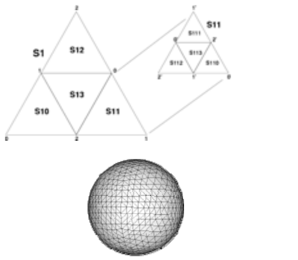
Concepts for single/multiwavelength astronomical data accessibility and reduction via relational database: a powerful resource for old and new projects

Luciano Nicastro, IASF-INAF, Bologna, Italy

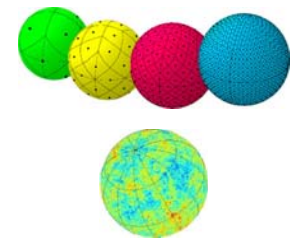


HTM: Hierarchical Triangular Mesh sky pixelization. Mainly used for objects catalogues. Good to perform cross-matching between databases. **"www.sdss.jhu.edu/htm/"**

Background

Nowadays midium-large size astronomical projects have to face the management of a large amount of information and data. Typically dedicated *data centres* manage the collection of raw and pre-processed data and consequently make them accessible to the (authorised) users. Access is performed either via (s)ftp or http (web) and typically foresees only files transfer. Also selecting the data of interest is usually performed acting on a few parameters (e.g. object name or coordinates). In a few cases, when large amounts of data are involved, no data transfer is allowed but the user can submit batch jobs that return the results of a particular analysis. In some cases the data are delivered to the user on tapes, DVDs, etc. In all cases the data acquisition, archiving, delivering, processing and the results accessibility are managed separately. Often the information are not collected into relational databases tables and when this happens, the delay between the date of collection and the archiving is of the order of days or even months. The same happens for the data production logging and project documentation. Luckily the use, in many cases, of standard file formats like FITS can help to track the data origin and processing status.

International projects like GRID and the International Virtual Observatory Alliance (IVOA) represent an effort to give a robust and standard framework to physicists and astronomers. However these projects size and ambitions cause them to proceed quite slowly and the potential users do not get immediate advantages from them. Small and medium projects/experiments tend to optimise the data management for their internal use only. Only international facilities (large optical telescopes, high energy satellites, etc.) put some effort into keeping a standard data format. This to make easier the exchange and exploitation of the data by using standard analysis packages.



HEALPix: Hierarchical Equal Area isoLatitude Pixelization. Mainly used for sky maps in CMB experiments but can also be used for any large area survey and sources catalogue. "**healpix.jpl.nasa.gov"**

Why a Database in Astronomy?

Here is a short list of answers:

- Can track in an ordered form what a project produces and let the rest of the world know it.
- Can manage all the information aspects of a project within a single framework.
- Don't need to worry about data management but concentrate on the analysis and interpretation.
- Make data accessibility uniform for all the Observatories/Projects from any internet computer.

In particular we have in mind a system with the following characteristics:

Input/Output

- Populating
  - From: any internet node
  - How: any s/w
  - What: data in various format
- Info/content log
  - From: any internet node
  - How: any s/w
  - What: tables in various format
- Retrieval
  - From: any internet node
  - How: any s/w
  - What: data, files, tables, graphics (1,2,3-d), images

User view

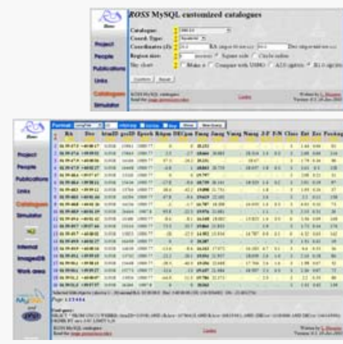
- Support
  - Ready to use pre-processed data
  - High level processing, automatic or *on demand*
  - Per user backup and restore
  - Data access security and replication
- Internet links
  - Direct access with custom s/w or Web based user interface
  - Virtual Observatory
  - Project Data Centre

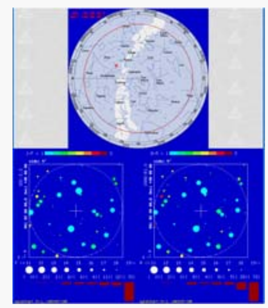
What do we propose?

The basic idea is that it is easier and more efficient to use databases for almost all the aspects related to a *modern* experiment/project in Astronomy. Archives with documents, s/w packages, data logs, observing schedules, objects catalogues, simulated data, quick look graphs, raw and processed data, *all* can be managed by a modern database server without caring about computer architecture, programming language, access security and even about data sharing, backup and restore.

We propose the use of a single package (see the **MCS** dedicated poster) to deal with all these aspects of a project. In particular an experiment status and management, its data acquisition and analysis results can be made accessible to any user on the internet in real time. Also the format of the selected data can be different depending on the user need. For example one can ask to get simply a PS graph without transferring the plotted data. The same if one wants to perform simple statistical analysis. In addition a user contributed library will allow more sophisticated analysis like simple fitting, sky mapping, coordinates and time conversions, astrometric calculations. Objects catalogues can be indexed either using the Hierarchical Triangular Mesh (HTM) or the Hierarchical Equal Area isoLatitude Pixelization (HEALPix). In other terms, the common tasks are performed on the *server* side whereas the *client* (the user computer) can concentrate on specific tasks on the retrieved data.

The fact that the user can use any programming language (or even an interactive shell) to interact with the database makes collaborations much more profitable and the (existing or new) software easily sharable. Data input/output can be performed in several standard formats, like XML, FITS and VOTable. The latter immediately makes accessible data and products to a Virtual Observatory. Noticeably communication between an **MCS** server and a VO allows a real time view and access to the Observatory products. In other words the Virtual meets the Real. We plan to perform "real" tests in collaboration with Institutions involved in the IVOA in the near future.





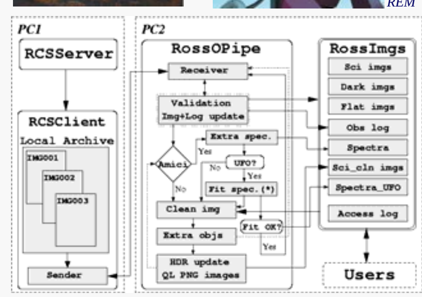
HTM indexed catalogues with ~10^9 objects each accessible from the internet: GSC 2.2, 2.3, USNO B1.0, 2MASS, UCAC2, … "**ross.iasfbo.inaf.it/"**

ROSS

ROSS is the **REM** (Rapid Eye Mount. robotic telescope) Slitless Spectrograph producing direct or dispersed (via an Amici prism) optical images. Observation logging and real-time image processing/archiving is performed accessing local and remote DBs. A schematic flow chart is shown below. As soon as the image is (pre)processed, it is available to the owner in the database. As usual it is accessible from any internet node. Typically the use of the dedicated web interface would be appropriate. Each user has his/her own account and can access only proprietary data. The observation log is freely accessible. As shown below, the web interface (written in PHP) allows a simple and fast access to the log and products (images and spectra). It is very easy to implement new facilities performing more tasks on images or spectra.

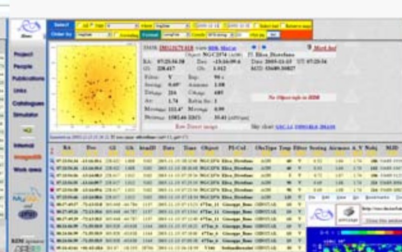
La Silla, Chile

REM

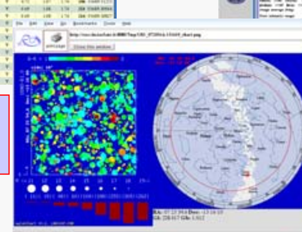


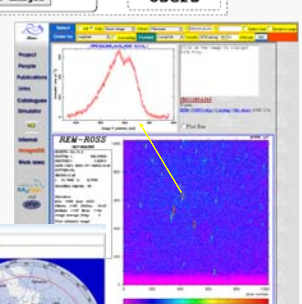
ROSS observation log: Updated in real time and freely accessible from the internet. **"ross.iasfbo.inaf.it/"**

ROSS images archive: Updated during daytime and accessible to authorised users (the images owner) only. Super-users privileged people can access all the images. Images selection and retrieval is very easy and does not use any extra disk space.



Interactive Analysis: Objects photometry and astrometry (approx) can be performed interactively via a custom version of **SExtractor**. Sky charts of the field, for various catalogues, need just one click.





Spectra: data (counts or flux) extraction and plotting requires just one click.