H2020-FETHPC-3-2017 - Exascale HPC ecosystem development

EXDCI-2

European eXtreme Data and Computing Initiative - 2

Grant Agreement Number: 800957

D4.7
Second report on EXDCI-2 technical workshop

Final

Version: 2.2
Author(s): Corentin LEFEVRE, Néovia Innovation
Date: 06/05/2020

Project and Deliverable Information Sheet

EXDCI Project	**Project Ref. №:FETHPC - 800957**	
	Project Title: European eXtreme Data and Computing Initiative - 2	
	Project Web Site: http://www.exdci.eu	
	Deliverable ID: <D4.7>	
	Deliverable Nature: <DOC_TYPE: Report>	
	Dissemination Level: PU *	**Contractual Date of Delivery:** 30 / 04 / 2020
		Actual Date of Delivery: 08 / 05 / 2020
	EC Project Officer: Evangelia Markidou	

* - The dissemination level are indicated as follows: **PU** – Public, **CO** – Confidential, only for members of the consortium (including the Commission Services) **CL** – Classified, as referred to in Commission Decision 2991/844/EC.

Document Control Sheet

Document	**Title: Second report on EXDCI-2 technical workshop**	
	ID: D4.7	
	Version: <2.2>	**Status:** Final
	Available at: http://www.exdci.eu	
	Software Tool: Microsoft Word 2019	
	File(s): EXDCI-2-D4.7-Second Report on EXDCI-2 technical workshop.docx	
Authorship	**Written by:**	Corentin Lefèvre, Néovia Innovation
	Contributors:	François Bodin, Université Rennes 1
	Reviewed by:	Jean-Philippe Nominé, CEA/ETP4HPC Katia Soto, PRACE
	Approved by:	MB/TB

Document Status Sheet

Version	**Date**	**Status**	**Comments**
1.0	28/04/2020	Draft	
2.0	04/05/2020	Reviewed	Implementations of the corrections and changes asked by the reviewers
2.1	05/05/2020	Final version	Minor editing
2.2	06/05/2020	Final version	Implementations of the corrections and changes asked by the reviewers

Document Keywords

Keywords:	PRACE, , EuroHPC JU, Research Infrastructure, Digital continuum, Vision, Roadmap, Artificial Intelligence, Big Data, Quantum Computing, HPC Ecosystem, BDEC

Table of Contents

List of Figures

References and Applicable Documents

[1] https://exdci.eu/
[2] https://prace-ri.eu/
[3] https://www.etp4hpc.eu/
[4] https://exdci.eu/sites/default/files/public/files/d4.6_v1.2.pdf
[5] https://eurohpc-ju.europa.eu/
[6] https://www.exascale.org/bdec/
[7] https://openai.com/blog/ai-and-compute/
[8] https://www.top500.org/system/179699
[9] https://www.seagate.com/fr/fr/our-story/data-age-2025/
[10] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[11] https://www.etp4hpc.eu/pujades/files/ETP4HPC_SRA4_2020_web(1).pdf
[12] https://exdci.eu/newsroom/news/bigdft4chem-wins-exdci-call-spin-initiatives-spi

List of Acronyms and Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
AQMO	Air Quality and Mobility
BDEC	Big Data and Extreme-scale Computing
CEA	Commissariat à l'Energie Atomique et aux Energies Alternatives
CHEESE	Center of Excellence for Exascale in Solid Earth
CI	Cyberinfrastructure
CNRS	Centre National de la Recherche Scientifique
CoE	Centres of Excellence for Computing Applications
CSA	Coordination and Support Action
DL	Deep Learning
EC	European Commission
EI	E-Infrastructure
EOSC	European Open Science Cloud
EPI	European Processor Initiative
EsD	Extreme scale Demonstrators
ETP4HPC	European Technology Platform for High Performance Computing
EU	European Union

EuroHPC JU	European High Performance Computing Joint Understanding
Eurolab4HPC2	Consolidation of European Research Excellence in Exascale HPC Systems 2
EXDCI-2	European eXtreme Data and Computing Initiative 2
FENIX	
FET	Future and Emerging Technologies
FLOPS	Floating Point Operations Per Second
FPGA	Field-Programmable Gate Arrays
GA	Grant Agreement
GDPR	General Data Protection Regulation
H2020	Horizon 2020 – The EC Research and Innovation Programme in Europe
HPC	High Performance Computing
HPC-GIG	High Performance Computing Governance Intelligence Gathering
HPDA	High Performance Data Analytics
ICT	Information and Communication Technology
IDC	International Data Corporation
IoT	Internet of Things
IP	Intellectual Property
LHC	Large Hadron Collider
LOFAR	Low Frequency Array
MAX	Materials design at the eXascale
MFF	Multi Financial Framework
ML	Machine Learning
NM	Nanometer
NVM	Non Volatile Memory
OPC	Organisation and Program Committee
PHIDIAS	Prototype of HPC/Data Infrastructure for On-demand Services
PPP	Public Private Partnership
PRACE	Partnership for Advanced Computing in Europe
Qbits	Quantum bits
R&D	Research and Development
SHAPE	SME HPC Adoption Programme in Europe
SKA	Square Kilometer Array
SME	Small and Medium Enterprise
SRA	Strategic Research Agenda
USA	United States of America
ZB	ZetaByte

Executive Summary

The EXDCI-2 second technical workshop was dedicated to reviewing the elements of what could be the European HPC landscape beyond the race for exascale. Scientific, technical and organisational trends have been assessed to forge the elements of a renewed version of the directions the European HPC ecosystem should take to fulfil its long-term objective to remain one of the best ones, and even a leading HPC ecosystem in the world, and therefore to contribute to accelerate and sustain its impacts towards scientific communities, industries and civil society.

On the technical and applications' sides, the main shift is to the upcoming of the digital continuum era where all elements, from the edge computing to the HPC are interrelated. The digital continuum will lead to a new generation of data / computing platforms that would have major impacts but also carry significative challenges. Also the question of interrelation of AI in the continuum has been one of the key discussions during the workshop.

On the organisation side, the creation of the EuroHPC JU has been a key asset for the structuration of the EU HPC ecosystem. The ecosystem has strengthened in the past years through the success of many activities, including the creation of the EuroHPC Summit Week, and novel initiatives appeared to take the most of the efforts already done within the European Union. Three types of connection are seen of utmost importance for the EU HPC community:

- Within the ecosystem itself
- Connection beyond its specialties with other domains such as AI, Big Data and cybersecurity
- Connection beyond the European area at the international level.

The end of Horizon 2020 and forthcoming Horizon Europe is also a key element structuring the EU HPC landscape.

The main output of the workshop is the preparation of a renewed EXDCI-2 vision that aims to address the above-mentioned challenges of the European HPC ecosystem that could lead to a shared European roadmap for the digital continuum era.

1 Introduction

The European project EXDCI-2 sets up actions to coordinate the European HPC ecosystem with important enhancements to better address the convergence of big data, cloud computing and HPC. EXDCI-2 strategic objectives are:

- The development and advocacy of a competitive European HPC exascale strategy
- The coordination of the stakeholder community for European HPC at the exascale

EXDCI-2 mobilizes the European HPC stakeholders through the joint action of PRACE and ETP4HPC. The project promotes global community structuring and synchronization in domains such as HPC, Big Data, cloud and embedded computing, for a more competitive related value chain in Europe.

To achieve its objectives, the project develops a HPC technology roadmap addressing the convergence with HPDA and the emergence of new HPC uses. The project also delivers application and applied mathematics roadmaps that will pave the road towards exascale simulation in academic and industrial domains. It develops a shared vision for the future of HPC that increases the synergies and prepares for targeted research collaborations. EXDCI-2 works to increase the impact of the H2020 HPC research projects, by identifying synergies and supporting market acceptance of the results. EXDCI-2 also contributes to the formal monitoring and impact assessment of the EU HPC programme (PPP run between 2014 and 2018).

The project also contributes to the international visibility of Europe and wants to improve HPC awareness.

In EXDCI-2, a transverse work package has been designed. Transverse coordination is important for the HPC community to share its visions alongside the HPC value chain, as well as to exchange with other communities, such as Big Data and IoT.

Among EXDCI-2 transversal actions, two technical workshops are organized during the course of the project. These workshops bring together all EXDCI-2 participants, as well as invited European and international HPC stakeholders. Through these dedicated meetings, all the participants can share information on various topics and aspects of HPC, including funding models of relevant research and needs/challenges with respect to efficient software for future exascale systems.

In the course of the EXDCI-2 project, the first project technical workshop was organized in Brussels on the 28th September 2018[1].

The second project technical workshop was organized in Barcelona on the 2nd and 3rd of December 2019.

The event was the occasion for all invited stakeholders to share over the work done during the project and its first results, as well as to prepare coordinated actions and exchange views on a new vision for the European HPC beyond exascale: the continuum computing paradigm.

[1] For more information, the related deliverable "D4.6 – First report on EXDCI-2 technical workshop" is downloadable online on the EXDCI-2 website.

2 Setting and Agenda

The second EXDCI-2 technical meeting took place in Barcelona on the 2nd and 3rd of December 2019 at the Universitat Politècnica de Catalunya. The on premises meeting was completed by remote connection setup with the aim to gather more expertise into the meeting.

The number of participants in the meeting was 21, of which 6 were women. 5 persons participated through conference call. The names and affiliations of the participants can be found in annex 5.1.

Figure 1 - Technical workshop group photo

Agenda of the workshop

The indicative agenda of the workshop is given below. Some minor modifications occurred during the course of the meeting.

Monday, December the 2nd, 2019

Aules Vertex – Room VS208 – Catering Room VS217

1st day – An assessment of the progress made by the EU HPC ecosystem towards exascale		
10:45 – 11:00	Welcome and registration	
11:00 – 11:45	Foreseen technologies in a convergent world of distributed models: Upcoming HPC technologies: Upstream technologies	
11:00 – 11:45	RTOs Upstream technologies and EPI evolution	Jean-François LAVIGNON
11:45 – 12:15	EuroHPCSW 2020	
11:45 – 12:15	Discussion on EuroHPCSW organisation and agenda	Serge BOGAERTS
12:15 – 12:45	BDEC	
12:15 – 12:45	BDEC-2 status and next steps	Mark ASCH
12:45 – 14:00	Lunch	
14:00 – 15:00	Foreseen technologies in a convergent world of distributed models: Upcoming HPC technologies: Legacy Codes	
14:00 – 15:00	Legacy Codes	Guillaume COLIN de VERDIERE
15:00 – 15:25	Break	
15:25 – 17:45	Foreseen technologies in a convergent world of distributed models: Trends and influencing factors	
15:25 – 15:45	Introduction	Michael MALMS
15:45 – 16:00	HPC and the digital continuum	Marc DURANTON
16:00 – 16:15	Application co-design	Erwin LAURE
16:15 – 16:30	Data everywhere	Gabriel ANTONIU
16:30 – 16:45	AI everywhere	Maria PEREZ
16:45 – 17:00	Trustworthy Computing	Jens KRUEGER
17:00 – 17:45	Open discussion	Michael MALMS

17:45 – 18:00	Break	
18:00 – 19:30	Evolution of the ecosystem	
	Discussion on post EXDCI2-CSA	Serge BOGAERTS
20:30	Dinner	

Tuesday, December the 3rd, 2019

Aules Vertex – Room VS208 – Catering Room VS217

2nd day – EU HPC Ecosystem towards and beyond exascale -Next steps and vision		
09:00 – 10:30	Foreseen technologies in a convergent world of distributed models: Applications	
09:00 – 10:00	The place of AI in the future, continuum e-infrastructure	Mark ASCH François BODIN
10:00 – 10:30	Foreseen activities on the applications' side	Stéphane REQUENA
10:30 – 11:00	Break	
11:00 – 12:15	Building on the ecosystem's results	
11:00 – 11:30	Scientific Case and Industrial Collaboration	Stéphane REQUENA
11:30 – 12:15	How best to foster the HPC ecosystem and leverage the results of EU projects - Results for the spin-off initiative - FET projects results - Areas where the EU HPC ecosystem can lead the promotion of de facto standards / standardization of APIs	Maike GILLIOT Thierry BIDOT Jean-François LAVIGNON
12:15 – 13:30	Lunch	
13:30 – 14:15	Elements for a new vision	
13:30 – 14:15	Towards continuum HPC – a new vision	François BODIN
14:15 – 14:45	EXDCI-1 and EXDCI-2 legacy	
	Lessons learned and takeaways	Serge BOGAERTS

3 Report on the 1st day – An assessment of the progress made by the EU HPC ecosystem towards exascale

3.1 Foreseen technologies in a convergent world of distributed models – Upcoming HPC technologies: upstream technologies

In the new context of the European Green Deal and the focus puts on strengthening European sovereignty, reflexions upon upstreaming technologies should not be conducted in silos – e.g they should not only assess the promises and steps of single technologies – but rather adopting a strategic viewpoint that includes the creation of new value chains for the EU HPC ecosystem.

EXDCI-2 gathered various stakeholders during a workshop held in Brussels on the 5th and 6th of November 2019 that made a review of upstreaming HPC technologies and identified three axes to increase the computional power:

- New representation of information (qbits, analog coding, spikes)
- New architectures (processor in memory, data flow, neuromorphic, graph computing, simulated annealing, quantum annealing, quantum computing)
- New devices (NVM, silicon photonics, memristive technologies)

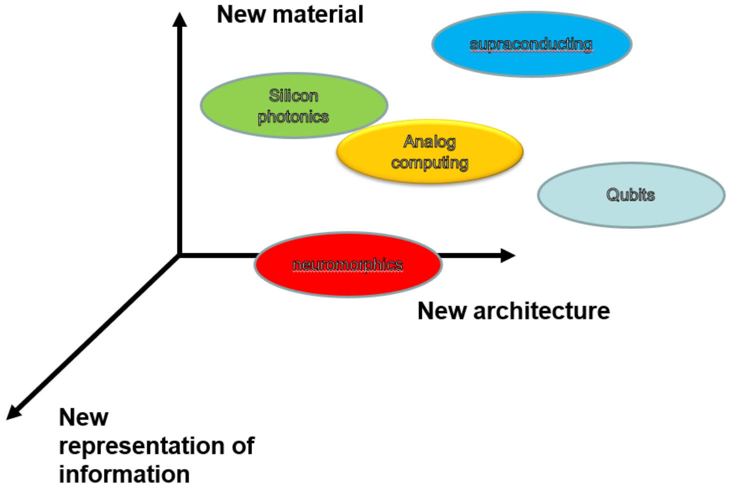

Figure 2 - Upstream technologies directions

Working on a combination of these axes, several solutions are under scrutiny or being developed with a research ecosystem already in place at the European level – communities are active on photonics or electronics for instance.

For each new path, and more importantly regarding their interrelations, it shall be assessed whether the investments to be made are worth it. These assessments shall answer the questions:

- What is the rationale to develop such a technology? (e.g who needs it?)
- What is its development timeframe? (and how it imbricates with other technologies' development?)
- How do we accelerate the uptakes of such a technology?

As a result, the EU shall be able to:

- Set a list of promising technologies/architectures that are relevant for future HPC/edge systems and meaningful to develop in Europe
- Establish indications on how the listed technologies/architectures could emerge

- Be in a position to write short but credible foreseen success stories for some of these high potential technologies.

The HPC ecosystem cannot be alone in such a task as the development of new technologies/architectures are mostly to be driven by other domains (e.g. Big Data) – which calls to pursue and strengthen the existing collaborations that exist between the EU HPC ecosystem and related domains.

3.2 EuroHPC Summit Week 2020

The EuroHPC Summit Week 2020 was prepared to take place from 23rd to 27th of March 2020 in Porto[2]. Four main contributors went together to organize this week which aims to gather the European HPC ecosystem with the objective to share its latest achievements and pave the way for its future.

Those contributors were the EuroHPC JU and the European Commission, PRACE, ETP4HPC and EXDCI-2.

EXDCI-2 took the opportunity of the second technical meeting to define the scientific and technical programme the project would propose during the week. This programme covers the structuring and trending topics shared among the EU HPC community and is organized in 6 sessions:

- Applications
- Delivering HPC in a sustainable manner
- Training
- Urgent Computing
- Continuum Computing
- Emerging technologies

The proposed sessions alongside with proposition of speakers was meant to be communicated to the EuroHPC Summit Week OPC, with an attention to invite initiatives and projects close to the EU HPC Ecosystem (FENIX, EOSC) and also link with the running CoEs (propositions have been received from CHEESE and MAX for instance).

The EuroHPC Summit Week is seen as one of the vectors to engage with other communities such as AI.

3.3 BDEC2

BDEC2 is an international think tank that was created to first address the convergence of compute and (big) data in the road to exascale. EXDCI-2 organizes the European participation to this global group, including the organisation of BDEC2 session held back-to-back with the EuroHPC summit week since 2019.

The BDEC’s role, as a think tank, is to prepare position papers and roadmaps that are disseminated through every main HPC ecosystems worldwide. BDEC2 continued the work done during BDEC-1 with a change of focus coming from convergence to the concept of continuum e-infrastructures.

[2] Note: due to the outbreak of the Covid-19 pandemic, the EuroHPC Summit Week 2020 has been postponed to 2021.

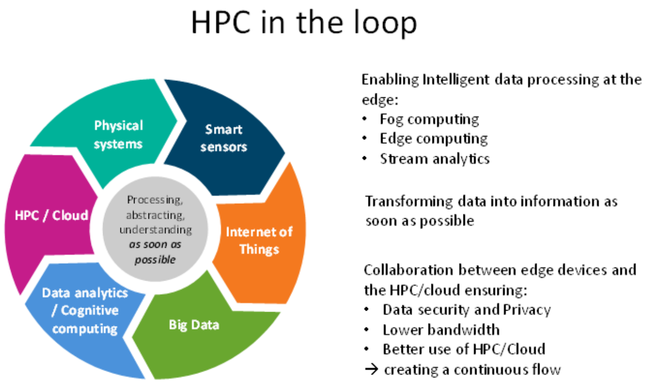

Figure 3 - Digital continuum - HPC in the loop

Digital continuum is identified as one of the main challenges to be tackled for the next HPC era as existing cyberinfrastructures were not designed to adequately deal with edge to cloud/HPC workflows, especially not extremely data-intensive ones.

The current objective of the BDEC2 is to work on Demonstrators that are proof-of-concept platforms. Such a demonstrator shall produce a working version of an international, federated, continuum-spanning platform that can be cooperatively operated and managed and that engages stakeholders at all levels.

These demonstrators will show examples of how identified technical and scientific problems caused by the digital continuum can be solved. The main identified issues are:

- An end-to-end problem (from AI at the edge to HPC in the cloud)
- A software stack problem
- A resource allocation problem due to the building of on-demand shared infrastructure
- A data movement and logistics problem, in both directions
- Robustness, security, sustainability and reliability of large, interlinked, composed infrastructures
- AI is everywhere, and new infrastructures must support monitoring and control; infrastructure learns (not just the applications).

Potential use-cases for such demonstrators have also been listed:

- Big Instruments:
  - Radio telescopes – LOFAR, SKA.
  - High energy physics – LHC.
  - Satellite data – Copernicus, SWOT, HIMAWARI, …
  - Climate, earth sciences, oceanography.
- IoT-like:
  - Personalized medicine.
  - Autonomous vehicles.
  - Predictive maintenance.
  - Precision agriculture.
- Digital Twins...

The BDEC2 current action is to continue its work with those demonstrators. A recent focus has been done on AI in the continuum with the identification of three interrelated categories.

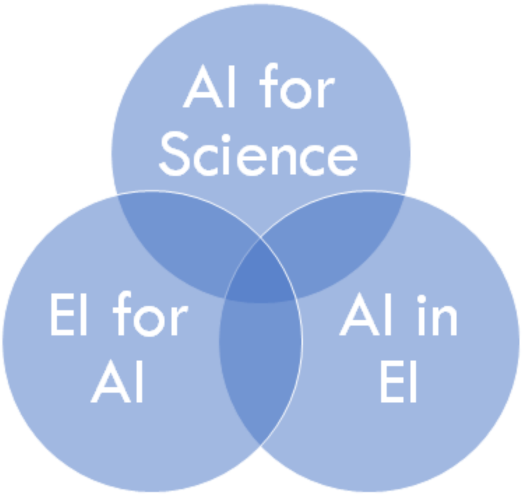

Figure 4 - AI in the digital continuum

- The role of AI for science is mainly on the applications' side with functionalities such as the steering of simulations, embedding machine learning in simulation methods, tuning application parameters, generate models to compare with simulations…

  Refactoring codes with AI has been discussed but does not seem to be operational in the short-term.
- The role of AI for the E-Infrastructure would be in improving system operations (on costs, reliability, security), as well as to improve applications and workflow performance. It could also lead to the implementation of closed-loop systems and test-beds.
- Finally, the role of EI for AI is distributed among five topics: programming, distributed service composition, data, communications, authentication and authorisation.

BDEC2 was supposed to continue its work with a workshop to be held in Porto from the 24th and 26th of March, 2020 and organized by the EXDCI-2 project.[3]

A community report is scheduled for the fall 2020 and shall be presented at Super Computing 2020 in November while spanning to all HPC ecosystems.

3.4 Foreseen technologies in a convergent world of distributed models – Upcoming HPC technologies: legacy codes

EXDCI-2 has set up a dedicated group to investigate the European legacy codes and the technical debt that the EU could have when moving to the next generation of HPC.

The inception of the group takes its roots on evaluating the cost of porting current application codes to the next generation of HPC systems, which has been so far seldom studied nor included and thus underestimated.

[3] Note: due to the outbreak of the Covid-19 pandemic, the BDEC-2 session in Porto has been postponed to October 2020.

The first problem to be solved was to define the complexity of this migration and to build a common metrics to address a wide diversity of situations. Looking at the European richness of HPC applications, a large variety of cases was to be addressed and compared.

The group's first action has been to select universal parameters that take into account the main difficulties faced by a development team in migrating codes to the next generation of machines. Those parameters should be easily determined by the concerned software development teams, also for an envisaged survey to be less time consuming and thus to collect broader answers from the European HPC ecosystem. Simplicity was the key, which, in return, means that the metrics would not be highly sophisticated, yet strong enough for the evaluation purposes.

Four parameters were retained to define the metrics:

- The age of the code in years – named A
- The length of the source code, in millions of lines, excluding third party libraries – named L
- The team capability or confidence to develop accelerated codes. This parameter is an estimation made by the concerned team and its value goes from 0 – the team cannot port the code – to 1 – the team has the expertise and skills to perform the job. This parameter is named TC
- The percentage of acceleration, from 1 for a code fully ported to accelerators to 0 for code not ported at all – named P

With these four parameters, a metric has been created - named "code Viscosity" by analogy to the viscosity of a fluid that quantifies its resistance to movement. The more viscous is a code, the more expensive it will be to do the port, both in terms of time to completion and in the manpower needed, the less it is prone to the adoption of newer technologies. In other words the more viscous, the higher the technical debt of the code that will have to be paid sometime in the future.

Code Viscosity – named V – is defined by the following formula:

$$V = \frac{A \times L}{TC \times P}$$

Figure 5 - Code viscosity formula

The group then decided to focus on the legacy codes that are defined by three criteria:

- They are old codes – e.g. more than 10 years old
- They are vital to their organizations
- They must be ported to the new generation of machines.

Due to their characteristics, legacy codes are particularly vulnerable when it comes to migrating to a new generation of machines. During their lifetime, these codes go through several generations of computers, several generations of programming environments, several standards for different APIs or other elements, several generations of software developers and are manipulated by changing teams over time. All these layers strongly impact the viscosity measure of a legacy code.

Legacy codes are not a new subject for the HPC centres: they already face these issues with codes in FORTRAN, but as time goes, we noticed that codes in C started to join the legacy code category.

In the group's study, a particular attention was paid to the legacy code category as critical assets for which extra efforts may be foreseen when it comes to migrating to the exascale.

The next steps for the EXDCI-2 legacy codes at the time of the meeting were to conduct a European survey towards the HPC ecosystem and thus establish a clearer view on the European technical debt and so propose adequate means to ensure that those codes evolve smoothly.

Such actions shall take account that the evolution of HPC architectures will most probably lead to a wide adoption of accelerated architectures, that fit to address the power, scaling and memory walls the HPC domain is currently facing.

3.5 Foreseen technologies in a convergent world of distributed models – Trends and influencing factors

While the focus is still put today on the race to exascale, it is necessary to prepare the next steps of the HPC and think beyond for the EU to be able to stand among the HPC leaders of the post-exascale era.

To be able to look beyond and establish a common vision, several questions shall be addressed:

- What is the next big thing afterwards that we need to get prepared for?
- Which are the trends for HPC technology, skills, expertise deployment?
- What are the most important disciplines and their impacts?
    - Pre-identified ones are: AI, Big Data, IOT, Cyber Security, Upstream Technologies, 5G
- How can we in Europe take an active role in the upcoming change process?
    - To the benefit of technology and application providers
    - To grow the expertise and skills level in the relevant domains

Such questions have been assessed in the process of producing the ETP4HPC SRA-4 (fourth version of this recurring Strategic Research Agenda) and have been intensively discussed among the HPC community. SRA-4 identified and consolidated a comprehensive matrix of current research clusters and research domains. Four new clusters and one research domain have been reviewed during the EXDCI-2 second technical workshop.

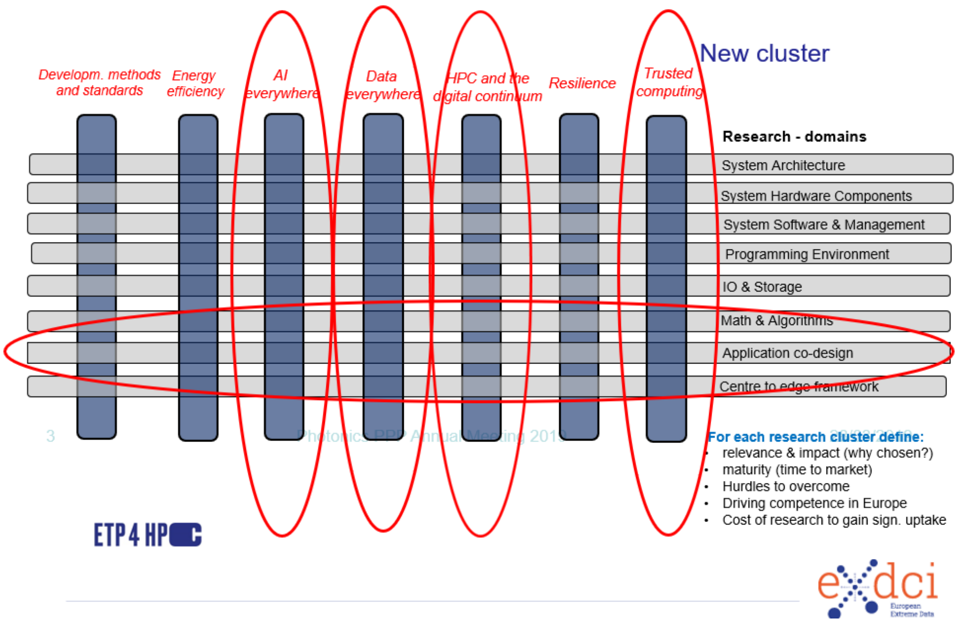

Figure 6 - HPC research matrix – ETP4HPC SRA-4

3.5.1 *HPC and the digital continuum*

The trending factors for this new research cluster come from the convergence of HPC, Big Data and AI:

- Current AI and Auto-ML require HPC capabilities for learning, but this is nourished by data provided "at the edge"
- "Digital twins" or Cyber Physical Systems will need that HPC will be "in the loop" with real world, coping with its (timing and security) constraints, stream processing, security and safety requirements
- Edge data will drive/tune numerical simulations
- HPC in the box or Embedded HPC

The impact of this trend is massive and makes it a hot spot in which to invest for several reasons:

- More and more applications are distributed
- Global energy efficiency requires reducing the amount of data transferred
- Importance of AI in the computing continuum
    - Deep Learning requires more and more Pflops
    - Auto-ML to compute the meta parameters of Deep Learning systems
- Importance of industrial applications (manufacturing, automotive, finance, communication, …) where HPC is more and more "in the loop"
- HPC capabilities more and more required to improve the efficiency of our digital world

The need for HPC is particularly striking when it comes to AI training. It is estimated that "*Since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a* ***3.5 month-doubling time*** *(by comparison, Moore's Law had an 18-month doubling period)*[4]"

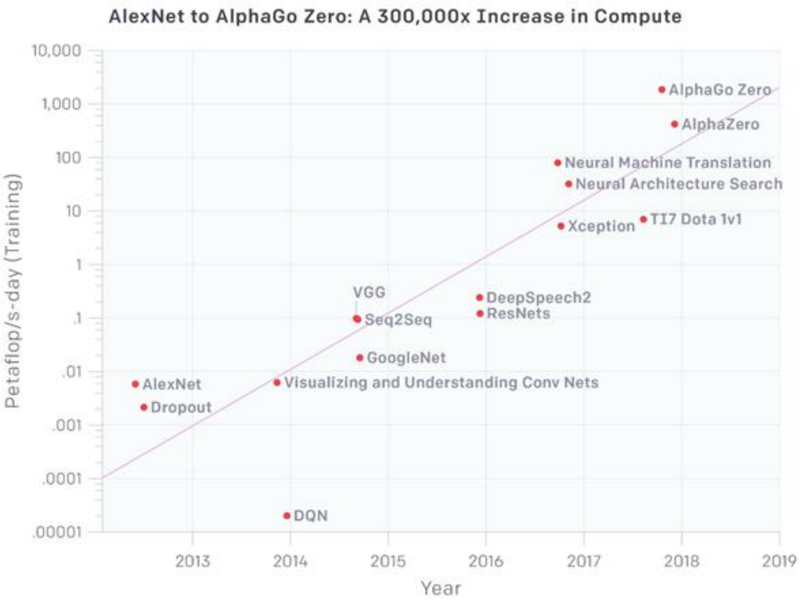

Figure 7 - Trends for AI training computing needs

Hurdles to overcome have been assessed during the SRA-4 preparation which are:

- Not anymore floating point simulations

[4] https://openai.com/blog/ai-and-compute/

- Real-time use, stream processing, less “batches”
- More open to external data, with the challenges of interoperability, security, data protection, interconnectivity, ….
- Different “silos” for HPC, Big Data and IA applications and software stacks
- Interdisciplinarity

Beyond scientific and technical problems to solve, it is foreseen that the access to machines and adequate computing resources is of extreme importance for the years to come. For example, the Jean Zay supercomputer partition dedicated to AI use[5] in France is currently fully booked. Addressing AI needs in the long-term will probably lead to design different machines addressing the needs of each scientific/industrial community. The ownership of such facilities is also of crucial importance for Europe[6].

3.5.2 *Application co-design*

The trends on the applications’ side are also related to the increase role of AI and (big) data, with AI playing a role in establishing new HPC fields beyond its traditional scientific domains as for example humanities, social science, finance…

An increasing uptake from the industry is also observed.

The main challenges identified are:

- Porting, adapting, optimizing for new architectures
    - Requires co-design, stable programming environments
- New opportunities through convergence of HPC, HPDA, and AI
    - More complex workflows, data handling
- Memory bandwidth and communication latency becomes more important than FLOPS
- Long term maintenance of applications and codes[7]
- Continued funding and stable (standardized) environments

Through the Centres of Excellence for Computing Applications (CoEs), discussions with concerned scientific community has shown that – though each of them is aware of such challenges – there was an unequal distribution of their needs and current status among the HPC foreseen developments, which led to create a “heat map”[8].

---

[5] https://www.top500.org/system/179699 - Jean Zay is operated by CEA/CNRS (FR) and is one of PRACE Tier-0 machine, with part of its architecture and operations dedicated to AI.

[6] It is reminded during the discussion that Google has actually more Deep Learning compute power *than any academic*

[7] Among which the question of legacy codes

[8] This heatmap has been developed further in the ETP4HPC SRA-4 released in March 2020 (pp.72 and 73)

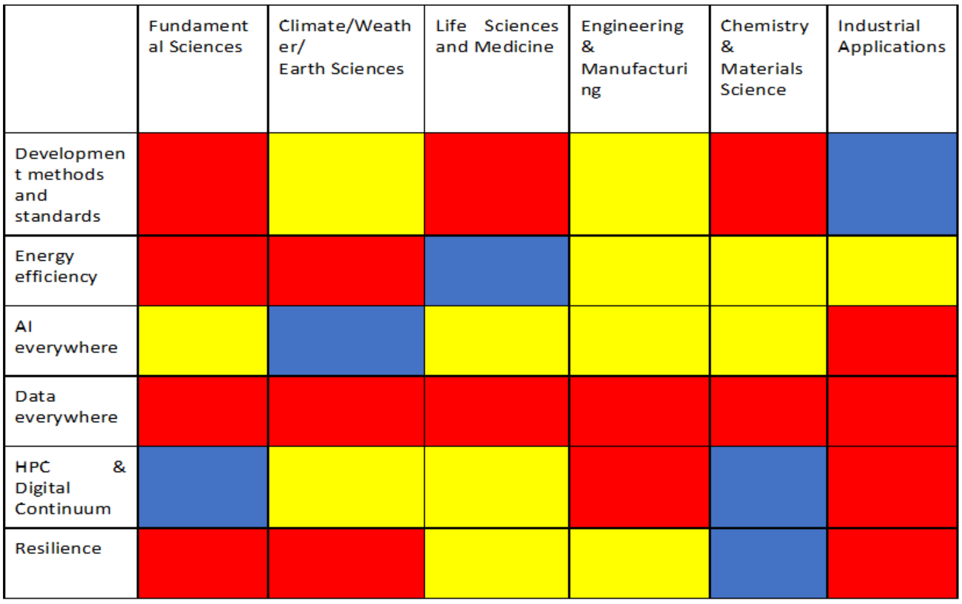

Figure 8 – Relative relevance of the research clusters to application co-design

Relevant — Very relevant — Highly relevant

This heat map helps to identify the progress made by sectors towards their applications and thus to identify the efforts to be done in the coming years.

3.5.3 *Data everywhere*

In this cluster, two main HPC trends have been identified.

The first one is a shift from a model/simulation-driven science to more data-driven science. The "data sphere" could reach 175ZB by 2025[9]. This covers major foreseen innovations and related markets such as driverless cars but also major scientific instruments such as the SKA. It is foreseen that HPC will take an increasingly part in the data-centric era by being one of the components of the digital continuum, either through HPC facility, HPC clouds or embedded HPC.

Impacts of the "data deluge" can have three main branches:

- Scientific simulations and workflows will also generate and work with enormous volumes of data
- Data needed by scientific workflows is now highly distributed geographically
  - Edge + Distributed "Fog" nodes + Clouds + HPC Centres
  - Use of Accumulated historical data and real-time data
- Various data processing techniques needed
  - HPC data processing techniques: in-transit/in situ processing
  - Big Data Processing Techniques (batch + real time/streaming)
  - Trend: support decentralization of (AI-based) analytics towards the edge
  - Cloud-based data access & associated APIs (S3, Swift, etc)

To address this cluster's challenge, several directions are identified which are:

[9] Source: IDC, sponsored by Seagate

- Requirement for high bandwidth between the various infrastructure entities (Edge, Fog, Cloud/HPC system)
    - This is still lacking in spite of in-transit and in-situ processing in HPC systems
    - It is estimated that 5G networks may play a role here
    - However data generation continues to outstrip the evolution of network bandwidths
- Data logistics and data life-cycle management
    - Answer the questions: when to move data, how long to retain data, etc.
    - Collecting provenance metadata becomes important
- Making data infrastructures a lot more resilient/performant at scale
- Data consistency across these infrastructures is still a big problem
- Data federation – for example: lack of unified APIs or need to cope with sensitive data

Several projects and initiatives are being conducted in the EU and worldwide on those subjects of crucial importance for the years to come.

3.5.4 *AI everywhere*

Three trends and influences of AI on HPC are discussed, which are the computing needs of deep learning, convergence of Big Data, HPC and AI and the development of AI-HPC applications not only for academic purposes but also for industrial applications.

The foreseen impacts are huge as AI is acknowledged as one the pillars of the 4th industrial revolution and leads to major changes alongside the HPC value chain (building of new hardware / leading to disruptive methods to write or to refactor software). AI will play an important role in developing actual markets but also to create new ones.

The research cluster "AI everywhere" has summarized the hurdles that shall be overcome in the coming years:

- Scalability of AI systems and algorithms, mainly linked to Big Data and demanding AI algorithms
- Interoperability of tools and software stack
- Ethical aspects[10]: Human-centric design: lawful, ethical, robust
- Liability of AI systems through complex decision-making processes, involving different parties
- Explainable AI built on trust and acceptance of AI applications

3.5.5 *Trustworthy computing*

Security is already and increasingly an important factor of data/computing platforms and services due to the implementation of the digital continuum. System of systems renew security requirements and methods that shall be addressed for the coming years.

Not only security plays a major role, but trust relies also on social acceptance and thus on how data are treated. GDPR compliance and handling of personal data will be a key point for civil society's uptake of digital continuum facilities.

Addressing this challenge requires first to establish security within the mindset of HPC: opening HPC facilities to the digital world lead to rethink security instruments and protocols that should

[10] High-Level Expert Group on Artificial Intelligence, ethics guideline for trustworthy AI, European Commission, 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

be integrated by design in all projects. On a second point: cross-sectorial and cross-disciplinary efforts will be required, and, at last, it is foreseen that crucial choices are to be made to balance security needs vs. energy efficiency vs. performance of the data / computing continuum.

3.5.6 *Open discussion*

The discussion points out that the same main drivers in the SRA-4 research clusters and domains lead to overlaps in the coming trends. Though it is estimated that the European HPC ecosystem is in line for the next generation of machines, it is not prepared yet for the digital continuum.

A specific focus is made on cybersecurity questions:

- An assumption shall be made that any continuum system will have one of its subsystems hacked at some time
- HPC centres are reluctant to open to the continuum – which is a psychological shift to be addressed among the technical challenges
- Also security questions will lead to crucial questions on security's responsibilities as blocks in the continuum will have to rely on each other

To push forward and engage further the discussion, a specific focus on cybersecurity was proposed during the continuum computing topic that EXDCI-2 proposes during the EuroHPC Summit week.

Sustainability is the other focus of the discussion, in particular regarding sustainable data and more importantly on the resources that shall be allocated to handle the increasing amount of them[11]. It is pointed out that a sustainable strategy for data and AI shall be drafted and implemented at the European level. At this scale only will it be possible to compete on a global market dominated by the major IT companies (for the US: Google, Apple, Facebook, Amazon and Microsoft).

3.6 Evolution of the ecosystem

Since the beginning of EXDCI-2, the European HPC ecosystem has moved, pursued its collaboration with other domains and its inner structuration. The main change in the EU HPC landscape is the creation of the EuroHPC JU.

Thus the discussion turned on establishing EXDCI-2 added value and searched to provide a high-level vision on the evolution of the needs of the ecosystem, as well as to assess the role – or future role – of EU HPC main bodies.

EXDCI-2 added value to the EU HPC landscape has been:

- A strong link with the research ecosystem
- Establishing and maintaining a larger and most comprehensive network
  - Inside the EU HPC ecosystem (links with CSAs: FocusCoE, Eurolab4HPC, HC-GiG)[12]
  - Beyond HPC boundaries (e.g link with Big Data, IOT, Cybersecurity…)

[11] Also see 4.3 – Elements for a new vision

[12] While it has been envisaged to strengthen the HPC ecosystem by merging its related CSAs, it is more likely than several coordination actions will continue and thus that the collaboration strategy shall be pursued.

  - Beyond the European area (BDEC)
- Establishing a *de facto* European HPC high-level network and organizing exchanges among several groups
- Taking a role in the creation and organisation of the EuroHPC Summit Week
- Structuring and organising the presence of the European Ecosystem in international events and promoting its results
- Structuring the EU HPC ecosystem for the mid-term and long-term by providing visions and roadmaps

All these elements are important to deliver the objectives of placing Europe among the current leaders in the field and to develop its Single Digital Market.

The discussions also defined the challenges that the EU HPC ecosystem will have to face in the coming years, with a specific focus on anticipating the actions to be taken beyond the exascale race. These challenges are:

- To think and organise the implementation of continuum computing
- To establish solutions for the cybersecurity on system of systems
- To move forward to address the environmental impact

While an important body, the current mission and resources allocated to the EuroHPC JU cannot cover all the challenges above (for example: international cooperation is not at the agenda of the JU), and therefore propositions shall be made on how to complete its work and structure the EU HPC landscape for the mid-term. These propositions shall naturally take into account the potential role of the existing HPC European networks around PRACE and ETP4HPC, but they shall also assess the resources that would be necessary[13].

EXDCI-2 will prepare and disseminate in the coming months those propositions as a high-level position paper.

[13] Particularly taking into account the time gap in European funding that could arise from the shift between two MFFs

4 Report on the 2nd day – EU HPC Ecosystem towards and beyond exascale – Next steps and vision

4.1 Foreseen technologies in a convergent world of distributed models: Applications

4.1.1 *The place of AI in the future, continuum e-infrastructure*

This first session of the second day was dedicated to review three questions, following the work done the previous days:

- What is AI for CI (cyberinfrastructure)?
- What is CI for AI?
- What will this mean for EuroHPC and Horizon Europe?

AI is foreseen to take a disruptive role in the future cyberinfrastructure, yet many questions are still unanswered. A (short) list of such questions were reviewed during the session.

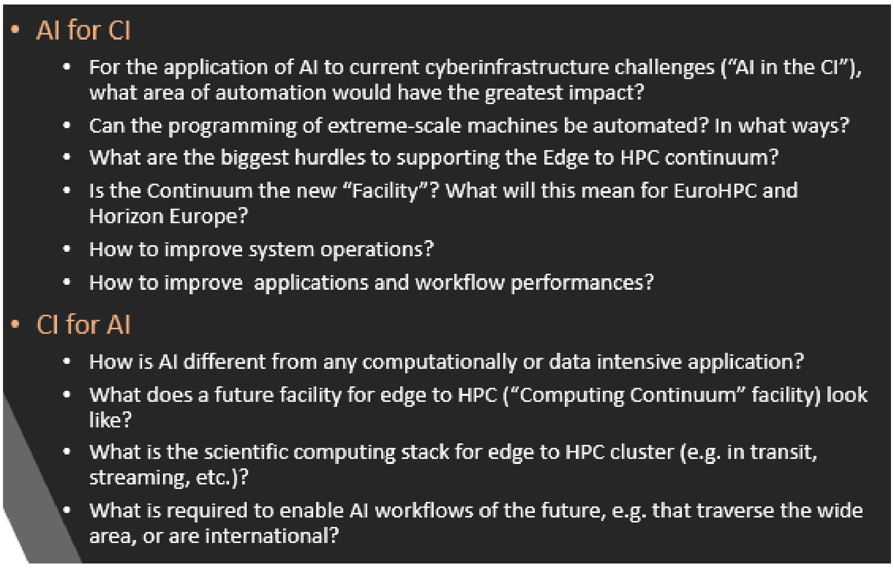

Figure 9 - AI for CI / CI for AI - a shortlist of current questions

Besides those questions, there is an organisational challenge as the HPC community is yet not enough included in the ongoing works carried on by the AI community. Several steps have been taken in this domain, but the interrelation shall rapidly grow.

The work to be conducted by the EU HPC ecosystem in this domain is to first build a statement of HPC needs and expertise, then to share it with the AI community to move forward and increase the inclusion of AI in the HPC strategy.

External competences shall also be sought on ethics (related to data), and extended relation be made with other projects that could be related to the cyberinfrastructure (EOSC).

EXDCI-2 scientific director will work on the above-mentioned statement to prepare the connections with the AI community.

4.1.2 *Foreseen activities on the applications' side*

On the applications' side, EXDCI-2 has an ongoing deliverable[14] on establishing a new roadmap of HPC applications and usages. This roadmap is based on the new convergence context and takes benefit of recently published vision documents (the PRACE Scientific Case for Computing in Europe 2018-2026, ETP4PHC SRA3, SRA4 draft, BDEC2…). The work has been closely conducted with those sources by inviting their experts to contribute to the roadmap.

Moreover, the new roadmap has mobilized the EU HPC ecosystem for new sources of information, for example through consultation of the CoEs and FET HPC projects. Other fields related to AI and Deep Learning have also been updated with this new angle, such as nuclear fusion, turbulent combustion and climate. The case of social sciences has also been taken into account.

One of the main assumptions is that applications have to take into account rapid changes in both hardware and software such as:

- Convergence of (new) architectures
- Hybrid approaches for HPC and Big data (HPDA, IA/MD)
- Coupling large scale instruments
- Urgent computing

The work conducted so far outlined the rise of hybrid methods based on two trends:

- Growth in the needs of compute power for the simulations based on physical laws
- Growth of applications using deep learning methods

Several examples of use have been documented in different scientific fields, including applications from CoEs.

The work conducted also outlined a progress in validation, verification and uncertainties quantification, while the need of training and retraining is growing.

Engagement with the HPC communities, in particular with the CoEs will be pursued with a potential focus on AI, ML and DL for 2020, while HPDA has been the key topic in 2019.

Finally, with the rise of new fields that need HPC capacities, it is estimated that current CoEs cover 70% of the HPC fields of applications. Hence, reaching the lasting 30% such as humanities will be a task to be conducted in the coming years.

4.2 Building on the ecosystem's results

4.2.1 *Scientific Case and industrial collaboration*

In the pathway to exascale, an evaluation of the industrial needs has been conducted in France that led to identify around 20 industrial applications that would use such capacity. It is expected that the same proportions will be found in other main HPC countries (Germany, Switzerland, United Kingdom…)

EXDCI-2 working group has detected new use cases and industrial needs, for example with the use of hyperscalers, linked with HPDA and based on HPC techniques. This results in the introduction of new providers that could be a threat for HPC providers.

[14] D3.1 « Roadmap of applications and usages" submitted to the EC on 26/02/2020

It is also expected in the coming years to have a burst in HPC uptake pushed by a growing demand from SMEs. On this side, specific projects implemented by PRACE (SHAPE) and ongoing work with EU Industry and Open R&D help to disseminate and increase the impact of HPC on the European economy. These efforts shall be extended, in particular in training and support that are key for the SMEs to rapidly benefit from HPC capacities.

As in the entire HPC ecosystem, the shift from compute-centric to data-centric facilities is a game changer for the industrial collaboration and will imply that a lot of new communities engage at some point with HPC centres. The need for support is thus going to be massive.

The EU HPC ecosystem can already have a look on ongoing projects that are working in liaison with the digital continuum such as the CEF projects AQMO and PHIDIAS. New machines such as Jean Zay have been or will be released that integrate the notion of "converged system" (HPC +HPDA + AI).

Also the EXDCI-2 working group considered that EuroHPC JU could become a one-stop-shop for European companies and propose a full range of services in that domain.

Finally, while quantum computing is announced as a game changer and carries on potential disruption for the computing world, investments in this area should be made following an in-depth risk / reward analysis.

4.2.2 *How to best foster the HPC ecosystem and leverage the results of EU projects*

While efforts have been put on funding HPC projects through several EU related-calls since 2014, whose first projects came to an end, it has been analysed that there was a missing part for the exploitation of their results, that could be related to a lack of post-project follow-up.

EXDCI-2 proposed two main initiatives to bridge this gap: conduct and disseminate an analysis of the results of the H2020 technology projects and the creation of a spin-off initiative.

Analysis of the results of the H2020 technology projects

A global and comprehensive analysis has been conducted on all FET HPC projects from the related 2014 call. 171 IP elements (IPs) were listed as results of those projects. These 171 IPs could potentially benefit to any member of the HPC ecosystem.

API	7
application optimisation	7
benchmark suite	6
demonstrator	8
hardware	20
report	2
software	114
training	7
Total général	**171**

Figure 10 - FET HPC 2014 results - Quantitative analysis

Étiquettes de lignes	API	application optimisation	benchmark suite	demonstrator	hardware	report	software	training	Total général
application developer	7						39	4	50
application developer / computing centre				8			4	1	13
application developer / end user							10		10
computing centre		1	4				5		10
end user		6				2	34	2	44
ESD					4		5		9
HPC system customer					1				1
HPC system provider					13		9		22
HPC system provider/ application developer					1		3		4
HPC system provider/ computing centre			2				4		6
HPC system provider/ Processor provider					1		1		2
Total général	**7**	**7**	**6**	**8**	**20**	**2**	**114**	**7**	**171**

Figure 11 - FET HPC 2014 results - potential beneficiaries

The conclusion of this analysis found that the EU created a rich reservoir of results with the potential to lead to develop projects both on a vertical (Extreme Scale Demonstrators – Pilot Systems) and horizontal (FPGA programming environment, Runtime, Energy management) integration.

The analysis has been disseminated through various channels such as a presentation during the EuroHPC Summit Week 2019, EXDCI-2 website and social media.... The first return of experience is that dissemination shall be pursued to maximize the impact of those projects and may include extra narrative communication such as the devise of success stories. A handbook of IPs / results produced can also be considered. A second return of experience showed that the team producing the IPs may need advice on how to industrialise their results – innovation mentors may be useful in such cases.

The FET results' study that EXDCI-2 conducts will also offer an opportunity to assess the results achieved by the European ecosystem to previous roadmaps such as the ETP4HPC SRA-3. Such a gap analysis could give precious insights on future orientations/implementations of the European HPC R&D programs to enhance its impact.

Such an action shall also be pursued towards the ongoing projects – FET HPC 2016 and 2017 as well as the projects selected at the ICT 11 call of 2018. Processor projects (Mont-Blanc 2020 and EPI) shall also be integrated.

Results of the spin-off initiative

The spin-off initiative was launched as a programme meant to facilitate the uptake of promising results before a project ends and therefore accelerate its impacts.

A call for proposal has been launched by an *ad-hoc* EXDCI-2 committee that received 5 applications – all being of good quality. The winner of the call was not known at the time of the meeting[15].

Nevertheless, this pilot sounds promising and could be followed up after the end of EXDCI-2, for example by being pursued by ETP4HPC. It could lead to propose and define a continuity funding to boost the outcomes of a project. This continuity funding may be implemented in the frame of the EuroHPC JU which, as an independent body, could carry on this initiative. Another idea coming from this pilot would be to install a dedicated HPC incubator.

As these propositions were discussed, new ideas emerged:

[15] EXDCI-2 announced the winner (BigDFT4CHEM) in February 2020

- Create a dedicated CoE to industrialize HPC projects' results in the area of software development
- Integrate the results of the pilot scheme into H2020 and Horizon Europe projects at large, for example by defining guidelines for the EC project officers to easily accept amendments dedicated to the further exploitation of a project's results

4.3 Elements for a new vision

Wrapping up elements discussed during the two days of the technical workshop, the following items were discussed as they will frame EXDCI-2 new vision:

- Taking into account the deluge of digital data: 175 ZB in 2025[16] represent a market far beyond HPC
- Natural resource limitations and environmental concerns, which lead to move forward towards sustainable HPC delivery. In particular the "HPC beyond exascale" needs to be built on resilient, reconfigurable, energy-efficient and modular systems.
- Cyber-security as a roadblock for deploying the "continuum". Efforts shall be put on a holistic approach, based on trusted communications – which also concern authentication. Privacy is also a key for social acceptance.
- Monitoring of the "continuum" to understand what is happening with complex workflows - this would especially be used to detect cyber attacks
- Meta-data, particularly regarding interoperability as they will need to be shared all over the continuum
- Available compute resources for AI & HPC, as AI deep learning training runs need a lot of compute capacities; this could lead to a shortage of resources and may become a very strong bottleneck.

One of the core messages to add is that the possible EU added value is not only to provide a one-stop-shop access for HPC facilities, but also to enable aggregation of the continuum. Training and support shall be included in this renewed vision.

A white paper on this renewed vision to pave the way for the future HPC landscape will be drafted by EXDCI-2 in the coming months.

[16] In terms of resources, a rough estimation of the costs of such a volume of data would be close to $3{,}5x10^{12}$€ for the long-term storage, while the HPC market is estimated to reach $45x10^{9}$€

4.4 Conclusion (lessons learned and takeaways)

The move of the HPC landscape towards the digital continuum and the development of a renewed vision are the key takeaways of the EXDCI-2 second technical workshop.

To remain efficient and competitive the European HPC ecosystem must permanently be adapted in a fast-moving environment. Evolution is foreseen both on the technical and organizational sides. For that, the EU HPC stakeholders (research communities and industry) need to rely on long-term visions while supporting a vivid networked community.

EXDCI-2 highlights that significant results have been achieved in the previous years that could pave the way towards these evolutions (creation of the EuroHPC JU, dialog with other EU initiatives, results of the FET HPC projects, successful delivery of European roadmaps, active participation to the BDEC2 global think tank…)

With an updated vision of the evolution of the European HPC landscape, EXDCI-2 also aims to give insights and advices for the preparation of the next generation of European supporting programs such as Horizon Europe.

The workshop experts also pointed out that new international collaborations should be foreseen in addition to existing ones – with the USA, Japan and China – for example with Australia, Canada, India and Saudi Arabia.

5 Annex

5.1 Attendance list

EXDCI2 2d Technical Meeting
2d and 3rd of December 2019, Barcelona

EXDCI2 2d Technical Meeting
List of participants

First Name	Last Name	Organisation	2d December 2019	3rd December 2019
Gabriel	ANTONIU	INRIA	Confcall	
Mark	ASCH	Université Picardie		
Rosa	BADIA	BSC		
Thierry	BIDOT	Neovia Innovation		
Serge	BOGAERTS	PRACE		
Paul	CARPENTER	BSC		
Guillaume	COLIN de VERDIERE	CEA		
Marc	DURANTON	CEA	Confcall	
Stelios	EROTOKRITOU	Cyprus Institute		
Maike	GILLIOT	ETP4HPC		
Renata	GIMENEZ	BSC		
Jens	KRUEGER	Fraunhofer	Confcall	
Erwin	LAURE	KTH	Confcall	
Jean-François	LAVIGNON	TS-JFL		
Corentin	LEFEVRE	Neovia Innovation		
Michael	MALMS	ETP4HPC		
Marjolein	OORSPRONG	PRACE		
Maria	PEREZ	Universidad Politecnica de Madrid	Confcall	
Stephane	REQUENA	GENCI		
François	Bodin	UR1		
Nikoleta	Kiapidou	BSC		

The EXDCI project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement n° 800957

EXDCI2 2d Technical Meeting
2d and 3rd of December 2019, Barcelona

The EXDCI project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement n° 800957