Building the MDMC-NEP Glossary

Rossella Aversa (KIT-SCC)
SIG Metadata and Ontologies, 23.02.2024
Outline

- Background and motivation
- The MDMC-NEP Glossary of Terms
- Methodologies for glossary development
- Overcoming the implementation challenges
- Adopted practices for maintenance and updates
- Next steps
- Conclusions
Background

H2020 NEP: Nanoscience Foundries and Fine Analysis (NFFA) – Europe Pilot

- Access to nanoscience research infrastructure
- Integrate synthesis, growth and manipulation of nanostructures with fine analysis, theory and simulation
- 22 international partners, 180 techniques

[https://nffa.eu/about/]
Background

- Helmholtz JL-MDMC: Joint Lab “Model and Data driven Materials Characterization”
- Central, correlative experimental methodology platform
- Multiscale and multidimensional characterization, analytics and simulation methods
- 3 Helmholtz Centers (KIT, FZ-Jülich, HZ-Hereon)
- [https://jl-mdmc-helmholtz.de](https://jl-mdmc-helmholtz.de)
Background

- Metadata Working Group:
  - Implement (meta)data management practices following the FAIR principles
  - Develop tools and infrastructure solutions guided by community requirements
  - Agree on common descriptions
  - Collaborate on interoperable results
  - [https://jl-mdmc-helmholtz.de/mdmc-activities/metadata-working-group/](https://jl-mdmc-helmholtz.de/mdmc-activities/metadata-working-group/)
Motivation

- Different communities
- Common aims
- Similar workflows
- Need of data reuse/exchange

- Importance of a common glossary

Enhances interoperability
Facilitates collaboration
Ensures consistency
Eases data exchange
MDMC-NEP Glossary

- 45 high-level terms
- Describes computational/experimental workflows
- Reflects the lifecycle of entities and data
- Framed in the management infrastructure of the involved projects
- Tracks basic provenance information
- Living document
- DOI: 10.5281/zenodo.10663833
MDMC-NEP Glossary: overview

MDMC-NEP Glossary of Terms. DOI: 10.5281/zenodo.10663833
MDMC-NEP Glossary: overview

Research User
Person, usually member of a Project, who conducts any part of the Study, in order to collect and/or analyse Research Data or is interested in reusing Research Data by a third party (e.g., Reference Data) with the final aim to extract insights that support the answer to some specific research question (i.e., Conclusions). Research Users may be assigned with a role (data curator, instrument scientist, team leader, team member).
MDMC-NEP Glossary: workflows

Experimental Workflow

- System
- Input
- Precursor
- Fabrication
- Sample Preparation
- Sample
- Measurement
- Sample Component
- Raw Data
- Institution
- Laboratory
- Settings
- Technique
- Equipment
- Instrument
- Consumable

agents  entities  processes

Computational Workflow

- Research Data
- Data Processing
- Processed Data
- Model Preparation
- Model
- Data Acquisition
- Raw Data
- System
- Simulation
- Calculation
- Data
- Institution
- Laboratory
- Settings
- Technique
- Equipment
- Instrument
- Research Software

MDMC-NEP Glossary of Terms. DOI: 10.5281/zenodo.10663833
MDMC-NEP Glossary: data analysis lifecycle

MDMC-NEP Glossary of Terms. DOI: 10.5281/zenodo.10663833
MDMC-NEP Glossary: (meta)data management

MDMC-NEP Glossary of Terms. DOI: [10.5281/zenodo.10663833]
Glossary development

1. Clarify objectives
2. Literature review and adoption
3. Collaborative sessions
4. Use tools
1. Clarify objectives

- Define scope and purpose: ensure relevance to specific research goals
- Identify target audience: enhance effectiveness and usability within the scientific community

14 techniques
2. Literature review and adoption

- Identify existing terminology: relevant to the research field
- Adoption: ensure alignment with the relevant terms
3. Collaborative sessions

- Brainstorming and exchange: involve participants
- Consensus building and decision-making: collaborative environment

2021 2022 2023 2024

- Weekly meetings
- Monthly meetings
- Thematic subgroups
4. Use tools

- Shared tables: collaborative work and comments
- Website: promote access and visibility

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Analysis</td>
<td>The identifiable action of processing Raw or Analyzed Data, which may be performed using Data Analysis Software, combined in chains or workflows. The Data Analysis is the most crucial part of any research. It summarizes collected data. It involves the interpretation of data through the use of analytical and logical reasoning to determine patterns, relationships, or trends.</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>Data Analysis is the process of systematically applying and/or logical techniques to describe and illustrate, collect, receive, and evaluate data. According to Shamoo and Resnik, various analytical procedures provide a way of drawing inferences from data and distinguishing the signal from the noise (statistical fluctuations) present. Data Analysis involves actions and methods performed to help describe facts, detect patterns, develop explanations, and interpret results. Data Analysis is a process of inspecting, cleaning, transforming, and modeling data with the goal of highlighting useful information, suggesting conclusions, and supporting decision making. Data Analysis involves several steps and approaches, encompassing domains under a variety of names, in different business, science, and social science domains.</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>A lifecycle stage that involves the techniques that produce synthesized knowledge from organized information.</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>This glossary defines and explains the high-level terms used in the context of the MDMC metadata management environment. The definitions of terms have been derived keeping in mind a balance between the specific applications of MDMC and the definitions available in other projects (NEP, IMC, ODE, CODATA, CASRAI). The MDMC glossary is intended to be a living document, subject to updates as required by the community. For any inquiries, please contact Dr. Rosalinda Avendaño.</td>
</tr>
</tbody>
</table>


MDMC-NEP Glossary of Terms

https://jl-mdmc-helmholtz.de/mdmc-activities/metadata-working-group/metadata-wg-topics/semantics/glossary/
Overcoming implementation challenges

- Perceived lack of background
- Resistance to change
- Drop of motivation
- Conflicting terms/feedback loops

Long-term vision, regular checkpoints
Intermediate results, application examples
Advantages, engagement in the development
Harmonization as much as possible, defined scope
Maintenance and updates

- Regular reviews and revisions: up-to date and aligned to the scope
- Incorporate new terms: dynamic resource
- Enhance accessibility and interoperability: vocabulary service

EVOKS: Collaborative online vocabulary editor
- Developed at KIT
- SKOS model
- Persistent identifier to each term
- Seamless integration of terms
- Centrally maintained
- Public read-only Skosmos instance
Next steps: glossary extension

- Describe each of the processes and entities in the common workflow

Minimal

Input → Precursor → Sample → Sample Component

Fabrication → Sample Preparation → Measurement → Raw Data

System

Sample description

Adopted from the Materials Data Vocabulary
DOI: 10.5334/dsj-2021-018

SEM, TEM, MRI, STM, SEM/FIB Tomography, Nano CT/micro CT
Next steps: from glossary to ontology

- Formalize relationships: add context to the knowledge representation
- Logical constructs (classes, properties, axioms): structured and machine-readable

PRIMA (PRovenance Information for MAterials science) Ontology

https://jl-mdmc-helmholtz.de/ontology/
Conclusions

- Smooth collaborative process
- Guided by community requirements
- Driven by communication with scientists for collecting information
- Common description, tracking data provenance
- Aligned, whenever possible, with existing terminology
- Can be adopted by other materials science facilities and projects, e.g., NFDI-MatWerk, NFDI4Ing, Helmholtz Metadata Collaboration (HMC)
- High-level frame for future in-depth descriptions
- Can be extended by integrating existing fine-grained ontologies
- Training is planned for bringing the vocabulary in use
Contacts: rossella.aversa@kit.edu

Acknowledgements to:

Founded by:
The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524; the Joint Laboratory Model and Data driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association: NFFA-Europe-Pilot (EU H2020 – n. 101007417); the research program “Engineering Digital Futures” of the Helmholtz Association of German Research Centers; the Helmholtz Metadata Collaboration Platform.