EMMC-CSA

European Materials Modelling Council

Report

"EMMC IntOP2018"
Workshop on Interoperability in Materials Modelling

6-7 November 2018, Freiburg, Germany

TABLE OF CONTENT

This workshop and report has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 723867

1. Summary

1.1 Description of the workshop and objectives

The IntOP2018 workshop was held from 6-7 November 2018 in Freiburg. The meeting was attended by 66 delegates representing a wide range of stakeholder communities.

Market and product-oriented industry problems are complex and require the integration of skills, methodologies and data from a range of disciplines, including different types of modelling, experimentation and data analytics.

Large amounts of materials data (calculated and measured) are being generated continuously in multiple research organisations and labs in many universities across Europe and the world. Increasingly, more of such data is being stored and curated in numerous scattered repositories. Interconnected, interoperable data resources are essential for enhanced utilisation of these data in advanced materials modelling workflows.

The objective of the workshop was to bring together experts from a range of communities including different types of modelling, data science, as well as materials characterisation and manufacturing to discuss in particular the interrelated topics of marketplaces, open simulation platforms and semantic knowledge organisation (especially ontologies).

The EMMC calls for establishment of a digital marketplace as innovation hub for the advancements of materials-based industries.

Materials Modelling Marketplaces support enhanced innovation by taking advantage of recent advances in information technologies to establish web-based online collaboration platforms to explore, learn, and create advanced materials modelling workflows covering all models and domains and enabling rapid transfer of materials modelling development and expertise to industry.

The workshop addresses both key technological and organisational human capital gaps.

The EMMC is seeking support of the entire materials modelling community for the establishment of common standards for access to all online materials modelling resources including data repositories of materials properties, online modelling workflows, translation, education and training services.

An Open Simulation Platform is formulated as a set of common standards and related tools that form the basic environment on top of which compatible and compliant simulation workflows can be developed and run.

Achieving interoperability should be based on common ontologies. Ontologies also provide for concerted access to repositories of data and of services, and will also enhance interoperability between materials models and integration of services.

The workshop also served as a continuation of the EC ontology meeting (https://emmc.info/events/ec-workshop-on-materials-ontology/) and hosted a kick-off session for the International Materials Ontology Interest Group led by the EMMC.

Webpage: https://emmc.info/events/intop2018/

1.2 Major outcomes and recommendations

The major outcomes of the workshop are:

- Digitalisation: The importance of materials in product development and the significance of advances in digitalisation and a structured approach to data generation (from experiment or simulation), processing, storage and analytics was strongly emphasized. It is very important to store data in a

structured way following an ontology to be able to apply some rules by which a decision can be made. It may take years to store data to such an extent that inference becomes a possibility. Only, if data can be stored in a very structured way, it becomes possible to find missing data or incoherent data.

- Industry Requirements: Industry presentations highlighted the benefits and need for interoperability, semantic data technologies, open simulation platforms and marketplaces as potential pre-competitive collaboration/prototyping environments. In particular, interoperability is key in the development and application of industrially relevant modelling workflows, the need for improved data transfer in particular between collaborating partners (based on open standards), the opportunities of virtual input structure generation and/or microstructure databases to enhance modelling and experimental workflows, data science/semantic technology-based data/information gathering.
- The marketplace concept has been well received and understood by the community. Key aspects of marketplaces were discussed or touched upon in a range of talks, during discussions and a panel session. These included
    - Translation: These services should be provided by persons who understand industrial problem in-depth, are good interpreters and can support non-expert users. These persons should be able to improve models/initiate model improvement and also validate software.
    - Licensing: from an SWO perspective, they would require licensing possibilities with the end-users in case these do not want not go via the cloud. Also, SWOs have existing systems in place and may not buy-in into licensing templates provided by the marketplaces. SWO mostly sell "all or nothing" even if end-users do not require 100% to all software capabilities. Marketplaces may be the advent of commercial APPs and thus, a new way of licenses. End-users would like to see OSP solutions on the marketplaces that embrace open source and commercial software. This may require different licensing schemes. Also, SaaS licensing requires further investigation.
    - Software products on the marketplace and commercializing of software via marketplaces
    - Marketplace platforms for industry collaborations: In the "*digital age*" new collaboration models are necessary.
      A market place platform needs to establish standard template contracts to cater for such collaborations, but these are difficult to negotiate.
    - Ontologies to drive marketplaces were outlined: OSP, Marketplace functions, services, etc.
    - Key problems that marketplaces should address were discussed
        - *Users:* provide community building, provide *everything* one needs around a problem (people, tools, data and process), provide training and "getting started -services", licensing templates, comprehensive collections of links to computer clusters with installed software, provide standards for input/output, provide a test framework, linking of open source and commercial software
        - *Database providers:* offer the database and get people to deposit new structures/data, find a community willing to support the database
        - *Commercial Software Owners:* marketplace as a vehicle to offer apps, provide license templates for the new way of operating with XaaS, Validation activities could build trust in the software, create a database for Materials Relations, facilitate marketing for small companies, a trainer license (i.e. 3rd party persons certified by a SWO)
        - *Open Source Software Owners:* build a community that can contribute to open source development

- Market places are a key infrastructure: the market place could be seen as key infrastructure that accelerates materials science and whose infrastructure is supported by the European Commission programmes; i.e. a "building" which would host a well-maintained infrastructure including all relevant software, hardware, databases etc. and ideally a team of highly skilled staff supporting the researchers who applied for "simulation time" when temporarily using this infrastructure.
- Marketplace Business Models: Business models for the operator as well as the service providers have been suggested and discussed. Among them, paying annual fees for certain standard services and paying extra for on-demand services would be one business model for the Marketplace operator. Generally, users tend to prefer payment through a single vendor point instead of various individual providers.

- Open Simulation Platforms (OSPs) Main points and outcomes:
  - Widespread support for the concept of OSP
  - All models and databases should link as even today a large amount of data is still transferred manually
  - Needs to permit the end-users to link their own tools and generate their own workflows
  - An open simulation platform (OSP) should be formulated as a set of common standards and related tools that form the basic environment on top of which compatible and compliant simulation workflows can be developed and run.
  - As an example of a current Simulation Platforms, AiiDA includes
    - … flexible and scalable informatics infrastructure to manage, preserve, and disseminate the simulations, data, and workflows of modern-day computational science.
    - … a python package for advanced users (powered by Jupyiter) that users install on their computer, that can submit jobs, and retrieve results. It comes with a local database and reports each step to it.
    - … predefined workflow for non-expert users which they can find in the AiiDA Lab (https://aiidalab.materialscloud.org/hub/login).
    - … being code-agnostic (to honor the concept on being open)
    - … a simple ontology (not yet MODA and EMMO compliant, work in progress)
- Machine learning and data-based approaches: Main outcomes
  - *Machine learning:* data needs to be annotated and organised before training a neural network, as raw data may lead to a prediction of low quality;
  - *Data-base approaches: status quo*
    - Usually data retrieved from patents, literature and websites via text mining are unstructured
    - Ontologies can aid to define what is relevant to a material class, the properties of interest, and how things are related (numbers, physical units, …)
    - Annotating is performed and relations are extracted.
    - Curation and common file formats often limited to one specific organisation
    - Actual text mining software is often more advanced in drug discovery
  - *Data-base approaches: where next?*

        - combine data bases/sources and build a kind of data warehouse, e.g. to combine the *property* data with crystal structure data, data from ab-initio calculations, data from in-house property characterisation, … to generate a "database of *property* knowledge".
        - Curation, to ensure new data is added correctly
- Semantic Assets and Ontology Development
    - Major step forward on EMMO development
    - Importance of governance highlighted
    - Setting up of EMMO training course/sessions was requested and will be considered by EMMC for 2019.
- Formation of Interest Group proposed and mechanisms outlined. General support received and positive comments about RDA mechanism based on past experience. As a result, we will further pursue setting up the Interest Group.

2. Progress report (main activities)

2.1 The digital Marketplace – a collaboration platform of the future[1]

2.1.1 What are the biggest problems that should be solved with the marketplace?

The information on materials modelling software is very siloed and one had to browse all the different software owners' (SWO) webpages to find information. Also, SWOs tend to list the functionality of their software but do not provide workflow solutions unless the customer is willing to purchase contract research or similar services. Software users (SWU)[2] would ideally like to try software on their problem before the buy and check if it actually can do what they want and also delivers results in a suitable fashion. A marketplace should provide such easily accessible workflows, ideally in a testbed scenario. This means there should be toy use cases attached which allows the SWU to not only test but also validate the software of interest.

Many SWUs are interested in multiscale modelling and argued that for a good workflow an expert for each model was needed to link them in a sophisticated manner. They were in favour that the prototype of a multiscale workflow indeed had to be built in such a fashion, but once built they would like to see the workflow on a marketplace so it can be modified, tested and exploited. The more advanced SWUs would like to see platforms where they can link a variety of tools and build their own workflows; this requires software, wrappers, pre and post processors to be standardised and the majority of the audience was in favour. SWUs see commercial software as a vital part of their work, so they would like marketplaces to attract commercial SWOs and ideally have all of them in one place.

The biggest problem for a commercial SWO will be that they have to deal with competitors on the marketplace and their wares will be in close proximity to each other. Here the marketplace had to function as a mediator to enable a constructive side-by-side scenario. The academic SWOs or those running a SME often have no access to marketing or business development and could see themselves to use a marketplace as a vehicle for advertising their solutions.

[1] The content of this section is compiled from input received in the workshop during talks and discussions (see Annex).

[2] SWU are Industrial End Users of the marketplace.

The marketplace will be also a vehicle for services around materials modelling, and for many users it will be of interest to have a marketplace as a single-entry point to these resources and find services and also collaborations.

Many attendees were interested in data and similar to software these are also very disparate. They would like to use the marketplace as a single-entry point into a wealth of data, databases, repositories, etc. It occurred, that database owners would like to feature on a marketplace as they can find users and also persons who would like to deposit data in their respective databases.

2.1.2 What are the expectations for the marketplace to bring materials modelling to the next level?

SWU: They would like to see a provision of platforms which are interoperable where they can test their workflows. They also expect a library of workflows to reuse existing knowledge. They would like to use the market place as a faster way of networking as it can take years to find and nurture collaboration:

- *Pose an open question* to the marketplace community and pay the ones who answer.
- *Incremental development*: work with an existing partner outside the firewall.
- *Productionise a workflow*: company staff across the globe use an external model
- *Collaboration across the value chain*: different partners run the simulations.

They would like to see a variety of software on offer with a comprehensive description of its capability.

SWO: most of them compete so they may only bring their software on the marketplace if there is a good mediation available. SWOs with complex software tools will be reluctant to split software from services and would like to avoid unsolicited testing by non-expert users. However, there should be space on the marketplace to offer APPS. Academic SWOs might have issues with paying a fee for having their tools on the marketplace.

Data: Marketplace users would also like to see curated data repositories with metadata present.

2.1.3 What are the top three features for achieving the above aims?

A marketplace …

- … should have everything a modeller needs in one place (tools, data, process, people)
- … should have one standard
- … should have an easy payment system

2.1.4 What requirements are needed for the marketplace regarding security / data privacy and confidentiality / performance / data storage?

SWUs are very security aware and they would like to see the marketplace having infrastructure that enables security. They would like to see encryption mechanisms in place, possibilities to access a marketplace behind a company's firewall, and a secure version of the cloud.

Both software and services should be quality controlled. There should be some acceptance criteria software has to pass before being offered on the marketplace and a validation should be possible. Toy cases should be in place so a potential user can test the software. The same should apply to service providers as a marketplace customer would like to know about reliability and quality before one chose to work with them.

Data storage/accessibility will be important and they should be stored in a structured way. Data should be made citable so users can refer to the creator and give credit. There will be some challenges as data is heterogeneous,

has a varying quality between organizations, projects, and data sets and data sharing is culturally not appreciated and encouraged. Data processing is often non-transparent. Solutions will have to be found on the community level were people actively use existing funding channels to advance digitalization by making data fit for this purpose and use several dissemination channels (papers, summer schools) to create momentum. There must be also solutions on a political level as databases require long-term funding and the responsibility for funding of public databases is currently undefined and passed between universities, faculties, libraries, computer centers, government and the EU.

2.1.5 What must the marketplace absolutely offer that you would use it every day?

A variety of stakeholders should be present and they must have endorsement so a marketplace customer can easily decide if they wish to enlist products or services.

The marketplace should offer a library of workflows or platforms where self-made workflows can be tested.
All software packages on the marketplace should come with a good description of what they can actually do, preferably using the same taxonomy so the user can easily compare. Commercial software should be present. SWO would profit from a common taxonomy as well as "scouts" (i.e. people hired by companies in US or Japan to link them to the newest technologies) may pick out there software more easily.

Translators, consultants or similar service providers should be software agnostic and understand an industrial problem and offer good solutions.
Databases and repositories linked to a marketplace should ideally provide annotated and clean data to permit work with AI.

The marketplace users would expect an easy billing system, similar to Amazon, i.e. one pays Amazon rather than the individual product provider.

2.1.6 What are services you would pay for on a marketplace?

- Access to comprehensive databases and repositories
- Testing workflows and/or new/unfamiliar software
- Cloud access if access to hardware is an issue
- Translator services
- Training

2.2 Interoperability based on common ontologies

Standards and ontologies are very important to make data and workflows truly interoperable
ViMMP and Marketplace work on ontologies covering the market place and all its services to achieve an agreement at the high level and to achieve compatibility with EMMO.
There was strong interest in and discussion about ontologies. Since the beginning of EMMC, the community has started to embrace the concept and value of semantic technologies. A range of contemporary ontology developments in the field and several were mentioned and discussed during the workshop:

- EMMO
- European Virtual Market Place Ontology
- WWC building energy ontology
- IAO Information Artefact Ontology (IAO), used for documents
- Gene ontology
- Industrial Ontologies Foundry (IOF)

https://sites.google.com/view/industrialontologies/home

- Characterisation Ontology

The semantic spectrum (from low to high) is important to enable different level of interoperability. Many semantic assets are for a particular type of material (domains) only. Materials are very complex and all relevant features have to be captured (modelling, characterisation, manufacturing (domain specific), materials process (4 dimensional), toxicology, …). There was agreement from stakeholders who work on such domain ontologies to make them compliant with the marketplace ontologies and the EMMO. The EMMC plans to work further on the TAXONDA (Taxonomy Documentations, see https://emmc.info/taxonda-dashboard/; 11 were registered at the time of the workshop) and people were invited to contribute.

2.3 EMMO

The first near-release version of EMMO was presented and discussed. It was noted that the EMMO development has come a very long way since the IntOp 2017 workshop one year ago. The ontology has been completely re-designed and put on a new upper ontology footing. EMMO is now a four-dimensional (space-time) ontology. Materials are represented as being a space-time, which enables processes and changes to be captured quite naturally. Also, EMMO now has firm and well-defined axioms clearly distinguishing sets (set theory) and fusion (mereotopology). Hence collections (based on sets) are clearly defined, and contrasted to mereotopological parts. The direct-parthood relation already introduced at the last workshop has been established and tested as the route to capture the multi-facetted and hierarchical structure of materials.

At the latest expert meeting, EMMO: Material Representation and Applications in Cambridge in September 2018 (https://emmc.info/events/emmc-csa-emmo-irag-workshop/ it was established that EMMO has the relevant expressivity and branches to cover the application (e.g. models, properties) and what relationships in the ontology are required to represent the use case (i.e. the relations between the material, the relevant models and/or properties).

The EMMO will also provide a dictionary of standardised concepts and relations that can be the basis for interoperability as needed for marketplaces. Interoperability is superior to compatibility: Interoperability can be defined as the ability of two or more systems to exchange information between them through a common representational system. Compatibility can be defined as the ability of two or more systems to establish a one-to-one connection between them.

The information that is exchanged between systems are concepts (i.e. user case and/or model) and data. In an interoperability environment there is a mutual interdependence between concepts and data. The concepts provide the meaning for a set of data (EMMO), and data sets cannot be exchanged without a linked concept that describes their meaning.

An open simulation platform can then be defined as a metadata-based architecture for interoperability that can be derived directly upon representation of concepts in a formal ontology.

2.4 Beyond the EMMC

There is a wide agreement within the community to carry on with the EMMO and other initiatives emerging from the EMMC-CSA. An interest group (International Materials Ontology Interest Group) was launched and 42 people had joined at the time of the workshop.

The ongoing efforts should be supported and RDA was suggested as a framework.
The RDA is a worldwide organisation and would offer an existing framework to keep the effort going.

There is already a Task group on Semantic Assets for Materials and an existing Interest Group in the field of Materials Data, Infrastructure & Interoperability.
The intention is to expand the objectives of this Interest Group to include Materials Ontologies, and to launch a Working Group with a specific objective regarding materials ontologies in April 2019 at the RDA plenary meeting in Philadelphia. Also, in the manufacturing and Product Life Cycle area, there is the Industrial Ontologies Foundry IOF,

https://sites.google.com/view/industrialontologies/home. An interaction opportunity was in spring 2019 in Oslo. Other suggestions were W3C, https://www.w3.org/.

The community is called up to further discuss which supporting framework to use. An organisation is needed to keep a group going and develop specific goals. Also, one needs to decide what happens to the EMMO. EMMO provides the upper level as well as high level of branches for materials, models, properties, characterization etc. Further details within these branches will need to worked on.

The participants of the workshop are also invited to suggest topics one should work on. Use cases for the EMMO are of interest to see what people know about their materials and their granularity. More cases should be collected. The marketplaces can be assets, too, as they are ongoing at the moment and have funds. They could prove the validity of the EMMO. It may be important to plan training on the EMMO and see how to move beyond August 2019, when the EMMC-CSA ends.

3. Conclusions

Actions

- Working further on a holistic solution for marketplaces, i.e. all relevant products and services under one umbrella
- Mediation of conflict between competitors to enable a productive side-by-side existence
- Workflows are important; could be offered like many project management marketplaces offer them, such as https://marketplace.atlassian.com/
- Databases and repositories should feature on the marketplaces
- A business model for the Marketplaces needs to be worked on to enable various stakeholders to join.
- Keep working on the marketplace ontologies and draw in more domain-ontology developers.
- Think about possibilities to keep the EMMO going. (RDA, new projects/funding, forming a consortium, …)
- Need for EMMO training.

The outcome of this workshop was compiled by	Gerhard Goldbeck (Goldbeck Consulting Limited) Alexandra Simperler (Goldbeck Consulting Limited)

Workshop organisers	GCL, FRAUNHOFER
Contributing beneficiaries	ACCESS, EPFL, GRANTA, HZG

EC-Grant Agreement	723867
Project acronym	EMMC-CSA
Project title	European Materials Modelling Council - Network to capitalize on strong European position in materials modelling and to allow industry to reap the benefits
Instrument	CSA
Programme	HORIZON 2020
Client	European Commission
Start date of project	01 September 2016
Duration	36 months

Consortium		
TU WIEN	Technische Universität Wien	Austria
FRAUNHOFER	Fraunhofer Gesellschaft	Germany
GCL	Goldbeck Consulting Limited	United Kingdom
POLITO	Politecnico di Torino	Italy
UU	Uppsala Universitet	Sweden
DOW	Dow Benelux B.V.	Netherlands
EPFL	Ecole Polytechnique Federale de Lausanne	Switzerland
DPI	Dutch Polymer Institute	Netherlands
SINTEF	Stiftelsen SINTEF	Norway
ACCESS e.V.	ACCESS e.V.	Germany
HZG	Helmholtz-Zentrum Geesthacht Zentrum für Material- und Küstenforschung GMBH	Germany
MDS	Materials Design S.A.R.L	France
QW	QuantumWise A/S	Denmark
GRANTA	Granta Design LTD	United Kingdom
UOY	University of York	United Kingdom
SINTEF	SINTEF AS	Norway
UNIBO	ALMA MATER STUDIORUM – UNIVERSITA DI BOLOGNA	Italy
SYNOPSYS	Synopsys Denmark ApS	Denmark

Coordinator – Administrative information	
Project coordinator name	Nadja ADAMOVIC
Project coordinator organization name	TU WIEN
Address	TU WIEN | E366 ISAS | Gusshausstr. 27-29 | 1040 Vienna | Austria
Phone Numbers	+43 (0)699-1-923-4300
Email	nadja.adamovic@tuwien.ac.at
Project web-sites & other access points	https://emmc.info/