

Mathematics for Digital Science

Opportunities and Challenges at the Interface of Big Data, High-Performance Computing and Mathematics

Introduction

Mathematics provides the fundamental supporting framework, as well as a common language, for the science and technology on which our modern societies now depend. Very little of the physics, biology and engineering of the past century would exist in the absence of mathematics. Today, as a result of the ongoing explosion of computing and communications technology, we see completely new challenges emerging for both science and mathematics: how can we handle and make useful the deluge of data on social, technical, economic, ecological, and technological systems? What kinds of new mathematical tools do we need to do so? How does mathematics need to change to help science, both fundamentally and its application? And how might new science and technology feed back to create new mathematics?

The European Commission is aware of these pressing questions, and is working to explore how its future programmes might best support these developments at the European level. As part of this activity, the commission has sought specific input from the mathematical and scientific community on how mathematics might help address the challenges of Big Data and near-future high performance computing (HPC). The Commission aims to use this input in its planning of the Horizon 2020 Work Programme 2016-2017. This report offers a summary of the most important themes emerging from two exploration activities, an online consultation and a follow-up workshop, both exploring opportunities and challenges at the interface of big data, high-performance computing and mathematics.

The online consultation (<https://ec.europa.eu/digital-agenda/en/content/mathematics-and-digital-science>) asked experts to address several specific topics:

1. The role of mathematics in big data
2. The role of mathematics in HPC, in particular exa-scale computing
3. The role of e-infrastructures in mathematics
4. The impact of applied and industrial mathematics on innovation
5. The preparation of the FET Proactive (HPC) and/or the e-Infrastructure Work Programmes 2016-17 under the Excellent Science pillar of Horizon 2020

The consultation was then followed by a one-day workshop in Brussels, bringing together roughly 50 experts to further elaborate ideas and identify key promising themes. At the workshop, Commission representatives encouraged participants to elaborate their thoughts on a set of topics closely related to the consultation:

1. How can mathematics help high-performance computing and “big data”?
2. How can high-performance computing and “big data” help mathematics?
3. What practical steps can be taken to make these influences take place?

The extensive material emerging from both the consultation and workshop touched on a broad range of topics. This report describes a set of six principal themes, along with their specific opportunities and challenges. It also offers further discussion of some slightly less prominent topics, which nevertheless appear to be of notable interest to the broad community.

Participants in the workshop were also asked to submit a short description of a vision for a specific research challenge that would help to realise the goals identified by the community. These contributions are included here as appendices to this report.

Main themes emerging from the consultation and workshop

1. Modelling, simulation and optimization (MSO)

Modelling, simulation and optimisation (MSO) has been called the third pillar for scientific progress and innovation, standing alongside experiment and theory. As technology-driven industries grow in complexity and innovation cycles become shorter, the efficient and timely development of industrial technology has come to depend on accurate methods for MSO. This will be even more pronounced in the context of the emerging importance of high-performance computing (HPC) and technologies and businesses associated with Big Data. Competitive industrial development will increasingly require a “virtual product,” which accompanies the real product and allows cheaper, faster and more effective product verification, risk analysis and optimization.

Research efforts on several challenges are required to encourage the rapid development of such methods, and to bring them into industrial practice:

Build interdisciplinary cooperation

Pushing Europe forward in MSO will demand a policy focus in building cooperation between mathematicians, engineers, scientists and leaders in industry. As mathematics plays a crucial role in all three elements of MSOs, development of mathematical methods should be emphasized. It is also important to create e-infrastructures to allow knowledge of new methods to spread quickly to the entire community of European science and industry. The European Commission should include mathematical MSO into all major technology funding streams.

Develop expertise in key technical areas

MSO is already useful, but its application in a number of fields will require technical advances. Important areas of focus include:

- Multi-physics and multi-scale systems,
- Combined discrete and continuous non-linear systems,
- Stochastic processes,
- Model reduction schemes for high-dimensional systems,
- Methods for handling inhomogeneous uncertainties in a model
- Coupling of models
- Methods for optimization with multiple criteria

Make MSO valuable in applications at the European level

The impact of MSO development could be particularly important for the future development and management of infrastructure networks for gas, electricity or communication. The European Commission could push for a major European research initiative in this area. This would allow a more holistic approach to controlling and optimizing national or supranational networks through planning informed through MSO, with large economic and social returns.

2. Develop methods to extract information/ structure from big data; develop analytics to make sense of data and complex systems (with topological concepts, etc.)

The data deluge affects almost every sector of commercial and public endeavour. The data itself is merely material; it is nothing without analytics. We need concepts, methods and practices which can extract valuable and actionable insights and meaningful knowledge from large volumes of data. It is also the case that algorithms handling large data sets work best when they are attuned to the structure of that data. The blind application of algorithms to data, ignoring structure, is both inefficient and prone to producing erroneous results.

Programmes to encourage European research in this area should focus on three specific themes:

Developing analytics to detect structure and correlation in ‘big data’

To be useful and trustworthy, analytics should be founded on rigorous mathematical concepts, ideas, and methods. The key challenge is to find analytics for producing data-driven insights in situations where structure is hidden and non-obvious. Progress in this area should inspire new products and services and expand the capabilities of companies and public institutes. The European Commission can stimulate such progress with initiatives that bring mathematicians into close contact with experts from big data, and by encouraging the use of a wide range of mathematical ideas and methods.

Develop better algorithms by focusing on structure

In areas such as computational chemistry, materials science, and weather forecasting, physics-inspired models start on scales of time and space that are much smaller than those of the desired outputs. Naïve or brute force simulation using off-the-shelf numerical methods often leads to the “pollution” of outputs as errors from small scale details percolate upwards to larger scales. This problem can be avoided with mathematically well-designed algorithms which respect the inherent structure of the data or physics involved (if that structure is known). Creating methods for designing structurally sound algorithms and simulation is crucial to the development of better real-time analysis and simulation tools. The European Commission should encourage research programmes at the European level to integrate mathematicians with research scientists in such areas.

Exploiting statistics in big data

For useful analysis, big data often needs to be made smaller. Many findings emerging from big data result from a small sampling of a fraction of the data collected. Statistics offers models and tools to screen and split big data in a ‘divide and conquer’ approach, so as to clarify how much of the data is actually useful. A focus on statistical methods for analysing big data would help us be “less wrong,” by ensuring that results produced by models and tools are not over-interpreted, or interpreted incorrectly. Gaining practical knowledge in this area requires European programmes to integrate experts in mathematical statistics with research in big data. Promising research topics include new inference theory and methods to reduce “noise” and the risk of false positives (which increases as data gets “bigger”), development of parallel algorithms for machine learning and of parallel data handling methods for statistical analysis.

3. Building infrastructure to stimulate innovation through mathematics

Mathematics can be helpful for HPC and big data, but is also important far beyond these areas. It is a key driver of innovation, yet new mathematical ideas typically do not move into industrial use immediately; they take time to spread through the sciences and engineering. The European

Commission has an opportunity to stimulate innovation directly by helping ideas to flow more quickly, in part by linking mathematicians more directly to people in industry. This could be done in different ways:

Create an information infrastructure to help mathematicians and companies find one another

Mathematical knowledge flows into practice fairly well in some large companies with their own research departments, but is rare for most small or medium-sized companies. Europe needs infrastructure designed to help companies and mathematicians find each other, and to further the exchange of mathematical ideas, including algorithms. It should also be easy for companies to publish their problems and attract the interest of mathematicians. One forum of this kind already exists: the European Consortium for Mathematics in Industry (<http://www.ecmi-indmath.org/>).

What industry needs is a "single-stop-shop" where all the information on availability of competencies can be found. This could be provided by an e-infrastructure at the European level, which would have a significant impact on innovation by increasing the flow of ideas into industry.

Make the most of pure mathematics

Many branches previously considered purely theoretical – including group theory, algebraic topology such as persistent homology, harmonic analysis, algebraic geometry, combinatorics and so on – are now seeing an increasing number of practical applications especially in the context of extreme data. It's important that we find ways to use and further develop this existing mathematics in confluence with today's data. The Commission should consider designing a call with the aim of bringing researchers from traditionally "pure" areas of mathematics into closer collaboration with application-oriented mathematicians, scientists and end users.

4. Help solve emerging challenges for efficient exa-scale computing

The European Commission is committed to supporting a strong push to realise exa-scale computing by 2020. This will, of course, require progress in a host of technologies, but will also demand rapid advances in a set of core issues in both applied mathematics and computer science. Progress in two general areas will play a significant role in determining the ultimate reach of HPC applications and their overall impact on society:

Encourage the development of mathematically sound software for exascale computing

Current software designs do not scale well for exa-scale applications. The European Commission should encourage a focus on the development of mathematically-informed algorithms and designs. Core technical issues demanding attention include:

- developing highly scalable solvers and numerical libraries
- developing energy-aware algorithms
- developing hierarchical algorithms exploiting all the levels of parallelism of the platforms
- learning to ensure resilience and correctness including reproducibility
- finding reliable and rigorous mathematical approaches for exascale system software
- solving problems of data intensive computing (big data management and data analytics)
- developing architecture-aware algorithms and software for exascale computing

Learn to quantify uncertainty in HPC applications with statistics

Statistics are needed to quantify uncertainty in new high-performance simulations, and thereby to assess their accuracy in a reliable way. The Commission should launch initiatives at the European level to advance research on uncertainty quantification issues for mathematical modelling and optimization.

5. Realise e-infrastructures for mathematics knowledge management/sharing

E-infrastructures can be helpful to mathematics, as well as to science and industry, in a number of ways. They can also act to catalyse further integration between disparate communities, and also to make existing mathematical knowledge more widely available and easy to find. The European Commission should sponsor initiatives pushing in four main directions:

Help mathematicians with access to computing resources on a Europe-wide basis

European e-Infrastructures can support the computational needs of mathematicians by providing computing resources (storage and processing) that may not be available to them locally. It is difficult to approach mathematicians at a European level, as most are supported by national e-Infrastructures. The Commission could help in building European networks of mathematicians with dedicated workshops.

Connect mathematicians more effectively with industry and also the rest of science

e-infrastructures for mathematical knowledge management and sharing should be designed to help companies and mathematicians find each other. Industry needs ways to find mathematical ideas, experts and links to possible solutions. Likewise, mathematicians need access to possible applications of mathematical ideas, or to new problems where ideas may be applied and developed further.

Help with continuous professional development in current mathematical methods

Scientists working in industry often ask for information about current mathematical methods and tools. To match these needs and to facilitate the pan-European industrial employment of these methods, an e-infrastructure for continuous professional development in current mathematical methods should be developed, including the utilization of modern e-learning tools. There is generally an insufficient awareness (from the industrial side) that mathematical tools are evolving very quickly.

Make past mathematical research widely available and searchable

Mathematical research data is rapidly becoming more widely and freely available, in part through the European Digital Mathematics Library <https://eudml.org/>, currently the largest collection of Open Access Publications in mathematics. It goes beyond traditional library services, for example, by including a formula search engine. Further developments exploiting big data techniques could empower researchers by making existing mathematical results (including software, algorithms, benchmark problems sets, etc.) much more easily available. Such infrastructures would facilitate the transition of mathematical content into technological innovation.

6. Impact on mathematics – new problems, challenges from HPC and big data

Mathematics has often sprung from the minds of mathematicians and only later found applications in science and engineering. Just as often, however, science and application have provided inspiration and challenges for mathematicians, resulting in new mathematics; the theory of random processes is one example. The European Commission could make such back-flow more likely with specific workshops or other activities.

Focus on identifying new mathematical challenges in emerging industrial and scientific areas. We

should expect this kind of back-reaction into mathematics to continue with big data and high-performance computing, which will pose mathematical problems of unprecedented kinds, demanding new concepts and tools for their solution.

Other notable sub-themes

Risk and statistics

In the area of financial risk, various mathematical models exist for risk assessment, based on statistical methods. However, the era of big data brings new challenges and opportunities. Two current challenges are to improve assessment of systemic risk in the financial sector in general, and to develop an overall risk assessment strategy for use in financial markets, insurance, and corporate governance including customer classification and individual pricing. Meeting these challenges would help address recurring issues such as the impact on markets of changes in the regulatory environment. The emergence of Big Data means that there are ever more possibilities for analysing risks. There are fundamental opportunities for mathematics to provide help in this domain.

Help in handling privacy/ anonymity issues

Privacy and anonymity issues threaten to become the ugly undoing of the promise of big data. Current solutions are often half solutions cobbled together on the fly with many shortcomings. Mathematics can help to provide more robust and sustainable solutions.

Mathematicians and communication

Communication has room for improvement in the community of mathematicians, especially in Europe. The impact of mathematics in tackling the big scientific and societal challenges and as part of the huge recent development in ICT is not sufficiently recognised. This is an issue that requires training and concerted effort to improve.

Summary of Sub-Group Contributions

The appendices (to follow, below) present the specific research challenges identified by workshop participants during an afternoon of focus within ten distinct subgroups. The topics of the various contributions were, in brief summary, as follows (the order is arbitrary):

Subgroup 1: Infrastructures for maths and innovation

Discussion and observations

We discussed different possible actions/instruments that the EC Directorate General CONNECT could take to involve existing initiatives at the interface of mathematics and industry in the fields of Big Data and High Performance Computing. Important examples include [EU-Maths-IN, the European Service Network of Mathematics for Industry, and ECMI](#) , the European Consortium for Mathematics in Industry and Innovation.

Recommendations (for the Commission)

Interdisciplinary networks: One of the main instruments that could be implemented is an open call for interdisciplinary networks involving mathematicians from different areas and computer scientists in order to solve specific problems in this field. It could also initiate *European Study Groups in Big Data and HPC* to bring industrialists and scientists from different areas, proposing open problems in some topic.

Other infrastructures: Other desirable outcomes include a Software Repository for HPC, a Job portal, & virtual job fair in mathematics for Big Data and HPC, and a Database of mathematical expertise and collaborations in HPC and BD. This will allow to advertise a one-stop-shop to facilitate access to mathematics collaborations related to BD and HPC especially for SMEs and to communicate on success stories.

Subgroup 2: Challenges for HPC/exascale computing

Discussion and observations

The evolution of high performance computing (HPC) to the Exa-scale brings tremendous opportunity to address industrial simulation, data processing challenges and fundamental research.

Recommendations (for the Commission)

Progress will require overcoming a number of key technical challenges, which require interdisciplinary collaboration. The Commission should initiate programmes to encourage such collaboration around the following themes:

- Enriching mathematical descriptions of solution algorithms to include the manner in which they are executed on large, complex computers
- Accounting for hardware scaling issues in formulating mathematical solution schemes
- Providing mathematical support for remote computing and data access.

Subgroup 3: Interfaces between maths and industry

Discussion and observations

People in industry generally do not know of new developments in mathematics, and mathematicians lack contact with industry. There are interfaces between the two at regional, national and transnational levels, but they are tenuous and need to be made much stronger.

Recommendations (for the Commission)

The Commission could help in a variety of ways, including:

- Seeding the foundation of a European database on mathematical expertise
- Building a central resource for finding jobs at the interface of mathematics and industry
- Linking initiatives in different geographical areas within Europe
- organization the collection of best practices, as well as their wide promotion

Subgroup 4: Organising outreach towards non-mathematicians

Discussion and observations

Non-mathematicians do not see the value mathematics brings to many different fields. This is, in part, because mathematicians only communicate internally, in a precise but mathematical language, even though this excludes non-mathematicians. Mathematics needs effective outreach, to demonstrate to scientists, political leaders, as well as the general public, the important role mathematics plays in science and society.

Recommendations (for the Commission)

Outreach should be promoted as a standard procedure, when communicating mathematics-related results. This could for example be achieved if powerful institutions (like the EU or other funding

bodies) make outreach an important focus of projects and encourage effective communication about project results.

Subgroup 5: Interfacing maths with science and engineering

Discussion and observations

The extraordinary growth of High Performance Computing (HPC) and Big Data (BD) rich science/engineering is creating revolutionary opportunities for mathematically driven advances in research and development. Capitalizing on the opportunities will require new mathematical, computational or statistical models and tools, and these can only be developed in close connection with other disciplines.

Recommendations (for the Commission)

The Commission should consider the full spectrum of the mathematical, computational, physical, engineering and life sciences, and should undertake both fundamental research and cutting-edge applied work. It should be demand oriented, and should aim to advance interaction and collaboration between mathematics and other disciplines.

Subgroup 6: Modeling, simulation and optimization

Discussion and observations

Innovation in industry and society is increasingly complex and occurs over ever-shorter innovation cycles. A key technology in facing this challenge is the integrated computational modelling, simulation and optimization (MSO) of systems, e.g. a digital factory, a complete vehicle, or a human heart. Although there are many success stories of the use of MSO, the full potential of MSO as an integrated discipline has not yet been fully realized.

Recommendations (for the Commission)

To create real value from MSO, it must be an essential part of every innovation project, together with HPC. The combination of MSO with HPC and Big Data would enable us to solve larger-scale problems, perform real-time simulations, solve inverse problems, address multi-scale problems and model, understand, master and optimize complex and rapidly changing networks. The European Commission can help by strongly encouraging the use and development of MSO as a research infrastructure, and including MSO aspects in relevant areas of the work programmes of Horizon 2020.

Subgroup 7: Quantum computing

Discussion and observations

Quantum computing promises to unleash a new era of technological innovation. Yet it has remained a promise only for some two decades, as the technology to make it a reality has grown gradually. New ideas in mathematics could play a crucial role in getting quantum computing past the critical threshold of applicability, which is now close.

Recommendations (for the Commission)

The European Commission could be influential by helping to start a road map for mathematics

applied to quantum computing. Key topics for focus include describing the topology of entanglement, assessing programming languages for quantum computing software, learning to re-write classical computing algorithms as quantum algorithms, and developing a theory of quantum Hamiltonian complexity.

Subgroup 8: Mathematics, complexity and data

Discussion and observations

The huge amount of data now being collected essentially reflects transactions between complex systems, such as the Internet, human brains, the environment, or biological systems. Extracting behavioural models in this context requires new mathematics, and new foundations to deal with digital science.

Recommendations (for the Commission)

The Commission could help with initiatives in this area, spurring efforts to apply mathematics in applications to HPC and Big Data. We need focussed research to develop new kinds of mathematical abstractions for HPC, as well as new processes that respect the structure of the data.

Subgroup 9: Repositories of mathematical knowledge

Discussion and observations

Mathematical results never become obsolete. This mathematical knowledge must be safely archived in long lasting interlinked repositories accessible by research mathematicians as well as other users of mathematical methods and results. Mathematical knowledge is a commons and should as such be curated by public (at least not-for-profit) entities and be available through open access. This is currently not the case.

Recommendations (for the Commission)

Strong support from the EC and the EU states is thus critical. The European Digital Mathematics Library (EuDML) currently holds a leading position in the ongoing global/world DML efforts, but is in need of continuing attention and work to ensure it remains so. Its leading position is threatened by an initiative supported by US funding agencies. The Commission could have a large impact by encouraging focussed efforts on the following tasks:

- Building a complete picture of existing data sources (digital libraries, data repositories, software, specialized tools, ...),
- Reviewing existing metadata standards and developing new ones
- Providing stable identifiers for interlinking resources
- Finding new maths-aware semantic tools to extract metadata from existing sources
- Defining standards for interoperability between different types of objects.

Subgroup 10: A multi-purpose software repository for mathematics in industry

Discussion and observations

Ever-faster progress in computer technology presents a number of challenges for mathematics, and we need novel concepts even to develop basic components such as mathematical solvers and to handle massive data. This challenge may be a basis for a wide range of collaborations among mathematicians focused on new opportunities arising from industry.

Recommendations (for the Commission)

To address this challenge, the European Commission should encourage a programme to develop generic software tools and solutions for a broad range of applications in science, technology and services, with the aim of building a repository of solutions for industry. Such a repository would also help education in computational and data sciences. Specific outcomes would include:

- An increase of the effective computing power available in massively parallel computational infrastructures
- Advances in industrial mathematics, especially in the development of new algorithms
- Associated advances in software development
- Progress in spreading open source software and in changing patterns of software usage
- The creation of large distributed repositories integrated by common indexing and access mechanisms

Conclusion

The results emerging from the online consultation and follow-up workshop clearly indicate that the mathematics community sees important challenges and opportunities at the interface between mathematics and high-performance computing (HPC) and Big Data. These work both ways, promising new mathematical ideas and techniques useful for industry and basic science in these areas, as well stimulating new developments in mathematics itself. The community believes that significant innovation for science and industry, with important economic impact, will come from a research and development focus in this area.

The European Commission will use the output summarized in this report as inspiration for the Horizon 2020 work programme for 2016-2017. The Work Programme is foreseen to be drafted during the first half of 2015 and approved during the second half.

Appendix: Sub-group Reports

On the following pages are the individual reports of the scientists reporting on the discussions and findings of the various sub-groups (ordering is unimportant).

Infrastructures for mathematics and innovation in big data and HPC

(José A. Carri)

[EU-Maths-IN](#), the initiative launched a year ago as a European network of national networks for interactions on mathematical research with enterprises, can serve to manage several European actions connected with Big Data (BD) and High Performance Computing (HPC) to develop, enlarge, and empower the mathematical community working in these fields. We discussed different possible actions/instruments that the EC Directorate General CONNECT could take to develop the involvement of mathematics, and in particular [EU-Maths-IN](#), in the fields of Big Data and High Performance Computing. Among them, we extract some of the main ideas discussed that could help for possible future calls in “Mathematics in Digital Science”:

- **Research Networks:** We think that one of the main instruments that could be implemented is an open call for interdisciplinary networks involving mathematicians from different areas and computer scientists in order to solve specific problems in this field. They could be medium size networks, consisting of 10/15 nodes, producing the suitable mobility of senior and postdoctoral researchers and building a sense of community in the field. They will allow structuring the relation between mathematics and data science in the near future. They will produce activities such as congresses, summer schools, training events incrementing the mobility and overall synergies between mathematics, Big Data, and HPC. We have in mind as a model the successful scheme of RTN (Research & Training Networks) in past EU Framework Programmes.
- Specialised **European Study Groups in Big Data and HPC:** The idea of these workshops, already tested with lots of success stories, is to bring industrialists and scientists from different areas, proposing open problems in some topic. The audience will be typically composed of senior researchers, postdoctoral, and PhD students that by the end of the week have to propose a “solution” or a “strategic plan” to solve the problem. This eventually will lead to the matching between mathematicians and HPC/Big Data scientists.
- **Software Repository in HPC.** Developing efficient software repositories is a challenging but essential issue. Mathematics permits to represent problems, algorithms or software in abstract way and universal language. This allows to increase/facilitate the portability (adaptation), to enable interdisciplinary transfer, and to anticipate the technological evolution.
- **Job portal, & virtual job fair** in mathematics for Big Data and HPC and more generally in Mathematics for Industry. The use of modern communication technologies such as internet meetings could allow for electronic job fairs inspired by successful recent experiences in France. This and the previous point could be proposed in some other call in e-infrastructures that we intend to apply.
- **Database of mathematical expertise and collaborations in HPC and BD**, this will allow to advertise a one-stop-shop to facilitate access to mathematics collaborations related to BD and HPC especially for SMEs and to communicate on success stories. It can serve as a way to collect from companies some challenging problems involving mathematics in BD/HPC. Moreover, it will be useful for enterprises but also to recognize the (already) involved research groups and to motivate others to join the effort.
- **Mathematics/Grid Computing**, Mathematics should be present and interact with all the

pyramidal structure of HPC ranging from team level to the national nodes and the pan-European ([PRACE](#)). Connections to the European Grid Infrastructure (EGI, www.egi.eu) should be encouraged.

We claim that a FET proactive devoted to MSO (Modelling, Simulation & Optimization) will allow implementing such concrete actions for the benefit of both mathematical science and innovation. Mathematics is essential for the development of Big Data and High Performance Computing. Indeed, Mathematics is the oxygen of the digital world. If it is there, you do not notice it. If it would not be there, you realize that you cannot do without.

Challenges for HPC/exascale computing (Anne Trefethen and David Standingford)

The evolution of high performance computing (HPC) to the Exa-scale brings tremendous opportunity to address industrial simulation, data processing challenges and fundamental research. Mathematical challenges associated with this new era in computing fall into a number of broad categories:

- **Enriching mathematical descriptions of solution algorithms to include the manner in which they are executed on large, complex computers.** This includes characterisation of the interconnections and node characteristics - including the network topology, data storage and process energy consumption. The performance of algorithm classes running on very large systems is now intimately linked to the way in which the computing systems are configured on a range of scales – from individual (possibly many-core) processors, local network connectivity, I/O and the latency of distributed architectures. These might be articulated in algorithmic terms in such means as communication avoidance and data ordering.
- **Accounting for hardware scaling issues in formulating mathematical solution schemes.**
 - As computers become more complex, the scale of the system is sufficiently large that even the smallest of probable failure becomes significant? Algorithms should be tolerant of such faults – either via detection and response, or via fundamental properties of robust iterative schemes that remove the impact of uncorrelated errors.
 - For certain classes of algorithm, increasing the task size leads to stability issues and error accumulation in standard methods. These need to be revised with appropriate numerical analysis.
 - Key attributes of certain physical process modelling must be conserved along with geometric structure regardless of scale (this includes total energy, etc. for numerical simulation).¹
- **Provision of mathematical support for remote computing, data access and curation.** Large expensive computers increase in utility and efficiency if they can be centrally located and managed, with remote access provided to a range of end users.
 - This introduces problems when the data being processed is commercially or nationally sensitive. Mathematics can provide support for encryption of data exchange – even homomorphic encryption where data is processed in an encrypted state.
 - Minimising the data that needs to be exchanged with a remote computer can make use of new forms of data compression, particularly where the nature of the data can be specified or inferred.
 - Real-time data at scale (for example the Square Kilometre Array – SKA) may require special mathematical treatment prior to ingestion to ensure suitability for a given processing task.
 - The need to keep data for long periods of time – to support product assurance or certification requirements – may rely on new mathematics for compression and / or

1

¹ Care should be taken to distinguish between algorithms where conservation issues become apparent at scale, and those where conservation is a more fundamental issue.

indexing

- **New classes of task can be regarded as computationally tractable.** Whereas for some disciplines, just getting a large numerical task to complete has been enough to challenge the available processing capacity, Exa-scale brings new opportunities:
 - Optimisation algorithms that use the primary numerical task within a larger task to search some parameter space can be made more effective with a range of mathematical tools.
 - Algorithms that have been developed that are suddenly tractable due to scale, randomised linear algebra and other approaches that are not a method of choice at smaller scale need to be revisited.
 - The data representation of variables within tasks can be recognised as stochastic, rather than deterministic – which offers the opportunity to propagate uncertainty information.
 - The a-priori parameters for numerical tasks can be updated with a range of mathematical devices for inverse problems. This opens the opportunity for commercial data assets in industry that are maintained throughout the life cycle of a product or service offering.
 - Tasks previously considered too large to consider at all are now coming in to scope. These include climate prediction and molecular dynamics.
 - Co-design – where computers and humans (or other systems) work together at scale via an appropriate mathematically formulated interface.

- **Data Field Theory.** New mathematics is required to represent multilevel dynamical systems with a primary focus on the data itself.
 - The mathematical assumptions underlying common software libraries (for example freely available linear algebra, data management and visualisation code) can be revisited.
 - The framing of applications from a data centric perspective will make use of new classes of mathematics – at least mathematics exposed to the wider community though publicly available and well-documented code bases and programming interfaces.
 - Mathematics for comparing very large datasets from disparate sources – assimilation, correlation and re-use.

Interfaces between maths and industry (Marie-Christine Sawley)

Today's interfaces exist at the regional, national and transnational level

- Such as the European study group AQMI, EU Maths-IN and AMIES in France for example
- Observations
 - Liaising with additional disciplines such as on a University campus, is very positive; the Marie Curie TN and ID are very good examples of fellowship benefitting from the immersion into broadened campus
- Recommendations
 - connect the dots of the initiatives at different geographical levels into a tiered system
 - the experiences are positive, and among the best known methods, having a local, committed facilitator plays a major role in the success
 - do not keep it closed onto the definition of HPC and BD only
- Recommended actions
 - Resources have to found for consolidating software basis and its lifecycle, beyond PhD duration
 - Put some resources aside for large scale validation on the fly accessing extremely large HPC systems
 - Build a EU job bank
 - Collect best practices, promote, broadcast
 - Build a European DB on mathematical expertise
 - Policies and regulations should be monitored, as they may dictate the perimeter of possible actions

Organising outreach towards non-mathematicians
(Andreas Wierse and Alberto Chierici)

The group agreed that there is a significant lack of awareness among non-mathematicians about what value mathematics gives in many different fields, how it influences the every day life of everyone. We also agreed that mathematicians themselves have a tendency to mostly communicate internally, i.e. inside their community; that they are proud of their very precise language, that allows them to communicate efficiently, but do not realise that this keeps non-mathematicians excluded.

There are some initiatives that already try to improve the communication about mathematics, for example:

- AMIES (Agence pour les Mathématiques en Interaction avec les Entreprises et la Société)
- Imaginary (<http://imaginary.org/>)

The main question is: How can we make more people interested in the mathematics that is underlying many well known problems/topics/fields?

Outreach seems the right word for what we have in mind. And successful outreach is extremely important for politicians: we could show them, what mathematics can do but we also have to be careful that it stays clear, what mathematics cannot do! This is especially relevant, when (numerical) simulations increasingly play an important role in the public.

We as mathematicians have to explain or summarize mathematics in an understandable way; we have to create support structures, e.g. for teachers, to support them in doing this.

The main goals that the subgroup concluded are:

- Outreach must be promoted as a standard procedure, when communicating mathematics-related results. This could for example be achieved if powerful institutions (like the EU or other funding bodies) make outreach a relevant part of projects and the communication about project results
- Popular topic(s) should be chosen and we should show how much mathematics is in there; this would be a good basis for communication and the improvement of the image of mathematics

Interfacing maths with science and engineering (Manolis Vavalis)

Objectives

To promote research and development at the interfaces of Mathematical sciences and other science and engineering disciplines on questions originated from, but not limited to, HPC-BD. To encourage new collaborations, as well as to support existing ones. Successful proposals will either involve the formulation of new mathematical, computational or statistical models and tools whose analysis poses significant mathematical challenges or identify innovative mathematics or statistics needed to solve an HPC-BD related important problem in other disciplines.

Background

The extraordinary growth of High Performance Computing (HPC) and/or strong Big Data (BD) rich science and engineering is creating revolutionary opportunities for mathematically driven advances in research and development.

Specifically, due mainly to size and stochasticity considerations, HPC-BD quite often require higher level of conceptualization, elucidating formalism and high level of reasoning. This commonly leads to abstraction, which in turn brings Mathematical Sciences much closer to related disciplines.

The interaction between Mathematics and other disciplines takes place through continuously evolving scientific interfaces. These interfaces range from ones that have been proposed centuries ago and still offer great value and create new opportunities, to new emerging ones that connect mathematics with other disciplines in either a new way or by composing existing interfaces into added values ones.

Activities should

- Consider the full spectrum of the mathematical, computational, physical, engineering and life sciences, and should undertake both fundamental research and cutting-edge applied work.
- Propose general frameworks as well as specific actions, which will allow us to explore and advance interfaces between disciplines leading to a demand-oriented federated framework for advancing the research and development in various disciplines.
- Result in theoretical foundations, practical considerations and specific software platforms.

In particular they should

- identify potential for changes in the existing interfaces and potential for generating new interfaces
- propose selected generic efforts, tasks, methodologies and platforms that will help us strengthen these interfaces, foreseen problems and remove obstacles
- analyze in depth selected cases that will emerge from the above two actions and have the potential to act as a roadmaps to other cases
- implement specific concrete actions concerning
- training for building interfaces between disciplines
- developing economic and other incentives
- popularization and public awareness actions
- utilizing other actions within this call perhaps through an additional CA project

Examples of areas of appropriate R&D include the following:

1. Mathematics and approximate computing: It should be based on the long standing observation that not all computations should be performed with the same accuracy. This poses new R&D related questions and opens new opportunities for collaboration between computational mathematics and computer architecture concerning floating point arithmetic, perhaps fully utilizing the early works of John Von Neumann and Jim Wilkinson.

2. Topology and smart grids: Electric grids considered as the most complex man-made system ever build. They need to obey physical laws and at the same time provide practical solutions on the particular needs of the society and the associated markets. It has been recently recognized that topology could play a protagonistic role on the open competitive energy markets that are emerging from the emerging smart grid. Topology is considered as an attractive way to deal with the prohibitively large amount of digital information that need to travel together with the watts and rapidly processed.

3. Large scale statistical mechanics for lipid bilayer membranes: ...

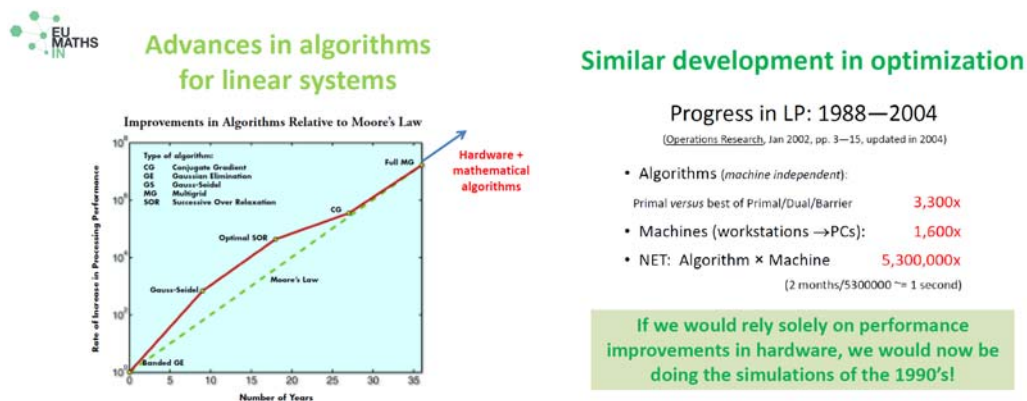
These areas are examples only. They are not meant to be inclusive. The work that is supported under this initiative must impact other science and engineering fields and advance mathematical sciences. Thus, collaborations between the mathematical scientists and appropriate scientists from other disciplines are expected. Other methods to ensure impact are also possible and should be specified in the proposal.

Modelling, simulation and optimization (Wil Schilders)

Future challenges for innovation in industry and society exhibit increasing complexity and at the same time have to obey to ever-shorter innovation cycles. One of the key technologies in this permanent fight is the use of computers at peak performance in an appropriate way, i.e. in the integrated modelling, simulation and optimization (MSO) frame. In competitive industry and in the top scientific research projects a full holistic approach is to be applied (e.g. to use MSO on a complete vehicle, a full digital factory, the human heart or the complete vascular system). To develop such a holistic approach one needs a mathematical model that allows simulating and optimizing the real product on virtual products via the use of high performance computing (HPC) tools.

Although there are many success stories of the use of MSO (see “European success stories in Industrial Mathematics”, Springer, 2011, ISBN 978-3-642-23848-2), the full potential of MSO as an integrated discipline has not yet been fully realized and hence its role in the creation of value continues to be severely overlooked. Simulations can provide unobstructed access and unveil hidden worlds, whereas optimization can avoid costly over-engineering.

It is well known that computers have become orders of magnitude faster over the past 20 years, but what is unknown is that the performance of mathematical algorithms has improved with even higher factors, as is shown in the following slides:



Together, these speed increases allow us to do the challenging simulations and optimizations that are currently being performed. If this would be left to computer performance only, we would now be solving the problems from the 1990's. Hence, in order to create real value from MSO, it must be an essential part of every innovation project, together with HPC. This holds, in a similar way, for Big Data. The Big Data Revolution is one of the main science and technology challenges of today. While this is multifaceted, mathematics is at the very core of the challenge – in ranking information from vast networks in web browsers such as Google, or identifying consumer preferences, loyalty or even sentiment and making personalised recommendations, the very scale of big data makes automation necessary and this, in turn, necessarily relies on mathematical algorithms. The

challenge is to derive value from signals buried in an avalanche of noise arising from challenging data volume, flow and validity.

The combination of MSO with HPC and Big Data would enable us to solve larger-scale problems, perform real-time simulations, solve inverse problems, address multi-scale problems and model, understand, master and optimize complex and rapidly changing networks. The inclusion of randomness and manufacturing uncertainties is also very important, and is adequately treated by new emerging areas in mathematics such as uncertainty quantification.

Highlighting MSO as a research infrastructure and/or future emerging technology would provide both the scientific and industrial research communities with an advanced way of using newest mathematical technology combined with high performance computer resources and give them a tool to systematically achieve new results of high impact in their fields. Establishing MSO as a future emerging technology will enable Europe to capitalize on the current European leadership in application-driven MSO, to strengthen European competitiveness in industrial innovation in providing industry with tools of higher precision within the same time scale, and to meet important future societal challenges. Moreover, many significant research projects could be brought into the breakthrough level.

Quantum computing (short text from email)
(Antonio Puertas-Gallardo)

The objective of my chaired subtopic was to accelerate the transition to a new way to do computing, moving from Boolean logic (classical) to superposition logic (quantum).

- Start a road map or an assessment of Mathematics applied to QC, for example topology of entanglement.
- Assessment of programming languages for Quantum computing software development
- Development of new quantum algorithms to solve problems now solved with classical ones.
- Translate/re-write the classical computing algorithms into Quantum Algorithms.
- Manipulating quantum information by geometric operators.
- Develop the Quantum Hamiltonian Complexity theory.

Mathematics, complexity and data (Emanuela Merelli)

This is a report on the little workshop on "New Mathematical Foundations for Big Data & HPC". The workshop was mainly focused on three questions:

1. Why do we need for new maths foundations?
2. Is the topology the starting point towards new mathematical methods for Big Data & HPC issues?
3. Do we have any idea of what kind of new theoretical framework can be proposed?

Premise by the chair:

The huge amount of data, that is nowadays being collected and for which we shall suite a proper HPC, essentially represents collection of transactions data, that are sequences of actions performed by entities of one or more complex systems interacting one each other in an environment, being Internet, the brain, the weather, a biological system. Thus, focusing on data, we deal with complex systems hidden in the data, especially on their emerging behaviors. Be able to extract behavioral models, it could require the capability to represent the collection of transactions through an interactive model of functions, or likely an interactive model for "embedded" algorithms (computers). The current computing machine are not suitable to this purpose, even if very well equipped with memory and potential power. Perhaps the mathematics behind the Von Neumann architecture need to be revised and a new one will lead us beyond Turing.

The discussion among the attendants began on posing the question: Why do we need new mathematics, new foundations to deal with digital science?. All agree on the necessity to find a new abstraction for HPC, suitable to represent problems as semi-structured mathematical description where the computational system is part of the 'model' of complex multilevel systems. A kind of self-organizing computation coming with the model. But also new processes that respect the structure of the data and the flow of the information.

The second analyzed question was: Is topology the starting point towards a new mathematics for facing the Big Data & HPC issues? This was widely discussed. Topology is just the branch of mathematics that deals with both local and global qualitative geometric information in a space, those properties that do not depend on coordinates but only on intrinsic geometric features. Coordinate-free qualitative analysis as consequences of dimension reduction on decisions. But also qualitative new foundation for HPC to solve new class of problems. Anyway the ability to get a mathematical interpretation to the equation that link data to information, information to knowledge: $\text{shape of data} = \text{information} + \text{computation}$.

A third question regarded the possibility to determine a theoretical uniform framework enabling us to extract the manifold hidden relations (patterns) that exist among data, as correlations depending on the semantics generated by the mining context. A possible approach, briefly discussed as Field Theory of Data, it proposes to exploit the above mentioned way of incorporating data in a topological setting, transferring and generalizing to the space of data notions inspired by physical (topological) field theories and harnesses the theory of formal languages to define a potential semantics to understand the emerging patterns. To this direction a possible computational abstract framework supporting HPC could be based on the S[B] paradigm, a new way of modeling complex system by entangling in a unique model the mathematical description of the structural and computational components by supporting data-based model evolution.

Repositories of mathematical knowledge (Fabian Müller and Marta Sanz-Solé)

Mathematical knowledge is a complex system, still primarily based on scientific publications, although other sources such as mathematical data, algorithms and software become more important. Since most mathematical research results are obtained from already known results by strictly logical deduction, combination and generalization, mathematical knowledge represents a continually growing building in which it is necessary that each floor is reached and no stone is lost. Due to the timeless validity of mathematical results, they never become obsolete. It is therefore absolutely necessary that the mathematical knowledge is safely archived in long lasting interlinked repositories accessible by research mathematicians as well as other users of mathematical methods and results.

Mathematical knowledge is a commons and should as such be curated by public (at least not-for-profit) entities and be eventually open access. Strong support from the EC and the EU states is thus critical. The group discussed ways to enhance access to mathematical knowledge amenable to storage in a repository. Such data may consist of literature, software, algorithms, data sets/computational results, animations, formalized mathematics and specialized databases such as [Mizar](#), [Coq](#), [GAP](#), [LMFDB](#), [OEIS](#), [Knot Atlas](#), [Atlas of finite groups](#), among many others. Due to the heterogeneous structure of these types of information, not all of them should be stored in a single repository. However, storage formats for this data should be designed with a view towards easy ways of citing and linking each other, both within a single and among several repositories.

It was emphasized that the [European Digital Mathematics Library \(EuDML\)](#) currently holds a leading position in the ongoing global/world DML efforts, but is in need of continuing attention and work to ensure it remains so. In particular, in view of the support of the WDML project by the US National Academy of Sciences (NAS) and funds from the Alfred P. Sloan Foundation, the EuDML needs strong support in order to maintain this leading position and to represent European interests in the WDML project. It can serve as a connecting and coordinating centre between other kinds of repositories, both already existing and newly created ones. The software database [swMATH](#) with its accompanying links to [zbMATH](#) articles is the first example where two different kinds of mathematical knowledge are interconnected.

Other areas like formalized mathematics are in need of unifying metadata standards that turn articles and input data (programmes) into machine readable and citable pieces of data. The [Module system for Mathematical Theories \(MMT\)](#) was mentioned as a possible candidate. A corresponding unified metadata standard for literature was designed in the EuDML project, which again can serve as a good example case. Since a large part of mathematical knowledge is contained within natural language text, semantic tools are needed in extracting it. Such tools need to be specially adapted to mathematics because of the widespread presence of formulas and also due to the fact that everyday terms are used in a specialized sense. In the end, such data mining should enable users to consult a knowledge base for, e.g.,

- finding large data sets with certain properties,
- ⊗ - finding algorithms applicable to a particular problem,
- ⊗ - finding people who already solved some problem (or a similar one).

Availability of full texts for machine processing is crucial in order to discover relations between papers and new paths in the literature. The EuDML already covers a critical mass of data, which should continue to grow, and offers an expertise in math-aware mining, which is a research area to be supported.

As a reasonable action plan for the next work programme, the group resolved on the following steps to be taken in order:

1. take stock of existing data sources (digital libraries, data repositories, software, specialized tools, ...),
2. review existing metadata standards and develop new ones where necessary,
3. provide stable identifiers for interlinking,
4. extract metadata from existing sources by employing new maths-aware semantic tools (e.g., using keywords from referenced literature for tagging other kinds of data),
5. define standards for interoperability between different types of objects.

For final thoughts the group concurred in that new knowledge in the form of previously unnoticed connections can be uncovered just by interlinking already existing sources of data. Therefore with respect to the question of what kinds of data should be collected in a repository one should aim to be as comprehensive as possible.

It should also be noted that while not all mathematical knowledge can be expected to be freely available (e.g. proprietary algorithms), it is usually in the interest of the rights owner to spread the word of its existence to as large an audience as possible. Thus at least the relevant metadata can be expected to easily find its way into a repository.

What beyond HPC & Big Data in digital science? (Emmanuel Harcourt)

Our idea was to think about how the resources of Big Data and HPC can be used without anyone knowing precisely how they work. We agreed that these resources should be available to anyone and people should be able to use them without being experts in their technical aspects.

To achieve this, we need abstract semantics, or mathematical semantics, created by experts in Big Data and HPC. These would make the resources usable by a wide variety of people, making it more likely that the full range of uses of these technologies will be found.