ASTRONET: Europe strategy for data management

D. Mourard, CNRS/INSU – FR , ASTRONET Coordinator

ESO/ESA Workshop

Science Operations 2015: Science Data Management

November 25th, 2015 - Garching

Rationale

The objective of this workshop is to present and discuss the **various approaches to science data management** in **spacecraft** missions and **ground-based** facilities for Astronomy..

The European Southern Observatory (ESO) and the European Space Agency (ESA) generate a significant fraction of science data from ground and space for the European astronomical community (and beyond). These data feed both **direct PI use** and use from the community at large through **powerful science archives**.

- Presentation of what is and what does ASTRONET
- The Science Vision and the Infrastructure Roadmap
- Focus on data issues with a few examples
- Conclusions

ASTRONET (FP6 & FP7)

www.astronet-eu.org

A consortium of funding agencies for European astronomy focusing on strategic planning and coordination

Aim

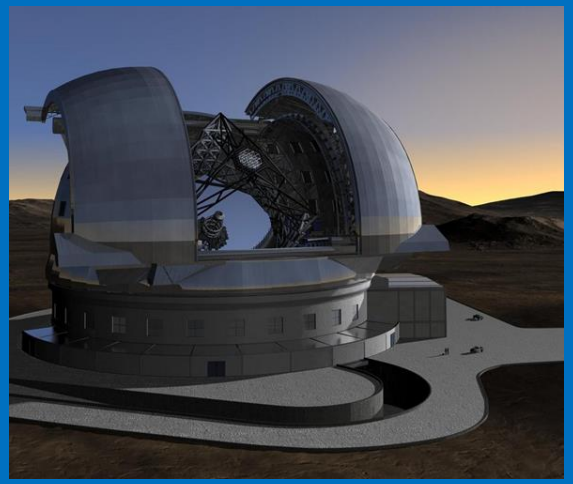
To develop the overall context that can assist national funding agencies and European organisations in taking science-based, rational, and coordinated decisions for the long-term benefit and cost-effectiveness of European astronomy.

ASTRONET is independent and involves all of Europe

Yes, a large part of Europe indeed

Implementation and Realization

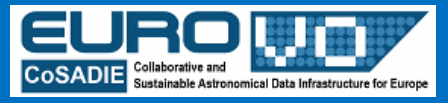
ASTRONET initiates, pushes and enables …



… focusing on efforts underlying the big facilities

Supporting efforts

Preparing for a sustainable Virtual Observatory in Europe



ETFLA: European Task Force Laboratory Astrophysics

Astrophysical Software Laboratory Working Group

ERTRC: European Radio Telescope Review Committee

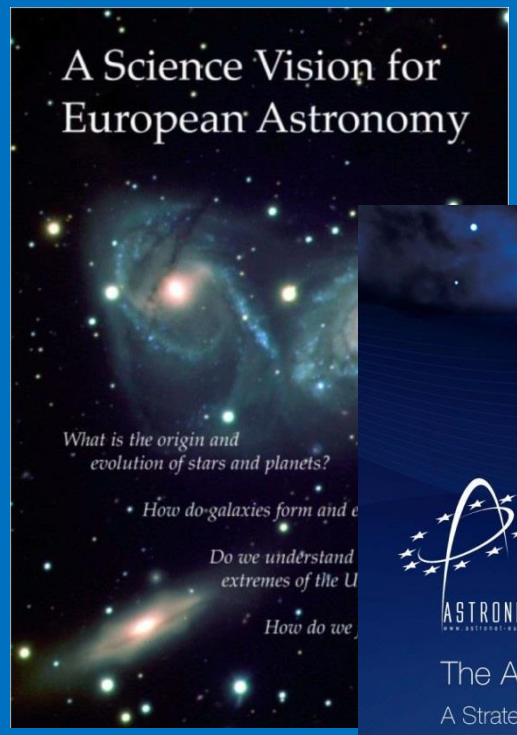
European Optical/IR Telescopes

ASTERICS H2020, cluster of infrastructures

Education and public outreach report

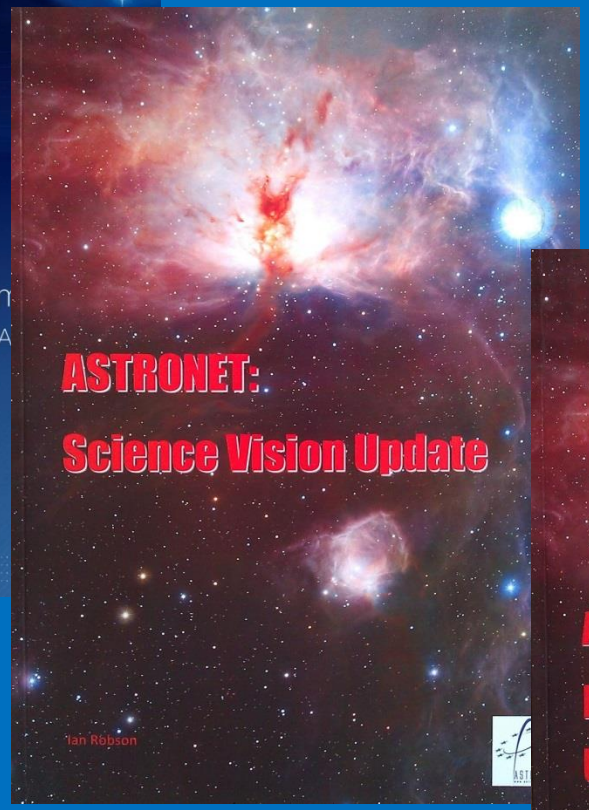
Key achievement

Science Vision and Infrastructure Roadmap



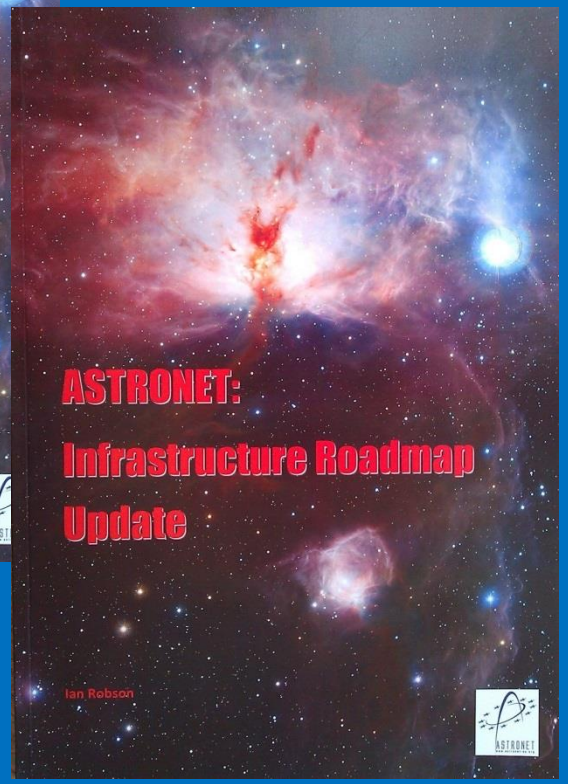
2007

2008



Feb 2013

Jul 2014



Science Vision Update

Panel A: Do we understand the extremes of the Universe?

Panel B: How do galaxies form an evolve?

Panel C: What is the origin and evolution of stars and planets?

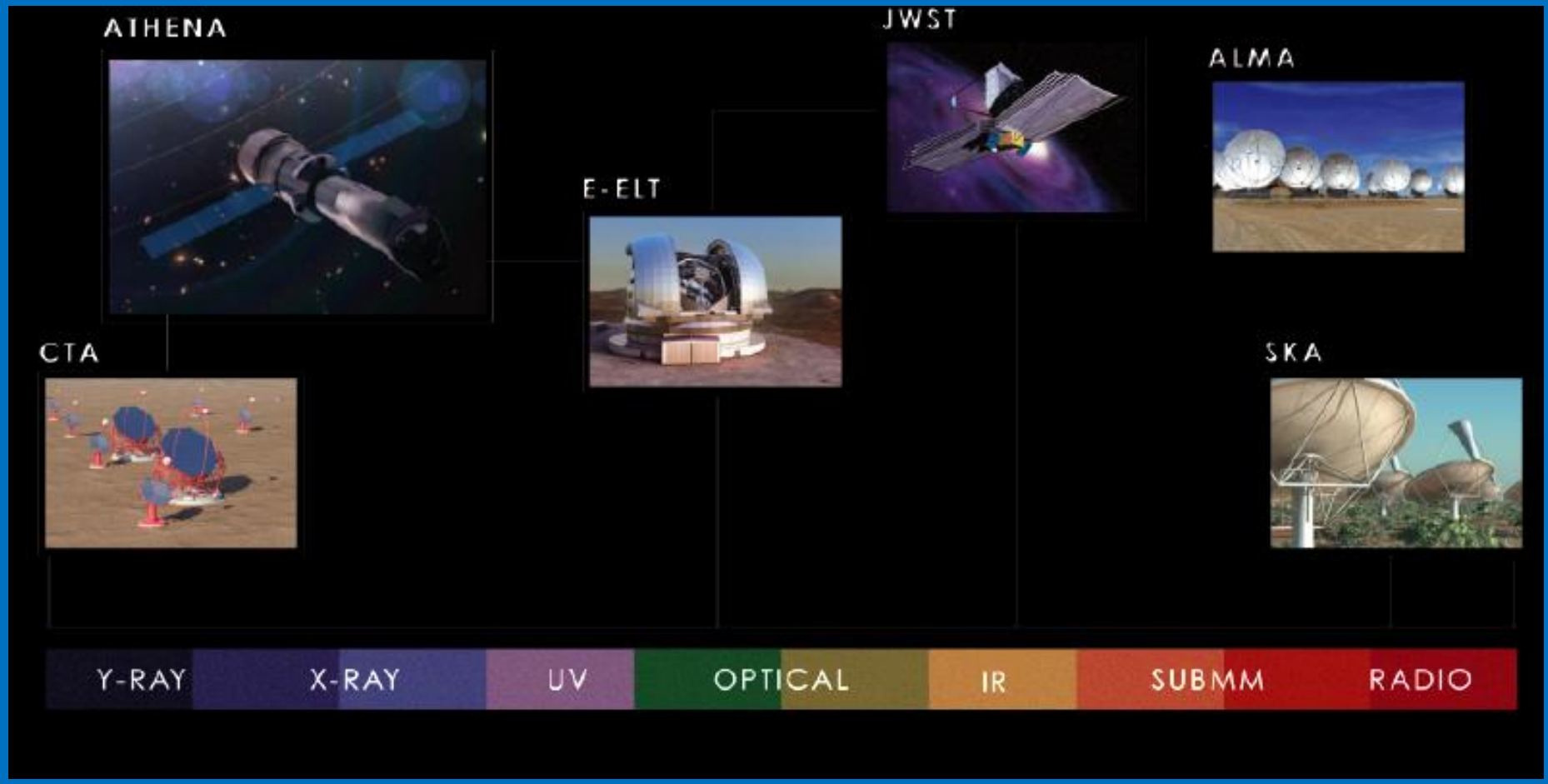
Panel D: How do we fit in?

- Good progress and significant new discoveries and advances: Exoplanets, AGN feedback, Milky Way filaments, solar weather, astrobiology
- The original questions still remain valid; probably some changes for the next edition
- Discoveries driven by host of new facilities from the ground and space

ASTRONET high-priority roadmap for new infrastructures

High-energy, Astrop & GW	UVOIR	Solar, Solar System & Lab
ASTRO-H (2016)	**ALMA (2011)**	**ExoMars (2016)**
eROSITA (2017)	**JWST (2018)**	**Solar Orbiter (2018)**
Athena (2028)	**WEAVE (2018), MOONS (2020), 4MOST (2020)**	**JUICE (2022)**
GW L3 (**2034**)	**EUCLID (2020)**	EST
CTA	**CHEOPS (2023)**	
KM3Net	**E-ELT (2024)**	
	PLATO (2024)	
	SKA Phase 1	

The new observatories in 2020-2030



Where do we go? The era of Big-Science astronomy?

ESA Cosmic Vision

ESO Long Term Plan

New ESFRI infrastructures:

EISCAT-3D, CTA, SKA, KM3NeT

National roadmaps completing the global landscape

Main challenges for ASTRONET and the agencies:

- How do we coordinate the deployment of these facilities and manage the flood of new data efficiently?
- How do we strike the right balance between new and existing facilities?
- How do we train a new generation of leading scientists for this new era?

Remarks at the global level

Growth of the multimessenger/multiwavelength approaches

Growth of the trans disciplinary aspects

- With particle physics
- With fundamental physics
- With chemistry
- With biology

Transverse activities

- Theory, simulations, high performance computing
- **Big data management, archives and science through the VO**
- Laboratory astrophysics

Coordination, education

Lots of existing "added value" scientific synergies, outstanding outcome

- Cluster evolution up to z~1 (Planck + XMM-Newton)
- Galaxy/AGN evolution (HST + Opt/NIR ground + X-ray) in selected cosmological fields.
- AGN feedback through winds & outflows (XMM-Newton + Herschel/ALMA)
- GRBs (Swift/Agile + everything)
- Molecular astrophysics (ISO/Spitzer/Herschel + ALMA + MIR instruments on ground telescopes + Rosetta)

And much more…

Today

- Complex molecules (ALMA, Rosetta)
- Planetary systems:
  - SS in situ missions
  - Observations of SS bodies
  - FIR/submm observations of distant planetary systems

In the future

- Multi-messenger observations
  - CR & gamma-rays
  - GW signals, followed up by multi-wavelength observations
- SgrA* (EHT, VLTI/Gravity)
- Chemical evolution

Panel A - High Energy Astrophysics, Astroparticle Astrophysics and Gravitational Waves

- ESA and national agencies need to plan for the retention of key skills and key teams for the **long lead-time missions** of Cosmic Vision.
- Strengthen **multiwavelength** collaborations through dedicated programmes and grants.
- Continuing R&D technological research activities remains of paramount importance to maintain European leadership in the field of high-energy astrophysics.
- In view of the excellent health of the XMM-Newton and INTEGRAL missions, this panel feels confident to strongly endorse, yet again, their continuation. Moreover, this panel welcomes the outcome of the recent NASA Senior Review, which has approved the continuation of SWIFT and Fermi operations for between 2 and 4 years.

PANEL B Ultraviolet, optical, infrared and Radio/mm Astronomy

- ESA, the EU and national agencies should address the potential for a more coherent funding arrangement for the **exploitation of scientific data** from space missions.
- it is important that the determination of Gaia's precise position from ground-based observations should be secured for the total lifetime of the mission.
- it is vital that national agencies ensure that **adequate funding is provided for data analysis** to ensure that Europe is best placed to maximise scientific return from the Gaia mission
- long-term missions usually require considerable study and technical development and it is important that adequate funding needs to be provided by ESA and National Agencies to support the preparatory R&D activities in the future. Areas that require special attention are e.g. the development of large, low-noise detector arrays, and the development of techniques that will allow high precision formation flying.
- the future optimization of the 2-4m class optical/IR telescopes in Europe requires further and ongoing work in order to maximise overall efficiency and cost effectiveness.
- A coherent long-term plan should be established under the auspices of ESO and the European Initiative for Interferometry during the coming two years. It should be built on the realizations of Gravity and MATISSE and prepare the future plans for enhanced high angular resolution capabilities in the ELT era and in complement to exoplanets and stellar physics space missions.
- a coherent long-term plan should be established under the auspices of ASTRONET and RadioNet during the coming two years. It should outline the scientific role of each of the facilities mentioned above in the ALMA era, develop an access strategy beyond the current Trans National Access (TNA) scenario, and it should define the future investments to be made on the basis of the scientific excellence of the projects that can be carried out. This is very urgent as the future funding for some of these facilities is currently under discussion/threat.
- before considering in any systematic manner perceived gaps and technology developments, it seems desirable to consider the creation of such a database, e.g. through ASTRONET. This should cover **developments both for instrumentation and for software**.
- the preparatory studies for new projects should include a verification of an advanced stage of technical readiness (TRL). This will help to reduce the risk of significant cost-overruns during the construction phase.
- it is critically important that these technology developments needed for the future in terms of key parameters (e.g. large-scale detector arrays) and high-tech solutions are explored in close collaboration with industry.

PANEL C: Solar Telescopes, Solar System Missions, Laboratory Studies

- the European Solar Telescope (EST) should be included in the ESFRI Roadmap in the current revision process.

PANEL D: Theory, Computing Facilities and Networks, Virtual Observatory

- the ASTRONET Board needs to determine the status of the **Astrophysical Software Laboratory** in the near-future
- there is a **need for continued investment in dedicated data facilities** across Europe to keep pace with the data increase.

A new era for Astronomy data?

The end product of all of these missions and facilities is **data** but the explosion in the quantity and quality generates:

- new challenges for storage, analysis, and distribution.
- An increased need for an efficient and interoperable access to this wealth of data, observed or computed in simulations, and curated by information services.

The **VO** is the Infrastructure for astronomy that provides the framework for seamless, **unified access to standardized astronomy data services**.

Significant progress since 2008:

- development of the VO,
- the continued take-up of VO techniques in astronomy data centres,
- and the current efforts to establish VO as a sustainable component of the astronomy infrastructure.

Analysis of a few examples from the Roadmap

- **Planck and Herschel**
  - they have both produced a wealth of data before running out of cryogenic coolant.
  - Both missions have already strongly impacted fields like cosmology, galaxy evolution and star formation, and will continue to do so through extensive data analysis that will continue for at least 5-10 years.

- **Gaia**
  - The direct Gaia data will be generated by the Data Processing and Analysis Consortium (DPAC);
  - Some 300 individuals in 15 European countries are involved in the processing, calibration and reduction of the raw Gaia data, before it becomes available for scientific analysis by the whole community.
  - The currently predicted total costs to the national agencies for data reduction and analysis amounts to about €15M/year (190 FTEs/year) over 14 years.

- **EUCLID mission**
  - Technically, the key components of the mission build on a significant heritage from other missions and the technological risk appears generally low …
  - …except for the Ground Segment, where the total amount of data and its processing represent one of the main challenges of the mission.

Additional recommendations

At mm and sub-mm:

- there is still a lot of room for further improvements by installing e.g. more sensitive, wider-bandwidth receivers, bolometric and heterodyne receiver arrays with larger numbers of pixels (like e.g. SCUBA-2), large format arrays based on new technologies (e.g. Kinetic Inductance Detectors–KIDs), and much more powerful spectral backends.
- Also, the software tools for data reduction and analysis need to be developed further, especially in view of a growing user community of non-specialists.

- At radio wavelengths:
  - The LOFAR data requires a dedicated, high performance central computing facility (recently upgraded) because of the unprecedented amount of radio astronomical data to be treated.
  - Key issues for the radio astronomy community will be
    - (i) how to maintain and even strengthen Europe's current position in radio astronomy by new technical developments like those mentioned above,
    - (ii) how to foster even more cooperation between the existing groups in order to optimise the use of existing radio telescopes,
    - (iii) how to attract and involve an even larger fraction of the European astronomical community in radio astronomical research. The necessary tools, including access to the facilities and data reduction and analysis support clearly need improved coordination, and new organizational and funding schemes will most likely have to be invented for this, in addition to the funding needed for Europe's participation in the SKA project.

Large volume of data

- The need to be able to analyze massive spectroscopic data sets is being reinforced with the new ground-based survey instruments of 4MOST, MOONS and WEAVE as well as the current Gaia ESO Survey

- Considering the size of the current multi-dimensional simulation data sets, parallel, non-data local algorithms also for data visualization and analysis are required. It is also foreseeable that in the field of simulations, even running individual simulations may become less and less the task of individual people or groups. The trend may be that joint analysis of a few, very big simulations, is undertaken by the community at large.

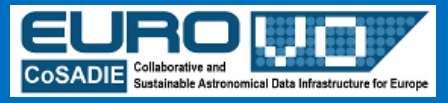
The growth of the open data principles

- The importance of 'open data' and e-Infrastuctures continues to be recognized by high-level European and International bodies.
- the G8 Science Ministers principles for open scientific research data emphasizing that data should be 'easily discoverable, accessible, assessable, intelligible, useable and wherever possible interoperable to specific quality standards.
- EC High-Level Expert Group on Scientific Data: 'Riding the wave – How Europe can gain from the rising tide of scientific data'.
- International Council for Science (ICSU) World Data System, founded in 2008
- The Research Data Alliance (RDA), founded in 2013, is a major initiative for sharing research data across borders and disciplines.

Astronomy is prominent in these organizations and astronomy is recognized as being at the forefront for data sharing and re-use in this fast evolving field.

Supporting efforts

Preparing for a sustainable Virtual Observatory in Europe



ETFLA: European Task Force Laboratory Astrophysics

Astrophysical Software Laboratory Working Group

ERTRC: European Radio Telescope Review Committee

European Optical/IR Telescopes

ASTERICS H2020, cluster of infrastructures

Education and public outreach report

Virtual observatory in Europe

- The relevance of the VO to European data centres is well established since many years:
  - EuroVO-DCA, EuroVO-AIDA
  - Participation to the IVOA
  - EuroVO-CoSADIE has defined the medium-term requirements to establish a sustainable European Virtual Observatory:
    - support to science users and data centres,
    - continued technological development of standards and tools,
    - education network.

    ➔ASTERICS DADI Project (Data Access, Discovery and Interoperability)

- Interoperability of tools via VO protocols includes access to images, catalogues across multiple distributed archives, the ability to perform billion-object cross matches, and complex queries on large tables.

- VO-like projects in several subdisciplines of astronomy, including planetary and heliospheric physics, interest of CTA for implementing their data in the VO. The development of the Virtual Atomic and Molecular Data Centre

Software and computing

- Computing within the VO is now more often described in terms of cloud computing. VO interfaces for asynchronous queries and remote job execution are in place and various implementations of VOSpace provides remote storage.
- One of the new developments is the requirement for supercomputing to operate some of the big observational facilities of this and the next decade.

- Progress continues to be made in the area of Data Networks and Data Grids. The trend for observatories and instruments to produce ever larger datasets from surveys in the petabyte category continues and there is a need for dedicated data facilities distributed throughout Europe to provide access.

  ➔ ASTERICS / OBELICS WP

- Astrophysical Software Laboratory:
  - little progresses have been made in this area; a working group has been established in order to make a census of the codes used by a substantial subgroup of the community.
  - 40 codes have been identified that are currently used by the astrophysical community in Europe for high-performance simulations in various fields of astrophysics. About 8 of these are sufficiently modular in design that they can be seen as major international framework for simulation studies in the field of astrophysics.
  - A plan for a future ASL has been proposed to the ASTRONET board.

ASTERICS: Cluster of RI in Astronomy
EU funded project led by M. Garret (NL)

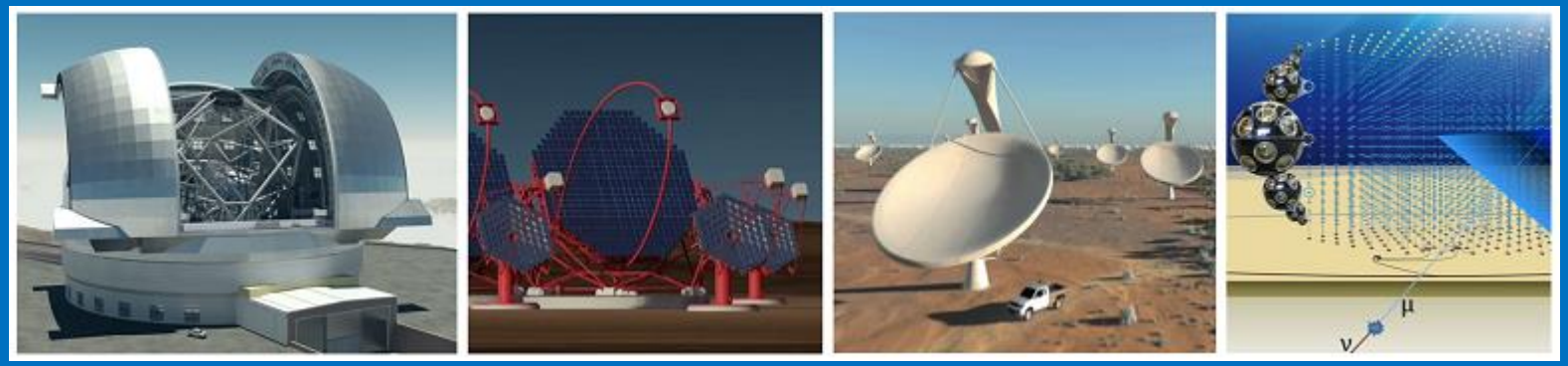
ASTERICS aims to address the cross-cutting synergies and common challenges shared by the various Astronomy ESFRI facilities.

Astronomy, astrophysics and particle astrophysics communities.

The major objectives of ASTERICS are to support and accelerate the implementation of the ESFRI telescopes, to enhance their performance beyond the current state-of-the-art, and to see them interoperate as an integrated, multi-wavelength and multi-messenger facility.

Five work packages:

- ASTERICS Management Support Team
- Connecting Locations of ESFRI Observatories and Partners in Astronomy for Timing and Real-time Alerts
- **Data Access, Discovery & Interoperability**
- Dissemination, Engagement and Citizen Science
- **Observatory E-environments Linked by common ChallengeS**

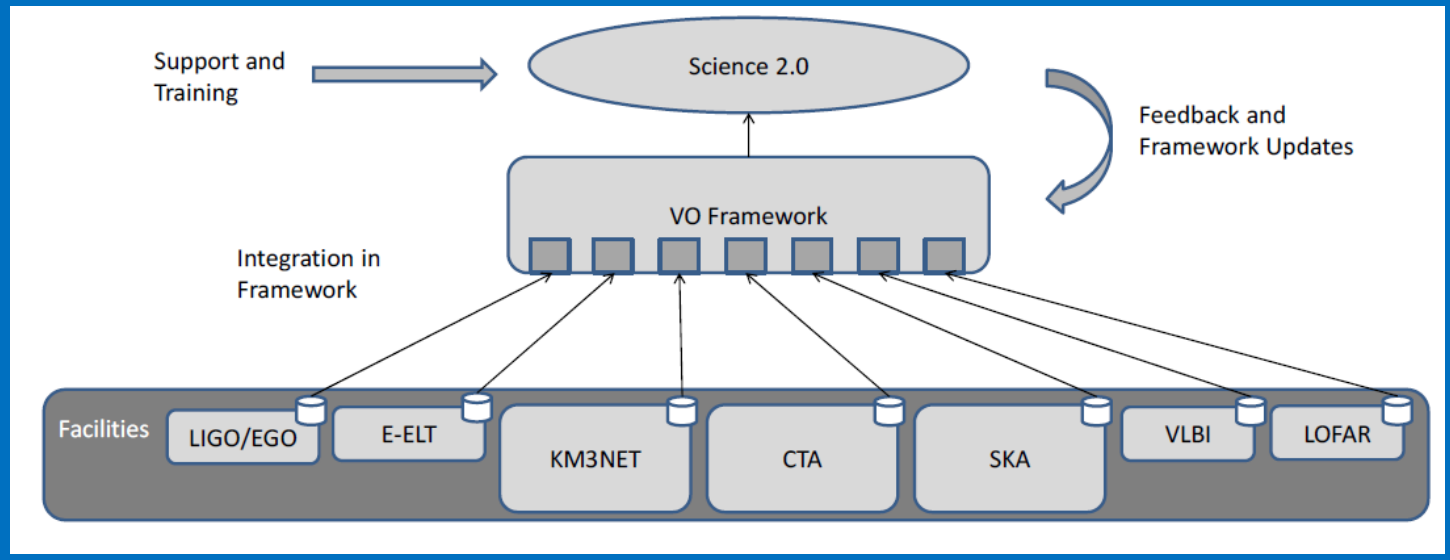


ASTERICS DADI WorkPackage (F. GENOVA)

- Train and support ESFRI project staff in the usage and implementation of the VO framework and tools, and make them active participants in the development of the VO framework definition and updates.

- Train and support the wider astronomical community in scientific use of the framework, in particular for pathfinder data, and gather their requirements and feedback.

- Adapt the VO framework and tools to the ESFRI project needs, and make sure European astronomers remain lead actors in the IVOA, influencing it in the interest of the European infrastructures and the European scientific community.



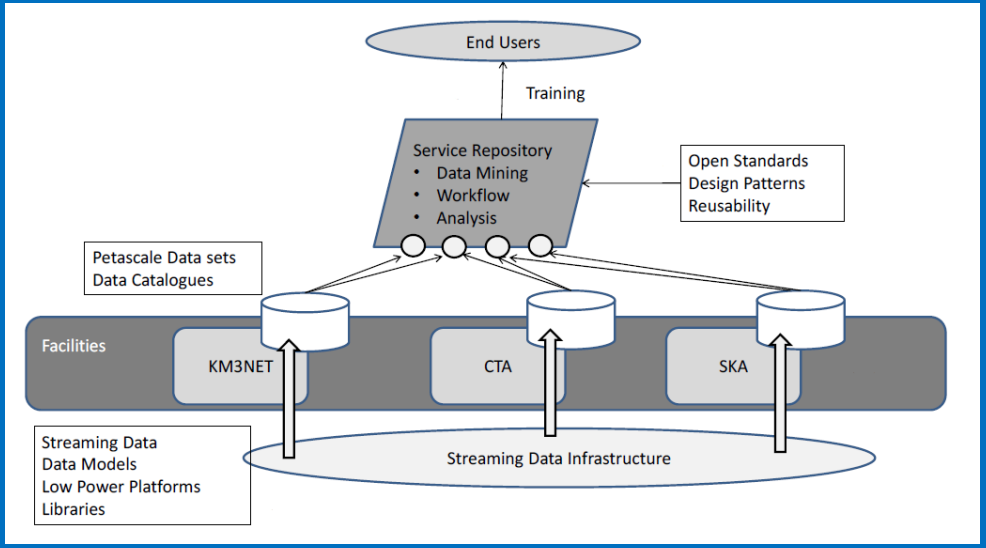
ASTERICS OBELICS WorkPackage (G. Lamanna)

- Train researchers and data scientists in the ASTERICS ESFRI and pathfinder projects to apply state-of-the-art parallel software programming techniques, to adopt big-data software frameworks, to benefit from new processor architectures and e-science infrastructures.

- Maximize software re-use and co-development of technology for the robust and flexible handling of the huge data streams generated by the ASTERICS ESFRI and pathfinder facilities.

- Adapt and optimize extremely large database systems to fulfil the requirements of the ASTERICS ESFRI projects. Cooperation with the ESFRI pathfinders, computing centres, e-infrastructure providers and industry will be organized and managed to fulfil this objective.

- Study and demonstrate data integration across ASTERICS ESFRI and pathfinder projects using data mining tools and statistical analysis techniques on Petascale data sets.



Conclusions and future actions

- Astronet actions are based on the SV+IR recommendations
- new ASTRONET is under construction

- Networking is important for ASTRONET
  - Asterics forum
  - EWASS meetings
  - …
  - Recommendations from this workshop are also important to shape and adapt our future working plan