Metadata Schema to support FAIR Data in Scanning Electron Microscopy

Reetu Joseph[1][0000-0002-1507-9327], Aditya Chauhan[2], Catriona Eschke[3][0000-0002-1033-144X], Ahmad Zainul Ihsan[4][0000-0002-1008-4530], Mehrdad Jalali[5][0000-0003-2465-4933], Ute Jäntsch[6], Nicole Jung[7][0000-0001-9513-2468], C. N. Shyam Kumar[8][0000-0003-4860-5327], Christian Kübel[5,10,11][0000-0001-5701-4006], Christian Lucas[3], Matthias Mail[5,10][0000-0002-9732-8453], Andrey Mazilkin[5], Charlotte Neidiger[5], Mirco Panighel[9][0000-0001-8413-5196], Stefan Sandfeld[4][0000-0001-9560-4728], Rainer Stotzka[1][0000-0003-3642-1264], Richard Thelen[2], and Rossella Aversa[1][0000-0003-2534-0063]

[1] Karlsruhe Institute of Technology, Steinbuch Centre for Computing (SCC), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
{reetu.joseph, rossella.aversa}@kit.edu

[2] Karlsruhe Institute of Technology, Institute of Microstructure Technology (IMT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

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

[4] Forschungszentrum Jülich, Institute of Advanced Simulation, Materials Data Science and Informatics (IAS-9) Wilhelm-Johnen-Straße, 52425 Jülich, Germany

[5] Karlsruhe Institute of Technology, Institute of Nanotechnology (INT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[6] Karlsruhe Institute of Technology, Institute for Applied Materials- Applied Materials Physics (IAM-AWP), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[7] Karlsruhe Institute of Technology, Institute for Biological and Chemical Systems - Functional Molecular Systems (IBCS-FMS), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[8] CEA Saclay, Service de Recherches de Métallurgie Physique (SRMP), 91191 Gif-sur-Yvette, France

[9] Consiglio Nazionale delle Ricerche - Istituto Officina dei Materiali (CNR-IOM), Area Science Park, Strada Statale 14, km 163,5, 34149 Basovizza (Trieste), Italy

[10] Karlsruhe Institute of Technology, Karlsruhe Nano Micro Facility (KNMF), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[11] Technical University Darmstadt, Department of Materials & Earth Sciences, Alarich-Weiss-Straße 2, 64287 Darmstadt, Germany

Abstract. The development and the adoption of metadata schemas and standards are a key aspect in data management. In this paper, we introduce our approach to a metadata model in the field of Materials Science. We present the specific use case of a metadata schema for Scanning Electron Microscopy, a characterization technique which is routinely used in Materials Science. This metadata schema is aiming to be a de-facto standard which will be openly available for reuse and further extension to other electron microscopy techniques.

Keywords: Data Management · Metadata schema · Materials Science · Scanning Electron Microscopy.

1 Introduction

Research in Materials Science is gaining more importance than ever, with applications ranging from the nanometer scale up to large meter-sized structures. Investigation of the structural, chemical, magnetic or optical properties and connection with the underlying microstructural aspects through detailed characterization have enabled the development of advanced materials with superior properties and functions. This, in turn, has made it possible to unravel hidden mysteries pertaining to surprising behaviors and to the invention of new materials. The huge number of experimental and computational techniques to study, characterize, and predict properties of materials results in a large variety of datasets and representations.

Materials Science is a highly multi-disciplinary research field where scientists often need to access data from more than one discipline in order to properly characterize materials, to understand the governing mechanisms, to answer critical research questions and to design novel materials. This aspect is particularly important in correlative characterization, where the task is to combine different types of information from co-referenced (in time or space) multi-dimensional data obtained using different measurement techniques. Each of these methods results in datasets that have to be combined and correlated in order to fully characterize materials and thereby to relate features and properties of different sample areas, across multiple length scales, or on different time scales. The exchange and combination of information can be facilitated if the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [51] are considered for the output of the measurement data.

Many groups within the broad field of Materials Science are already developing metadata schemas and ontologies, driven by their individual organizational goals and guided by the open and FAIR data initiatives (e.g., EOSC [14], Horizon 2020 [24], Horizon Europe [16], Turning FAIR into reality [17]). However, to promote interoperability, a harmonization approach to find a common description of the data coming from the measurements needs to be developed and should be adopted as far as possible.

Motivated by this, we started a coordination effort between the two projects we are directly involved in: the NFFA (Nanoscience Foundries and Fine Analysis) EUROPE Pilot (NEP) [38] and the Joint Lab "Integrated Model and Data Driven Materials Characterization" (MDMC) of the Helmholtz Association [22]. Our final objective is to describe the entire workflow of an experimental project, providing rich metadata to make data interoperable. This is an incremental process; as a first use case, we developed a metadata schema to describe Scanning

Electron Microscopy (SEM) measurements. This schema is aiming to become a de-facto standard for SEM and to pave the way towards further extension to other electron microscopy techniques.

2 Approach and Methodology

2.1 The Approach to a Communal Model

Any metadata model represents a simplified and idealized subset of the reality. The particular choice of the idealization depends on the specific purpose of the model. To be able to design a communal model, we compared the aims of the two above mentioned projects [38, 22]: NEP users are interested in tracking the lifecycle of data that are collected in nanoscience experiments and then get archived [4]; the Joint Lab MDMC focuses on correlative characterization, which implies to conduct experiments and to put into relation different results. It appears very clearly that the two communities share a similar need, which can be satisfied by the proposed metadata model: a contextually rich description of the experimental lifecycle, from sample preparation to data analysis, which we define as a *study* following the core concept in [33].

Figure 1 shows the basic steps of a simple study, in which only a single measurement is performed. In reality, measurements from two or more different techniques are usually engaged; however, the steps involved in the workflow for each of them remain the same. At each stage of the process, we expect to collect the appropriate metadata describing the context of the experiment at that particular phase.

In this work, we focus on describing a measurement (green box in the second block in Figure 1), which is usually performed using an instrument (yellow box in the second block) to measure a sample (light blue circle in the second block) and to generate raw data (light blue circle in the third block). In section 3 we will show how this structure can be applied to the specific SEM use case.

2.2 The State-of-the-Art Landscape

We surveyed the current landscape, looking for existing schemas or standards we could adopt. The well known high level schemas, such as crossref [7], Dublin Core Metadata Element Set [10], DataCite [36] and schema.org [45], are intended to be generic, rather than customized to the needs of any particular discipline. Nevertheless, community efforts to extend schema.org to the field of Materials Science are currently under development [46].

A particularly interesting core model is the Core Scientific Metadata Model (CSMD) [33, 49], which is oriented towards Facilities Science, to capture high level information about scientific studies and the produced data. Some elements of CSMD also occur in the first version of the NFFA-Europe metadata framework [4], which we used as the basis for our model shown in Figure 1. Schemas such as the Open Geospatial Consortium (OGC) SensorML [2] and the Alfred

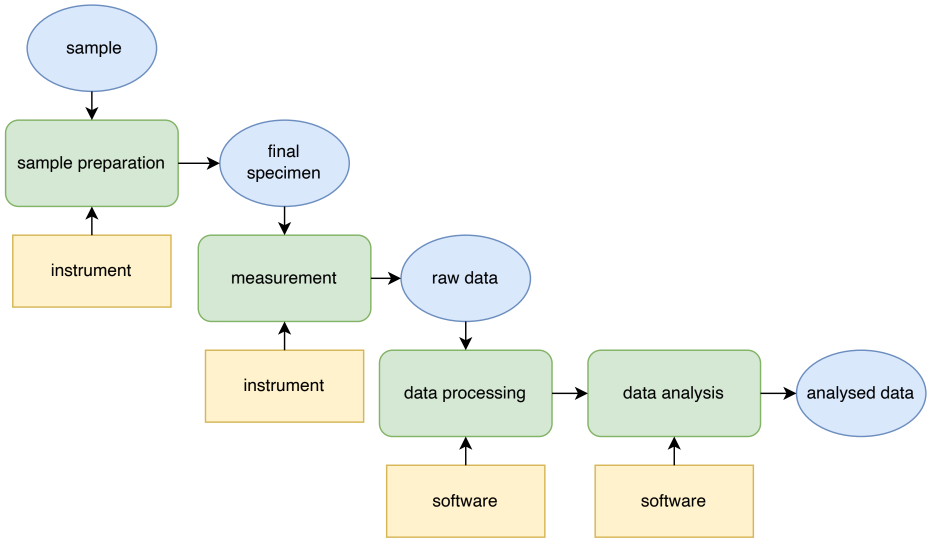

Fig. 1. Idealized workflow describing the basic steps of a simple study.

Wegener Institute (AWI) Sensor Information System [47] provide a conceptual model and a technical metadata description, as well as eXtensible Markup Language (XML) and JavaScript Object Notation (JSON) encodings for metadata of sensors and measurement processes. In addition to that, the Research Data Alliance (RDA) [42] "Persistent Identification of Instruments" (PIDINST) Working Group facilitates a community-driven solution for persistent identification of instruments [48]. In our schema, we describe the instrument metadata to be compliant with the PIDINST schema, where applicable.

We also reviewed the state of Electron Microscopy initiatives with the hope of reusing existing schemas describing the SEM technique; while there are a number of valuable efforts such as [8, 41, 40, 44, 18], none of them is fully in line with the purpose of our schema. In particular, the Electron Microscopy Public Image Archive (EMPIAR) schema [28] is designed to describe image sets; SEM appears here as a technique used to produce the raw data, but apart from the pixel size no additional instrument settings are included. A more exhaustive description of SEM appears in the NanoMine XML schema [53] as specification of the materials characterization equipment used, methods and experimental conditions. The main focus of this schema is on the representation of polymer nanocomposites and not so much on instrument description.

Core and domain-specific controlled vocabularies (e.g., [37, 19, 1]) and ontologies (e.g., [15, 6]) in the Materials Science and Engineering community also exist. In particular, MatONTO [5] is developed to represent materials, properties, structures, and processes in the Materials Science domain. Moreover, it leverages and maps the available metadata, e.g. the Crystallographic Information Framework (CIF) dictionary [8], into an ontology module. The authors in [52] developed an ontology based on MatML [32] to enable the data interoperability and exchange. Another ongoing effort to develop a core ontology in Materials Science is the European Materials and Modelling Ontology (EMMO) [13] developed by the European Materials Modelling Council (EMMC) [12]. It provides a common semantic framework for describing materials, models, and data with the possibility to extend and adapt it to other domains. The recent work in [25] has shown the development of a level-domain ontology of crystalline materials defects (i.e., a dislocation ontology). Even though we do not exclude the possibility to consider such efforts in future extensions, e.g. mapping and re-engineering the schemas into an ontology, we currently focused on a metadata schema describing SEM measurements.

In this context, a promising initiative is NeXus [30], which was originally developed as an international standard for neutron, X-ray and muon facilities, with major focus on recording data directly taken from experimental equipment. The NeXus data format offers a flexible metadata structure, together with an extendable glossary of domain-specific terms. In our schema, we attempted to adhere to the NeXus naming convention as much as possible, and thereby to allow basic interoperability.

3 The SEM Metadata Schema

3.1 The Implementation

The proposed schema is constructed as a hierarchy of relevant information blocks, which we call *groups*, resembling the NeXus groups [30] and adopting NeXus naming convention, when applicable. Figure 2 shows how the hierarchy of the groups is structured in the SEM schema. In Section 3.2 all the elements used in each group are reported.

Many integrations to the NeXus model have been done, in particular on the description of the dimensional quantities, which should contain at least value, uncertainty, and units. We defined all these quantities using a complex type `dimensionalDetail`, which we adopted from the MatML schema [32] and can be mapped onto the schema.org [45] `QuantitativeValue`. It is worth mentioning that the similar complex type `ScalarUncertainty` in Nanomine [53] contains an element called `data`, which describes the data distribution. This does not fit in our SEM schema, however it might be relevant in future extensions to spectroscopy. It is worth noticing that we decided to define a specific type for each quantity rather than a unique one, in order to provide a controlled list of allowed values and a default value. The complete XML implementation can be found at [34].

Each property of the type `identifier` resembles the PIDINST [48] structure, including both an `identifierValue` and an `identifierType`, such as URL, ROR [43], GRID [20], ISNI [27], DOI [9], Handle [21]. Including ROR IDs in metadata potentially enables more efficient discovery and tracking of publications by institutions and makes unambiguous affiliation information widely and freely available.

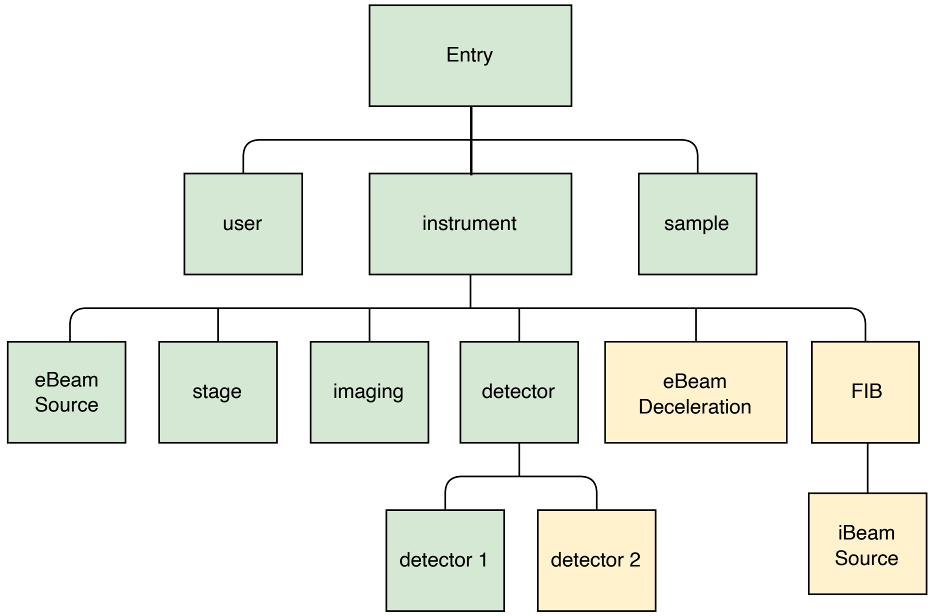

Fig. 2. The hierarchy used in the schema to describe a SEM measurement. Green and yellow boxes represent required and optional groups, respectively.

3.2 The Structure of the Schema

In this Section, we describe each of the main groups constituting the SEM schema. The hierarchy of these groups is illustrated in Figure 2. The complete XML implementation of the schema can be found at [34].

The `entry` level. The `entry` level is the root element of the schema, and resembles the NeXus [30] `NXentry` base class definition. It contains all the metadata describing a single measurement. Being the design of the schema modular, an entry for each measurement can be included in case of multiple measurements.

The `user` group. The `user` group contains the contact information of the user responsible for the measurement. The design of the schema is modular: if more than one user is involved, it is possible to add as many `user` groups as needed. The metadata properties have been selected from the NeXus [30] `NXuser` group, adopting the same naming convention. it is worth noticing that our user properties map well to the ones of DataCite [36] `Contributor`.

To indicate the role of the user, the `role` property is included with a controlled list of values which have been selected from DataCite [36] and adapted to the specific needs of our projects: in particular, we mapped `Project Leader` to `Principal Investigator` and we created a new definition for `Work Package`

`Leader` fitting our context. A Common Vocabulary for Nanoscience Data Management was already developed for NFFA-Europe [4]: from there, we adopted the definitions of `Instrument Scientist` and `Research User`. One new term (`Site Leader`) was recently added to meet the requirements of the NEP project. In the following, we list the definitions which differ from DataCite:

- `Instrument Scientist`: A person, or a group of them, who manage a particular instrument, or a set of them [4]. This is the person who usually performs the measurement and possibly the data analysis;
- `Research User`: A person, a group of them, or an institution (organization) who, within a project, conducts experiment on one or more laboratories using one or more instruments in order to collect and analyze research data, or is interested in data collected or analyzed by other research users on the same or other laboratories. `Research User` may be assigned with a role, e.g. a designation as `Principal Investigator` [4];
- `Site Leader`: A person in charge of a measurement technique at the location of a laboratory;
- `Work Package Leader`: A person responsible for the coordination, supervision, and implementation of the specific activities in the project.

The details of the user affiliation have been selected from the Crossref schema version 5.0 [7] and are listed in a subgroup called `affiliation`.

The `sample` group. The `sample` group includes any information describing the sample on which the measurement is performed, similar to the NeXus [30] `NXsample` group. Depending on the particular research context, multiple `sample` groups can be added to the same measurement.

We intentionally did not cover any details on material structure or properties, as many ontologies exist already for this purpose. In future developments, a standardized description of the material is going to be included. A promising option seems to be the Materials Science and Engineering related vocabulary [1] currently used in the Materials Resource Registry [31], which was developed at the National Institute of Standards and Technology (NIST) [39] and through the RDA [42] "International Materials Resource Registry" Working Group [26].

The `instrument` group. The `instrument` group describes the collection of the components of the instrument. Following the NeXus design [30], each component is by itself a group, as sketched in Figure 2, and will be described separately in the following. This structure has many advantages: it makes the schema modular, flexible, and easily extendable. Moreover, we included the unique identification of the detectors (via `detectorID`) and the `componentGeometry`, which offers a way to put into relation the coordinates within the Cartesian coordinate system of the laboratory, which we term the *global coordinate system*. Positions, directions, coordinates, uncertainties, and units refer to this coordinate system unless stated otherwise.

The `source` group. The `source` group gives details about a beam source. It is designed to be reused in multiple contexts, as the structure of the schema is flexible: in our case, we use it to describe both the electron beam source in the `instrument` group and the ion beam source in the `FIB` group (see Figure 2).

The `stage` group. The `stage` group is a required group describing the stage settings during a measurement. As it is not included in the NeXus design [30], the selection of the appropriate properties and their names was based on the user experience and on the comparison of the metadata provided by different manufacturers.

The `imaging` group. The `imaging` group lists the imaging settings of the instrument during the measurement. It does not describes the output file, as e.g. the EMPIAR schema [28]; it rather includes properties such as `isCorrelationImage` to indicate whether the image produced in the measurement is used for correlation with another image. If so, `coordinates` is useful to make the images refer to the same global coordinate system.

The `detector` group. The `detector` group describes the details and the settings of the detector used during the measurement. If multiple detectors are used, they can be put under the `detectors` group (see Figure 2), reflecting the NeXus design [30]. The SEM schema contains at least one detector, with the further option to include a second one in case of signal mixing. The metadata properties of this group were tailored to be applicable for all kinds of detectors that can be mounted on the SEM, such as detectors for Energy-Dispersive X-ray Spectroscopy (EDS) or Wavelength-Dispersive X-Ray Spectroscopy (WDS).

The `eBeamDeceleration` group The `eBeamDeceleration` group is an optional group describing the settings of the deceleration applied to the electron beam, if any.

The `FIB` group. The Focused Ion Beam (`FIB`) group describes the details and the settings of the FIB system attached to the SEM, if any. Among the metadata properties, the `gasInjectionSystem` subgroup provides the option to include the details of a gas injection system (GIS) which can be used for deposition, charge compensation or enhanced milling. As shown in Figure 2, an optional `source` group describing the ion beam source may be added.

4 Discussion

Scanning Electron Microscopy is an almost ubiquitous characterization technique: in the NEP consortium, 16 partner institutes out of 22 have an SEM in their laboratories, and within the Helmholtz Joint Lab MDMC the ratio is comparable. A very similar situation can be probably found in any Materials Science related laboratory involved in experimental characterization. Thus, a schema tailored for SEM is strategic and of great importance: currently, it can be seen as the only feasible way to systematically and consistently reproduce or reuse SEM measurement data. Such a schema also represents the first step towards true interoperability among different characterization techniques.

Nonetheless, a number of practical challenges exist: e.g., different manufacturers have different ways to provide metadata; dealing with data produced by multiple instruments may become problematic, even more so when trying to closely relate them to each other, as it happens in correlative characterization. The proposed schema is cross-platform and includes options to account for the variation in settings between different models; the future use of a vocabulary service will ease the mapping of metadata used by different manufacturers and, in turn, the exchange of data between communities which still have their own terminology.

A positive aspect of our schema is that it is flexible and adaptable, suiting the needs of the Materials Science community. For instance, an instrument can be upgraded by mounting extra detectors or sources. Our modular design offers the possibility to include additional groups in order to accommodate further extensions and system upgrades. Moreover, the suggested properties allow for a versatile description. As an example, the same schema can cover measurements performed in different configurations: SEM (secondary electron or back-scattered electron), EDS, WDS, cathodoluminescence or environmental-SEM (e-SEM).

Having a rich metadata description offers many advantages to the users, but might also be demanding. First of all, the sheer number of parameters to be set can be overwhelming. To tackle this issue, many fields can be automated to take input from an Electronic Lab Notebook (ELN). Chemotion [50], Kadi4Mat [3], eLabFTW [11] and the Hereon ELN under development at the Helmholtz-Centre Hereon [23] seem promising tools to be adopted within our projects to digitize the process of recording the metadata associated with a measurement.

If such an ELN is not available, the individual metadata properties have to be entered manually, representing a demanding task for the user. A Graphical User Interface (GUI) would provide a simple access point, and the integration of a metadata editor [29] would facilitate the user in filling the fields as well as in validating the metadata document against the provided schema.

To simplify the introduction of a new data management practice for the users, the current number of required properties in our schema is intentionally limited; an example of a metadata document containing only required fields is available at [34]. Moreover, controlled lists are implemented and default values are pre-entered for most parameters. This choice, based on user experience, limits the flexibility but facilitates adoption and semantic interoperability. New values in the controlled lists could be added in future, to meet the needs of the community.

It is worth mentioning that unique URIs for metadata elements can constitute a basis for the further sharing of data records as Linked Open Data, although the actual implementation of it is beyond the scope of this work.

5 Conclusion

The schema presented in this paper offers an effective description of SEM measurements, striking a balance between the necessary and available instrumental parameters. The implementation is based on the experience of more than ten expert users from five different institutes, spanning across three countries. By trying to bring a consensus among multiple research groups working independently of each other, this schema accommodates the needs of the Materials Science community in an inclusive way.

The result we presented provides a missing *middle layer* in data management, which fills the gap between high-level core schemas and extremely detailed discipline-specific ontologies. To ensure reusability, reproducibility and interoperability, we adopted already existing terminologies to a great extent, as reported in Sections 2 and 3.

As a matter of fact, a standard way to describe measurement techniques in Materials Science is strongly needed. We suggest that our metadata schema may act as a first de-facto standard in this direction. To improve its visibility and promote its adoption, we plan to register it to the RDA Metadata Standard Catalogue [35].

As a next step, we aim at extending the structure of this schema to other characterization techniques such as Transmission Electron Microscopy (TEM), Scanning Transmission Electron Microscopy (STEM), Atomic Force Microscopy (AFM), Scanning Tunneling Microscopy (STM), Secondary Ion Mass Spectrometry (SIMS) and Atom Probe Tomography (APT), which are offered in the NEP [38] and MDMC [22] catalogues. This will pave the way to a broad variety of new applications in materials characterization, which would be difficult to manage and to automate otherwise.

Our work on metadata schemas is continuing, as well as the activity to receive feedback and create consensus in the Materials Science community. It is our intention to improve and maintain this schema, by integrating the use of the metadata to be compliant with new recommendations or standards which may be established, e.g. from RDA [42] or NIST [39].

Acknowledgements. This work has been supported by the Joint Lab "Integrated Model and Data Driven Materials Characterization" (MDMC), the research programs "Engineering Digital Futures" and "Materials System Engineering" of the Helmholtz Association of German Research Centers and the Helmholtz Metadata Collaboration Platform.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101007417 within the framework of the NFFA-Europe Pilot (NEP) Joint Activities.

A.I. and S.S. acknowledge financial support from the European Research Council through the ERC Grant Agreement No. 759419 MuDiLingo.

References

1. Becker, C., Audus, D., Bartolo, L., Brinson, C., Campbell, C., Traut, Z.: Materials registry vocabulary draft (2017), https://rd-alliance.org/groups/working-group-international-materials-resource-registries.html, last accessed: 2021/08/09
2. Botts, M., Robin, A., Hirschorn, E.: OGC SensorML: Model and XML Encoding Standard (2020), http://docs.ogc.org/is/12-000r2/12-000r2.html
3. Brandt, N., Griem, L., Herrmann, C., Schoof, E., Tosato, G., Zhao, Y., Zschumme, P., Selzer, M.: Kadi4mat: A research data infrastructure for materials science. Data Science Journal **20**(1), 8 (2021)
4. Bunakov, V., Griffin, T., Matthews, B., Cozzini, S.: Metadata for experiments in nanoscience foundries. In: Kalinichenko, L., Kuznetsov, S.O., Manolopoulos, Y. (eds.) Data Analytics and Management in Data Intensive Domains. pp. 248–262. Springer International Publishing, Cham (2017)
5. Cheung, K., Drennan, J., Hunter, J.: Towards an ontology for data-driven discovery of new materials. In: AAAI Spring Symposium: Semantic Scientific Knowledge Integration. pp. 9–14 (2008)
6. Cheung, K., Hunter, J., Drennan, J.: Matseek: An ontology-based federated search interface for materials scientists. IEEE Intelligent Systems **24**(1), 47–56 (2009)
7. Crossref schema version 5.0 (2020), https://gitlab.com/crossref/schema/-/blob/5.0/schemas/common5.0.xsd, last accessed: 2021/08/09
8. Crystallographic Information Framework, https://www.iucr.org/resources/cif, last accessed: 2021/08/09
9. International DOI Foundation, https://www.doi.org/, last accessed: 2021/08/09
10. Dublin Core Metadata Initiative (DCMI): DCMI Metadata Terms, https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, last accessed: 2021/08/09
11. eLabFTW, https://www.elabftw.net, last accessed: 2021/08/09
12. European Materials Modelling Council (EMMC), https://emmc.info, last accessed: 2021/08/09
13. The European Materials & Modelling Ontology (EMMO), https://github.com/emmo-repo/EMMO, last accessed: 2021/08/09
14. EOSC Portal, https://eosc-portal.eu/about/eosc, last accessed: 2021/08/09
15. Erkimbaev, A.O., Zitserman, V.Y., Kobzev, G.A., Trakhtengerts, M.S.: A universal metadata system for the characterization of nanomaterials. Scientific and Technical Information Processing **42**(4), 211–222 (2015)
16. European Commission: Horizon Europe. Research and innovation funding programme until 2027, https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en, last accessed: 2021/08/09
17. European Commission, Directorate-General for Research and Innovation: Turning FAIR data into reality: final report and action plan from the European Commission expert group on FAIR data. European Commission (2018), http://publications.europa.eu/publication/manifestation_identifier/PUB_KI0618206ENN, oCLC: 1111111608
18. European Light Microscopy Initiative, https://elmi.embl.org, last accessed: 2021/08/09
19. FBbi - Biological Imaging Methods Ontology, http://cellimagelibrary.org, last accessed: 2021/08/09
20. GRID, https://www.grid.ac/, last accessed: 2021/08/09

21. Handle.net Registry, http://www.handle.net, last accessed: 2021/08/09
22. Helmholtz Association, https://www.helmholtz.de/en/research/, last accessed: 2021/08/09
23. Helmholtz-Zentrum Hereon, https://www.hereon.de/index.php.en, last accessed: 2021/08/09
24. Horizon 2020 (2013), https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020, last accessed: 2021/08/09
25. Ihsan, A., Dessì, D., Alam, M., Sack, H., Sandfeld, S.: Steps towards a dislocation ontology for crystalline materials. In: Proceeding of the Second International Workshop on Semantic Digital Twins. vol. 2887. CEUR Workshop Proceedings (2021)
26. International Materials Resource Registries Working Group, https://www.rd-alliance.org/groups/working-group-international-materials-resource-registries.html, last accessed: 2021/08/09
27. ISNI, https://isni.org/, last accessed: 2021/08/09
28. Iudin, A., Korir, P., Salavert-Torres, J., Kleywegt, G., Patwardhan, A.: EMPIAR: a public archive for raw electron microscopy image data. Nature Methods **13**, 387–388 (2016)
29. Metadata Editor. Karlsruhe Institute of Technology (KIT). Steinbuch Centre for Computing (SCC) - Data Exploitation Methods (DEM) group, https://github.com/kit-data-manager/metadata-editor, last accessed: 2021/08/09
30. Könnecke, M., Akeroyd, F.A., Bernstein, H.J., Brewster, A.S., Campbell, S.I., Clausen, B., Cottrell, S., Hoffmann, J.U., Jemian, P.R., Männicke, D., Osborn, R., Peterson, P.F., Richter, T., Suzuki, J., Watts, B., Wintersberger, E., Wuttke, J.: The NeXus data format. Journal of Applied Crystallography **48**(1), 301–305 (2015)
31. Materials Resource Registry, https://materials.registry.nist.gov, last accessed: 2021/08/09
32. MatML schema, http://www.matml.org/schema.htm, last accessed: 2021/08/09
33. Matthews, B., Sufi, S., Flannery, D., Lerusse, L., Griffin, T., Gleaves, M., Kleese, K.: Using a Core Scientific Metadata Model in Large-Scale Facilities. International Journal of Digital Curation **5**(11), 106–118 (2010)
34. Metadata Schemas for Materials Science, https://github.com/kit-data-manager/Metadata-Schemas-for-Materials-Science, last accessed: 2021/08/11
35. RDA Metadata Standard Catalogue, https://msc.datamanager.kit.edu, last accessed: 2021/08/09
36. Metadata Working Group DataCite: Datacite metadata schema documentation for the publication and citation of research data and other research outputs. version 4.4 (2021), https://schema.datacite.org/meta/kernel-4.4/, last accessed: 2021/08/09
37. the Description of Nanomaterials, C.V.W.G.O., Rumble, J.: Uniform Description System for Materials on the Nanoscale, Version 2.0 (2016). https://doi.org/10.5281/zenodo.56720, https://doi.org/10.5281/zenodo.56720
38. NFFA-Europe Pilot (2021), https://www.nffa.eu, last accessed: 2021/08/09
39. NIST - National Institute of Standard and Technology, https://www.nist.gov, last accessed: 2021/08/09
40. The Open Microscopy Environment, https://www.openmicroscopy.org, last accessed: 2021/08/09
41. Documentation of the Physical Information File (PIF) schema, http://citrineinformatics.github.io/pif-documentation/index.html, last accessed: 2021/08/09

42. Research Data Alliance, https://rd-alliance.org, last accessed: 2021/08/09
43. Research Organization Registry (ROR), https://ror.org/, last accessed: 2021/08/09
44. Sarkans, U., Chiu, W., et al.: REMBI: Recommended Metadata for Biological Images - enabling reuse of microscopy data in biology. Nature Methods (2021)
45. Schema.org, https://schema.org, last accessed: 2021/08/09
46. Schema.org Extensions, https://schema.org/docs/extension.html, last accessed: 2021/08/09
47. AWI Sensor Information System, https://sensor.awi.de, last accessed: 2021/08/09
48. Stocker, M., Darroch, L., Krahl, R., Habermann, T., Devaraju, A., Schwardmann, U., D'Onofrio, C., Häggström, I.: Persistent Identification of Instruments. Data Science Journal **19**(11), 18 (2020)
49. Sufi, S., Matthews, B.: A Metadata Model for the Discovery and Exploitation of Scientific Studies. In: Talia, D., Bilas, A., Dikaiakos, M.D. (eds.) Knowledge and Data Management in GRIDs. p. 135–149. Springer US (2007)
50. Tremouilhac, P., Nguyen, A., Huang, Y., Kotov, s., Lütjohann, D.S., Hübsch, F., Jung, N., Bräse, S.: Chemotion eln: an open source electronic lab notebook for chemists in academia. Journal of Cheminformatics **9**(54) (2017)
51. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data **3** (2016)
52. Zhang, X., Hu, C., Li, H.: Semantic Query on Materials Data Based on Mapping MatML to an OWL Ontology. Data Sci. J. **8**, 1–17 (2009)
53. Zhao, H., Wang, Y., Lin, A., Hu, B., Yan, R., McCusker, J., Chen, W., McGuinness, D.L., Schadler, L., Brinson, L.C.: NanoMine schema: An extensible data representation for polymer nanocomposites. APL Materials **6**(11), 111108 (2018)