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

Powered by Bing Image Creator

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

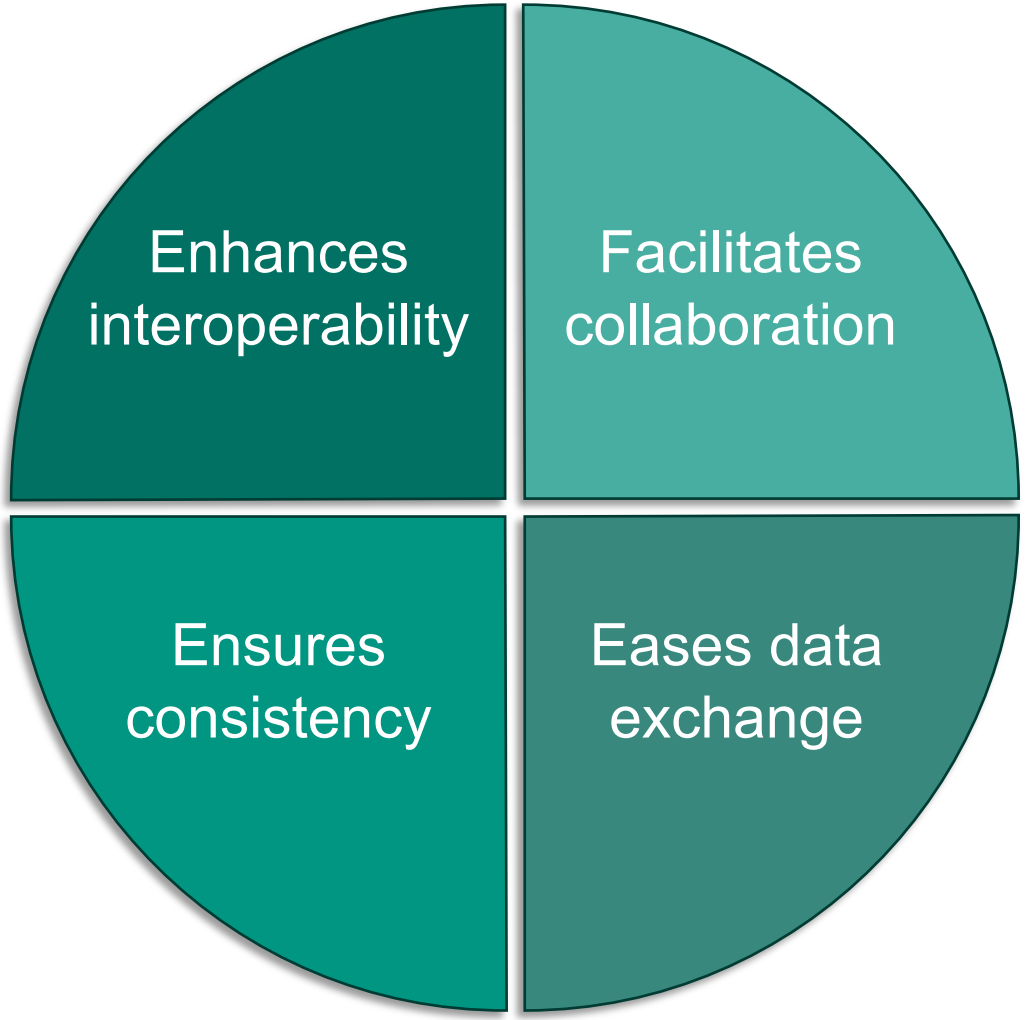
Background

- 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/

Motivation

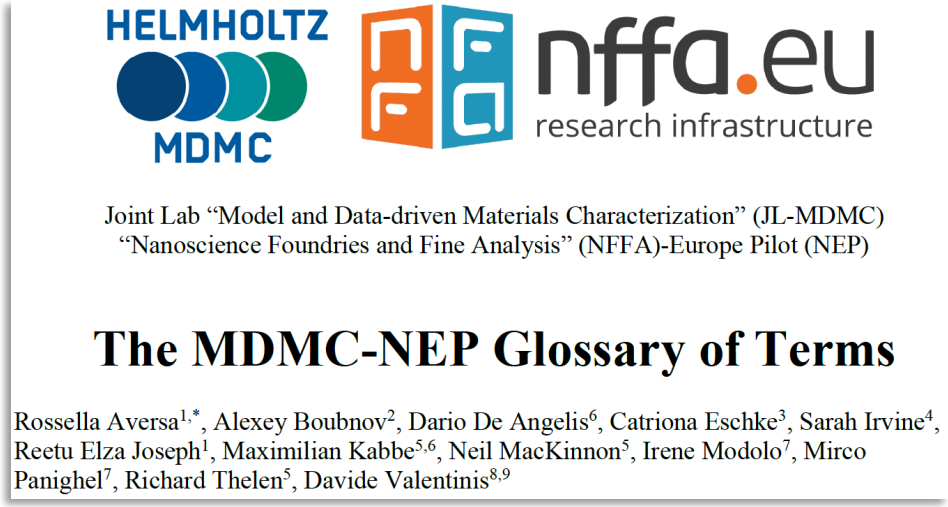
- Different communities
- Common aims
- Similar workflows
- Need of data reuse/exchange

- Importance of a common glossary

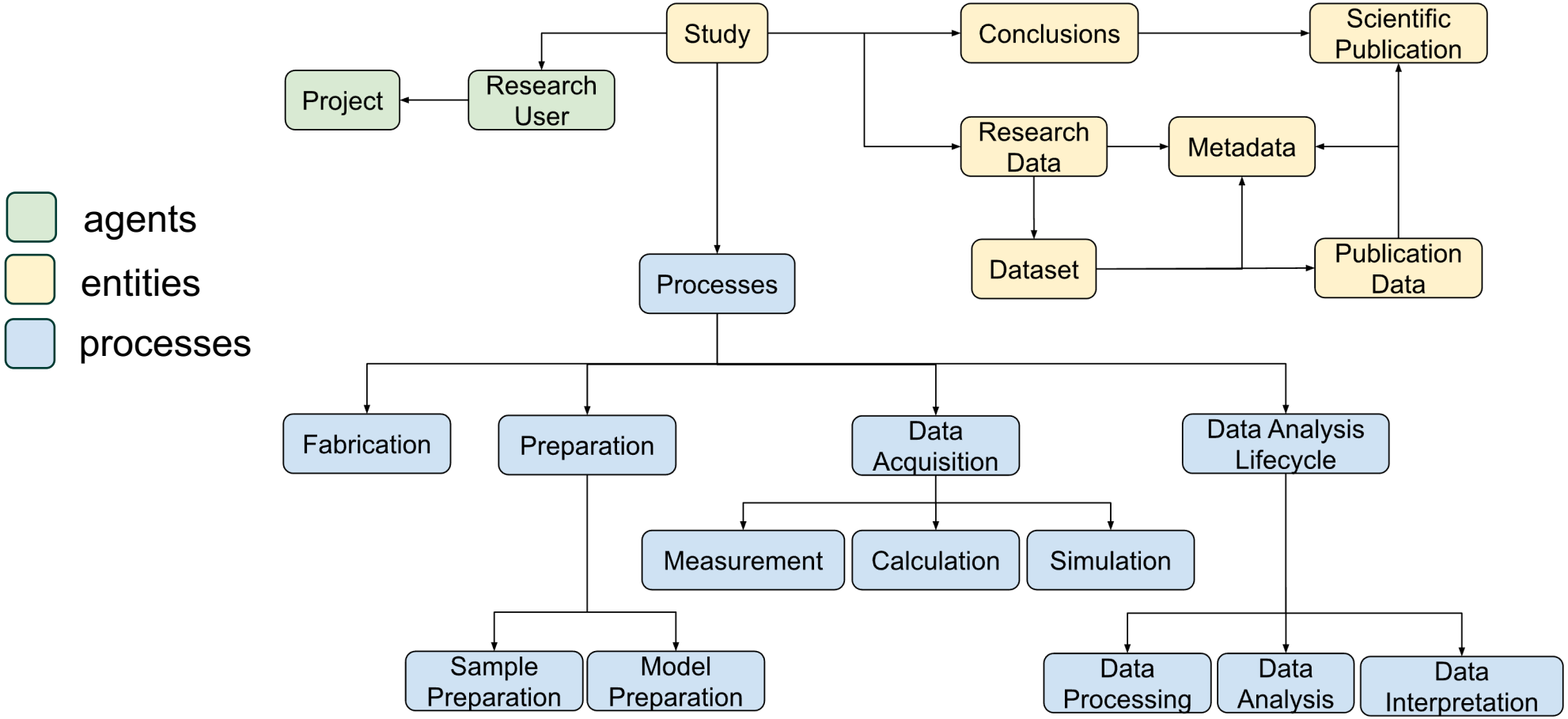


MDMC-NEP Glossary

- 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

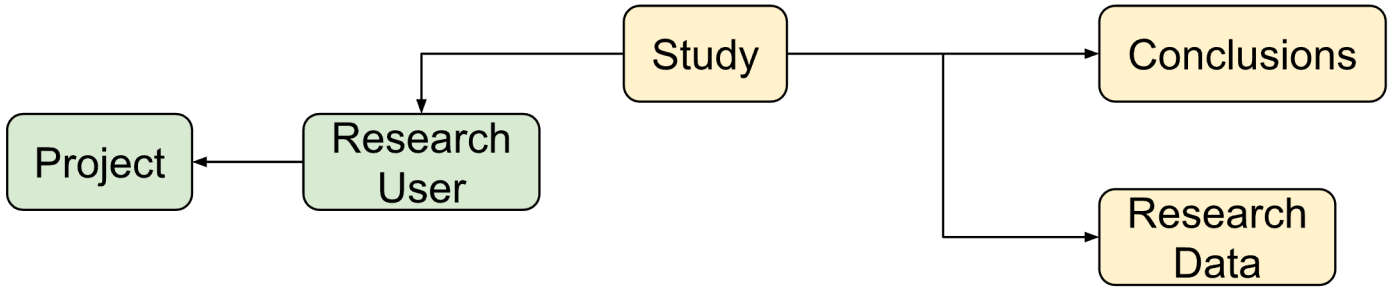


MDMC-NEP Glossary: overview



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

MDMC-NEP Glossary: overview

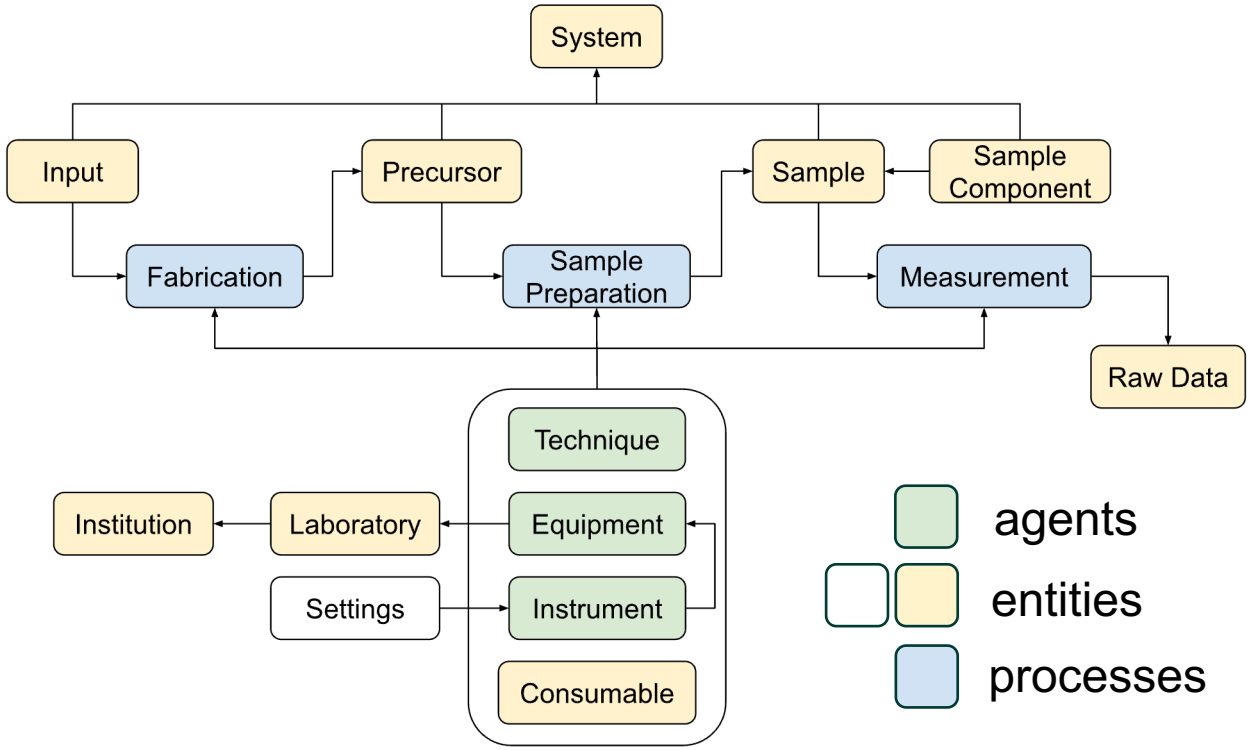


Research User
Person, usually member of a **Project**, who conducts any part of the **Study**, in order to collect and/or analyse **Research Data** or is interested in reusing **Research Data** by a third party (e.g., **Reference Data**) with the final aim to extract insights that support the answer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

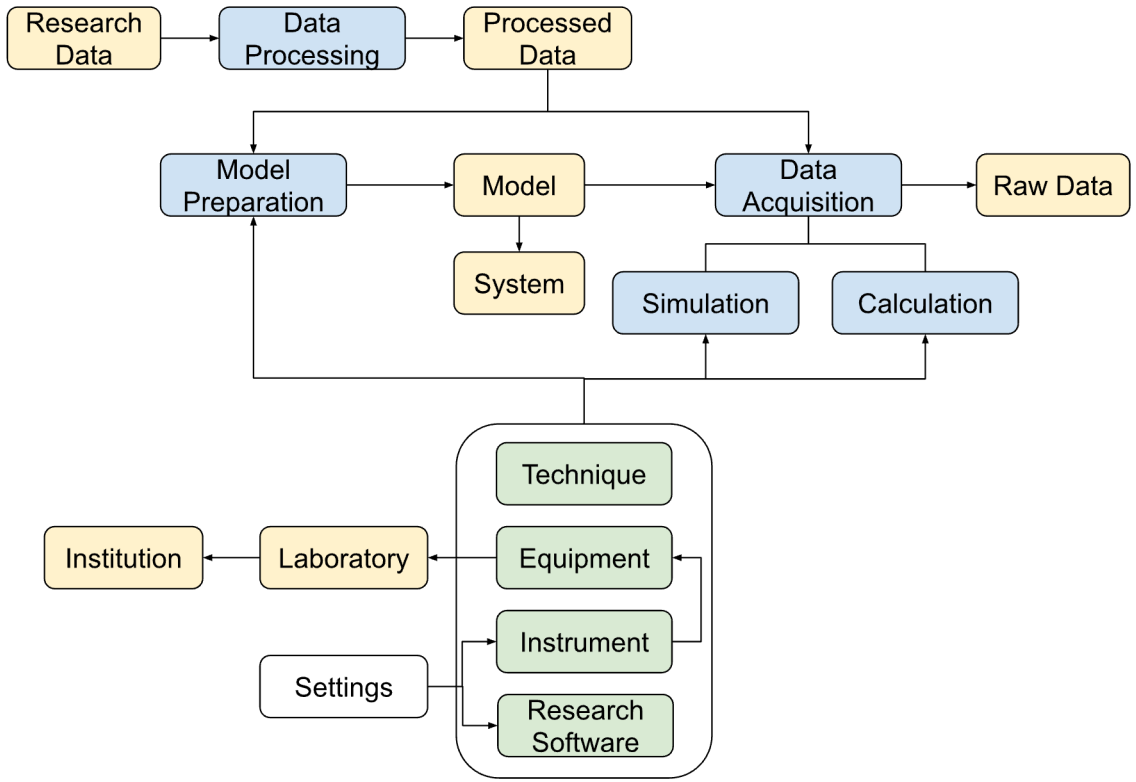
MDMC-NEP Glossary of Terms. DOI: 10.5281/zenodo.10663833

MDMC-NEP Glossary: workflows

Experimental Workflow

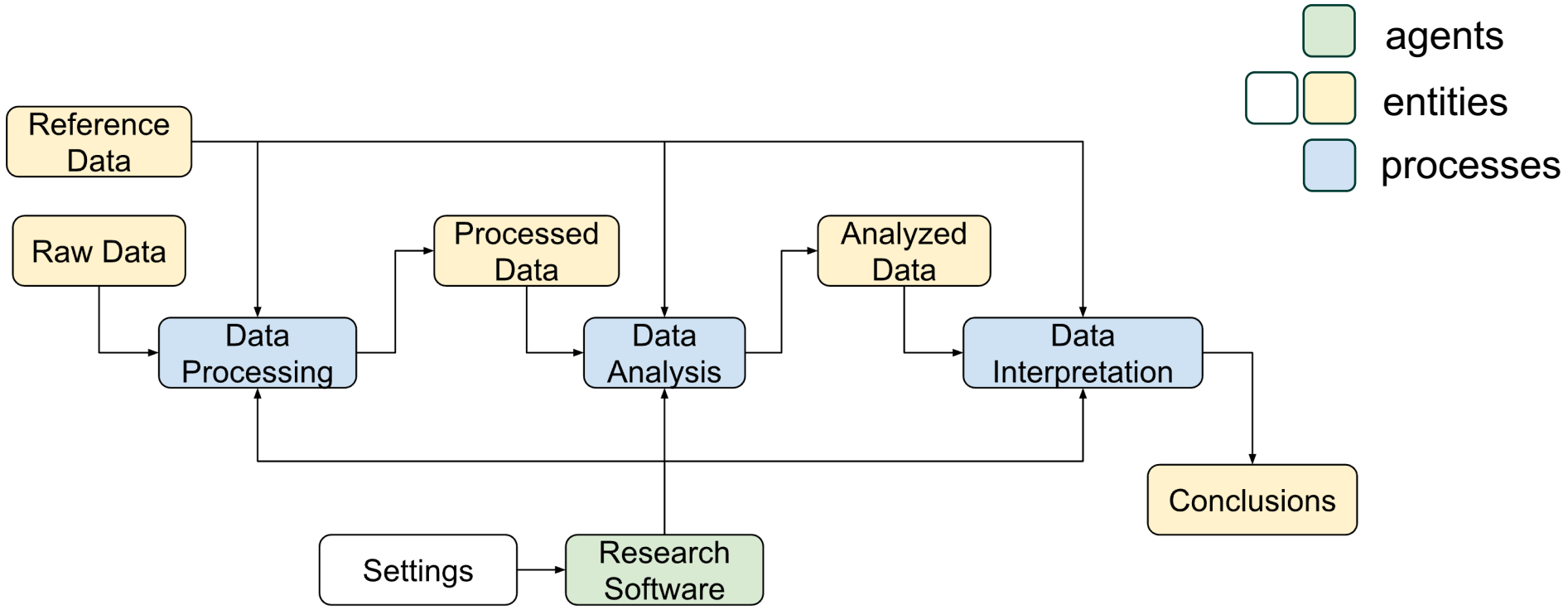


Computational Workflow



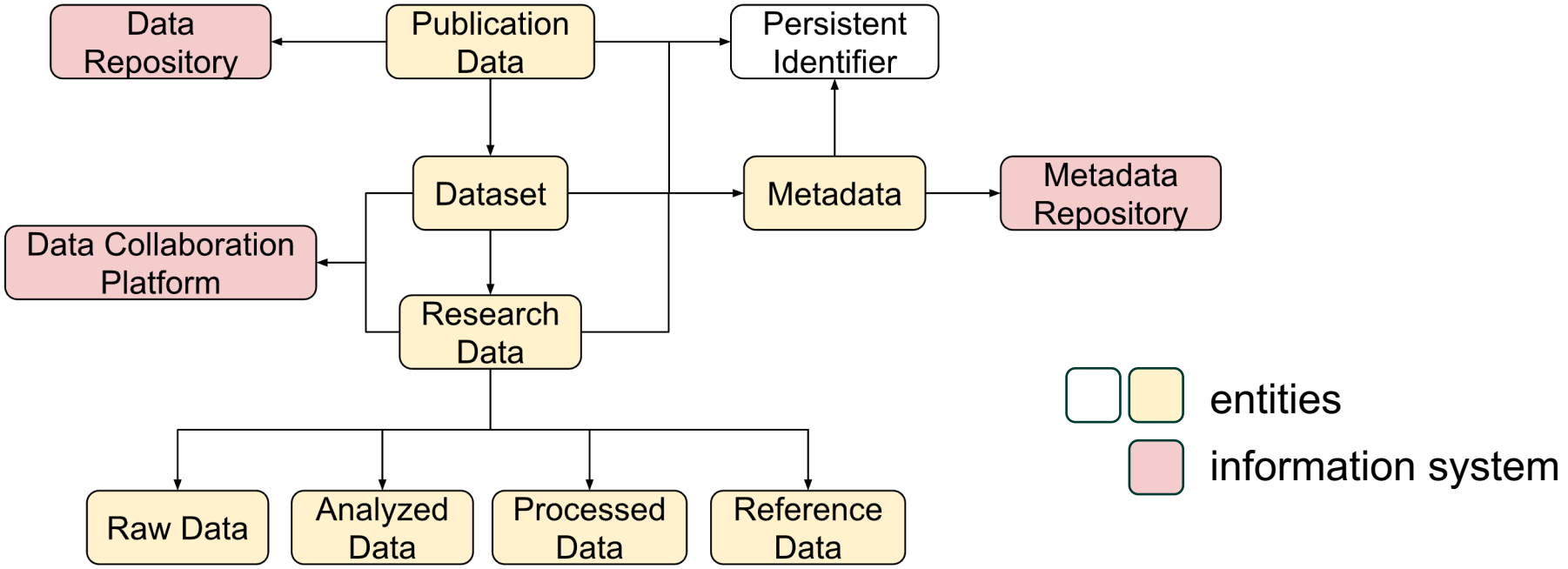
MDMC-NEP Glossary of Terms. DOI: 10.5281/zenodo.10663833

MDMC-NEP Glossary: data analysis lifecycle



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

MDMC-NEP Glossary: (meta)data management

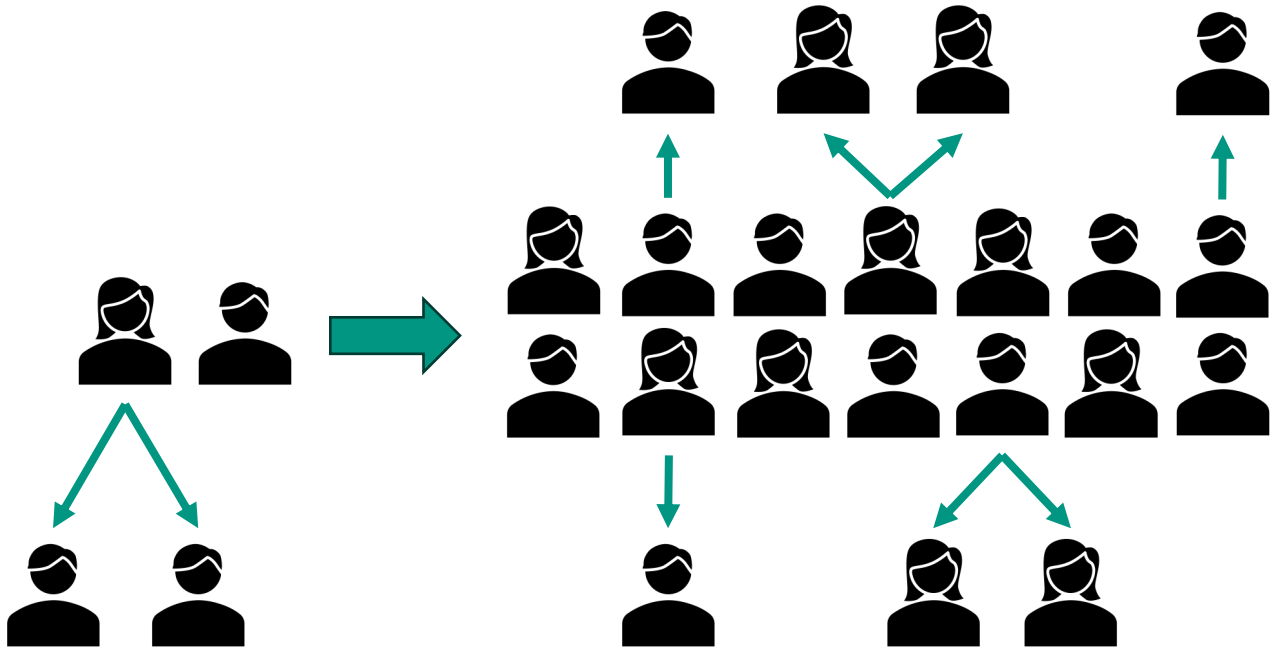


MDMC-NEP Glossary of Terms. DOI: 10.5281/zenodo.10663833

Glossary development

1. Clarify objectives

- 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

Crossref DublinCore

Metadata for Experiments in Nanoscience Foundries

Vasily Bunakov, Tom Griffin, Brian Matthews & Stefano Cozzini

HMC

DataCite Schema.org

Using a Core Scientific Metadata Model in Large-Scale Facilities

Brian Matthews, Shoaib Sufi, Damian Flannery, Laurent Lerusse, Tom Griffin, Michael Gleaves, and Kerstin Kleese

Metadata4Ing: An ontology for describing the generation of research data within a scientific activity.

Arndt, Susanne[1]; Farnbacher, Benjamin[2]; Fuhrmans, Marc[3]; Hachinger, Stephan[4]; Hickmann, Johanna[5]; Hoppe, Nils[2]; Horsch, Martin Thomas[6]; Iglezakis, Dorothea[7]; Karmacharya, Ashish[3]; Lanza, Giacomo[8]; Leimer, Sophia[9]; Munke, Johannes[4]; Terzijska, Džulia[10]; Theissen-Lipp, Johannes[11]; Wiljes, Cord[12]; Windeck, Jürgen[3]

NeXus MatPortal.org

Materials Design Ontology

Materials Resource Registry

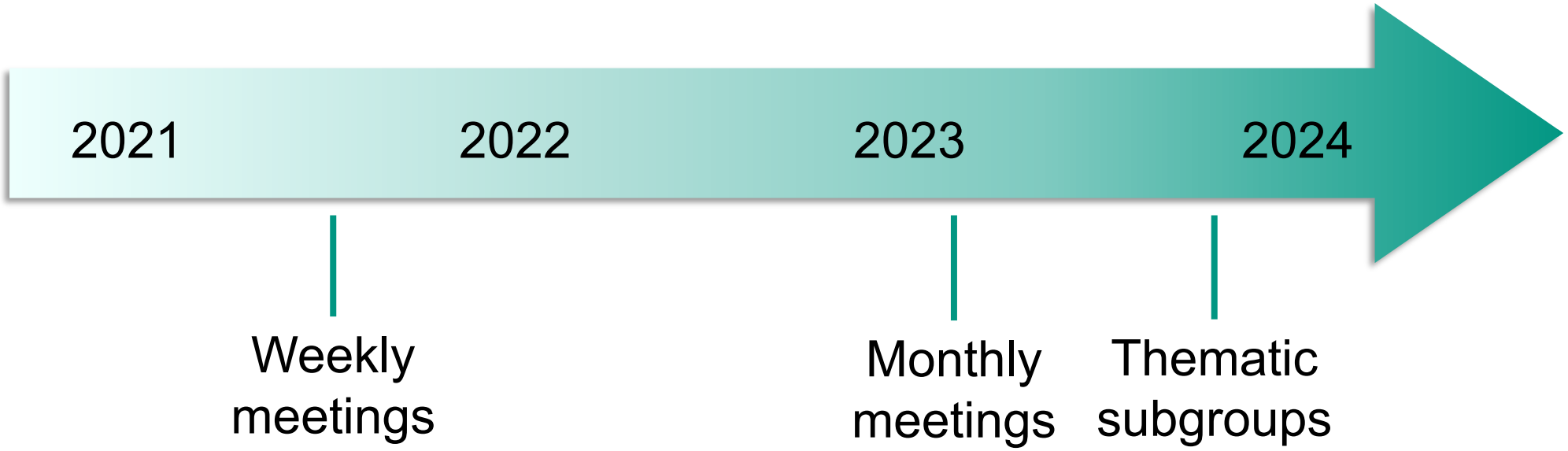
Part of the Materials Genome Initiative

CODATA CASRAI RDM Terminology

PhySH - Physics Subject Headings

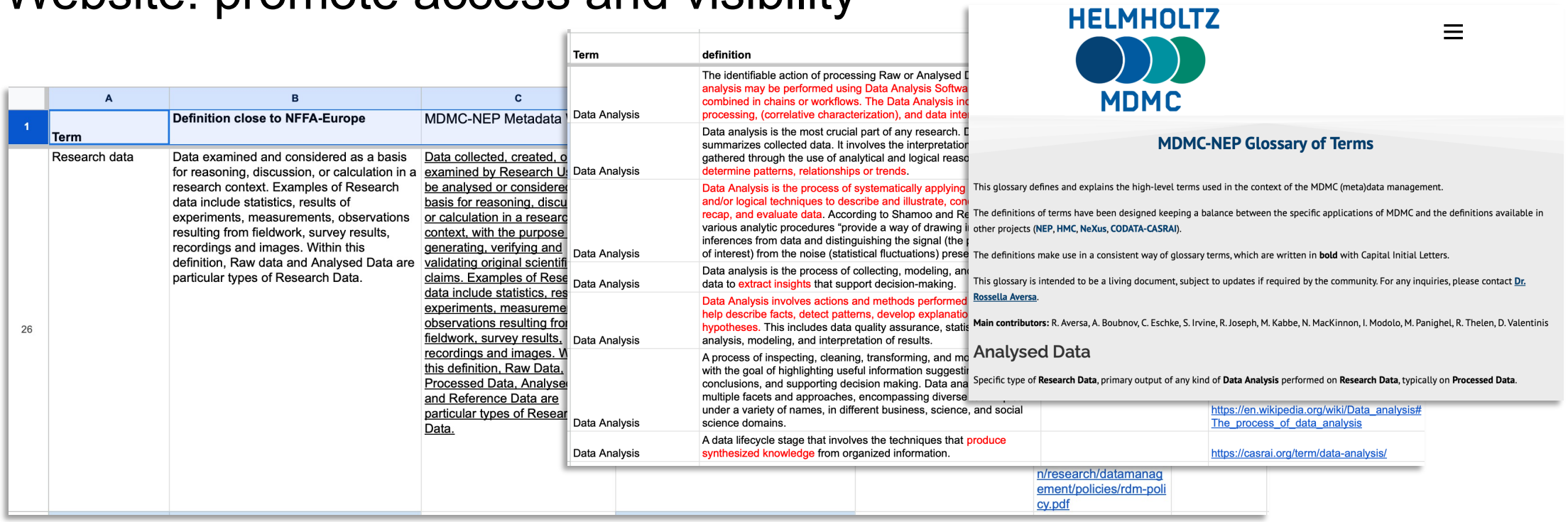
3. Collaborative sessions

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



4. Use tools

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



	A	B	C
1	Term	Definition close to NFFA-Europe	MDMC-NEP Metadata
26	Research data	Data examined and considered as a basis for reasoning, discussion, or calculation in a research context. Examples of Research data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, recordings and images. Within this definition, Raw data and Analysed Data are particular types of Research Data.	Data collected, created, o examined by Research U be analysed or considere basis for reasoning, discu or calculation in a researc context, with the purpose generating, verifying and validating original scientifi claims. Examples of Rese data include statistics, res experiments, measureme observations resulting fro fieldwork, survey results, recordings and images. W this definition, Raw Data, Processed Data, Analyse and Reference Data are particular types of Resear Data.

Term	definition
Data Analysis	The identifiable action of processing Raw or Analysed D analysis may be performed using Data Analysis Softwa combined in chains or workflows. The Data Analysis inc processing, (correlative characterization), and data inte
Data Analysis	Data analysis is the most crucial part of any research. D summarizes collected data. It involves the interpretatior gathered through the use of analytical and logical reasc determine patterns, relationships or trends.
Data Analysis	Data Analysis is the process of systematically applying and/or logical techniques to describe and illustrate, con recap, and evaluate data. According to Shamoo and Re various analytic procedures "provide a way of drawing i inferences from data and distinguishing the signal (the p of interest) from the noise (statistical fluctuations) prese
Data Analysis	Data analysis is the process of collecting, modeling, an data to extract insights that support decision-making.
Data Analysis	Data Analysis involves actions and methods performed help describe facts, detect patterns, develop explanatio hypotheses. This includes data quality assurance, statis analysis, modeling, and interpretation of results.
Data Analysis	A process of inspecting, cleaning, transforming, and mo with the goal of highlighting useful information suggesti conclusions, and supporting decision making. Data ana multiple facets and approaches, encompassing diverse under a variety of names, in different business, science, and social science domains.
Data Analysis	A data lifecycle stage that involves the techniques that produce synthesized knowledge from organized information.

n/research/datamanag ement/policies/rdm-poli cy.pdf

https://en.wikipedia.org/wiki/Data_analysis#The_process_of_data_analysis

https://casrai.org/term/data-analysis/

HELMHOLTZ MDMC

MDMC-NEP Glossary of Terms

This glossary defines and explains the high-level terms used in the context of the MDMC (meta)data management.

The definitions of terms have been designed keeping a balance between the specific applications of MDMC and the definitions available in other projects (**NEP**, **HMC**, **NeXus**, **CODATA-CASRAI**).

The definitions make use in a consistent way of glossary terms, which are written in **bold** with Capital Initial Letters.

This glossary is intended to be a living document, subject to updates if required by the community. For any inquiries, please contact Dr. Rossella Aversa.

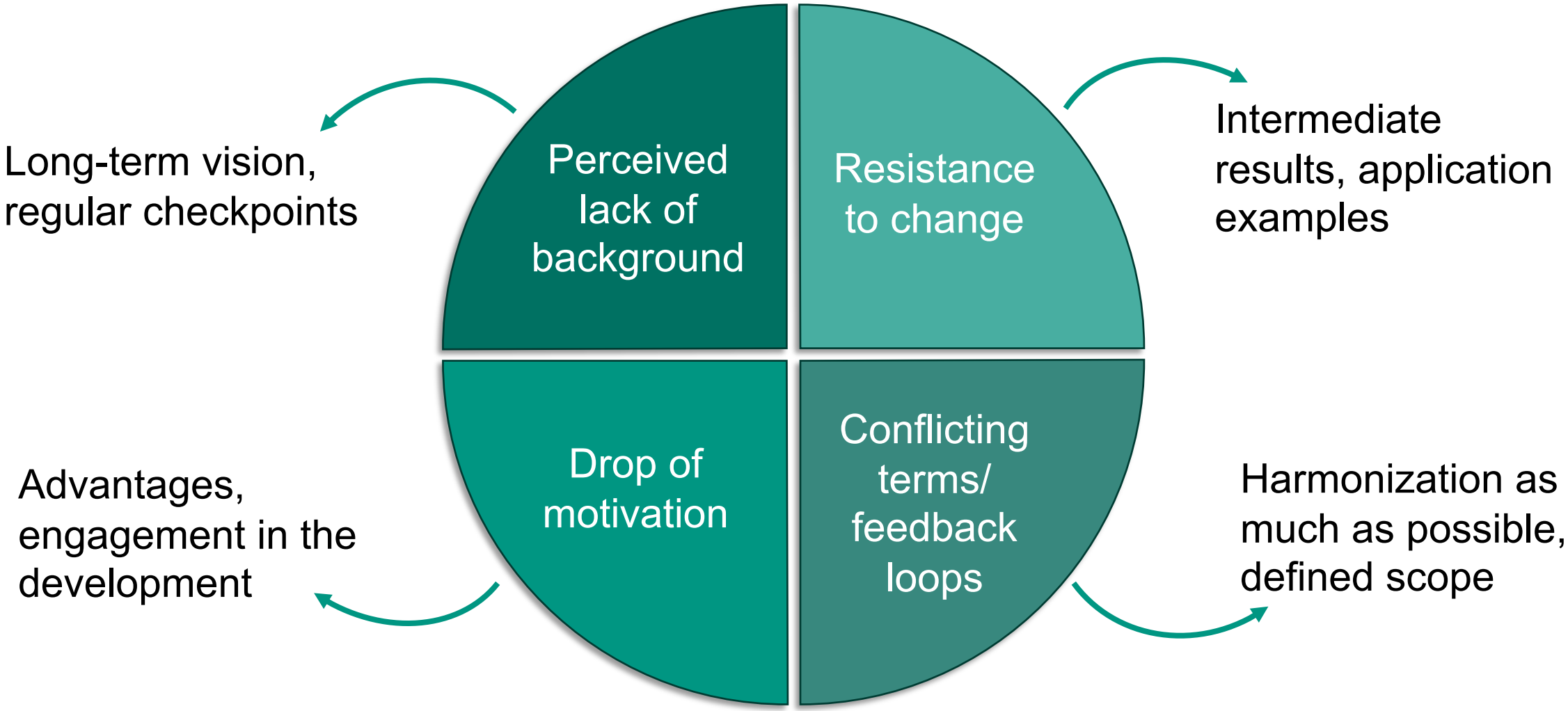
Main contributors: R. Aversa, A. Boubnov, C. Eschke, S. Irvine, R. Joseph, M. Kabbe, N. MacKinnon, I. Modolo, M. Panighel, R. Thelen, D. Valentinis

Analysed Data

Specific type of **Research Data**, primary output of any kind of **Data Analysis** performed on **Research Data**, typically on **Processed Data**.

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

Overcoming implementation challenges

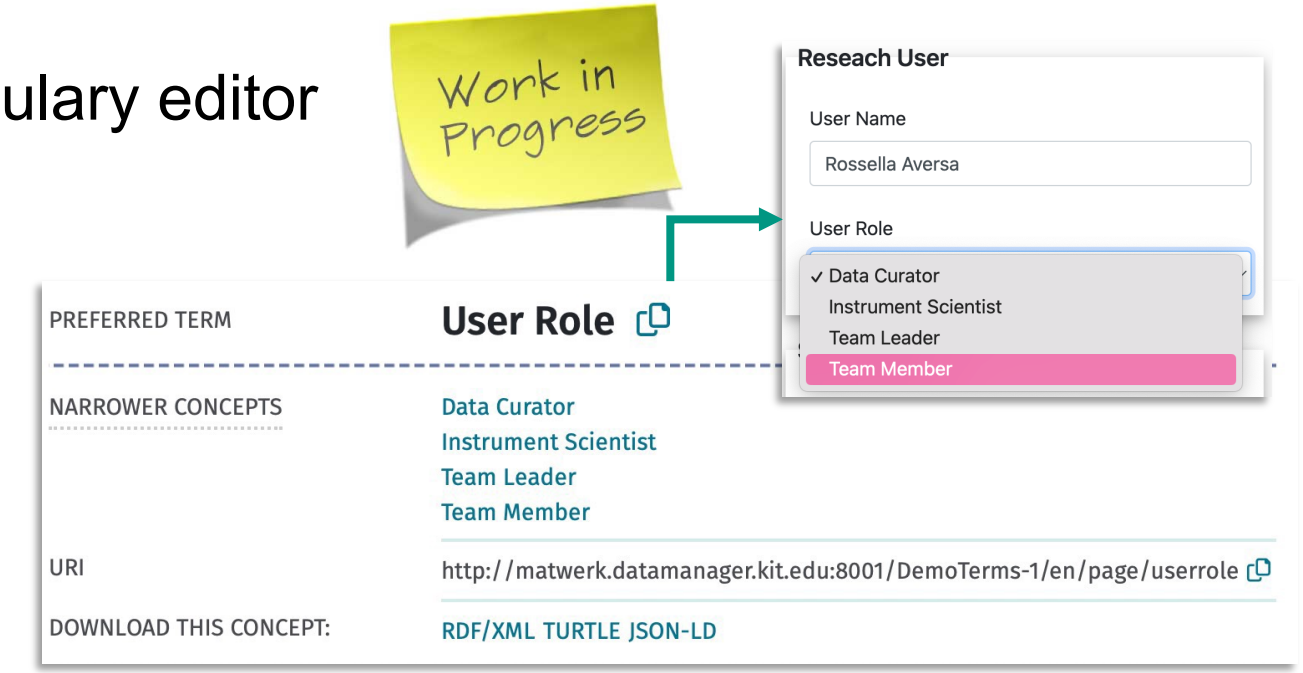


Maintenance and updates

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

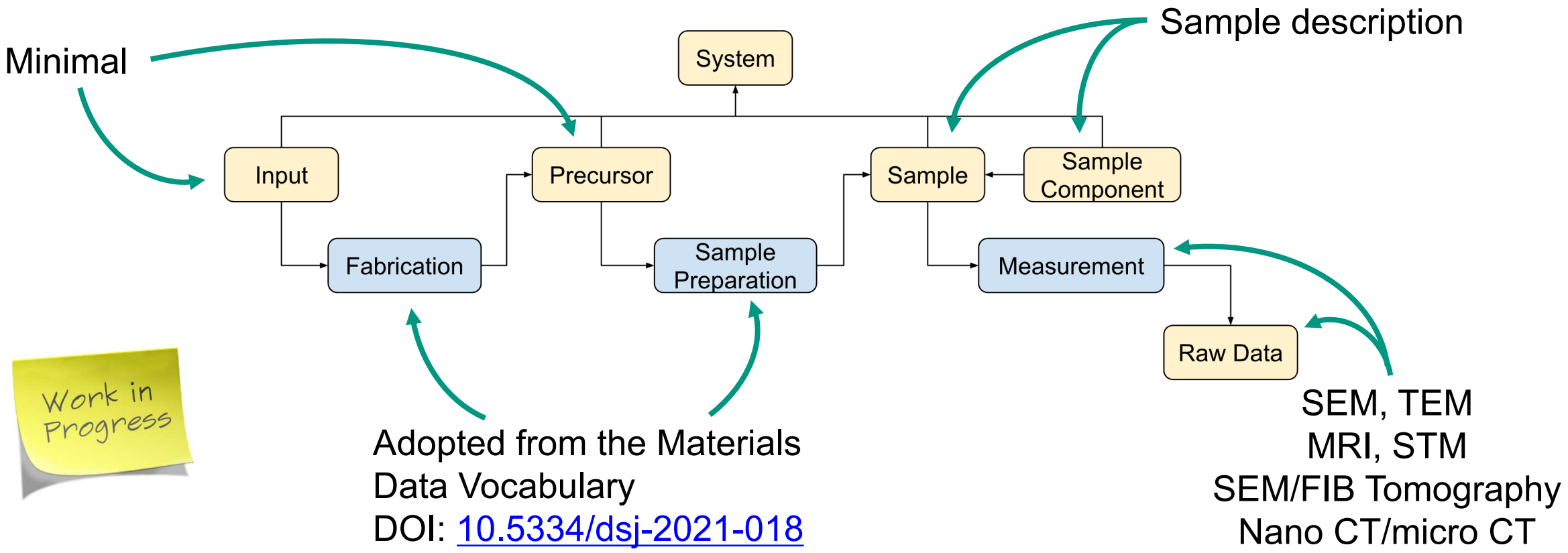
EVOKS: Collaborative online vocabulary editor

- 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

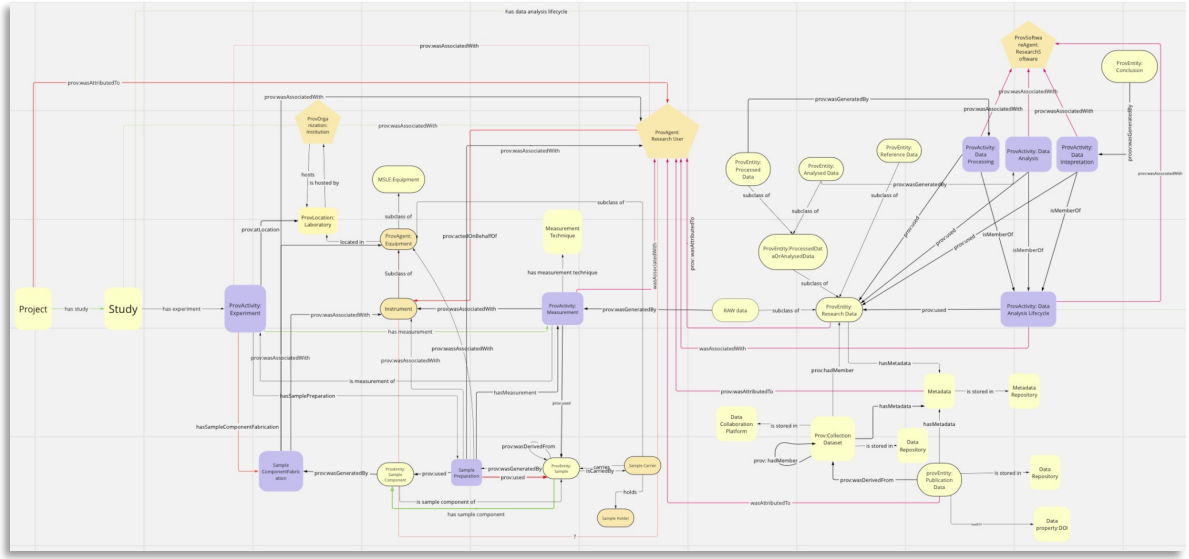
- Describe each of the processes and entities in the common workflow

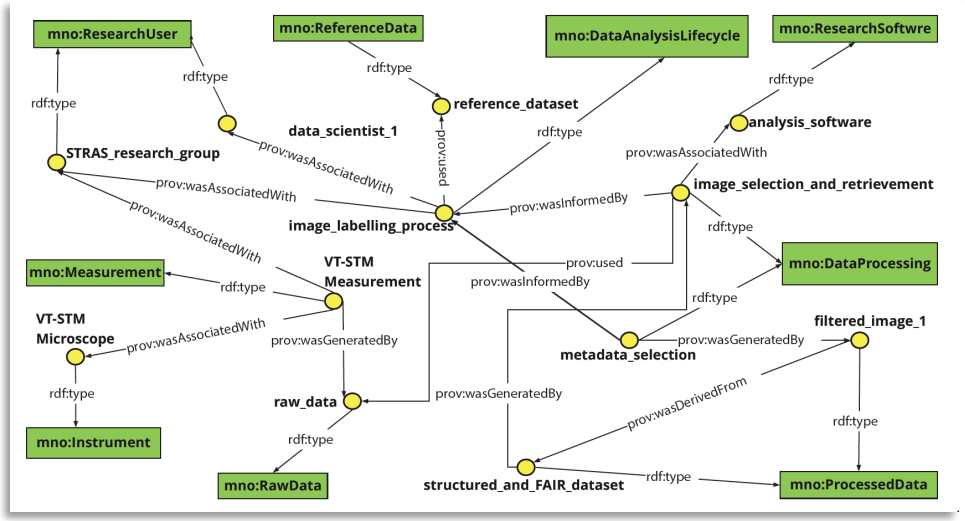


Next steps: from glossary to ontology

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

PRIMA (PRovenance Information for MAterials science) Ontology





Work in Progress

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:
G. Abdildina, N. Blumenröhr, F. Ernst, V. Haltmann, M. Inkmann, T. Jejkal, A. Kirar, E. Vitali, NEP JA6, JL-MDMC Metadata WG.

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.