Metadata for nanotechnology: interoperability aspects

Vasily Bunakov and Brian Matthews

Published version information

Citation: V Bunakov and B Matthews. "Metadata for nanotechnology: interoperability aspects." In: E Garoufallou, S Virkus, R Siatri, D Koutsomiha (Eds.) **Metadata and Semantic Research**. MTSR 2017. Communications in Computer and Information Science, vol. 755. Springer (2017): 247-252.

DOI: 10.1007/978-3-319-70863-8_24

The final publication is available at Springer via https://doi.org/10.1007/978-3-319-70863-8_24

Metadata for nanotechnology: interoperability aspects

Vasily Bunakov[1] and Brian Matthews[1]

[1] Science and Technology Facilities Council, Harwell OX11 0QX, UK
{vasily.bunakov, brian.matthews}@stfc.ac.uk

Abstract. The work outlines the landscape of emerging metadata models for nanotechnology. A gap analysis and possible cross-walks for a few metadata recommendations are presented. The role of interoperability in the design of metadata for nanotechnology is discussed.

Keywords: nanotechnology, metadata models, metadata interoperability

1 Introduction

Nanotechnology is no more a futuristic vision but established a permanent presence in the economy [1]. The fast pace of nanotechnology innovation blurs boundaries between research and industry; another noticeable trend is an intensive use of computer simulations for modeling nanomaterials which requires computation on the industrial scale.

Nanotechnology research and innovation deal with substantial amounts of data of all sorts. It may be data about nano-materials (either physical or computer-simulated) or it may be contextual data that is important for the research or industrial management, for complying with health and safety regulations, as well as for keeping records about the provenance of materials and products. Managing these various data requires good metadata which presents challenges both from a metadata design perspective and from an operational perspective of metadata quality and its semantic interoperability.

This work presents the effort of the Nanostructures Foundries and Fine Analysis (NFFA-EUROPE) project [2] on metadata design and, specifically, on the relation of this effort to other metadata initiatives. The main purpose of this work is to reflect on the position of the NFFA metadata model in a bigger landscape of metadata for nanotechnology and identify possible connections of the NFFA model to the elements of this landscape that can be further discussed with the respective projects and initiatives.

We first outline the design of the NFFA metadata model [3], then introduce other metadata models for nanotechnology with possible cross-walks between each of them and the NFFA model, then discuss the strengths of each metadata model and refer to the auxiliary collaborative effort that can contribute to the design of quality metadata for nanotechnology.

2 NFFA metadata model

The Nanostructures Foundries and Fine Analysis (NFFA-EUROPE) project [2] brings together European research laboratories with the aim to provide seamless access to experimental equipment and computation for nanoscience researchers across the borders. The project organizational and IT infrastructure offers a single entry point for research proposals, and a common platform for the access to data resulting from the research. Both physical and computational experiments are intentionally in scope, as they often complement each other.

A novel metadata model has been developed in NFFA as a part of a Joint Research Activity that unites, on one hand, organizations that develop an IT infrastructure and on the other hand, the project partners who are involved in the actual running of physical or computational experiments and therefore can supply their requirements for metadata design and then validate the resulted model in their operational environments.

The collaborative and multi-aspect process of metadata design in NFFA involved the development of a common vocabulary in the first place which served as a cornerstone for the loosely-coupled but semantically unambiguous enterprise architecture with the inclusion of both technological and organizational aspects of the project. The vocabulary was designed and validated with multiple stakeholders in order to ensure that common concepts can cover both physical and computational experiments.

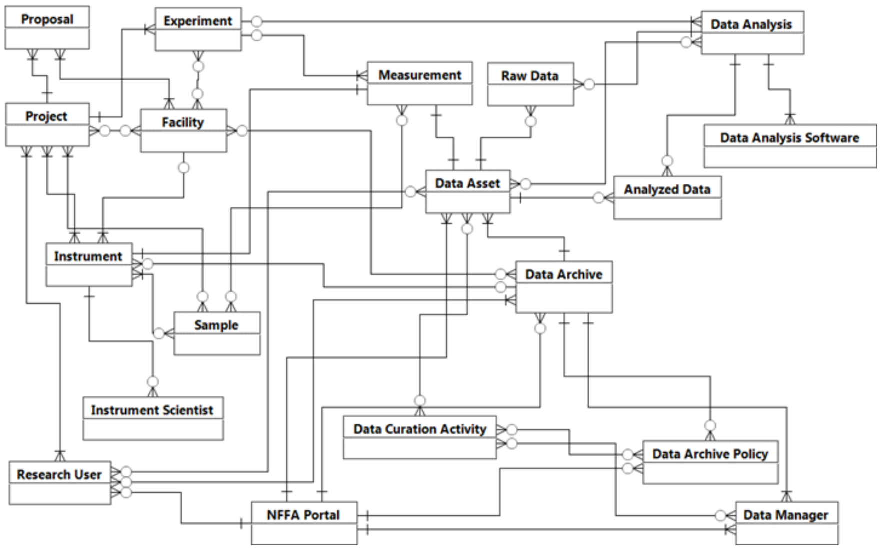

Fig. 1. Entity-relationship diagram for NFFA metadata.

The common vocabulary was then used for the definition of information entities with clear semantic boundaries, and for the definition of relationships between them. The resulted entity-relationship diagram is presented in Figure 1.

Such entities as Proposal, Project, Facility, Research User or Instrument Scientist represent organizational aspects of nanoscience experiments. Such entities as Experiment, Measurement, Instrument and Sample relate to the actual conduct of the experiment. Other entities represent data management, data curation and data analysis aspects of the experiment. These three interconnected aspects: organizational, experimental and data-related are conceptual pillars of the NFFA metadata model with its main purpose to capture all significant aspects of the experiment so that data stored in the common archive could be given a rich context. This contextual richness is important for sensible data reuse and data interoperability for a variety of business cases, as an example when the same user conducts a few experiments across different facilities then wants to combine the data for a common analysis.

Overall the NFFA data model is aimed at a contextually rich description of nanoscience experiments lifecycle, with data management and data analysis considered to be essential parts of this lifecycle. The model is designed to reflect on both physical and computational experiments and on data resulted from them. The more detailed description of the model and its design considerations is given in [3]. Based on this model, various serializations (metadata formats) can be developed for their implementations in particular IT platforms.

3 Other metadata models and semantic assets for nanotechnology

A significant effort has been made by CODATA [6] and VAMAS [7] who established a joint working group for the development of a uniform description system for nanomaterials. The group was international, also multi-discipline with the inclusion of representatives from physics, chemistry, pharmacology, ecology, engineering and other branches of research and technology. Through a number of workshops, the working group developed an elaborated recommendation [4] which, similarly to the NFFA metadata model [3], does not specify a data format but rather presents a structure of concepts that can be applicable for developing data formats and ontologies, for reporting research results, and for other practical uses.

The main focus of the CODATA-VAMAS model is on a nano-object with the metadata categories (sections) for the description of the object shape, size, physical structure, chemical composition, crystallographic structure and surface description. The model also pays attention to characterization of a collection of nano-objects with the captured concepts of a collection composition, size distribution, association type and topology. The model attempts to address the problem of the nano-objects production and testing, too, describing typical steps involved in those processes.

Cross-walks between NFFA and CODATA-VAMAS models are possible using three NFFA model entities: Sample, Experiment and Measurement. Sample can be related to nano-objects and collections of them in the CODATA-VAMAS model, Experiment can be related to nano-object production steps and Measurement to testing steps. To enable metadata cross-walks, either the respective entities of the NFFA model can be developed and presented as containers that include metadata definitions from CODATA-VAMAS model, or alternatively, these entities can serve as wrappers with pointers to the uniquely identifiable instances defined by the use of CODATA-VAMAS model.

Another prominent effort has been made by NOMAD (NOvel MAterials Discovery) Laboratory, a European Centre of Excellence (CoE) [8] and is focused on modelling the computation for nanoscience. NOMAD maintains a large repository of input and output files for computer-simulated materials, and has developed metadata for it [5]. Unlike NFFA where metadata model has been derived through rounds of communication with nanoscience practitioners and IT architects but overall has been designed in a top-down manner with the expectation that the model can be adopted beyond research communities who initially contributed to it, the NOMAD approach to metadata design is quite different and can be called opportunistic, as metadata elements are defined looking into the actual results of computational experiments. The NOMAD call this *a posteriori* approach with the main advantage of it that all significant properties of data can be captured; a few hundred metadata elements have been defined this way.

In order to implement this opportunistic or *a posteriori* approach, NOMAD construct the names of metadata elements on-the-fly depending on the concepts discovered in the results (data output) of a particular computational experiment. In addition to these metadata elements that can be called “topical keys”, e.g. “energy_total_potential” name is a key for the corresponding data value, NOMAD consider a hierarchy of descriptors for the runs (executions) of a computer program that are related to particular software configurations, to the results of computation, as well as to theoretical methods used. This gives very context-rich descriptions of the computations actually performed.

Cross-walks between NFFA model and NOMAD one are possible via the NFFA Experiment entity that can relate (again, as a container or a wrapper with references to external definitions) to NOMAD “topical keys” that describe a particular experiment, as well as NFFA Measurement entity that can be related to the NOMAD definitions of program runs. The Sample entity of the NFFA model can relate to input data in NOMAD, and Data Asset to the output of NOMAD computation.

The identified cross-walks across the mentioned metadata models are compiled in the Table 1.

Table 1. Cross-walks across NFFA, CODATA-VAMAS and NOMAD metadata models.

NFFA concept	CODATA-VAMAS concept	NOMAD concept
Experiment	Nano-object production steps	Series of s/w runs
Measurement	Nano-object testing steps	S/w run
Sample	Nano-object or collection of objects	Input data
Data Asset		Output data

It is noticeable that CODATA-VAMAS model is heavily focused on the description of Samples (nano-objects) and the processes directly related to Samples such as production or testing steps but it does not care about data involved in the Experiments or Measurements.

On the opposite, the NOMAD model cares a lot about data; especially about Data Assets resulted from the computational experiment as this data is a source for the extraction of key-value metadata pairs in NOMAD.

Neither CODATA-VAMAS nor NOMAD care much about the organizational environment where experiment or production is conducted, whilst the NFFA model pays a detailed attention to such environment with a few entities like Facility, Proposal or Project catering for this. The data lifecycle in the archive has decent means of its description in the NFFA model but not as such in CODATA-VAMAS or NOMAD model.

Overall, the three models have some overlaps which make it possible to specify the above mentioned crosswalks but otherwise the models are complementary to each other. The levels of coverage of a few key aspects of nanotechnology by the three models is presented in Table 2 with the following gradations: Conceptual coverage (where there is at least one concept that can be potentially expanded), Detailed coverage (where there is enough interconnected concepts to cover the aspect) and Unadressed (when there is nothing or very little in the model to address the aspect).

Table 2**.** Conceptual coverage of nanotechnology experiments by NFFA, CODATA-VAMAS and NOMAD metadata models.

Nanotechnology aspect	NFFA model	CODATA-VAMAS model	NOMAD model
Nano-object	Conceptual	Detailed	Detailed *
Computation	Detailed **	Unaddressed	Detailed
Experiment lifecycle	Detailed	Conceptual	Conceptual
Data lifecycle	Detailed	Unaddressed	Conceptual

*) For in silico (computer simulated) nano-objects only but key-value metadata pairs are potentially applicable to physical objects, too.

**) As all NFFA model concepts are formulated in view of their dual application to physical and computational experiments.

Apart from the NFFA, CODATA-VAMAS and NOMAD metadata models, other semantic assets such as vocabularies and ontologies can be used to complement or augment the meaning of metadata concepts or their particular attributes. To better address the metadata interoperability challenge, the collaboratively developed semantic assets should be given a preference before the industry-led specifications. The Interest and Working Groups of the Research Data Alliance [9] can be the right forums for such collaborations, with the particularly relevant effort of RDA/CODATA Materials Data, Infrastructure & Interoperability Interest Group [10] and RDA Materials Registry Working Group [11].

4 Conclusion

What we have considered in this work is the problem of interoperability of the NFFA metadata model with other prominent metadata models for nanotechnology that are not immediately represented in the NFFA project but can contribute to quality metadata design.

We look forward to fruitful discussions with the CODATA-VAMAS and NOMAD communities, as well as with other metadata practitioners in nanotechnology and material science who can contribute to the design of interoperable metadata.

5 Acknowledgements

This work is supported by Horizon 2020 NFFA-Europe project www.nffa.eu under grant agreement no. 654360.

References

1. Vance, M.E., Kuiken, T., Vejerano, E.P., McGinnis, S.P., Hochella, M.F., Jr., Rejeski, D. and Hull, M.S. Nanotechnology in the real world: Redeveloping the nanomaterial consumer products inventory. Beilstein Journal of Nanotechnology, 6, 1769-1780 (2015). http://dx.doi.org/10.3762/bjnano.6.181
2. Nanostructures Foundries and Fine Analysis (NFFA-EUROPE) project. http://www.nffa.eu
3. Bunakov V., Griffin T., Matthews B., Cozzini S. Metadata for Experiments in Nanoscience Foundries. In: Kalinichenko L., Kuznetsov S., Manolopoulos Y. (eds) Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2016. Communications in Computer and Information Science, vol 706. Springer, Cham (2017). doi:10.1007/978-3-319-57135-5_18
4. Uniform Description System for Materials on the Nanoscale, Version 2.0. CODATA-VAMAS Working Group On the Description of Nanomaterials. 25 May 2016. doi:0.5281/zenodo.56720
5. NOMAD Meta Info. https://metainfo.nomad-coe.eu/nomadmetainfo_public/info.html
6. CODATA: Committee on Data for Science and Technology. http://www.codata.org/
7. VAMAS: Versailles Project on Advanced Materials and Standards. http://www.vamas.org/
8. NOMAD (NOvel MAterials Discovery) Laboratory, a European Centre of Excellence (CoE). https://nomad-coe.eu/
9. Research Data Alliance. https://www.rd-alliance.org
10. RDA/CODATA Materials Data, Infrastructure & Interoperability Interest Group. https://www.rd-alliance.org/groups/rdacodata-materials-data-infrastructure-interoperability-ig.html
11. RDA Materials Registry Working Group. https://www.rd-alliance.org/group/international-materials-resource-registries-wg/post/rda-materials-registry-working-group