EXDCI
European Extreme Data & Computing Initiative

ACM Europe Conference: EXDCI Final Event 2017
https://exdci.eu

Coordination: Sergi Girona (sergi.girona@bsc.es)
WP Leaders: M. Asch, *F. Bodin*, R. Gimenez,
C. Inglis, M. Malms, JP Nominé, S. Requena

The EXDCI project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 671558

Introduction

- Important investment from the European commission in HPC
  - ~700 M€ euros to be invested between 2016-2020
  - Call for projects driven by roadmapping efforts and community feedbacks

- EXDCI *support action* is to help in building a strong European HPC *ecosystem*
  - Build relations and collect feedback from the ecosystem
  - Provide ecosystem recommendations
  - Promote exchanges between CoEs, FETs, ecosystem at large

EXDCI Contributions (not exhaustive)

- Roadmapping
  - Strategic Research Agenda
  - Extreme scale Demonstrator
- Recommendations
  - Technical-level, application-level and ecosystem wide
  - For SMEs and startups
- Toward the ecosystem
  - Events (EHPCSW), dynamic interactions
  - Analysis (KPI)
- International
  - Liaison with other initiatives: BDEC, BDVA, Eurolab4HPC, …

Covered already in previous EXDCI talks

EXDCI, a CSA in a Rapidly Evolving Context

- Since EXDCI beginning a lot has happened
- As planned, new FETHPC projects and CoEs
- New international initiatives
  - IPCEI
  - European Open Science Cloud
  - EuroHPC
- Exascale still a moving target
  - Inputs from the ecosystem (e.g. CoEs)
  - Better understanding of *the Exascale transition*
  - Pathways to convergence data and compute
  - Extension of the use of supercomputers (e.g. urgent computing)
- Increased international competition

We rarely had such exciting time!

Overview of the Presentation

- EXDCI Technical Context
- Data at the Core of the Exascale Transition
- The European HPC Ecosystem (from EXDCI)
- A Vision of the Future of Supercomputing
- Conclusions & Perspectives

EXDCI Technical Context

Petascale to Exascale

- Petascale to Exascale transition is raising many issues
  - Not only related to technology
  - Not happening in isolation
  - In a context of scientific (observational) data deluge
- Well summarized in the USA National Strategic Computing Initiative (NSCI)
  - “NSCI seeks to drive the convergence of compute-intensive and data-intensive systems”
- We are potentially on a paradigm change denoted Exascale but meaning computing *generation transition*

Petascale to Exascale cont.

- Peta-Exa transition is not similar to Tera-Peta transition
  - This is a disruption mainly due to parallel model issues (compute and IO)
  - And the need to deal with *large amounts* of data from multiple sources (scientific instruments, simulations)
- Some questions are
  - Can one platform fit all?
  - Is the "*Cloud*" a relevant solution?
  - How to move data around (or not)?
  - How to integrate Big Data technologies?
  - How to manage resources?

Why Peta-Exa is Not Similar to Tera-Peta Transition?

- The main Tera-Peta transition was performed before during the Giga-Tera transition
  - Adaption of codes to distributed memory machines
  - Tera to Peta was smooth and with minimum (side-) effects for most HPC users
- Data issue is changing the game for Peta-Exa
  - New software stack and algorithms
  - Questions the discovery process (e.g The Fourth Paradigm)
  - Data analytics and machine learning
  - Data localization

What Exascale is Not

- Exascale == 10^{18} flops of interest for a small community
  - Such as LQCD and field based on embarrassingly parallel methods (e.g. Monte-Carlo)
- Exascale transition for most people is not about the next increment in machine features
  - The next generation of machines is likely to create a practice and organization disruption
  - It is easy to compute anywhere (c.f. PRACE) but moving data around is (very) slow, if feasible
  - Adherence to a system (including storage and networking) is likely to increase

Exascale-Wise Applications Characterization*

1. Workload
2. Workflow
3. Code
4. Scalability
5. Operating System
6. I/O
7. HPC Community
8. Hardware
9. Visualization
10. Interactivity
11. Data management and analysis
12. Impact on Science/Society

*Computer science point of view

Exascale Transition Impact on Codes

	Tera-to-Peta	Peta-to-Exa
Off-the-Shelf	ISV in charge	Not addressed (Market?)
In-House	Update of the codes, but no significant effort compare to Giga-to-Tera	Too many technologies for an in-house team. Will need to add software engineers, etc.
Languages	Fortran, C/C++, OpenMP, Cuda, OpenACC, OpenCL	Fortran, C/C++, OpenMP, OpenACC, OpenCL, DSL, interpreted languages, task support, …
Runtime	Accelerator support	Runtime must handle more resources
Sustainability	No significant changes	More specialized codes (e.g. DSL) , code architecture rendered obsolete
Complexity	Code refactoring to make it accelerator friendly	Heterogeneous software stack to address the data and task management
Portability	Not simple, but achievable with careful design	Very dependent on workflow management and application architectures
Performance	Retuning (expensive in some cases) possible	Tuning is going to be Hell (too many dimensions), energy tuning?*

*1w/year ≈ 1€, 10% of 20MW/Y → 2MW/Y → 2M€/Y → ~20 persons/Y

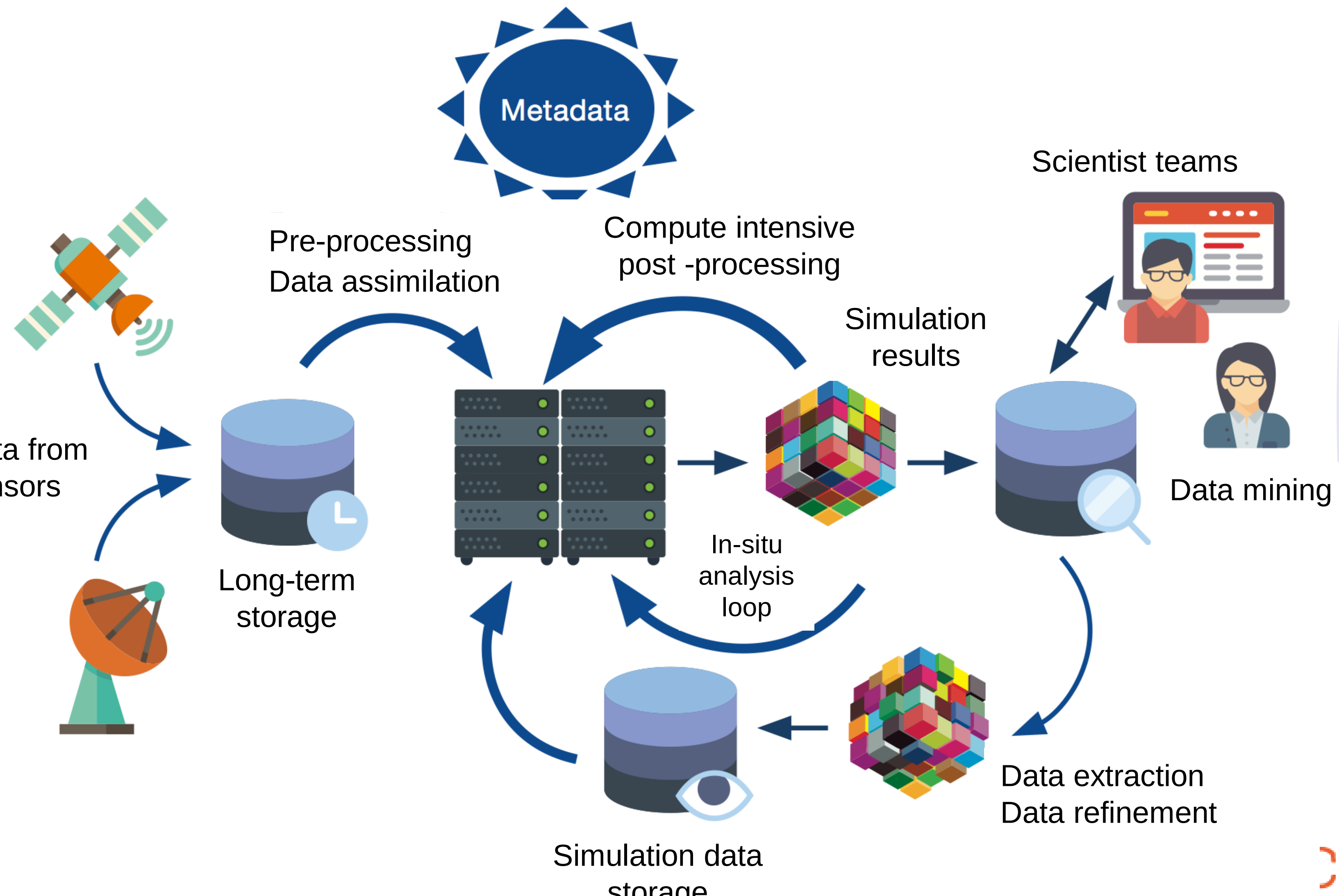
Exascale Transition Impact Example with Workflows

	Tera-to-Peta	Peta-to-Exa
Complexity	Code coupling	More multiphysics, multiphase models, data assimilations, data analytics, edge computing, …
Heterogeneity	Mostly homogeneous	Mix of data analytics and simulation, heterogeneous bricks
Localization	All in one system	Data may come from large scientific instruments, or a large number of small instruments
I/O constrained	Solvable issue	Cannot move the data around, not sure it can be solved
Allocation	Batch mostly	Batch, interactive (guided simulation and analysis), (soft) real-time* (visualization, …)

*Big data assimilation for Extreme-scale NWP, Takemasa Miyoshi

Data at the Core of the Exascale Transition

Data Life Cycles



From « *Les big data à découvert* », CNRS Éditions, 2017

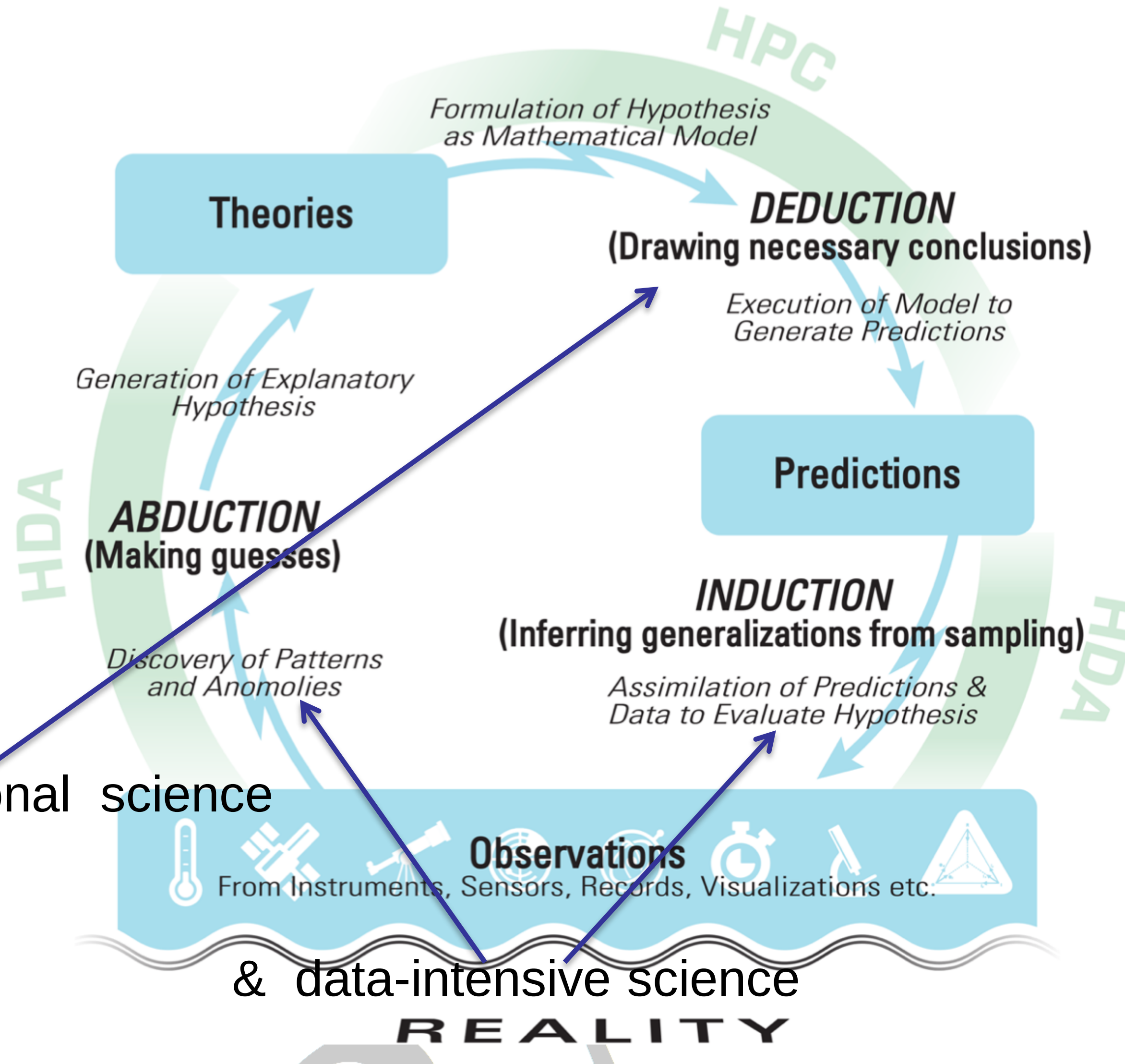
Big Data, Why Bother?

- Big data happens when current practices to handle data become inoperative
  - Large volume does not define it
- Sensors, observations
  - Producing a deluge of data
  - Cheaper and faster sensors (e.g. genomics, IoT, motes -smart dust and wireless sensing networks-)
- Strong and active economy sector driving many technology evolutions
  - Cannot afford to redevelop HPC specific data analytics (and AI) software stack

A Converged Scientific Process

From the BDEC “Pathways to Convergence” Report
BDEC Committee

This model as a plausible discovery hypothesis for physical laws



Combines computational science

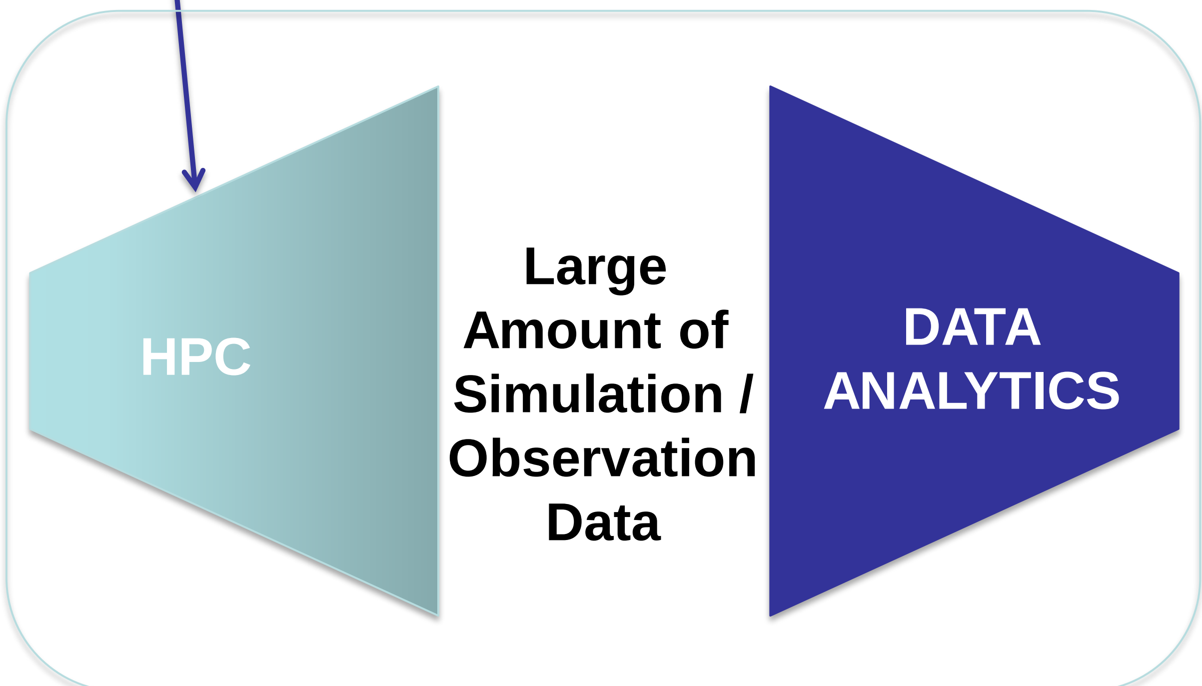
& data-intensive science

Convergence of HPC and Data Analytics

This part of the process can only be partially automated (humans in the loop)

Large Experimental data

Small Amount of Model Description Data



Small Amount of Extracted Data

Where to Compute? Where to Store? Where to analyze?

- Unclear, related to the workflow organization
  - Strongly constrained by I/O
  - Dictated by data location and processing
- New compute in storage approach
  - e.g. Percipient storage (SAGE Project), ability for I/O to accept computation
- Computing inside the scientific instruments
  - Edge/fog computing

(too) many complex choices?

Exascale Transition is an Holistic Issue

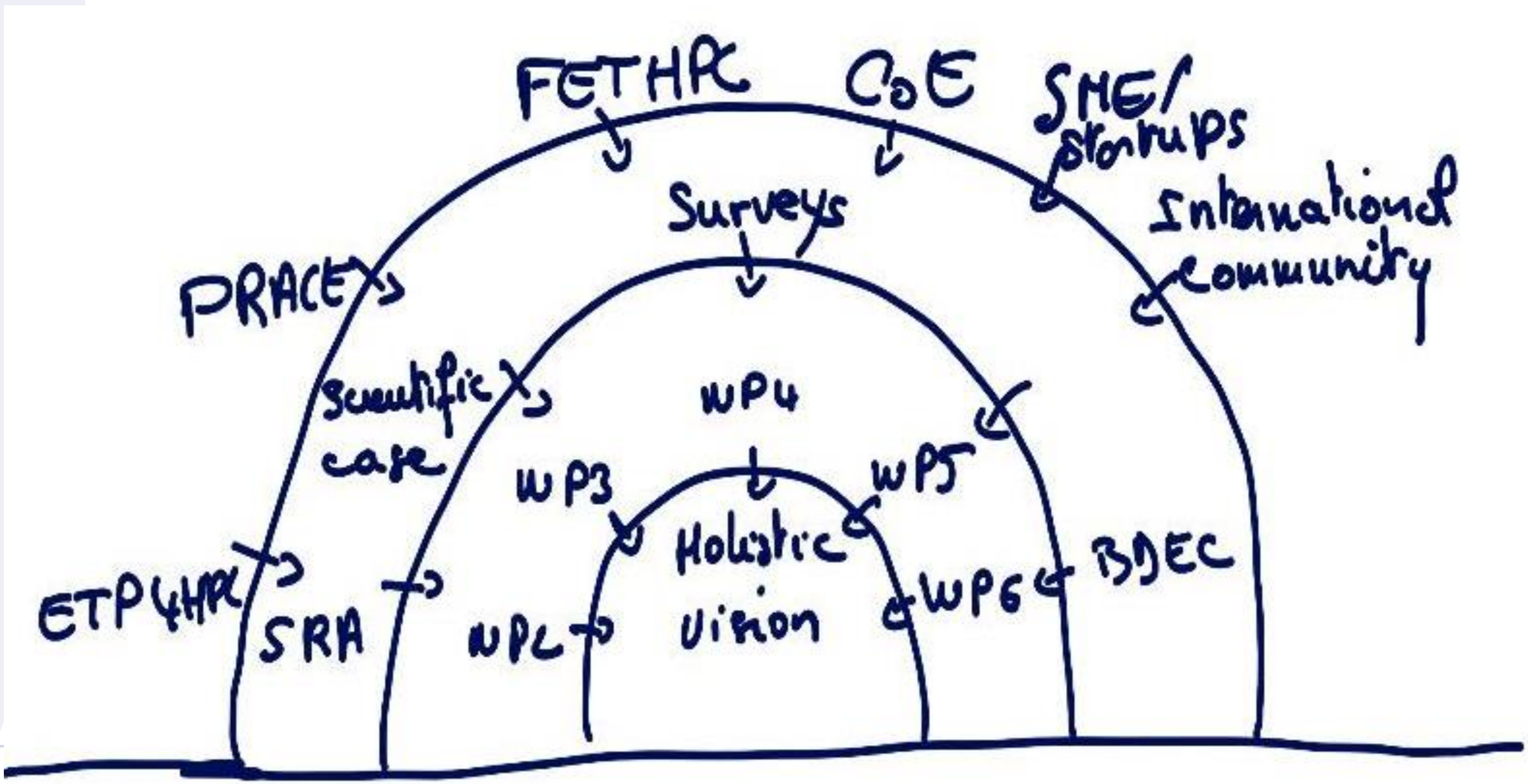
- Four elements
  - Applications
  - Software stack
  - Hardware technology (mostly vendors community)
  - HPC centers

- Supposedly being specified in a co-design process
  - An Ecosystem effort scattered on multiple organisms, projects (e.g. CoE)
  - Extreme scale Demonstrator (EsD)

The European HPC Ecosystem (from EXDCI)

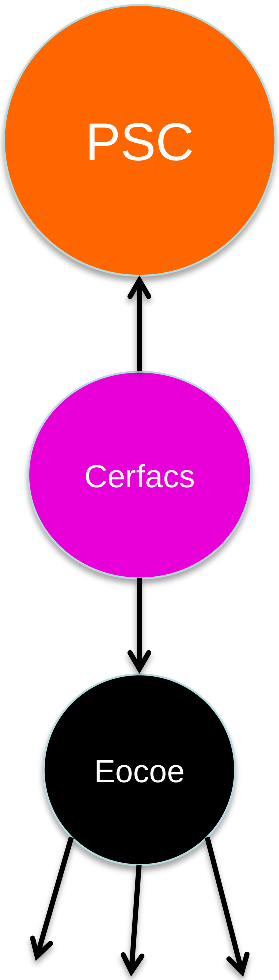
EXDCI View of the Ecosystem

- How does the Ecosystem contributes to the recommendations?
- What is the coverage of EXDCI actions?
- What type of stakeholders do we have?



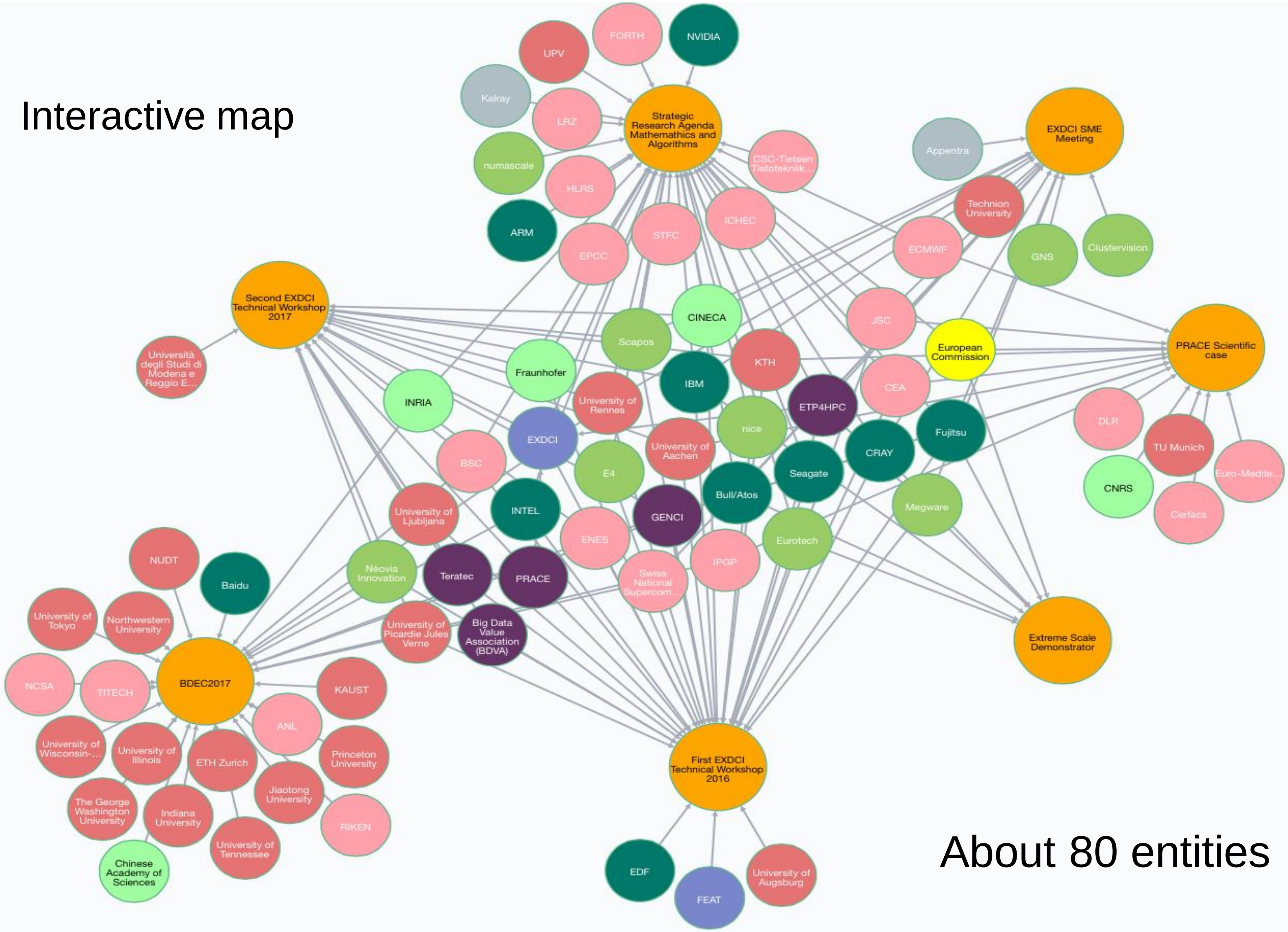
An Ecosystem Cartography Attempt (D4.7)

- Showing EXDCI impact on the Ecosystem
  - How do we connect stakeholders
- Graph representation
  - Two types of nodes
    - EXDCI Events that generate a common production
      - SME workshop
      - SRA, PSC, EsD
      - BDEC 2017
      - EXDCI Technical Workshops
    - Stakeholders, CoE, (FET HPC soon)
  - Edges represent the "participate to" relationships
- Interactive to highlight different points of view
  - Participation at the periphery
  - Central stakeholders
  - …



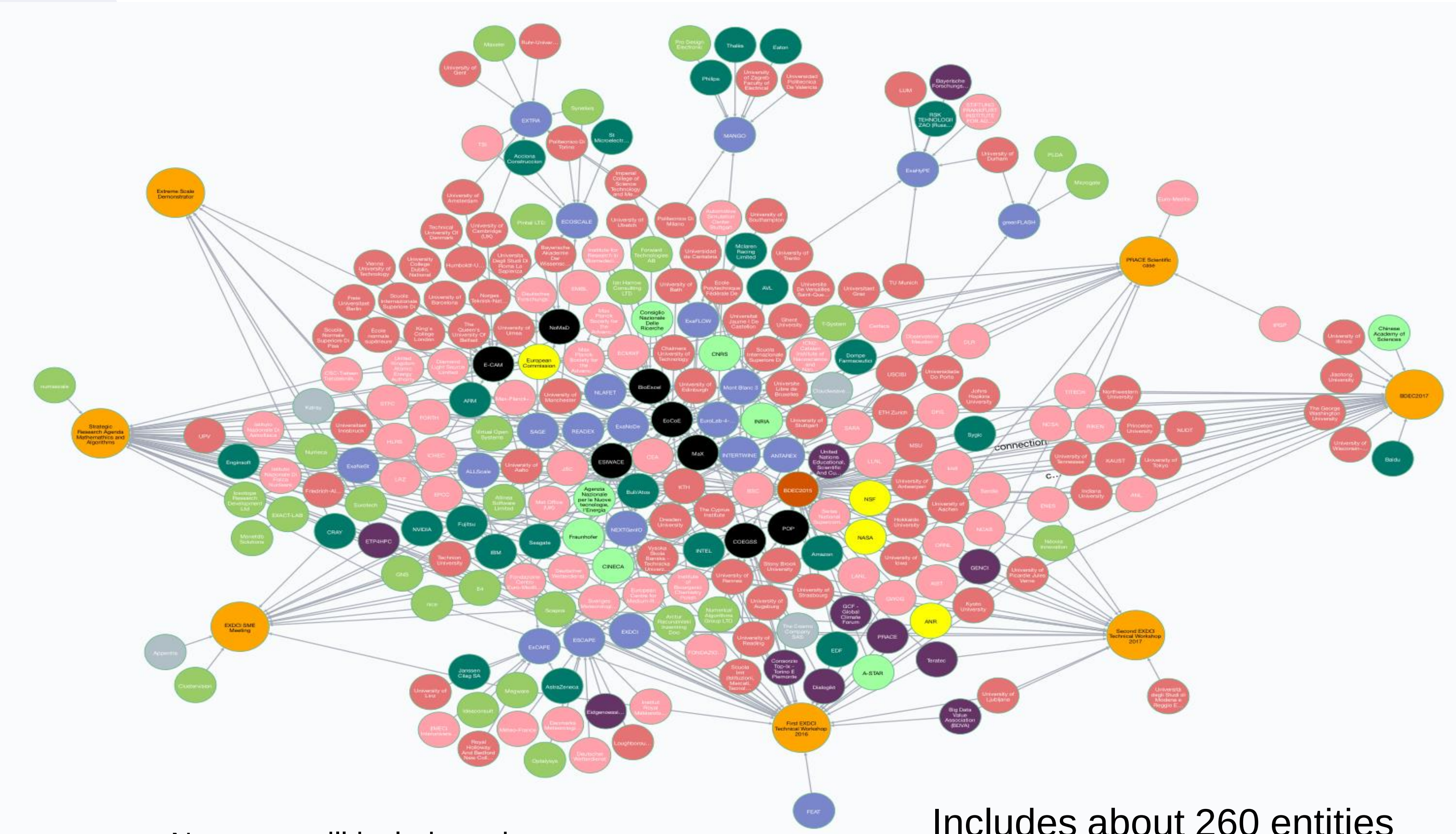
Entities Active in EXDCI Related Activities

Interactive map



About 80 entities

A Map of the Ecosystem as Connected by EXDCI



Includes about 260 entities

Next step will include topics.

A Vision of the Future of Supercomputing

This is where you throw rocks

A personal view

Combining Edges, Data Centers, Supercomputers-1

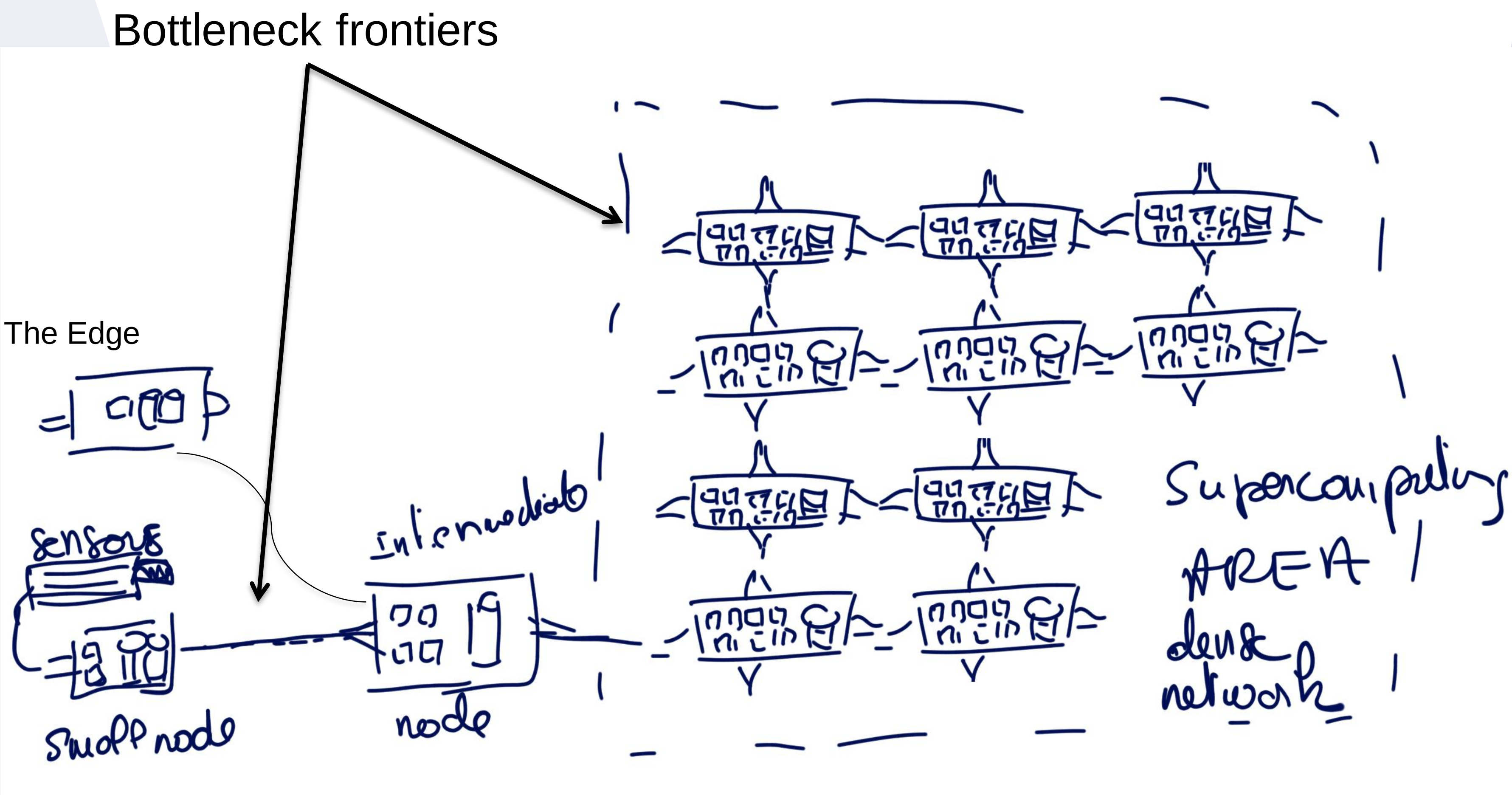
- In BDEC we have been struggling in combining data and compute
  - See the excellent "pathways to convergence" document and its evolution
- Current/next scientific challenges don't fit current infrastructure
  - HPC centers, Clouds, Edge, Data, Compute, …
  - We have frontiers where we need a continuum
  - Data location is the dimensioning parameter
    - Large data volume (from large scientific instruments, from simulations) cannot move efficiently (and to go where?)
  - Archiving data is expensive **1PetaByte ~ $60k a year**
    - Data analysis/reduction needed everywhere → compute

Combining Edges, Data Centers, Supercomputers-2

- Very difficult to extend current infrastructure concepts to fit the future needs
  - If data cannot be moved around, then it is a design game changer
  - Complex distributed workflow is the future rule
- We need an homogeneous global infrastructure based on connected nodes
  - All nodes provide compute, storage and communication capabilities
  - Nodes are qualitatively equivalent
  - Nodes are not quantitatively equal
- Internet topology oblivious routing scheme not adequate

Inspired from the paper "Interoperable Convergence of Storage, Networking and Computation" Micah Beck, Terry Moore, and Piotr Luszczek University of Tennessee, Knoxville, Tennessee, August 12, 2017

Combining Edges, Data Centers, Supercomputers-3



Combining Edges, Data Centers, Supercomputers-4

- Bandwidth and latency between nodes can vary
- A supercomputer node is not different from an edge node (capabilities only differs)
  - Uniformity of nodes helps to create complex workflow that includes all sorts of tasks (i.e. sensing, simulation, analysis, visualization)
- Function of a node is by destination not by nature (they are all the same anyway), of course adequacy of the capabilities of a node helps
- A supercomputer is defined by a set of nodes that are strongly connected by high bandwidth, low latency links
- There are no frontier between the core supercomputers and its edges or intermediate nodes or the storage nodes
  - Complex workflow can be distributed in a uniform way
- Storage is just high capacity (NVM) reliable nodes
- There exists a global (world wide) data addressing scheme

Examples

- A supercomputer does an exascale computation, the data remains stores on the nodes for a month to allow analysis before it can be deleted (while other simulations are running)

- Data storage is at the edge as a form of permanent storage because it is convenient to keep raw data for a while
  - Data from sensors are progressively analyzed to be used in simulation

- Intermediate nodes are used as storage and compute addendum (in-transit computing)

Conclusions & Perspectives

Some EXDCI Contributions - 1

- EHPCSW
  - Now an effective European HPC venue
- Extreme scale Demonstrator
  - Five workshops were held with different HPC and Big Data **stakeholders to further define and disseminate the concept of "Extreme scale Demonstrators"**. Such EsDs are targeting the integration and use of pre-exascale HPC system prototypes based on R&D results out of H2020 and FP7 projects.
- International
  - EXDCI has been actively promoting international collaboration with all FET-HPC projects and **has led the BDEC international roadmapping effort**. The resulting "Pathways to Convergence" report will be presented at SuperComputing'17 at Denver in November 2017.
- PRACE Scientific Case
  - PRACE Scientific Steering Committee to issue in **2018 a third version of the PRACE Scientific Case**, helping PRACE, the pan European HPC research infrastructure, to deploy (pre)Exascale systems and to shape new HPC and data services.

Some EXDCI Contributions - 2

- SME / Startup
  - We have put all stakeholders (HPC center, startups, SME, Constructors, ...) around the table to **issue a set of very concrete recommendations** to help SME growth and new startup to emerge.
- Strategic Research Agenda
  - Within the EXDCI project two issues of the Multiannual Roadmap for HPC Research under the H2020 framework will have been generated. A collaborative process involved other work packages of EXDCI as well as external HPC stakeholders for this Strategic Research Agenda elaboration. **The European 'Bible' of HPC Technology!**
- Talent Generation and Training
  - EXDCI has worked to address the shortage of HPC-skilled staff in the European workforce. The **HPC Careers Case Studies feature the personal stories** of enthusiastic HPC experts.
- Analysis of the Ecosystem
  - Setup and implementation of an **impact assessment methodology** which was applied to the EU HPC ecosystem progress monitoring. This contributed to the HPC Public Private Partnership mid-term review and assessment in 2017.

Why Exascale Projects Matter?

- Scientists don't do research on abstract machines
  - and buying the next generation of supercomputers is not a good enough approach to stay competitive
- Exascale is not an incremental change
  - Requires **the community to adapt** to a new way of using computing and storage → pre-exascale machines important for this reason
  - **This takes time, risks and many resources**
  - The later adaptation starts, the more expensive it is (e.g. impact on competitiveness, hiring, …)
- Industry is dependent on the academic community to produce trained PhD and engineers for cutting-edge technologies
  - Not happening without early access to a word class research HPC infrastructures based on advanced technologies
- The lack of EU hardware technologies creates strong dependencies on other countries
  - In a context of high competition, access is not denied but may be delayed

Exascale is more than One Order of Magnitude to Go

- Can one platform fit all?
  - Efficiency very dependent on data fluxes
  - Probably not, but the *Cloud* is not the magic bullet, the Big Data technology alone neither

- Paradigm changes are more complex to manage and more expensive than incremental ones
  - Organizational issues
  - Need the scientists on-board but also industry
  - Require data scientists with a different cultural background*
  - Project-based calls for proposals better adapted for incremental changes

*"We develop algorithms, we don't have time to deal with C/C++ or MPI"

Final EXDCI Recommendation

Focus and organize the current efforts in a way that is closer to an integrated industrial project in order to ensure a successful delivery of European Exascale level sustainable systems capable of serving a convergence based scientific discovery process.