Astrophysical Virtual Observatory

-

a GRID application

R. Albrecht - P. Benvenuti

ESAGRID WORKSHOP
ESSTEC, October 25th, 2002

Astrophysics's peculiarities

- Observational science
- The information is carried by photons (of all wavelengths !)
- The Lab is the Universe
- Experiment's conditions are supplied by the observations of "families" of objects
- Astronomers are flooded with data
- Statistical approach is becoming essential

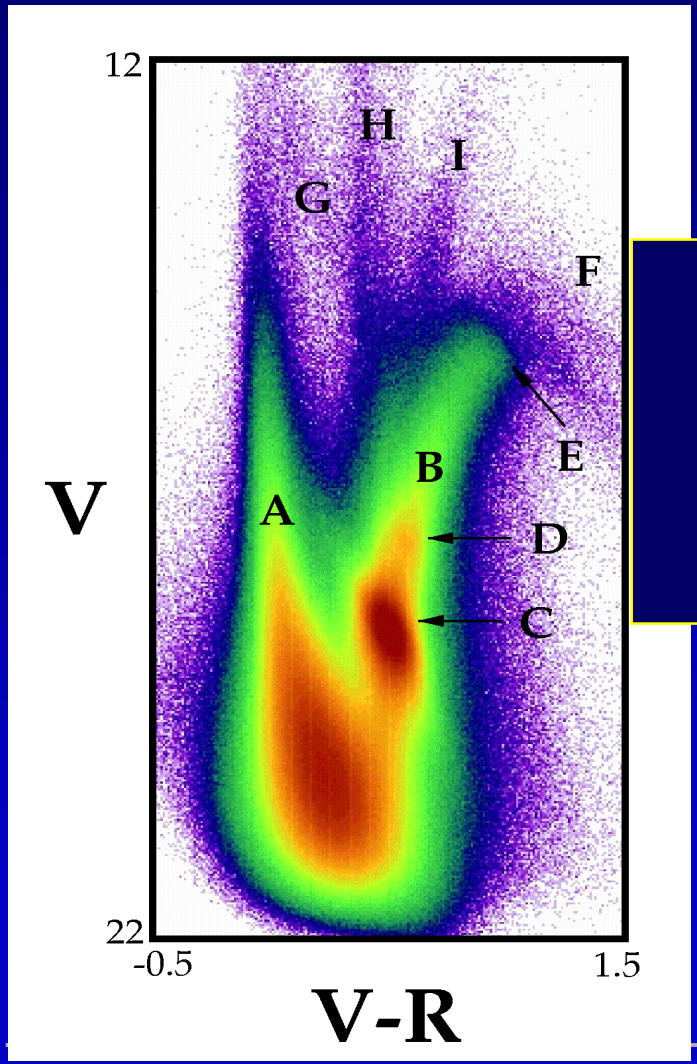
Opportunities for New Science

Mining Large Data Sets as a discovery tool:

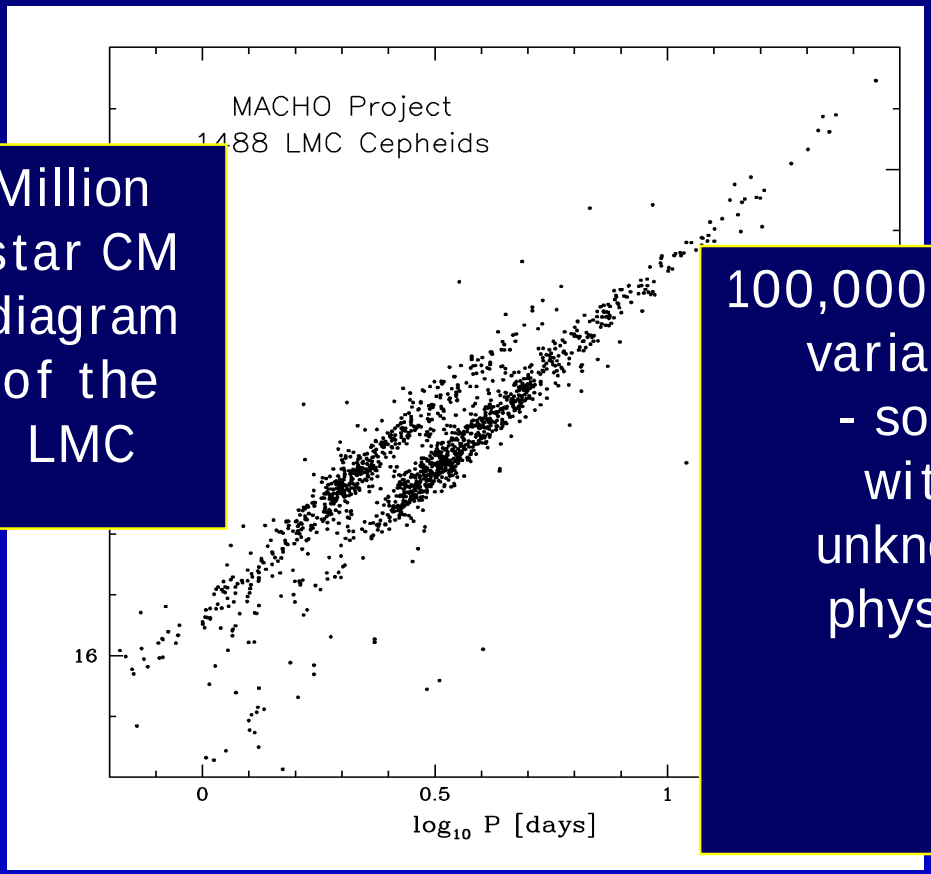
- Statistical Significance
- Rare Events
- The Unexpected

The issue: joining Large Data Sets

Statistical Significance

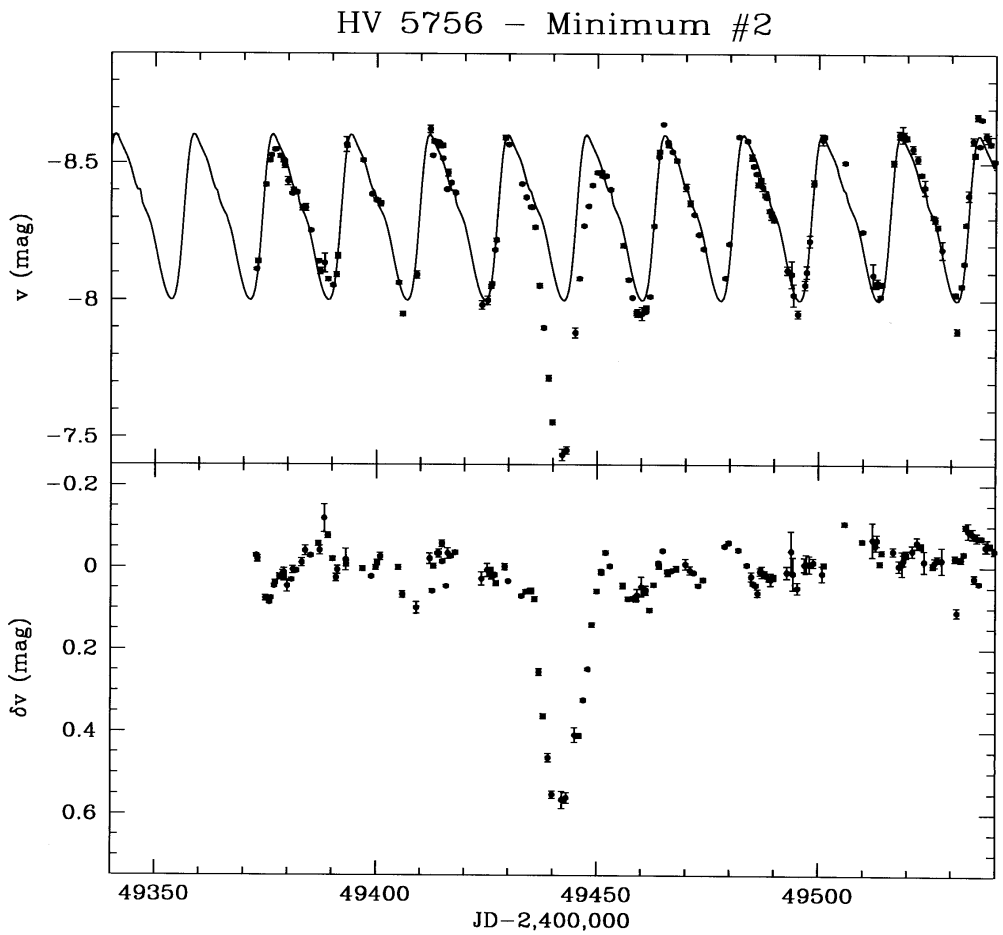


9 Million star CM diagram of the LMC



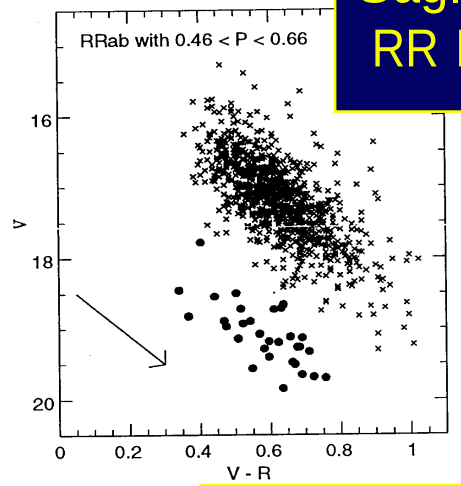
100,000 new variables - some with unknown physics

The Rare

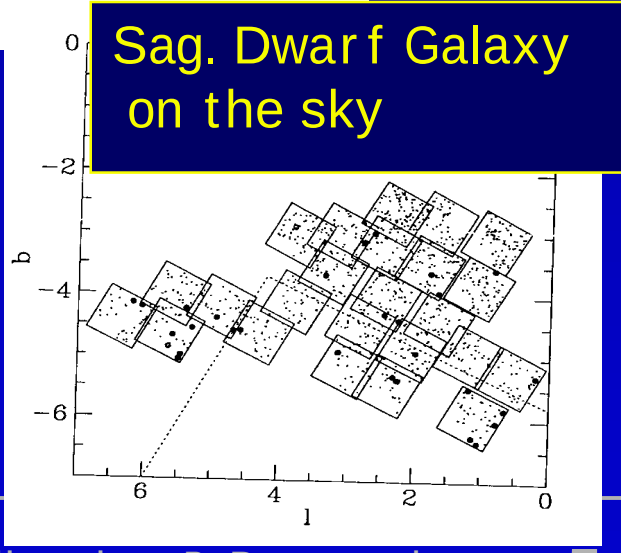


Eclipsing Classical Cepheid

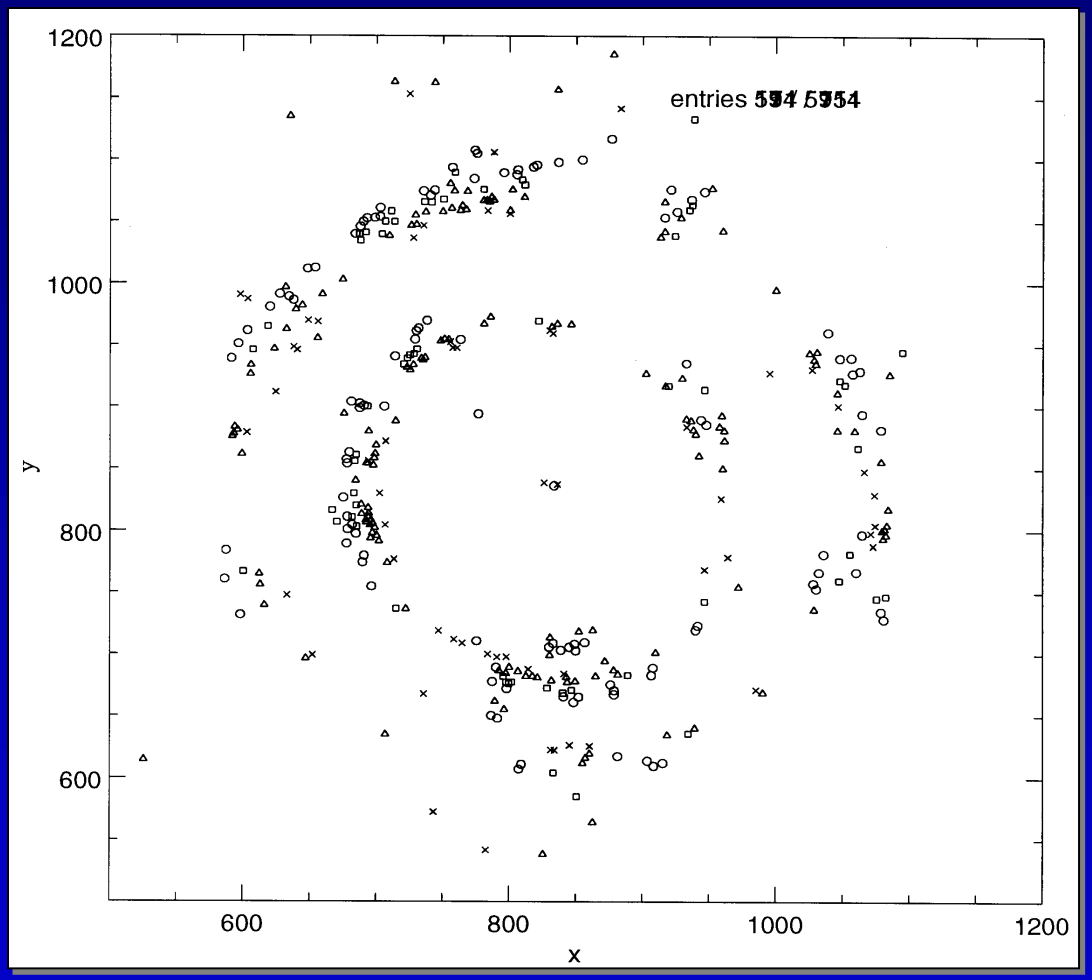
Sag. Dwarf Galaxy
RR Lyr.



Sag. Dwarf Galaxy
on the sky



The Unexpected



Dark Matter Bubbles?

Why a Virtual Observatory ?

- On-line digital Archives do exist
- Friendly web interfaces for browsing and retrieving data do exist
- Analysis tools and affordable workstations do exist
- What do we need more than that ?
  - …maybe just grants to buy more workstations and data-slaves…

Reality is different !

- Calibration, data quality and data description are highly non-uniform across Archives
- Merging multi-archive data is possible, but it is still a manual, tedious affair
- New instruments and surveys generate huge quantity of pixels (to be processed and analyzed…)
- Algorithms and analysis tools that can operate on distributed large data sets do not exist yet

How a VO can help?

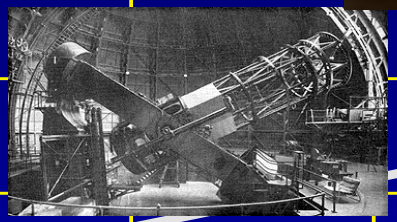
- It will make archives really inter-operable
- It will create a new generation of algorithms and procedures
- It will offer technical expertise to its Users

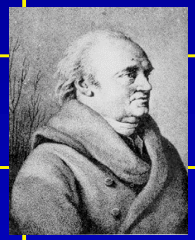
A technical issue…

- The increase in the data acquisition rate is larger than the increase in network bandwidth and local computing power
- The GRID approach offers a solution:
  - The data stay where they are
  - Processing is distributed
  - The network is used to transfer the results, not the raw data

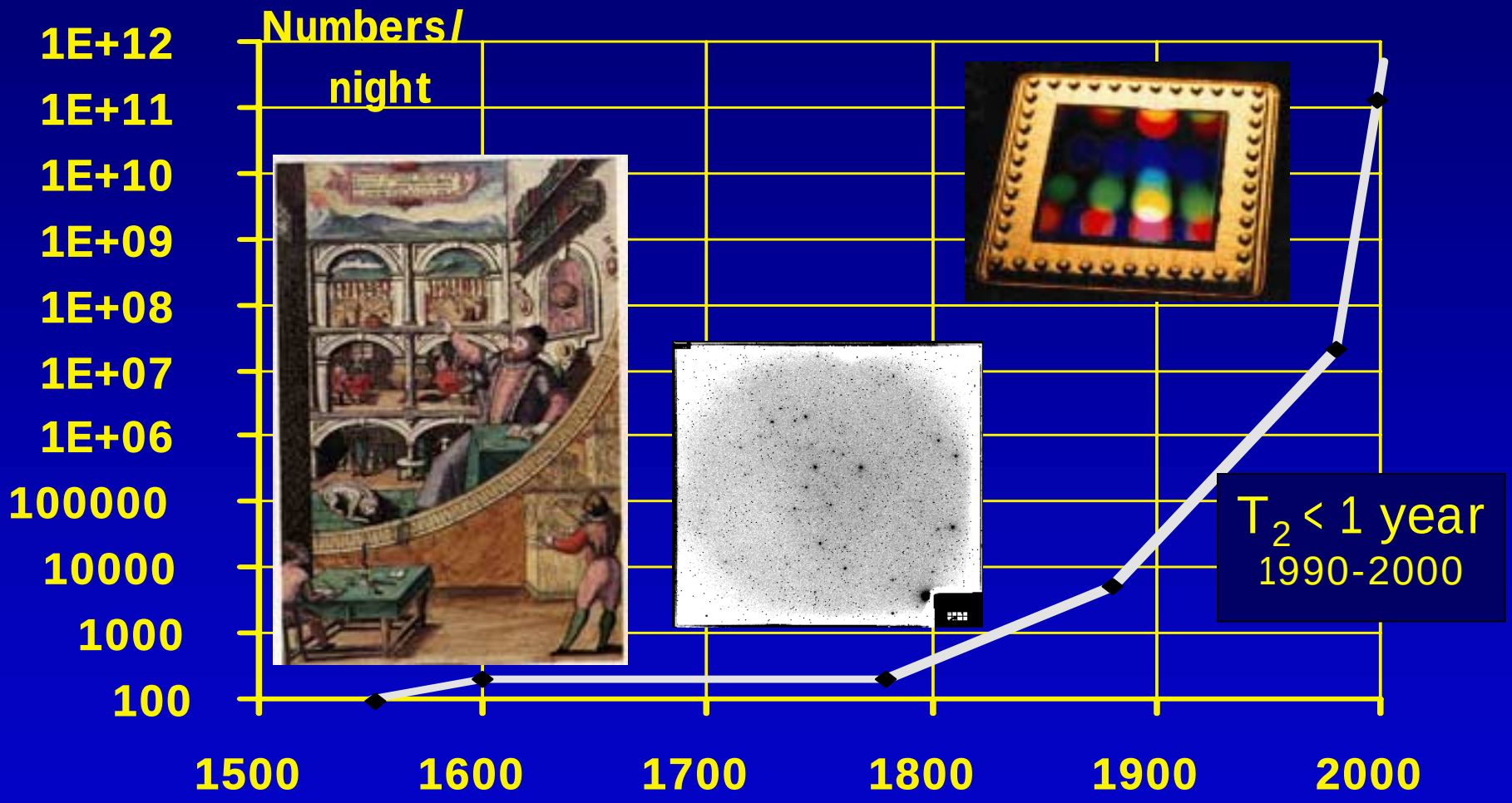
Eyes on the Sky



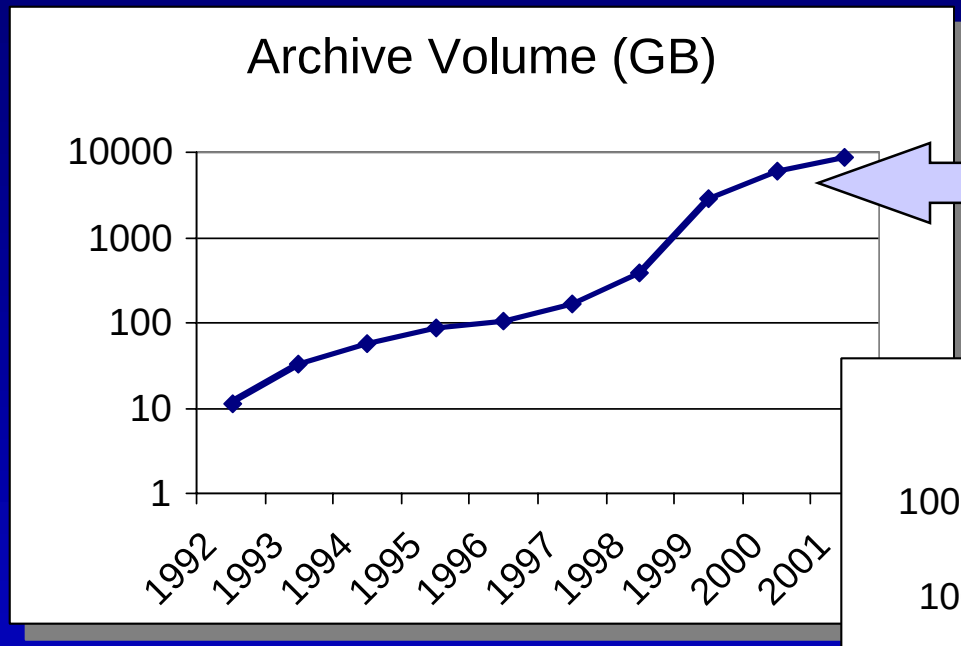




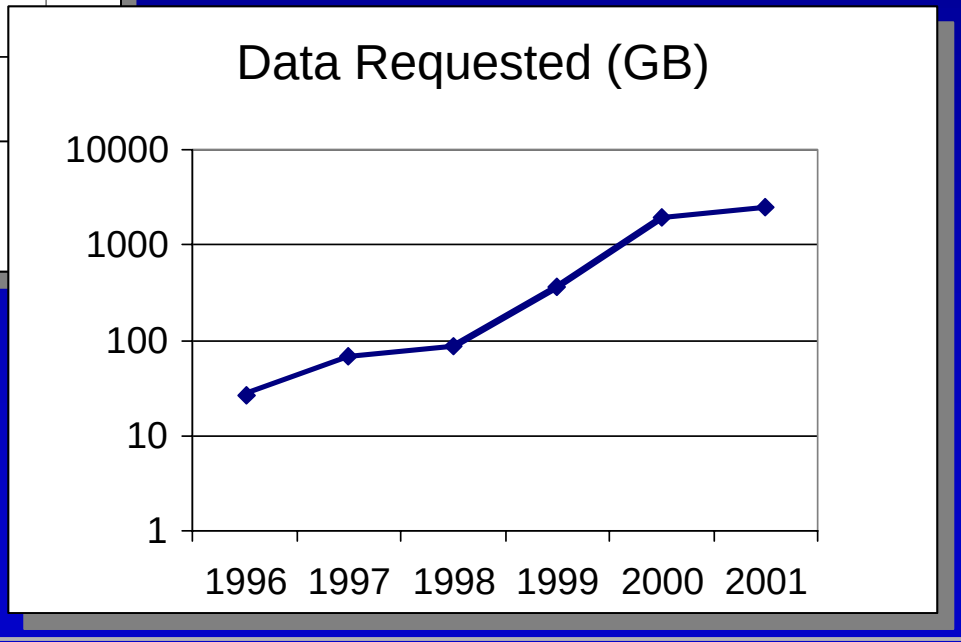
Numbers from the sky



Science Data Volume

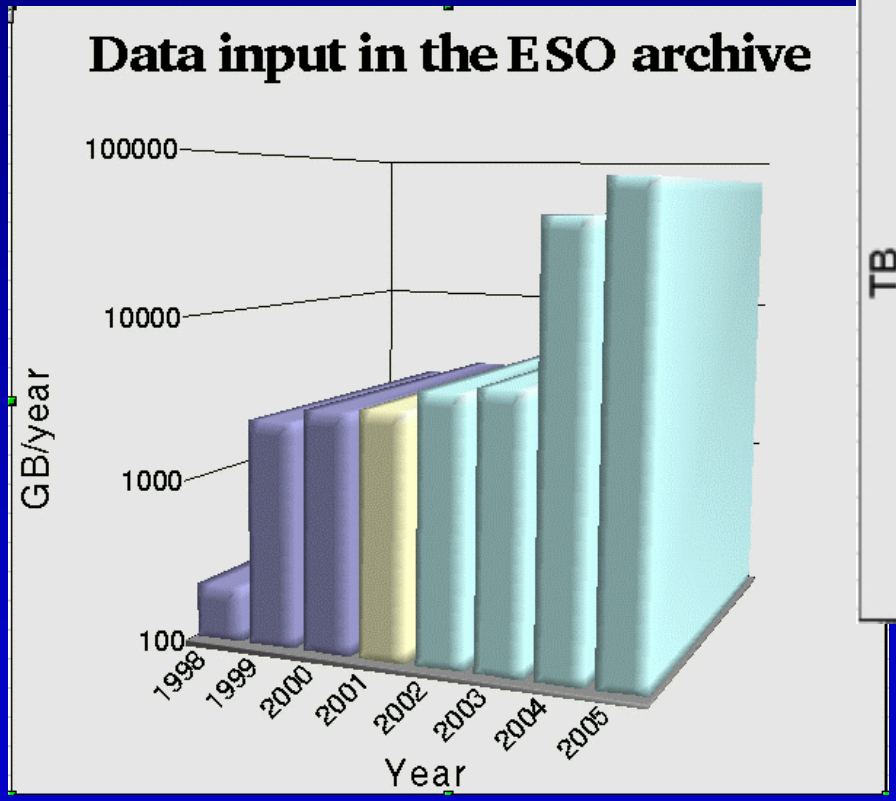


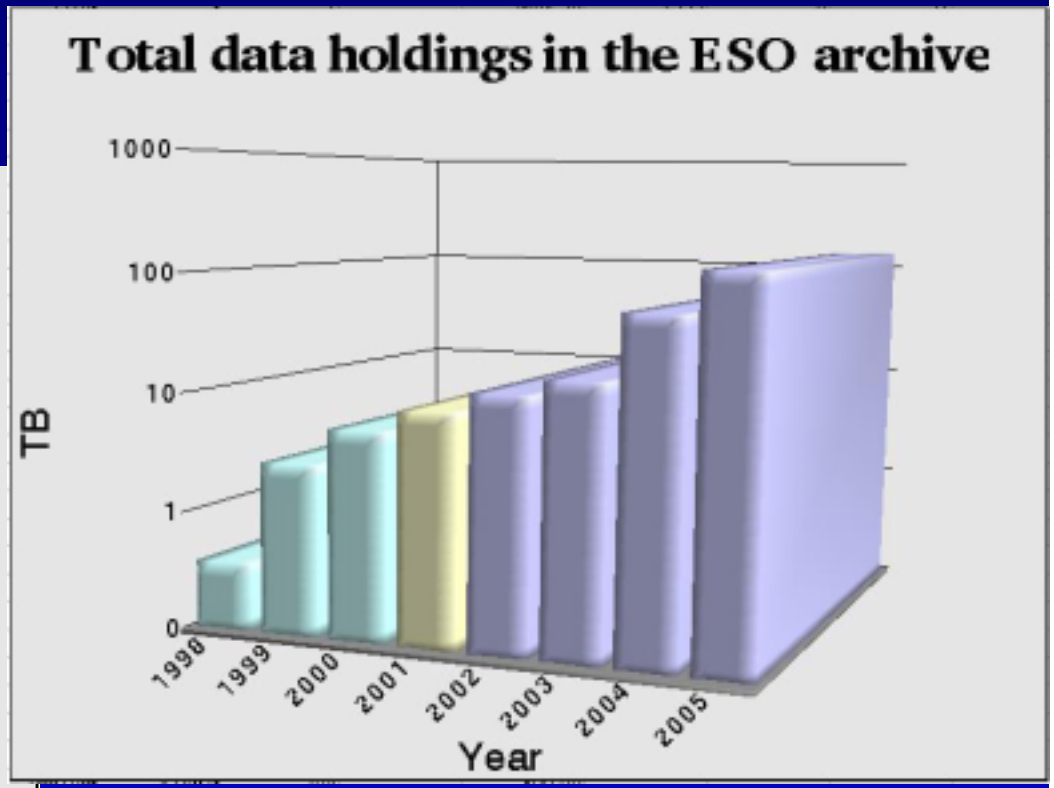
$T_2 < 12$ months



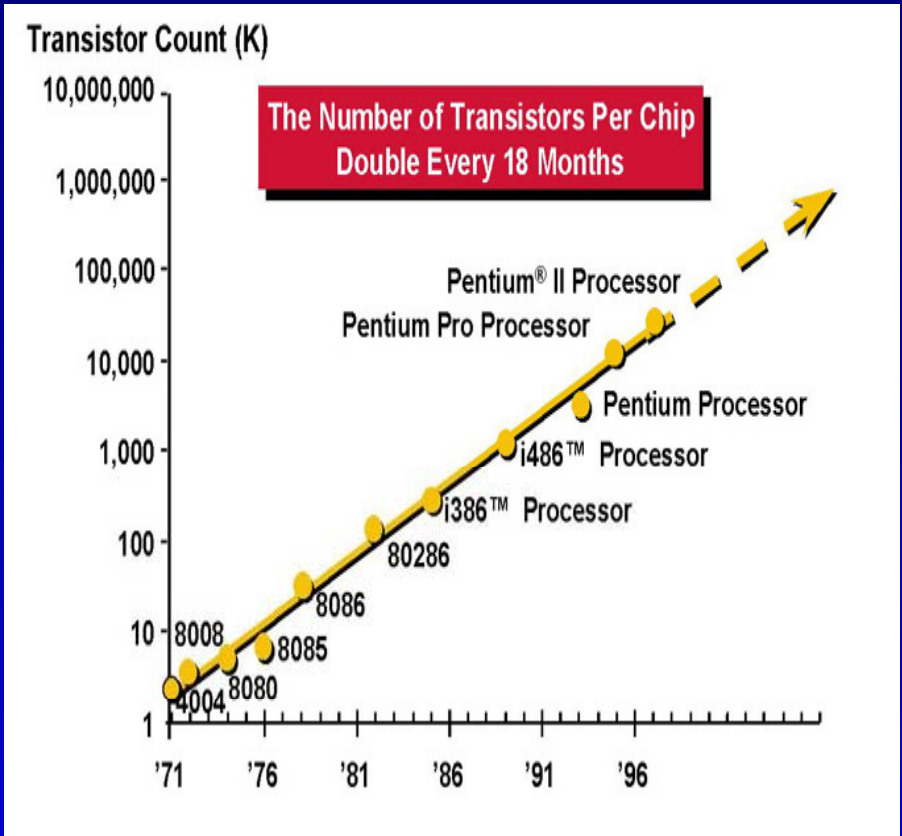
ESO/STECF
Science Archive Facility

Projected Growth



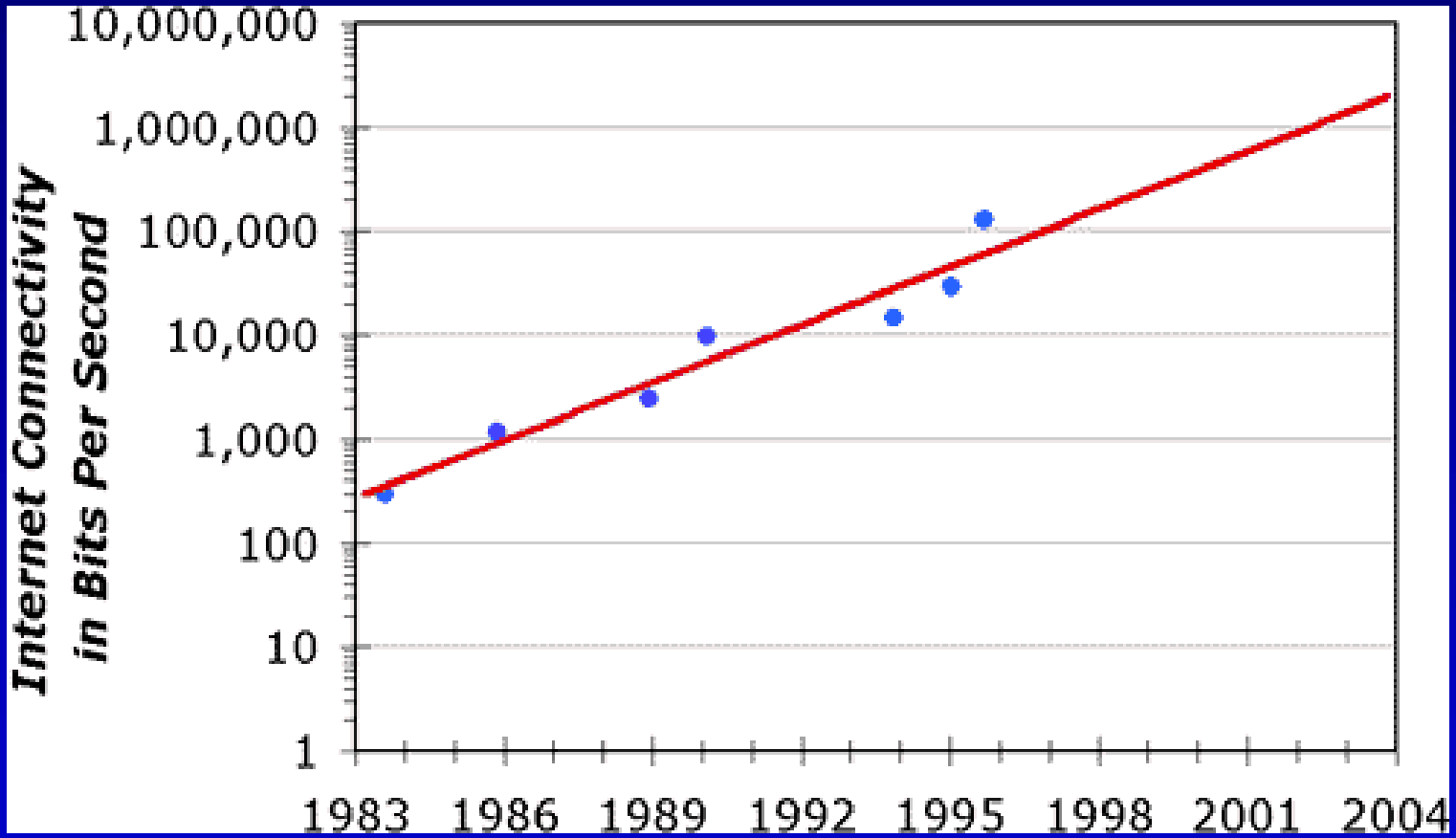


Compute Power



- Moore's Law
- T_2~18 mths
- Strategy to exceed this: parallelism
- Implies : bigger,distributed, shared systems

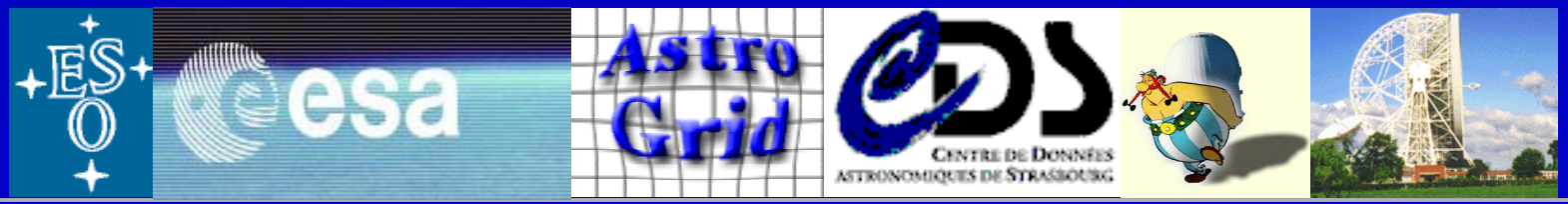
Network Bandwidth



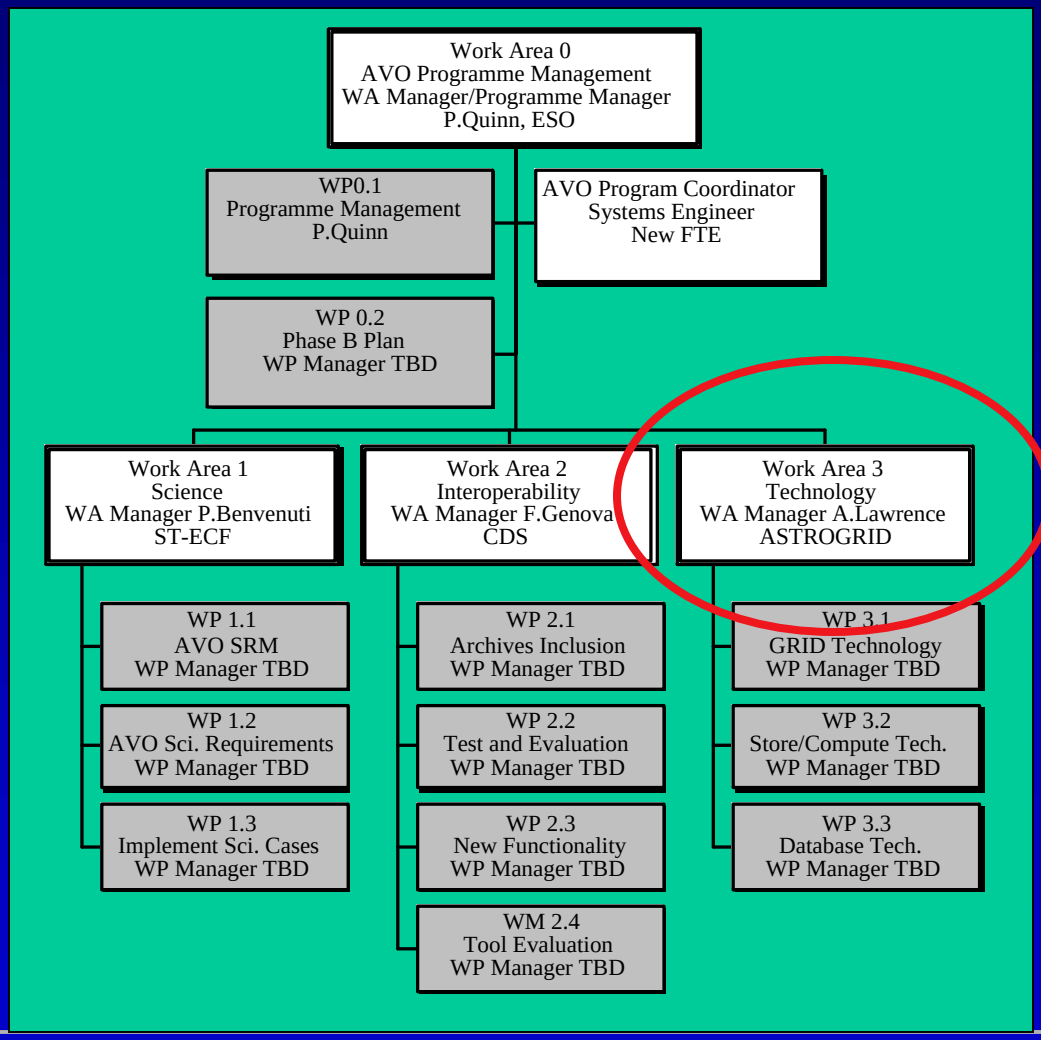
Nielsen's Law : T_2 ~ 20 months

AVO Status

- EU RTD 6th FP Proposal submitted in February 2001
- AVO approved with EU funds ~2 Million €
  (total budget ~ 4M €)
- Contract started in November 2001 - 3 Year Phase A study
- 9 NEW POSITIONS for 3 years over 6 institutions
- Total VO funding AVO+NVO+ASTROGRID = $21 million (US)



AVO Work Program

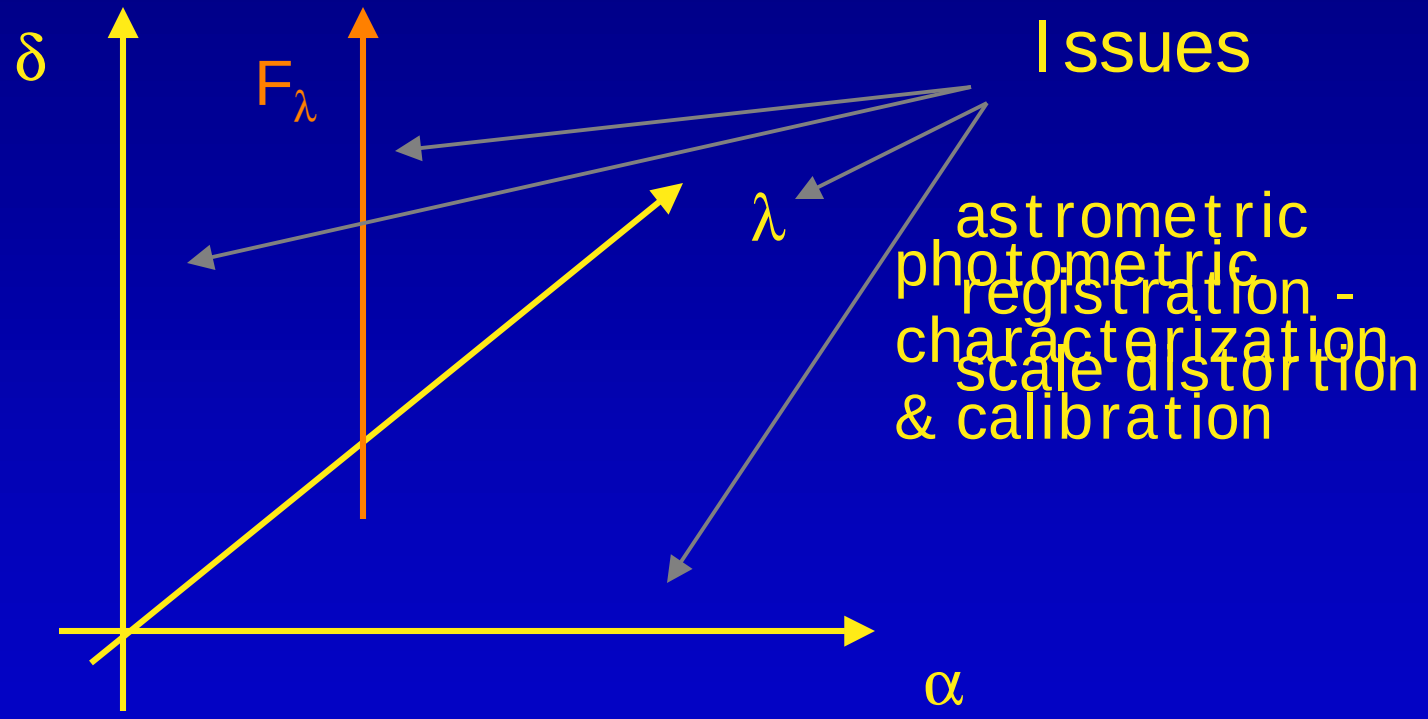


Work Areas :

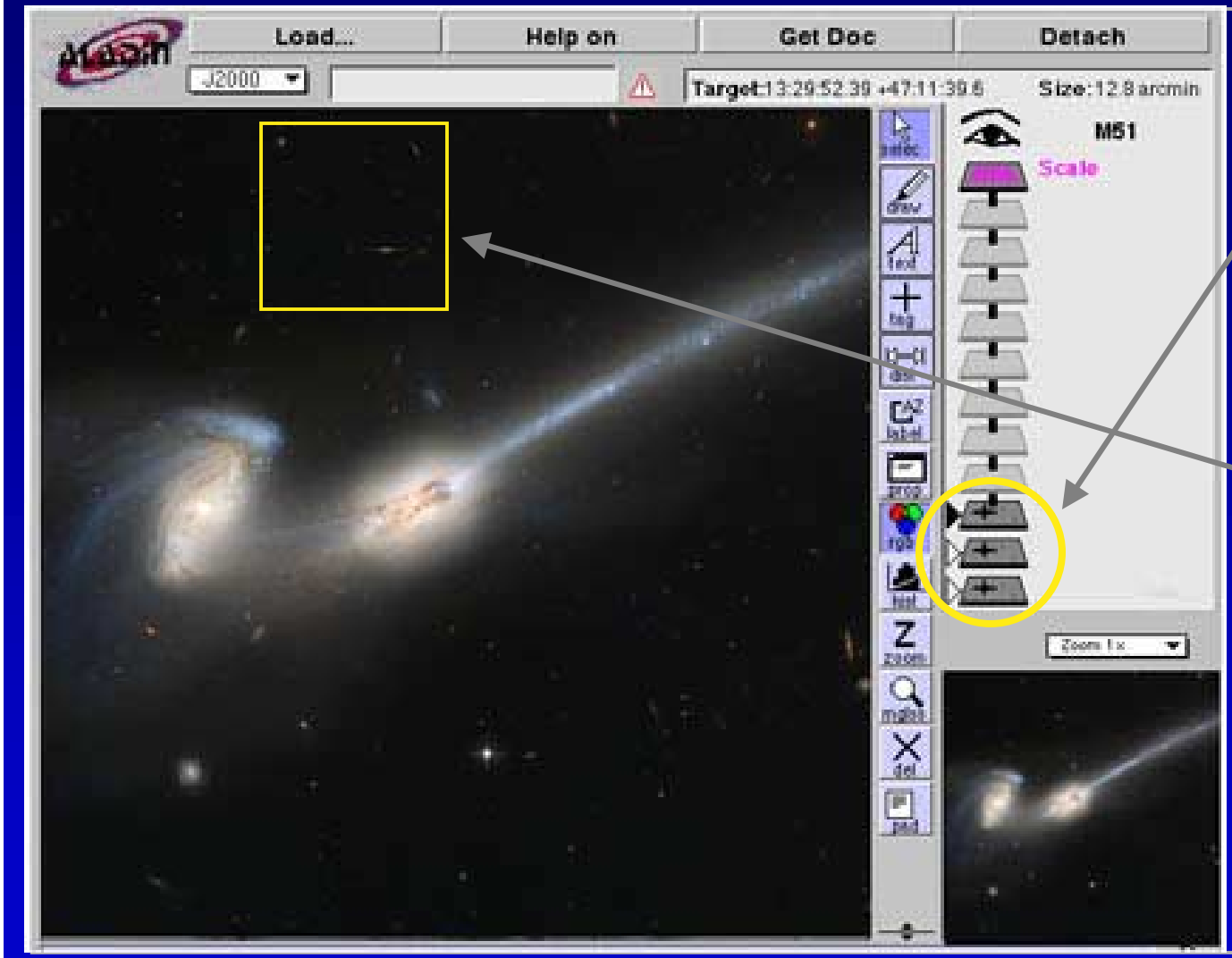
1. Science Use Cases and Requirements
2. Interoperability deployment and demonstration
3. Technology needs
   - GRID systems
   - Scalable storage and computation
   - Databases

AVO at the bottom

- A self-documented data hypercube



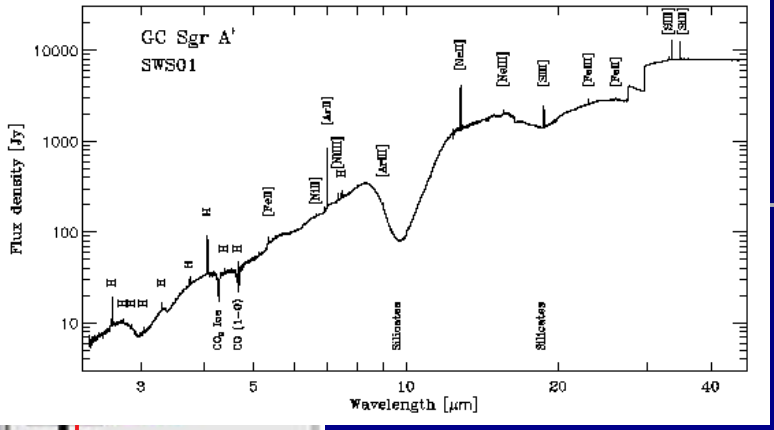
An evolving prototype

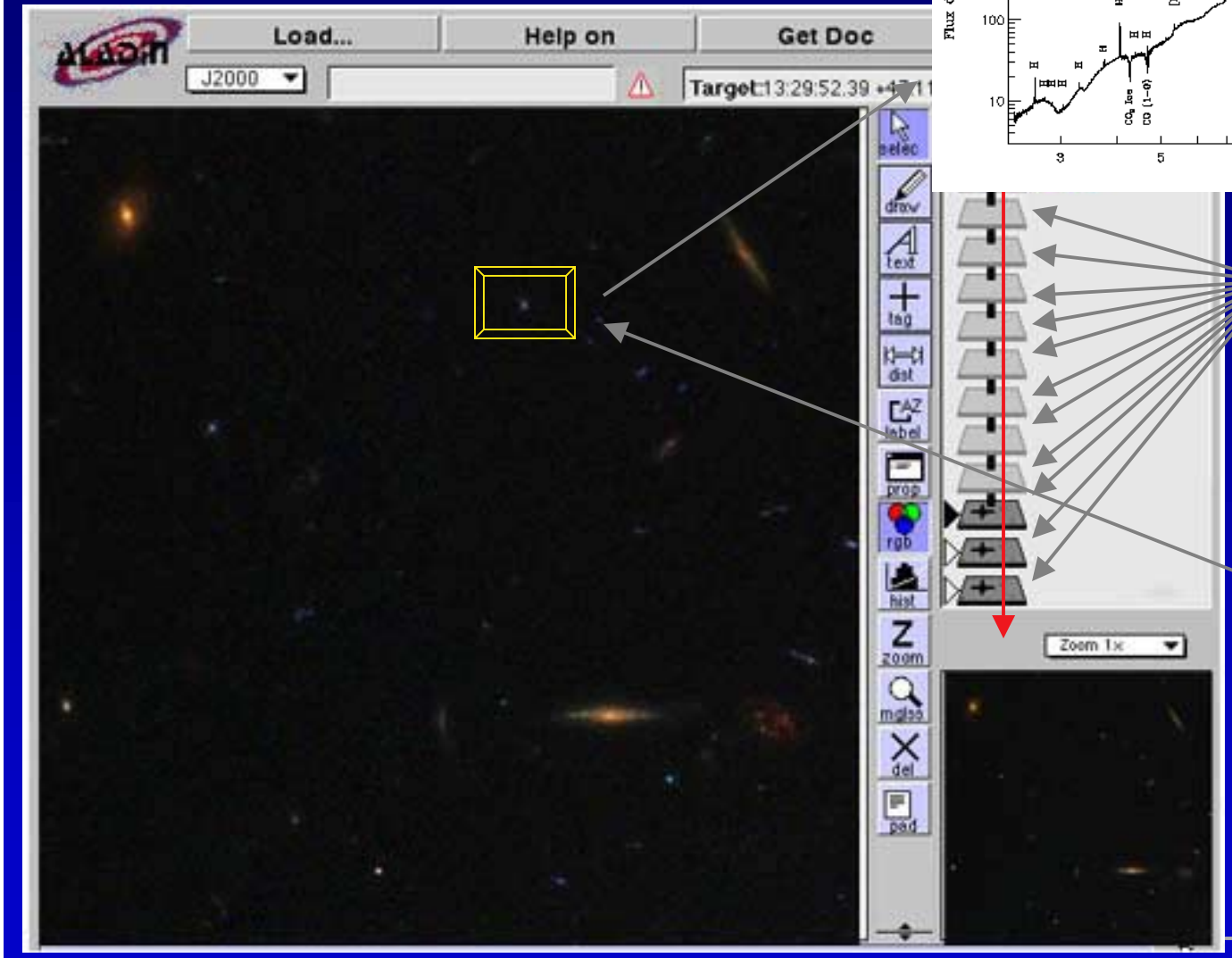


...a suitable choice of layers for previewing and browsing

...make a cut-out

The cut-out



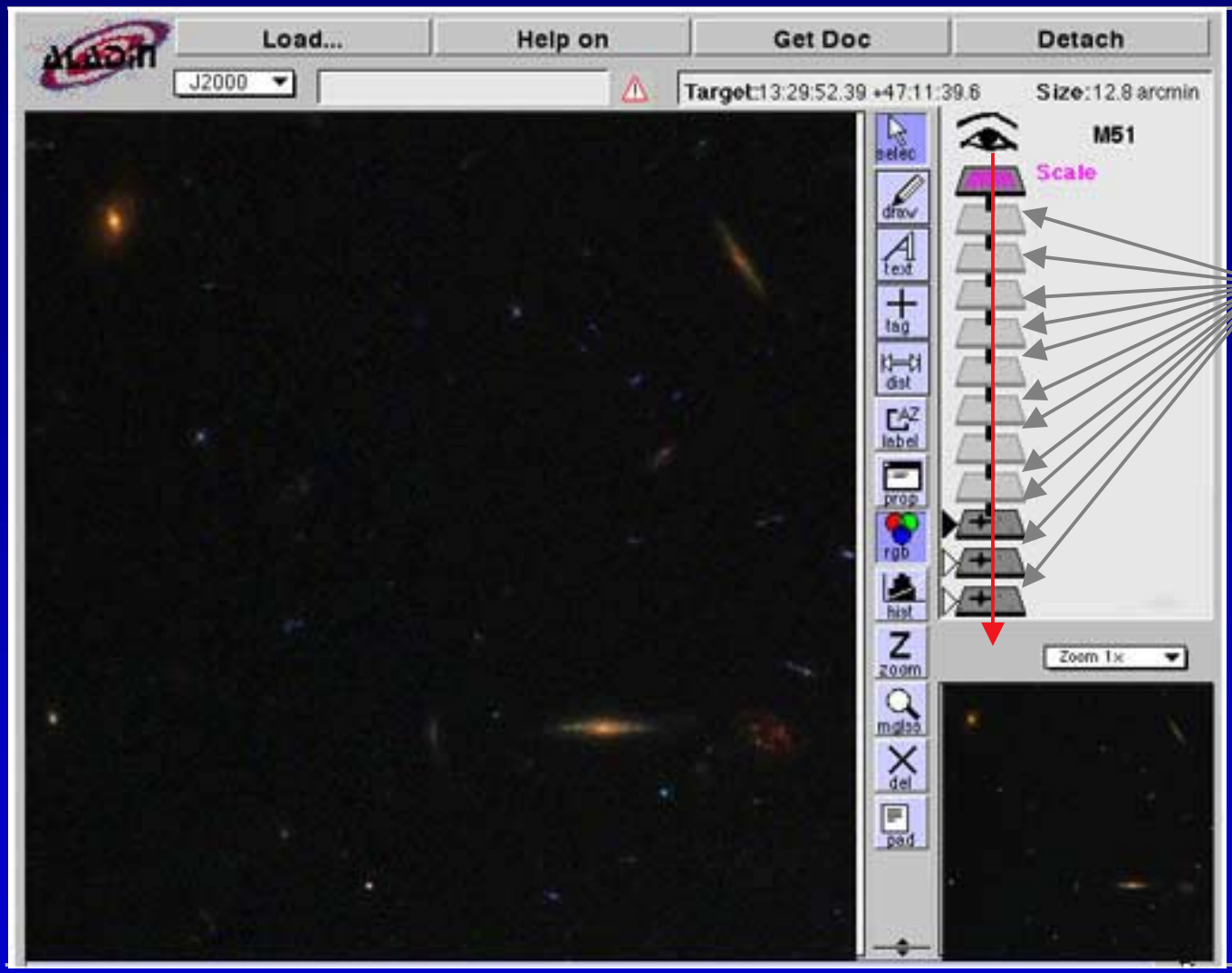


layers are loaded automatically

...further tools will provide additional capabilities

- sliding cursor gives an interactive Spectral re-extraction
- fitting Energy sets Distribution

AVO & GRID



…the multi-wavelength layers will be accessed remotely

…the data processing approach will be offered as a web service

Major International VO Roadmap Milestones

- First Science Demonstration Jan'03
  - Simple, limited data sets, use VOTable
- Global science demo. August'03 (IAU)
- Intermediate demo. Jan'04
  - First use of grid-like capabilities
- Complex science demo. Jan'05

AVO Web Pages

http://www.eso.org/projects/avo