Multidisciplinary High Performance Data Analysis Forum: summary and recommendations

P.-H. Cros[1], M. Daydé[1,3], O. Marques[2]

[1]Informatics Research Institute of Toulouse – Toulouse University – CNRS

[2] Scientific Representative at CNRS

[3]Lawrence Berkeley National Laboratory

Introduction

- Multidisciplinary High Performance Data Analysis for Societal Challenges Forum held at Palais des Evêques, Saint Girons, France, on June 17-19, 2018.

- Organized by Informatics Research Institute of Toulouse (IRIT), Lawrence Berkeley National Laboratory (LBNL), with support of the LabEx CIMI (Centre International de Mathématiques et Informatique de Toulouse).

- Report authored by the forum's organizing committee with input from attendees.

- Participation at the forum by invitation only including decision-makers in research laboratories and academy, all faced with problems related to societal challenges, coming from France, Japan, USA, Germany, Spain, …

REPORT

"Multidisciplinary High Performance Data Analysis for Societal Challenges" Forum

June 17-19, 2018
Saint-Girons (Palais des Evêques), France

Editors :
Pierre-Henri Cros - Michel Daydé - Osni Marques

Motivations for the Forum

- October 9, 2014, workshop hosted at Embassy of France in Washington:
    - Organized by french Embassy, CNRS, CNES, IRIT, and LBNL.
    - To discuss potential collaborations between EU and US in areas involving big data, with an anticipated great impact to society.
    - Participants included representatives from R&D communities, US funding agencies (DOE, NITRD, NIST, NSF), and from Europe (France, Germany, Italy, Netherlands and United Kingdom).

- Plenty of other examples when setting up projects in transportation, heath, biology, environment, …

- Practises related to data in CNRS
    - Better identification of practises and needs
    - Promotion of a transdicisplanry synergy around data

Big Data for Society workshop on Oct 9, 2014

Published on Wednesday September 24, 2014 , Lire cette page en français [fr] , Share

CHALLENGES AND PERSPECTIVES ON COLLABORATIONS BETWEEN THE EUROPEAN UNION AND THE UNITED STATES IN BIG DATA FOR SOCIETY.

▸ When: October 9, 2014, 8:30 AM - 2:00 PM
▸ Where: Embassy of France in the United States, 4101 Reservoir Road NW, Washington, DC

Livre Blanc sur les Données au CNRS
État des Lieux et Pratiques

Mission Calcul et Données (MiCaDo)

Janvier 2018

Conclusions of CNRS report on practises related to data

- **Organization, management, scientific exploitation / valorization of data produced ithin CNRS are major challenges: amounts of data explode**
- **Big Data concerns all CNRS research institutes at various levels**:
  - Well established practices and *stewardship* of data (IN2P3, INSU, INSB, INSHS).
  - Structuration going on to answer to the needs (INEE)
  - Still embryonic consideration of the data except for some TGIR (INC, INP, INSIS)
  - Data and data analytics are research topics (INSMI, INS2I)
- **Logistics and Stewardship of data are two of the main challenges in addition to recruit or train more data scientist**
- **Important issues: interdisciplinarity, platforms, user support, scalable data analysis workflows, multi-sources data, convergence HPC / HDA** (High-End Data Analysis) **in complex workflows**
- **Define new strategies / architectures**
- **Energy efficiency !**
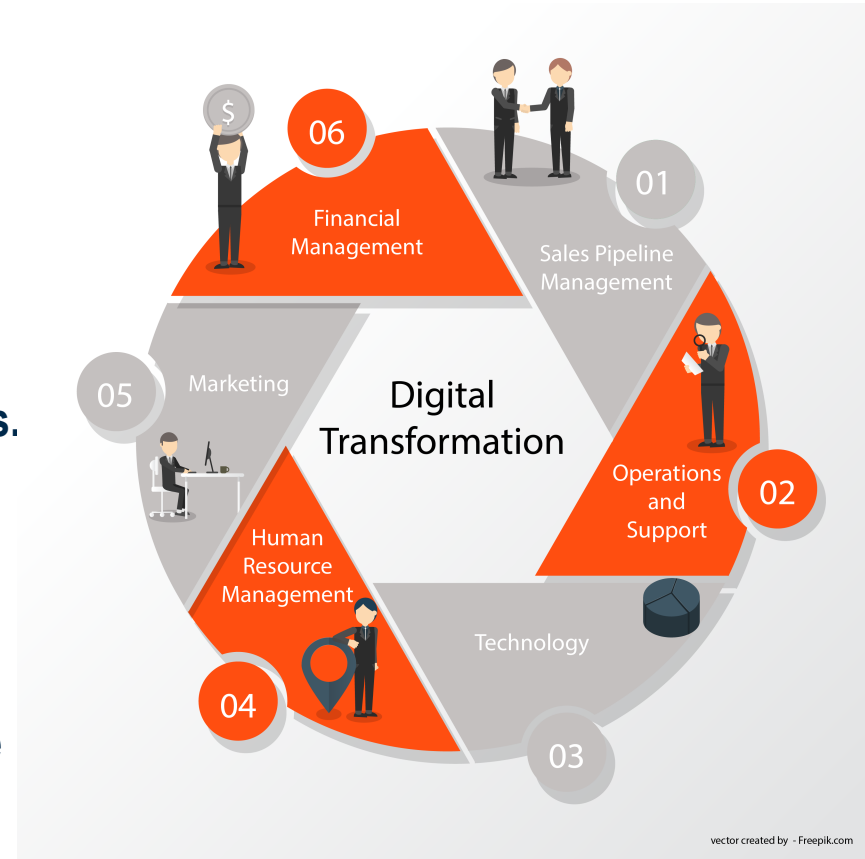
A new paradigm for the research

- ⦿ Scientific activity undergoes an epistemological upheaval = > new forms of production of knowledge and emergence of several sub-disciplines.

- ⦿ **New fields of investigation born at the interfaces of the scientific disciplines:**

  - ○ **bioinformatics**, **computational neurosciences**, **cyber-security**, **digital humanities**, **geoinformatics**, **e-health**...

  - ○ Example : **astroinformatics that incorporates astronomy, astrophysics, computer science and signal processing**

  - ○ …

Digital transformation

- **Integration of digital technology into all areas of society, science and industry**

- **It changes fundamentally organization:**

  - **Research:**

    - **New scientific paradigms**

  - **Industry**

    - **Operation and value delivery to customers.**

  - **Society:**

    - **Government, mass communication, medicine**

  - **It's also a cultural change that requires organizations to continually evolve: challenge the status quo, experiment, and resilience to failure**

- **Two main current revolutions:**

  - **Big Data => High performance data analytics**

  - **Artificial Intelligence**

06 Financial Management
01 Sales Pipeline Management
02 Operations and Support
03 Technology
04 Human Resource Management
05 Marketing

Digital Transformation

vector created by - Freepik.com

*-disciplinarity

- Intradisciplinary: working within a single discipline.

- Crossdisciplinary: viewing one discipline from the perspective of another.

Multidisciplinary: people from different disciplines working together, each drawing on their disciplinary knowledge.

- Interdisciplinary: integrating knowledge and methods from different disciplines, using a real synthesis of approaches.

- Transdisciplinary: creating a unity of intellectual frameworks beyond the disciplinary perspectives.

Agenda of the Forum

Monday June 18:

- Morning from 9:00: keynote talks.
  - Welcome and Introduction: Michel Daydé, Head of IRIT
    - Horst Simon, Deputy Laboratory Director for Research, Chief Research Officer, Lawrence Berkeley National Laboratory
    - Tetsuya Sakurai
    - Jacques Crémer, Toulouse School of Economics
    - Robert Bojdak, EC
  - Lunch
- Afternoon
  - Discussion in small working groups for producing contributions to the brown report that will be the main results of the Forum.
  - Evening : reception and banquet

Tuesday June 19:

- The morning will be devoted to the restitution of the working groups. The synthesis of the discussion will give rise to the brown report
- Lunch then transportation organized back to Toulouse

REPORT

"Multidisciplinary High Performance Data Analysis for Societal Challenges" Forum

June 17-19, 2018
Saint-Girons (Palais des Evèques), France

Editors :
Pierre-Henri Cros - Michel Daydé - Osni Marques

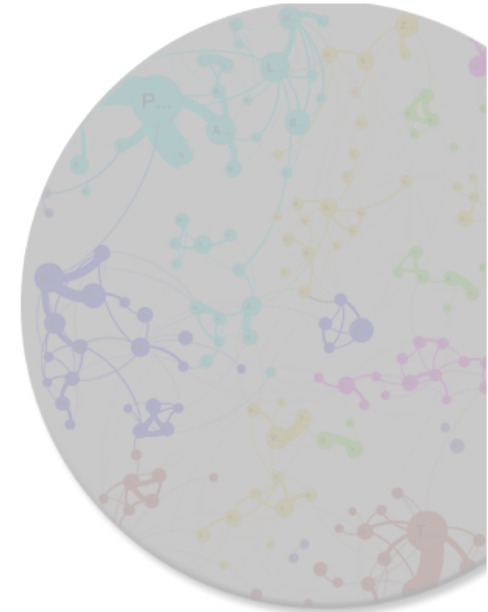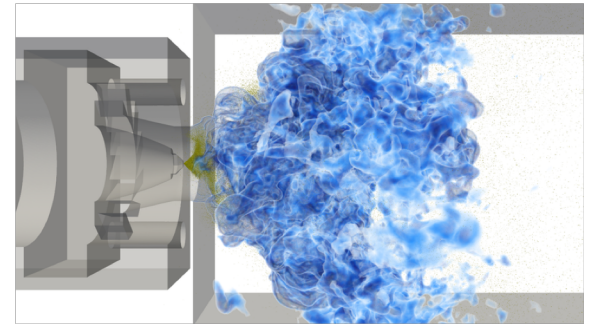# Goals of the Forum

MERCATO burner with YALES2, V. Moureau, CORIA

- Transition from knowledge gained through computer simulations to knowledge gained from experimental and observational data

- Foment discussions about ongoing or yet to-be-implemented mechanisms for data collection and processing related to societal challenges: Is one possible approach to form teams regrouping data scientists, HPC users, specialists of the considered area, jurists, people from social sciences, … ?

- This implementation needs to be discussed andmportance of assessing the potential impact of such mechanisms, particularly on the management of teams working on those challenges by:
  - Measuring impact of involvement of different professions and skills through new metrics,
  - Identifying appropriate communication strategies to ensure stakeholder input,
  - Estimating potential impact that this approach may have in our research organizations and in the future call for proposals at national or international levels.

- Understanding and predicting features that may play a key role in societal challenges, e.g. personalized medicine or economic well-being, require specific approaches to ascertain the data to be processed.
  - E.g. datasets can be created by different disciplines, in non-uniform ways, and results need to be validated by end-users.

Insights on keynote talks

Keynote talks provided a variety of examples. Inputs from participants reinforced the need for an increased multidisciplinary HPDA.

HPC and Sustainability - Horst Simon (LBNL)

- Emphasized how science is poised for transformation from lone scientist, to scientific teams (e.g. around large scale instruments), and now to innovations arising from the use of AI (sometime from lone scientists).

- Experimental science changing toward regrouping computing, experiments, networking and expertise into "Superfacility".

- HPC is not any longer exclusively devoted to numerical simulations: it is more and more concerned by Data Analysis (and AI ?).

- Three use cases in sustainability perfectly illustrated Forum motivations:

  - Sustainable water, energy and food production applied to the watersheds in Sierra Nevada and Upper Colorado: demonstrates how HPC and AI can improve understanding and prediction of the behavior of watersheds throughout the world.

  - How machine learning can contribute to actionable and sustainable agriculture.

  - Modelling coupled urban systems for example using CityBES data and computing web platform: demonstrates that modelling building in cities can be an exascale computing problem

HPC and Sustainability

Horst Simon

Deputy Lab Director for Research

with Kathy Yelick, Susan Hubbard, Ben Brown, and Mary Ann Piette

Lawrence Berkeley National Laboratory

hdsimon@lbl.gov

IRIT Forum, June 18, 2018

U.S. DEPARTMENT OF ENERGY Office of Science   - 1 -

Insights on keynote talks (con't)

Inter-Disciplinary Research Hub for Simulation, Data Analysis, and AI - Tetsuya Sakurai (Tsukuba Univ.)

- Talk introduced "Society 5.0" concept proposal for an advanced, future and human centered society where the integration of the cyber and physical spaces is realized through AI, Internet of Thing (IoT), robotics and Big Data.

- Goal: achieve a society providing necessary goods and services to anybody at any time regardless of region, age, gender, language or other limitations, i.e. to achieve economic growth, well-being and overcome societal challenges at the same time, for prosperity of global communities.

- Big Data and Artificial Intelligence play crucial role e.g. to provide improvements in healthcare and caregiving

Inter-Disciplinary Research Hub for
Simulation, Data Analysis, and AI

Tetsuya Sakurai
Director, Center for Artificial Intelligence Research (C-AIR)
University of Tsukuba

Economists and Data Analysis - Jacques Crémer (Toulouse School of Economics)

- Talk made the link between economics and data analysis with examples in IT illustrating the role and the implication of economists: Google, Uber, Amazon.

- Theoretical studies of economy are also applied to the digital economy: incentives and auctions, cost allocation and fairness, how big a comparative advantage does data give to incumbent firms, competition policy, …

- The talk pointed out the differences between economists and scientists from engineering:
  - Economists are intently aware of incentives, they do not think that their models represent reality.
  - Data analytics and high performance computing are not sufficient since it is an old idea that big all-inclusive models solve all the problems while economists typically believe that a multiplicity of models is a good thing.
  - Also causality and correlation are not the same thing and economic policy changes the systems (e.g. housing prices).

Conclusion: need for organic communications and multidisciplinary cooperation that cannot be obtained in disciplinary silos organizations.

The European Commission's 2018 Data Package: Towards a Common European Data Space - Robert Bojdak (European Commission, DG CONNECT)

- Talk involved several points: 2018 Data Package, Artificial Intelligence Strategy, EuroHPC Joint Undertaking and the current and future funding opportunities.

- Commissioner Mariya Gabriel has declared: "*Data lies at the core of the 4th Industrial Revolution. This is an essential resource for economic growth, competitiveness, innovation, creation and society's progress in general*". Big Data can provide societal benefits: easier and better lives for individuals and allow to address societal challenges. It is also expected that it will contribute to economic growth with the potential to double the size of the data economy (4% of EU GDP by 2020) and improves the efficiency of most sectors of the economy.

- Data workers expected to increase from 6 million in 2016 up to more than 10 million by 2020. Number of Data companies could increase from 255 000 in 2016 up to 360 000 in 2020 while the economy value could increase from almost 300 billion € up to 739 billion € during the same period (source European Data Market study).

- Which explains the EC communication entitled "*Towards a common European data space*" that states that data economy will flourish if different types of data are accessible and reusable: across borders, for and by different types of organisations (private, public, research) and for and by different sectors (e.g. energy, manufacturing, health etc.).
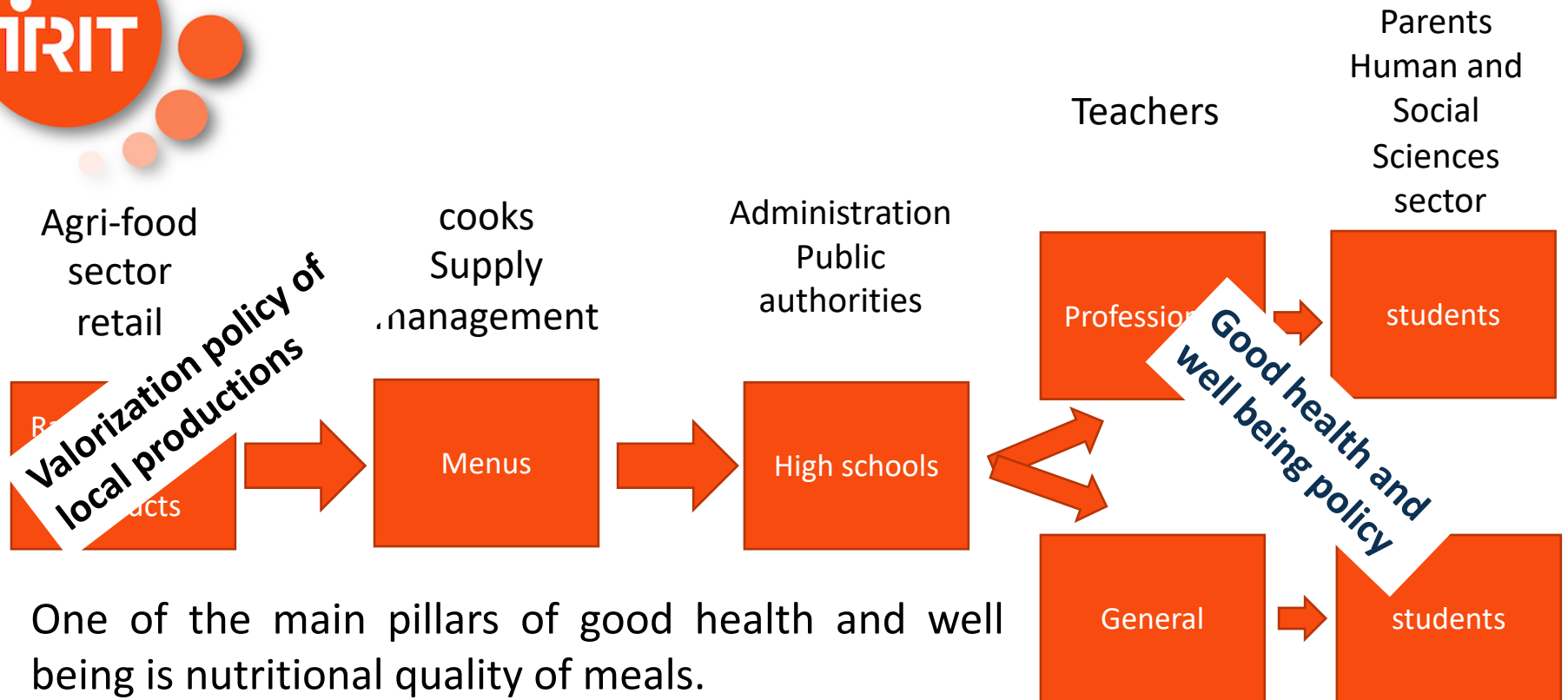

The European Commission's
2018 Data Package:
Towards a Common European Data Space

Robert BOJDAK

European Commission, DG CONNECT

HPDA Forum, Saint Lizier, 18-19 June 2018

Example of application at IRIT

Good Health and Well Being in Occitania Region

Parents
Human and Social Sciences sector

Teachers

Agri-food sector retail

cooks
Supply management

Administration
Public authorities

Professional

students

Valorization policy of local productions

Good health and well being policy

Menus

High schools

General

students

- One of the main pillars of good health and well being is nutritional quality of meals.
- Quality and diversity of menus to achieve appropriate nutritional quality are primarily a matter of education and that from an early age.

Occitania provides 22 million meals a year to high school students and wants to evaluate, strategies to simultaneously develop a local policy of good health and well-being and a policy of valorization of local productions of its agri-food sectors

Where is the difficulty ?

o **As other labs, IRIT is involved on some « hot » topics** : Big Data, Internet of Things, Artificial Intelligence, security…

o **Deeply involved in various areas** : security, biology, agronomy, health, social sciences and humanities, smart city…

o **Big Data and AI has an impact on structuration of our scientific landscape**

o *How to answer to scientific / industrial and societal challenges with HPDA ?*
  o Not enough providing numerical libraries, managing processing facilities, software tools, …
  o We need to set up a whole ecosystem: processing and storage capabilities, network, visualization, skills and user support
  o But still not enough

o *Develop interaction with other disciplines:*
  o Consortiums involving experts of the considered area, humanities, law, ethic, economy, mathematicians, computer scientists , …
  o This should be reflected in our practices, in call for proposals, ….

Strong change of paradigm !
Are we ready and organized for that ?

Discussions in the round tables

- Compilation of the topics discussed in the round tables:
  - Relevance and opportunities offered by HPDA
  - Data Analysis and Machine Learning (ML)
  - Do AI and deep learning solve everything?
  - Societal issues (humans versus data)

Conclusions and recommendations

- Forum participants concurred that:
  - Access to datasets as rich as possible is essential.
  - Setting up teams composed of members from different disciplines is equally essential, and that composition of such teams should be a function of the societal challenges to be addressed.
- Different projects are already addressing data access and processing issues. Forum participants agreed to work on potential solutions for the creation of adequate research teams to tackle specific problems. Often, practitioners work in silos, which hinders the formation and the work of teams with a broader expertise.
- Outcome of the first forum can be summarized by the two following recommendations:
  - **Recommendation 1**: *the role played by call for proposals*. Considering the numerous call for proposals announced each year, the participants recommended a clear identification of the topics or calls targeting societal challenges, for which funding and implementation would require an interdisciplinary approach.
  - **Recommendation 2**: *the importance of work recognition*. To motivate the work of interdisciplinary teams, the participants recommended that collective research results be included as one of the elements of the evaluation of team members' scientific work.

Next Forum middle of June 2019 will be discuss the potential impact of interdisciplinary approaches in our organizations and associated work environments