BDEC = Convergence of Big Data & HPC

EXDCI
WP6 International Liaisons

Mark Asch
Université de Picardie, Amiens

ORAP #38– October 18th, 2016

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

What is EXDCI?

- EXDCI is a 30-month project starting from September 2015 with
  - a budget of € 2.5 million
  - with 173,5 PMs
  - funded in call FET-HPC 2014.
- EXDCI's objective is to support the coordination of the development and implementation of a common strategy for the European HPC Ecosystem.

Strategic Goals of EXDCI

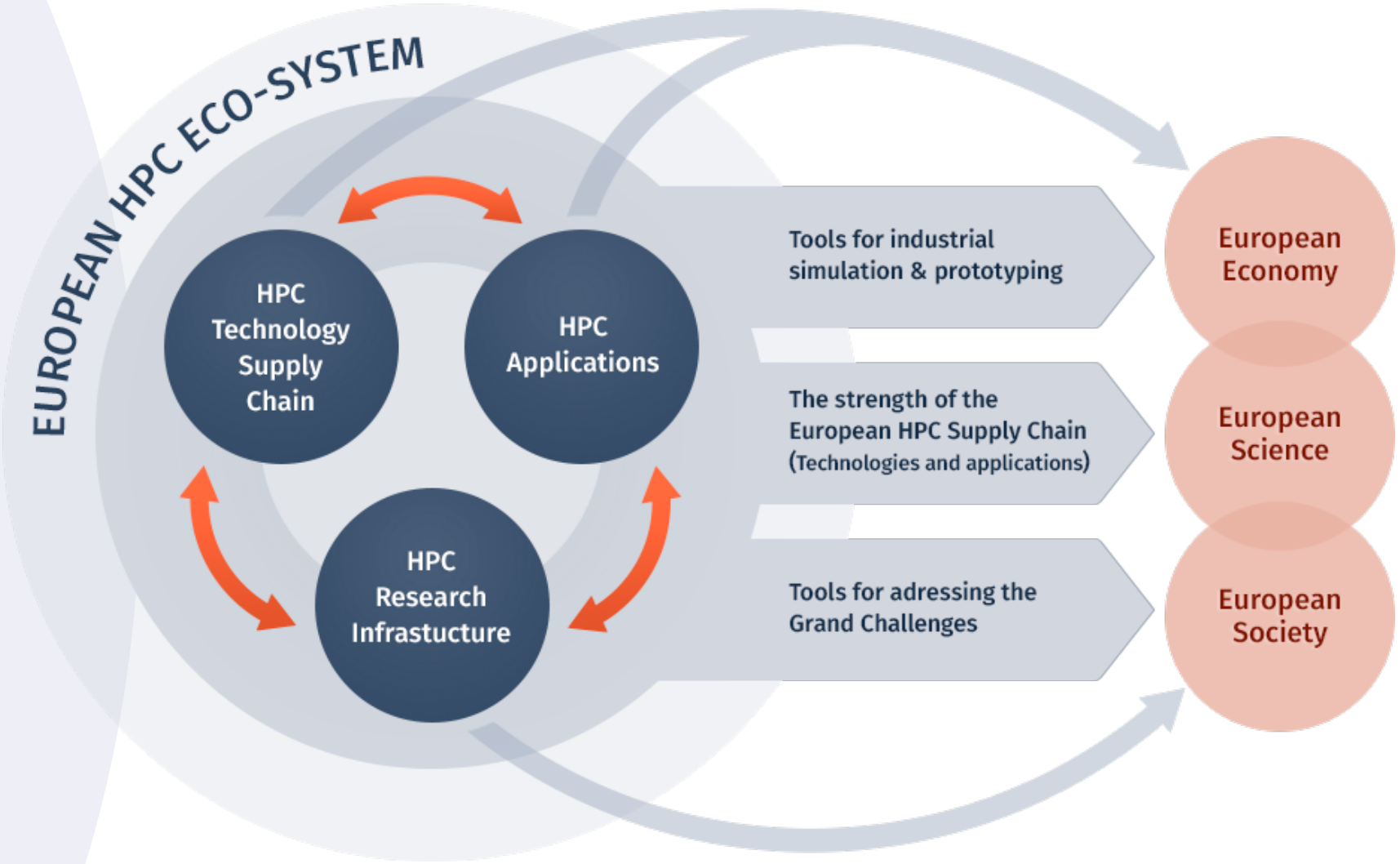
Development of a common European HPC Strategy - EXDCI will support the implementation of a common European HPC strategy through the coordination of activities of stakeholders such as the European Technology Platform for HPC (ETP4HPC), PRACE, application owners and users (including emerging HPC applications), the European exascale computing research community, the open source HPC community, and other related activities in other parts of H2020.

EXDCI aims to support the road-mapping, strategy-making and performance-monitoring activities of the ecosystem, i.e.:

- ✓ **Producing and aligning** roadmaps for HPC Technology and HPC Applications;
- ✓ **Measuring** the implementation of the European HPC strategy;
- ✓ **Building and maintaining** relations with other international HPC activities and regions;
- ✓ **Supporting** the generation of young talent as a crucial element of the development of European HPC.

exdci European Extreme Data & Computing Initiative

European HPC Ecosystem



What is BDEC not?

IESP: 2009-2012

International Journal of High Performance Computing Applications, **February 2011**, vol. 25 no. 1 **3-60**

The International Exascale Software Project Roadmap[1]

Jack Dongarra, Pete Beckman, Terry Moore, Patrick Aerts, Giovanni Aloisio, Jean-Claude Andre, David Barkai, Jean-Yves Berthou, Taisuke Boku, Bertrand Braunschweig, Franck Cappello, Barbara Chapman, Xuebin Chi, Alok Choudhary, Sudip Dosanjh, Thom Dunning, Sandro Fiore, Al Geist, Bill Gropp, Robert Harrison, Mark Hereld, Michael Heroux, Adolfy Hoisie, Koh Hotta, Yutaka Ishikawa, Zhong Jin, Fred Johnson, Sanjay Kale, Richard Kenway, David Keyes, Bill Kramer, Jesus Labarta, Alain Lichnewsky, Thomas Lippert, Bob Lucas, Barney Maccabe, Satoshi Matsuoka, Paul Messina, Peter Michielse, Bernd Mohr, Matthias Mueller, Wolfgang Nagel, Hiroshi Nakashima, Michael E. Papka, Dan Reed, Mitsuhisa Sato, Ed Seidel, John Shalf, David Skinner, Marc Snir, Thomas Sterling, Rick Stevens, Fred Streitz, Bob Sugar, Shinji Sumimoto, William Tang, John Taylor, Rajeev Thakur, Anne Trefethen, Mateo Valero, Aad van der Steen, Jeffrey Vetter, Peg Williams, Robert Wisniewski, and Kathy Yelick

Abstract

Over the last twenty years, the open source community has provided more and more software on which the world's High Performance Computing (HPC) systems depend for performance and productivity. The community has invested millions of dollars and years of effort to build key components. But although the investments in these separate software elements have been tremendously valuable, a great deal of productivity has also been lost because of the lack of planning, coordination, and key integration of technologies necessary to make them work together smoothly and efficiently, both within individual PetaScale systems and between different systems. It seems clear that this completely uncoordinated development model will not provide the software needed to support the unprecedented parallelism required for peta/exascale computation on millions of cores, or the flexibility required to exploit new hardware models and features, such as transactional memory, speculative execution, and GPUs. This report describes the work of the community to prepare for the challenges of exascale computing, ultimately combing their efforts in a coordinated International Exascale Software Project.

Keywords

High Performance Computing, Software Stack, Exascale computing

[1] The International Exascale Software Project was organized by and has received ongoing support from a variety of national agencies: In the United states, the Department of Energy Office of Advance Scientific Computing Research (DOE-ASCR) and the National Science Foundation Office of CyberInfrastructure (NSF-OCI); In France, the Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), Agence nationale de la recherche (ANR), INRIA and Teratec; In the United Kingdom, Engineering and Physical Sciences Research Council (EPSRC); In Japan, The University of Tsukuba, RIKEN, Kyoto University, Tokyo University and the Tokyo Institute of Technology. Corporations contributing to the staging of different IESP meetings have included Cray, EDF/EESI, IBM, Intel, Fujitsu Ltd., and NVIDIA.

What is BDEC?

- IESP (2009-2012)
  - IESP Roadmap
- BDEC (2013-...) = { EU, USA, JAPAN, CHINA }
  - EESI – European Exacscale Software Initiative
  - EESI2 – European Exacscale Software Initiative
  - EXDCI

1.Create an international collaborative process focused on the co-design of software infrastructure for extreme scale science, addressing the challenges of both extreme scale computing and big data, and supporting a broad spectrum of major research domains, 2.Describe funding structures and strategies of public bodies with Exascale R&D goals Worldwide 3.Establishing and maintaining a global network of expertise and funding bodies in the area of Exascale computing

EXDCI WP6

Task 6.3: EXDCI in Europe and Worldwide (aka BDEC)

ORAP #38– October 18th, 2016

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

Achievements to date

- BoF in SC'15.
- 2-day BDEC workshop in Frankfurt, June 2016.
  - 70 invited participants from 10 countries.
  - 6 keynotes, 4 national roadmaps (EU, USA, JP, CN), 21 presentations.
  - Dynamic breakout sessions.
- ½-day BDEC workshop in ISC'16, June 2016.
  - Reporting on BDEC workshop.
  - Perspectives and roadmaps.
  - Discussion on new roadmap.
- Intensive drafting of a new international roadmap.
- BoF in SC'16

BDEC @ Europe

Conclusions to date

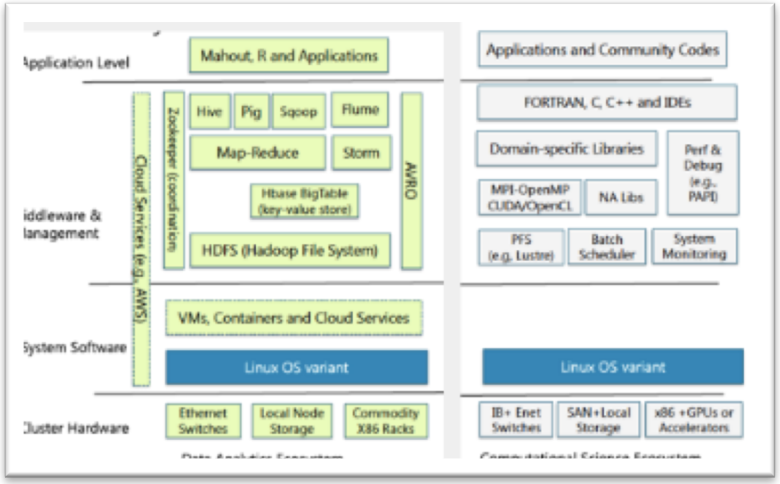
- BDEC is alive and well: Europe, through EXDCI, is an central player in international road-mapping.
- EXDCI has been an important engine.
- Europe is the only continent that has this type of CSA.
- Sharing of international perspectives is vital for future research and funding strategies in e-infrastructures.
- Convergence of Big science and Big data is going to play an increasingly important role on the global scene.

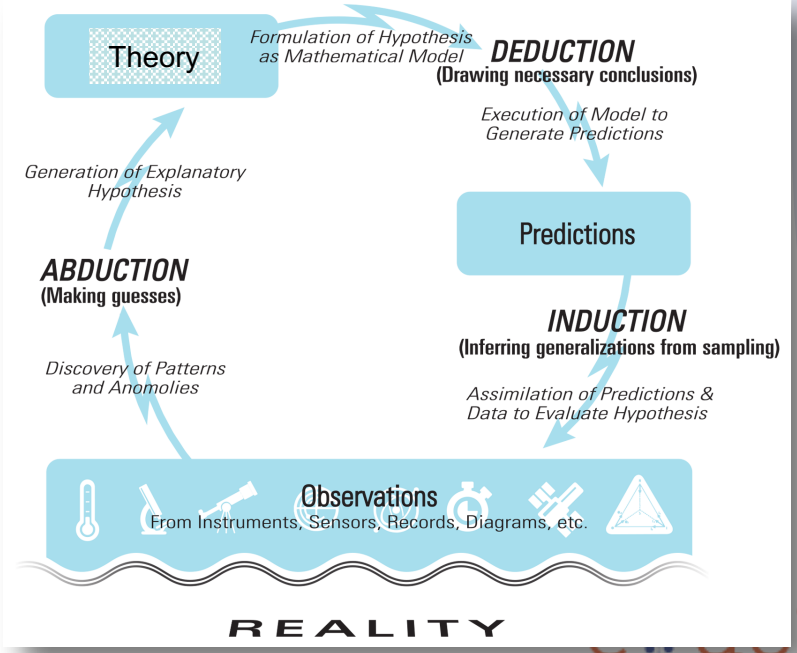
Going forward

- BoF in SC'16 where an initial draft of the new report will be presented.
- Coordination of contributions to the final draft, entitled "Pathways to Convergence" – see next slides.
- Publication of the Pathways document in the most highly visible journals.
- Publication of white papers targeted at funding organisations and decision makers.
- Close collaboration with the EOSC initiative.
- Close collaboration with the BDVA cPPP.
- Preparation of 2017 BDEC meetings (USA, Japan).

Pathways to Convergence

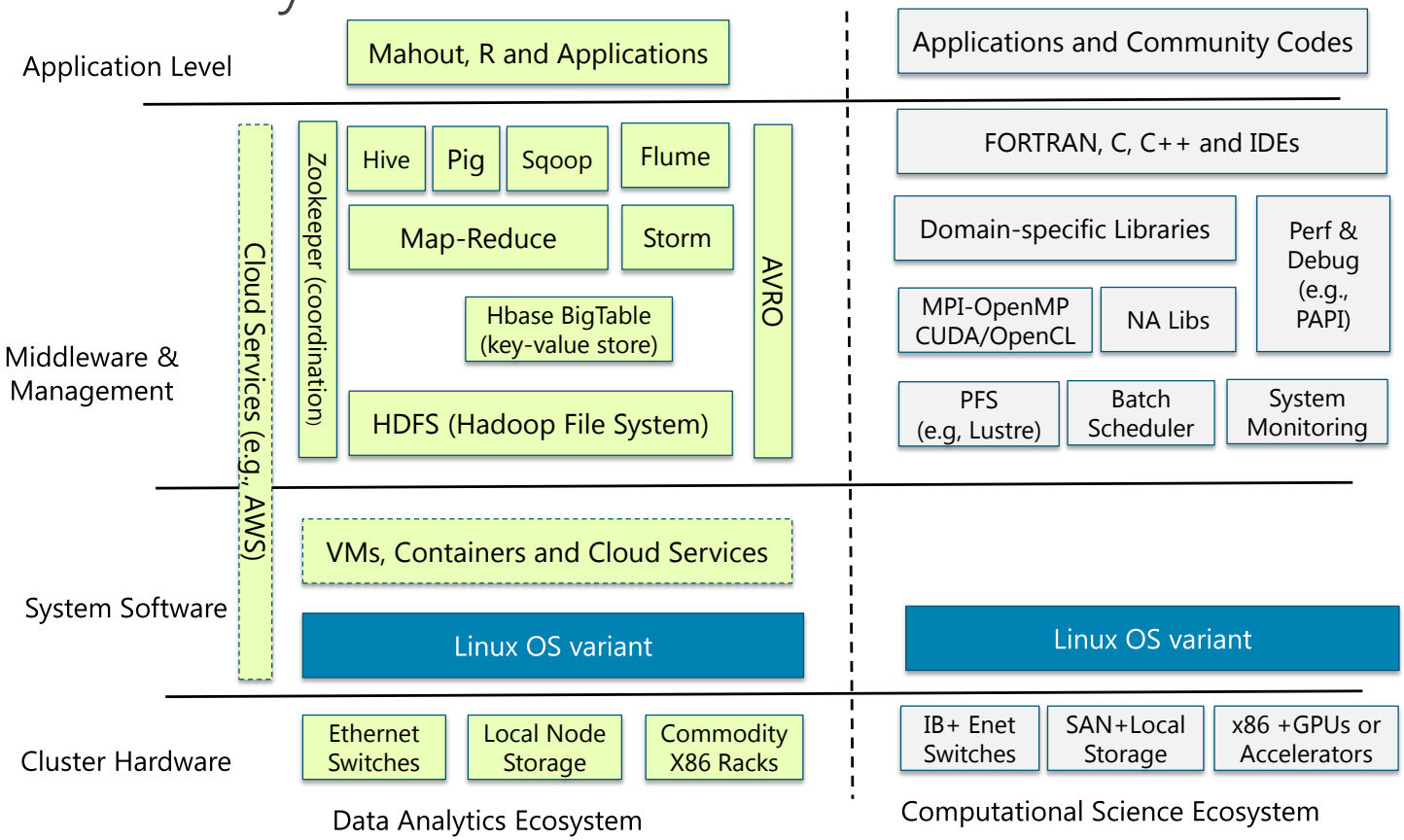
- What is the problem?
  - 2 separate software stacks – 1 for HPC, 1 for Big Data.
  - Impossibility to continue investing in both.
  - Outdated paradigm for scientific process.
- What is the solution?
  - Convergence of the software stacks.
  - Coordinated investment strategies.
  - Updated scientific process paradigm…





Bifurcation ahead:

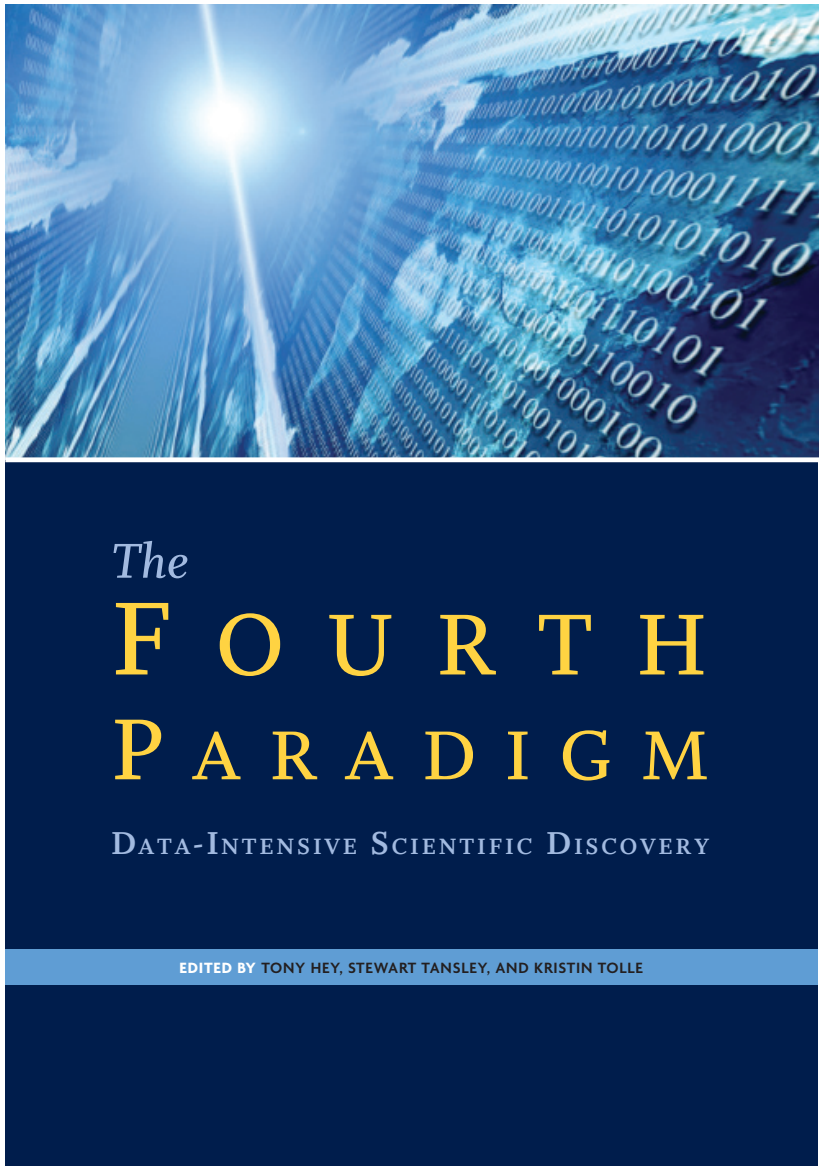
Two ecosystems



10

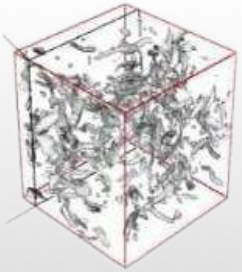
Credit: J. Dongarra, D. Reed

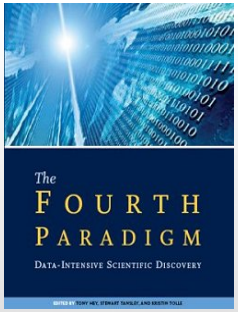
2009:


The
FOURTH
PARADIGM
DATA-INTENSIVE SCIENTIFIC DISCOVERY
EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

4th silo…

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$





Experimental	Theoretical	Computational	The Fourth Paradigm
Thousand years ago *Description of natural phenomena*	Last few hundred years *Newton's laws, Maxwell's equations…*	Last few decades *Simulation of complex phenomena*	Today and the Future *Unify theory, experiment and simulation with large multidisciplinary Data* *Using data exploration and data mining (from instruments, sensors, humans…)* *Distributed Communities*

Crédits: Dennis Gannon

Pathways to Convergence: outline of the report (1)

- Introduction
- A Common Context for Planning Cyberinfrastructure Convergence
  - A More Analytical Model of Scientific Inquiry
  - Confronting the General Problem of Data Logistics
    - Alternative 1: Streaming
    - Alternative 2: Content Delivery Networks (Processing In-transit)
    - Alternative 3: Computing at the Edge (at the Source)
    - Other material relevant to this section:
  - Software Infrastructure for Science as Social Investment
    - Stakeholders in Software Infrastructure
  - What is to be gained from a converged cyberinfrastructure?

Pathways to Convergence: outline of the report (2)

- Pathways to Convergence: What are the Alternatives?
  - The classical paradigm of a common infrastructure: The hourglass architecture.
  - BDEC target goal onto the BD stack; use HPC technology underneath
  - Continue in the Bifurcated world, but do containerization to achieve some commonality
- An Architecture for the New Stack
  - Context for Architecture Section
    - Pragmatic meaning of convergence: Building a new hourglass
  - Containerization: The new "narrow waist"?
  - Containers with adequate HPC and security capabilities

Pathways to Convergence: outline of the report (3)

- Trends in the Hardware Platform Substrate
- Applications in the Scientific Research Process ("Applications")
  - Taxonomy of Application/workflow patterns and templates
  - Science at the boundary of observation and simulation
  - Numerical laboratories
  - Multimodal sensing of the same reference object
- Layers and Components in the Scientific Software Stack
  - Math Libraries, Software Ecosystems for app development
- Operations and Systems Management
  - Radically improved resource management
  - QoS and Performance tools
- Conclusions and Recommendations

European Extreme Data & Computing Initiative

Thank you!

mark.asch@u-picardie.fr

https://exdci.eu

ORAP #38– October 18th, 2016

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