Virtual Observatory: Observational and Theoretical data

Patrizia Manzato

INAF-Trieste Astronomical Observatory
Astronomy is facing a data avalanche

Multi-Terabyte (soon: multi-Petabyte) sky surveys and archives over a broad range of wavelengths

Billions of sources, hundreds of attributes per source
The changing face of observational astronomy

- **Large digital sky surveys** are becoming dominant source of data in astronomy: > 100 TB, growing rapidly:
  - SDSS, 2MASS, DPOSS, GSC, FIRST, NVSS, RASS, IRAS, QUEST, GALEX, SST; CMBR experiments; Microlensing experiments; NEAT, LONEOS, and other searches for Solar system objects;
  - Digital libraries: ADS, astro-ph, NED, CDS, NSSDC;
  - Observatory archives: HST, CXO, space and ground-based;
  - Future: PanSTARRS, LSST, and other synoptic surveys; astrometric missions, detectors;

- **Data sets** orders of magnitude larger, more complex, more homogeneous than in the past;

- **Roughly 1 TB/Sky/band/epoch**
  - Human Genome is < 1 GB, Library of Congress ~ 20 TB
Toward a “new astronomy”

- Past: Observations of small, carefully selected samples (often with \textit{a priori} prejudices) of objects in one or a few wavelength bands
Toward a “new astronomy”

Future: Multi-wavelength data for millions of objects, allowing us to:

- Discover significant patterns from the analysis of statistically rich and unbiased image/catalog databases;
- Understand complex astrophysical systems via confrontation between data and sophisticated numerical simulation;
Toward a “new astronomy”

- Discovering new phenomena and patterns in these datasets will require simultaneous access to multi-wavelength archives, advanced visualization and statistical analysis tools.
The Virtual Observatory is...

- A set of international standards to share complex data;
- A modular set of tools to work with distributed data;
- A simple environment to publish data to;
- An essential part of the research astronomer’s toolkit;
- A catalyst for world-wide access to astronomical archives;
- A vehicle for education and public outreach;

R.J. Hanisch & P.J. Quinn, “International Virtual Observatory Alliance”, http://www.ivoa.net/pub/info/
The Virtual Observatory is NOT...

- A replacement for building new telescopes and instruments;
- A centralized repository for data;
- A data quality enforcement organization;
VO data

I step: VO started dealing with observational data, see "All VO" in Aladin tool;

II step: now in the VO there is some prototypes for theoretical data, see VOTech project:
- cosmological simulation;
- stellar Spectra simulation;
- Tracks and isochrones simulation BaSTI;

You can provide your collection of data into the VO;
IA² - TNG / LBT

Italian Astronomical Archive center
http://wwwas.oats.inaf.it/IA2/

- TNG ~ 3500 CD / 270000 (~10 MB one) images of the TNG (Galileo National Telescope)
- LBT - SDT (Large Binocular Telescope - Science Demonstration Time) blu camera, red camera: 1000000 (~80 MB) images;
- will arrived LBT data;

M51, TNG image

06 Dec 2007
Winter School, P.sso Tonale - P.Manzato
IA² – Theoretical VO

http://wwwas.oats.inaf.it/IA2/ITVO/ and http://itvo.oact.inaf.it

- **Gadget2** N-body+SPH,
  ~1.2 TByte,
  40000 CPU hours on 64 processors of the IBM-SP4 machine,
  102 snapshots;
- **Enzo** N-body+AMR;
- **Fly** N-body;
- Work in progress: a new web portal for the BaSTI stellar evolution simulation performed with **FRANEC** code.

Temperature Graphic of simulated galaxy clusters
GRID and HPC

Two infrastructures for running expensive CPUs calculations:

- **GRID** ([http://wwwas.oats.inaf.it/grid/](http://wwwas.oats.inaf.it/grid/))
  a computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities. (*Carl Kesselman, Ian Foster “The Grid: Blueprint for a New Computing Infrastructure” 1998*)
  
  Contacts: vuerli@oats.inaf.it for the Grid workerNode in Trieste
  i.e.: Planck pipeline tests performed with Grid infrastructure.

- **HPC**=High Performance Computing, CPUs cluster.
  i.e.: CINECA ([http://www.cineca.it/en/index.htm](http://www.cineca.it/en/index.htm)) is a Large Scale Facilities, supercomputer center.
  It was used for running Gadget2 simulation stored into IA2 web site.

You can ask for tutorial.
Frequently used VO tools

- **Aladin** (CDS, Strasbourg) 2-D images visualizer
  http://aladin.u-strasbg.fr/aladin.qml;
- **VisIVO** (CINECA) N-D data (HDF5, Gadget, VOTable, fits table, ASCII table, raw binary) http://visivo.cineca.it/;
- **TOPCAT** (Starlink) to create plots
  http://www.star.bris.ac.uk/~mbt/topcat/;
  (can query the DB of the Millennium simulation, G. Lemson et al., “Halo and Galaxy Formation Histories from the Millenium Run: Public realise of a VO-oriented and SQL-queryable database for studying the evolution of galaxies in the ΛCDM cosmology” Astrophy/0608019, 2006)
- **Specview** (STScI) spectra
  http://www.stsci.edu/resources/software_hardware/specview ;
- **VOSpec** (ESAC) spectra http://esavo.esa.int/vospec/ ;
- **Plastic Hub** to connect all these tools
  (http://www.ivoa.net/Documents/latest/PlasticDesktopInterop.html)
- Etc.
Aladin tool

Aladin is able to search and open observational and theoretical 2D images.
Spherical Cutout

A cutout of a FLY output file performed by Catania@ITVO http://itvo.oact.inaf.it/ and visualized with VisIVO tool (new features: http://eurovotech.org/twiki/pub/VOTech/DS6PlanningStage06/VisIVOSTage5report.pdf)
Conclusions

- the VO is the fundamental infrastructure that enable the astronomers of XXI century to use the totally investments of keeping for long time the astronomical data.
- so the VO becomes a key element for every new astronomical facility:
  - guarantee the maximum scientific return of investments
  - distribute the knowledge to all scientific community.
- Merge technological and scientific group to collaborate inside the IVOA;
- I invite you to try to use the VO tools. We want a feedback or suggestions from scientists!!!
  - e-mail to: IA2@oats.inaf.it or mailing lists of IVOA;