# EMMC-CSA

European Materials Modelling Council

Report

# Sessions on

# Interoperability and Integration

within the

# EMMC International Workshop 2019

FEB 25-27, 2019, Vienna Austria

## TABLE OF CONTENT


1. EXECUTIVE SUMMARY ........ 3
1.1 Description of the deliverable content and objectives ........ 3
1.2 Major outcome ........ 3
2. PROGRESS REPORT (MAIN ACTIVITIES) ........ 5
2.1 Beyond the Models: Applying Semantic Technologies Across the Enterprise ........ 5
2.2 Digital Transformation of Materials R&D ........ 5
2.3 Open Modelling Framework ........ 6
2.4 International Materials Ontology ........ 8
3. CONCLUSIONS ........ 10
4. ANNEX 1: REPLIES TO THE SURVEY ........ 10
4.1 Digital Transformation of Materials R&D ........ 10
*4.1.1 How can we increase European manufacturing and process industry capacity to extract knowledge from data? ........ 10*

*4.1.2 How can we support leadership by digital transformation of data into knowledge? ... 11*

*4.1.3 How can we achieve that manufacturing companies remain competitive (new products, high-added value) on the basis of knowledge extracted from wide-spanning data sets? ... 11*

*4.1.4 How can we promote European benefits from data management, standardised documentation enabling interoperability and linking tools for enriching and enhancing the usability of data? ... 12*

*4.1.5 What actions are proposed for a common semantic knowledge base in materials? ... 12*

*4.1.6 What practical ways forward are there for industry to reap the benefits of digitalisation in materials? ... 12*

4.2 Open Modelling Framework ... 13

*4.2.1 What modelling frameworks, open or otherwise are the most important ones to you and your community/organisation? ... 13*

*4.2.2 What kind of standardisation is needed to use an OSP for interoperability? ... 13*

*4.2.3 How would you want to use an OSP to fully automatise data transfer between models? ... 13*

*4.2.4 Is full flexibility with tools that can link to it important to you? ... 13*

*4.2.5 What would an OSP have to offer so that you would use it regularly? ... 13*

*4.2.6 What would an OSP have to offer so that you would enable your software to link to it? ... 14*

4.3 Session 8: International Materials Ontology ... 14

*4.3.1 Are ontologies important to you? ... 14*

*4.3.2 How can we get all relevant stakeholders on board? ... 14*

*4.3.3 What is needed to keep the international materials ontology going (RDA, new projects/funding, forming a consortium, …)? ... 14*

*4.3.4 Can we achieve “interoperability”? What does this mean to you? ... 14*

*4.3.5 What sort of training/ dissemination/ would you like to see? ... 15*

The EMMC-CSA project has received funding from the
European Union‘s Horizon 2020 research and innovation programme
under Grant Agreement No 723867

# 1. Executive summary

## *1.1 Description of the deliverable content and objectives*

At the EMMC International Workshop 2019 in Vienna, three discussion sessions and one plenary talk were organised by EMMC Working Group "Interoperability and Integration". The plenary presentation entitled "Beyond the Models: Applying Semantic Technologies Across the Enterprise" was given by Eric Little (Osthus, US). The three session topics were:

- Digital Transformation of Materials R&D (Session 2)
- Open Modelling Frameworks (Session 5)
- International Materials Ontology (Session 8)

In addition, prior to the workshop an online survey with relevant questions for each session was circulated to the participants.

The main outcomes of the plenary, discussion sessions and surveys are outlined in this report. In addition, the

the replies to the survey are summarised in Annex 1 and the Rapporteur reports of the three sessions can be found in Annex 2. Available online are the plenary presentation slides and impulse presentation slides from each of the three sessions, see https://emmc.info/emmc-international-workshop-2019-workshop-presentations/

The objective of Session 2, Digital Transformation of Materials R&D, was to explore current and emerging approaches to transform the development, processing and applications of materials in European industry. The aim is to increase European capacity to extract knowledge from materials and manufacturing data. In order to achieve that, a strategic approach to digitalisation is required involving stakeholders from materials, manufacturing and digital technologies.

The objective of Session 5, Open Modelling Framework, was to present and discuss approaches to open modelling frameworks and find common widespread support for the concept of an OSP. We need to find ways to link all models and databases openly as even today a large amount of data is still transferred manually or in non-standard means.

The objective of Session 8: International Materials Ontology, was to focus on the semantic foundations for the field of materials, including modelling, characterisation, processing and manufacturing. Many semantic assets are for a particular type of material (domains) only. Materials are very complex and all relevant features and their interrelations have to be captured. Thus, it is important to build semantic foundations for the field of materials science, including modelling, characterisation, processing and manufacturing and also find agreement with stakeholders who work on domain ontologies to make them compliant with a common framework.

## *1.2 Major outcome*

The major outcomes of the workshop are as follows.
Digital transformation is seen as one of the key enablers in materials and manufacturing for future competitiveness. Data and Models are key elements of digital transformation. However, there is a need increase the European manufacturing and process industries capacity to utilise models and extract knowledge from data. In order to achieve that, current barriers and obstacles need to be overcome. These include domain boundaries for data and models, for example that data in existing databases don't lend themselves to complex queries from materials science domain experts. Data are in siloes and there is a lack of semantic annotation and even a lack

of awareness and organisation of semantics that already exists. Data integration and semantics are seen as key. However, to embark on Digital Transformation of Materials R&D one immediately gets confronted with the task of changing the culture of Materials R&D. This change should happen with many small changes that will make a big difference in the end. Semantic knowledge bases have found their way already into some materials manufacturers and offer a proprietary but not yet a common solution. The community will have to make efforts to evidence why a “common” solution is the way forward and enter into a dialogue with all relevant stakeholders. To ease the achievement of this, we good and practical tools and techniques are needed. A common solution is to require that data should be collected in common pools as far as possible. In the plenary Eric Little proposed an alternative whereby data stay at their source, and flag any semantic information to a rich semantic layer which operates on top of the data, creating an environment for cross silo analysis.

In any case, to advance materials and manufacturing research, data and modelling resources must be combined more and better than today. This requires interoperability and the development of as well as agreement upon an ontology able to represent the relevant domains, e.g. materials at all levels of granularity, processes, characterisation, properties, models.

The key importance of ontologies in general was endorsed and the importance of ontology as a tool was shown in both the principle of the setup of EMMO, in applications and in approach to big data analysis. The need for ontology interoperability was emphasised. Awareness has been increasing and communities are starting to interact and collaborate. The three impulse sessions described a global ontology effort (Industrial Ontologies Foundry), a materials ontology effort in Japan enabling integration of heterogeneous data sources into a common framework and EMMO as an emerging ontology for the applied sciences. The importance of strong governance and sustained development was emphasised. Global collaboration on materials ontologies is being instigated under the umbrella of the Research Data Alliance.

Regarding Open Modelling Frameworks, a large number of well-established integrated simulation environments exist, including commercial closed systems, commercials systems with APIs and open source environments. However, there is a lack of truly open modelling frameworks that enable end users to integrate any modelling or pre or post processing tools on a widely agreed, open standard footing for all types of models (electronic, atomistic, mesoscopic, continuum). At the very heart of an open modelling framework is the Open Simulation Platform (OSP) which 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.

The three impulse talks show three different approaches to work with OSPs. Tata Consultancy Services (TCS) approach has been to start with an analysis of the respective industry workflows and the important “elements of knowledge”. These are conceptualised and ‘ontologised’, leading to a digital representation and hence a “meta-modelling” integration framework. Models and data can then be tied into this framework. Hence in contrast to many existing simulation frameworks, the semantics comes first.

VMAP (Virtual Material Modelling in Manufacturing, http://vmap.eu.com/) aims to gain a common understanding of, and interoperable definitions for, virtual material models in CAE. A key point is that even once there is a common understanding of concepts and metadata, efficient interoperability also requires tackling of formats and detailed standards.

SimPhoNy has integrated everything from atomistic to continuum models. It has wrappers and facilitates communication between itself (i.e. the sematic OSP) and whatever tool with its own standards. The wrappers are based on semantics using EMMO, semantic CUDS and APIs. Besides these excellent efforts, what is still

missing in an agreement to have a universal OSP. To reach this, ontologies and semantics are a must as well as common standards.
For both users and software owners, there should not be too many open simulation platforms available, because users will want a high number of tools available through the OSP, while the software owners do not want to make their software compatible with more than one to a few platforms. This requires that existing and future OSP projects work together.

# 2. Progress report (main activities)

## *2.1 Beyond the Models: Applying Semantic Technologies Across the Enterprise*

In this plenary presentation, Eric Little pointed out that Big Data (i.e. more and more data) will not provide necessarily more answers to questions especially if the data is messy and unstructured. The new concept to think about will be "Big Analysis"
Also, a difference between Machine Learning (ML) and ontology must be noted: ML is based on mathematics and statistics, thus ML can help with clustering data while ontology is based on logics and can help with reasoning (deductive, inference). Together they can provide "Big Analysis". ML on its own can miss important links, which can be filled by ontology built with domain expert knowledge (inform/constrain possible relationships).
Abductive Reasoning (not done by ontology alone) could answer questions such as, "something unexpected happens: what could be the cause?" The aim is to answer questions without having all the facts.
Considering the question whether all data can be represented in an ontology, one realises (for many practical reasons) that never all data can be organised to the level of ontology. However, there is a whole semantic spectrum (e.g. vocabularies, metadata, taxonomies) which could be utilised. One therefore needs to know where data live on the semantic spectrum and utilise existing semantic information.
Higher sematic levels open up the possibility of building a smart interface layer between siloed data sources and an analysis functionality (e.g. SPARCL queries). That interface layer works semantically by referring to a light weight ontology which performs as an annotation model for the data. Hence data stay at its source, and a semantic layer on top of the data provides integration and enables Big Analysis.
It was noted in discussions that EMMO can be the tool to build such an interface layer ontology and annotation model in the materials and modelling domain.

## *2.2 Digital Transformation of Materials R&D*

Digital transformation is seen as one of the key enablers in materials and manufacturing. A useful scheme for evaluating the status of materials modelling as a key component of digitalisation of materials R&D was presented (taken from Dr B. Ziebarth's Impulse talk)

- Level 0: no modelling; Edison's trial and error
- Level 1: Statistic and stochastic approaches
- Level 2: Guidance by modelling
- Level 3: Existing modelling data are coherently stored and exploited
- Level 4: Strong integration of modelling
- Level 5: Fully autonomous, modelling based discovery and development

Data is one of the key ingredients in digital transformation, and a large amount of data exist already. Unfortunately, there are some barriers that prevent full utilisation of the data available:

- Data often remains in silos, and is hard to get out, hard to find and identify
- Quality of data is varying, and it is time consuming to evaluate
- Data must be well curated, and this requires standards
- There is a need for a cultural change to share data, and the best way to achieve this is to create use cases
- Modelling and experimental databases should work together by uniting of databases, leading to larger datasets and hence supporting machine learning. Recommendations for achieving integration of existing operational nanotechnology resources were presented. In particular, it includes the barrier that data are conceptualised in different ways in different domains, hence calling for common/interoperable semantics and ontologies.
- Data integration is key, and we should move from an unstructured to a semantic level
- Adding semantics to data is a key step in making data FAIR (findable, accessible, interoperable and reusable). However, there is a considerable barrier due to the effort involved.
- Adding metadata and semantics greatly improves the quality of results. This was illustrated with examples such as glass transition temperature which is a complex concept where simple data mining leads to a wide scatter of results. Adding metadata leads to higher value outcomes.
- Ontologies are regarded as complicated and adoption is in particular hindered by a shortage of experts
- Semantic queries (e.g. using SPARQL) based on ontologies are much more powerful and 'natural'. They put the domain scientist in control, rather than relying on complex and often limited SQL database queries.

To increase the European manufacturing and process industries capacity to extract knowledge from data, they need to disseminate the results to encourage others to do the same. Knowledge should not stop at the corporate boundaries. To ease the achievement of this, we need good tools and techniques to access the data. A common solution is to require that data should be collected in common pools as far as possible. However, Eric Little suggested an opposite approach. Let the data stay at its source, in in whatever format it is, and instead build a semantic layer on top of the data from which it can be accessed and operated on.
Case studies and dissemination of success stories are important to get more and more stake holders involved. There is a need for a common semantic standard, and the leading software providers should lead the way by adopting such a common standard, but all stakeholders should be heard/included to ensure broad impact. On a cultural level (i.e. taking into account different domains and different companies) change needs to be managed carefully taking into account the need for proprietary solutions while managing the need for sharing and openness at least on the level of semantics.

## *2.3 Open Modelling Framework*

The three impulse talks show three different approaches to work with OSPs. Tata Consultancy Services (TCS) use a meta-modelling concept and have product, material and process. They move to the subject level such as steel gear forging and the lower level is an instance. This is their semantic basis and they use this for integration, as it is extensible. VMAP (Virtual Material Modelling in Manufacturing, http://vmap.eu.com/) aims to gain a common understanding of, and interoperable definitions for, virtual material models in CAE.

SimPhoNy has integrated everything from atomistic to continuum models. It has wrappers and facilitates communication between itself (i.e. the sematic OSP) and whatever tool with its own standards. The wrappers are based on semantics using EMMO, semantic CUDS and APIs. Besides these excellent efforts, what is still missing in an agreement to have a universal OSP. To reach this, ontologies and semantics are a must as well as common standards

A large number of well-established integrated simulation environments exist, including commercial closed systems, commercials systems with APIs and open source environments. However, there is a lack of truly open modelling frameworks that enable end users to integrate any modelling or pre or post processing tools on a widely agreed, open standard footing for all types of models (electronic, atomistic, mesoscopic, continuum). At the very heart of an open modelling framework is the Open Simulation Platform (OSP) which 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.

Characteristics of an OSP:

- The foundation of an OSP is an open semantic standard described in an ontology, such as the EMMO.
- An OSP enables the integration of codes into a simulation workflow.
- An OSP supports all aspects of a simulation, including pre-and post-processing and computational representations of models (including solvers).
- An OSP provides standards for the basic communication (metadata) between the varieties of components and software tools in a simulation workflow. The communication and data standards must be compatible with the ontology.
- Different levels of linking and coupling can be supported by wrappers.
- An OSP itself is separate from any pre-/post-processing or modelling codes or applications.
- All information exchanged through the OSP must be represented with an ontology-compliant standard.

The key factor here is to create open standards for how models (and data) should work together, including:

- Interoperability
- Ontology
- Standards
- Use cases
- KPIs of use cases
- Collaboration with software owners

Systematic knowledge representation can significantly reduce development time. Examples were shown in the presentation by B.P. Gautham (TCS Research). The approach has been to start with an analysis of the respective industry workflows and the important “elements of knowledge”. These are conceptualised and ‘ontologised’, leading to a digital representation and hence a “meta-modelling” integration framework. Models and data can then be tied into this framework. Hence in contrast to many existing simulation frameworks, the semantics comes first.
Simulation Platforms that are close to OSP standards are emerging, and example have been discussed, in particular SimPhoNy which has integrated everything from electronic to continuum models. It facilitates workflows between components.

Creating standards and agreeing on vocabulary can be challenging due to different agendas from different stakeholders, but there is no single tool to capture a manufacturing process, so tools need to be combined to be successful. Standards for interfaces and data transfer between tools is hence highly important. Efforts by the VMAP project in the engineering simulations domain were discussed. Clearly, even once there is a common understanding of concepts and metadata, efficient interoperability also requires tackling of formats and detailed standards.

For both users and software owners, there should not be too many open simulation platforms available, because users will want a high number of tools available through the OSP, while the software owners do not want to make their software compatible with more than one to a few platforms. This requires that existing and future OSP projects work together.
Standards are important, and should be put in place, but only where it makes sense.

- Interoperability means:
- the ability to exchange data and metadata while agreeing on certain rules that govern their meaning and interpretation.
- seamless integration of all relevant tools.
- that data of different domains relies on a common semantic standard and can thus be linked.
- (technically) the possibility of combining two or more computer programs into single, consistent solution.
- (semantically) that results from one program must be automatically convertible to be inputs to a second program. That includes documentation and tools for automatic validation
- the frictionless transfer and processing of information from one application to another.
- a structured data repository underlying all modelling and simulation

## *2.4 International Materials Ontology*

To advance materials and manufacturing research, data and modelling resources must be combined more and better than today. This requires interoperability. A key to achieve this is, amongst other things, is the development of as well as agreement upon an ontology able to represent the relevant domains, e.g. materials at all levels of granularity, processes, characterisation, properties, models. In manufacturing, the importance of ontologies is growing in the time of Industry 4.0. The aim is to form a “Triangle” of Product – Manufacturing – Service which are connected by the Internet of Things (IoT) and cyber-physical systems. Efficient information is key to such systems and ontologies are a key enabler to achieve that.
Ontologies must be truly global, and initiatives like the Industrial Ontologies Foundry work towards an open robust global ontology framework (presentation by Dimitris Kiritsis). IOF’s mission is to have a suite of highly interoperable ontologies which should be modular and thus, easily extendable. From IOF perspective, materials ontologies are at the mid-level, alongside e.g. machine and tool ontologies or manufacturing process ontologies. Apart from the challenge of establishing an ontology an even larger task is maintaining one. Some of the reasons why ontologies are not always successful include:

- that there are too many
- people develop them ad hoc
- everybody thinks it is easy to do so
- definitions are a problem
- there is an issue with interoperability between different ontologies

- there is no top-level management and no really commitment
- short half-life as there is no long-term funding
- poor documentation
- poor training

Using and developing ontologies requires specialists, and currently not many companies have the required personnel to take advantage of the possibilities that are presented.
Ontologies are important to the stakeholders involved, especially those who are trying to harmonize and analyze data across different sources, but they can be reluctant to adapt due to:

- There are many ontologies, is this the right one?
  - Show the benefits
  - Provide support (tools)
  - Use one common base ontology
- Must create use cases
- Many users will not join until a critical mass is reached

A long-standing example of materials ontology with practical applications is the ontology-based materials information integration with a focus on structural materials, developed and maintained by NIMS/NEDO in Japan (talk by Toshihiro Ashino). The ontology enables the integration of heterogeneous databases. A very powerful aspect Is the integration of mathematical representations which enables the unambiguous representation and identification of materials relations, important for interoperability in modelling applications.
The European Materials & Modelling Ontology (EMMO) was introduced as a tool for representing the physical world, giving a key role to the interpreter/domain expert. Thus, EMMO is an ontology tool for applied science rather than "the truth about the world". EMMO is based on 4 Primitives, i.e. Taxonomy (is_a relations), Mereotopology (parthood and slicing relations), Semiotics (representation, has_property relations) and Set Theory (Membership relations). It is structured in modules and its core include the abstract conceptual level (set/item), geometric/topological level (time, space, hybrid) and physical level (4D spacetime). Further branches include Semiotics, Materials, Models, Maths, Characterisation and Properties. In the materials branch, a "direct parthood" relation is introduced to handle granularity such that levels of granularity can be traversed in a direct rooted tree which is important for cross-scale interoperability. Also, the semiotic approach enables quantum systems to be represented in a way that is compatible with different interpretations (i.e. Copenhagen, De Broglie-Bohm) and approximations (e.g. Born–Oppenheimer).

There is a recommendation to create an organisation around EMMC, and to get EMMO published to ensure a long-term commitment for an international material ontology.
In summary, the key importance of ontologies in general was endorsed and the importance of ontology as a tool was shown in both the principle of the setup of EMMO, in applications and in approach to big data analysis. The need for ontology interoperability was emphasised. Communities are starting to interact and are open to collaboration.

# 3. Conclusions

In addition to the points summarised in Section 1.3, the following Actions are highlighted:

Actions

*Digital Transformation of Materials R&D:*

- There is a need to establish a 'change framework' to support industry in digital transformation, taking a wide range of aspects into account.
- Semantics and Ontologies play a key role and can put the domain expert in the driving seat. Better integration of semantic technologies with data analytics is needed.
- Digitise data, make them available, harmonise them, and create a European data warehouse.
- Verified and validated tools to extract knowledge from data need to be made available to industry, and their benefits are clearly assessed.
- The users need to be assured that any costs incurred to adopt them is outweighed by the benefits these tools bring.
- A standardised EU toolbox and related documentation should be aimed for.

*Open Modelling Frameworks*

- EMMC and SWOs should work close together to make sure that the latter are actually interested in an OSP and could envisage their code running on one.
- Standards are needed but should only be put in place where they make sense.
- Existing OSP projects should work together.
- Involve end users to see what interoperability they are aiming for.
- OSPs need to be made very user friendly

*International Materials Ontology*

- The EMMC needs to establish a long-term sustainable association to support the ontology effort
- The EMMC needs to take charge and curate the EMMO, as in standards, governing, training and dissemination activities.
- The EMMC must get involved into the digitalisation process.
- The EMMC should work with all stakeholders and excel in stakeholder management
- The EMMC should provide case studies that show what people can do with the EMMO
- The EMMC should further tackle interoperability and what is needed to make it happen.

# 4. Annex 1: Replies to the survey

## *4.1 Digital Transformation of Materials R&D*

### 4.1.1 How can we increase European manufacturing and process industry capacity to extract knowledge from data?

We should enlighten the materials modelling community about methods and objectives. There is a need of manufacturing and industry R&D activities. We could establish protocols for various specific needs such as material discovery from past data. We could provide proven frameworks and demonstrate benefits. However, these best-case examples and interoperability of data, should not end at the corporate boundaries. A value proposition is needed to aid decision makers to buy in to the often very intensive to digitize data. We need dissemination activities to make people wanting to extract knowledge which will lead to increase the usage of data.

We should provide a common semantic basis to easily extract the relevant information by end-users. Tools and techniques are needed to extract information from data but such tools must be easily accessible and frictionless.

A creation of a common pool of data from disparate sources is seen as necessity. However, we must take into account the varying formats of data from different sources and define an ownership. It also needs to be established if this harmonisation of data is a responsibility of the data provider or the data processor.
We need to be aware that data is geared to answer specific problems; so just the right level of data should be stored.
The EMMC community should decide on a framework that holds data on physics/chemistry/materials level. This has not been done. Regarding the EMMO we should not concentrate on the high level, but go for the mid-level which is very practical. Everybody should be able to plug extensions to the EMMO, if they wish so. EMMO can be seen as a tool to enable interoperability, and to do this, a minimum number of classes and relations should suffice.

### 4.1.2 How can we support leadership by digital transformation of data into knowledge?

We should promote of Centers of Excellence in Materials Modelling that can take this leadership. One must increase the visibility of modelling and create a vision which is then consistently communicated at all levels.
Case studies will be relevant that can show how knowledge deducted from data could have a clear business and strategic impact. These case studies should also showcase the tools needed to extract knowledge from data and offer real scenarios to prove their value.
To enable this, it is still pertinent to make data available and accessible. Also, sharing data, ideally as open or "semi-open" data is beneficial. Tools need to be provided to extract knowledge from data, but they must be verified and validated.

### 4.1.3 How can we achieve that manufacturing companies remain competitive (new products, high-added value) on the basis of knowledge extracted from wide-spanning data sets?

Before starting this, is pertinent to understand a company's current market position - its customers, suppliers, competitors, etc. Based on this information, one can judge where and what risks should be taken and how a targeted extraction of knowledge from data can be conducted and be useful. A company may ask for enabling development of new materials (material discovery), better products (learning from past designs and performance) and finding robust process regimes – and if these needs can clearly be transformed into data analytics or queries. Cases Studies can provide information around the opportunities of knowledge extraction, and thus, if these case studies are validated a company can try to replicated them.
Probably, a new positive taxonomy, such as “Knowledge bases" instead of "Data Sets" could be beneficial. Knowledge-based systems should be easy to use for technologists to allow as many stakeholders as possible to take part in the process.

### 4.1.4 How can we promote European benefits from data management, standardised documentation enabling interoperability and linking tools for enriching and enhancing the usability of data?

Industry needs evidence that this could work, so demonstrator projects with real life use cases could aid. The promotion will only work if one manages to generate sufficient success stories, where the benefits/usefulness are clearly shown. These should be disseminated at conferences and applied research conventions. Also, academia could play a role in education and raising awareness in future employees. Most likely an EC regulation could help to start this off, but an actual business model and success stories for this very business model are necessary at some point.

This promotion will not work if one does not first build trust with respect to the data. This comprises promoting, monitoring, improving data exchange practices and formats. Encouraging these activities inside a European Center of Excellence could be a start and also a provision services like data management and knowledge acquisition should be offered.

### 4.1.5 What actions are proposed for a common semantic knowledge base in materials?

Common semantics can only happen if all stakeholders agree, so it will be pertinent to offer more dissemination towards all communities involved. Also, encouraging meetings between different communities should be an action item.

Create a Consortium with these stakeholders to improve adoption and practice of such a semantic knowledge base. Get leading software providers such as Siemens, Dassault Systèmes, ANSYS, etc. on board to adopt such a common semantic basis for their developments and platforms. In the beginning, there will be EC or national funding necessary to aid with this transformation.

The stakeholder need evidence why a common semantic database is beneficial and why they should undertake the efforts to move from internal to common/joined efforts.

The common semantic knowledgebase must be developed with the users in mind, and one needs to research how far companies already went by themselves. Only then tools should be developed that also must be able to work with already established company internal efforts and honour confidentiality issues.

### 4.1.6 What practical ways forward are there for industry to reap the benefits of digitalisation in materials?

The first practical way forward will be to digitize data, make them available, harmonise them, and basically create a European data warehouse. The knowledge (extracted from these data) should make available multiple types of processes and product optimisation in order to offer RoI, such as saving costs or improve performance/functionality. The businesses/manufacturers should be enticed by case studies that are evidencing why digitalisation is the way forward. According to an outcome of the 2019 Industry Days in Brussels, it is vital to leave no stakeholder behind, so all practical ways forward need to be inclusive. For instance, a powerful user-friendly digital platform that talk the users' language can be beneficial.

Verified and validated tools to extract knowledge from data need to be made available to industry. However, to minimise the risk of these tools not being adopted, it is critical that their benefits are clearly assessed. The users need to be assured that any costs incurred to adopt them is outweighed by the benefits these tools bring.

To avoid an unsolicited explosion of tools and solutions a standardised EU toolbox and related documentation should be aimed for.

## *4.2 Open Modelling Framework*

### 4.2.1 What modelling frameworks, open or otherwise are the most important ones to you and your community/organisation?

The most promising modelling frameworks in molecular modelling and simulation are those that are based on physically inspired, yet straightforward building blocks, such as Lennard-Jones and Mie sites, point charges, and point multipoles. Combined to model classes with a few (say, up to five) free parameters, such that the model class as a whole can be characterised quantitatively.

Many frameworks are established within research groups, e.g. for sharing numerical models developed by different researchers. In a particular case, these can be run on a cloud environment for easy deployment, and can work with any FEM solver. There are also some international frameworks, such as the AiiDA python framework for high-throughput computations and provenance tracking, as well as the AiiDA lab simulation platform.

However, some participants do not see immediate need of OSPs.

### 4.2.2 What kind of standardisation is needed to use an OSP for interoperability?

Standards should be in such a way that different codes can be integrated, workflows can be described, input and output data are standardised, processes can be aligned, and different file formats can be converted.

There is doubt that standards can be imposed for all single models, so standards should be proposed for nodes which are concerning ontologies and data. Software Owners need to be part of a standardisation attempt, so that they are in agreement with the OSPs and actually want to adopt the standards. Semantic technology based on the EMMO and derived marketplace-level and subdomain-specific ontologies, i.e. a strong semantic base with flexibility is seen as beneficial.

### 4.2.3 How would you want to use an OSP to fully automatise data transfer between models?

Ideally users want to see automated data transfer as this could mean that multi-objective simulations could be conducted utilizing different models available on an OSP. Realistically, this will be a stepwise process and one should start with an OSP as a backbone of communication with plugins for commonly used other systems and over time bring in more tools and systems to use the same standards. Fully automatisation may not be wanted everywhere, for example file-based I/O will be necessary for communicating with platforms where a data base access is limited, such as nodes on HPC platforms. For this purpose, there are many good solutions available.

### 4.2.4 Is full flexibility with tools that can link to it important to you?

This depends on if a user works with a plethora of tools and is highly interested in multiscale modelling and requires different tools to satisfy their modelling. For some domains this may be not relevant if the user only works with a particular set of tools which seldom interfere.

### 4.2.5 What would an OSP have to offer so that you would use it regularly?

- Easy to use
- User-friendly interface and transparent processes
- Compatibility with common programs/tools and data standards
- Rapid prototyping
- Computational resources to run models remotely
- Provision of enough automated post-processing to be useful.

### 4.2.6 What would an OSP have to offer so that you would enable your software to link to it?

- It should be seen as the preferred OSP within a community
- A high enough number of potential users to justify necessary efforts
- In addition to standardisation of input and output data for the OSP, having access to anonymised data sets for simulations conducted using our own software through the OSP.
- Should be able to use existing workflow managers, e.g. AiiDA

## *4.3 Session 8: International Materials Ontology*

### 4.3.1 Are ontologies important to you?

Ontologies are very important some of the stakeholder who are analysing data across different sources and aim for harmonisation, etc. Ontologies are important for validation and data exchange Also, stakeholders involved in predictive maintenance in manufacturing and product sector will answer with yes. There is great interest in an ontology that is easy to handle and can cover use cases in materials modelling and can (partially) solve interoperability problems in the industrial domain. Very few stakeholders would negate the importance of ontologies and this related to a low digitalisation maturity.

### 4.3.2 How can we get all relevant stakeholders on board?

It is important to hit critical mass before all relevant stakeholders come on board. This could be achieved by demonstrating widespread expressions of interest, and confirming that a particular ontology will be THE ontology that the community will be following for the foreseeable future.
One needs to effectively show the ontology's utility and benefits accrued in other communities. This could happen developing relevant business use cases at various levels and covering various aspects of the lifecycle of a Product Service Systems. One should then actively promote a common base ontology and provide convenient support (e.g. tools) in order to derive or build compatible domain ontologies

### 4.3.3 What is needed to keep the international materials ontology going (RDA, new projects/funding, forming a consortium, ...)?

The EMMC should turn into a consortium/organisation to keep things moving. Collaborations between all stakeholders are necessary and they should even go beyond the materials domain. Clever funding needs to be in place as ontologies *per se* are unlikely to attract grants.
Training and dissemination are important and also evidence of active use of an ontology. The EMMO should be published and also the relevant tools so it can be used and real-world use cases should be added. It needs to be demonstrated that companies or large software providers are using it for real.

### 4.3.4 Can we achieve "interoperability"? What does this mean to you?

*Interoperability means:*

- The ability to exchange data and metadata while agreeing on certain rules that govern their meaning and interpretation.
- Seamless integration of all relevant tools, i.e. the frictionless transfer and processing of information from one application to another.
- That data of different domains relies on a common semantic standard and can thus be linked.

- (technically) the possibility of combining two or more computer programs into single, consistent solution or (semantically) that results from one program must be automatically convertible to be inputs to a second program.
- A structured data repository underlying all modelling and simulation

The stakeholders believe in interoperability but first semantic interoperability has to be achieved to guarantee a better buy in. Otherwise interoperability remains limited to certain scales. This requires discipline and that it has to be done in an orchestrated way at global level. As a start one could utilise the principles on which IOF is built.

### 4.3.5 What sort of training/ dissemination/ would you like to see?

EMMC should form a working group on Materials Modelling didactics. This group should formulate recommendations, down to the level of example syllabi, to be used in curricula across the globe. The group should be ready to act as a consultant to universities and other institutions that aim at establishing curricula in Materials Modelling and closely related fields, or at modernizing their existing curricula. One should argue in favour of creating a funding landscape for the development of Open Educational Resources in Materials Modelling, and push agencies such as EC, BMBF, etc., to establish such programmes.

The training required should comprise introductory materials to ontology courses, tools usage, courses on ontologies and formats for data exchange, training and clear documentation of the EMMO and routes to derive domain ontologies, and best practices for ontology development. The dissemination required could happen in regional workshops, tutorials, online video series and via published use cases.

| | |
|---|---|
| Authors | AlexSimperler & Gerhard Goldbeck (both Goldbeck Consulting Ltd.) |
| Contributing EMMC-CSA partners | FRAUNHOFER, POLITO, SINTEF, ACCESS, GRANTA, MDS |

| | |
|---|---|
| EC-Grant Agreement | 723867 |
| Project acronym | EMMC-CSA |
| Project title | European Materials Modelling Council -<br>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 | SINTEF AS | 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 |
| Synopsys / QW | Synopsys ( former 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 |
| Email | nadja.adamovic@tuwien.ac.at<br>emmc-info@tuwien.ac.at |
| Project web-sites & other access points | https://emmc.info/ |