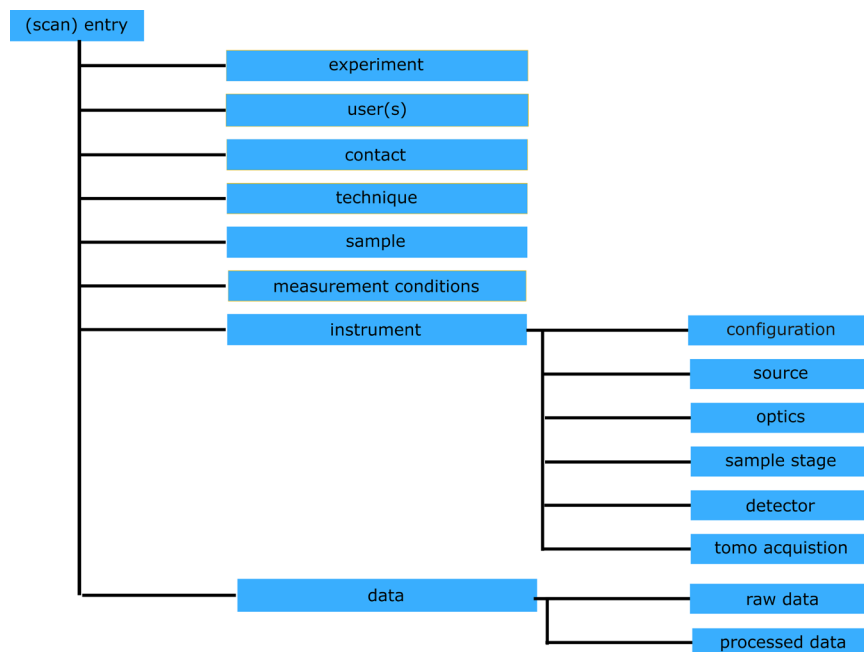


Fig. 1 The hierarchy used in the schema to describe a scan measurement (entry).

conducted at a synchrotron X-ray source are significantly more extensive, and somewhat separate to measurements performed on a commercial laboratory-based tomography setup. One reason for this is that synchrotron beamlines are constructed according to specific demands with respect to their capabilities and the properties of the delivered X-ray beam from the synchrotron. This makes each beamline a unique device with respect to its configuration. For this reason, the proposed schema was then split into two. A lab CT schema (https://metarepo.nffa.eu/api/v1/schemas/lab_ct) was created which may be more generally applied to microCT and nanoCT experiments alike, on commercial instruments within a laboratory setting. The SRnCT schema used as the basis for the project in this paper has been designed specifically for nanoCT experiments performed at a synchrotron source. Whilst it is intended to serve as a basis for other synchrotron beamlines, the schema has some additional details tailored for the specific nanotomographic endstation of the beamline P05¹⁸ at PETRA III at Deutsches Elektronen Synchrotron (DESY, Hamburg), which is operated by Helmholtz-Zentrum Hereon. The schema has been initially implemented in the XML schema format.

Following the general structure described in¹⁶, the primary hierarchy of the template is outlined as shown in Fig. 1. Each information block is here referred to as a *group*. The main groups are described below:

Entry, or Scan Entry: The entry level is the root element of the schema, resembling the NeXus¹⁹ NXentry base class definition. It contains all the metadata describing a single measurement, *i.e.*, a single SRnCT scan.

Experiment: This group represents the overarching information pertaining to the overall experiment which may comprise multiple measurements. At a synchrotron, this is generally referred to as the ‘beamtime’, which is generally linked to a beamtime proposal and assigned a unique identifier by the synchrotron facility. This identifier is also linked to the location of data storage. At many synchrotrons including PETRA III, a JSON file containing the beamtime proposal data together with user data, is created before the beginning of the experiment period. The same information may be duplicated here for each scan entry within the same experiment.

Users: This group represents the contact information of the user(s) responsible for the measurement, together with the indicated role of the user. The metadata properties were selected from the NeXus¹⁹ NXuser group, adopting a similar naming convention. The Primary Investigator is usually the lead research user, while the Applicant is the person who submitted the beamtime proposal. It is typical for multiple users to be present during a beamtime to assist with measurements.

Contact: The local contact is a designated person named from the beamline staff being responsible for helping in the beamtime organization before the experiment and assisting during the beamtime either by helping in performing the experiment or in case of problems.

Technique: This group states the measurement technique, which is in the case for the implemented beamline: TXM - absorption contrast, TXM - Zernike phase contrast, TXM - inline phase contrast or Near Field Holotomography.

Sample: This group describes the sample information, and can include any information describing the sample on which the measurement is performed, similar to the NeXus¹⁹ NXsample group. This could be linked to the sample provenance by including a persistent uniform resource locator (PURL) if available.

Measurement Conditions: This group describes the conditions of the sample environment, which is used for in situ testing, for example the usage of any special cells such as a flow cell, furnace or load frame. The environment itself is included together with the relevant physical parameters such as flow rate, temperature, or pressure.

Prefix	IRI
dash	http://datashapes.org/dash#
dcterms	http://purl.org/dc/terms/
foaf	http://xmlns.com/foaf/0.1/
hash	http://purls.helmholtz-metadaten.de/herbie/hash/#
mbs	http://purls.helmholtz-metadaten.de/herbie/mb/mbs/#
nfdi	http://nfdi.fiz-karlsruhe.de/ontology/
pmd	https://w3id.org/pmd/co/
prima	https://purls.helmholtz-metadaten.de/prima/core#
prima_experiment	https://purls.helmholtz-metadaten.de/prima/experiment#
prov	http://www.w3.org/ns/prov#
qudt	http://qudt.org/schema/qudt/
rdfs	http://www.w3.org/2000/01/rdf-schema#
sh	http://www.w3.org/ns/shacl#

Table 1. Namespace prefix bindings used in the text.

Instrument: The instrument group describes the collection of the components of the beamline. Similar to the NXInstrument group, this group is modular, whereby each component is by itself a group, and described below¹⁹.

Configuration: This group contains information pertaining to the geometric setup of the beamline. For the nanotomography endstation at P05, two general configurations are possible, for the techniques of either Near Field Holography (NFH) or Transmission X-ray Microscopy (TXM)^{18,20,21}. The current version of this template requires a selection of one of the two configurations/techniques. For the groups that follow, this selection then affects which parameters should be included for the Optics group such as geometric parameters including distances of the X-ray lenses or similar.

Source: This group describes the source. The information included in this group may be split into electron source (and corresponding information regarding *e.g.* storage ring current) details, and X-ray source or insertion device (*e.g.* insertion gap) details.

Optics: This group describes the optical components used for the respective configuration, such as Fresnel zone plates.

Sample Stage: This group describes the sample stage motor positions.

Detector: This group describes the physical detector specifications.

Tomo Acquisition: Whilst related to the detector, this group specifically describes the key tomographic acquisition parameters. Also related to the raw **data** format.

Data: The data group describes the output datasets of the measurement. This includes information pertaining to the raw data acquired (data storage location, data format, etc) as well as any processed data which is generated after the necessary steps are performed such as phase retrieval and tomographic reconstruction.

Implementation of schema in Herbie. *Development of application ontology.* For semantic annotation of all metadata of the beamtime experiments, an application ontology was developed based on the PRIMA ontology. Like the metadata schema, this ontology is based on terms from the MDMC-NEP Glossary of Terms²². The ontology was implemented in the web ontology language OWL 2²³. Table 1 lists all prefixes used in the following. IRIs under <http://purls.helmholtz-metadaten.de/herbie/> will be registered with the Persistent Identifiers for Semantic Artifacts service (PIDA, <https://purls.helmholtz-metadaten.de/>) provided by the Helmholtz Metadata Collaboration (HMC). As the majority of the beamline users originate from the material science community, we chose the PRIMA ontology as it was developed there and is well-aligned with the mid-level ontology PMDCo and the top-level ontology PROV²⁴. Other viable choices for top-level ontologies would have been SIO²⁵ or OBI²⁶, which were not considered due to their focus on biomedical imaging.

For each mandatory block within the metadata schema, a matching subclass within the class hierarchy of PRIMA was added. For example, the “entry” block gives rise to the `mbs:ScanEntry` class which is a subclass of the `prima_experiment:Measurement` class of the PRIMA ontology, or the “instrument” block is matched with the `mbs:BeamlineSetup` class, a transitive subclass of the `nfdi:Specification` class which is used in PRIMA as well. Similarly, nested blocks like “tomo acquisition” or “detectors” are mapped to the classes `mbs:TomoAcquisition` or `mbs:DetectorSetup`.

To enable structuring the resulting knowledge graph in line with PRIMA specification and to ensure high re-usability of collected data, content within one metadata block might be distributed among several classes. For example, the “instrument” block specifies properties “instrumentName” and “facilityName”, as well as properties like “configuration” and “detector”. The first two describe properties of the used beamline, which are a separate semantic concept from the “instrument” section covered by `mbs:BeamlineSetup`. Therefore, a class `mbs:Beamline` (subclass of `prima:Instrument`) and a class `mbs:Facility` (subclass of `prima_experiment:Laboratory`) were added.

Data properties such as “pixelSize” within the “imagingDetails” give rise to subclasses of `pmd:Value-Object`, in this case `mbs:ImagePixelSize`. PRIMA specifies that for each such datum the respective

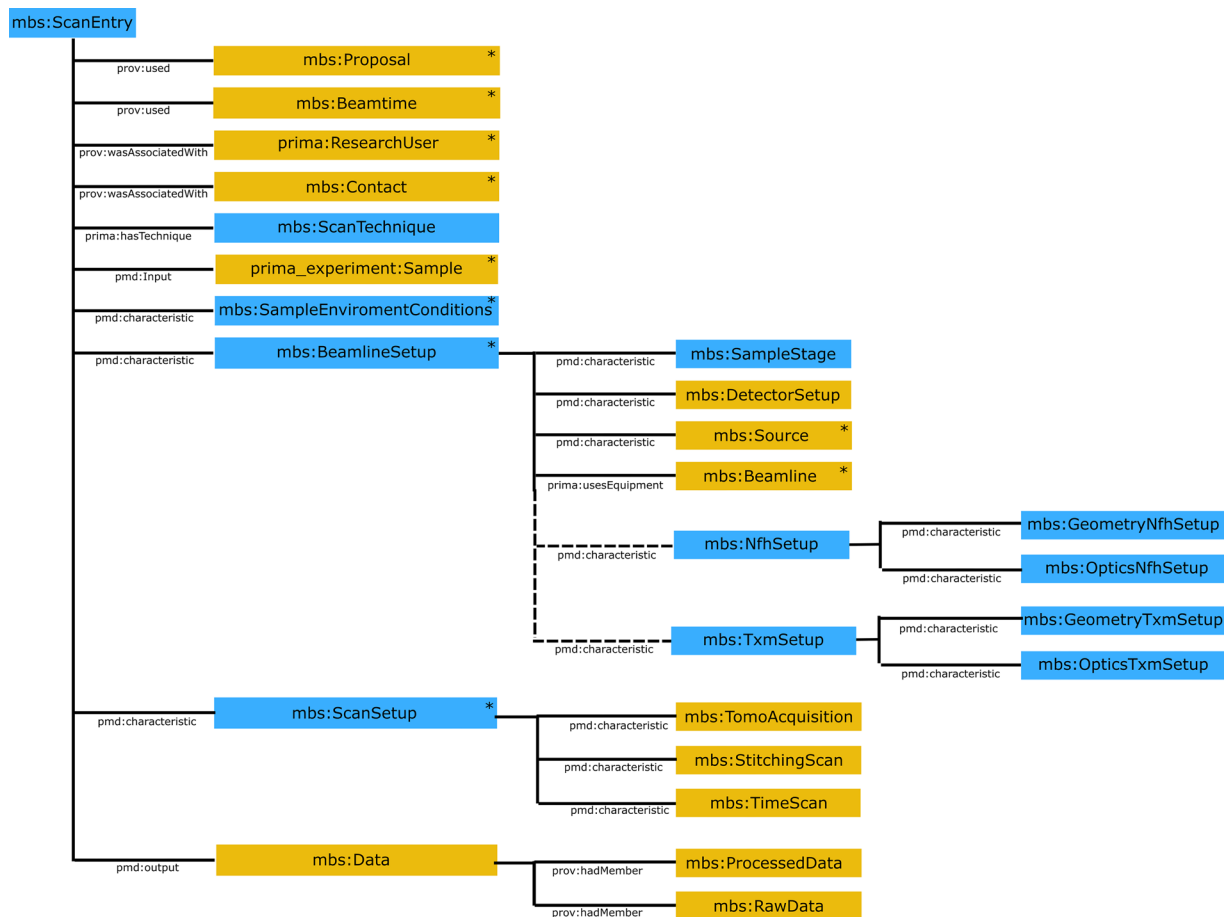


Fig. 2 Hierarchy implemented in the SHACL-shapes to describe the experiment. The classes/shapes denoted by a star are single forms to be filled. Shapes marked in orange can most likely be reused in the development of a logbook for another beamline.

`pmd:ValueObject`-subclass is instantiated and linked to a `qudt:Quantity` instance, which references the actual numerical value alongside its unit.

Implementation of SHACL shapes. All metadata was recorded using the semantic lab notebook Herbie¹¹ whose user interface is configured by the ontology together with SHACL shapes (<https://www.w3.org/TR/shacl/>). Herbie creates usable web forms for each node shape which specifies an `sh:targetClass`. Data entered *via* such a form will be stored within an RDF knowledge graph. Additionally, these SHACL documents can also be used to validate externally provided knowledge graphs.

In order to structure the metadata collection process SHACL shapes were created following the strategy: All metadata blocks were segmented into sections, each of which could be created by the recording scientist. The general hierarchies and connection of the SHACL shapes are depicted in Fig. 2. For each section, a SHACL document was created containing one root SHACL node shape specifying an `sh:targetClass`, e.g. the class `mbs:Beamline`, and containing property shapes for all required and optional parameters, such as its name or facility, as can be seen in Fig. 3.

Nested metadata blocks were either inlined by nesting node shapes inside property shapes with `dash:editor dash:DetailsEditor`. These will be displayed as nested forms by Herbie. Alternatively, only a property shape with `dash:editor dash:InstancesSelectEditor` and referencing the class via `sh:class` was added, if data within these blocks was to be reused among several experiments, as is the case for the `mbs:BeamlineSetup`. Such a property shape will be rendered by Herbie as a dropdown menu or a list of choice chips, depending on the number of selectable instances within the already existing knowledge graph.

Fields for the different parameters were created by including a property shape for the respective `pmd:ValueObject` subclass. Its `sh:maxCount` was set to 1 and the `sh:minCount` to 0 or 1, depending on whether the parameter was optional or not. This information was drawn from the original schema and ultimately decided by the scientists, technicians and data stewards. As the semantic distinction of these parameters was done *via* subclassing `pmd:ValueObject`, the `sh:path` property of the property shape was usually set to the same property, in most cases `pmd:characteristic`. Therefore, the property shapes use `sh:qualifiedValueShape` to qualify all instances at the reused path, and `sh:qualifiedMinCount/sh:qualifiedMaxCount` are used instead of `sh:minCount/sh:maxCount`. At the `sh:qualifiedValueShape` property

```

:BeamlineShape a sh:NodeShape ;
  sh:targetClass mbs:Beamline ;
  hash:documentRoot true ;
  sh:property shared:rdfsLabel ;
  sh:property shared:rdfsComment_multipleLine ;
  sh:property :BeamlineShape_facility ;
.

:BeamlineShape_facility sh:order 2 .

:BeamlineShape_facility a sh:PropertyShape ;
  sh:path prov:atLocation ;
  sh:qualifiedValueShape [
    sh:class mbs:Facility ] ;
  sh:qualifiedMinCount 0 ;
  sh:qualifiedMaxCount 1 ;
  dash:editor dash:InstancesSelectEditor ;
.

```

a

Name
Petra III

Comment

Optional

Facility ▼

Optional

b

Fig. 3 Example of a shape implementation and resulting user interface for generating instances of the `mbs:Beamline` class. In this example, a root node shape is generated specifying a `sh:targetClass`, e.g. `mbs:Beamline`, containing property shapes such as the facility name. (a) Excerpt of SHACL implementation. (b) Resulting web form in Herbie.

a node shape is embedded with `sh:class` set to the subclass of `pmd:ValueObject` corresponding to the parameter. Inside this node shape, the shapes for the actual numerical value and the unit are added. To simplify the development, a set of reusable shapes was extracted, which contain property shapes for the `qudt:value` and `qudt:unit` properties. Despite this rather elaborate setup, each of these top-level property shapes is displayed by Herbie as a simple numerical text input with the respective unit. Figure 4 shows the implementation for the `mbs:ImagePixelSize`, as well as the resulting input field within Herbie. The implementation of the shared `quantityValue_decimal_NanoM` shape was omitted for brevity.

In cases where the set of required parameters varies depending on the class of experiments (configuration of the beamline), separate node shapes were created for each of these sets and then combined via `sh:or` in a top-level node shape. This was done for the `mbs:BeamlineSetup` which has subclasses `mbs:NfhSetup` and `mbs:TxmSetup` and whose instances require different parameters to be recorded. Figure 5 shows the structure of the SHACL document and how in these cases Herbie renders a set of segmented buttons. Thus, the user can select one of the variants and is shown only those fields pertaining to the variant. Note that Herbie uses `sh:or` instead of `sh:xone` to support use cases where one variant is an honest extension of another.

In cases where triples in the resulting data graph should be automatically generated, instances of `sh:SPARQLRule` were included. For example, an `rdfs:label` was auto-generated for each `mbs:User` by concatenating their given and family name as well as their email-address. The concrete implementation can be seen in Fig. 6.

To improve the general understandability of the generated web forms, the `sh:order` and `sh:group` features of SHACL were used to fix the ordering of the input fields as well as grouping them into sections. Moreover, the `sh:name` and `sh:description` properties were used to adjust the label and info text of the

```

:DetectorSetupShape_ImagePixelSize a sh:PropertyShape ;
  sh:path pmd:characteristic ;
  sh:qualifiedValueShape [
    sh:class mbs:ImagePixelSize ;
    sh:property shared:quantityValue__decimal__NanoM ;
  ] ;
sh:qualifiedMinCount 1 ;
sh:qualifiedMaxCount 1 ;
.

```

a

Image pixel size	
23.28	nm

b

Fig. 4 Example of a shape implementation and resulting user interface on the basis of entering the image pixel size property. (a) SHACL implementation via qualified property shape. (b) Resulting input field in Herbie.

```

:BeamlineSetupShape a sh:NodeShape ;
  sh:targetClass mbs:BeamlineSetup ;
  # properties required in all beamline setups
  sh:or ( :NfhSetupShape :TxmSetupShape ) ;
.
:NfhSetupShape a sh:NodeShape ;
  sh:class mbs:NfhSetup ;
  # properties required for the NFH setup
.
:TxmSetupShape a sh:NodeShape ;
  sh:name "TXM"@en ;
  sh:class mbs:TxmSetup ;
  # properties required for the TXM setup
.

```

a

NFH	✓ TXM
-----	-------

b

Fig. 5 Example of a shape implementation and later web form for parameters depending on the setup of the beamline. One of two different configurations `mbs:NfhSetup` and `mbs:TxmSetup` can be selected. (a) SHACL implementation via `sh:or` construct. (b) Resulting segmented buttons in Herbie.

```

:UserShape a sh:NodeShape ;
  sh:rule :UserShape_rdfsLabelRule ;
.
:UserShape_rdfsLabelRule a sh:SPARQLRule ;
  sh:condition [
    sh:property [ sh:path foaf:familyName ; sh:minCount 1 ] ;
    sh:property [ sh:path foaf:givenName ; sh:minCount 1 ] ;
    sh:property [ sh:path foaf:mbox ; sh:minCount 1 ] ;
  ] ;
sh:prefixes foaf:, rdfs: ;
sh:construct """
  CONSTRUCT { $this rdfs:label ?label }
  WHERE {
    $this foaf:familyName ?familyName .
    $this foaf:givenName ?givenName .
    $this foaf:mbox ?mbox .
    BIND(CONCAT(?givenName, " ", ?familyName, " (", ?mbox, ")") AS ?label)
  }
  """ ;
.

```

Fig. 6 SPARQL SHACL rule for automatically generating an `rdfs:label` from other user input.

```

CONSTRUCT {
  ?entry
    ns:entryID ?entryID ;
    ns:startTime ?startTime ;
  .
} WHERE {
  ?entry
    a mbs:ScanEntry ;
    dcterm:identifier ?entryID ;
    prov:startedAtTime ?startTime ;
  .
}

```

Fig. 7 SPARQL construct query for partially transforming Herbie's knowledge graph into a tree resembling the XML schema's structure.

rendered input elements for property shapes covering paths from external ontologies, e.g. `foaf:familyName` or `rdfs:label`. A mapping of the metadata schema to the SHACL shapes is provided in the supplementary information Fig. 9.

Interoperability to other ELNs and connection to metadata schema. Once all data for the beamtime experiment is collected inside a knowledge graph using Herbie, the possibility exists that this data can be exported into an XML document adhering to the metadata schema. This option ensures broader technical and semantical compatibility, as other parties which only adhere to the metadata schema, would be able to make use of the created knowledge graph during the creation of the ELN. This is done by querying the knowledge graph with a SPARQL construct query, which produces an RDF graph that contains all required data for the schema in a tree structure, and then serializing this RDF graph into an XML document which undergoes minor post-processing step removing RDFa-related tags. The SPARQL query is created in the following way: Its CONSTRUCT part is a one-to-one resemblance of the tree structure of the XML schema with variables for each datum. The WHERE clause then maps every datum to the respective part in the knowledge graph. To ensure that only one tree is created for each `mbs:ScanEntry` instance, each node in the CONSTRUCT tree is given a unique IRI, either by binding an appropriate IRI in the knowledge graph, or by generating a new IRI from these. See Fig. 7 for an excerpt of the SPARQL query.

Beamtime experiment. The capabilities and versatility of the ELN was tested in an in situ beamtime at the nanotomography endstation P05 at PETRA III (DESY, Hamburg, Germany). The aim was to investigate the degradation of magnesium-based wires for biomedical applications under physiological conditions. Thus, the experiment was conducted using a custom bioreactor-coupled flow-cell setup optimized for in situ SRnCT imaging. Detailed information on the experimental methodology and environment can be found in¹⁵. Initially, the selected magnesium-based wire is immersed in a flow of ethanol (EtOH) for sterilization and one tomographic scan is obtained to capture the initial sample shape. Subsequently, the medium is changed to simulated body fluid (SBF-JL) according to Bohner *et al.*²⁷ as degradation medium and multiple tomographic scans are recorded over time. For all measurements, a temperature of 37 °C is set, as well as a flow rate of 2 ml/min and a pH of 7.4.

This experiment was selected to evaluate the ELN's performance and usability specifically examining the resulting entry for this experiment. The web browser Mozilla Firefox (Mozilla Corporation, San Francisco, USA, Version 132.0.2, 64-bit) was employed for accessing the Herbie web application. As the application resides on the internal network of the Helmholtz-Zentrum Hereon, secure access from the beamline was ensured *via* a VPN connection established using GlobalProtect (Palo Alto Networks, Santa Clara, USA, Version 6.1.2-83). When accessing the platform, a dedicated workspace was created specifically for the beamtime, enabling systematic testing and application of the developed ontology. The workspaces are only used for overview purposes. Data and forms can be exchanged and shared between the areas at any time.

To record the measurement in the ELN, the web forms generated from the SHACL shapes were filled in, beginning with the top-level "Scan entry", as illustrated in Figure 9. The hierarchical levels were sequentially completed. Missing instances were created in a step-by-step manner along with the respective sub-forms. Once the required resources were created, they appeared in the top-level entry form. As an example, in Figure 9, the "beamtime" is initially a missing instance. By filling out this instance with its "Beamtime ID" and creating its sub-instance "proposal", the "beamtime" instance can be submitted. Since we have already created the "proposal" instance for the submission of the "beamtime", the proposal can also be selected in the top-level instance (Fig. 9).

The generated instances and associated parameters are designed to remain permanent within the workspace. Thus, instances within classes such as "beamtime" remain unchanged for reuse in subsequent scans. This feature significantly reduces the workload for future experiments by minimizing repetitive data entry. However, if parameters change over the course of experiments or beamtimes, new instances and sub forms can be generated accordingly.

Application of competency questions. Finally, the developed application ontology was tested by translating each of the competency questions shown in Table 2 into a SPARQL select query and running this query against the knowledge graph which was generated during the beamtime experiment. The questions were used

	Competency question
1	What were the monochromator energy, flow rate, and system temperature during the measurement?
2	Which medium was used for the degradation?
3	What flow rate was applied?
4	Where is the raw data of the scan stored?
5	Which measurements were performed using a degradation cell and a bioreactor?
6	What FZP distance was used for the FZP “QP040B.01”?
7	How often was the TXM or NFH technique applied?
8	What were the magnification and pixel size of the TXM or NFH experiments?
9	What was the “sample_in_position”?

Table 2. Competency questions for testing the ontology.

to test the ELN but can also serve to communicate data extraction perspectives in a human readable way. To show case the versatility and utility we have collected questions pointing either to the experimental setting of the beamline (e.g. Q1), the sample environment (e.g. Q2) or experimental overarching questions like the frequency of measurement types (e.g. Q7).

Results

A finalized ELN entry for a nanoCT scan of the magnesium-based wire immersed in EtOH is presented in Figure 10 and the corresponding knowledge graph in Figure 8. Figure 10a illustrates the top-level instance “Scan entry” along with its subordinated instances and properties. This overview highlights the hierarchical organization of metadata fields and their corresponding sub-classes. By following the hierarchical structure, sub-level instances, such as the “Beamline setup” can be accessed for a more detailed examination of specific properties e.g. “Image pixel size”, as shown in Fig. 10b. These sub-level instances can be inspected either by following the hyperlink in Figure 10a or by searching for the instance in the classes overview menu on the left side of Fig. 10a. This allows for a comprehensive inspection of all sub-levels and properties.

During the testing process, the system exhibited stable performance with no major technical issues observed. The testing was performed by scientists and technicians who have been partially involved in the schema implementation. It was reported to be very easy to use, as it only required filling in pre-defined forms, which were well aligned with the scientific process. In particular, as all the semantic elaboration had been offloaded into the SHACL shapes, users could focus on entering just the parameters. Also splitting the data among entries which could be entered before the experiment and entries which were entered during the performance proved useful. While the ELN proved to be effective for the experienced users that participated in development, further (qualitative) usability testing is necessary to assess its accessibility for a broader range of users, particularly those unfamiliar with the system. To enhance user-friendliness, the integration of automated selection for instrument-dependent parameters, such as “x pixels” which depend on the selected detector, can be generated automatically. This minimizes manual input and improves the system’s adaptability to various experimental setups.

The evaluation and knowledge extraction of parameters is performed via competency questions that probe the knowledge graph. Figure 11 and Table 3, and Figures S1, S2, and S3 in the supplementary information, show the queries and results for some example competency questions. These questions were selected as they assess different parts of the experiment. The manner in which the data was structured within Herbie enables the extraction of deeply nested parameters without knowing the full tree structure, which simplifies later reuse. This nesting was performed with only a few properties, e.g. `pmd:input` or `pmd:characteristic` and the semantic distinction of parameters was then achieved by having a subclass for each. So, for example in Figure 11, the system temperature of the `mbs:ScanEntry` is retrieved by querying the property paths `pmd:characteristic+` for an instance of `mbs:SystemTemperature`.

After the successful extraction of temperature and flow rate for example (question 1), further correlations can be established by connecting them to the results of the analysis and interpret differences between measurements. Similarly, the extracted parameters can be used for beamline alignment. A beamline setup is always slightly different due to the inherent complexity of such instruments: it consists of many components including for example monochromators, slits, optics, sample stages. All of them need to be aligned with sub-micrometer precision to achieve nanometer resolution, thus, even thermal drifts may change parameters. In principle, the X-ray coming from the synchrotron might be delivered at different positions, or the reconfiguration of an optical element might change parameters. Thus, competency questions like the question about the usage of the beamline, e.g. the magnification and pixel size (question 8), would enable the identification of settings where similar configurations were already used.

The time needed to set up all semantic documents for the performed experiments highly depends on the familiarity of the scientists with the used semantic technologies RDF, OWL and SHACL. Similarly, if expert programmers were to implement the semantic documents, their understanding of the experimental workflow would be integral to the usability of the product. To provide a more objective overview of the required effort the number of classes, properties, node shapes and property shapes that had to be defined in order to facilitate the experiments can be quantified.

The decision to use PRIMA was made to achieve maximum overlap with other in-house developments and cooperation partners. As PRIMA itself is based on the PMD core ontology as well as the ubiquitous PROV

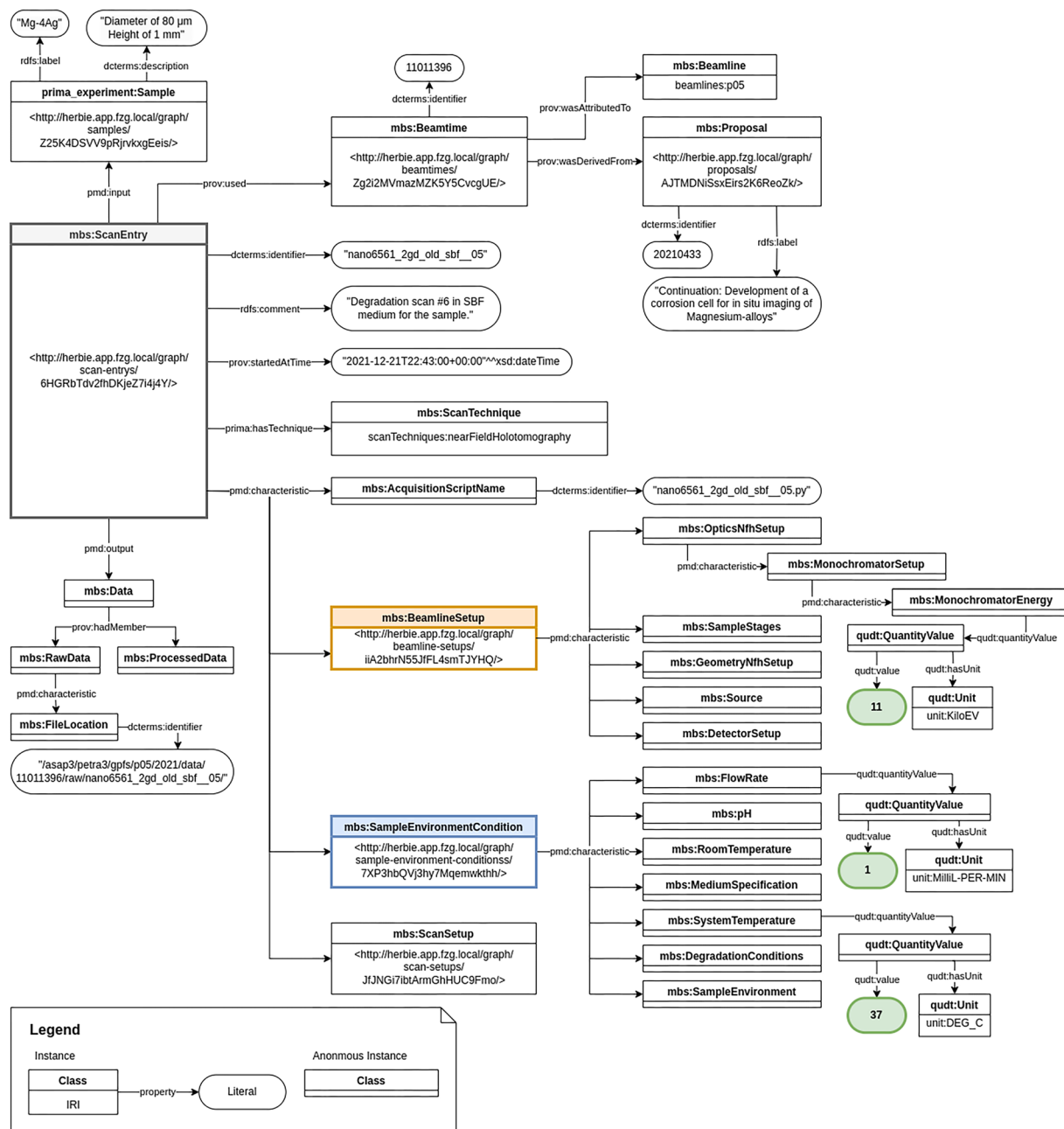


Fig. 8 The corresponding knowledge graph of a measurement at the P05 showing the scanEntry along with specific entries for the measurement conditions and beamline setup.

ontology, a high-level of interoperability is achieved. In the presented use case 146 new OWL classes were defined, see Table 4 for a count with respect to their parent class. The majority were subclasses of `pmd:ValueObject` (71), *i.e.* classes corresponding to individual numerical or textual data values. And although each of these requires some elaboration in the SHACL implementation, they would be highly reusable when modeling a related experimental setup where *e.g.* used equipment has the same configuration parameters. Sub classes of `prima:Instrument` (12), `prima:System` (9), as well as `prima:Equipment` (7) are also good candidates for reuse.

All SHACL documents contain a total of 159 node shapes of which 58 are actual named node shapes. Of these, 25 are node shapes specifying a `sh:targetClass`, and hence will be picked up by Herbie as an independent web form, see supplementary information Table 1. The majority of these root node shapes are for subclasses of `prima:Equipment` (10), `prima:System` (5), and `prima:Instrument` (3), and therefore reusable in a similar experimental setup. Finally, 185 property shapes were defined with a `sh:path` distributed as can be seen in supplementary information Table 2.

Top-level instance with no available instances for required properties

Creating the instances for required properties

Top-level instance with available instances for required properties

Submitting the beamtime instance with the required properties

Fig. 9 Schematic of the workflow using the ELN at the beamtime. Initially the the top level instance “Scan entry” had no required properties. By creating and submitting the missing instances, the necessary properties and instances can now be selected to complete the top-level instance. This process reduces the amount of necessary context switches for the user.

Discussion

Creating the metadata schema which serves as the foundation to describe the information needed from an experiment, was possible for the scientists themselves, as this step did not require experience with ontologies and resulted in a tree structure which was close to how metadata would have otherwise been stored, *e.g.* in a spreadsheet. But the schema differs from the realized structure within the ELN. By looking at Figure 9 in the supplementary information, which show the structure of the hierarchy of the metadata schemata and SHACL shapes, it is obvious that both trees mostly overlap but have differences. These differences are a result of specific considerations during the implementation and adoption of special properties of the beamline P05 or usability. For instance, the information on the data acquisition is a child of the class “Instrument” in the metadata schemata and was moved to the top level in the ELN implementation. This decision was made as during a beamtime,

Top-level instance “Scan entry” with all instances and required

Sub-level instance “Beamline Setup” with all instances and required properties

Sub-level instance “measurement conditions” with all instances and required properties

Fig. 10 Final ELN entry for a nanoCT scan. a illustrates the top-level instance “Scan entry” with its sub-instances and the class selection menu on the left. b and c show the sub-level instances “Beamline setup” and “Measurement conditions”, respectively, with their required properties, which can be accessed via the hyperlink or the selection menu in a. Moreover, the addressed competency questions (c.f. Table 2) are highlighted in green boxes.

the setup of the beamline usually does not change, whereas parameters on the scan, depending on the sample, might change. Due to this a user would be required only to create a new instance of `prima_core:DataAcquisition` rather than generating a whole instance of `mbs:BeamlineSetup` again. In the experiments, filling out the complete beamline configuration took about 43 minutes. However, once this step is completed and only the individual measurements performed during the beamtime need to be entered, the required time

```

SELECT ?entry ?energy ?energy_unit ?flow_rate ?flow_rate_unit ?temperature ?temperature_unit
WHERE {
  ?entry a mbs:ScanEntry ;
  pmd:characteristic+ [ a mbs:MonochromatorEnergy ;
    qudt:quantityValue [
      qudt:value ?energy ;
      qudt:hasUnit / rdfs:label ?energy_unit ] ] ;
  pmd:characteristic+ [ a mbs:FlowRate ;
    qudt:quantityValue [
      qudt:value ?flow_rate ;
      qudt:hasUnit / rdfs:label ?flow_rate_unit ] ] ;
  pmd:characteristic+ [ a mbs:SystemTemperature ;
    qudt:quantityValue [
      qudt:value ?temperature ;
      qudt:hasUnit / rdfs:label ?temperature_unit ] ] .
}

```

Fig. 11 SPARQL query to answer the competency question 1: What were the monochromator energy, flow rate, and system temperature during the measurement?

entry	energy	energy_unit	flow_rate	flow_rate_unit	temperature	temperature_unit
<http://...ZFz/>	11	"keV"	1	"mL/min"	37	"°C"
<http://...BsC/>	11	"keV"	1	"mL/min"	37	"°C"
<http://...j4Y/>	11	"keV"	1	"mL/min"	37	"°C"
<http://...8xp/>	11	"keV"	1	"mL/min"	37	"°C"

Table 3. Results of the SPARQL query to answer the competency question 1: What were the monochromator energy, flow rate, and system temperature during the measurement?

External class	No. of child classes	Example
pmd:ValueObject	71	mbs:MonochromatorEnergy
http://nfdi.fiz-karlsruhe.de/ontology/Specification	27	mbs:MonochromatorType
prima:Instrument	12	mbs:Beamline
prima:Setting	10	mbs:MagnificationFactor
prima:System	7	mbs:SampleStage
prima:Equipment	7	mbs:Detector
prima:Project	2	mbs:Proposal
prima_dataset:RawData	2	mbs:MotorsLogFile
pmd:Object	2	mbs:StorageRing
prov:Agent	1	mbs>Contact
prima:Technique	1	mbs:ScanTechnique
prima_dataset:Dataset	1	mbs:Data
prima_dataset:ProcessedData	1	mbs:ProcessedData
prima_dataset:ReferenceData	1	mbs:ReferenceImages
prima_experiment:Laboratory	1	mbs:Facility
prima_experiment:Measurement	1	mbs:ScanEntry

Table 4. Number of classes in the application ontology for each external class from the mid- and top-level ontologies.

is reduced to approximately 3 minutes per sample, as the beamline configuration is automatically linked to the sample metadata. Nevertheless, each structure may be projected on the other, as shown above, and offers the degree of freedom needed for a successful and efficient integration. It would be possible to create further mappings of the original metadata schema to other ontologies to facilitate other scientific groups' needs. The XML schema could then serve as a shared data transformation coupling.

In this work, the presented metadata schema can be extended to include many additional optional parameters, owing to the modular structure design. To adapt the used metadata setup to similar experiments, the ontology would require extension with the required classes and a corresponding extension of the set of SHACL documents would be necessary.

The schema and implementation could for example be generalized to be applicable for synchrotron radiation-based microCT (SR μ CT) measurements. Usually, SR μ CT measurements require fewer optical components, thus, a smaller template may be sufficient. To make the schema applicable for SRnCT measurements at different synchrotrons, further customized input may be required. Depending on the exact beamline layout different optics and other hardware components are used. Of course, a more generalized approach is feasible which would result in less customized entries and layout. However, this would also require less restricted entries and could lead to undefined string formatting of most entries. This data diversity would be a significant disadvantage for automated downstream analysis, for example Jalali *et al.*²⁸ discussed that a high standard of data quality was essential to train a materials science specific large language model.

Depending on the exact overlap with the original setup a lot of code may be reused. Node shapes for the `qudt:Quantity`-instances can be directly reused if already present for the specific units or shapes for generating common properties to generate instances of `mbs:Proposal` or `mbs:Detector`. More specialized subclasses of classes already covered by a SHACL document which do not require any additional properties in the knowledge graph, can be included by adding the subclass to the ontology and a `sh:or` construct to the existing node shape. This then makes it possible for the user to select the more specific class in Herbie. An example is the `mbs:BeamlineSetup` which can be either an `mbs:NfhSetup` or an `mbs:TxmSetup`, both setups require similar parameters but also have their individual requirements. Figure 5 shows how they are combined in one SHACL document using `sh:or`. Figure 2 also indicates those SHACL shapes with orange which can be reused immediately as they are for collecting the metadata at another beamline. Clearly, the major part of the code might be reused and expanding the already existing shape would consequently require significantly less work.

Extracting reusable parts of the schema into their own SHACL documents has the advantage that elements like specific *e.g.* optical elements can be reused among several scans even from different beamtimes and have not to be entered a second time. Also, this decreases the risk of different notations and increases reproducibility and findability within the ELN. Additionally, if a separate SHACL document for the required class exists, Herbie renders a button next to the dropdown/choice chips to let the user conveniently create an instance with new parameters for the specific class.

When adapting or extending the setup, the portions of the SPARQL construct query for transforming the knowledge graph into an XML document can be reused as well. The post-processing steps do not have to be adjusted. It is worth mentioning, that the SPARQL query – like every data transformation pipeline – potentially has to be adjusted if the XML schema, the ontology or the SHACL documents change.

In order to achieve FAIRness of metadata it is necessary to specify its semantics and validate its structure. Popular ELNs typically allow to immediately upload any not rigorously structured document. In these setups, usually, another framework is required to either extract well-structured and semantically annotated data or to run additional checks on the entered data, which validate conformance to some defined schema. So the workload of making all data FAIR is done in a post processing step. In our approach, we follow the concept of constructing an ontology and structuring the metadata up front before the ELN can be used. To ensure that metadata is findable, each dataset and resource within the ELN is assigned a globally unique Internationalized Resource Identifier (IRI). This persistent identifier guarantees unambiguous referencing and facilitates efficient indexing. Furthermore, every metadata record includes the corresponding beamtime identifier, enabling direct association with experimental sessions. The ELN provides search functionality that allows users to locate metadata through intuitive queries. In addition, SPARQL will be available for advanced semantic queries, supporting precise retrieval of information across the knowledge graph. Accessibility is achieved through consistent use of standard web technologies and semantic protocols. Metadata is stored as an RDF-based knowledge graph, which can be retrieved in all standard serialization formats and accessed via SPARQL queries. Continuous online availability of the metadata is planned to ensure persistent access, although this functionality is currently under discussion. Interoperability is a central design principle of the ELN. All metadata is rigorously semantically annotated using a formally defined ontology developed in collaboration with the materials science community. This guarantees semantic consistency and compliance with established community standards. Consistent use of RDF representations further enables a potential data exchange between different platforms and tools. Reusability is ensured through adherence to semantic guidelines and community standards. Each metadata record is explicitly linked to the beamline numbers, providing clear provenance information. Furthermore, the ELN keeps track of all changes made to the metadata set. The ontology and schema design inherently support rich metadata descriptions, enabling downstream applications such as data mining, machine learning, and reproducibility studies. As a result, our approach leads to a very accessible ELN, which, out of the box, produces FAIR data, offers advanced semantic search capabilities, and automatically links all its data to ontologies.

In conclusion, the presented ELN forms are able to collect the main features of a nanoCT experiment at the nanotomography endstation of the beamline P05 and to collect all metadata needed to describe the experiments. Because of the structured and ontology-based approach, the metadata is completely semantically annotated. As such, the ELN also represents a valuable option to consider when a data management plan (DMP) is set up. As a data management plan defines how the primary, evaluated and metadata is stored, the herein described tool provides a toolbox following the FAIR principles and, thus, complies with the demand of a DMP.

In the future, different improvements are desired to ensure longterm use of the ELN forms. Currently, all data regarding the beamline information, such as motor positions, have to be filled in manually. However, many of the information are already logged in beamline specific files (either as ASCII file or as a NEXUS file) as they are required to describe the experiment. In future we are aiming to extract such information automatically so that the user needs to provide information on the sample and specific environments only. To do so, dedicated auxiliary scripts are needed to extract the information from log files depending on the file types and their structure. Competency questions formulated via natural language can be transformed into SPARQL queries, where

the defined ontology acts as a dictionary, and the SPARQL query can search the knowledge graph accordingly, without limitations in the depth of the knowledge graph. This can facilitate the extraction of comprehensive data from the ELN and might be subsequently used for analysis by machine learning to find e.g. relationships between specific parameters for optimization. However, translating these questions into SPARQL queries is still a task for experts, and thus a bottleneck. This might change in the future as the advancement of LLM agents might be a chance to leverage competency questions without being an ontology expert or fluent in query languages.

Due to the versatility of Herbie and the possibility to already adapt Herbie for the initial stage during e.g. sample production, it is possible to comprehensively map the sample provenance. Additionally, future developments should focus on including the whole image processing workflow from phase retrieval, tomographic reconstruction, application of filters, segmentation and quantification within the ELN in an ontology-based manner. In doing so, a feedback loop between sample manufacturing and functionality can be established.

As pointed out the parts of the SHACL shapes and ontology can be directly reused for other beamlines at the PETRA III storage ring. An RO-Crate containing the ontology and all SHACL shapes is available at²⁹. This can be imported into a Herbie instance to obtain the same set of web forms for collecting data in similar experiments¹¹. Naturally, the ongoing development should be kept up to date with other metadata management efforts such as the DAPHNE4NFDI (Data from PHoton and Neutron Experiments for NFDI) project³⁰. Importantly, interoperability between different implementations and ELN solutions is required and could be done using a tool like ELNdataBridge³¹, and an option for export/import of XML documents should be added. Moreover, mapping to data catalogues such as SciCat (<https://www.scicatproject.org/#documentation>) may be envisioned. Finally, to facilitate and accelerate SHACL shape creation, a graphical user interface may be introduced for this purpose.

Data availability

All metadata describing the NanoCT experiment to populate the ELN and test the competency questions is available for download at²⁹ in `nanoct-v2.zip`, under the path `/data/input/knowledge_graph.ttl`. The Herbie framework can be downloaded at¹¹.

Code availability

The ontology (`/src/mbs/1.0.0.ttl`) which defines the used terms, the SHACL documents that define the shapes of the form in the ELN (`/src/mbs/shapes`), and SPARQL queries (`/src/queries/competency_questions`) for all competency questions, and the RO-Crate for setting up a Herbie instance (`mbs.zip`) can be download for direct use at²⁹.

Received: 24 April 2025; Accepted: 9 March 2026;

Published online: 19 March 2026

References

- Wilkinson, M. D. *et al.* The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* **3**, 160018, <https://doi.org/10.1038/sdata.2016.18> (2016).
- Rubacha, M., Rattan, A. K. & Hosselet, S. C. A review of electronic laboratory notebooks available in the market today. *JALA: Journal of the Association for Laboratory Automation* **16**, 90–98, <https://doi.org/10.1016/j.jala.2009.01.002> (2011).
- Bird, C. L., Willoughby, C. & Frey, J. G. Laboratory notebooks in the digital era: the role of ELNs in record keeping for chemistry and other sciences. *Chemical Society Reviews* **42**, 8157–8175, <https://doi.org/10.1039/c3cs60122f> (2013).
- Higgins, S. G., Nogiwa-Valdez, A. A. & Stevens, M. M. Considerations for implementing electronic laboratory notebooks in an academic research environment. *Nature Protocols* **17**, 179–189, <https://doi.org/10.1038/s41596-021-00645-8> (2022).
- Goddard, N. H., Macneil, R. & Ritchie, J. eCAT: online electronic lab notebook for scientific research. *Automated Experimentation* **1**, 1–7, <https://doi.org/10.1186/1759-4499-1-40> (2009).
- Carpi, N., Minges, A. & Piel, M. eLabFTW: An open source laboratory notebook for research labs. *J. Open Source Softw.* **2**, 146, <https://doi.org/10.21105/joss.00146> (2017).
- Kanza, S. *et al.* Electronic lab notebooks: can they replace paper? *Journal of Cheminformatics* **9**, 1–15, <https://doi.org/10.1186/s13321-017-0221-3> (2017).
- Kanza, S., Gibbins, N. & Frey, J. G. Too many tags spoil the metadata: investigating the knowledge management of scientific research with semantic web technologies. *Journal of Cheminformatics* **11**, 23, <https://doi.org/10.1186/s13321-019-0345-8> (2019).
- Wolstencroft, K. *et al.* SEEK: a systems biology data and model management platform. *BMC Systems Biology* **9** <https://doi.org/10.1186/s12918-015-0174-y> (2015).
- Sansone, S.-A. *et al.* Toward interoperable bioscience data. *Nature Genetics* **44**, 121–126, <https://doi.org/10.1038/ng.1054> (2012).
- Kirchner, F. *et al.* Herbie - The Semantic Laboratory Notebook & Research Database <https://doi.org/10.5281/zenodo.1825444> (2026).
- Hogan, A. *et al.* Knowledge Graphs <https://doi.org/10.1007/978-3-031-01918-0> (Springer International Publishing, Cham, 2022).
- Claro, P. I. C. *et al.* From micro- to nano- and time-resolved x-ray computed tomography: Bio-based applications, synchrotron capabilities, and data-driven processing. *Applied Physics Reviews* **10**, 021302, <https://doi.org/10.1063/5.0129324> (2023).
- Zeller-Plumhoff, B. *et al.* Utilizing Synchrotron Radiation for the Characterization of Biodegradable Magnesium Alloys-From Alloy Development to the Application as Implant Material. *Advanced Engineering Materials* **23**, 2100197, <https://doi.org/10.1002/adem.202100197> (2021).
- Reimers, J. *et al.* Development of a Bioreactor-Coupled Flow-Cell Setup for 3D In Situ Nanotomography of Mg Alloy Biodegradation. *ACS Applied Materials & Interfaces* **15**, 35600–35610, <https://doi.org/10.1021/acsami.3c04054> (2023).
- Joseph, R. *et al.* Metadata schema to support FAIR data in scanning electron microscopy in Supplementary Proceedings of the XXIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2021): Moscow, Russia, October 26–29, 2021. Ed.: A. Pozanenko 265 <https://doi.org/10.5445/IR/1000141604> (2021).
- Jacobsen, A. *et al.* FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence* **2**, 10–29, https://doi.org/10.1162/dint_r_00024 (2020).
- Fleener, S. *et al.* Pushing the temporal resolution in absorption and Zernike phase contrast nanotomography: Enabling fast in situ experiments. *Journal of Synchrotron Radiation* **27**, 1339–1346, <https://doi.org/10.1107/S1600577520007407> (2020).
- Könnecke, M. *et al.* The NeXus data format. *Journal of Applied Crystallography* **48**, 301–305, <https://doi.org/10.1107/S1600576714027575> (2015).

20. Flenner, S. *et al.* Hard X-ray nano-holography with a Fresnel zone plate **28**, 37514–37525 <https://doi.org/10.1364/OE.406074> (2020).
21. Flenner, S. *et al.* Hard x-ray nanotomography at the P05 imaging beamline at PETRA III in Developments in X-Ray Tomography XIV **12242**(SPIE, 2022), 122420L <https://doi.org/10.1117/12.2632706>.
22. Aversa, R. *et al.* The MDMC-NEP Glossary of Terms <https://doi.org/10.5281/zenodo.10663833> (2024).
23. W3C OWL Working Group. OWL 2 Web Ontology Language Document Overview (Second Edition). W3C Recommendation 11 December 2012. <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/> (2012).
24. Lebo, T. *et al.* PROV-O: The PROV Ontology. W3C Recommendation 30 April 2013 <http://www.w3.org/TR/2013/REC-prov-o-20130430/> (2013).
25. Dumontier, M. *et al.* The SemanticScience Integrated Ontology (SIO) for biomedical research and knowledge discovery. *Journal of Biomedical Semantics* **5**, 14, <https://doi.org/10.1186/2041-1480-5-14> (2014).
26. Bandrowski, A. *et al.* The ontology for Biomedical Investigations. *PLoS One* **11**, e0154556, <https://doi.org/10.1371/journal.pone.0154556> (2016).
27. Bohner, M. & Lemaitre, J. Can bioactivity be tested in vitro with SBF solution? **30**. issn: 0142-9612. <https://doi.org/10.1016/j.biomaterials.2009.01.008> (2009).
28. Jalali, M. *et al.* Large language models in electronic laboratory notebooks: Transforming materials science research workflows. *Materials Today Communications* **40**, 109801, <https://doi.org/10.1016/j.mtcomm.2024.109801> (2024).
29. Kirchner, F. *et al.* Accompanying data and software for “An ontology- based description of nano computed tomography measurements in electronic laboratory notebooks” Version v2. <https://doi.org/10.5281/zenodo.18253604> (2026).
30. Barty, A. *et al.* DAPHNE4NFDI - Consortium Proposal June 2023. <https://doi.org/10.5281/zenodo.8040606>.
31. Starman, M. *et al.* ELNdataBridge: Facilitating Data Exchange and Collaboration by Linking Electronic Lab Notebooks via API. *Journal of Cheminformatics* **17**, <https://doi.org/10.1186/s13321-025-01024-1> (2025).

Acknowledgements

The SRnCT schema presented in this work was initially conceived as an additional technique represented within the collection of metadata schemas created by the MDMC-NEP Metadata Working Group. This is a collaboration between two projects: the Joint Lab “Integrated Model and Data-Driven Materials Characterization” (JL MDMC) and the Nanoscience Foundries and Fine Analysis (NFFA)-EUROPE Pilot (NEP). The authors thank Dr. Reetu Joseph for her work in the creation of the metadata schema. The authors acknowledge the following funding: the Joint Laboratory “Model and Data-driven Materials Characterization” (JL MDMC), a cross-centre platform of the Helmholtz Association, the Helmholtz Imaging project SmartPhase (ZT-I-PF-4-027), the Helmholtz Metadata Cooperation projects MetaSurf (ZT-I-PF-3-076), and MetaCook (ZT-I-PF-3-046), Bundesministerium für Bildung und Forschung (BMBF) ErUM-Data project “KI4D4E: Ein KI-basiertes Framework für die Visualisierung und Auswertung der massiven Datenmengen der 4D-Tomographie für Endanwender von Beamlines” (Förderkennzeichen 05D23CG1), Deutsche Forschungsgemeinschaft (DFG) Project with the number 526239533 and the EU’s H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe PILOT.

Author contributions

S.I., R.A., A.B., C.L., S.F. and I.G. developed the metadata schema. F.K., D.C.F.W., S.S., J.R., A.L.M., T.M.W., C.E., B.Z.P. developed the ontology and implemented the SHACL Shapes. S.S., J.R., S.F., I.G. and A.L.M. documented all metadata during the beamtime. F.K., D.C.F.W., and C.E., created competency questions alongside its SPARQL implementation. F.K., D.C.F.W., S.S., J.R., A.L.M., T.M.W., C.E., B.Z.P. have collaboratively written the initial draft of the manuscript. In the following all authors actively contributed to the refining and enhancement of the draft in terms of clarity and accuracy. The research was overseen and checked for fitting to the research portfolio by R.W.R. and B.Z.P.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-026-07052-2>.

Correspondence and requests for materials should be addressed to F.K. or C.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026