Giornate INAF 2019

TECNOLOGIE ABILITANTI: SOFTWARE, CALCOLO E TRATTAMENTO DATI

Ugo Becciani - OACT
ICT - INAF

Napoli 27 Marzo 2019

General Considerations

This is not an epoch of changes but the start of a change of epoch.

- Infrastructures are **rapidly evolving** in a new generation of more and more powerful computing elements
- Software and analysis tools **must be re-engineerized** to be still a good instrument for the new generation of Computing infrastructure

Main questions for the next few years

- **What are the main scientific challenges we are facing with the tools we have today?**
- **What are the main innovations in computer science we will have in the next future?**

- **Which steps do we have to take to be ready for the new generation of Hw and Sw?**
- **What new professional figures will INAF need in the next few years?**
- **What level of investment should we make?**

INAF - ICT Main Topics

The main INAF - ICT fields and challenges of our community can be summarized on four main topics (as described in the Strategic Vision document)

- **Control and real time software**
- **Open Data, Virtual Observatory and Data archives**
- **Data mining: Astroinformatics & Astrostatistic**
- **Big Data: HPC and new technologies**

Missing:

- **Data Reduction pipelines**
- **Analysis software**
- **Numerical simulations and modelling**

Control & RealTime software

Control & RealTime in astrophysics
driven by

a. instrumentation of **growing complexity**
b. **new generation of telescopes** and larger observatories

These facilities allow performing observations at higher spatial, temporal and spectral resolution, on wider fields and/or at increased depth.

As a consequence, advances in several technological fields will be required: optics, materials, detectors and high-performance **software for instrument control**, etc.

The embedded **real time SW** (for both instrument control and data processing) is a **fundamental component** of any astronomical instrument.

The new facilities will require control SW (close to) real-time to achieve the required performance and to target the desired science results.

STRATEGIC VIEW: Once the instrument HW is finalised, **the only component that can be still modified** to solve problems and to **optimize the instrument performances** is the **on-board control** and **pre-processing SW**.

Main approaches

- R&D: construction of **coherent frameworks** supported by space agencies
- **Multi-purpose frameworks** (e.g. the TANGO control system) supported by Scientific Institutions for relevant ground based experiments

INAF has a strong involvement in many projects of observing facilities and instruments and high level of expertise with ***leading role*** for both the approaches

Perspectives

- **Coordination and collaboration** to form a dedicated working groups to share either already implemented applications or already acquired expertise ⇒ keeping the costs under control and increasing the **overall efficiency**.
- This will **consolidate INAF as reference institution at European level**.

Archive Infrastructures @ INAF

- INAF has a **large experience** in the design, development and deployment of **large archives** like IA2 and SSDC (with ASI).
- The astronomical data has **peculiarities** that requires the development of **dedicated archives and access techniques.**

Priorities

- ***Investments on new technology for efficient access*** to online and **preserved** data;
- **Long Term preservation;**
- **Keywords**: interoperability, *federation between data and computing*, Virtual Observatory

FAIR: Findable, Accessible, Interoperable and Reusable

- **Virtual Observatory** and the International Virtual Observatory Alliance **have been the aggregators for the efforts on "Open Science" and FAIR-ness** in Astronomy and Astrophysics**.**
  - Data FAIRness has never been an issue in astrophysics **INAF is already a leader institution**
- The **European Open Science Cloud** effort is the main EU investment on opendata and cross-domain interoperability: INAF is involved in EOSC and related projects

PERSPECTIVE

- **We need to close the gap in data curation and metadata enrichment**, **defining a common strategy to deal with observed/simulated data**
  - from its first generation up to the end dissemination.
- We need to build a specialist of data-aware researchers alongside data scientist and data stewards.

Anything we want to know about a particular problem is specified and encoded within the data (i.e. data-driven).

The final goal of Big Data Science is to find those encodings, patterns and knowledge nuggets.

But Big Data Science requires:
ik. Efficient cross-matching.
Self-adaptive learning machines

- Characterize the known: Feature selection, Parameter space analysis & exploration
- Assign the new from the known: Regression, classification, supervised learning
- Explore the unknown: Clustering, unsupervised learning
- Discover the unknown: Outlier detection and analytics (serendipity)
- Benefits of very large datasets:
  - Best statistics of “typical” events, cross-correlation, automated search for “rare” events

The changing landscape of AI research

Challenge #1: it will be **prohibitively difficult to transport the data** to the user application.

Challenge #2: laboratories are useful **to measure and collect big data** in a systematic, controlled, repeatable fashion.

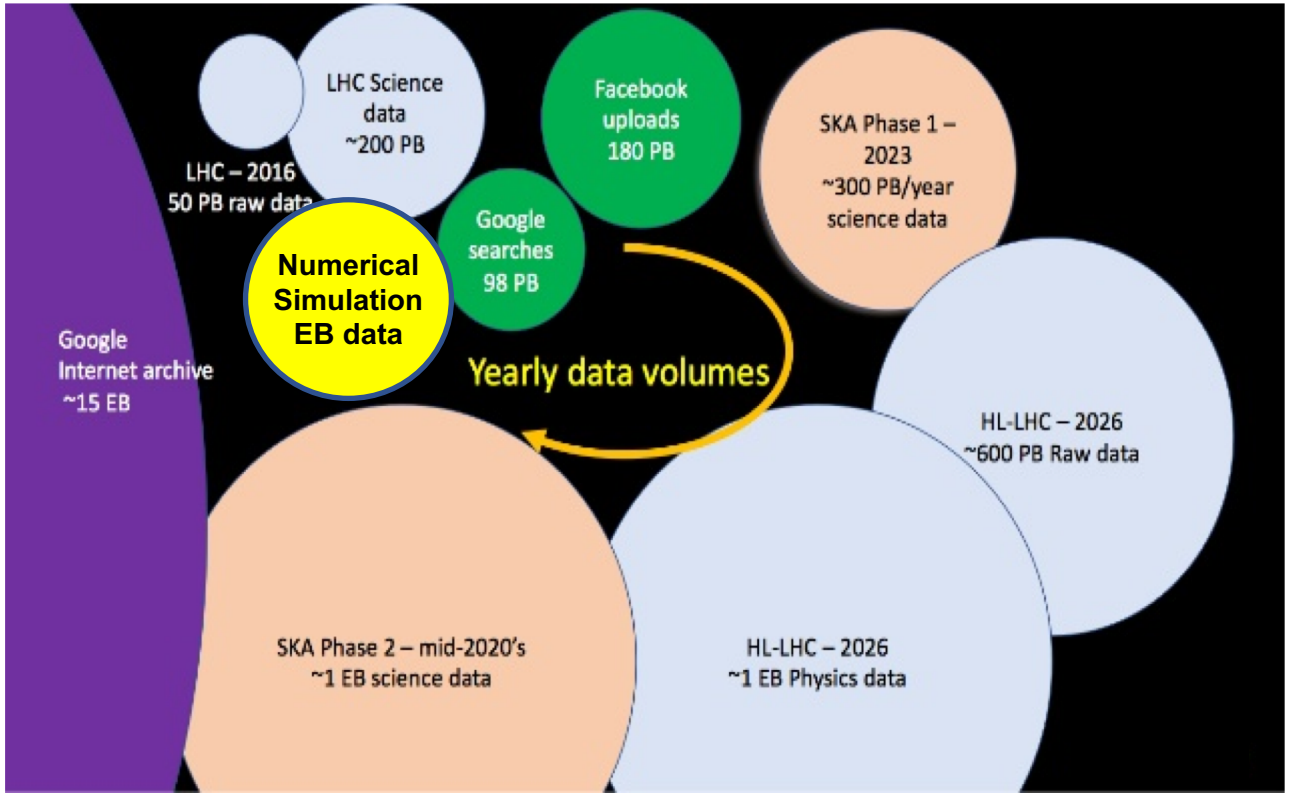
Challenge #3: we must be ready when huge of data will come. Mock data must be provided to ensure that data analytics methods will be compliant, efficient and scalable.

PERSPECTIVE

- ***Ship the code to the data:*** *We need Distributed Data Mining methodology…*
- **automatic self-adaptive methods are required to explore and cross-correlate their data**
- ***improve simulations and infrastructures to make intensive tests on your code!***

HPC and new technologies : the Big data challenge

Surveys will increase in data size, numerical simulations in complexity



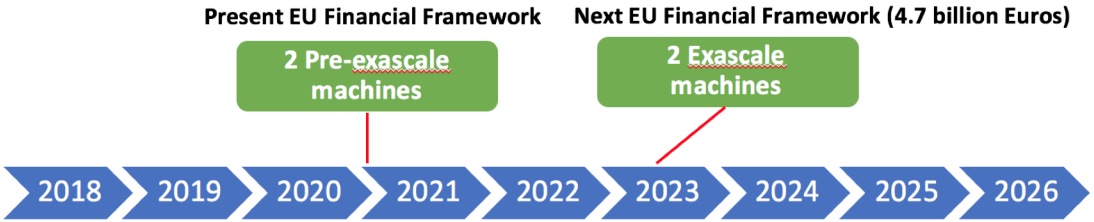
Two computational Infrastructures coordinated by the **ICT Unit**: **MoU with CINECA** (2017-2020) and **CHIPP pilot project**: National Tier-2 level infrastructure

INAF is involved in many e-infrastructure H2020 projects and initiatives in ICT fields.

Cloud @ EU EOSC (European Open Science Cloud): support and develop open science and open innovation in Europe and beyond. **6.7 Billions Euros**

HPC present: EuroHPC JU **6.1 Billion Euros to** Build Next-Generation Supercomputer

HPC Exascale computing: it is not (only) about scalability and Flops performance. New accelerators (FPGA, GPU, TPU), *non volatile memory NVRAM, High Trou. I/O,* ***Low Power***



Perspective

- The infrastructure will change rapidly and we **must use different models** (Cloud, HPC, HTC, HPDA) for same or different tasks
- Engineering for tools must change accordingly with **new algorithms and techniques**.

Network and services

- INAF institutes and observatories are linked with **GARR 1 Gbit/s** or 10 Gbit/s for particular situations.

Perspectives

- SKA Regional Centers in Europe will require:
  - **link at 100Gbit/s with GEANT**
  - GEANT dedicated link with Australia and South Africa
- BELLA Project: submarine cable to link GEANT with South America (ESO telescope, Alma, CTA)

Digital identity for INAF staff and collaborators (name.surname@inaf.it)

- Administrative services (user portal, salary docs, travels ……)
- Networking services EduRoam, Idem, Google (mail, drive …), Indico

Perspectives

Digital identity will be more and more required to use scientific services (access to databases, proposal submissions, acquisition of observational data, computing, etc..)

Crucial aspects to maintain the level of excellence

- **Which steps do we have to take to be ready for the new generation of Hw and Sw?**

New *instrument control and data processing software* development needs to follow the innovation of more powerful hardware.

The challenge will be to use collaborative environments to share and distribute algorithms, libraries and tools

The Astrophysics Big Data era is arriving. We are not yet totally ready for this challenge. We need to **start specific agreements/plans** (also at government level) **to obtain funds for e-infrastructure** also to be ready for FP9.

FP9 program will be the **main opportunity for networking and research** and innovation.

Conclusion 2/ 2

- **What new professional figures will INAF need in the next few years?**

  Big Data and Computing, Instrument control and data processing software*:* joint teams that can share new competence and software development.

  Astronomers must have (moderate**) computing skills, specific engineering teams (**HPC experts, data specialists, data analysts, and more)
  Training and high formation & **Opportunities**: Working on this field ***must*** open full careers opportunities

- **What level of investment should we make?**

  e-infrastructure to support research: 1.2 MEuro per year for current INAF infrastructure (GARR, Archive, Computing), **fundraising** for at least ***3-5 Million Euros per year*** should be foreseen.

FINAL CONSIDERATION - ICT point of view

We are not a "scientific usecase"

We cannot delegate other institutions by losing the specific skills we have in ICT