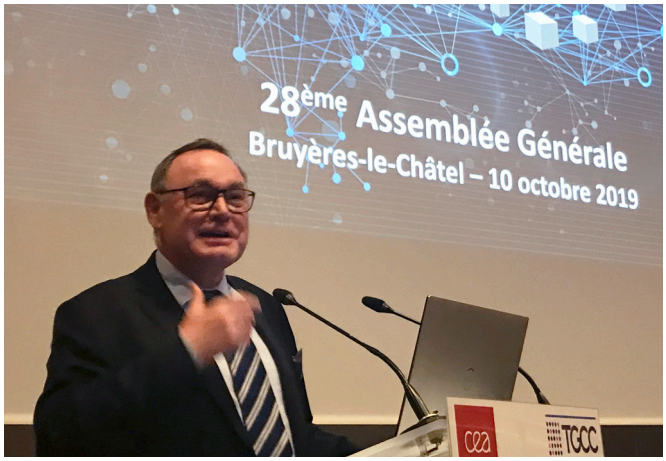
Europe and Genomics on the frontline

Held early this fall, our General Assembly provided an opportunity to take one complete view on activities and prospects of the Teratec Association, particularly in the context of EuroHPC established by the new European Commission. It was also an opportunity to meet Jean-Marc Grognet, General Manager of Genopole reminding genomics mission and its advances, who also explained why HPC is essential in the work of the Genopole and why he wanted to work closer with Teratec.



©TERATEC

The 28th General Assembly of **Teratec** took place on October 10 in the amphitheater of the CEA's TGCC. After welcoming participants, representing 55 of the 85 member companies of the Association, as well as the CEA hosting this GA and the staff of **Teratec, Daniel Verwaerde** new President of the Association since October 2018 presented the agenda, emphasizing the intervention of **Jean-Marc Grognet** as General Manager of **Genopole**.

*"We have a long-standing relationship with the **Genopole,** yet it was our joint decision to strengthen the collaboration between our two structures for just over a year now. This is why we have pleasure to welcome **Jean-Marc Grognet** who will present his vision of our partnership".* The General Assembly then started according to its regular schedule.

1. Toujours de nouveaux membres

Hervé Mouren, Director of Teratec presented candidacy from four new members to join Teratec in 2019: **NAG; Tweag.io; Western Digital** and **Xilinx.** « *And things are really coming up to speed as NAG starts today with one full day seminar gathering more than 30 participants on the Teratec Campus* ».

Joining Teratec on 2019:
- **NAG**
- **Tweag.io**
- **Western Digital**
- **Xilinx**

The four companies have been unanimously introduced within the Association. Adding to our current membership, Teratec now hosts 85 organizations. Each single new member was then invited to present its activities.

• NAG : available libraries for all applications

Starting with **François Cassier,** Account Manager Europe of Numerical Algorithms Group Ltd (**NAG**) explained what this structure is about and how it works. *"Collaboration has been in NAG's genes since its creation in 1970, as it is based on a collaborative project between several British universities in the field of numerical computation. We became an independent commercial company in 1976 and notably a non-profit organization, living off income from its products and services. We achieved a turnover of £8.4m in our last financial year and now employ 70 people".*

©TERATEC

With an HPC entity in Manchester, this British structure based in Oxford has subsidiaries in the USA and Japan, as well as offices in France and Germany. NAG has 3,000 customer sites worldwide operating in industrial, financial and oil markets.

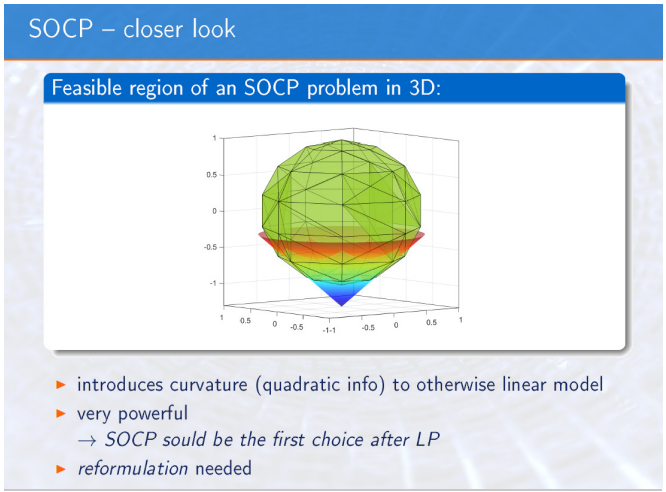
>>

NAG also works with processor manufacturers (***AMD, ARM, Intel***...) and software publishers.

« *Being a non-profit structure protects us from acquisitions, which also allows us to have a long-term vision of our missions. They have been of two kinds for 50 years: to offer our customers robust and efficient digital solutions to solve their technical problems on the one hand and, on the other hand, to help them bring the next generation of numerical computing users within their communities* ».

NAG is well known for its library which includes hundreds of mathematical and statistical components, that developers integrate into their own computing and data analysis applications to improve performance in Fortran, C, C, C++, Python, Matlab and other environments. « «*Our library is widely used by the* ***CEA's TGCC*** *hosting us for this General Assembly".*».

In addition to this library, NAG offers: high-performance Fortran compilers for error detection; **DCO** tools dedicated to algorithmic differentiation (calculation of code derivatives) as one key alternative to finite difference; **Origami** tool, providing a development and execution environment for Grid and Cloud Computing.



©NAG

NAG also offers a number of digital services such as development-oriented algorithms or other HPC-based code aiming at modernization and parallelization, technology assessment and benchmarking, as well as consulting in the implementation of HPC systems or for supply assistance.

« *We have a long-standing collaboration history with scientists and engineers around the world in the private or public, academic or industrial sectors. For example, we are currently working with the* ***Universities of Lancaster, Manchester, Oxford, Sheffield*** *and* ***Warwick*** *to enrich the NAG library with new functions, or with* ***Stanford University*** *on mathematical optimization. We also have a partnership with the* ***University of Aachen*** *in Germany on algorithmic differentiation for industrial and financial applications. Finally, we are working on European projects such as* ***POP 2*** *which as we know aims to improve codes, or* ***POEMA*** *with* ***INT Marie Curie f****or polynomial optimization* ».

NAG is very much involved in the digital community, since it participates in many normative bodies (**Fortran** and **MathML** languages, Lustre **OpenSFS** file system, **Blas & Lapack** libraries). We also collaborate with many universities to promote careers in science, technology, engineering and mathematics, and we contribute to the **Women in HPC** initiative.

Addressing a question from the audience, François Cassier said: «*We do indeed work mainly around CPUs, but we also have a consulting activity around GPU accelerators*».

- **Tweag.io : Open Source for fast-running, reliable and sustainable applications**

Then **Benjamin Robin**, Chief Financial Officer came to present **Tweag.io** and its activities. *"We are a software innovation laboratory created in 2013 by its current president* ***Mathieu Boespflug,*** *and a small group of mathematicians and experts in formal methods and distributed systems, who originally worked for the American pharmaceutical company* ***Amgen.*** *Eventually they developed an Open Source module for the* ***Haskell*** *community, an old functional programming language now resurfacing with the ability to integrate different types of languages into the same code."*

©TERATEC

This developer know-how has been noticed by industrialists who have entrusted other developments to the team, which now counts about forty engineers. « *Thanks to new technologies, we can easily find talent wherever it is, which explains why we now have employees of about twenty nationalities most of whom working from their countries of residence. Thanks to these different areas of expertise, we are able to meet challenges of our clients in very different fields (AI, Blockchain, Bioinformatics, Cloud Infrastructures...)* ».

Such complementarity of different areas of expertise allows Tweag.io to rely on 80% international customer base, supporting it from the first developments to the implementation of their applications on generally specialized themes. « *Our commitment to work for our partners is based on three qualifiers: Fast in execution; Correct in construction; Maintainable in time. This is how we have achieved two major contributions. Our team includes the inventors of* ***Nix*** *a programming language that consistently and reproducibly describes resource configuration systems in the cloud. It also includes the main contributors to the Haskell language, which rigidity experienced in the 80s and 90s is now forgotten, making its way into all subjects that require an extremely high level of data accuracy (medical analyses, Blockchain...)* ».

Tweag.io has developed a number of libraries and tools in Open Source that allow you to switch from one language to another, especially between **R** and Haskell. Tweag.io also

>>

works on transcriptions between the **Bazel** open source production engine, developed by **Google**, and Haskell, or between the **Spark** and Haskell open source computing framework.

« *We are working with our customers who generally play the game, to ensure that the majority of the projects we work on are Open Sourced. For example, we worked with* ***Pfizer*** *on a reproducible scientific process to sequence and analyze CRISPR genes in an automated way to reduce costs and delays. In bio-pharma, we are working with the Lyon-based* ***Novadiscovery*** *entity for which we have created an in-silico simulation tool of human physiology in Haskell, in order to reduce the cost of clinical trials by predicting the biochemical interactions involved in a treatment, so as to eliminate a certain number of physical tests, for what these can be managed* ».

©TWEAG.IO

« *We are working on Blockchain with* ***IOHK****, the creator of the* ***Cardano*** *crypto-currency, on the* ***Plutus*** *intelligent language that it uses as a support tool. Finally, we worked with the CEA on the industrialization of the analysis and processing of logs, bugs and anomalies generated by supercomputers which are, given the volume of data we are talking about, absolutely huge in terms of human processing, while in MLP we are able to tell the machine to find certain keywords, cross-check the information, eliminate waste and present to operators only what ''deserves'' to be processed* ».

• Western Digital : coping with booming of data volume

Davide Villa, Director EMEAI Business Development of **Western Digital** then described the activities of his company which, in addition to WD external drives, also includes the SanDisk brands, leader in USB sticks and **HGST,** formerly **IBM Storage** for mass storage.

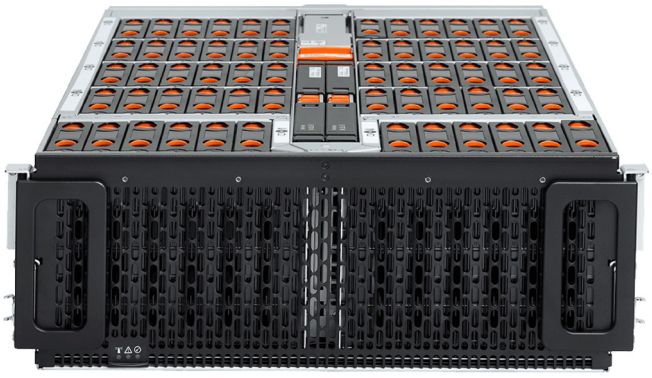
« *Our history is synonymous with innovation from first hard drives that appeared at IBM in 1956, to the very first Flash memories with* ***SanDisk*** *in 1988, to the 96 layer QLC NAND storage introduced in 2018 and backup ZB data generated today by the largest applications. Innovation is our culture that translates into Western Digital having the 4th largest commercial patent portfolio in the world, behind* ***Amazon, Google*** *and* ***Apple*** *».*

©TERATEC

An innovation that goes from chips to complete systems. « *We are one of the remaining 5 global manufacturers of Flash memories, for which we internally develop our silicon, firmware, controllers that are used in SSD semiconductor static disks and HDD hard disks. Yet beyond these components, we assemble them to make integrated systems that require a lot more storage and performance (Big Data and Fast Data)* ».

In enterprise systems, Western Digital's offering ranges from HDD and SSD storage disks to the most complex systems up to 1.4 inches in 4U racks. And the need is enormous: « *Today, 90% of the data that exist are less than 2 years old. And the trend is not going to change. It is therefore necessary to offer the market ever more efficient storage systems to absorb this phenomenal amount of data, which is currently in the order of 2.5 EB produced per day or 1018 or one billion of billion bytes!* ».

According to **IDC**, 35 ZB, or 1021 of data have already been produced worldwide, while the global installed storage base is currently able to absorb 5 ZB only. And the situation will not improve because the same organization predicts that by 2025, 175 ZB of data will have been created, while all suppliers combined should be able to increase the global storage capacity from 5 to 22 ZB. « *So we are under enormous pressure to develop larger storage systems more quickly to absorb most of this data, especially considering Western Digital represents 43% of the world's current storage capacity* ».



©WESTERN DIGITAL

>>

• Xilinx : FPGAs to accelerate data processing

Thomas Boudrot, Sr. Director Business Development of **Xilinx** closed the introductory session of new members. « *My role today is to evangelize our FPGA know-how in the world of Data Center and HPC. Our turnover is $3 billion for 4,400 employees and more than 20,000 customers. We invented the Field Programmable Gates Array (FPGA) or programmable gate network in post-production stage about 30 years ago. We are active in 8 major markets: Telecoms, which still represent more than 36% of our revenue; but also Aerospace & Defense; Industrial & Medical; Test, Measurement & Emulation; Automotive & Transport; Audio, Video & Broadcast; Consumer; Data Center & HPC* ».

©TERATEC

This last segment turned up to FPGA for the past 2 years. The very nature of the component is a massively parallel architecture, which makes it possible to respond better than traditional CPU-based architectures with ever-increasing processing needs. « *That's why, rather than continuing to offer components, we have developed a PCI Express card-based acceleration platform,* ***Alveo,*** *that can be placed directly in any server to accelerate computing, networking or even storage* ».

The development of these solutions is due to several factors: we are moving from CPU-Centric to personalized and distributed computing; the advent of Artificial Intelligence and Machine Learning, coupled with other applications is changing the situation. « *GPUs can do very well in Machine Learning alone, but if coupled to specific applications (genomic analysis; video encoding/decoding; compression/encryption...) then FPGA flexible structure becomes much more suitable. This is what we address with our acceleration cards that are deployed either in the Cloud (****Microsoft Azure, Amazon, Alibaba, Tencent****...), or On-Premises at the customer's site* ».

Depending on the application, the FPGA acceleration factor for one single CPU can range from 5 to 90. For example, a genome analysis, which took 23 hours for the American specialist **Illumina**, is now reduced to 12 minutes by adding an FPGA-based acceleration card to the calculation server. We're approaching real time!

©XILINX

One of the obstacles mentioned to the use of the FPGA is the need for programming. The fact that it can only be programmed in VHDL or low-level language has limited its use to date. « *At the beginning of October, we announced the Vitis software development platform dedicated to programming our FPGAs in* **C, C++, Python** *languages, capable of using libraries of Open Source mathematical functions, which will be mapped to our acceleration cores. And if you use frameworks for Machine Learning such as* ***TensorFlow, Cafe*** *or even* ***FFmpeg*** *for video and maintain these frameworks, our tool will directly convert your code and map it to our platforms, allowing users to target many applications such as Data Analytics, Image Processing, Video Processing, Machine Learning, Life Sciences & HPC, Financial Computing, using Open Source libraries or developed by specialized partners* ».

2. Digital challenge for Genomics in medical sciences

Daniel Verwaerde then welcomed **Jean-Marc Grognet**, Genopole's Chief Executive Officer. «*Jean-Marc Grognet is a great scientist. A biologist by training, he spent a good part of his career at the CEA, before taking over the management of the Genopole in 2017. We have since strengthened our ties and considered that a partnership between Genopole and Teratec would be beneficial for both our structures, as digital technology plays an increasingly important role in decoding the genome. That is what he wanted to remind us.*».

©TERATEC

"You introduced me as a biologist. I will therefore move away from my comfort zone by talking to you about digital technology and the benefits of this rapprochement between Genopole and Teratec, and by explaining why there is a renewed need for digital in the exploration of genomics in a global way" Jean-Marc Grognet began.

« *The Genopole is a vision that was developed 20 years ago to create a bio-cluster, i.e. an innovation campus that brings together the knowledge triangle (higher education and academic research actors with industrialists) around a focused theme of great importance (genomics) in the tightest possible physical area (Evry, Corbeil, Courcouronnes). And we also put something essential in the middle of the triangle, the patient.*

The first reason is historical because the ***AFM-Telethon*** *which fights neuromuscular diseases had its first laboratories on the Evry site. The second reason is that the Centre Hospitalier Sud-Francilien (CHSF), the largest hospital in Ile-de-France outside Paris with 1,000 beds and 3,000 carers is at the heart of our territory* » explains Jean-Marc Grognet.

>>

• 5,600 professionals around Genomics

And 20 years after, the Genopole innovation campus is a true success with 5,600 people involved, including 2,400 direct jobs. They are located in 16 academic laboratories under the supervision of **INSERM, CNRS, CEA, Evry Paris Saclay** and **Paris-Sud Universities** (1,000 people), as well as in 96 labeled companies (1,500 people) and 29 advanced technological platforms..

These make it possible to provide a laboratory or a company relying on technical skills, with state-of-the-art equipment that they would not necessarily have means to finance, helping them to operate and share with all stakeholders on each site, thus facilitating exchange of knowledge. For example, the **Ecole des Mines de Paris** has a laboratory working on materials in Evry which has been equipped with electronic microscope, now accessible to biological applications for Genopole member laboratories and companies.

®Christophe Hargoues - Genopole

There are Grandes Ecoles as well: École nationale supérieure d'informatique pour l'industrie et l'entreprise (**ENSIIE**); **Telecom Sud Paris**; Institut Mines-Télécom Business School (**IMT BS**).

Finally, it is an economic success since Genopole related companies raise an average of 70 to 80 million euros per year. The latest operation to date with **Ynsect**, a specialist in insect proteins created 5 years ago at the Genopole, has just raised $125 million.

Today the Genopole is 5 campuses which should quickly be joined by 2 others which will also house production units, in addition to laboratories.

• Genopole's DNA « is » DNA

« *Genomics has been a way of answering the very simple question of hereditary characters in nature and the noted exceptions for the past century. Why is it that the child of a blue-eyed couple does not have blue eyes? It was also known that some of the diseases had a genetic component that genomics was trying to explain. The answer was found in the DNA molecule located in the nucleus of our cells* ».

Measuring an average of 2 m, DNA is a long chain which has an impact on the amount of information carried by this molecule. The nucleus of each cell is structured into 23 pairs of chromosomes in humans. Proteins associated with DNA are present in these chromosomes. This DNA molecule is made up of a sequence of 4 types of molecular motifs (or bases) that represent all the information carried by the cell, or 3 billion base pairs!

« *The total DNA of an organism represents its genome. Almost all of the 70,000 billion cells that make up the human body have the same DNA. Deciphering the genome means acquiring these 3 billion pieces of information, or 3x10⁹* ».

It is sufficient that the sequence of the DNA molecule contains errors at one of the bases for the capacities of the proteins produced to be modified, leading to hereditary anomalies or genetic diseases.

®Christophe Hargoues - Genopole

• Sequencing human genome

Until the 1990s we tried to understand these mechanisms, but the genome had to be sequenced. If we tried to take the molecule and read each DNA fragment, at a rate of 1 second for each base, it would take about 100 years to read an individual's entire genome!

Hence the use of massively parallel computing, cutting DNA into a multitude of monobrin fragments. High-speed sequencers executing complex protocols read the sequence of bases of each fragment. The DNA fragments read are then reassembled by computer analysis. The computer reconstructs the genomes and stores them in large databases.

To carry out the first sequencing of human genome, the **Human Genome Project** was set up in 1990. About ten laboratories around the world have each dedicated themselves to sequencing a chromosome. For France, the sequence of chromosome 14 was revealed in 2001 by **Genoscope** (**CEA** laboratory located within the Genopole). The sequencing of the entire human genome was completed in 2003.

The next step is to find the meaning of this message by trying to find genes in the genome, specific sequences of bases, about 22,000 for the human being.

• From 100 M$ to 100 $ in 20 years' time

This sequencing has a cost. While first sequencing by the Human Genome Project could be estimated at $100 million, Moore's Law was then pursued gaining a factor of about 2 every 18 months until

>>

2007. Sequencing then costed about $10 million. « *That's when the American company Illumina found a shattering method to massively parallel the sequencing, which drastically reduced the costs to $1,000 per sequencing today* ».

« *This means that we are now able to have genome sequencing at a price similar to a complex biological analysis. This is becoming acceptable in conventional medical practice. The question now is no longer whether we will reach $100 one day but in how long, like 3 years, 6 years?* »

In addition to the drop in cost, the delay has also dropped drastically. The latest sequencing machines process 48 human genomes in parallel in 44 hours. « *With molecular technologies, robotics, computing and Artificial Intelligence, biology is undergoing a methodological revolution. We are entering the era of Big Data and sequencing for all* ». This allows multiple new applications to be considered.

First of all, to explore the diversity of the living world by better understanding living species, that is genomics. But also, discovering complex ecosystems formed by a whole community of organisms thanks to their DNA is metagenomics. It will also help to understand the mechanisms of living organisms by determining the function of genes identified in genomes, it is functional genomics. Finally, we will better understand the human being and his health and that is medical genomics.

• Genomics calls for the comeback of great explorers

Genomics also allows us to see what we didn't know we were witnessing before. For example, the **Tara Oceans** expedition, with its hundreds of thousands of water samples taken from all the world's seas and the sequencing carried out identified 117 million genes, more than half of which were previously unknown. We can thus identify species that we had never seen before! Similarly, the sequencing totally changed the vision of the intestinal microbiota. From a few species of bacteria identified 40 years ago, we have grown to several hundred today and many more remain to be discovered!

©S.Bollet-Tara Expeditions

• Tailoring the medical approach

Sequencing each individual's genome will revolutionize medical practices by personalizing diagnosis and treatment. But for this to be effective, it will be necessary to be able to know and interpret in real time the information contained in the genome, hence the growing need for computing power and performance. « *The idea is to be able to process heterogeneous data flow more and more quickly using algorithms, particularly Artificial Intelligence and bioinformatics tools in a context of increasing medical interdisciplinary practices to make the right diagnosis* ».

©Tara Expeditions

Analyzing its genetic characteristics and knowing the genomic rearrangement of a tumor can provide valuable information to physicians to guide treatments. To do this, it will be necessary to work upstream on the acquisition, interpretation, integration and presentation of data, so as not to be overwhelmed with massive and useless information.

This genomic or personalized medicine is divided into two main branches. Genome analysis will make it possible to precisely detect an individual's predispositions to pathologies according to his genetic identity card and to carry out preventive medicine by giving him behavioral advice or even prescribing preventive treatments to reduce the risks.

Genome analysis will also make it possible to detect rare diseases by making the right diagnosis on the first time. In the case of a proven disease, genome analysis will make it possible to evaluate different treatments and choose the one that will be most effective and fastest with the least side effects, namely curative medicine. « *In the last two years or so in oncology, it has been possible not simply to treat a kidney or digestive pathology but apply genetic mutation with a dedicated drug regardless of the location of the tumor. This is a radical change in the vision of treatment* ».

All this is at the heart of the great **France Médecine Genomique 2025** plan, partly born in Genopole, which aims to provide France with a large number of platforms dedicated to sequencing that will send the information obtained to a Data Analyzer Collector (**CAD**) which will process and interpret it for each patient and feed a research centre, **Crefix**, which will validate procedures and devices.

>>

• Genomics and Big Data

Unlike many Big Data applications where we process some information on a very large number of individuals, in genomic medicine we process a very large number of descriptors (genetic code, mutations...) on a very small number of individuals.

« *It is estimated that we will have 1021 bytes of information to process per individual to multiply by several hundred thousand patients per year. In addition to the processing power, this will also require very large storage resources of about 10 ExaBytes per year* ».

This also raises questions because if acquisition is no longer an issue, the data must remain confidential yet to be attached to a single medical file in order to be of relevant interest. How do we guarantee this confidentiality? To whom do these data belong, to the patient, to the doctor who prescribed them, to the sequencing center, to the community and therefore to the State? Who will be responsible for keeping them and with what sustainability, because today's information could be useful to treat a pathology in 30 years' time? Who will interpret them and who will reinterpret them as science evolves?

• No genomics without digital technology

« *This is why we have approached Teratec more closely and you have understood that the medicine of the future will be based on genomics, so that is also why we need ever more efficient computing technology with secure processing of very large volumes of data, while guaranteeing their sustainability over the very very long term. These are concerns that are at the heart of Teratec's activities. We are also planning to create the world's first "digital genomics" institute to meet these needs because we have some of the most powerful sequencing centers operating in Europe as well as structures such as* ***Teratec*** *on our territory, with very strong skills in high-performance digital computing (HPC)* » concluded Jean-Marc Grognet.

Following a question from **Christian Saguez**, Jean-Marc Grognet clarified the fact that studying the genome would not deal with everything in personalized medicine. « *Genomics alone will not provide the answer to everything. Over the past two decades, it has been discovered that not everything lies in the genome. Contrary to what we thought, our genome is not immutable, while we can experience changes in our genome which could be transmitted to other generations through our life experience and our exposure to external factors (ionizing radiation, chemicals...)* ».

This goes against the Lamarck/Darwin dogma where Lamarck defended the heredity of the acquired characters while Darwin said it is immutable. « *We thought Lamarck was wrong, but now we realize that the answer is much less firm. It has been shown, for example, that a mouse can transmit to its offspring and to their offspring, the fear resulting from a trauma. Therefore there is a transmission of the acquired characters. It was also discovered that viruses could integrate into the genome and thus be transmitted to a subsequent generation. In addition to genomics, tomorrow's medicine will therefore incorporate other parameters such as life experience and exposure to external factors. This implies the processing of multiple data, both personal and collective* ».

Following a question from **Karim Azoum** on the use of Machine Learning and AI in Genopole's work, Jean-Marc Grognet clarified: « *I don't know if we can talk about Artificial Intelligence, but we also have a whole activity that consists in knowing the interaction networks between genes. This part is extremely important because the interaction of gene in networks has a responsibility in pathologies. Other institutions are also working on the prediction of pathologies* ».

3. Europe major initiatives

Daniel Verwaerde then took the floor to present the Association's activity report.

©TERATEC

« *With regard to major European initiatives and following European elections of recent weeks, the highlight has been the change of the Commission moving from the* ***Junker Commission*** *(2014-2019) to the* ***Van Der Leyen Commission*** *(2019-2024). This is now reflected in fairly significant changes in activities of our association's members and how they are linked both in terms of HPC and more generally in terms of information technology. In the previous Commission, HPC and* ***DG Connect*** *being responsible for it were linked to the Commissioner for Digital Society and Economy,* ***Mariya Gabriel,*** *who reported to the Vice-President for the Digital Single Market. This paved the way to the creation of the Joint Undertaking* ***EuroHPC****, independent of the Commission and led by DG Connect. The current change for this activity in HPC and more generally for the digitization of the economy and society is considered even more important by the new Commission. Among its two Vice-Presidents, the quasi-operational management will be carried out by Margrethe Vestager in charge of preparing Europe for the Digital Age, and no longer by a simple Commissioner. Executing this role under her responsibility, HPC and DG Connect will be attached to the Commissioner for the Internal Market, Industry, Defence and Space*».

• Preparing the future of EuroHPC

It should be recalled that EuroHPC brings together the Commission and 29 Member States with a budget of around €1 billion over the period 2019/2020 half of which is financed by the Commission and by €7 billion for the period 2021/2027. The mission is twofold: to acquire very large computers in Europe and be in the Top 5 worldwide, making them available to researchers and industry; and to develop in Europe a complete industrial capacity to produce, sell, install and maintain these supercomputers. *"Mariya Gabriel used to say that 30% of the world's supercomputers were used in Europe and that her ambition was for 30% to be produced in Europe* ».

>>

The EuroHPC programme is progressing rapidly. For machines with 150 to 200 PFlops capacity, 3 sites have been selected (Kajaani in Finland, Barcelona in Spain, Bologna in Italy) with tenders being launched, and installations should take place by the end of 2020. The total budget over 5 years is around €650 million. « *France and Germany are not part of these three consortia, preferring to devote for the next phase which will involve exaflops machines 5 to 10 times more efficient". 5 other machines from 5 to 30 PFlops will also be acquired for a total budget of 180 M€ over 5 years. They will be installed in Luxembourg, Portugal, the Czech Republic, Slovenia and Bulgaria from mid-2020. "As these sites do not necessarily have a strong track record in operating this type of machine, Teratec members could there find new opportunities with this expertise* ».

« *We are currently preparing the evolution of EuroHPC for the coming decade in Brussels. It is planned to continue to purchase machines so that there will always be one of the 5 most efficient machines in Europe. We also want to have a technological chain at the required level to achieve this with one adapted application development ecosystem. Finally, the Commission wants all these machines to rely on extremely fast interconnection mode to facilitate access and use* ».

Teratec's role in EuroHPC is threefold. « *First of all as I have just mentioned this, we actively participate in 2 Advisory Groups (Infrastructure Advisory Group, **INFRAG**; Research and Innovation Advisory Group, **RIAG**), whose mission is to draft strategic documents for the selection of platforms and the definition of the technology roadmap, in order to continue to compete on equal footing with the other world leaders in the field. We are also working on drafting the"Vision Paper 2021/2030" devoted to the EuroHPC roadmap* ».

©EUROHPC

• Expressing the needs of users

In parallel, on request from the Commission and DG Connect, Teratec was asked to produce a document expressing the industrial need for the use of EuroHPC supercomputers. By so doing, the Commission objective is to define its policy of opening up these supercomputers to industry, from SMEs to large groups.

Thirdly, *"We are also participating in 3 major Calls for Proposals launched last July: EuroHPC 03 "Industrial software codes for extreme scale computing environments and application"; EuroHPC 04 "Competence centres"; EuroHPC 05 "Stimulating the innovation potential of SME's" as a continuation of the **FORTISSMO** European project, cousin of the French **SiMSEO**)* ».

« *In this context, Teratec is obviously ready to act as facilitator of project development by its members within the framework of various EuroHPC actions* ».

• Becoming French HPC Competence Center

The EuroHPC 04 action aims to create a single HPC Competence Centre in each European country. It will act as an interface with the Commission and the other Member States to implement European policy in this area, providing the following services to all countries' economic actors:

- Disseminate the uses of HPC/HPDA technologies;
- Offer a complete package of services (consulting, specific studies);
- Facilitate manufacturers' access to HPC/HPDA hardware and software capabilities;
- Provide support for the deployment of tools;
- Provide training and information;
- Collect information to measure the contribution of the Competence Centre.

« *The French Government has appointed Teratec towards Brussels authorities as its banner for French Competence Centre. We have therefore made a proposal to the French public authorities to organize it with **CERFACS,** whose work with industrialists over the years makes it more than legitimate within such structure. The organizing committee will have 5 members: **Teratec** (Presidency); **CERFACS** (Vice-Presidency); **France Industries**, governing the 17 French industrial branches; **GENCI,** as contracting authority for large computing equipment for research; **Maison de la Simulation**, an organization that includes the **CNRS,** the **CEA** and the Universities of **Paris-Saclay, Paris-Sud** (UPS) and **Versailles Saint-Quentin-en-Yvelines** (UVSQ)* »,

The Competence Centre's objective is to have a permanent link with the EuroHPC «Mirror Group», as well as with major research organizations able to provide technologies and knowledge (**CEA, CNRS, INRIA, AFneT, AMIES**, technical centers, competitiveness clusters, IRT...). Similarly, the Competence Centre will bring together the industrial world and those of education and research through regular meetings. The budget (EuroHPC and State) will be €1 million per year, which will make it possible to finance 8 people with 200,000€ investment.

« *The European Union and EuroHPC are important for us because we feel that the French-French channels will become European, and if we want our action to continue and benefit the French economic frameworks it is extremely important to play the EuroHPC game completely* », concluded Daniel Verwaerde on this chapter.

4. Latest updates on Teratec activities

Hervé Mouren, Director of Teratec then took over with sharing an overall focus on current activities of the Association.

• European programmes

« *For the past 4 years, Teratec has been participating in European programs such as **POP, Excellerat** and **Focus CoE**. POP is one of the Centers of Excellence (CoE) of the H2020 Research Framework Program. With cross-sectoral reach, it is dedicated to application performance. It's a remarkable mechanism* ».

>>

POP allows any organization to meet up with other relevant bodies to do performance analysis in order to optimize their software. They analyze the performance of a code for free and tell the author what to do and how to improve it significantly. This service is paid for by the European Commission, which has allocated a budget of €7 million. « *I strongly urge you to use this service to enhance the performance of your codes* ».

Another European program, **Excellerat** is a Centre of Excellence promoting digital simulation in engineering applications using 6 calculation codes and 11 typical user cases. Finally, **Focus CoE** is a kind of control tower that looks at how the results of all European sectoral CoEs are developing.

« *Among these operations, all of which of high quality, plus the launch of the EuroHPC operations adding to the role Teratec will have as a Competence Centre, our image will be transformed. Teratec will continue being seen as the French Competence Centre on these subjects, and with further request from the Commission to remain involved in the coordination mechanism of the various Competence Centres in Europe. Thus, our European image will also be strengthened for us to be granted very significant resources* ».

Despite these ambitious European aims, we must not forget French activities. « *The **SiMSEO** program which ended at the end of September was a great success. We were asked to step into the simulation arena with 100 SMEs in the worlds of mechanics and construction. We tipped over 360 of them! To this end, we relied on both professional organizations and specialized publishers to provide catalogues of validated software and services offers. SMEs have invested more than €2.7 million, 50% of which was subsidized by **BPI France** as part of the **Investissements d'Avenir*** ».

• Industrial initiatives

There has been a slight slowdown in the launch of industrial initiatives this year, as it is now necessary to bring forward those launched.

The one on the **Teratec Quantum Computing Initiative** (TQCI) which brings together major companies (**Atos, CEA, EDF, IFP, Total** ...) should be reinforced by a public initiative, visible at the end of the year, aimed to give us the opportunity to build a real industrial project.

©TERATEC

On Additive Manufacturing, a first work was done showing the breadth of interest for the subject and the points of application that one could imagine in an industrial project.

As for the others (Autonomous Systems, Natural Resources, Agri-food), they are in the start-up or design phases.

• Teratec Campus

Hervé Mouren then mentioned the positive functioning of the Teratec Campus with 22 structures installed to date, including 4 research laboratories and bringing together 250 people.

He highlighted an initiative that will take place on the Teratec Campus on November 28: « *We will be hosting 200 students to meet the recruitment needs of companies, whether they are small or large. The Universities of **Versailles Saint-Quentin-en-Yvelines** (UVSQ) and **Leonard de Vinci** will bring their final year Master students to the Campus with their teachers, as part of their curriculum. They will meet Teratec member companies who will present their know-how to this audience* ».

Teratec Campus

Technology Entreprises

• Atos • CEA • ESI • Intel • Egis •

Business Incubators

• AS+ • Assystem Technologies • Avise Consei • CesiMat • • Distène • ETP4HPC • Fullscale • Ingérop • Mantenna • • MicroTrans • NumTech • ParaTools • RTSolution • • Sema Software • Synomen • Teratec • WizYoo •

Industrial Research Laboratories

• Extreme Computing • Exascale Computing Research • • Intel Big Data Analitycs • CesiMat •

• Communication and Forum Teratec

Teratec's mission for communication follows its normal course with: publication of a 7th supplement of *L'Usine Nouvelle* with *Industrie & Technologies* devoted to **One year of simulation** and distributed in 55,000 copies; the update of the Teratec brochure; the distribution of 6 Newsletters between January and June 2019 distributed to 60,000 professional contacts

« *I would like to highlight the very good performance of the **Forum Teratec 2019** with some original features, including the roundtables which have received very positive feedback and should be repeated next year. Concerning the workshops and the exhibition, we reach the limit of the system in the current premises. We will therefore have to think about how we will develop it* ».

He then announced that the **Forum Teratec 2020** should take place on June 16 and 17, 2020 at the Ecole Polytechnique. Calls for papers, workshops, sponsorships and exhibitors are being launched.

Finally, **Hélène Bringer Garlain**, the Association's Treasurer confirmed that the 2019 budget is being implemented as planned.

Daniel Verwaerde then closed this General Assembly, recalling that the Association exists only thanks to its members and their active participation, pointing out that he remains attentive and listening to everyone.

Jean-François Prevéraud