

# INAF Big Data Challengers

R. Smareglia

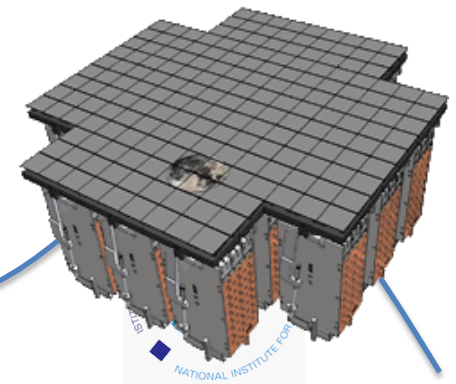
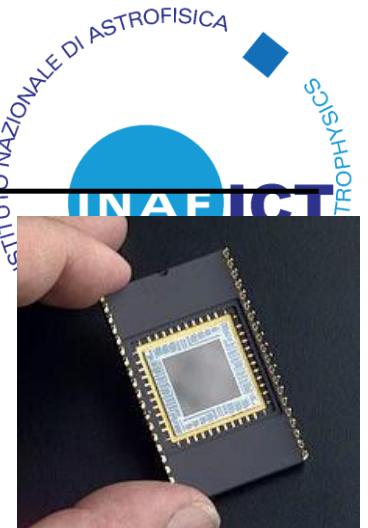
Head of ICT & Science Data Management office

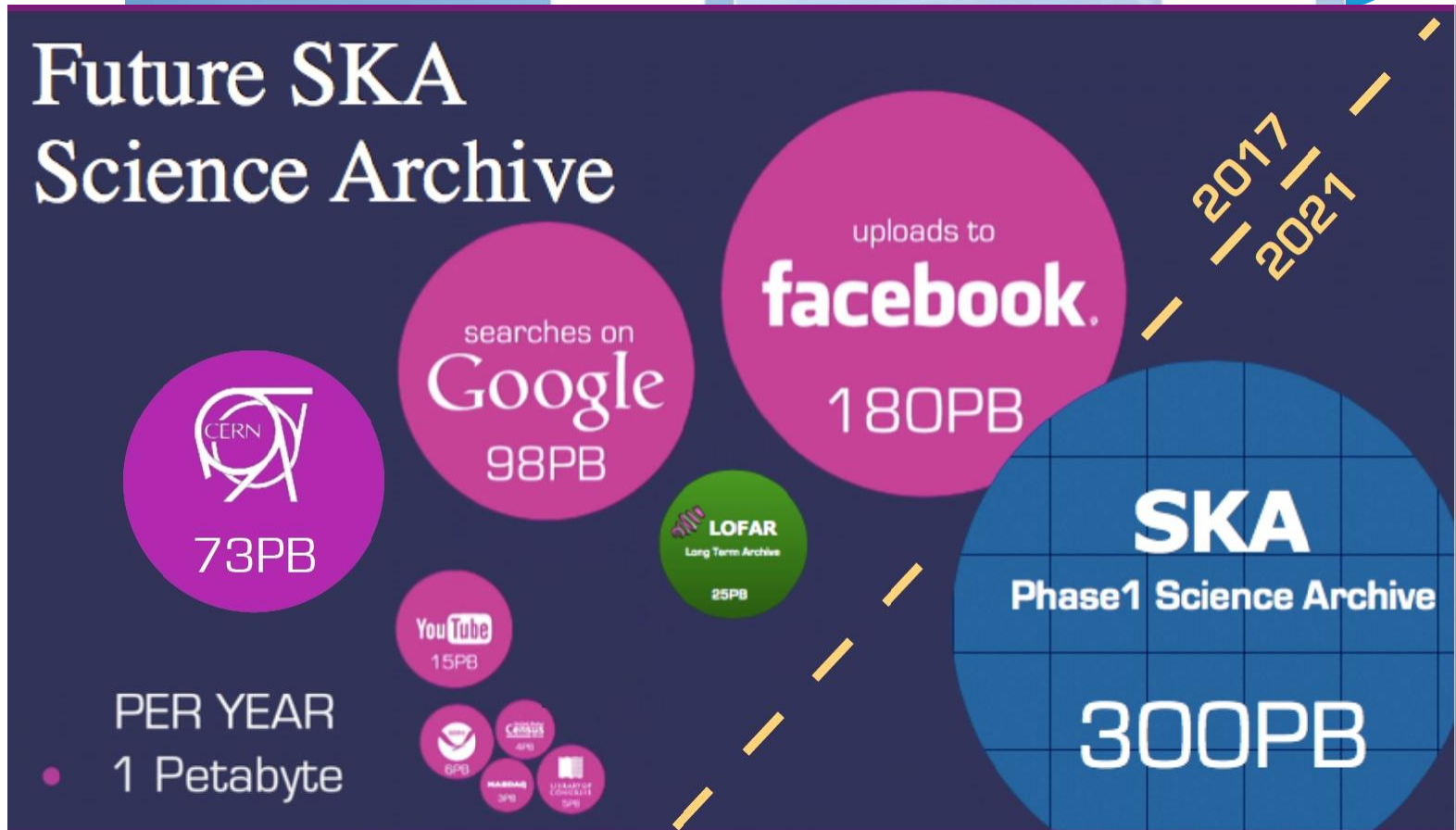
Second Italy-Ukraine Meeting in Astronomy

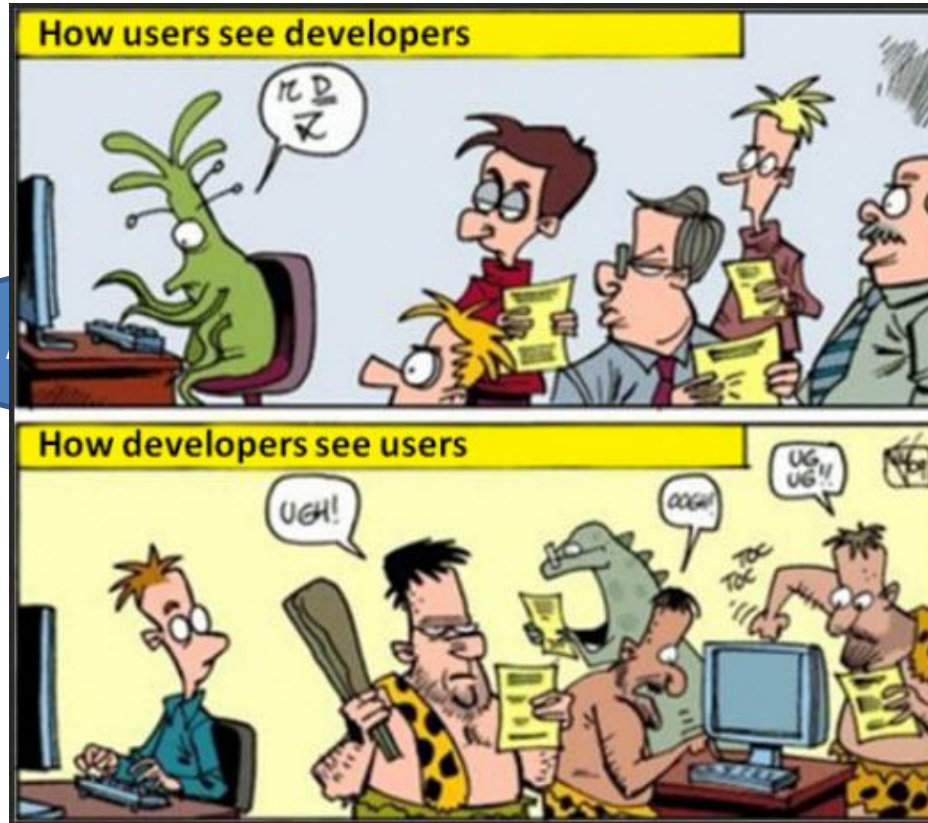
Kharkiv - 25-27 September 2018

# Data rate evolution

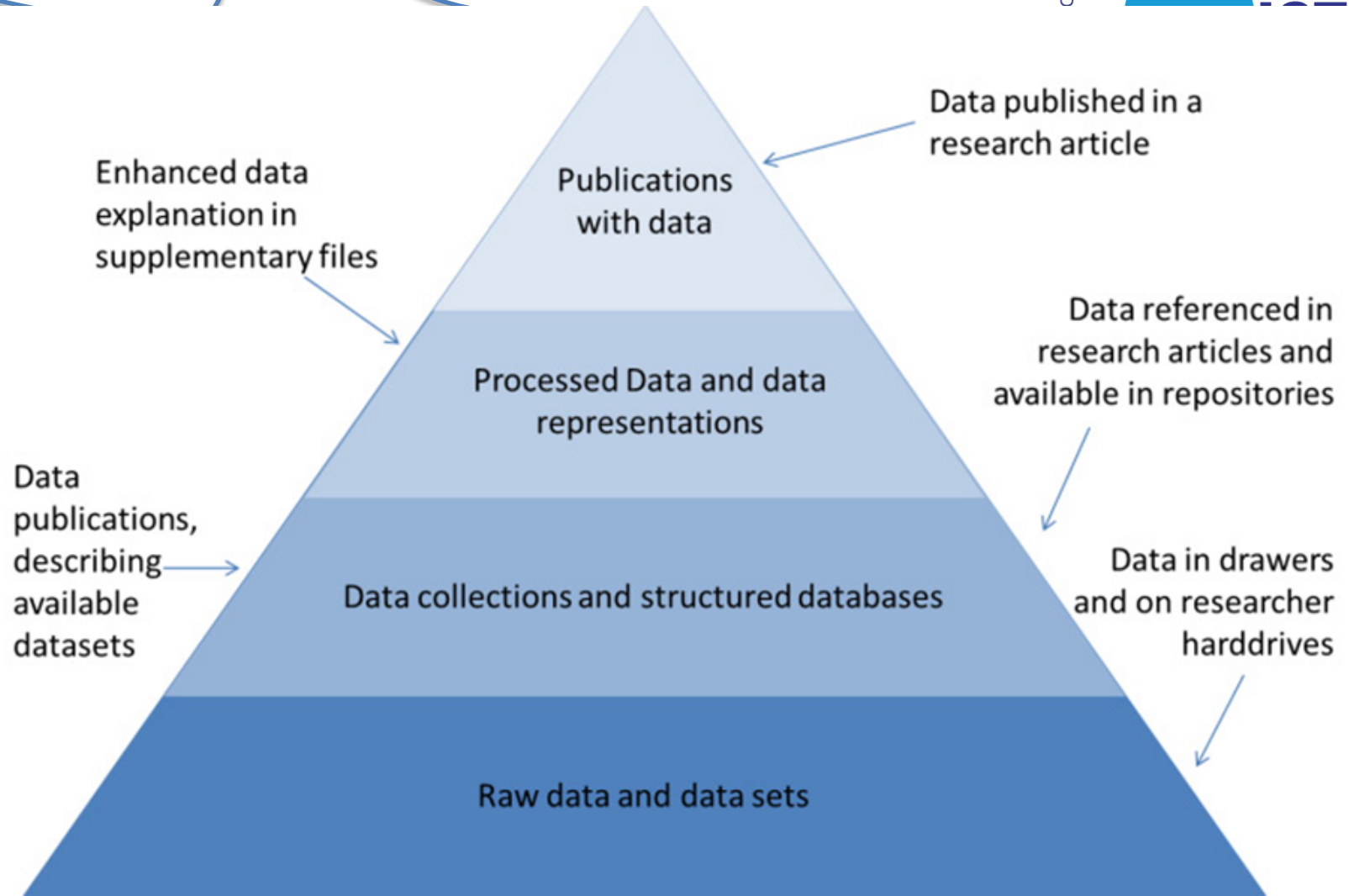
- Space based data limited by bandwidth
- 100 Mbps max (X-band), but
- INTEGRAL (2002): 1.2 Gbyte/day
- Hubble space telescope (1990): 15 Gbyte/day
- Gaia (2013), Euclid (2021): ~50 Gbyte/day
- Ground based: fast increase through fast read-out systems, multiple charge-coupled devices (CCDs)
- 1990s: 1 Mbyte / CCD frame
- LSST (2021): 3 GByte / exposure (15s)



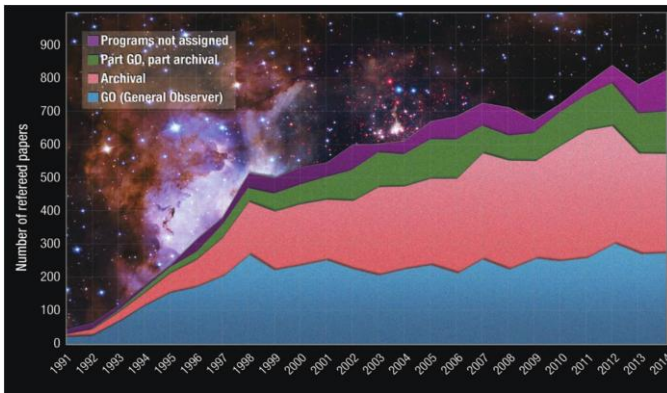




ting



# Data Archive: Why it's important



HST Newsletter: “At the present time, approximately **half of the refereed publications** based on Hubble observations are derived purely **from archival data**, and, every year, this number is slightly higher than the number of publications based on new observations. .... the Hubble Archive has become a goldmine for the astronomical community...”

## Archive are:

- Data Management
- Data Curation
- Data Preservation

## Archive are not:

- Data Sharing

- Archives ..

- Archives ( > 50, like VST) .. preservation



- IA2 (main INAF archive)

- LBT, TNG, ..
- Radio Archives
- Ground base

- ASI Science Archive - ASDC

- International projects
  - IVOA
  - Open Universe
  - EOSC
  - Open Science



# Interoperability: Open Access - Open Science



*Open Access and Open Science is one of the MUST of the EU/H2020 funding project policy*

- *The **European Open Science Cloud** (EOSC) pilot project, in which INAF is involved, will support the first phase in the development as described in the EC Communication on European Cloud Initiatives [2016].*
  - *It will establish the governance framework for the EOSC and contribute to the development of European open science policy and best practice;*
  - *It will develop a number of pilots that integrate services and infrastructures to demonstrate interoperability in a number of scientific domains; and*
  - *It will engage with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research*

# Interoperability: IVOA



- “Allow **astronomers** to interrogate multiple data centers in a seamless and transparent way”
- “Give **data centers** a standard framework for publishing and delivering services using their data.”

+

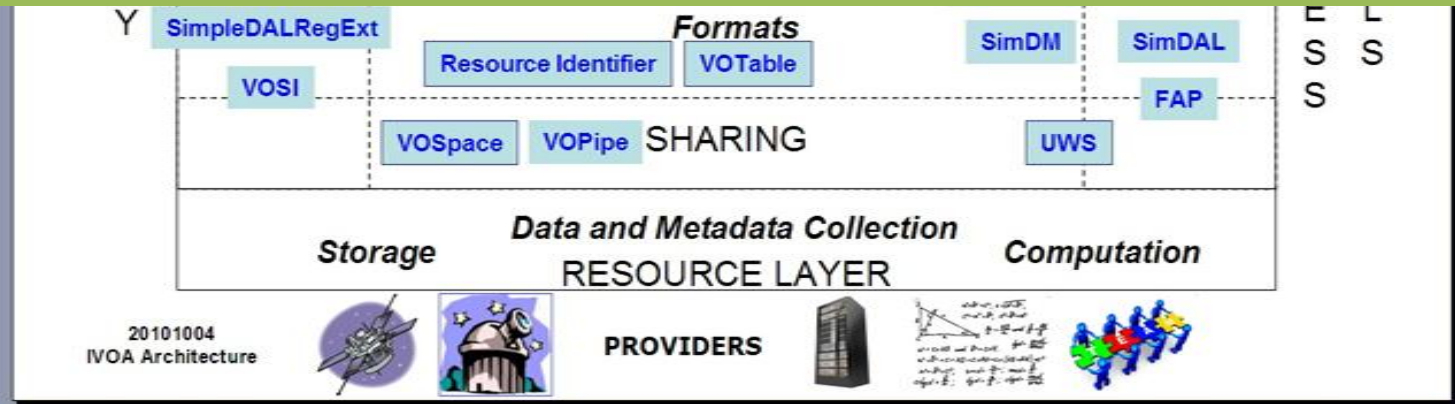
- Give **data centers** a standard framework for integrating interoperable authorization
- Allow **astronomers** to interrogate multiple data centers in a seamless and transparent authenticated way





Each Astro RI provides an archive of datasets in physical units (i.e. reusable); whenever the IVOA/FITS standards are used, data are FAIR

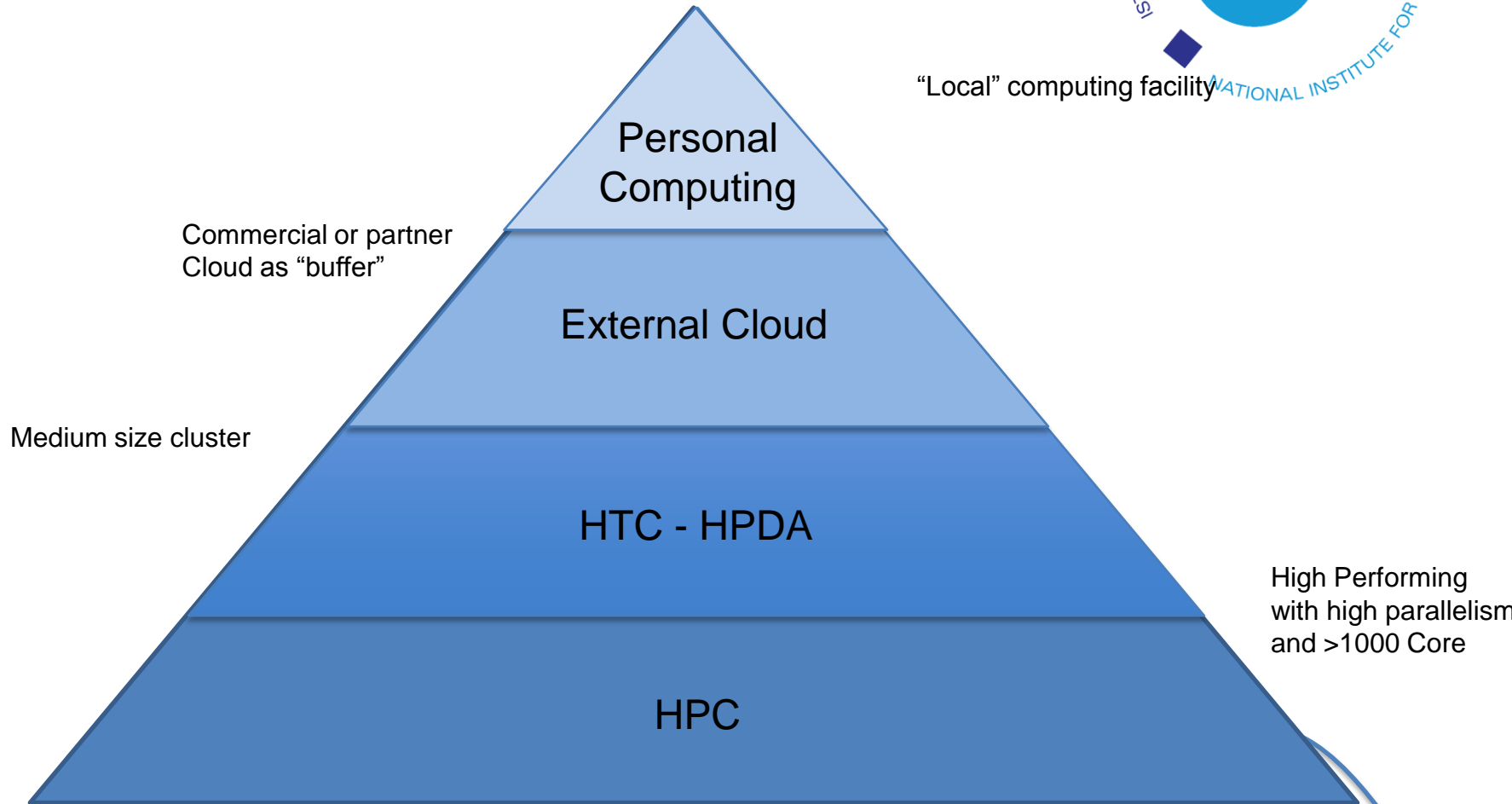
**FAIR: Flexible, Accessible, Interoperability, Reusable**



# The "pyramid" of computing



"Local" computing facility



What's happened in the future in HPC

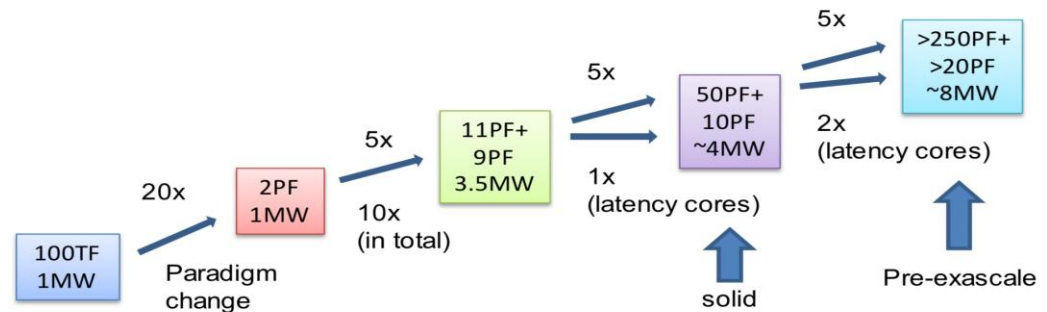


INAF have a MOU with CINECA for using computing infrastructure (50 Million core/hour)

## Cineca “sustainable” roadmap toward exascale

Main evolution, from “classical” CPU to accelerators (GPU, TPU, ...)

Old HPC programming will not more usable with the new architecture



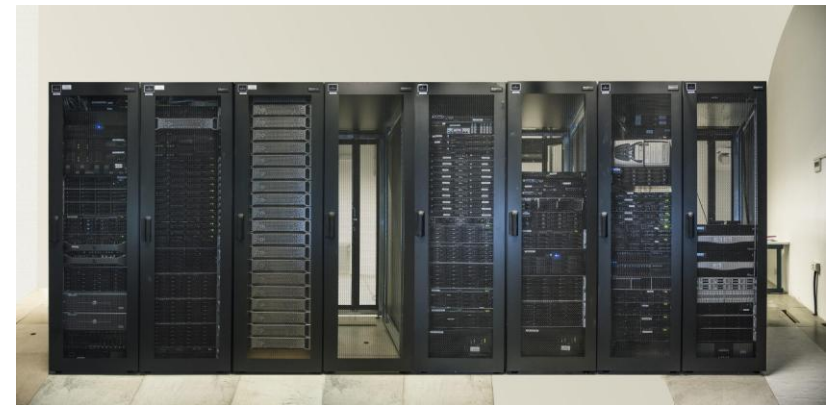
2009	2012/2013	2016/2017	2019/2020	2021/2022
IBM SP6 Power6	Fermi IBM BGQ PowerA2	Marconi Xeon + KNL	Marconi + PPI4HPC + ICEI (PPI-HB)	EuroHPC

# INAF - HTC “Computing Facility”



INAF have some dedicated e-infrastructure for a projects:

- VST
- Planck
- LOFAR
- GAIA
- ...

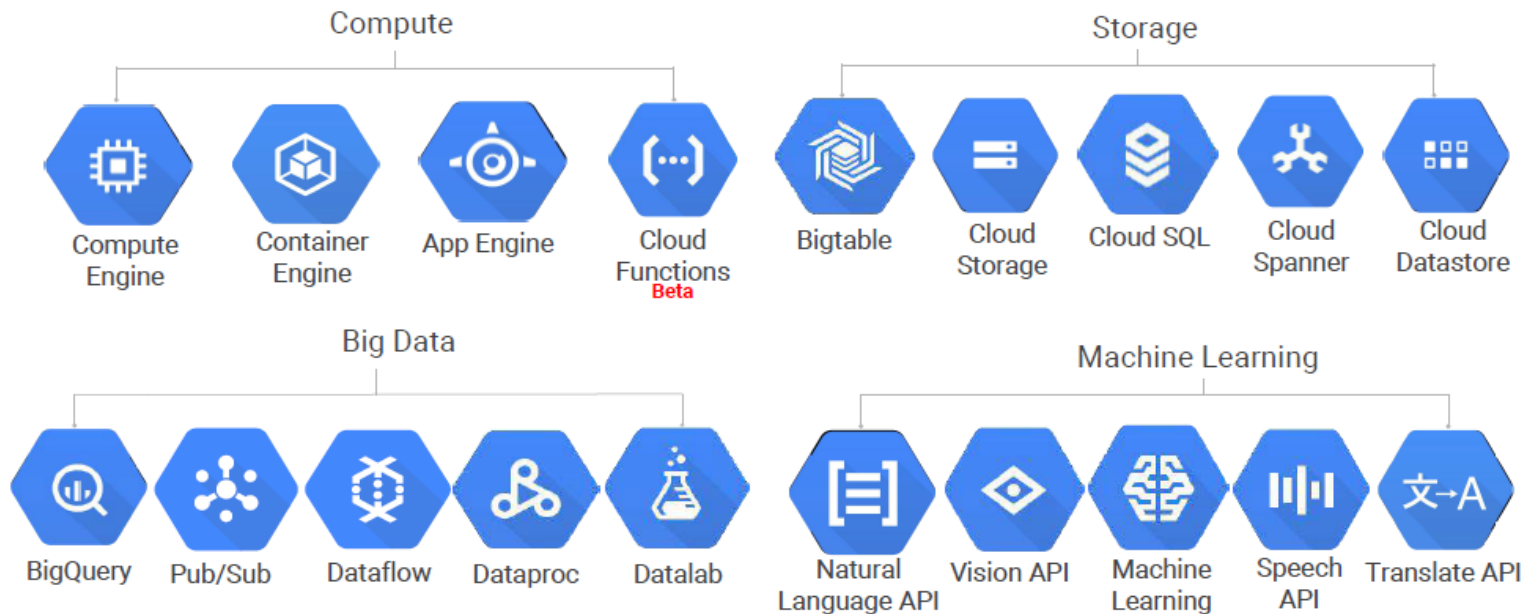


But have also a generic e-infra: CHIPP

Core fisici	800 <sup>1</sup>	200	192 <sup>2</sup>
Mem/Core	6.4 Gby	4 Gby	2.6 GBy
Network interno	Infiniband 56 Gbps	Infiniband 10 Gbps	FastEthern et
Fast Storage	240 Tby	70 Tby	24 Tby <sup>3</sup>
Long Storage	> 16 Tby	> 16 Tby	None
CHIPP (%) Core- ours	> 40%	70%	90%

- Main GCP services

## Google Cloud Platform



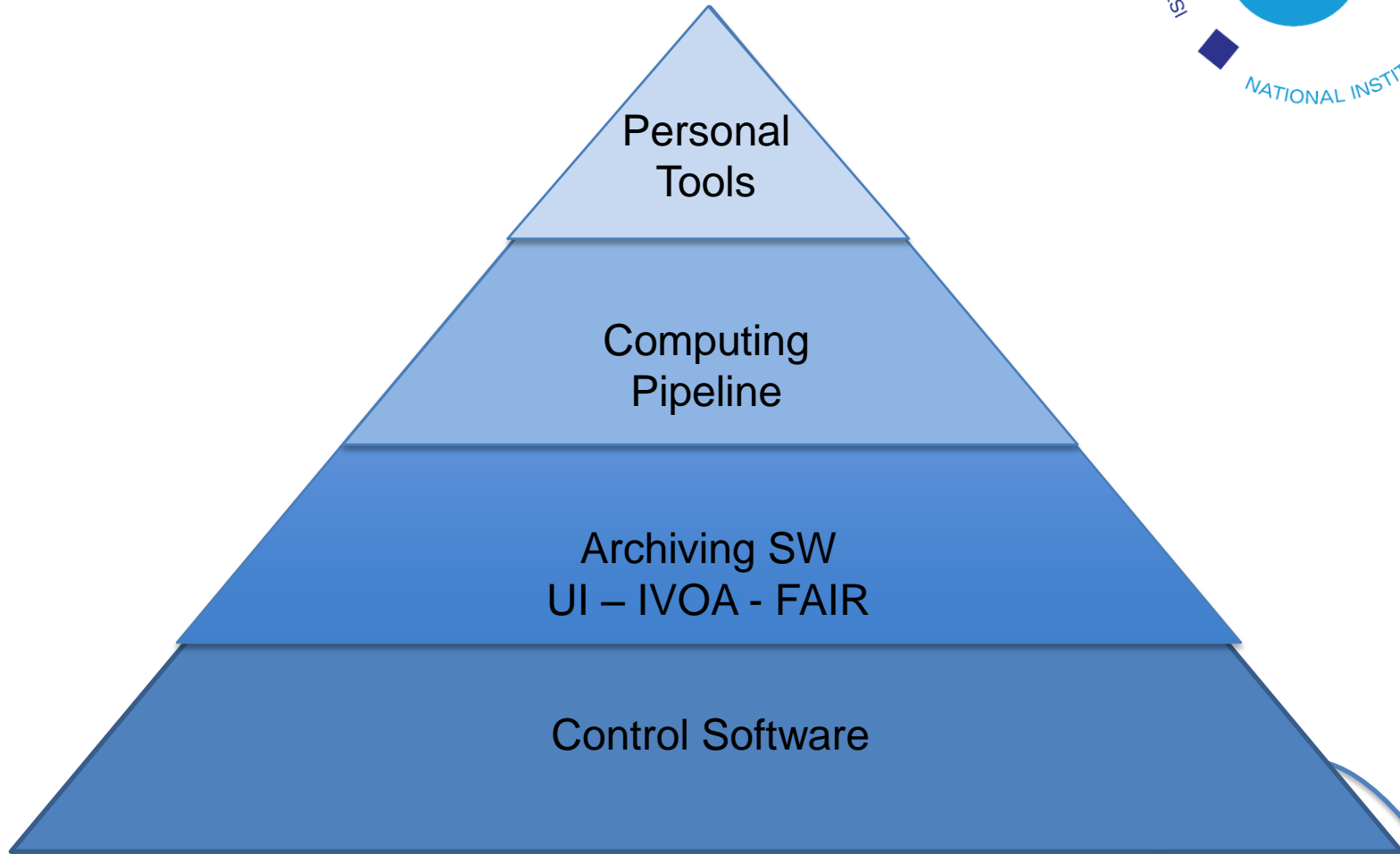
## 9 Proof of concepts (PoCs) proposed by our scientists

1. HTC computing and software containerisation for DIAMONDS (M. Landoni)
2. HTC computing for DIAMONDS with Kubernetes (M. Landoni)
3. GPU computing for Adaptive Optics (M. Landoni in collaboration with OA Arcetri)
4. HPC computing for GADGET (G. Taffoni)
5. HPC computing for Exoclimates (G. Taffoni)
6. HPC computing for GAIA GSR Solver (U. Becciani, A. Vecchiato)
7. Computing for ALMA (M. Massardi, A. Giannetti, S. Burkutean)
8. Workflow execution for GIANO@TNG pipeline (A. Bignamini)
9. Euclid LE3 software in the Google Cloud Platform (D. Tavagnacco)



Google Cloud

# The “pyramid” of software



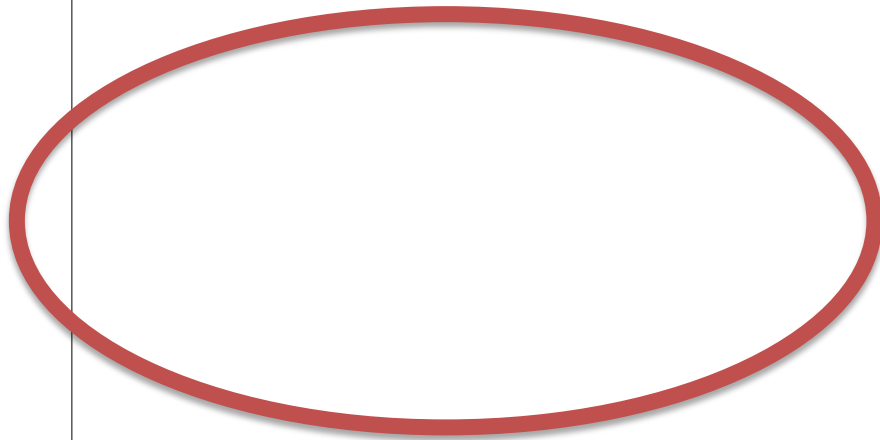
# Intellectual Propriety



- INAF develop more than 100 Package Software, some released to the consortium, some Open, some... to a friends
- Open Source code should be release under Open License:
  - GNU GPL v3
- Or must be released with a DOI - Digital Object Identifiers: (persistent, globally unique, resolvable)
- Can be assigned to **publications, data, software**
- Resolvable by prepending <https://doi.org/>
- Citable (unambiguously)
- As a URL, avoids link rot
- Machine-readable when cited
- Repositories offer (linked) metadata (for humans and machines)

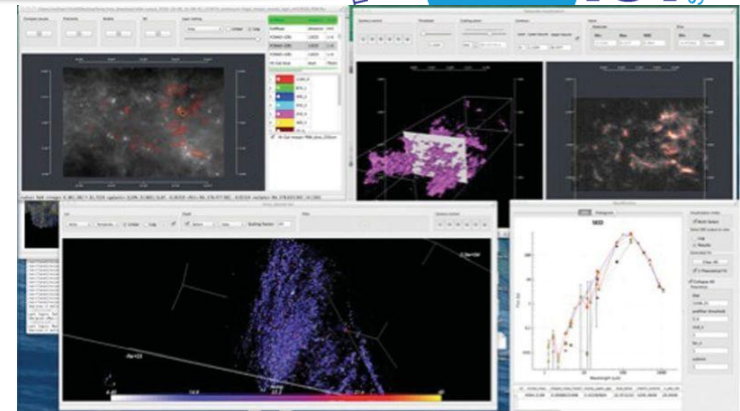
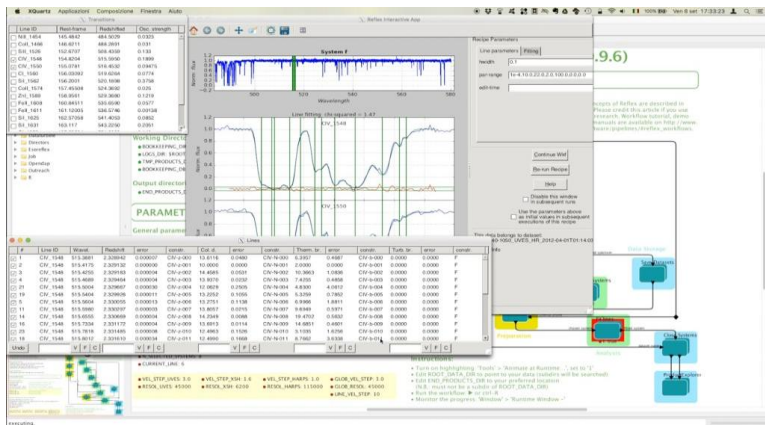


# From Control to Data



Courtesy: J. Monari

## Visual Analytics tools (Science Portal)



## ESPRESSO Data analysis SW

## Data Mining tools (Dame)

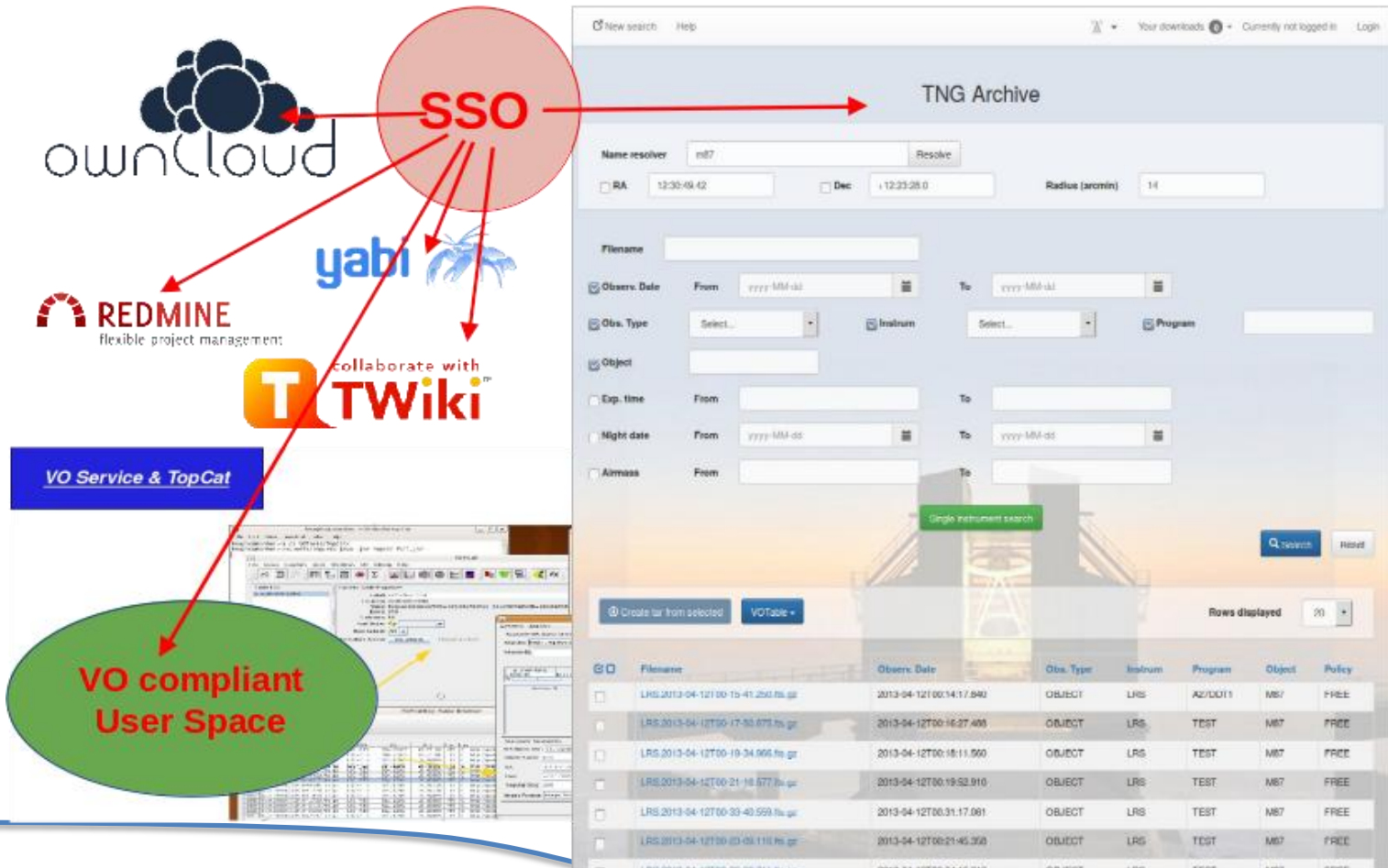
The screenshot shows the Dame Data Mining & Exploration web portal. The page features a navigation menu with items like 'Home', 'Software Services', 'Science Cases', 'Publications', 'Education & Lectures', and 'Who's who'. A 'Welcome on DAME official web portal!' message is displayed, along with a 'Latest News and Events' section listing dates such as 'March 15-17, 2016' and 'Apr 4-6, 2016'.

# SSO access

## Single Sign On



LDAP INAF  $\leftrightarrow$  IDEM  $\leftrightarrow$  EduGain



Thanks for your attention!



n

## Vacancy note:



- ***“Development of support software for archiving of astronomical data managed by the Italian Center for Astronomical Archives (IA2)”***
  - <http://www.oats.inaf.it/index.php/it/2014-09-12-12-50-59.html>
  - <https://www.ia2.inaf.it>
  - E-mail: [cristina.Knapic@inaf.it](mailto:cristina.Knapic@inaf.it)
- ***Coming soon: “Web interface developer/s”***
  - <https://www.ia2.inaf.it>
  - E-mail: [cristina.Knapic@inaf.it](mailto:cristina.Knapic@inaf.it)
- ***“Development of the Observation Preparation Software for the VLT MOONS spectrograph”***
  - [http://www.iasf-milano.inaf.it/detail\\_Job.cgi?51](http://www.iasf-milano.inaf.it/detail_Job.cgi?51)
  - <https://vltmoons.org/>
  - E-mail: [Bianca.qarilli@inaf.it](mailto:Bianca.qarilli@inaf.it)