EMMC-CSA
European Materials Modelling Council

Report

Expert meeting on open source, free software and commercially supported software for materials modelling

25-26 October 2018, Montrouge, France

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. Executive summary

1.1 Description of the workshop and objectives

We are currently witnessing the transition of materials modelling from an academic research activity into an industrial engineering practice. This evolution presents opportunities and risks of far-reaching consequences for the recognition of European academic research institutions and the competitiveness of the European industry. Associated with this development are issues such as industrial acceptance of materials modelling, human skills and resources, relationships between industry and academia, and the best sources of software tools.

In this context, an expert meeting was held on the 25th and 26th of October 2018 in Montrouge (Paris) France to focus on two primary forms of software development and deployment, namely free and open source software and proprietary commercial software. For two days, 13 experts from universities, governmental research laboratories and software companies, namely eight external experts and five members of the EMMC-CSA from Goldbeck Consulting, ACCESS e.V., QuantumWise/SYNOPSYS, and Materials Design SARL elaborated and discussed the synergies and conflicts of the two software approaches. The discussions were based on the extensive experience represented by this expert group, where almost all participants had a background in code development and distribution in the field of materials modelling.

The main objectives of this expert meeting were (i) to solicit the different viewpoints of the various stakeholders, (ii) to identify the pros and cons of the various practices, and (iii) to formulate recommendations.

To achieve these objectives, the workshop was divided into five sessions:

1. The lay of the land:
    a. What exactly do we mean by "open source, free software, and commercially supported software"?
    b. Who are the main players in open source software and commercially supported software for materials modelling?
    c. What are the synergies and conflicts between these types of software?
2. How are the industrial needs for materials modelling met by the different types of software?
3. How do we see the best solutions in the future?
4. What are the main obstacles to overcome for successful industrial deployment of modelling software?
5. Synthesis of discussion points and joint outline of report

Overall, the character of the expert meeting was very much like brain storming. As expected, the discussion revealed different opinions and was in parts quite controversial. This is also reflected by the present report.

Detailed information on the expert meeting can be found at the webpage of the workshop: https://emmc.info/events/expert-meeting-on-open-source-and-commercially-supported-software-for-materials-modelling/

1.2 Major outcomes and recommendations

The major results of this expert meeting are:

1. Materials modelling software is playing an increasing strategic role in the time of digitalization and "digital twins". It's a European strength and opportunity.
2. Successful software for materials modelling has an expected lifetime of many decades. This long-term nature requires a sound legal and business foundation, namely:
    a. The ownership of software must be clearly established.
    b. The license models need to be carefully thought through to ensure a sustainable development and maintenance of the software and impactful exploitation by both academic and industrial end-users.
    c. Solutions need to be considered for different forms of software including individual software components, comprehensive codes and software platforms. Various business models exist and should be analysed ranging from those built on services around open source software to proprietary solutions.
3. Both free and open source software and commercially supported software have their place. It is important that all parties seek to maximise the synergies and minimise conflicts. This is achieved by a clear understanding of the requirements and needs of the software developers and those of the end-users. The motivations of academic software developers and of commercial software providers are different. This complementarity should be used as strength, not as cause for conflict.
4. Conflicts can arise when academic research codes are directly distributed to industry (including non-European companies), especially when there is no or only a low license fee.
5. Education and training encompassing scientific/technological as well as business aspects are essential to achieve optimal synergy between pioneering academic software developments and successful commercial deployment and value creation.
6. Dissemination of success stories is very important.

2. Meeting Report

2.1 The lay of the land

This first session aimed to set the stage for the subsequent sessions. In particular, the discussions centred about the definitions of the terms open source, free software, and commercially supported software, which are often used without a clear understanding. This session also aimed at identifying the main players involved in the development, distribution, and use of software especially for materials modelling. Finally, the discussion extended to the different and possibly diverging interests of these key players as well as possible synergy effects arising from the complementarity of interests.

2.1.1 What exactly do we mean by "open source, free software, and commercially supported software"?

At the outset of this session the basic definition of the term "software" was discussed. According to the European Materials Modelling Ontology (EMMO) as introduced by the EMMC, "software" is not just a code but comprises everything needed for its use. This led the expert group to the general distinction of software and data. It was concluded that eventually this distinction depends on the view of the particular stakeholders. The group agreed that "software" is everything provided by the developers and includes computer programs as well as internal data like forcefield or potential parameters defining the interactions between atoms or like pseudopotentials or projector-augmented wave (PAW) potentials reflecting part of the single-particle potential in DFT calculations.

Most often, this internal data comes in hardwired and parametrized form. In contrast, "data" comprises all input data provided by the user necessary to perform an actual calculation. Taken together, "software" and "data" exploit the relationships between chemical composition, structure, and materials properties.

When it comes to the definition of the terms "open source", "free software", and "commercially supported software", ambiguities and misunderstandings show up. In particular, the term "free software" refers to the freedom of certain rights, which are guaranteed to the user on receiving the software, whereas it does not mean "no cost", in which case it would be usually termed "freeware". Specifically, free software is defined by the so-called four freedoms formulated by Richard Stallman, the founder of the Free Software Foundation (FSF), in 1986 and published by the FSF as follows [1]:

- Freedom 0: The freedom to **run** the program for any purpose
- Freedom 1: The freedom to **study** how the program works, and change it to make it do what you wish
- Freedom 2: The freedom to **redistribute** and make copies so you can help your neighbour.
- Freedom 3: The freedom to **improve** the program and release your improvements (and modified versions in general) to the public, so that the whole community benefits.

The freedom of software is guaranteed by a variety of software licenses, which may be used for a particular software as, e.g. the BSD, MIT, Apache licenses as well as the license models distributed by the FSF as the well-known GNU General Public License (GPL) and the GNU Lesser General Public License (LGPL). All these licenses are called permissive as they guarantee the maximum of rights to the licensee. However, it should not be overlooked that the GNU Public Licenses, especially the GPL, come with a considerable restriction of the freedom of the licensee due to the so-called copyleft condition, which requires that the rights guaranteed by the GPL must be preserved in all copies and modified versions derived from the software. As a consequence, if a developer wants to combine their own proprietary software with software coming under the GPL, the own software must also be distributed under the regulations of the GPL and, hence, will no longer be proprietary. This unbalanced distribution of freedom is avoided by the other licenses mentioned above.

Materials modelling software originating from academic research groups as free software is typically open source software, which is software with source code that anyone can inspect, modify, and enhance (https://opensource.com/resources/what-open-source). However, open source software is not necessarily free and can still be commercially distributed. Proprietary software can be used with any license model chosen by the developer and owner of the software [1]. With respect to the source code, options range from providing source code for personal use by the customer and end user (e.g. Gaussian, VASP, CASTEP) to closed source models with use of the executables free of charge or partially free of charge depending on regional markets and software as a service (SaaS) and, finally, to closed and fully commercial distribution.

Obviously, license models become important in relations with different markets in the world involving different legal rights of developers, software owners, software distributors as well as end users. This may lead to complex interrelations between stakeholders.

2.1.2 Who are the main players in open source software and commercially supported software for materials modelling?

Obviously, the stakeholders can be divided into different groups in different ways. From the perspective of software life-cycle one distinguishes developers, owners, distributors, and end users. Of course, there can be overlap of these groups, e.g., a developer can at the same time be the owner of the software and may even distribute the software. In contrast, from an organisational perspective, academic, non-profit, and commercial stakeholders can be distinguished.

The expert group estimated that worldwide about 600 codes in the field of materials modelling exist, about 400 of them focusing on post-processing of calculated results. Often used codes include (in alphabetic order and without any claim for completeness) Abinit, BigDFT, CASTEP, CP2K, DALTON, DMol, $DMol^3$, DPOLY, FHI-Aims, FLEUR (Jülich group), FPLO (Dresden group), Gaussian, GIBBS, Gromacs, GULP, LAMMPS, Molcas, Molpro, MOPAC, NEMO, NWChem, OMEN, Orca, Quantum Espresso, QuantumATK, SIESTA, Turbomole, VASP, and Wien2k. A more complete list of quantum chemistry and solid-state physics software can be found at https://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid-state_physics_software.

In the field of thermodynamic modelling, Thermo-Calc and COSMO-Therm are examples. A large spectrum of software is available for continuum modelling. In addition, Paraview, KITware, and Dream3D are well-known examples in the field of visualization software.

Some of these codes are commercially distributed by a number of companies including Dassault/Biovia, SCM, Materials Design, Culgi, Synopsys/QuantumATK, Cosmologic, Schrodinger, Gaussian, Scienomics, FACCTS, Molcas, MDLab, Ansys, Comsol, NanoAcademia, and JSOL.

As a matter of fact, most of the relevant codes for materials modelling have been developed in Europe and Europe also hosts most of the software companies active in this field as, e.g. Dassault Systems, Siemens, Hexagon MSC, ESI and others.

2.1.3 What are the synergies and conflicts between these types of software?

Overall, synergy effects arising from the collaboration of stakeholders from academia and commercial companies prevail even though the main goals of scientists working in these two domains differ considerably. Specifically, academic research leads to the development of, e.g. new methodologies, libraries (BSD, LGPL) (lapack, libxc, fftw, etc.) as well as to the evolution of standards, which strongly advance the field. Of course, modules and libraries, which can be linked by many codes (like libxc), bear the potential danger of error distribution. If undetected, errors in such libraries would "infect" all codes linked to them. However, the chances of detecting errors are larger due to better testing when more codes are using the same libraries. For this reason, competition between different codes for being the most accurate, fastest, and the one with most features, is important and likewise advancing the field. Needless to mention that validation and verification play an important role at this stage.

In return, academic research may benefit from commercial distributors and end users. Commercial distributors advance the use of materials modelling in industry through their presales staff which in typical sales cycles demonstrate to industrial end users how materials modelling can be used to address high value problems in the industry. Often some of the model methodology is lacking and, in such cases, they often bring back the understanding of industrial problems to academic groups and to initiate new developments, which may even be funded by industry. Hence, commercial support is important for creating the market as well as for financial return. It is helping to make choices in research activities, creates visibility for technology transfer, and may thereby even lead to more publications and citations.

Nevertheless, conflicts arise from the competition of commercially supported and free and open source software (FOSS) distributed directly by academic researchers especially to industry. The latter causes considerable distortion of the market since the development costs of these researchers are covered by public funds. This aspect is exacerbated if FOSS is made freely available to non-European companies. License schemes like the GPL add to the market distortion since they may undermine the development of proprietary software. Even worse, free and open source software may affect the development of new software. An example is the open source software package LAMMPS distributed by the Sandia National Laboratory in the U.S. The availability of this code makes it very difficult for software companies to bring alternative programs to the market. Yet, this might be viewed as only a commercial loss of opportunity without cutting scientific incentive to develop software, which might be motivated by the goal to increase academic reputation or gaining control over one's own code. Still, this point of view was questioned by some of the experts. Overall, the controversial discussion of these issues clearly revealed different perspectives.

In addition, it was expressed that if free and open source codes would not show an exceptionally good performance there would still be enough room for commercial competition. Of course, competing software development should not just replicate FOSS code but offer added value such as new features or increased performance. Finally, it was mentioned that FOSS still leaves enough commercial opportunities in service provision. Again, the discussion showed quite different points of view.

It was also pointed out during the discussion that support of FOSS may also play an important role. In particular, disturbances of the market situation may arise from public funding the development of FOSS, which is then openly distributed as stated above with the example of LAMMPS, but they may also arise from public funding of long-term support, as is also the case for LAMMPS. Insofar, this example was regarded as somewhat exceptional by some experts.

The experts acknowledged that both synergies and conflicts exist between free and open source software and proprietary commercial software. Synergies come from the fact that FOSS can help in the creation and dissemination of new computational approaches, algorithms, and ideas. Conflicts can arise in the marketing and the financing of ongoing support. Furthermore, there is a conflict in the objectives of university researchers and the purpose of commercial companies. Academic research aims to advance and disseminate knowledge without necessarily seeking practical applications while commercial companies need to focus on creating direct economic value and generating enough income to create stable employment and to invest in new developments.

2.2 How are the industrial needs for materials modelling met by the different types of software?

At the beginning of the session the expert group discussed the industrial needs for materials modelling software. First of all, software must be as general as possible and at the same time robust, it should warrant confidentiality of industrial research and the results generated by the software must be accurate, reproducible, and traceable. In addition, modelling software for industrial use needs to be able to handle complex systems while delivering high speed and computational efficiency. Important factor are continuous support and training over long periods (years or even decades). Finally, a code should offer a wide spectrum of functionalities and allow for easy integration of new functionalities. Last, but not least, interoperability with other software packages is clearly of value for end users.

Beyond these more general aspects there exists a variety of needs depending on the industry sectors, company size (large companies, SMEs), and business models. Usually, larger companies have a larger time horizon than small companies, which often have more specific short terms needs. Moreover, larger companies have enough resources to develop some of the solutions themselves, whereas smaller companies need to integrate their tools into simulation flows with tools from other companies, on which they thus depend. Yet, overarching industrial priorities are the near-term optimisation of existing materials and processes and the development of new materials.

The needs of industrial end users are met quite differently by distributors of free and open source software (FOSS) and commercial distributors. While the former provide source-code at low license costs if any at all, the

software products are often "bleeding edge" and thus may lack full verification, validation, and user-friendliness. At the same time, support is often via user forums and not provided by the distributors themselves. In contrast, commercially supported software is much easier to use, fit for purpose, and comes with qualified training and long-term support. In addition, it is often integral part of larger software environments, which provide interoperability and close linking to related software enabling, e.g. true multiscale modelling. Confidentiality requirements are also better respected due to customer-tailored support rather than support by open user forums.
Actually, free and open source software can create more benefits for large industrial research organisations than for SME's that don't have in-house experts to take full advantage of research-type codes.

2.3 How do we see the best solutions in the future?

Hot topics identified by the expert group as having the largest impact on future developments are standardization, interoperability, and interdisciplinary communication and teamwork. This includes the necessity of a common language, respecting the complexity of each domain, which is provided, e.g. by the European Materials Modelling Ontology (EMMO). Standardization is complemented by the creation of more and more libraries for standard tasks, which can be used by different software packages. Interoperability will benefit from the introduction of standard data exchange models like HDF5, which include metadata and allow easy exchange of data between different codes. Of course, standardization bears the risk of large software distributors becoming monopolistic, since they have large and experienced materials modelling groups with strong impact on new developments in materials modelling software.

2.4 What are the main obstacles to overcome for successful industrial deployment of modelling software?

The expert group identified the following major obstacles for widespread industrial deployment of materials modelling software:

- Lack of experienced and well-educated scientists and engineers in the field of materials modelling
- Lack of maturity of materials modelling software regarding accuracy and complexity
- Lack of connectedness of materials modelling and industrial end products
- Lack of sufficient number of convincing examples demonstrating economic benefit
- Small market segment size of materials modelling
- Unrealistic expectations on the side of the industrial end users
- Lack of good estimation of error bars
- Lack of appreciation of the value of materials property simulations; much higher appreciation in the field of FEM simulations in mechanical, chemical and electrical engineering communities
- Lack of high throughput simulations on smaller scales, databases
- High initial investment (human resources, software licenses, education and training of scientists and engineers)
- Cultural hurdles, reservations against simulations on the side of experimentalists

The following order of magnitude estimates of the market size as based on a number of previous studies were made by the expert group:

- New materials simulation: 100 M€ (electronic and discrete modelling)

- Materials modelling: 1000 M€ (continuum level)
- Design of products: 10000 M€ (system level)

According to the experts, materials modelling is mainly used in the US, in Japan, Europe, and South Africa, with strong emphasis on the US and Asia regarding the electronics industry. In particular, Japan is very strong in materials modelling.

Means to overcome the above-mentioned obstacles include dedicated university courses on materials modelling, improved training for industry researchers. In addition, synergy effects of the collaboration of academic and commercial software owners should be used to educate research scientists in industry. The expert group regarded it highly important to provide and deploy successful illustrations of how materials modelling software is able to solve high-value problems cheaper and/or better than using only experiments. However, it should be also made clear that simulations are meant to complement and focus experiments rather than to replace them. In general, much emphasis should be put on demonstrating the commercial benefit of materials modelling using convincing examples.

Of course, it is understood that continued investment in the development of more powerful, more accurate, more efficient, and more comprehensive modelling software will be the basis for future progress and this is best achieved by a productive collaboration between academic research efforts and professional software development and support.

2.5 Synthesis of discussion points and joint outline of report

The last session of the expert group meeting focused on digesting the outcome of the previous sessions and formulating conclusions and recommendations. These are summarised in the following section.

3. Conclusions

Key messages:

- Academic research and teaching are an indispensable foundation for the maturation of the field.
- Education should cover both understanding of materials properties and convey knowledge about methods for materials modelling.
- Successful case studies demonstrating the economic benefits of materials modelling to understand, predict, and optimize materials properties are an important prerequisite of deployment of materials modelling in industry. Corresponding contributions from industrial researchers are welcome.
- Demonstration of complementarity of experiment and modelling is essential as is the use of digital twins of materials.
- Preconfigured workflows for standard tasks as well as calibrated and validated software tools are needed.
- Software should increasingly comprise knowledge thereby accompanying the educational transition from passive (libraries, books) to active media.
- Standardization and modularization of software as well as the generation of libraries, e.g. libxc is essential for materials modelling software deployment.
- Institutional support for software of public interest is required.
- A European roadmap for the deployment of materials modelling software should be created.
- A fruitful dialogue between academic developers, commercial software providers and industrial end users of materials modelling software should be continuously encouraged.

- Materials modelling software as a way to communicate and advance knowledge to the benefit of society should be more appreciated.
- Synergy effects from combining academic free and open source software and commercial products should be exploited more effectively.
- Better integration of academic and commercial software providers is needed.

4. Annex: Agenda

4.1 Scope

The European Materials Modelling Council was installed to foster the exploitation of atomistic simulations and materials modelling software by the European industry. This effort includes measures to accelerate the transformation of materials modelling software mainly developed by academic researchers into software tools, which could be readily used by industrial end users. Currently, materials modelling software ranges from programs written by Ph.D. students and post-doctoral researchers to commercially supported comprehensive software platforms. In this context, license models selected in the course of the code development play an important role regarding the deployment of software and the acceptance of materials modelling software by commercial companies. This expert meeting will address the aforementioned issues especially with respect to open source and commercially distributed software. The expected outcome of this Expert Meeting will be a report on the strengths and pitfalls of different models and recommendations to increase the impact of materials modelling in the European industry. This will be achieved by impulse talks by leaders in the field and by focused discussions.

4.2 Participants

13 experts from universities, governmental research laboratories and software companies.

4.3 Agenda

Thursday, 25 Oct:

12:00 – 13:30	*Registration, cold lunch (sandwiches)*
13:30 – 13:45	Welcome and opening remarks
13:45 – 15:15	Session 1: the lay of the land What exactly do we mean by “open source, free software, and commercially supported software”? Who are the main players in open source software and commercially supported software for materials modelling? What are the synergies and conflicts between these types of software?
15:15 – 15:30	*Coffee break*
15:30 – 17:00	Session 2: How are the industrial needs for materials modelling met by the different types of software?
17:00 – 17:15	*Coffee break*
17:15 – 18:45	Session 3: How do we see the best solutions in the future?
19:30 – 21:00	*Dinner and informal discussions*

Friday, 26 Oct:

9:00 – 9:15	Summary of Day 1
9:15 – 10:45	Session 4: The main obstacles to overcome for successful industrial deployment of modelling software
10:45 – 11:00	*Coffee break*
11:00 – 12:30	Session 5: Synthesis of discussion points and joint outline of report
12:30	Meeting adjourns
12:30	*Lunch (optional)*

5. References

1. V. Eyert and K. Stokbro, White Paper on Standards of Modelling Software Development, EMMC, 2017.

2. A. C. Powell IV and R. Arroyave, Open Source Software for Materials and Process Modeling, JOM 60, 32 (2008).

3. G. Wilson, D. A. Aruliah, C. T. Brown, N. P. C. Hong, M. Davis, R. T. G. Guy, S. H. D. Haddock, K. D. Huff, I. M. Mitchell, M. D. Plumbley, B. Waugh, E. P. White, and P. Wilson, *Best Practices for Scientific Computing*, PLoS Biol. 12, e1001745 (2014).

4. A. Morin, J. Urban, P. D. Adams, I. Foster, A. Sali, D. Baker, and P. Sliz, *Shining Light into Black Boxes*, Science 336, 159 (2012).

5. A. Morin, J. Urban, P. Sliz, *A Quick Guide to Software Licensing for the Scientist-Programmer*, PLoS Comp. Biol. 8, e1002598 (2012).

The outcome of this workshop was compiled by	Volker Eyert (Materials Design) Kurt Stokbro (Synopsys) Erich Wimmer (Materials Design)

Workshop organisers	MDS, SYNOPSYS

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/