The European Materials Modelling Council

Venue

St John's Innovation Centre

Cowley Road, CB4 0WS Milton

United Kingdom

This event has received funding via the EMMC-CSA project from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 723867

Sponsors

The National Centers of Competence in Research (NCCR) are a research instrument of the Swiss National Science Foundation

Responsible for content:
Gerhard Goldbeck, Ernst-Dieter Janotka

EMMC-CSA

The European Materials Modelling Council

is organising the

EMMC

Workshop on Interoperability in Materials Modelling

November 7-8, 2017

Cambridge / United Kingdom

Venue

St John's Innovation Centre

Cowley Road, CB4 0WS Milton

United Kingdom

Purpose and objective of the workshop

The purpose is to discuss recent developments in interoperability approaches in materials modelling and related fields, following on from discussions at the First EMMC International Workshop (Notes from that event will be available to workshop delegates). In particular, the workshop will focus on semantic interoperability based on a future European Materials Modelling Ontology (EMMO). Definitions of some of the terms in such an ontology are the subject of a CEN Workshop Agreement.

Communication standards between models and databases will also be discussed, including initial requirements for cataloguing simulations in data repositories, and general requirements for Translation and Training components with a view to integration into future Materials Modelling Marketplaces.

EMMC is seeking support of the wider materials modelling community for the development of a European Materials Modelling Ontology as a basis for interoperability and domain specific metadata.

Participants

Representatives from the academic and industrial materials modelling community covering different types of models and applications, database repository owners and project representatives. The workshop is limited to 50-60 experts.

Organising team

Gerhard Goldbeck (gerhard@goldbeck-consulting.com)
Goldbeck Consulting / EMMC-CSA

Adham Hashibon (adham.hashibon@iwm.fraunhofer.de)
Fraunhofer IWM / EMMC-CSA

Alexandra Simperler (alex@simperler-consulting.com)
Goldbeck Consulting

Ernst-Dieter Janotka (ernst-dieter.janotka@tuwien.ac.at)
TU Wien / EMMC-CSA

Agenda

TUE, Nov 7, 2017

Arrival, Refreshments 10:00-10:30

Introduction 10:30-10:40

Introduction to the Workshop

Gerhard Goldbeck (Goldbeck Consulting Ltd) and
Adham Hashibon (Fraunhofer IWM)

Session 1

STATUS AND REQUIREMENTS FOR INTEROPERABILITY

Talk 1 10:40-11:00

Data and modelling integration at Dow

Hein Koelman (Dow Chemical)

Talk 2 11:00-11:20

Materials Modelling and Interoperability – Siemens PLM Vision

Stijn Donders (Siemens PLM)

Talk 3 11:20-11:40

Ontology requirements for software realisation

Wolfgang Wenzel (KIT and Nanomatch)

Talk 4 11:40-12:10

Interoperability approaches and implementations in current EU Projects

Borek Patzak (Czech Technical University; CompoSelector Project), Adham Hashibon (Fraunhofer IWM, FORCE Project), Jesper Friis (SINTEF, NanoSim Project)

Discussion 12:10-12:30

Lunch 12:30-13:30

TUE, Nov 7, 2017

Session 2

ONTOLOGIES FOR INTEROPERABILITY

Talk 1 13:30-14:00

Introduction to Basic Formal Ontology and Industry Ontologies Foundry

Barry Smith (University of Buffalo)

Talk 2 14:00-14:30

Big Data transforms into Big Analysis: the convergence of Formal Semantics & Data Science in Life Sciences

Heiner Oberkampf (Osthus)

Talk 3 14:30-14:50

Ontologies and rule-based knowledge in Knowledge-Driven Optimisation

Piotr Maciol (AGH University of Science and Technology)

Talk 4 14:50-15:10

Using ontology to augment measurements data with physico-chemical simulation for industrial application

Amit Bhave (CMCL Innovations)

Discussion 15:10-15:30

Tea/Coffee Break 15:30-16:00

Talk 5 16:00-16:30

European Materials Modelling Ontology (EMMO)

Emanuele Ghedini (University of Bologna), Adham Hashibon (Fraunhofer IWM), Jesper Friis (SINTEF), Gerhard Goldbeck (GCL), Georg Schmitz (ACCESS), Anne de Baas (EC DG RTD NMBP)

Interoperability Discussion and Action planning 16:30-17:30

Moderator: **Gerhard Goldbeck**

Dinner at Hilton Hotel Cambridge 19:30-21:30

WED, Nov 8, 2017

Session 3

DATA AND DOCUMENTATION

Talk 1 09:00-09:30

Materials Modelling Data and Documentation:
terminology, classification and ontology towards Digital Single Market

Anne de Baas (EC DG RTD NMBP)

Talk 2 09:30-10:00

Simulation documentation with Materials Modelling data tables (MODA): portal demo

Adham Hashibon (Fraunhofer IWM)

Talk 3 10:00-10:30

NOMAD Metadata for all

Fawzi Mohamed (Fritz-Haber-Institut and NOMAD Project)

Tea / Coffee break 10:30-11:00

Session 4

MATERIALS MODELLING MARKETPLACES

Talk 1 11:00-11:30

Workflows and data integration, vision and sustainability

Nicola Marzari (EPFL)

Talk 2 11:30-12:00

On system thinking, knowledge synthesis and data-driven analytics

Guido Smits (DataStories Int.)

WED, Nov 8, 2017

Talk 3 12:00-12:30

European Materials Modelling Marketplaces

Welchy Leite Cavalcanti (Fraunhofer IFAM, VIMMP Project), Adham Hashibon (Fraunhofer IWM, MarketPlace Project), Gerhard Goldbeck (GCL), Nicola Marzari (EPFL, MaterialsCloud), Sergio Lopez Lopez (SCM, Fortissimo Project)

Lunch 12:30-13:30

Talk 4 13:30-14:00

Connecting to infrastructure

Jörg Meyer (Steinbuch Centre for Computing, KIT)

Talk 5 14:00-14:30

Building a materials modelling marketplace: challenges for SME's and research organisations

Didrik Pinte (Enthought)

Panel discussion 14:30-15:10

Materials Modelling Marketplaces including ontology, repository, workflow management, curation and sustainability

David Cebon, Fawzi Mohamed, Hein Koelman, Welchy Leite Cavalcanti, Nicola Marzari, Adham Hashibon, Sergio Lopez Lopez

Chair: **Anne de Baas**

Tea/Coffee Break 15:10-15:30

Marketplaces Discussion and Action planning 15:30-16:00

Moderator: **Adham Hashibon**

Closing remarks 16:00-16:15

Gerhard Goldbeck, Adham Hashibon, Anne de Baas

Delegates

First Name	Last name	Affiliation	Country
Doaa	Altarawy	Virginia Tech - The Molecular Sciences Software Institute	USA
Shahriar	Amini	SINTEF Materials and Chemistry	Norway
Eva	Augenstein	Fraunhofer Institute for Mechanics of Materials IWM	Germany
Bijish	Babu	Swerea MEFOS	Sweden
Salim	Belouettar	Luxembourg Institute of Science and Technology - LIST	Luxembourg
Luca	Bergamasco	Politecnico di Torino	Italy
Andrea	Berto	Granta Design Limited	United Kingdom
Amit	Bhave	CMCL Innovations	United Kingdom
Nicola	Bianco	CMCL Innovations	United Kingdom
Ludovic	Briquet	Johnson Matthey	United Kingdom
Christopher	Broderick	Tyndall National Institute	Ireland
Vincenzo	Carravetta	Istituto per i Processi Chimico-Fisici - IPCF	Italy
David	Cebon	Granta Design Limited	United Kingdom
Gareth	Conduit	Cambridge University	United Kingdom
Anne	de Baas	European Commission	Belgium
Stijn	Donders	Siemens PLM	Belgium
Donna	Dykeman	Granta Design Limited	United Kingdom
Ian	Egby	Dassault Systèmes Biovia	United Kingdom
Omer	El Fakir	Imperial College	United Kingdom
Mads	Engelund	ESPEEM	Luxembourg
Hans	Fraaije	CULGI	United Kingdom
Jesper	Friis	SINTEF Materials and Chemistry	Norway
Jacob	Gavartin	Schroedinger	United Kingdom
Emanuele	Ghedini	University of Bologna	Italy
James	Goddin	Granta Design Limited	United Kingdom
Gerhard	Goldbeck	Goldbeck Consulting Limited	United Kingdom
Thomas	Hagelien	SINTEF Materials and Chemistry	Norway
Felix	Hanke	Dassault Systèmes Biovia	United Kingdom
Adham	Hashibon	Fraunhofer Institute for Mechanics of Materials IWM	Germany

First Name	Last name	Affiliation	Country
Kersti	Hermansson	Uppsala University	Sweden
Barbara	Holtz	Dassault Systèmes Biovia	United Kingdom
Sigve	Karolius	Norwegian University of Science and Technology	Norway
Christoph	Kloss	DCS Computing	Austria
Hein	Koelmann	Dow Benelux B.V.	Belgium
Natalia	Konchakova	Helmholtz-Zentrum Geesthacht - Centre for Materials and Coastal Research	Germany
Ivan	Kondov	Karlsruhe Institute of Technology - KIT	Germany
Ronan	Le Goff	Centre Technique Industrielde la Plasturgie et des Composites	France
Welchy	Leite Cavalcanti	Fraunhofer Institute for Manufacturing Technology and Advanced Materials IFAM	Germany
Lars-Erik	Lindgren	Luleå University of Technology	Sweden
Sergio	Lopez	Scientific Computing & Modelling NV	The Netherlands
Piotr	Macioł	AGH University of Science and Technology	Poland
Sami	Majaniemi	VTT Technical Research Centre of Finland Ltd	Finland
Nicola	Marzari	École polytechnique fédérale de Lausanne	Switzerland
Jörg	Meyer	Steinbuch Centre for Computing - KIT	Germany
Fawzi	Mohamed	Max-Planck-Gesellschaft / Fritz-Haber-Institut	Germany
Felipe	Murphy Armando	Tyndall National Institute	Ireland
Jessica	Nash	Virginia Tech - The Molecular Sciences Software Institute	USA
Heiner	Oberkampf	OSTHUS	Germany
Borek	Patzak	Czech Technical University	Czech Republic
Mitev	Pavlin	Uppsala University	Sweden
Mike	Payne	University of Cambridge	United Kingdom
Ben	Pellegrini	Intellegens	United Kingdom
Didrik	Pinte	Enthought	United Kingdom
Heinz	Preisig	Norwegian University of Science and Technology	Norway
Bart	Rijksen	Dow Benelux B.V.	The Netherlands
Kevin	Roberts	University of Leeds	United Kingdom
Struan	Robertson	Dassault Systèmes Biovia	United Kingdom
Fabio	Sacchoni	Tiberlab	Italy
Peter	Schiffels	Fraunhofer Institute for Manufacturing Technology and Advanced Materials IFAM	Germany

First Name	Last name	Affiliation	Country
Georg J.	Schmitz	Access e.V.	Germany
Alexandra	Simperler	Goldbeck Consulting Limited	United Kingdom
Guido	Smits	DataStories International NV	Belgium
Christopher	Thompson	IBM Research - Daresbury Laboratory	United Kingdom
Ilian	Todorov	Science & Technology Facilities Council - STFC	United Kingdom
Amangeldi	Torayev	Université de Picardie Jules Verne	France
Jeremy	Turvey	Dassault Systèmes Biovia	United Kingdom
Martin	Uhrin	École polytechnique fédérale de Lausanne	Switzerland
Wolfgang	Wenzel	Karlsruhe Institute of Technology - KIT	Germany

Discussion Notes

Topic of the workshop

This joint workshop of the EMMC "Interoperability and Integration" and "Repositories and Marketplaces" Working Groups focuses on interoperability aspects covering materials models, repositories and marketplaces.

Introduction to interoperability

All stakeholders of materials modelling face barriers regarding access to and utilisation of the wide range of models, data and software tools, as well as materials modelling information and expertise. More efficient access and effective utilisation requires interoperability between all of these parts.

Interoperability brings materials modelling communities closer together, boosts collaborative science and enables increased leverage of advances in materials modelling for the benefit of European Industry.

Interoperability is also very closely connected to digitalisation. Progress from electronic formats to truly digitalised representations enables new ways of operating, collaborating and benefitting from materials modelling.

Digital Materials Modelling Marketplaces are an example of gathering all parts and allowing them to cooperate.

Materials Modelling Marketplaces

Materials Modelling Marketplaces are digitalised systems which integrate a range of tangible and intangible components to support innovation based on materials modelling. Marketplaces utilise web-based platforms in order to link various materials modelling activities including repositories, modelling workflows, simulation tools, expertise, training, translators, etc.

Existing and emerging repositories and materials modelling marketplaces call for additional actions ensuring coherency and efficiency of information management and exchange. In particular linking various marketplaces and data repositories requires interoperability to facilitate common unified access and retrieval of data and information.

There is also a need for efficient and lean management and curation of data and knowledge across different platforms. These in turn pose additional requirements for deep interoperability that go beyond models, reaching out into data and information management in general.

Different types of interoperability

In general terms, interoperability is defined as the ability of making data collected in one system usable within the framework of a second system without further intervention by humans.

In our context, we define interoperability as the ability of a model (either physics or data based model) to make use of information generated by a simulation with another model or from a database (see Review of Materials Modelling Version 6).

While interoperability refers in general only to the *exchange of information* without reference to *how this is implemented*, in practice there are *three main levels of interoperability realisations*, **syntactic**, **semantic**, and **cross domain** as shown schematically in Figure 1.

Syntactic interoperability means that you make your tools understand the input and output from each other. This works, but only case-by-case and is not robust to changes.

Semantic interoperability is the ability to communicate shared meaning. It provides much more flexibility, but requires means to describe the data (using metadata) that link each data element to at least a controlled shared vocabulary, and preferably an associated ontology. Typically, these vocabularies and ontologies are limited to a certain domain. Semantic interoperability enables codes to be written that can operate on various data resources and add "understanding" (semantics!) to the data, promoting them to the level of *information*.

Cross-domain interoperability refers to an even higher level of semantics and enables interoperability between domains and ontologies.

The objective of this meeting

The objective of this meeting is to discuss how semantic and cross-domain interoperability could be addressed by EMMC. First steps are already undertaken and the European Materials Modelling Ontology (EMMO) and the basic metadata schema (described below) are two ongoing efforts to address semantic and cross-domain interoperability.

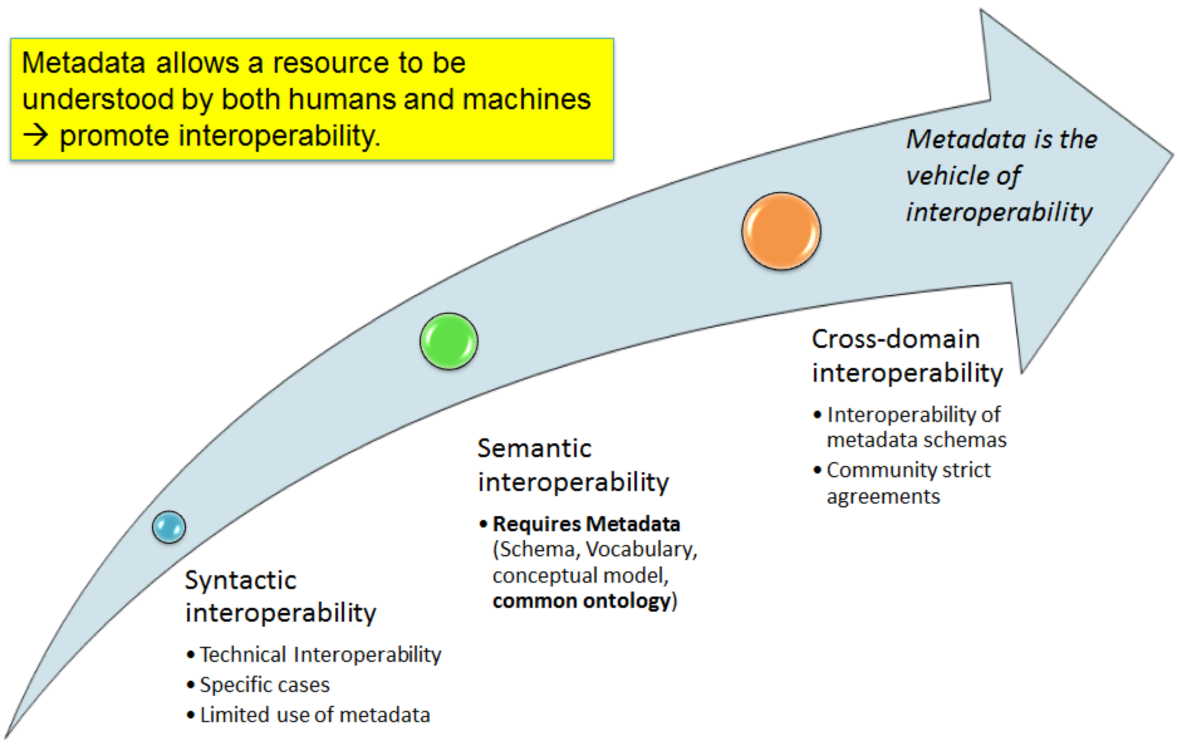

Figure 1. Levels of interoperability.

European Materials Modelling Ontology (EMMO)

EMMO is an emerging action to bring to materials modelling the same benefits that similar ontologies have brought to the fields of bio-informatics and cheminformatics.

EMMO is an ongoing effort to create an ontology for materials modelling and thereby pave the road for semantic interoperability within the field of materials modelling. The aim of EMMO is to be generic and to provide a common ground for describing materials models and data that can be adapted by all domains within EMMC.

As illustrated in Figure 2, there are several levels of structuring before reaching a proper ontology. EMMO builds further on the vocabularies and taxonomies already established in the RoMM[1] and further elaboration by the EMMC through the CEN Workshop on materials modelling terminology, classification and metadata[2].

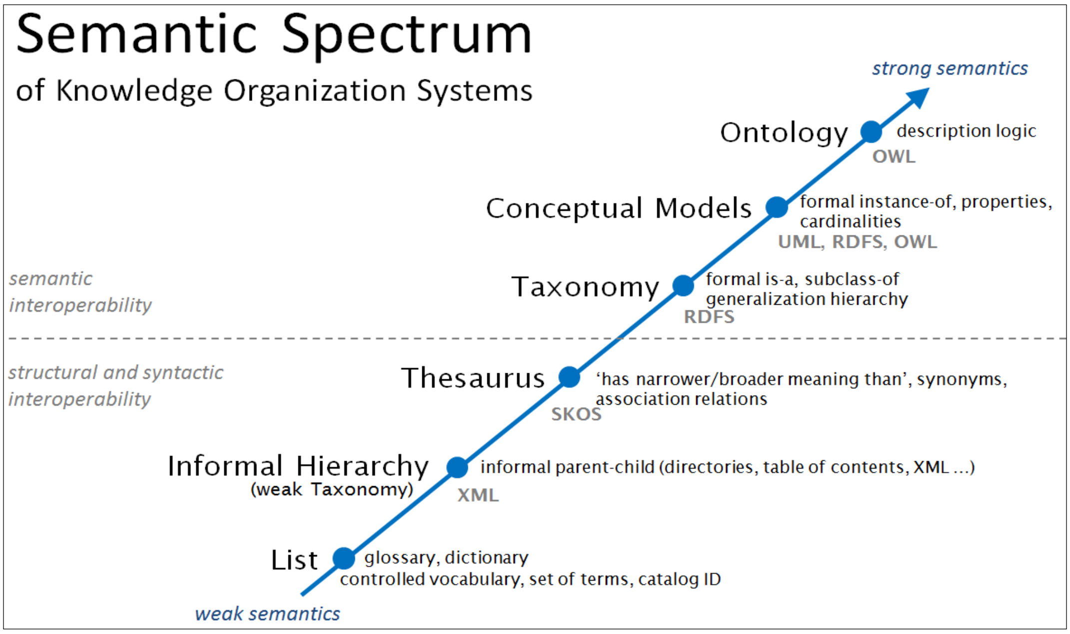

Figure 2. *Levels of semantic systems. Source: Geoff Gross, Osthus*.

The EMMO is built on top of the Basic Formal Ontology (BFO)[3]. BFO is a top-level ontology which includes the highly general representation of categories and relations common to all domains, and is built in particular for the 'material world'. Its widespread use will ensure that domain specific ontologies are built in a consistent manner and share common high-level universals.

BFO will be introduced as part of the presentation by Barry Smith.
See also http://ontology.buffalo.edu/smith/ for a wide range of resources.

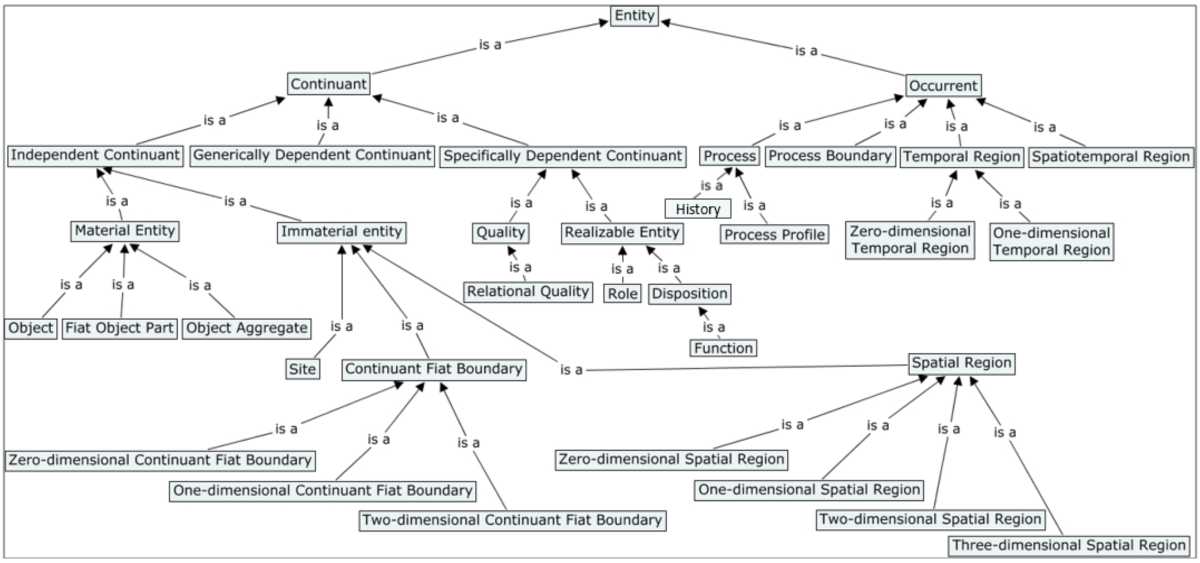

Figure 3. *Taxonomy for the Basic Formal Ontology (BFO). Source: BFO Reference*[4]

There are a wide range of ontologies based on BFO, in particular Gene Ontologies, Chemistry Ontologies etc. However, there are also ontologies of importance to our field that are not BFO based, such as OntoCAPE (chemical engineering/processing) and Engineering Materials Ontology (see: CEN CWA 16762:2014).

The top-level universals of BFO are Continuants (things that exist throughout time) and Occurrents (Processes in time) (see Figure 3). EMMO development is concerned at this stage with Continuants, which include branches covering the Material Entities that we want to model, and the types of Objects used to describe them, the physical laws underlying the physics based models, the models (physics and data based) and the mathematical representations.

Figure 4 shows the branch of the draft EMMO detailing the taxonomy for the Material Entity branch.

The 'engineered material' represents the 'User Case Material' in the MODA that is to be studied. An instance of the 'engineered material' can be a particular sample of alloys, metals, plastics, etc. When describing and modelling such a material, granularity and perspectivism are important. The EMMO includes the objects representing four granularity levels: subatomic, atomistic, mesoscopic and continuum. The modeller hence describes the 'engineered material' by means of these types of objects and/or their respective aggregates. For example, a piece of metal can be described as a solid continuum object or an atom aggregate. Not included in the view in Figure 4 are the parthood relationships. Objects can have other objects as parts. For example, a molecule object has atoms as parts.

Further development of granularity and perspectivism are issues high on the agenda to be incorporated. Also, the ontology of the knowledge issue "model" (a Generically Dependent Continuant) is on the agenda.

The current status of EMMO will be presented at the workshop.

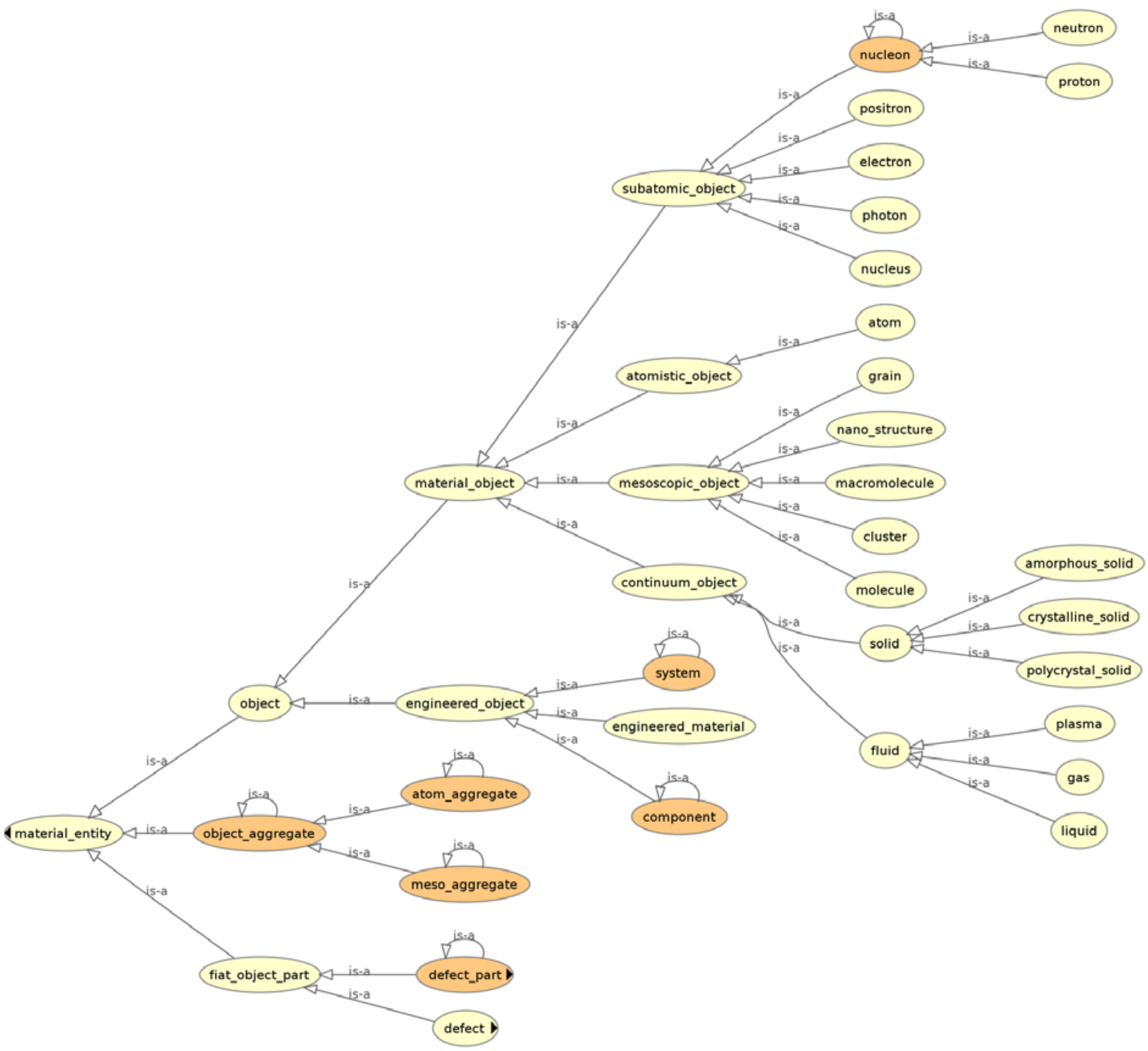

Figure 4. Branch of the **draft** EMMO, showing the taxonomy for the Material Entity branch.

Objectives of EMMO

The goal is to demonstrate the efficiency and practicality of an ontology based approach to interoperability.

- Simplified and more efficient communication across the field of materials modelling.
- Improved communication of materials modelling across other science and engineering disciplines.
- Improved communication of materials modelling to industry.
- Easier use and re-use of data from, and information about materials modelling.
- Interoperability, including between models and databases and marketplaces.
- Creating a semantic reference for the specifications of file-based schema ("data formats") such as the hierarchy and keywords in HDF5.
- Support the process of translating industrial problems into problems that can be simulated with materials models.
- Assist workflow development where several models can interoperate, including on marketplaces.
- Streamline data curation across multiple repositories with provenance and data governance models.

Discussion points and questions

- **Ontology has had a big impact on biosciences over the last decade allowing exchange and curation of data and enabling rapid innovation.**
    - How can materials sciences adopt a similar approach?
    - What are the already ongoing efforts in engineering and materials?
    - What are the low hanging fruit?
- **EMMO is designed to provide the semantic framework that software vendors can use to build interoperable solutions.**
    - What are the software owner needs?
    - What collaborations or consortia can be put in place to further the common goals?
    - What can we learn from other fields and initiatives (e.g. Allotrope)?
- **EMMO should support industry in managing and utilising models and data better as 'digital assets'.**
    - What are industry requirements?
    - How can collaborations with industry built to ensure this goal is achieved?
- **EMMO should facilitate the integration of materials modelling with other digitalisation efforts in industry.**
    - What efforts are needed to make this happen in an efficient and effective manner?
- **EMMO is a combined top-bottom and bottom-up activity, it requires commitment from the whole community to endorse it and use it.**
    - What are the potential bottlenecks and barriers for a wide endorsement and use?
    - What governance structure should be put in place of the EMMO?
- **Cross domain or semantic interoperability requires ontology, but this is usually used to annotate data, while data repositories and marketplace need to cater for the curation and transfer of data.**
    - Is this sufficient to achieve interoperability and linking between repositories?
    - Do we need an ontology of "operations, methods, and actions" on data, in terms of post, pre-processing, and accessibility? (these can be manifested, e.g., by ontology of application programming interfaces (API's))
- **Marketplaces contain much more than data: also, models, workflows and "social networking" components:**
    - Does the interoperability between marketplaces require an auxiliary ontology for workflows? Should this be part of the EMMO?
    - Are there ontologies out there already for workflows that we can use?
    - How to create and curate the mathematical representations of the model equations?

Future outlook – basic metadata schema

The EMMO aims to be an ontology that covers the domain of materials modelling, and in real applications there will be a need to connect to other domains. In order to implement cross-domain interoperability we must be able to translate between ontologies, which requires a common language for describing each ontology. The Web Ontology Language (OWL) is one such a language, which is used by Protégé to express BFO. OWL itself uses the Unified Modelling Language (UML) to define its structural elements, which is based on the Meta-Object Facility (MOF). This is a very general framework that is difficult to grasp for people not specialised in computer science. We therefore suggest creating a simplified subset of MOF, called **basic metadata schema**, with the following three structural elements:

- keyword-value pairs (properties)
- dimensions (provides arrays and lists)
- relations

In addition to means of uniquely referring to the metadata descriptions (as name, version, namespace triplets) and sane versioning of changes. We believe that these structural elements are easy to grasp and yet sufficient to describe all relevant metadata, ontologies and actual (realised) data. A formal UML diagram for the proposed basic metadata schema is shown in Figure 5.

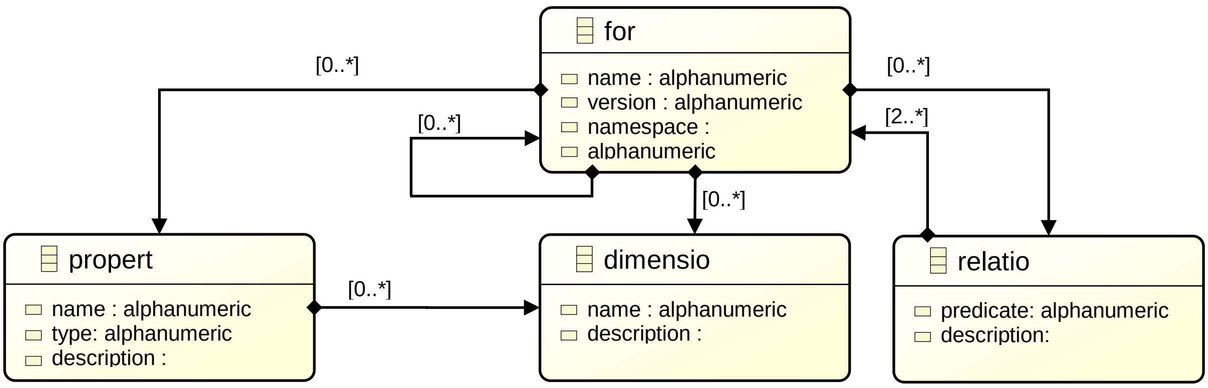

Figure 5: A formal UML diagram for the proposed basic metadata schema.

Appendix: Terminology

Metadata can be defined as “data describing the context, content and structure of records and their management through time”[5]. They provide information that allows for categorization, classification and structuring of data[6]. A well-established example is the Crystallographic Information File (CIF)[7] which provides metadata for atomistic structures and properties.

A **metadata schema** can be defined as “a logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values”[8].

A **controlled vocabulary** is a way to describe knowledge for subsequent retrieval. It mandates the use of predefined, authorized terms that have been preselected by the designers in contrast to natural language vocabularies, which have no such restriction.

A **taxonomy** is a hierarchical classification of the terms in a controlled vocabulary. It captures no complex relationships between elements, except subclass/superclass relations.

An **ontology** is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain[9]. Ontologies aim to define which entities, provided with their associated semantics, are necessary for knowledge representation in a given context[10]. Ontologies and related information technology provide an opportunity to share a common understanding of the structure of information within a specific domain, the possibility to reuse domain knowledge, to make domain assumptions explicit and to analyse domain knowledge[11].

References

1 Review of Materials Modelling. What makes a material function. DOI: 10.2777/417118.
2 https://materialsmodelling.com/2016/12/20/cen-workshop-on-materials-modelling-terminology-classification-and-metadata/.
3 R. Arp, B. Smith, A.D. Spear. Building Ontologies with Basic Formal Ontology. MIT Press 2015. ISBN: 978-0-262-52781-1.
4 BFO Reference, see https://github.com/BFO-ontology/BFO
5 ISO 15489-1 s 3.12
6 G.J. Schmitz et. al.: Towards a metadata scheme for the description of materials – the description of microstructures Science and Technology of Advanced Materials 17:1(2016) 410-430
7 http://www.dcc.ac.uk/resources/metadata-standards/cif-crystallographic-information-framework
8 Building a metadata schema – where to start. ISO/TC 46/SC11N800R1
9 https://en.wikipedia.org/wiki/Ontology_(information_science)
10 Thomas R. Gruber (1993). Toward principles for the design of ontologies used for knowledge sharing. Originally in N. Guarino and R. Poli, (Eds.), International Workshop on Formal Ontology, Padova, Italy. Revised August 1993. Published in International Journal of Human-Computer Studies, Volume 43, Issue 5-6 Nov./Dec. 1995, Pages: 907-928, special issue on the role of formal ontology in the information technology.
11 David Lamas, Metadata and Ontologies, 2011; https://www.slideshare.net/davidlamas/metadata-and-ontologies