RECOMMENDATIONS OF THE 'EXTREME DATA AND COMPUTING INITIATIVE – 2' (EXDCI-2) PROJECT

Issued by: **EXDCI-2** (The Extreme Data and Computing Initiative - 2 Project)

June 2021

Table of Contents

EXDCI-2 AND THE CHARACTER OF ITS RECOMMENDATIONS

The main objective of the **European Extreme Data & Computing Initiative 2** (**EXDCI-2**) project (March 2017 – December 2020, building on the work done within an analogous previous project called 'EXDCI') was to support the development and implementation of a common European High-Performance Computing (HPC) strategy by coordinating the operation of the European HPC ecosystem in order to achieve Exascale capabilities. The emphasis of the project was on road-mapping and analytical activities, tasks aiming to support Small- and Medium-sized Enterprises (SMEs), facilitating innovation, cross-domain and international collaboration and the organisation of events (e.g. conferences, seminars, etc.)

The **European HPC ecosystem** is understood as a set of stakeholders, resources and expertise in three main areas: infrastructure, technology and applications. EXDCI-2 covered a period which was marked by increasing interacting of the HPC ecosystem with other technologies, thus reflecting the convergence of HPC, Big Data, Cloud, Artificial Intelligence (AI), Internet of Things (IOT), Cybersecurity and Mathematics.

The project's work was divided into **Work Packages** (WP) the results of which were reported in **Deliverables** – documents summarising the work done and the main conclusions. Most of the deliverables contain **recommendations** in relation to the further development of the European HPC ecosystem and the progress of selected areas of the technology or the practice, i.e. proposals for activities that would raise the overall competitiveness of the European HPC ecosystem or its parts.

The objective of **this report** is to summarise these recommendations with a view to implementing adequate follow-up activities in the form of future projects or initiatives. Not all work packages produced recommendations and these are the ones that did:

- **WP 2 - Competitive HPC technologies ecosystem** (including: Preparing the input for Strategic Research Agenda 4 (SRA-4), Liaison with upstream technologies, Synergies with other computing markets, Coordination of HPC technology actions, Technology SMEs development)
- **WP 3 - Excellence in HPC applications and usages** (including: Roadmap of HPC applications and usages, Engagement with HPC users communities and Centres of Excellence (CoEs), Preparation of industrial codes to Exascale)
- **WP 4 - Transversal actions** (including: Transversal vision, Coordination with European actions on Mathematics and Machine Learning, Liaison and best practices analysis from the European pre-Exascale and Exascale system, Analysis and assessment of the European Union (EU) ecosystem, Legacy codes and software modernisation/transition, Transverse coordination and technical workshops organisation, HPC Outreach)
- **WP 5 – International Development** (including: International liaison and workshops, Presence of Europe in HPC-HPDA (High-Performance Data Analytics) standardisation and recommendations to promote European technologies)

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

HOW TO READ THIS REPORT?

The recommendations have been extracted from the respective EXDCI-2 deliverables and other reports, analysed and grouped. As some of the recommendations are limited in scope (i.e. they may refer to a single project or task), this document focuses on those that require a coordinated effort, cross-project or cross-organisational collaboration, or those the impact of which is deemed strategic. Also, the findings included in the project's reports differ in granularity, character and – of course – format. It is a mixed bag: some of the deliverables delineate specific **actions** to be implemented, while others name the **issues** to be resolved, contain **observations** regarding the state-of-the art the current **trends**, or the **priorities** of future work. Understandably, it is the former that have the most value and this document places an emphasis on those recommendations that can become part of an action plan, while trying to synthesise other forms of feedback.

Furthermore, in order to facilitate the readability of this document, it comprises the project's **Key Findings** (as an attempt to provide a concise, high-level statement of the recommendations) and **All EXDCI-2 Recommendations**, regardless of their original form.

Lesser-scope recommendations
In addition to the easily identifiable findings, the work of the project (e.g. meetings, workshops, conferences) also produced some recommendations that apply to isolated issues, domains, technological areas. These are scattered across different deliverables (and not mentioned explicitly in this document), for example:

D3.4 - First report on the organisation of WP3 workshops during HPC Summit Week 2019
D3.5 - Second report on the organisation of WP3 workshops during EuroHPC Summit Week 2020

KEY FINDINGS

The project's recommendations can be classified as **Ecosystem-Level Recommendations** - those whose implementation would call for an ecosystem-wide action, such as through tasks included in an ecosystem coordination project or Technology-Level Recommendations – those which affected selected, isolated technology areas and do not call for broader coordination and which are described in the All EXDCI-2 Recommendations chapter.

The most crucial **Ecosystem-Level Recommendations** are:

- There is clear need for the continuation of the **road-mapping** activities in the area of technology (Strategic Research Agenda, **SRA**) and applications due to the sheer volume of the recommendations produced by either activity and their role in the subsequent work programme definition process.
- The **coordination of the ecosystem** should take place on an on-going basis, including the technological work of the TransContinuum Initiative (**TCI**) as a way to orchestrate the development of the most advanced European digital technologies and collaboration between the projects.
- There is a need to **verify the progress and quality** of the work done within the ecosystem – this applies to the results of the projects and the structure of the funding or road-mapping framework.
- There are some areas in which Europe could attain **leadership** (upcoming technologies – electronics and photonics) or which should be addressed because of Europe's apparent **weakness** in the given field or which need to be addressed because as a **critical factor** enabling future progress (standardisation).

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

- In general, **any future coordination work should stem from a carefully carried-out strategic analysis identifying the main strengths, weaknesses, opportunities and threats present in the ecosystem.**
- The investment needed to accompany the disruption in the HPC domains is enormous and underestimated in many aspects such as **legacy codes** value preservations.
- The use of HPC resources is also fundamentally questioned in order to integrate data stream from large scientific instruments, IoT, etc. Large-scale "digital twins" are still to be implemented and there is a lack of tools and coherency in the HPC and data infrastructures to allow for the **development and deployment of efficient complex workflow application**.

The areas with the most significant Technology-Level recommendations are:

- **HPC technology research and development** – the provision of the European HPC Research and Development priorities
- **HPC applications and usages** - advocating the development of hybrid approaches combining "traditional" HPC and Machine Learning (ML)
- **Transversal cooperation** – addressing the drastic impact of data on the design of applications, adjusting the role of the European Research Council, reflecting the importance of legacy code and software modernisation report, solutions to promote HPC/scientific related careers
- **Coordination of the technology research action in Europe** – adjusting the Technology Readiness Level (TRL) and the scope as well as promoting the user centric character of future projects, maintaining a central project result repository
- **Synergies between electronics, photonics and HPC** – taking advantage of the opportunity to achieve global leadership in selected technologies
- **Standardisation** – addressing the lack of European presence in the global standardisation effort

A final recommendation concerns **international-level actions**.

- The presence of Europe in **international strategic actions**, regarding the future of HPC and Big Data **convergence**, is essential and must continue to be supported. This is particularly important in the context of a future, federated data ecosystem that will enable international collaboration on major societal challenges (pandemics, natural disasters, food and water, etc.) and HPC-data intensive applications (SKA, LHC, Copernicus, LIGO, et al.).
- Support European presence in **international standards organisations** on software and hardware for HPC and HPDA.

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

ALL EXDCI-2 RECOMMENDATIONS

The following is a list of **all recommendations** developed by the project as they appear in the respective deliverables. Please bear in mind that some of the work packages and some deliverables do not contain any – if this is the case, the given work packages or deliverable does not feature in this list. For a full list of work packages and deliverables, please refer to Annex 1: All EXDCI-2 Deliverables.

WP 2 - Competitive HPC technologies ecosystem

D2.1 - HPC-HPDA technology roadmap

The recommendations of this deliverable are included in the **Strategic Research Agenda (SRA)** document published in March 2020. Due to the size of the document, it is not possible to quote all its findings. The SRA is the work of the technology stakeholders of the European HPC ecosystem and it delineates the **priorities** for European HPC technology research across the following domains:

- System Architecture
- System Hardware Components
- System Software and Management
- Programming Environment
- I/O & Storage
- Mathematics & Algorithms
- Application Co-design
- Centre-to-edge-framework

The SRA also contains suggestions regarding organising and managing the work programme and related calls. There are use cases illustrating the role of HPC in future complex workflows.

D2.2 - Report on trends and potential synergies between electronics, photonics and HPC

The main conclusion is that research projects involving upstream technology providers and HPC teams could deliver potential new solutions for HPC systems. The following **actions** are put forward.

- Establish a continuous dialogue between photonics, electronics and HPC communities under the supervision of Photonics21, AENEAS and ETP4HPC.
- Undertake small actions to specify research objectives, benchmarks and test data sets at the interface of two research communities.
- Work on European specifications for the integration of heterogeneous chips.
- Launch a research programme to develop new ideas coming from upstream technologies to provide new solutions for upcoming HPC systems.

D2.3 - Report on Big Data, embedded and edge computing and HPC synergies

This report recommends **the continuation of the work of the TransContinuum Initiative** in order to facilitate the collaboration between European associations and projects interested in strengthening the European digital infrastructure that is essential for the implementation of many European priorities such as the 'Green Deal', the Horizon Europe Missions and the Destination Earth project. It also stipulates that work continue to **analyse**

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

industrial and scientific use cases reflecting the Digital Twin concept. The outcome of this work should be used to identify the building blocks of the next recommendations for the European R&D priorities.

D2.4 - Report on coordination of the technology research action in Europe

This report **identifies actions in the following thematic blocks**: The character of future projects, the quality of the Strategic Research Agenda and Intra-project collaboration:

The character of future projects:

- Integration projects with a high TRL level. These projects should be structured around the potential HPC system providers and help to push new technologies toward the market.
- Horizontal projects with high TRL level and with an objective to produce software of pre-production quality level. The projects should aggregate different technology pieces to provide software with sufficient coverage and usability level.
- Actions targeting the potential "users" of the technologies. There are at least three kinds of "users" to be considered: application communities, application developer communities and computing centre operational teams.

The quality of the Strategic Research Agenda:

- The gaps in the coverage provided by the SRA: Integration of security in HPC systems, Support of virtualisation to open HPC system usage, Emergence of performance metrics for the new applications
- Most of the milestones are too ambitious to be achieved within a single FET-HPC project.
- The milestones should be stated independently from any potential solution
- They should be considered mandatory steps toward efficient Exascale systems (and now post-Exascale systems)
- They should highlight new trends where more innovative research is needed

Intra-project collaboration:

- Maintain a global survey and data base of the results of the HPC technology projects to build a global vision which can be used to update the HPC strategy; EuroHPC team could do this analysis or delegate it to the relevant organisation(s)/project in the HPC ecosystem;
- Set up calls with higher TRL objectives for the technologies to enter the virtuous circle of continuous improvements by their user bases; These calls could be:
    - Integration projects with the objective to deliver a complete HPC solution;
    - Horizontal projects with the objective to develop a layer that could be used by several European computing centres or application developer communities
- Develop a programmatic approach of the research programme with a strong focus on the strategic axes as the EPI;
- Implement a framework to facilitate strong cooperation between the projects selected.

D2.5 - Feedback on new experimentation to unlock R&I project (unavailable on the EXDCI-2 website)

The following **high-level actions** and issues were named:

- Activities facilitating the collaboration between the public and the private sectors need to be engaged during the lifetime of the R&I project. There is a real need for support and funds, accessible fast without a high overhead in administrative matters.
- The current work programme has difficulties in coping with adaption and changes during the project lifetime. Very probably, the regulations of the upcoming framework Horizon Europe will not differ substantially to this regard.
- Apply cascade-funding mechanisms.

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

WP 3 - Excellence in HPC applications and usages

D3.1 - Roadmap of HPC applications and usages

This deliverable advocates the development of **hybrid approaches** combining "traditional" HPC and Machine Learning (ML) to address the following trends:

- Increasing importance of hybrid-modelling approaches, either by using ML techniques to solve more efficiently parts of HPC models (parameterisation of subgrid scale phenomena, solvers and preconditioners, ...), or by developing model-based (physics-based) ML;
- There is a need for more-and-more resource infrastructures allowing at the same time efficient numerical simulation of physical phenomena and treatment of massive data, calling in turn for resources where different types of processors are associated, e.g. CPUs for HPC and GPUs or other types of accelerators for converged HPC/ML workloads;
- Hybrid training should be supported so that application developer teams can address all aspects of these new methods.
- The evolution of resource infrastructures must also be combined with the fact that HPC facilities, either as concentrated centres or of cloud-types, are more and more integrated inside a global cyber-infrastructure, from places where the data are being produced to the place where they are used, stored and archived.
- Such improved simulation methods will require more detailed validation, calling in term for sophisticated post- processing in relation with massive validation data.
- Co-design process between hardware and software developers and application developers appears to be less developed within Europe as compared with the USA and, even more, with Japan. Addressing and supporting co-design issues in this way would largely facilitate efficient use of Exascale converged facilities for a number of applications.

D3.2 - First Report on joint brainstorming sessions among scientific and industrial users communities

This report provides the following two directions for future actions:

- Due to the rapid convergence between HPC, HPDA and AI as a result of the explosion of data generated by large scale instruments or numerical simulations, PRACE and the European computational ecosystem must develop new architectures and services addressing mixed HPC/AI workloads.
- EU-wide efforts should be made in education and training to develop new skills in data science and numerical simulation and train and retain a new generation of researchers and technologists in both science and industry.

D3.3 - Second Report on joint brainstorming sessions among scientific and industrial users communities

This report includes two general **observations**:

- 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 organisational sides. For that, the EU HPC stakeholders (research communities and industry) need to rely on long-term visions while supporting a vivid networked community.
- 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.

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

WP 4 - Transversal actions

D4.1 - Transversal vision report

This reports states that the European HPC ecosystem should address the following **trends**:

- A drastic impact of data in the design of applications - The introduction of big data in the discovery process is driving application to be implemented as large-scale distributed workflows that need to be deployed on a large set of systems, each one having its own idiosyncrasies (e.g. quantum accelerator) that questions the way application are implemented.
- An opening of HPC to other users (and thus other usages)
- The emergence of potential highly disruptive hardware; coping with hardware evolution, service-oriented architectures to address usage evolution
- The necessity to federate infrastructure and re-organise the entire ecosystem.

D4.2 - Mathematics and machine learning report

The **actions** suggested in this reports are centred on the role of the **European Research Council (ERC)**:

- Broaden the ERCs remit, to enable larger number of small grants to younger mathematicians and the appearance of mathematics in other panels.
- Launch dedicated research funding programs in Mathematics for AI, Mathematics for Industry and Mathematics for HPC.
- Set up a generous co-funding, or matched-funding program based on Public- Private-Partnerships, where the EC creates the conditions for bottom-up research projects, building upon teams that exist already in industry, universities and research institutes.

D4.5 - Assessment for legacy code and software modernisation report

The following **issues** were named:

- The European code base represents a significant amount of money that cannot or should not be wasted.
- The team sizes are too often too small to envision a full rewrite in a timely manner.
- The codes are critical to their owners and, therefore, they cannot be discarded through obsolescence due to a change of supercomputers.
- The majority of the codes rely on the availability of FORTRAN (FORTRAN90). This raises the concern with respect to teaching, as FORTRAN is no more studied during the academic syllabus.
- Funding needs to be dedicated to code porting by teams in addition to the funds given to the Centres of Excellence (CoE).
- The proposed viscosity formula is a crude yet effective way to get a feel of the investments needed in the future to preserve the European portfolio of codes.

D4.8 - HPC Outreach report

This report contains the following **recommendations**:

- To promote HPC/scientific related careers, **STEM** (science, technology, engineering and mathematics) vocations have to be promoted starting in primary school. All exchanges during the workshop do show that this is possible and well-suited assuming that the teachers are supported by scientists.
- Students from the age of 16 begin should have a minimum knowledge of writing programs for supercomputers.

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

- Teaching programming at school works much better if they can apply it at what they are studying otherwise.
- Python is popular. Pedagogically interpreted languages avoid dead time that helps the students to focus. It is recommended that from the age of 11 pupils begin to learn Python and that every year gradually increase the level of complexity with the aim of reaching a level of data management or automatic learning and supercomputing before the university level.
- It is necessary to emphasise about the benefits of mathematics and consequently the computer skills personally and in all careers. Children must be able to explain parents.
- We must implement the knowledge of programming in all the STEM subjects and evaluate them in the examinations from the age of ten.
- Nano computer based platform (e.g. Raspberry PI) can provide a very rich support for children to experiment computer science and in particular in parallel programming.
- Integrating data, AI and IoT related knowledge could be a great support in primary school to develop environment related pupils projects. Collaborations between universities and primary school to set up such initiative.
- The sequence STEM at primary school, programming in high school, HPC at the university must be carefully staged and coordinated.

WP 5 – International Development

D5.1 - Report on the first international workshop

The **actions** suggested in this report are:

- Draft a design for a distributed services platform for science to serve as shared software
- Need for an infrastructure for the growing continuum of computing devices and data sources on which future science will rely.
- Organise and develop an international demonstration of the feasibility and potential of the Digital Continuum Platform (DCP) (prototype implementation)
- Develop a corresponding “shaping strategy” addressing all relevant stakeholders and moving the community toward convergence on a standard DCP specification.

D5.2 - Report on the second international workshop

The **actions** suggested in this report are:

- Design a data ecosystem, based on international standards, for the security and transfer of data from devices to computing machines. This ecosystem should emphasise the concepts of interoperability and portability, to cross geographical frontiers, administrative domains and scientific fields. An essential part of this ecosystem is the need to conceive and develop transcontinuum workflow-enabling software stacks.
- Investigate how machine learning can play a role at all stages of the digital continuum, all the way from the edge to the centre.
- Adapt HPC centres to this new, data-centric science and the machine learning algorithms on which it relies.
- Build, together, a comprehensive shared, sustainable e-science e-infrastructure to address the major societal challenges in the UN’s Sustainable Development Goals in their Agenda 2030. Not forgetting fundamental science challenges.

D5.3 - Presence of Europe in HPC-HPDA standardisation and recommendations to promote European technologies

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

This work on standards led to two sets of recommendations. The first set targets the optimisation of the current efforts without a significant increase of resource:

- Achieve a continuous survey of emerging standards, produce maps of new standard organisations relevant in the context of the digital continuum, maintain all maps up to date and disseminate this information.
- Establish a network of interested people around each European standardisation expert to share the information about the standard activity.

The second group of recommendations aims at achieving European leadership in this domain:

- Implement a specific HPC organisation with sufficient financial and human resources, whose goal will be to develop European HPC presence in standard organisations.
- Put in place a process to help consortia organise their presence in a standard organisation or write emerging standard specifications.
- Establish a process for participants to apply to European funded projects for a specific funded effort to facilitate the emergence of a standard.
- Identify and launch standardisation projects in strategic domains, to establish a level playing field, to favour European player's new initiatives and to encourage green IT.

CONCLUSIONS

The EXDCI-2 project was deployed in a highly evolving landscape, marked by the establishment of the EuroHPC Joint Undertaking as a permanent mechanism for a sustained development of all pillars of European HPC. The main paradigm change over the duration of the project was the way the production and use of data were reshaping the scientific activities and the applications relying on HPC technologies. As a consequence, the project put forward the TransContinuum Initiative in order to analyse the nature of the complex workflow supporting new scientific applications. EXDCI-2 also contributed to applications and technology roadmaps and identified numerous recommendations in the area the software, hardware technologies and infrastructure. The recommendations stemming from the work of project confirm that the European HPC ecosystem requires continuous coordination (the main task performed by EDXCI-2 and its predecessor EXDCI), road-mapping activities and addressing selected areas which affect its competitiveness.

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

ANNEX 1: ALL EXDCI-2 DELIVERABLES

All public deliverables are listed at this address: https://exdci.eu/resources/public-deliverables

	WP 2 - Competitive HPC technologies ecosystem
D2.1	HPC-HPDA technology roadmap
D2.2	Report on trends and potential synergies between electronics, photonics and HPC
D2.3	Report on Big Data, embedded and edge computing and HPC synergies
D2.4	Report on coordination of the technology research action in Europe
D2.5	Feedback on new experimentation to unlock R&I project
	WP 3 - Excellence in HPC applications and usages
D3.1	Roadmap of HPC applications and usages
D3.2	First Report on joint brainstorming sessions among scientific and industrial users communities
D3.3	Second Report on joint brainstorming sessions among scientific and industrial users communities
D3.4	First report on the organisation of WP3 workshops during HPC Summit Week 2019
D3.5	Second report on the organisation of WP3 workshops during HPC Summit Week 2020
	WP 4 - Transversal actions
D4.1	Transversal vision report
D4.2	Mathematics and machine learning report

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.

D4.4	Assessment on the ecosystem report
D4.5	Assessment for legacy code and software modernisation report
D4.6	First report on EXDCI-2 technical workshop
D4.7	Second report on EXDCI-2 technical workshop
D4.8	HPC Outreach report
	WP 5 – International Development
D5.1	Report on the first international workshop
D5.2	Report on the second international workshop
D5.3	Presence of Europe in HPC-HPDA standardisation and recommendations to promote European technologies
	WP 6 – Dissemination
D6.1	First report on Dissemination Activities
D6.2	Second report on DisseminationActivities

The EXDCI-2 project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957.