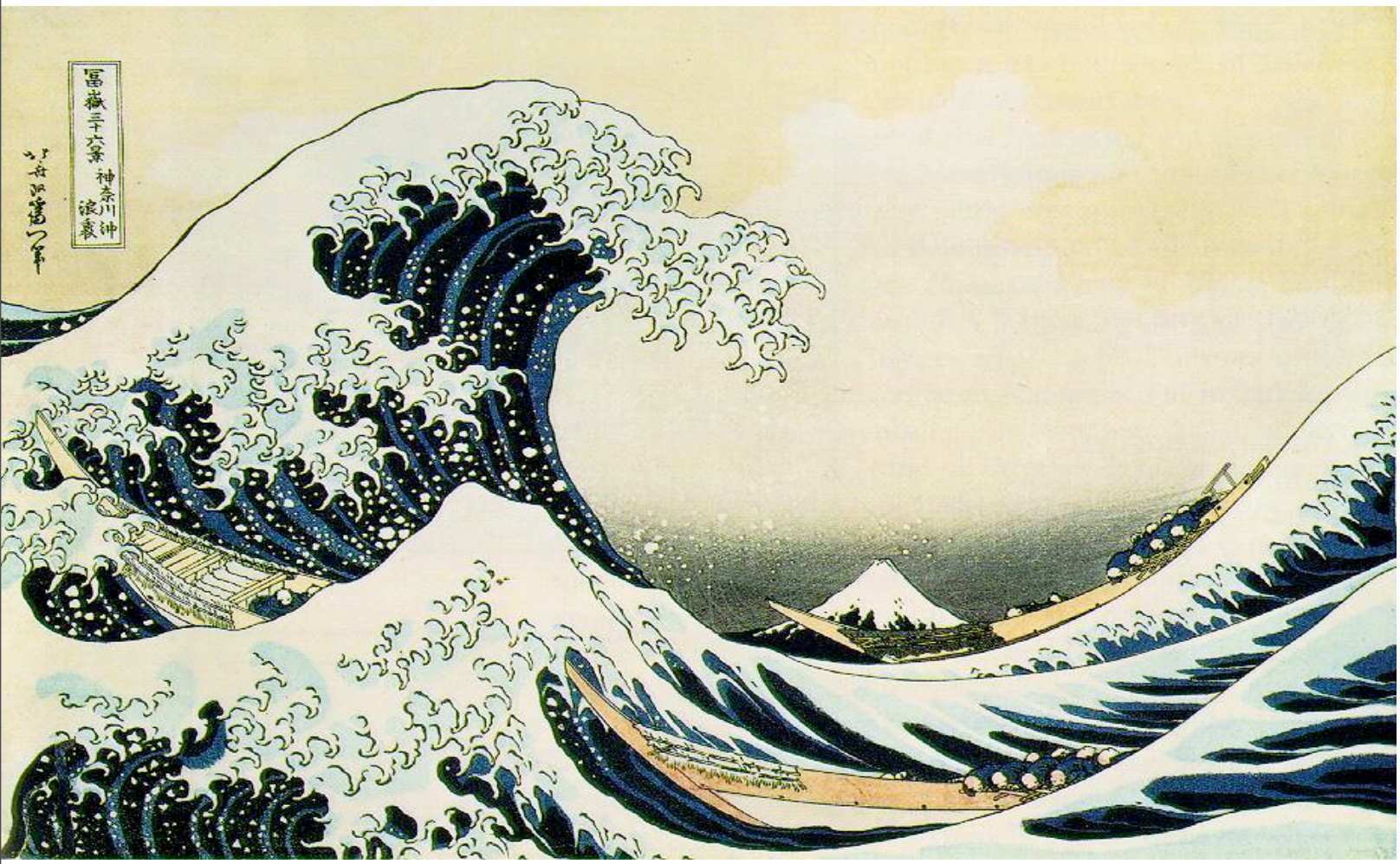
# Making the Grid and the Virtual Observatory Interoperable

*Dr. Giuliano Taffoni (INAF Trieste, Italy)*

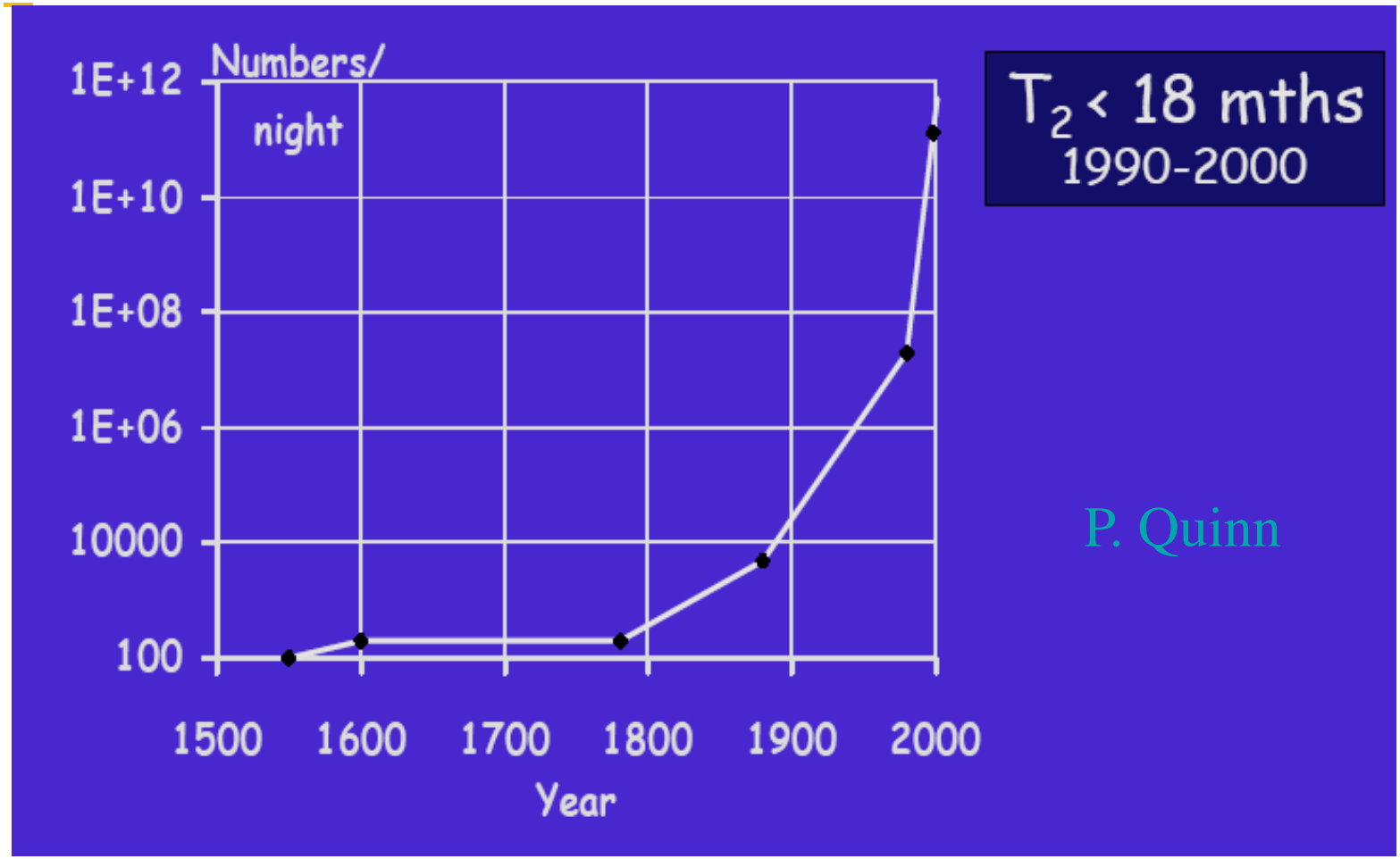
# Data Avalanche

- **Astronomy is Facing a Major Data Avalanche:**
  - Multi-Terabyte Sky Surveys and Archives (Soon: Multi-Petabyte), Billions of Detected Sources, Hundreds of Measured Attributes per Source …



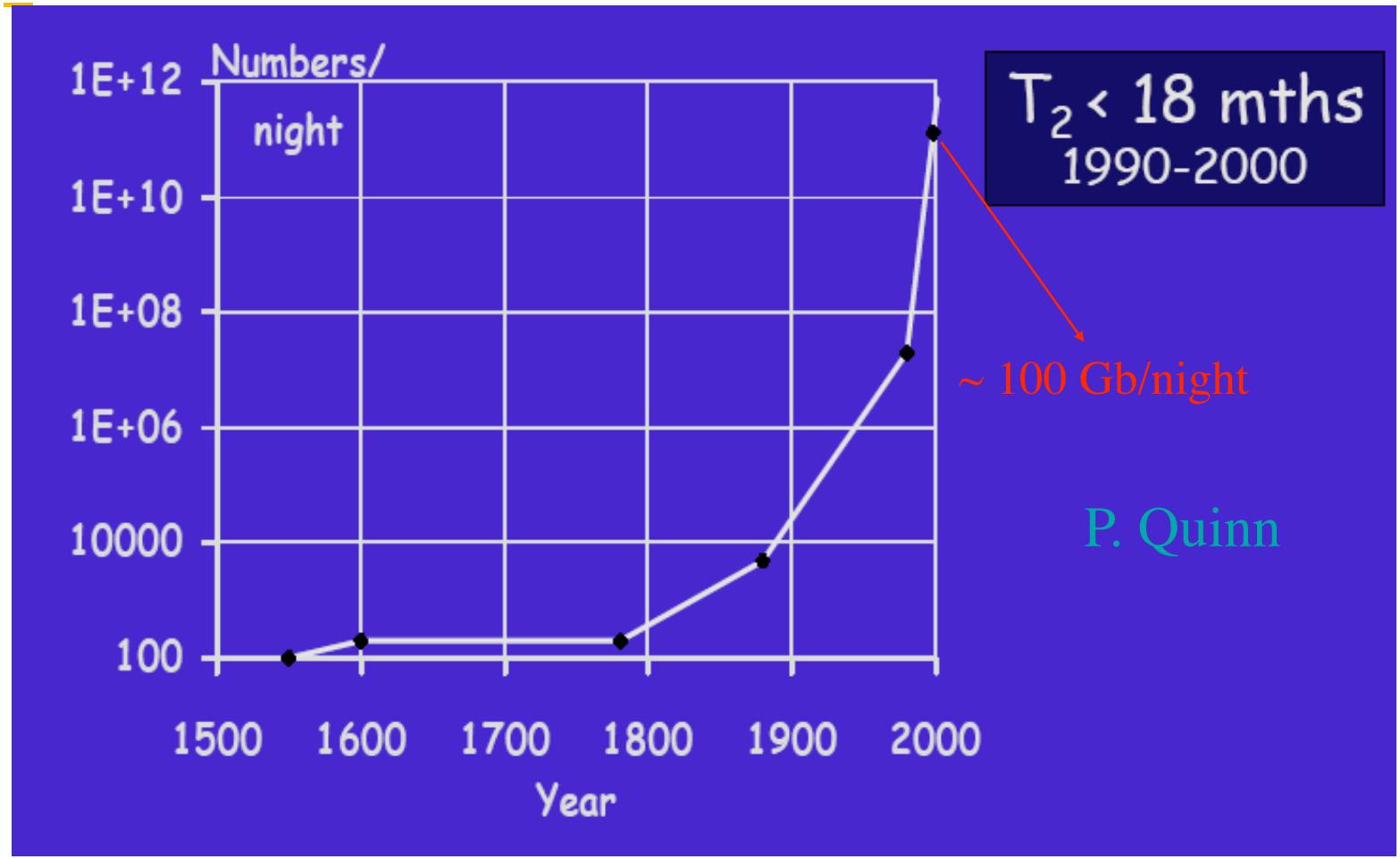
# New Trends in Astronomy

- **Astronomical Data is growing exponentially**



# New Trends in Astronomy

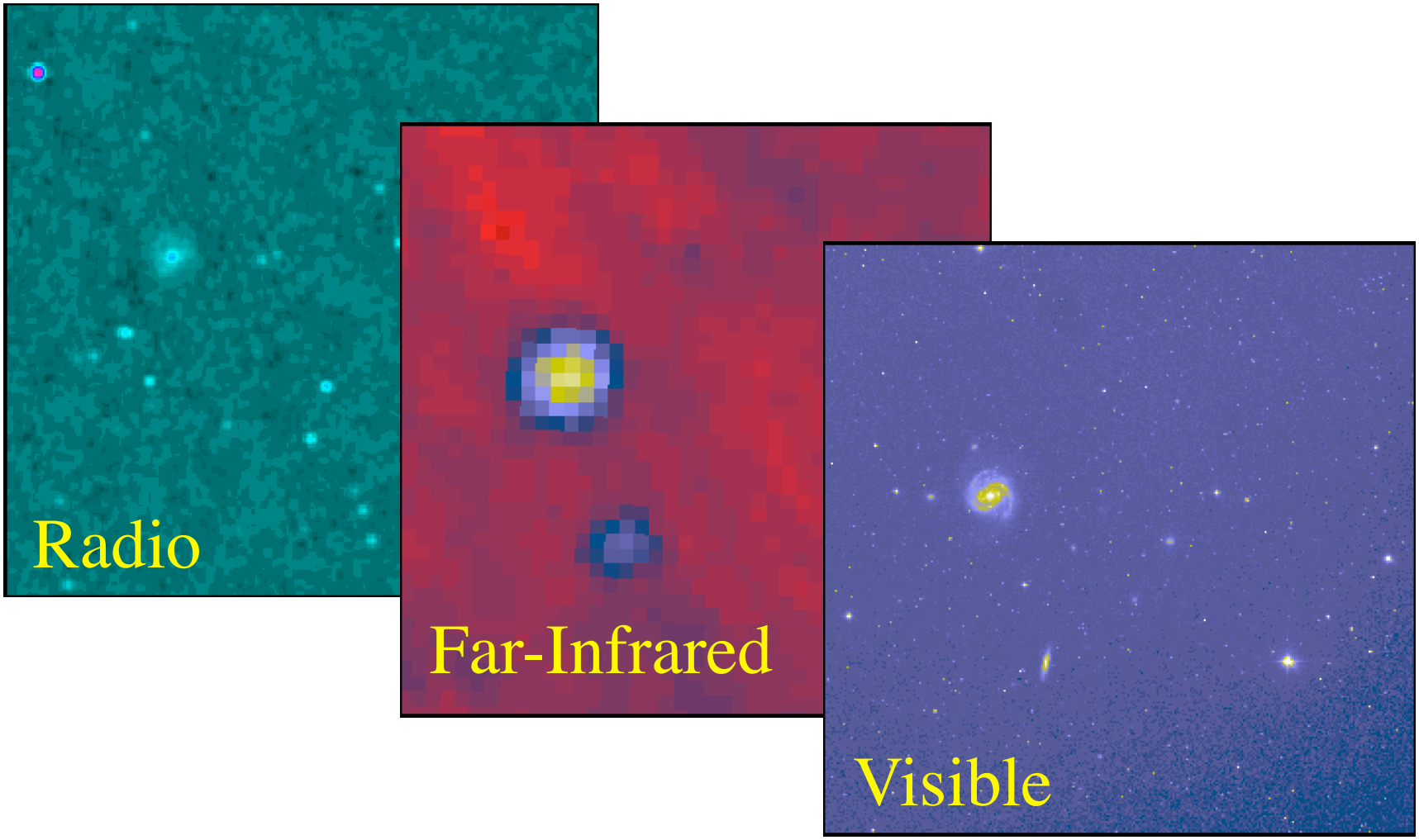
- **Astronomical Data is growing exponentially**



# The Challenging Aspects...

- Large digital sky surveys are becoming the dominant source of data in astronomy: currently > 100 TB in major archives, and growing rapidly
- Typical sky survey today: ~ 10 TB of image data, ~ $10^9$ detected sources, ~ $10^2$ measured attributes per source
- Data sets orders of magnitude larger, more complex, and more homogeneous than in the past
- Roughly 1+ TB/Sky/band/epoch
  - NB: Human Genome is ~ 1 TB, Library of Congress ~ 20 TB
- Spanning the full range of wavelengths, radio through x-ray: a panchromatic, less biased view of the universe
  -

# ...and a panchromatic view



# The Breakdown in Making Science

- Understanding of Complex Astrophysical Phenomena Requires Complex and Information-Rich Data Sets, and the Tools to Explore them...
- This will lead to a change in the nature of the Astronomical Discovery Process...
- which requires a novel research environment for Astronomy

## The Virtual Observatory

# The Virtual Observatory concept

- A response of the astronomical community to the scientific and technological challenges posed by massive data sets
- Federate the existing and forthcoming large digital sky surveys and archives, and provide the tools for their scientific exploitation
- A dynamical, interactive, web-based research environment for the new astronomy with massive data sets
- Technology-enabled, but science-driven

# The Virtual Observatory concept

- **A response of the astronomical community to the scientific and technological challenges posed by massive data sets**
- **Federate the existing and forthcoming large digital sky surveys and archives, and provide the tools for their scientific exploitation**

- **What is a Virtual Observatory?**
  - Dynamic collection of hardware, data and software working in harmony to solve arbitrarily large and complex astronomical problems. The VObs is the middleware and tools for Astronomers.

# So what...?

- **Vobs opened  new perspectives in Astronomy:**
  - Web: all documents of the world inside your computer
  - VO: all astronomical databases in the world inside your computer

- **Concrete example:**
- **Find all the observations of a given source available in all astronomical archives in a given wavelength range**
- **Tell me which ones are in raw or processed form**
- **Allow me to retrieve them**
- **If raw, give me access to the tools to reduce them on-the-fly Very time consuming, if at all possible, at present**

# So what...?

- **Vobs opened  new perspectives in Astronomy:**
  - Web: all documents of the world inside your computer
  - VO: all astronomical databases in the world inside your computer

- **Concrete example:**
- **Find all the observations of a given source available in all astronomical archives in a given wavelength range**
- **Tell me whi m**
- **Allow me to**
- **If raw, give em on-the-fly**
  **Very time consuming, if at all possible, at present**

**What is missing?**

# Virtual Observatory Grid

- **The Virtual Observatory is a Grid;**

# Virtual Observatory Grid

- The Virtual Observatory is a Grid;

# Virtual Observatory Grid

- The Virtual Observatory is a Grid;

**Objection, objection, objection, objection**

# Grid Essential

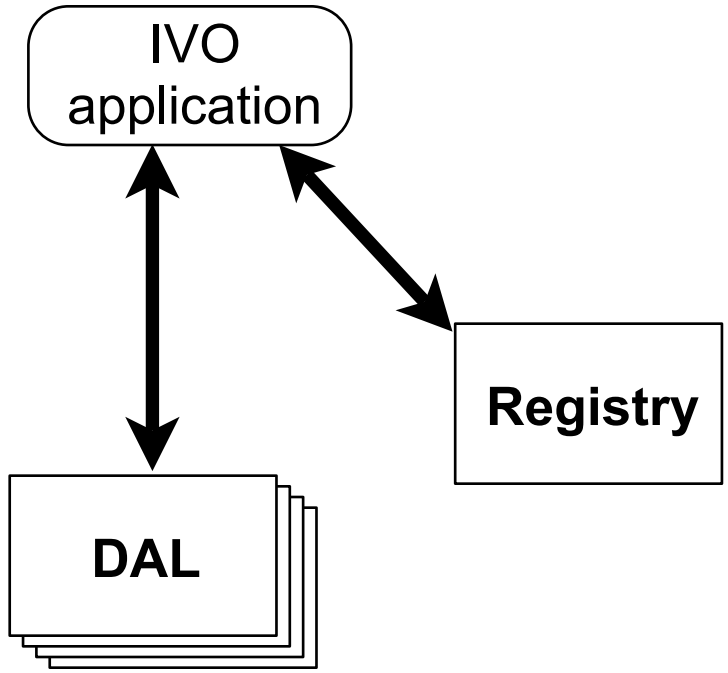
- “You can't be a real country unless you have a beer and an airline. It helps if you have some kind of a football team, or some nuclear weapons, but at the very least you need a beer”.

  » Frank Zappa

- You can't be a real Grid unless you have a **commodity** and a **discovery** mechanism. It helps if you have some kind of **middleware** or some supercomputers, but at the very least you need a commodity.

# Virtual Observatory is a Grid

- [x] Commodity
- [x] Discovery
- [x] Middleware
- [ ] Computational resources
- [ ] Supercomputers



# What is it missing?

- I need to make complex data miming calculations on my data.
- I need to compare Observational data with the result of my Code...
- ...but I need to run it somewhere
- I need to make some data reduction.

- Theoretical Virtual Observatory
  - Theoretical data produced on the fly
  - Comparison between theory and observation

# What is it missing?

- I need to make complex data miming calculations on my data.
- I need to compare Observational data with the result of my Code...
- ...but I need to run it somewhere
- I need to make some data reduction.
- Theore
  - Theo
  - Comp

**Maybe I can use the Grid!**

# EuroVO: the VObs in Europe

- The idea of the Euro-VO is to make it feel as if all the astronomical data and tools are available on the astronomers desktop, even though they are actually located on systems spread out over the whole of Europe and even the rest of the world.
- **EuroVO TECH**
  - responsible for completing the design work and feasibility studies on the backbone software components that will make the Euro-VO possible.
- **EuroVO DCA**
  - Coordinate and assist European Data Centers;
  - Produce a knowledge GRID (data + services)
  - **Coordinate with national and international GRID projects**

# EuroVO DCA

- Interest area: massive and distributed computing, Grid computing;
- Promote coordination between GRID(s) and VObs;
- Point of view of Data Centers;
- GRID(s) through Data Centers.

# In practice

- How can Data Centers benefit of of GRID computing?
- How can Astronomers can benefit of Grid computing?

# Partners

| Name | Affiliation | Project |
|---|---|---|
| G. Taffoni | INAF | DCA |
| M. Sponza | INAF | DCA |
| P. Osuna | ESAC | DCA |
| R. Alvarez Timon | ESAC | DCA |
| G. Lemson | MPG | DCA |
| J. Zuther | MPG | DCA |
| H. Enke | IAP | AstroGrid-D |
| E. Solano | INTA | DCA |
| J. Santander Vela | IAA-CISC | Spanish Grid |
| A.Schaaff | CNRS | DCA |
| K. Noddle | LU | DCA |
| G. Rixon | IAC | AstroGrid |
| E. Valentyn | NOVA | DCA |
| A. Belikov | NOVA | DCA |

EGEE VO: dca.euro-vo.org

# Coordination

- Keywords
  - Interoperability
  - Usability
  - Re-usability
- Useful Informations from GRIDs:
  - tools and services already developed!
  - problems already faced
  - dead-end already encountered
- Why not to use them instead of re-inventing?
  - the SSO example (see later)

# Making VObs and Grid interoperate

- Auth & Auth
- Data Management
- Job Management
- Information Systems

- Single-sign-on
- VOSpace
- Workflow
- Registries

# A first problem

# Auth & Auth

# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS
  - Shibboleth
  - etc...

# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS
  - Shibboleth
  - etc...
- Single-sign-on

# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS

# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS
  - Shibboleth

# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS
  - Shibboleth
  - etc...

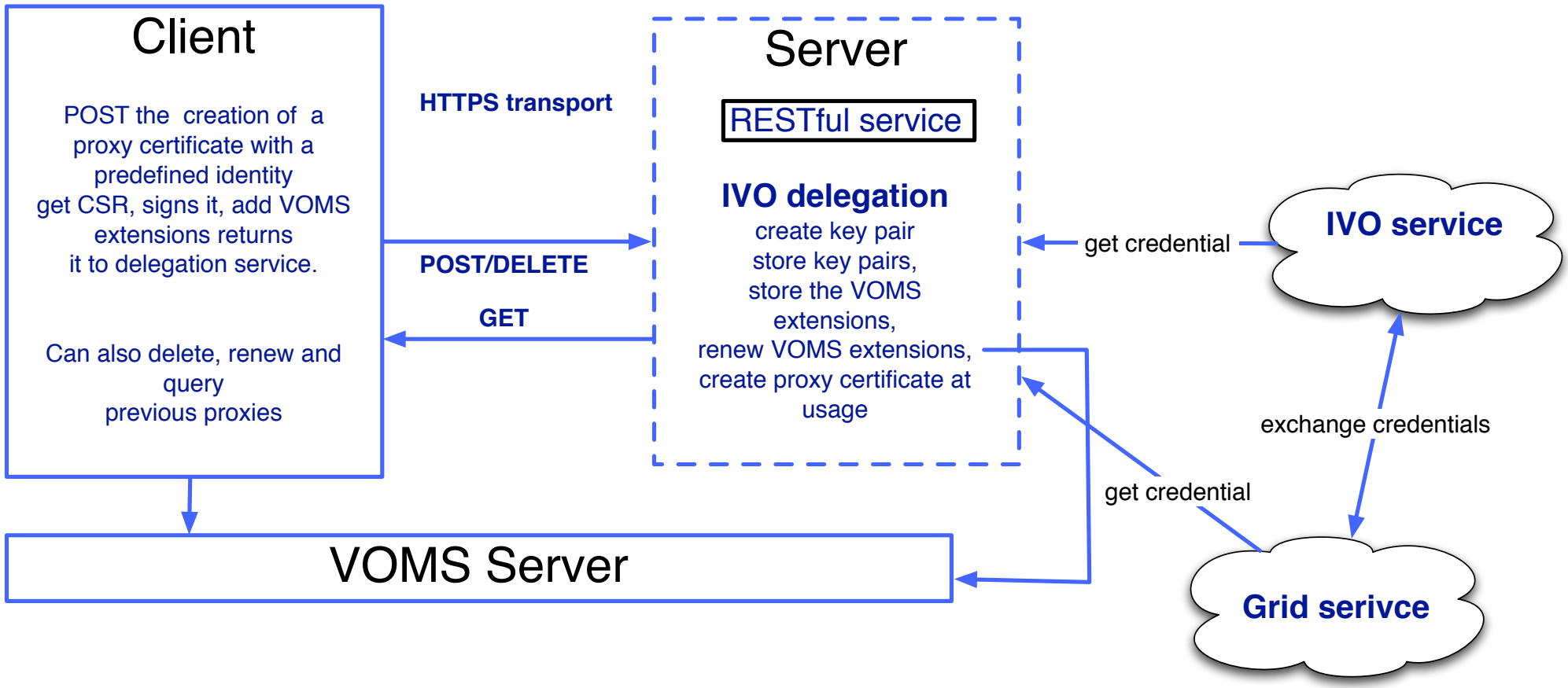
# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS
  - Shibboleth
  - etc...
- Single-sign-on-double-functionality

# Auth & Auth

- Authentication and authorization mechanisms:
  - VOMS
  - Shibboleth
  - etc...
- Single-sign-on-double-functionality
- User certificate, host certificate, application certificate...

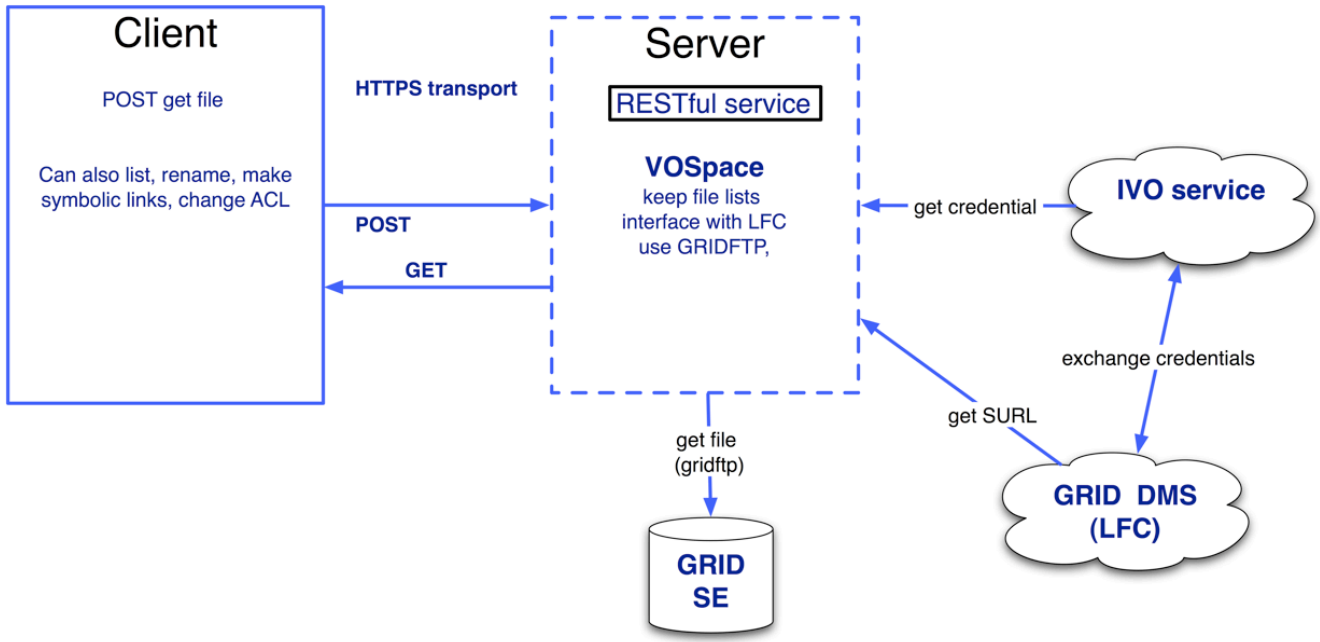
# IVObs Delegation Service



# Data Management

- **VOspace is the Data access protocol for VObs**
- **Web service interface**

- **The problem:**
  - share data from Grids and VObs.
  - we have EU Grid infrastructure that provides storage space for the Astronomical Data



# VObs and the Grid: WMS

- **Work in progress to interface the prototype of workflows systems with gLite WMS.**

- **NOTICE:**
  - different approach: it is not the user that submits its job but the Data centre that provides a service (...an application)

# Conclusions

- To build the bridge between the Grid and the VObs it is necessary to make interoperable the suite of standards and web services of the VObs with tools and services of the Grid
- The work in progress impacts some key aspects like:
  - authentication and authorization mechanisms to gain access to VObs resources (data) and Grid resources through a single authentication transaction (single sign-on);
  - access to both VObs resources and Grid resources simultaneously, in a transparent way to the final user and in both directions (from the VObs to the Grid and from the Grid to the VObs)
- two approaches foresee the ability to provide “wrapped” science applications, either legacy code or new, as services in an application server (from the VObs to the Grid) or the ability to federate VObs components (astronomical databases) as embedded resources of the Grid

# Conclusion

- **Interoperation is a two way procedure**
  - we are taking care that VObs adopts Grid (de jure/de facto) standards
  - The Grid should account for VObs requirements.
- **We should meet and discuss:**

# Conclusion

- **Interoperation is a two way procedure**
  - we are taking care that VObs adopts Grid (de jure/de facto) standards
  - The Grid should account for VObs requirements.
- **We should meet and discuss:**

**EuroVO and Grid workshop**

**9-11 April 2008 Garching (Munich)**

**http://wwwas.si.inaf.it/eurovow2008**

# Acknowledgements


- Prof P. Padovani
- Prof. P. Queen
- Prof. F. Genova
- and the IVOA (www.ivoa.net)