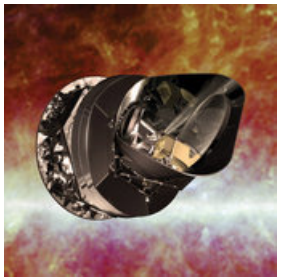
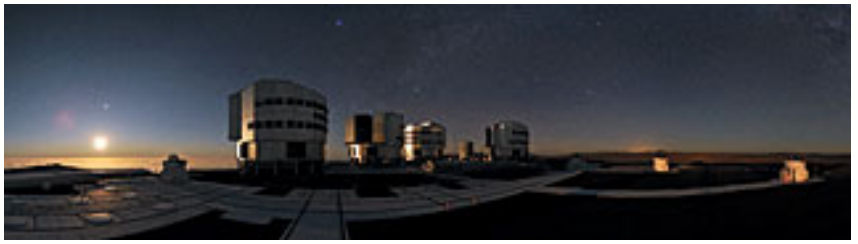
How astronomy shares and reuses scientific data

Françoise Genova

CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG

Astronomy research infrastructures





And data!

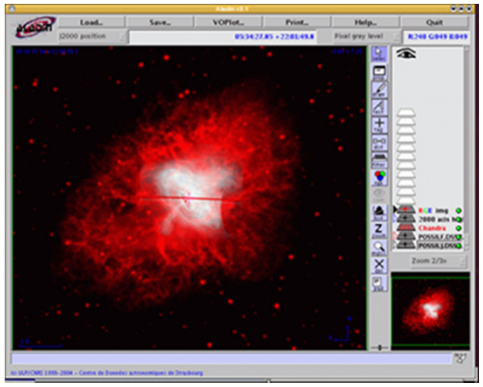
Astronomical data

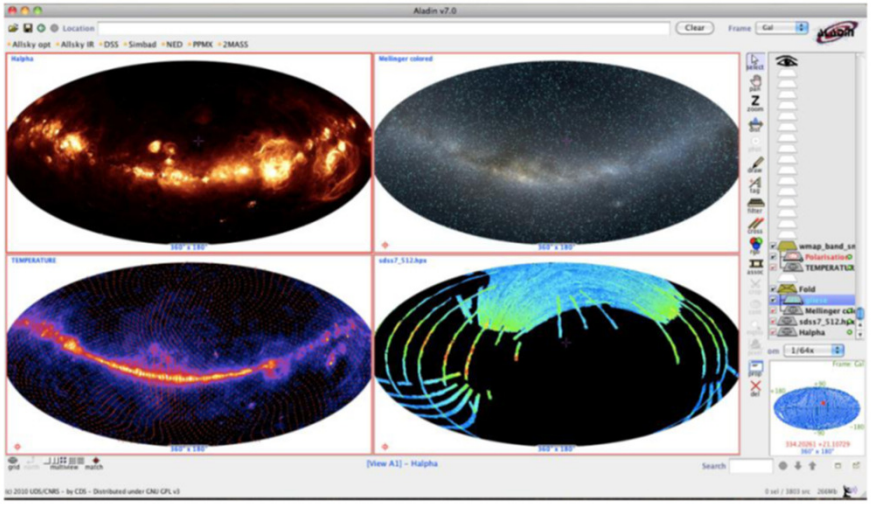
- Observations from ground- and space based telescopes (in general competitive calls for proposals)
- Sky surveys (homogeneous data set with up to billions of objects, measurements, images, spectra)
- Modelling results
- Data from publications
- Value-added data bases, which gather homogenized information in particular from publications

  e.g. SIMBAD, names and papers where the object is cited:

  8 000 000 objects, 22 300 000 object names, 380 000 references, begun ~1970

Why sharing data?





At the core of astronomy scientific needs !

- **Multi-wavelengths, multi-technique astronomy**
- **Time variability**
- **Comparison of theoretical models with observations**
- **Etc.**

Optimize the scientific return of the large infrastructures

Data is available and used

- **Change of paradigm done: astronomers use remote distributed data in their everyday work**
- **More papers from data retrieved from archives than from original observations (HST, …)**
- **More than 800.000 queries/day on the CDS services in 2014 – only one of the components of the astronomical data network**

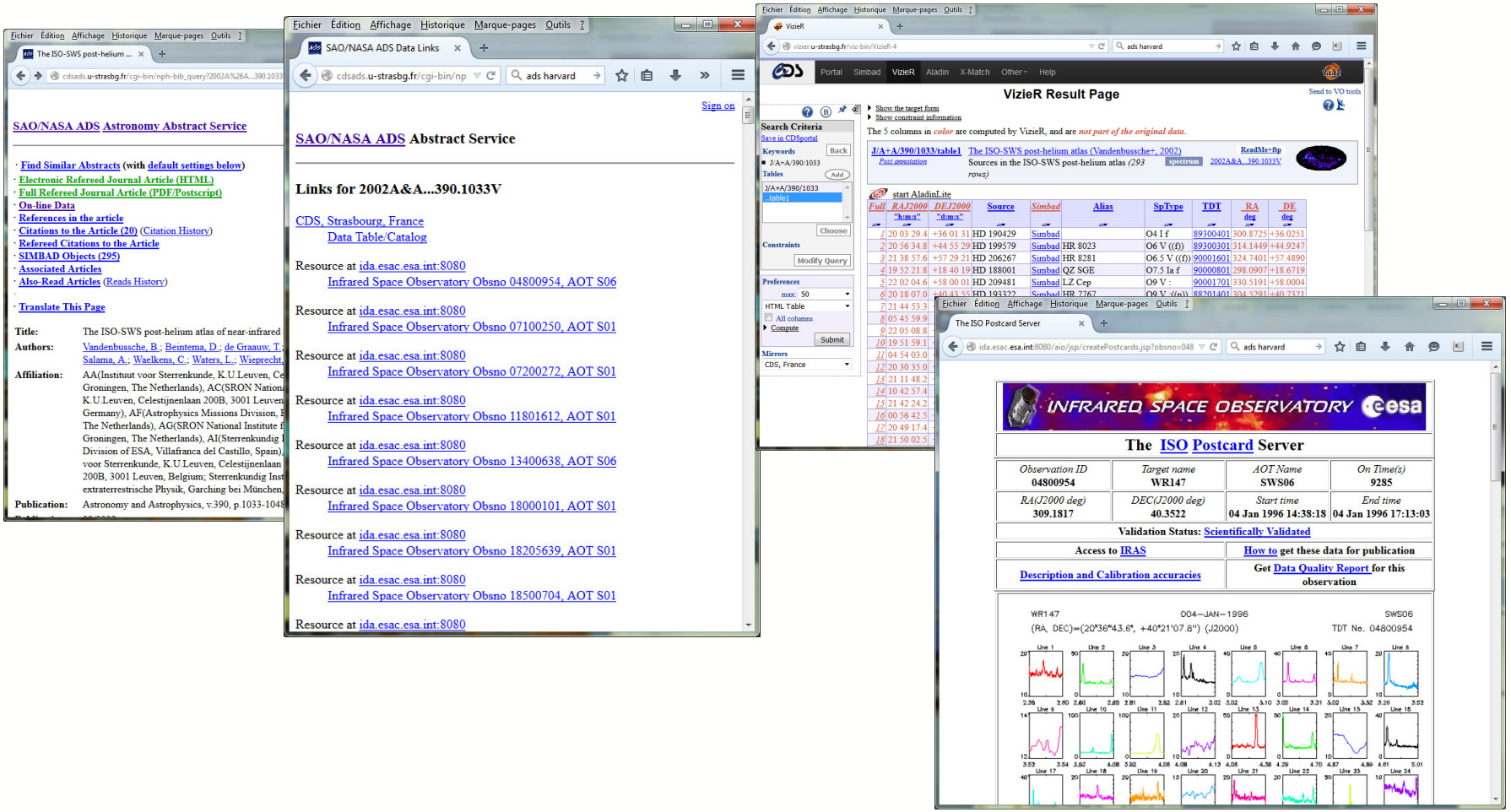
How? Basic elements

- A common data format since the 70s (FITS)
- Strong tradition of international collaboration
- Open data (in general after a proprietary period)
- Driven by community needs (on-line observation archives, on-line services)

Networking and interoperability

- **Networking of on-line resources from 1993-4 (added-value services, journals, archives)**

Early networking, still in use

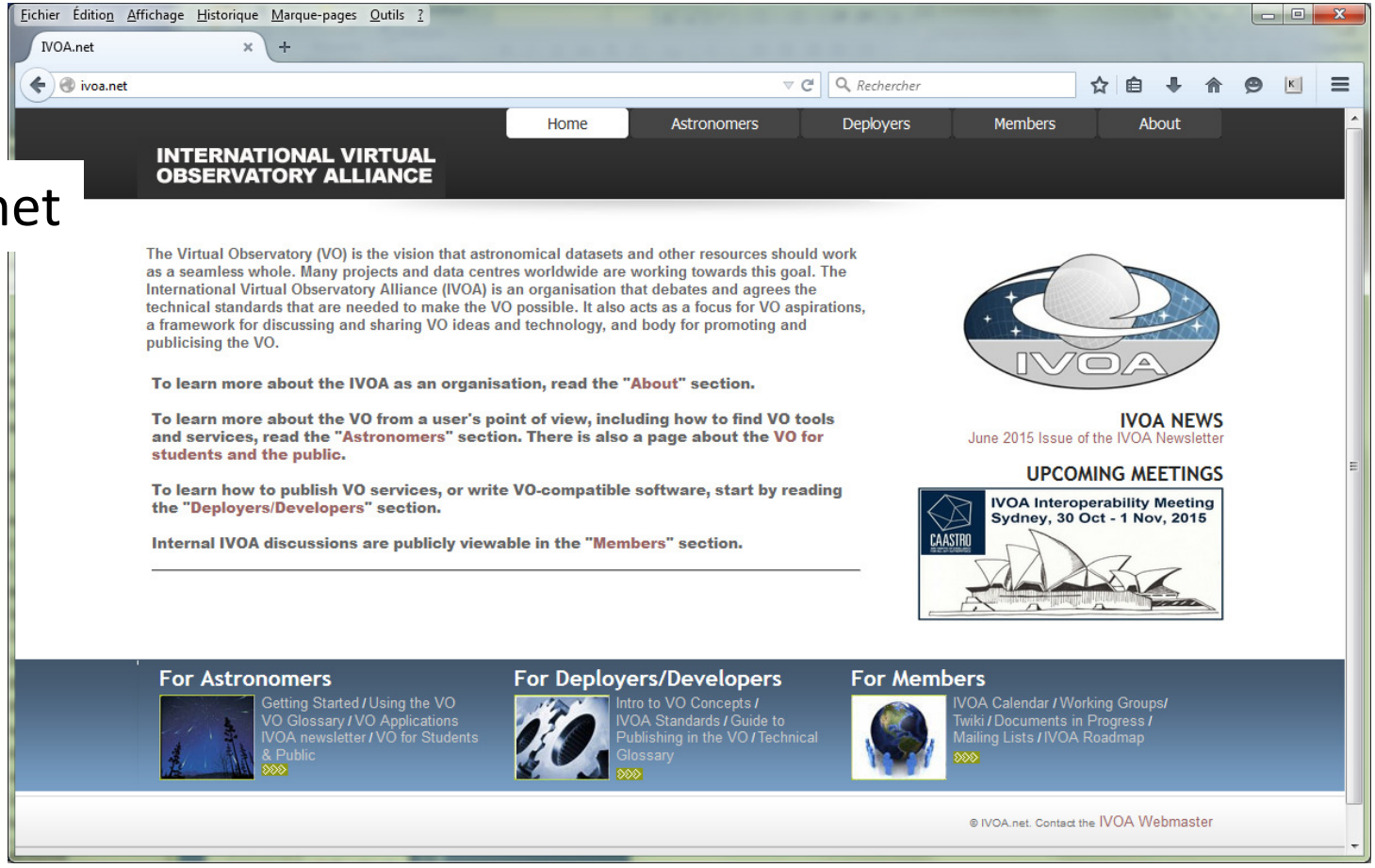


Networking and interoperability

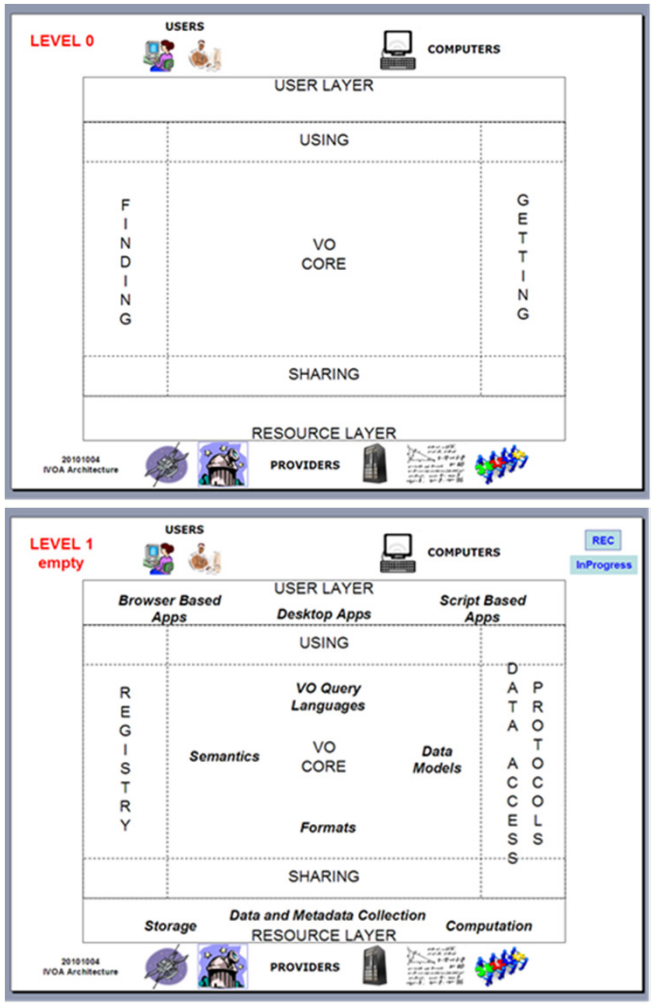
- **Networking of on-line resources from 1993-4 (added-value services, journals, archives)**
- **Seamless access to on-line data (~2000)**
  The astronomical Virtual Observatory
- **The VO framework : standards and data access tools – discover, access, use data**
- **Standards defined by the International Virtual Observatory Alliance (IVOA)**
  - **Procedure inspired from W3C**
  - **When possible generic elements (OAI-PMH, SKOS/RDF)**

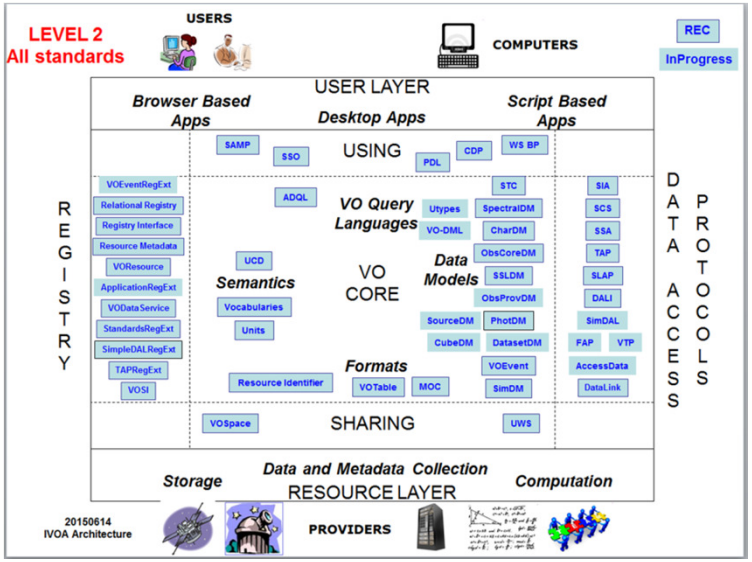
The Virtual Observatory

http://ivoa.net



The IVOA standard framework

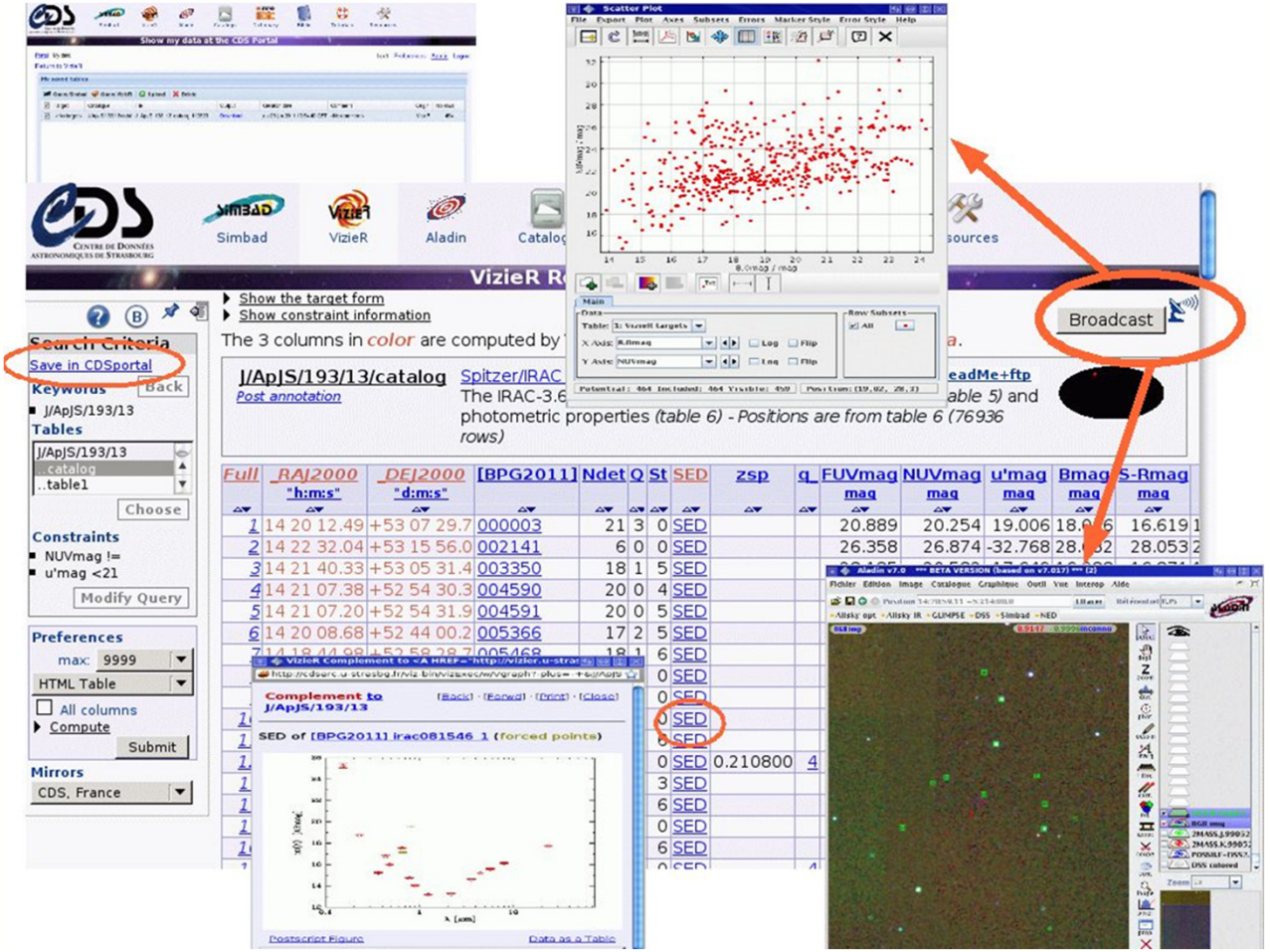




An inclusive and open framework

- **No central point, a multi-polar world, a global endeavour**
- **“Open” and inclusive model**
  - **A thin interoperability layer on top of the data holdings**
  - **Anyone can register a data service or build a tool (more than 100 “authorities” with a registered service)**
- **The VO is invisible but used because people use the services and the tools!**
- **Data providers also imbed VO building blocks in their archives and services**

Interoperable tools and data services



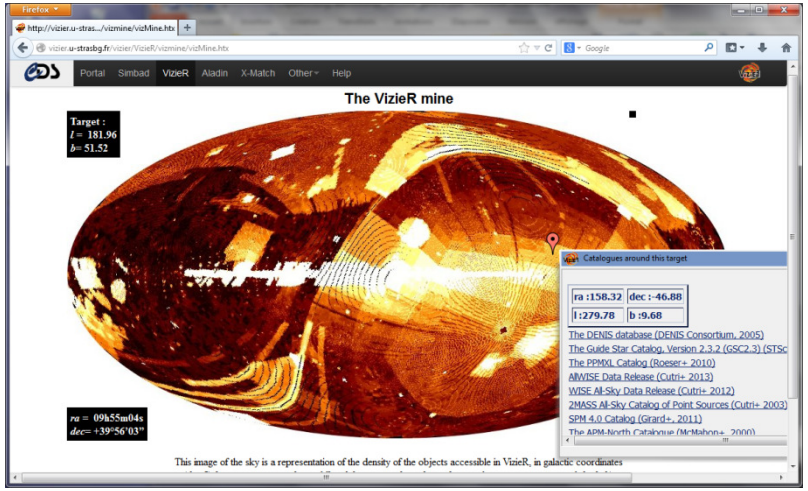
Keys for success

- **Key for success (science users): seamless access to data AND interoperable tools**
- **Key for success (data providers)**
  - **More visibility for their data**
  - **No need to reinvent the wheel, people already worked**

Big and smaller data = astronomical data

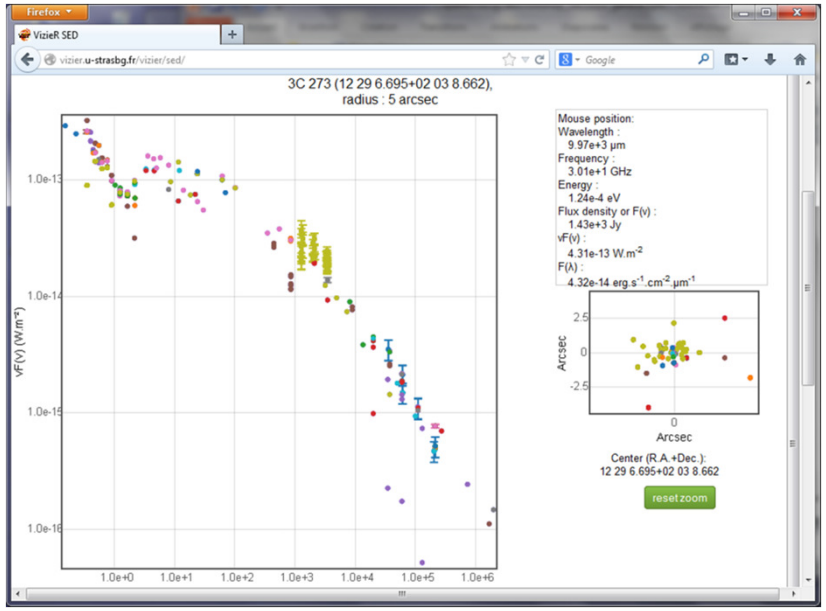
- **Observatory archives + disciplinary data centres**
- **Also data from publications**
  - **Agreement between CDS and the journals (started in 1993 with *Astronomy & Astrophysics*)**
    - **tabular data from publications (also images, spectra, time series)**
    - **together with catalogues from sky surveys, space missions (up to 2 billion rows)**
    - **14 000 "catalogues", i.e. data sets**
  - **Homogeneous metadata describing the very heterogenous content**

« Long Tail » data in VizieR



Data validated
by a publication
Fully discoverable
and usable
Together with the very large surveys

"Photometry viewer":
Spectral points
extracted from the
collection



The European Virtual Observatory

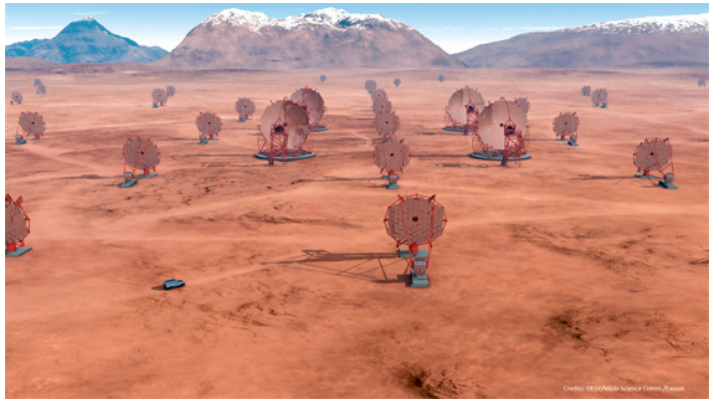
- Coordination of VO activities in Europe
- Three pillars
  - Support for the data providers to provide their data in the VO framework
  - Support for the astronomers in their scientific usage of the VO
  - Technological activities to update the VO framework of standards and tools
- Several EC-funded projects

EURO VO
CoSADIE Collaborative and Sustainable Astronomical Data Infrastructure for Europe

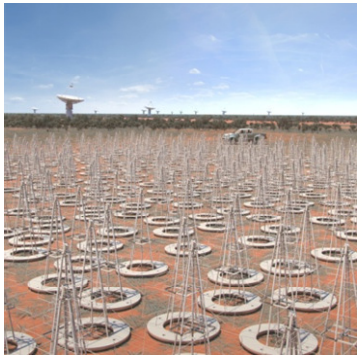
Next step: ASTERICS WP4 – large projects

- Data Access, Discovery and interoperability
- Make the ESFRI and pathfinder project data available for discovery and usage by the whole astronomical community, interoperable in a homogeneous international framework, and accessible with a set of common tools.
- Fully aligned with the current IVOA priorities

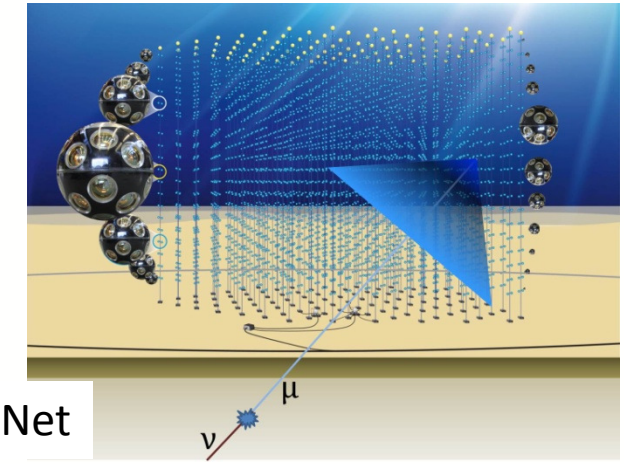
Who is involved



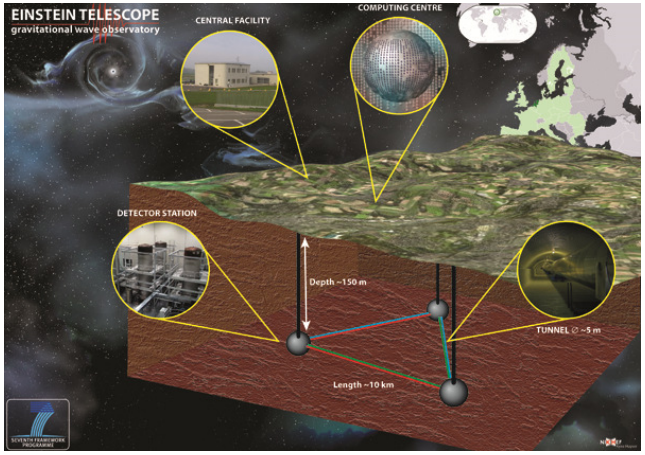
CTA



SKA



KM3Net



Astronomy ESFRI & Research Infrastructure Cluster

Who is involved

- Euro-VO partners, i.e. VO initiatives from France, Germany, Italy, Spain, UK
- Representatives of ESFRI and pathfinders
- Astronomy & Astroparticle physics, including new messengers
- ESO is associated to the project
- ESA (ESAC) is working in close collaboration with Euro-VO

Conclusions

- The way astronomers do science changed a lot with availability of open data on-line and interoperable data and tools
- Data in astronomy includes « Big » and smaller data
- Smaller data = research results, also discoverable, usable and used
- « Long Tail » = huge diversity, also characteristic of Big Data

Conclusions

- User-centered approach
  - Technology is a tool, not an aim
  - Data sharing in astronomy is science-driven, not technology-driven
- Evaluation of relevant new technologies to identify whether they fit our needs
- Beware of the buzz around seducing but volatile technologies – implementation is resource consuming and technologies must be « sustainable enough »!

Conclusions

- Elements of the VO framework are customized and reused by « nearby » disciplines – planetary studies, the Virtual Atomic and Molecular Data Centre
- The generic elements (registry of resources, vocabulary concepts) allow astronomy data infrastructure to interface with the generic data framework

Sharing scientific data – Open Science

- Strong Open data statement (among others) of G8 Ministers of Research in June 2013
- More and more demand from funding agencies (at least for a data management plan)
- Astronomy has been a pionneer and shares lessons learnt in the Research Data Alliance

The Research Data Alliance

- Founded in March 2013 by EC, NSF and Australia
- 3000 members from more than 100 countries
- Bottom-up work to tackle all the aspects of scientific data sharing, technological as well as « sociological »
- Have a look at

  rd-alliance.org!