Big Observatories - Big Data

the approach to astrophysical data and open science in the CTA, E-ELT and SKA era

Filippo Maria Zerbi

(INAF Science Director)

INAF

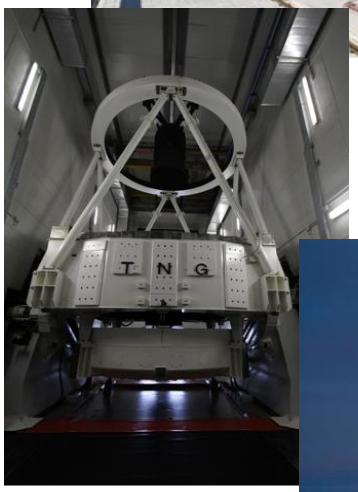
- INAF is the Italian national public research institute aimed to:
  - carry out, promote and exploit the scientific and technological research in the fields of astronomy and Astrophysics;
  - promote and favor the transfer of internally developed technologies to industry;
  - pursuing excellence at international level.

INAF and Space Astrophysics

- INAF collaborates with Space Agencies (mainly ASI but also ESA, NASA, JAXA and others) in the development, construction, operation and data exploitation of space missions for
  - Astrophysics and Observation of the universe
  - Exploration of the Solar System
- INAF develops jointly with space agencies key technologies to enable and/or improve scientific payloads for space missions.
- INAF contribute to Space Astrophysics Data Handling (e.g. in ASI-SSDC with INFN).

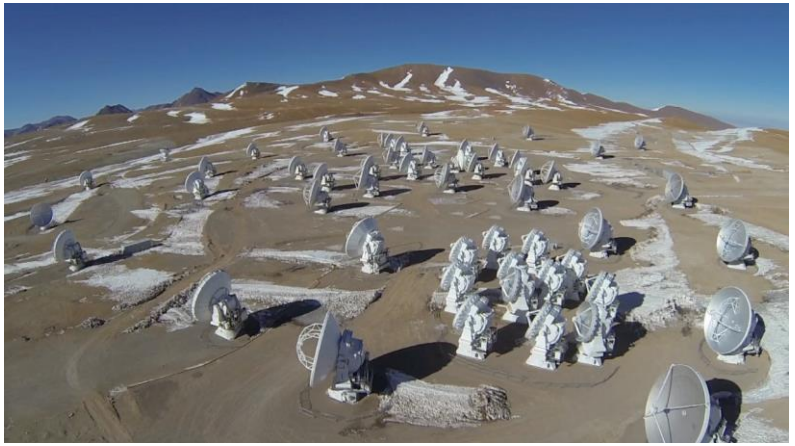
Ground Based (Optical)

- Italy through INAF is part of ESO
  - Access to ESO Facilities (including VST)
  - Contribute to the ESO Telescopes set of instrumentation (including data reduction pipelines)
  - Active part of the ESO users' community (fresh and archived data)
- INAF Operates Proprietary (whole or per share) facilities
  - Telescopio Nazionale Galielo (TNG)
  - Large Binocular Telescope (LBT)



Ground Based Radio

- VLBI
  - INAF Operates single dish antennas in Medicina and Noto and the Sardinia Radio Telescope SRT (in collaboration with ASI) near Cagliari.
  - These antennas are part of the VLBI network
- ALMA
  - Through ESO INAF has access to ALMA in Chajnantor.
  - An Alma Regional Center (ARC) is located in Bologna

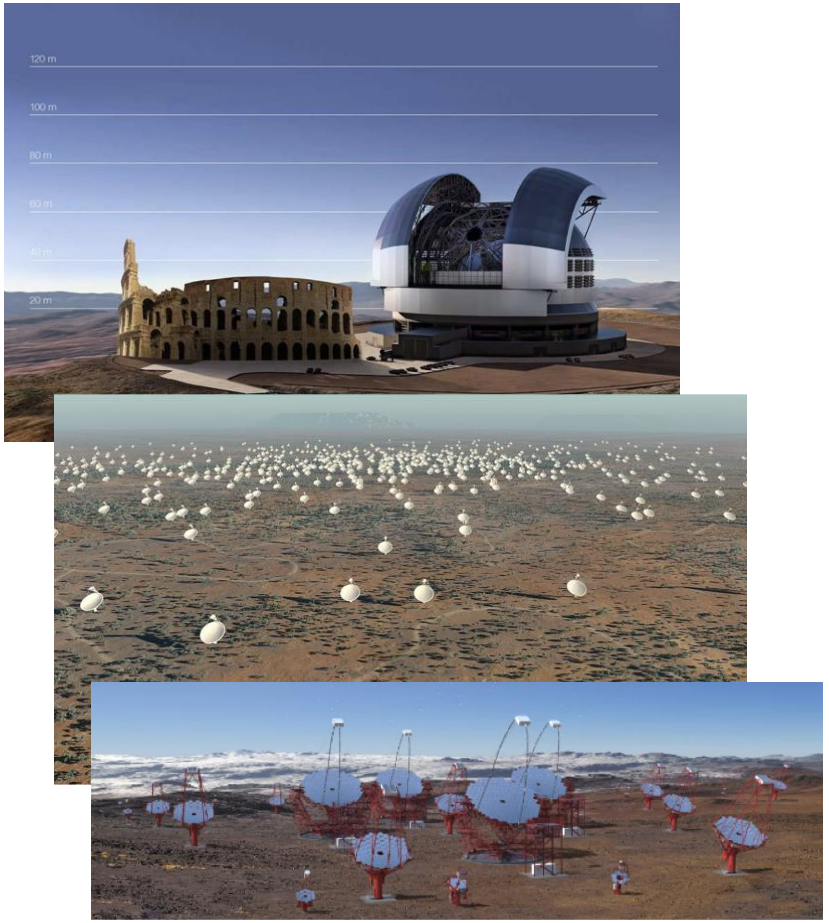


Ground Based Gamma

- INAF is partner in high Energy ground based experiments and observatories:
  - MAGIC Observatory in la Palma.
  - Pierre Auger Observatory in Argentina

Future Infrastructures

- INAF is strongly committed in the definition and construction of the major European future generation ground based research infrastructures:
  - E-ELT European Extremely Large Telescope
  - SKA Square Kilometer Array
  - CTA Cherenkov Telescope Array
- These facilities present intriguing ICT challenges, however they are not necessarily “big data” producers.



Astrophysical data in INAF

- The INAF Community (as and with other communities of Astrophysicists) is at the same time a producer and a user of Astronomical Data.
- Therefore we seek and foster
  - Data Preservation
  - Data (Open) Accessibility
  - Data Interoperability
- In optimal synergy and agreement with OU Proposal
  - (1) Promoting the robust provision and permanent preservation of science-ready data;
  - (2) Advancing calibration quality and statistical integrity;
  - (3) Fostering the development of new centralized services, both large and small, to exploit the interconnectedness of the modern Internet through new web-ready data;

Key Questions

- What should we to preserve ?
- For how long should it be preserved ?
- Why should we make it available ?
- To whom should we make it available ?

What should we preserve ?

- The concept of Science Ready Data in Astrophysics is debatable
- Many (not all) astronomical data comes from the equivalent of trawling
- You select in the net (pipeline) the fish you want to eat but that doesn't mean other fishes are suited to somebody else soup !
- Science-ready "pipelined" data could restrict the potential use of other users or future users

Why make data available ?

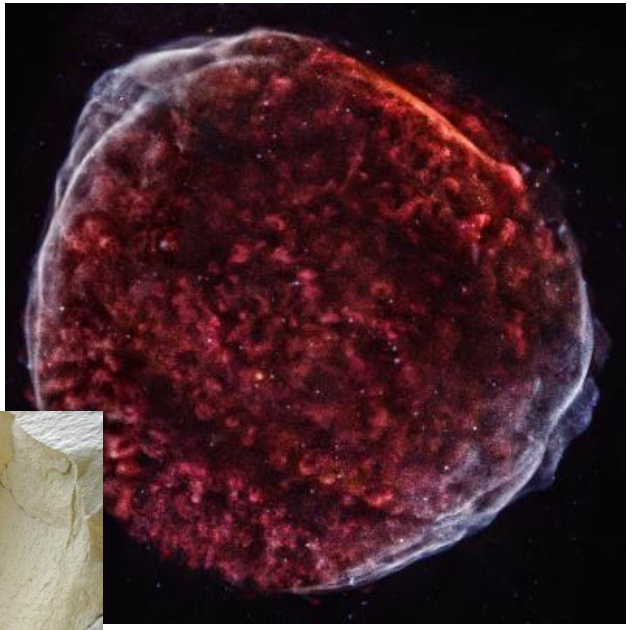
- The net almost always catches more fishes than the fisherman can or wishes to eat.
- There are however a lot of less-fortunate fishermen that could enjoy the meal if invited.
- Data (like fishes) stink after some time. There is no advantage to keep them reserved if the owner cannot use them.

Available to whom ?

- Tuna-fish hunters will not know what to do with a cod-fish by chance in the net. Restricted access does not maximize science return.
- Moreover something strange or serendipitous may have fallen in the net one would not even consider as a fish.
- Possibly a specimen of a new specie that will be classified only years later.
- I order to maximize the outcome the catch should be made available to ALL and preserved for the longest possible time.

For how long ?

- Astronomers are interested in preserving their data in time scales longer than any possible reference of the digital era.
- Ideally in a scale of millennia

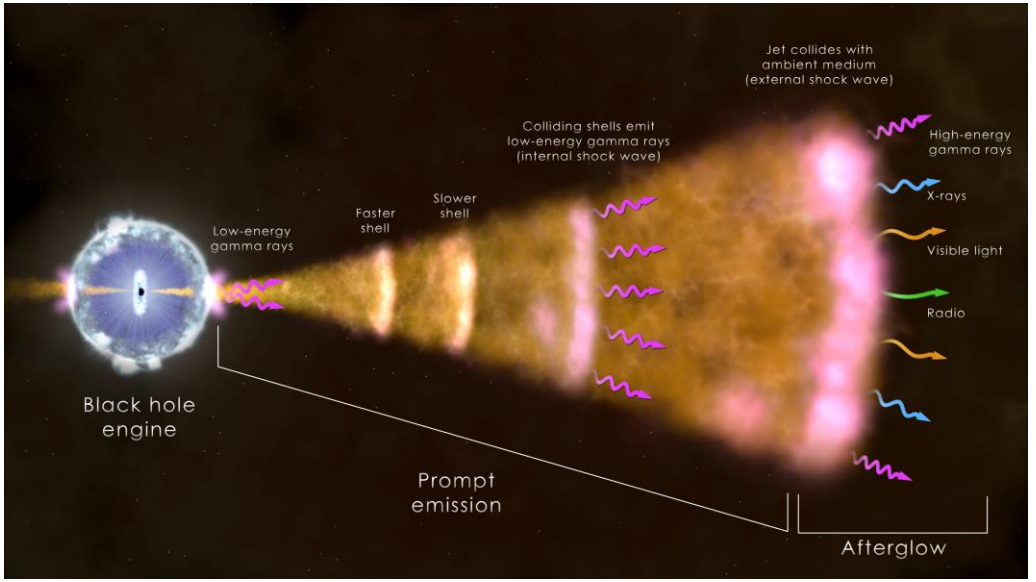


Chandra image of SN 1006

Anasazi recording of SN 1054

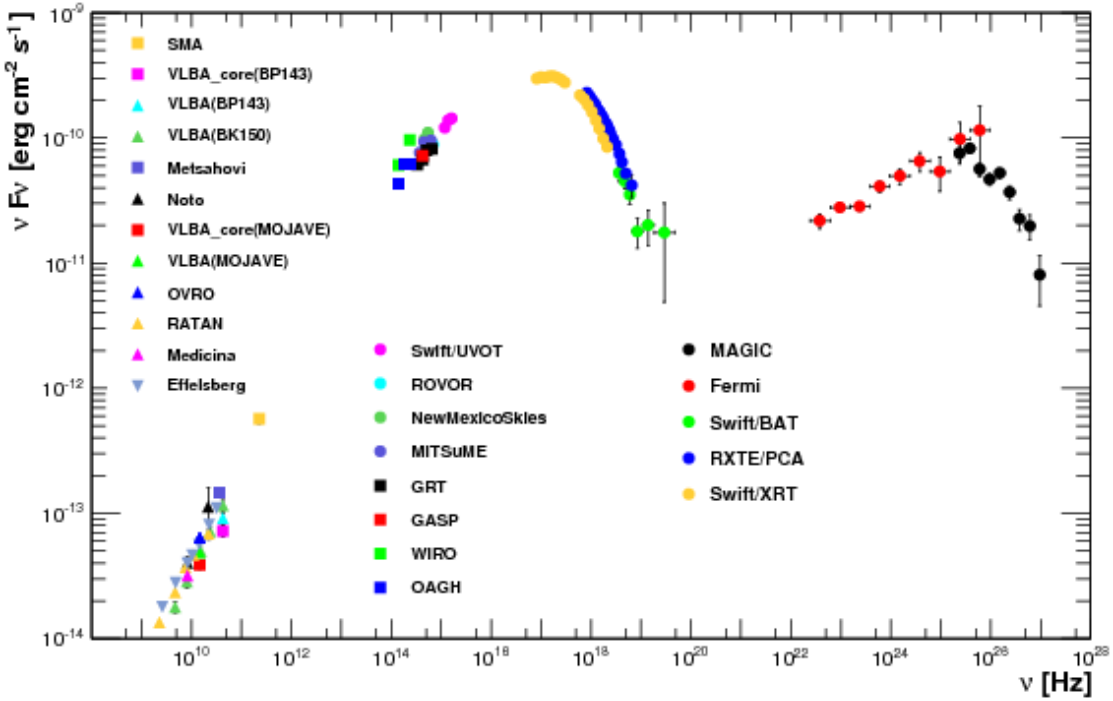
Data Inter-operability

- Astrophysics is a Natural Science
  - We observe nature without manipulating it
  - Nature often “surprises” us
  - A science for which “serendipity” is a central concept.
- “Correlation” is essential
  - Astrophysics is multi-wavelength
  - Recently Multi-messanger (Nutrinos, GW)



Inter-operable Data-bases

- Data collected with different observing equipment are inter-operable if
  - The position of the source is known with known precision (down to milli-arcsec)
  - The time of the emission is know with great precision (milliseconds)



What are we (already) doing

- INAF is involved in OU related initiaties.
  - ASTERICS/OBELICS
    - WP4/DADI Standardization and Interoperability
    - OBELICS/WP3 New Standard on Big Data
    - WP2/DECS Education and Citizen Science
  - EOSC-pilot
    - Data Inter-operability
  - INDIGO-DataCloud
    - Cloud data archiving
  - Others…

Conclusions

- A definition of “Science-ready” that preserves serendipity and post-analysis.
- Easy access and operability of Data Bases
- DB open to everybody in the shortest possible time
- Data preserved for the longest possible time
- Data archived with inter-operability in mind.

- THANKS !