

# Data management in Astrophysics

Data, Policy, standards and tools

C. Surace

Laboratoire d'Astrophysique de Marseille



INTERNET,  
SCIENCE,  
TECHNOLOGIES...

## La déferlante des OCTETS



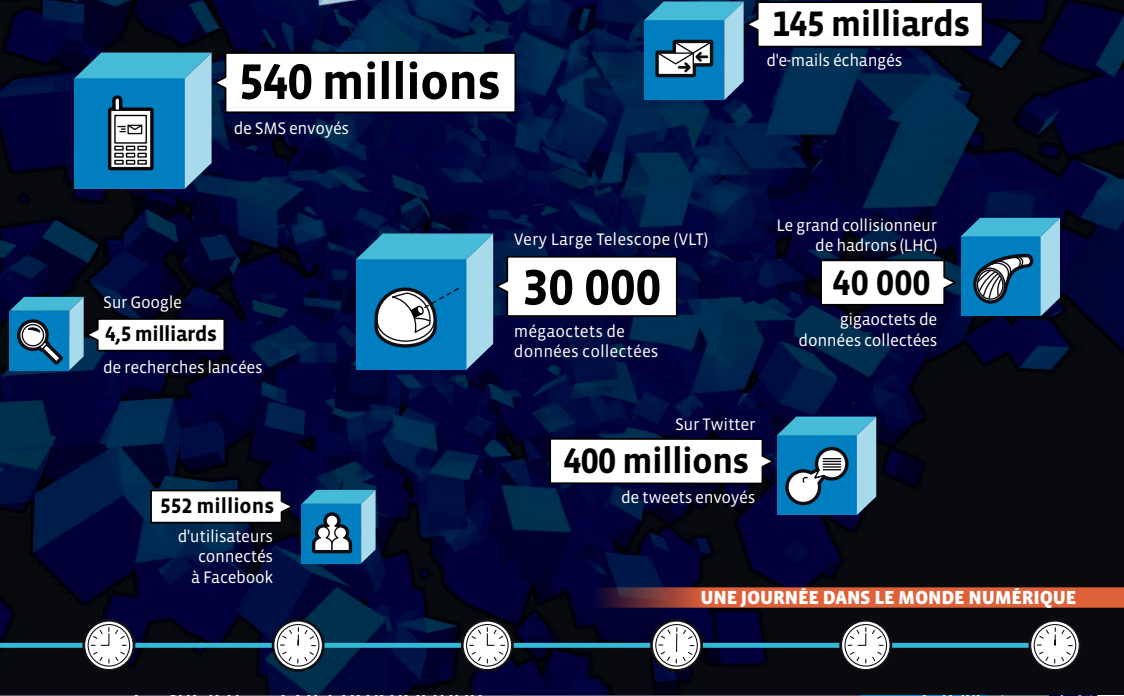
**L'événement**  
Philippe Descola,  
Médaille d'or  
du CNRS 2012

**Le grand entr...**  
Les cinquan  
du  
franco-alle

N° 269 | NOVEMBRE-DÉCEMBRE 2012

L'enquête

21



UNE JOURNÉE DANS LE MONDE NUMÉRIQUE



C. Surace - Data preservation

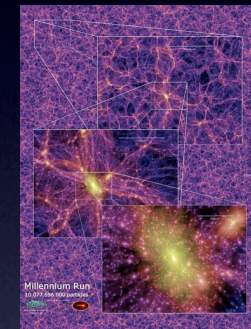
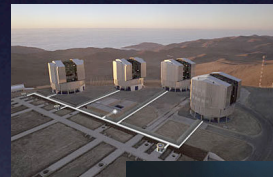
LAMI  
LABORATOIRE D'ASTROPHYSIQUE  
DE MARSEILLE

CESAM  
11110001  
11100001  
11001101



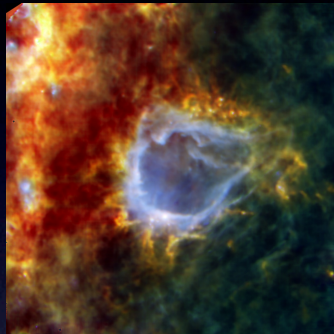
# Data Characteristics

- Different providers
  - Simulators, (Millenium,...)
  - Telescopes,
  - Satellites
- Diversity
  - Different kind of products

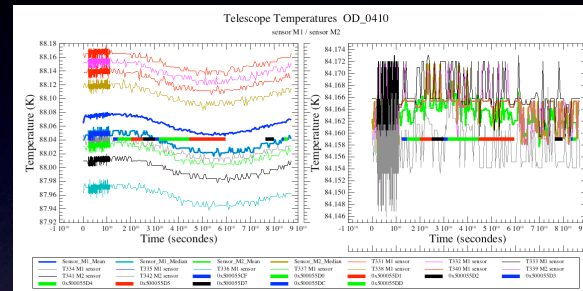


# Data in Astrophysics

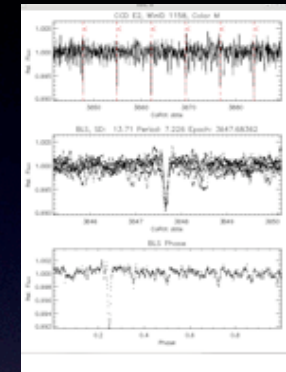
2D Images



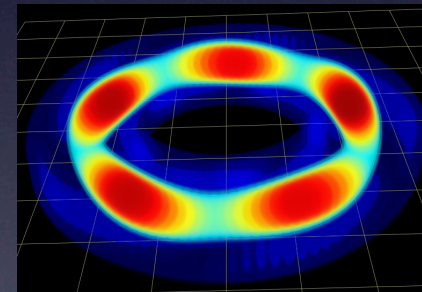
Spectra



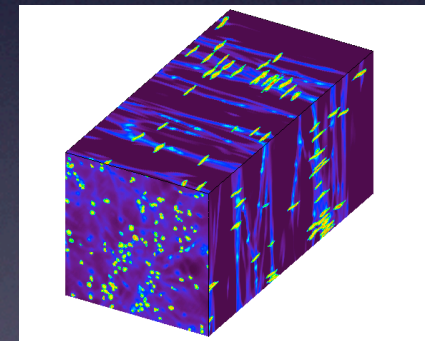
Time series



Simulation data



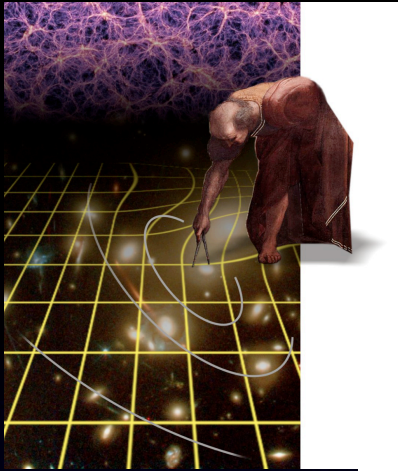
3D Cube data



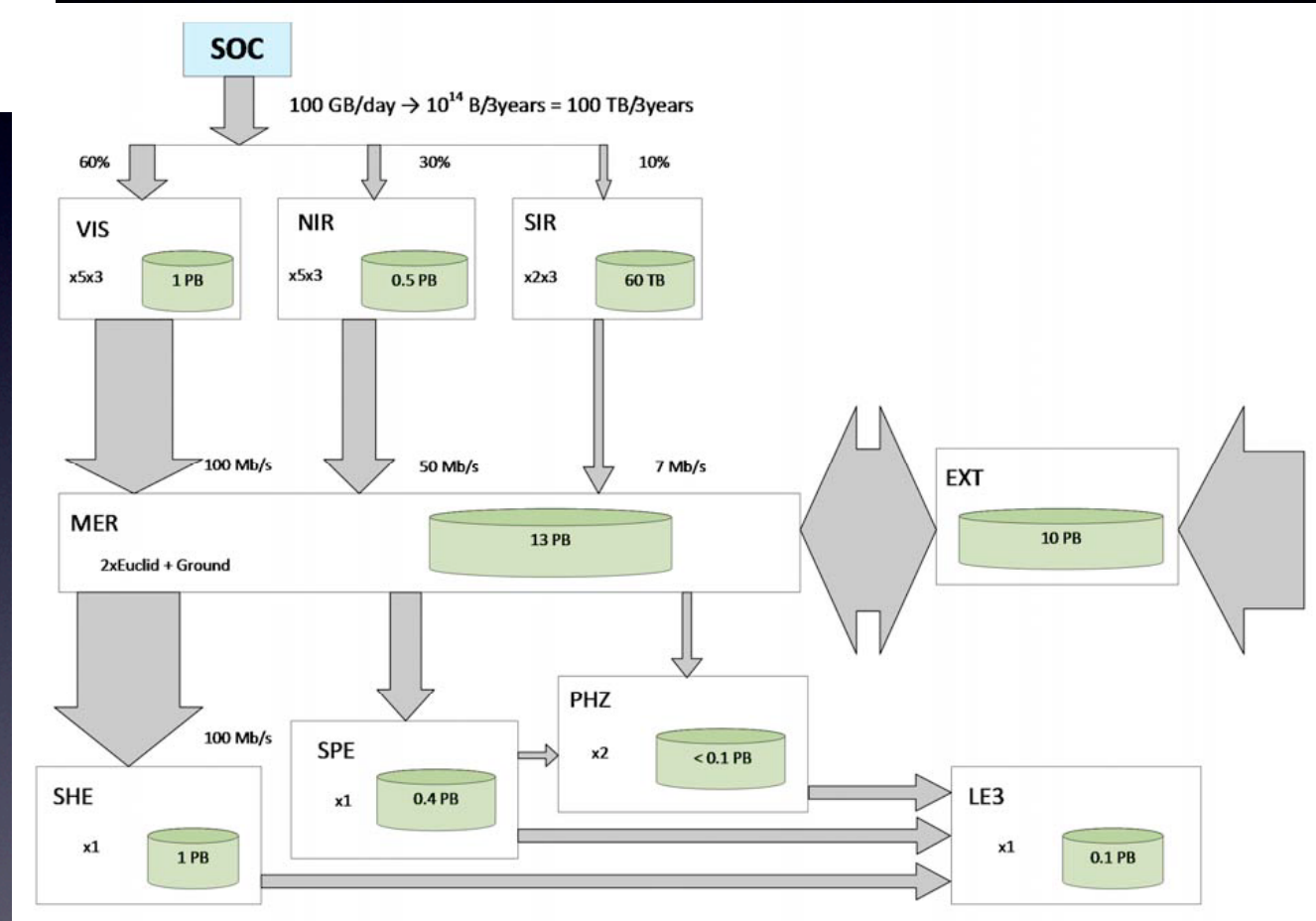
Tabular data

| RA              | Dec            | RAErr               | DecErr               | RA_Mean             | Dec_Mean             | RA_Min          | Dec_Min          | RA_Max           | Dec_Max          | RA_Mid           | Dec_Mid          | RA_MinErr        | Dec_MinErr       | RA_MaxErr        | Dec_MaxErr       | RA_MidErr        | Dec_MidErr       |
|-----------------|----------------|---------------------|----------------------|---------------------|----------------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| 186.61071916638 | 62.22266377957 | 0.00070513822064476 | 0.000532591732024145 | 0.00070513822064476 | 0.000532591732024145 | 156.46669553507 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 | 0.89665301747295 |





# Euclid





# Next data

- LSST (2020) : 3 Billions pixels every 17 seconds.
- 140 petabytes



Pan-Starrs : 6,000 deg<sup>2</sup> per night.

several terabytes per night for the full telescope.  
hundreds of Peta bytes.

Software techniques are therefore being developed to extract the information

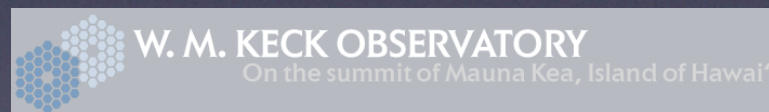
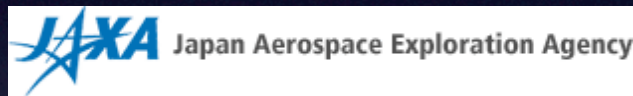
LSST : <http://www.lsst.org/lsst/>

Pan-Starrs : <http://pan-starrs.ifa.hawaii.edu/>



# The actors 1

## the providers and agencies



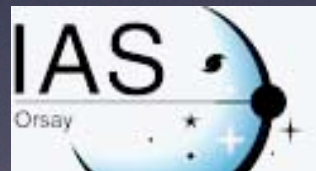
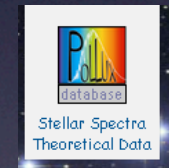
# The actors 2

The international and national collaboration





# The national organisations the SO-5



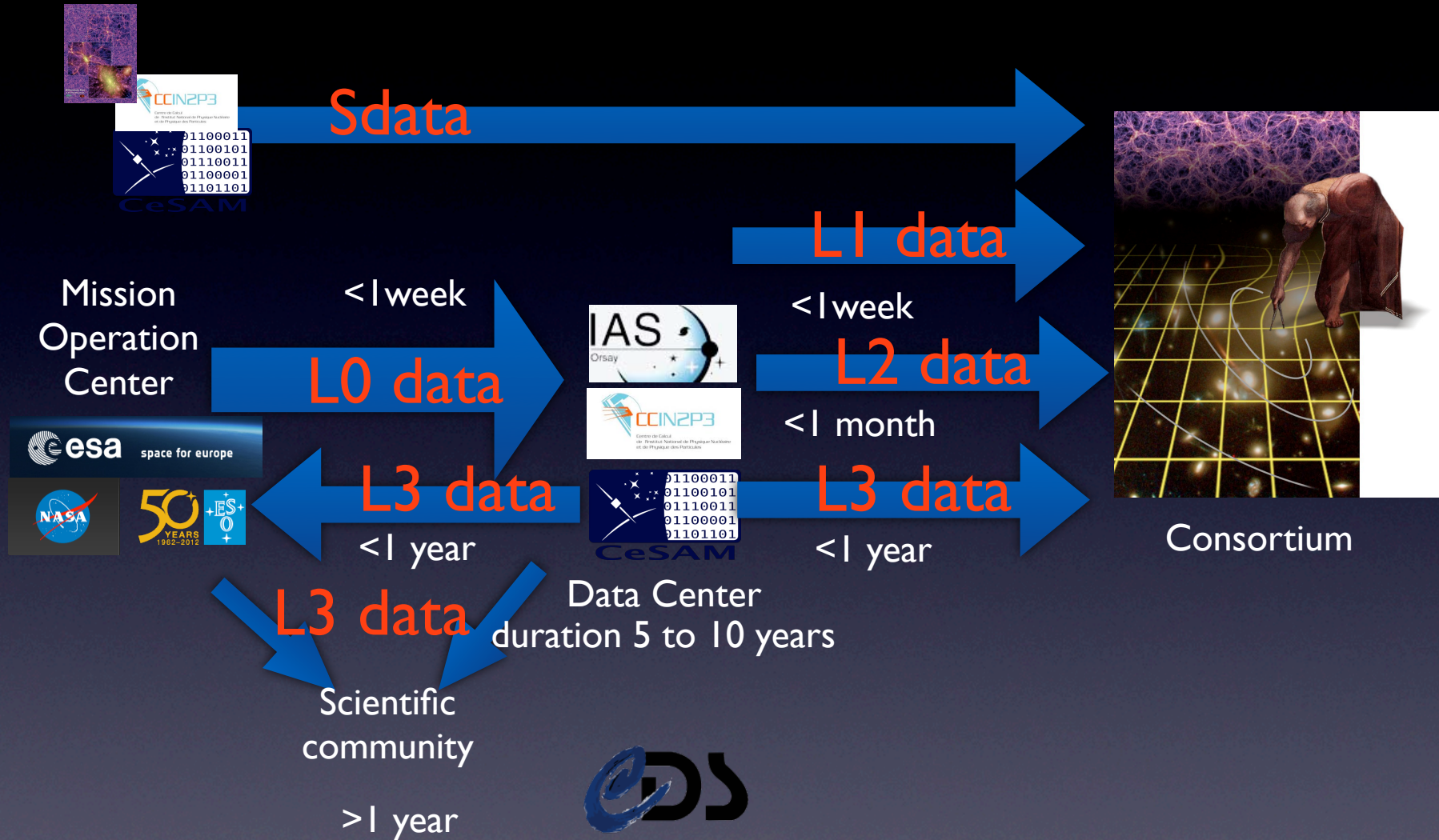
# Data Levels

- Sdata - Simulation data
  - Simulated Data
- Level 0 - Raw data
  - Data from telescopes, satellites
  - stored in the Mission Operation Center
- Level 1 data
  - Data corrected from instrument features
- Level 2 data
  - Scientifically calibrated data
- Level 3 data
  - Data cross correlated with ancillary data

N  
4 N  
N  
0.2 N  
0.2 N



# Data Flux



# Formats and Archives

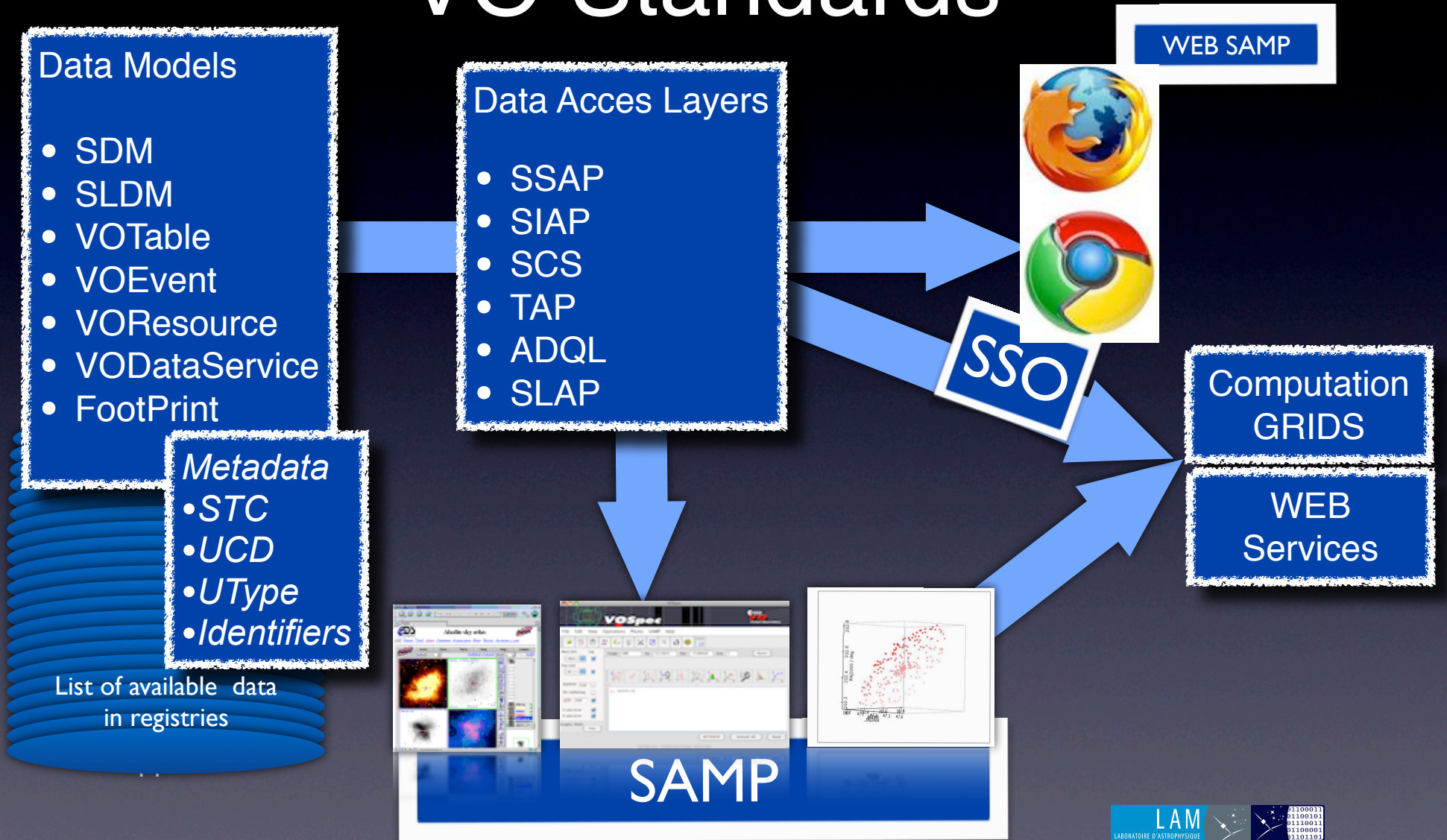
- FITS : Flexible Image Transport System
  - Initial release 1981 <http://fits.gsfc.nasa.gov>
- VO Format : Virtual Observatory formats
  - Characterisation of metadata <http://www.ivoa.net>
- Simulation formats
  - GADGET (Octree) - RAMSES (AMR) (HDF5) - GASOLINE (SPH)



# VO Standards and access protocols based on XML exchange format

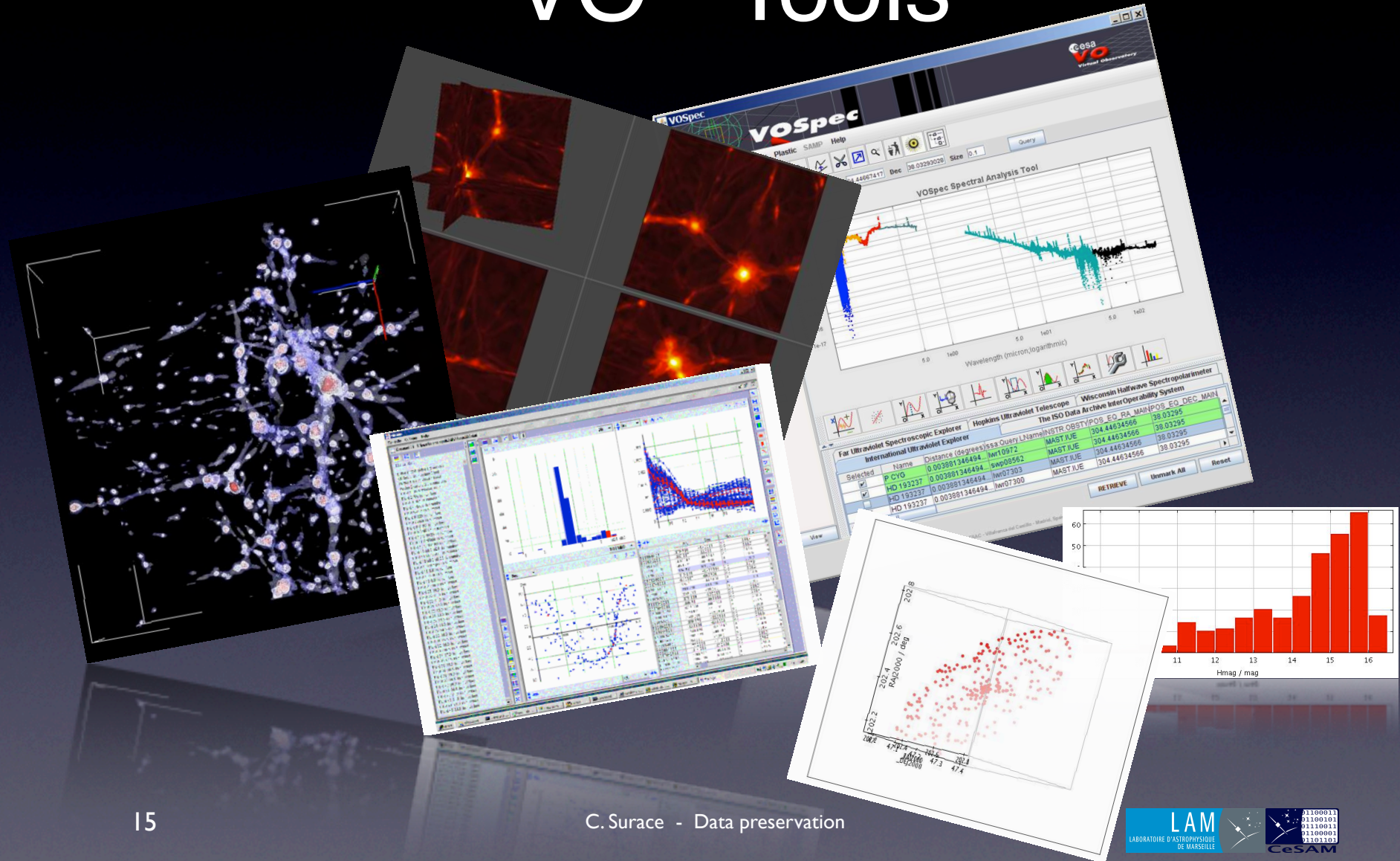
- DataModels
  - STC : Space Time Coordinates
  - Spectrum : describes a spectrum
  - characterisation
- in development :
  - Observation Provenance
  - VOEVENT (will replace GCN)
  - Simulation

# VO Standards



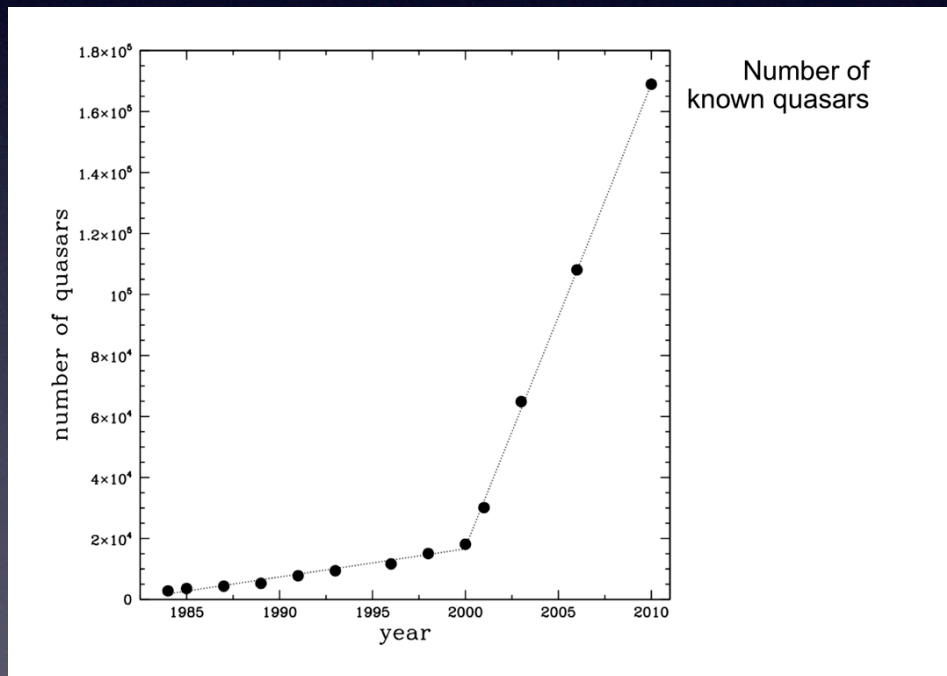


# VO - Tools



# Goals

- The final goal of the VO is to facilitate and foster astronomical research and astronomers are its ultimate users
- Therefore, scientific requirements should drive the IVOA process



Paolo Padovani interop meeting 2010



# A non exhaustive list (from IVOA)

- **Portals and queries :**
  - NVO Portal Services US National Virtual Observatory
  - DataScope Broadcast query
  - NOAO NVO Portal NOAO Image Visualization Discovery Tool
  - VODesktop A resource-centered desktop client for VO: includes VOExplorer, Query and Task Runner, Astroscope, Myspace Browser
  - Octet CVO Observation Catalog Exploration Tool
  - VOCat - VOIndia A catalog data interface tool
- **Tabular DATA Analysis**
  - VOPlot - VOIndia A tool for visualizing astronomical data
  - TOPCAT Tool for OPerations on Catalogues And Tables
  - STILTS Command-line tools for table/VOTable manipulation
  - Treeview A viewer for hierarchical structures
  - VOStat - VOIndia A tool for statistical analysis of astronomical data

# What else

- PRO
  - easier and easier to use
  - more and more data
  - scientific goals drive technical developments
- TO IMPROVE
  - Quality
  - Curation
  - Semantics
  - High performance computing
- TODO :
  - include ALL data
  - easy to use portals (<http://cdsportal.u-strasbg.fr/>)
  - SEDs, Photometry
  - Data Mining (AMADEUS - Mastosdons)
  - ObsTAP
  - Promising cloud computing



# Preservation Environment

- preservation of data
  - FITs, VO, GADGET,... and Archives
- preservation of accessibility of data
  - VO, Web Services and tools
- preservation of discovery capabilities
  - Data Mining, Data Visualisation
- preservation of knowledge
  - WF4ever (<http://www.wf4ever-project.org/>) / VO Workflow - Patterns

