Workflows in Astronomy: General Requirements and new implementations in SCI-BUS project

Ugo Becciani

MIUR: Italian Ministry of Scientific Research

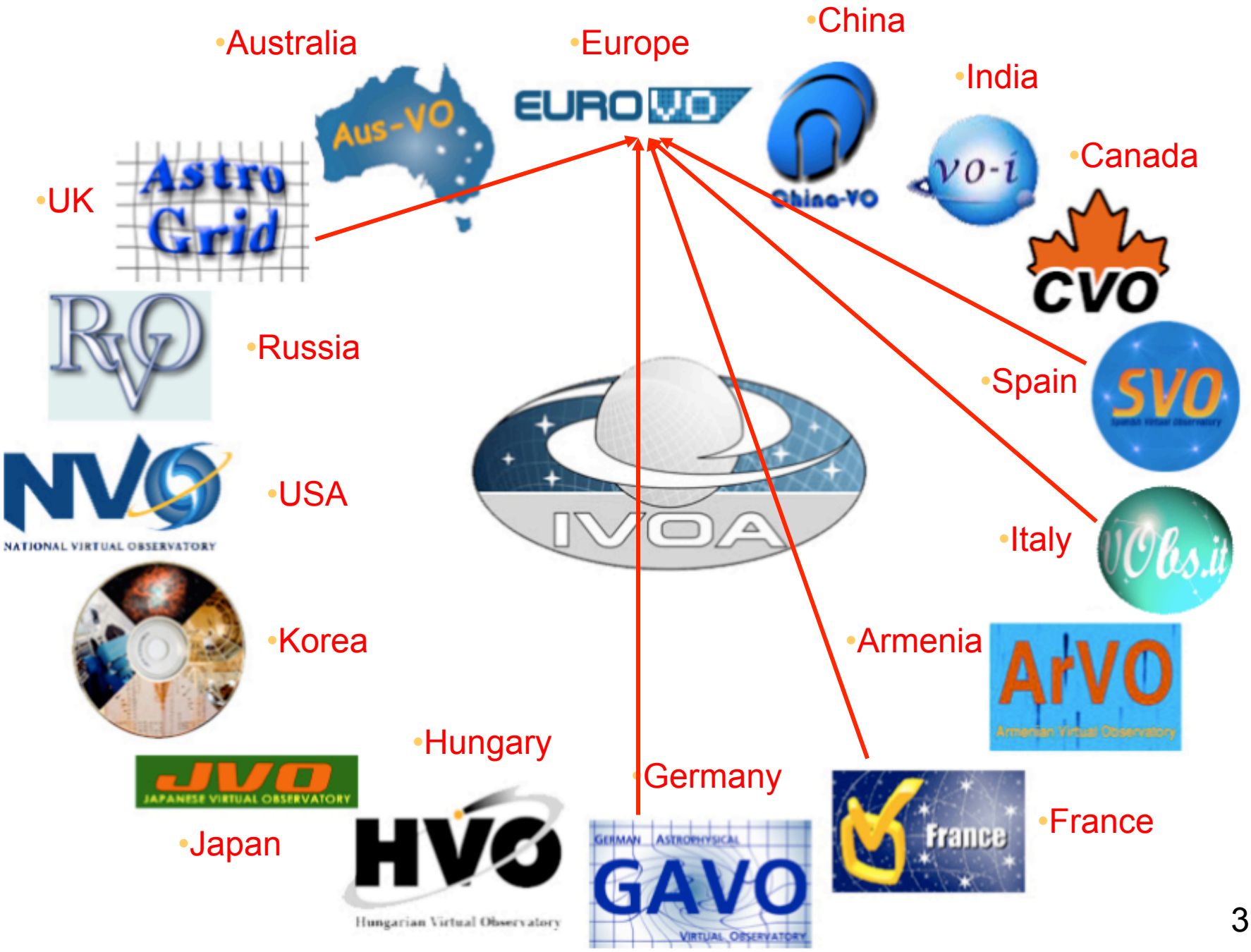
Workshops on e- Science Workflows
Budapest 9-10 Febraury 2012

What are the expectations of an astronomer on e-Infrastructures?

All scientists are equal… but are some "more equal" than others?

1. Need for democratisation of astronomy
   - Overcoming the digital divide, education
   - Providing chances to work with top-class data
   - Creating chances for new kinds of science
2. The Virtual Observatory as an initiative to evolve from astronomy to e-astronomy
   (specifically supported by the ASTRONET reports)

- 17 Member Organizations



The IVOA

- **Mission:** *"To facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory"*
- **Objectives:** standardization of data/metadata/sw, data interoperability methods, and list of available services (provided by national/international projects)
- Worldwide initiative started in 2002, 17 projects

Memories of the past

A&A community (VRC) using Grid technology since 2003:

- Pioneering the access to databases (GDSE)
- Pioneering instruments control via Grid (Astro-IE)
- ***Data Analysis and simulations***

- A&A participates to the EGEE, EGEE-II, EGEE-III, EGI-Inspire

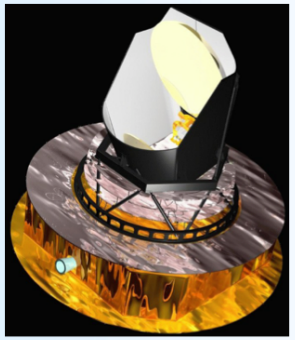
- Interoperability with other e-Infrastructures (VObs)

Successful stories

- Instrument Element: instrument control via Grid infrastructure.
- G-DSE: accessing databases via Grid
- A&A science gateway portal (2003!!!) for data reduction and analysis
- Several tools for cosmological simulations (FLASH, P-Gadget, Pluto, FLY etc)

The Early Days

- Planck Satellite Mission to measure the Cosmic Microwave Background (CMB) radiation:
  - Pre-launch simulations activities
    - Involves both EGEE and DEISA
  - Monte Carlo simulations (post-launch)

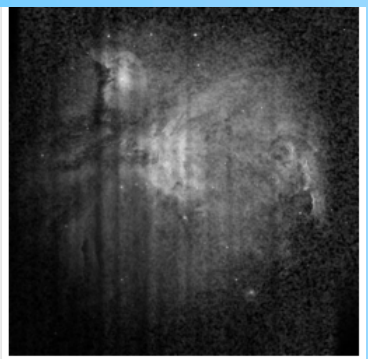


- MAGIC Telescope:
  - CORSIKA Monte Carlo code
    - Hadronic showers
    - Instrument response = simulated data sample

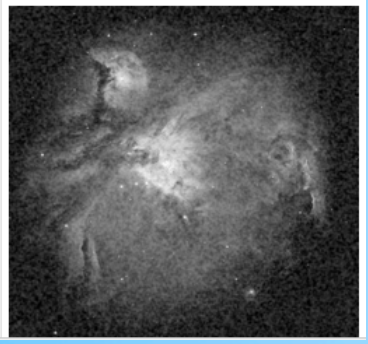
Workflows ➔ Pipeline data reduction (image calibration)

The data reduction process is important for photometry as the contributions from **dark** current, **bias** and **flat field** alter the number of counts in a frame.

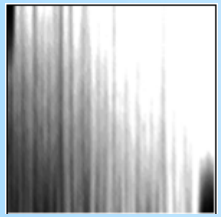
The general process for data reduction is outlined below with the aid of exaggerated mock results for illustrative purposes (image of M42 taken by the Hubble Space Telescope.)

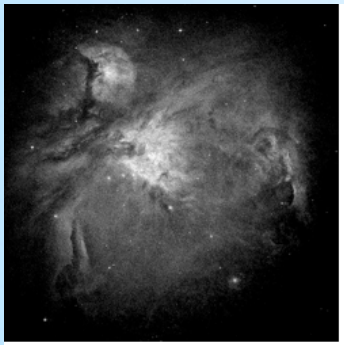


Raw Data

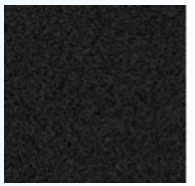


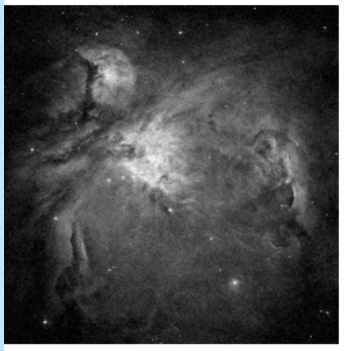
The first step to data reduction is to subtract a **master bias frame** for the same CCD temperature as the image





Next, a model of the dark current's contribution must be subtracted (master dark subtraction)





Finally, the image must be divided by the master flat frame. This removes any artefacts present due to uneven illumination of the CCD (generally vignetting due to the telescope)

Workflows ➔ Pipeline data reduction (spectra reduction)

The first part of spectra reduction (pre-reduction) has basically the same steps than the image reduction. The true extraction of the spectrum from the processed image proceeds according the following steps.

1. Trace of the spectrum along the pixel diagonals (low order polinomial fitting of the position of maximum intensity)

2. Summing up of all signal at the same wavelength position

3. Subtraction of the sky background spectrum formed at the border of the stellar spectrum

4. Wavelength calibration through the emission lines of calibration lamp spectrum (usually Th-Ar lamps). It determins the calibration relation pixel-wavelength

Workflows ➔ Pipeline data reduction (LSS cosmology)

The data reduction process on a large scale cosmological simulation

1) Build catalogue of raw particles' positions and velocities

2) Define a set of criterion for particle group extraction (density threshold, min/max number of particles, escape velocity, etc,)

3) Create a representation of the density field extracted from the particles position (density field representation)

4) Extract the groups (using the points 2 and 3 above)

5) Compute the group properties: center of mass position and velocities, global mass, inertia momenta etc.

6) Visualize the results

Workflows ➔ Pipeline data reduction (planetary transit search)

Corot and Kepler arte currently used for planetary transity search from space.

1) Reconstruction of the photometric time series from data taken in different conditions

2) Filtering data (direct or iterative): remove (reduce) intrinsic stellar variability due to stellar rotation, flares, eclipses by stellar companion, etc..

3) Box Least Square (BLS) or matched filters search for transit: light dips due to the planetary transit ➔ Estimate the probability of a trinsit-like signal: FAP (False Alarm Probability)

4) Select the target as an object of interest if FAP included in $[10^{-6}, 10^{-4}]$

Follow-up of the objects of interest (candidate planets).

VisIVO

Desktop

VisIVOServer

--fformat votable /home/user/demo/vizier.xml

.....

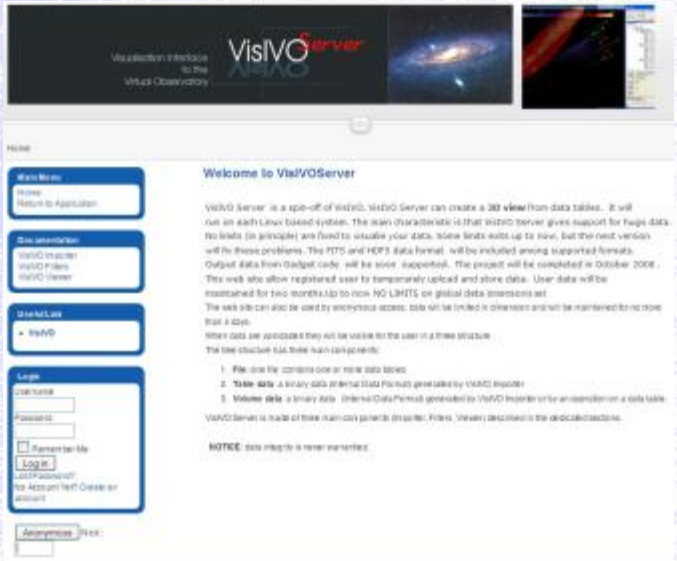
--x x --y y --z z --color --colortable --colorscalar scalar0 --glyphs sphere

Linux

Mac Osx

Windows

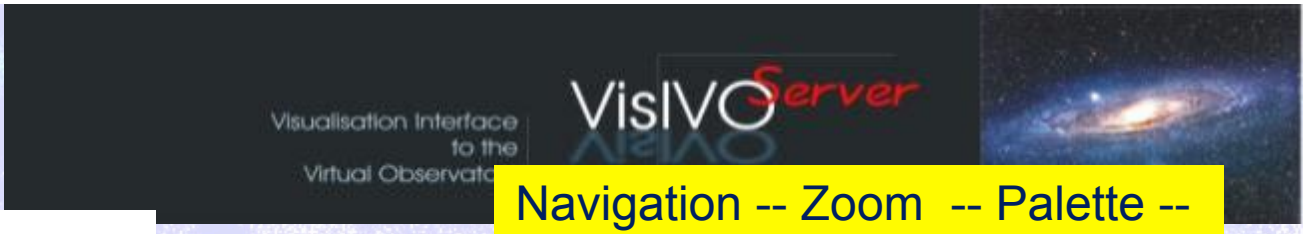
VisIVOWeb



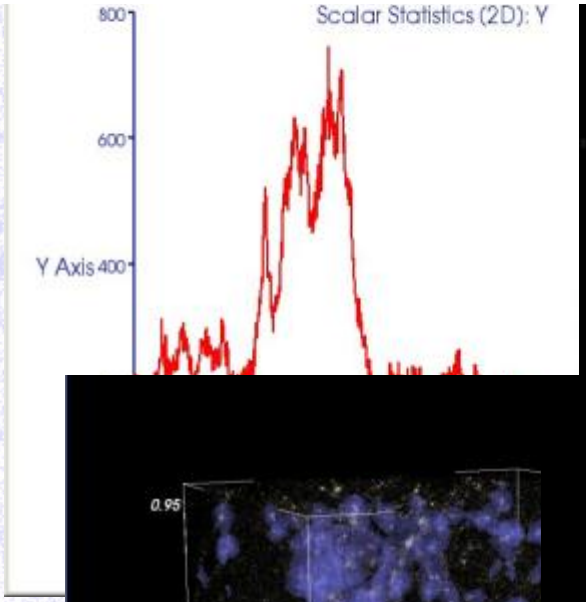
VisIVO C/C++ Library

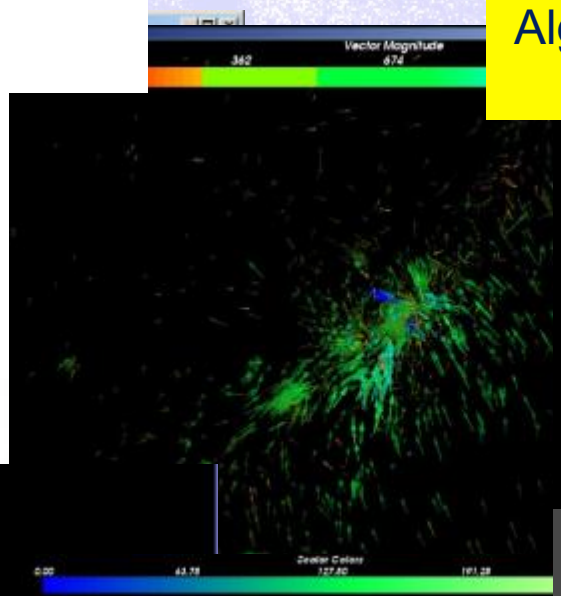
Closely integrated, complementary and independent !

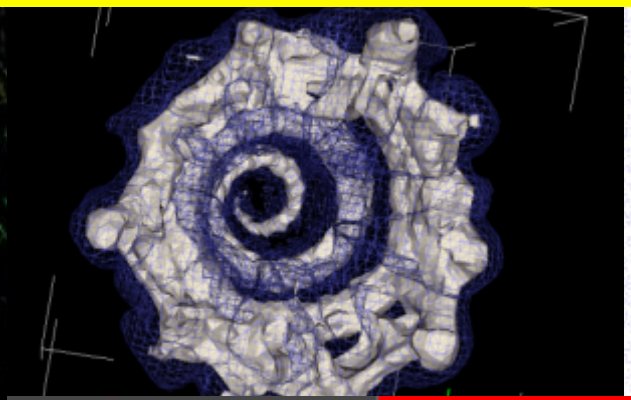
Visualisations

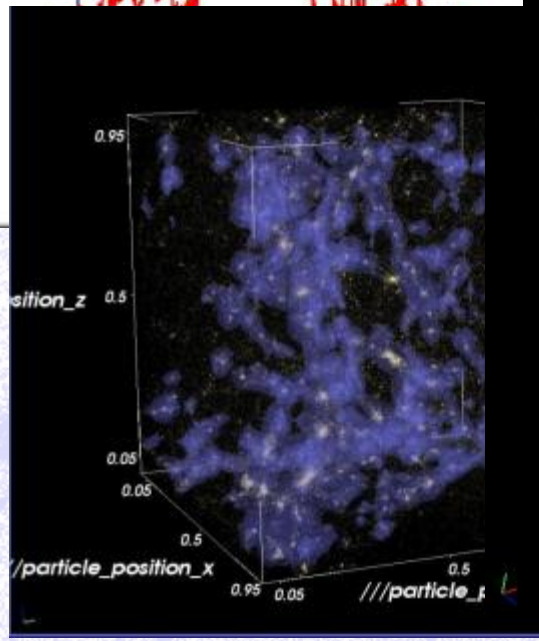


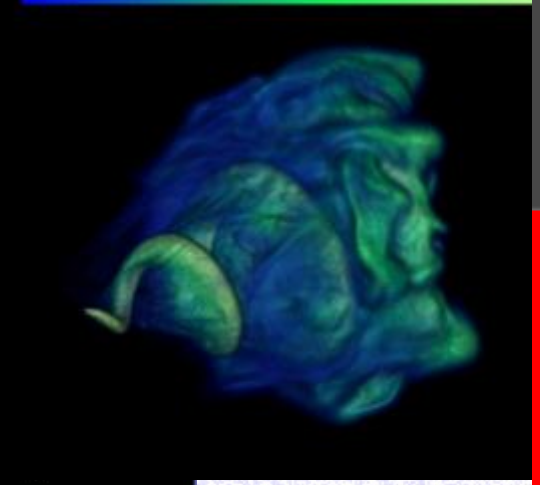
Navigation -- Zoom -- Palette -- Algorithms -- Data selection -- Picker op.

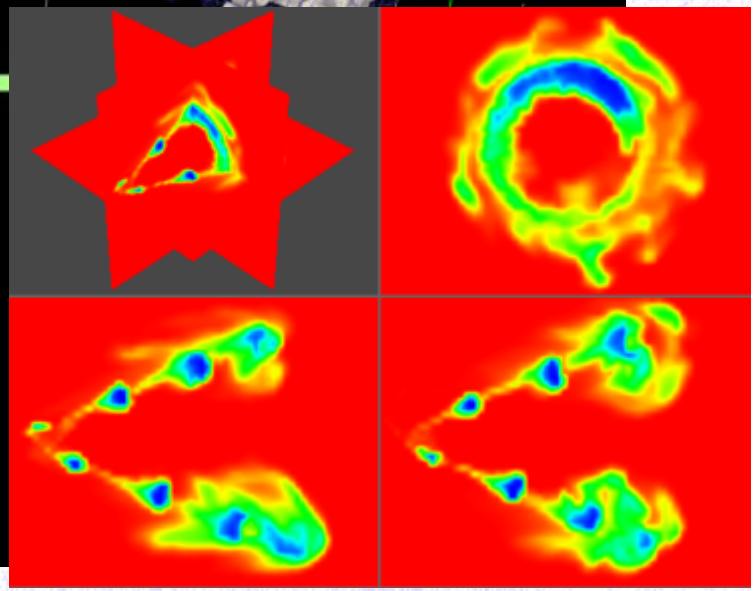






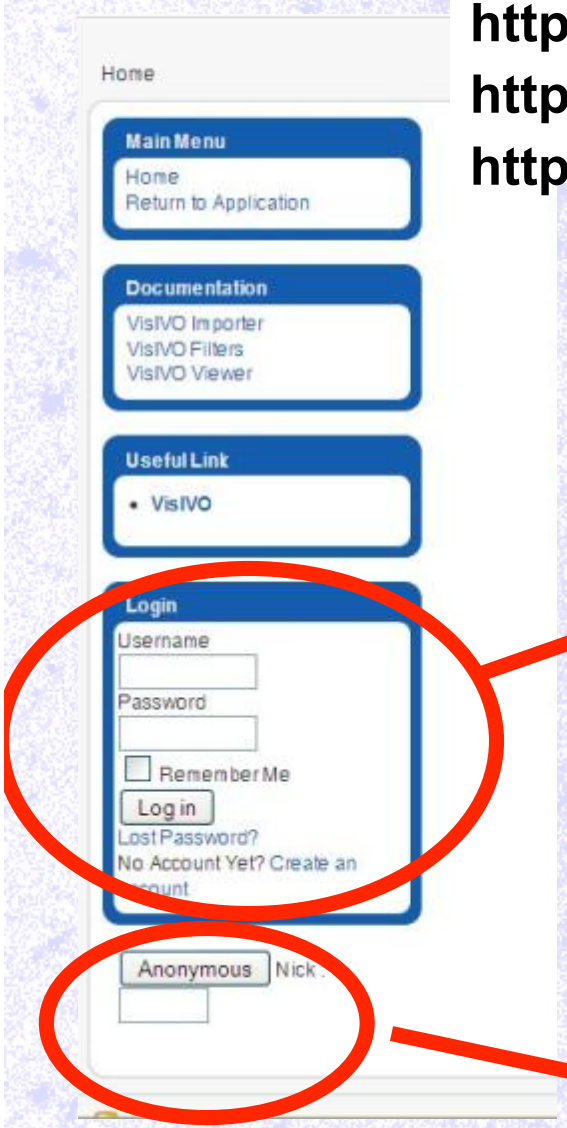


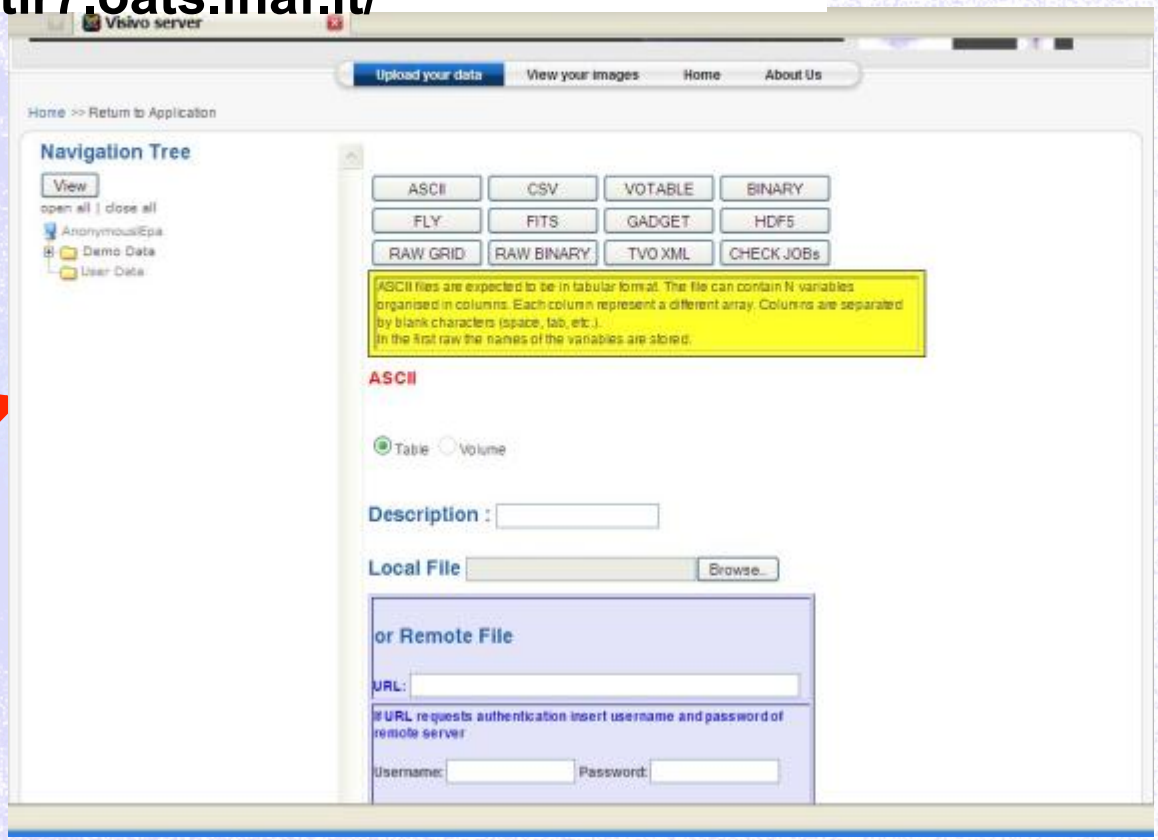


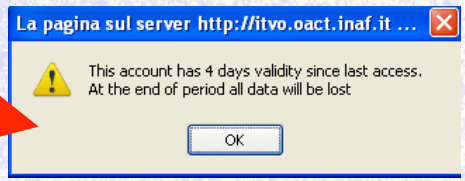


http://visivoweb.oact.inaf.it
http://visivo.port.ac.uk
http://palantir7.oats.inaf.it/

VisIVOWeb

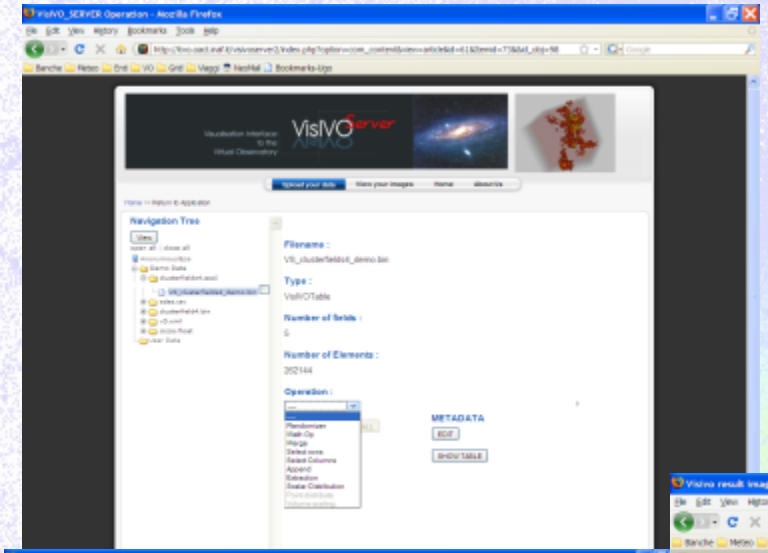


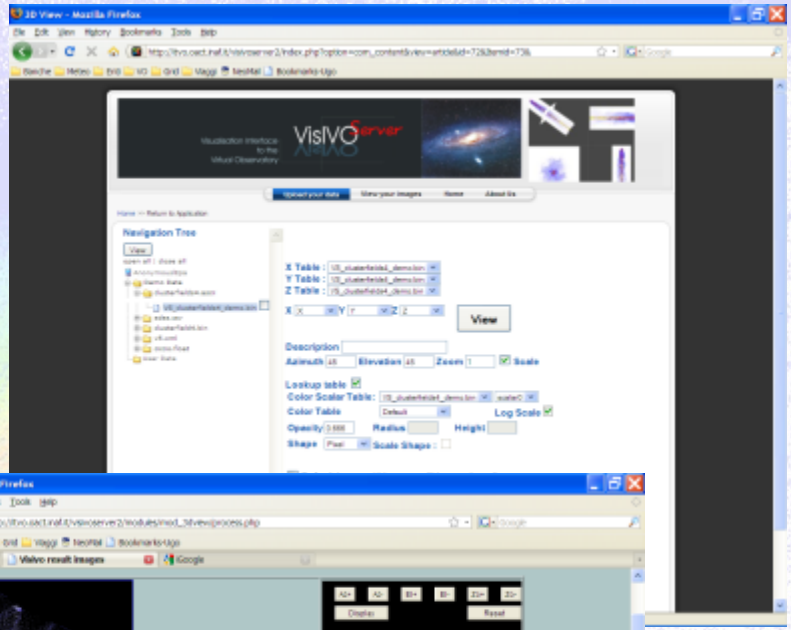


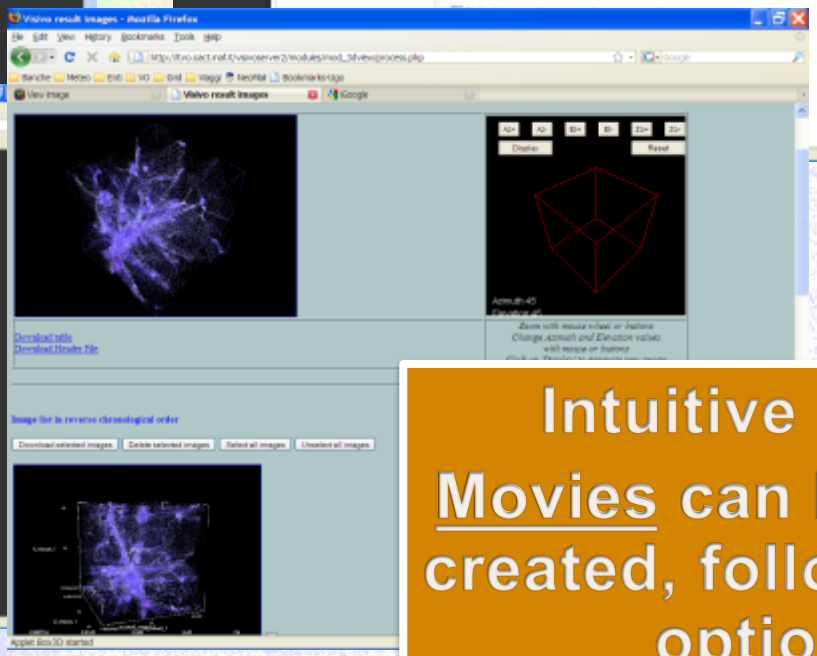


Visualisation Interface to the Virtual Observatory

VisIVO Server







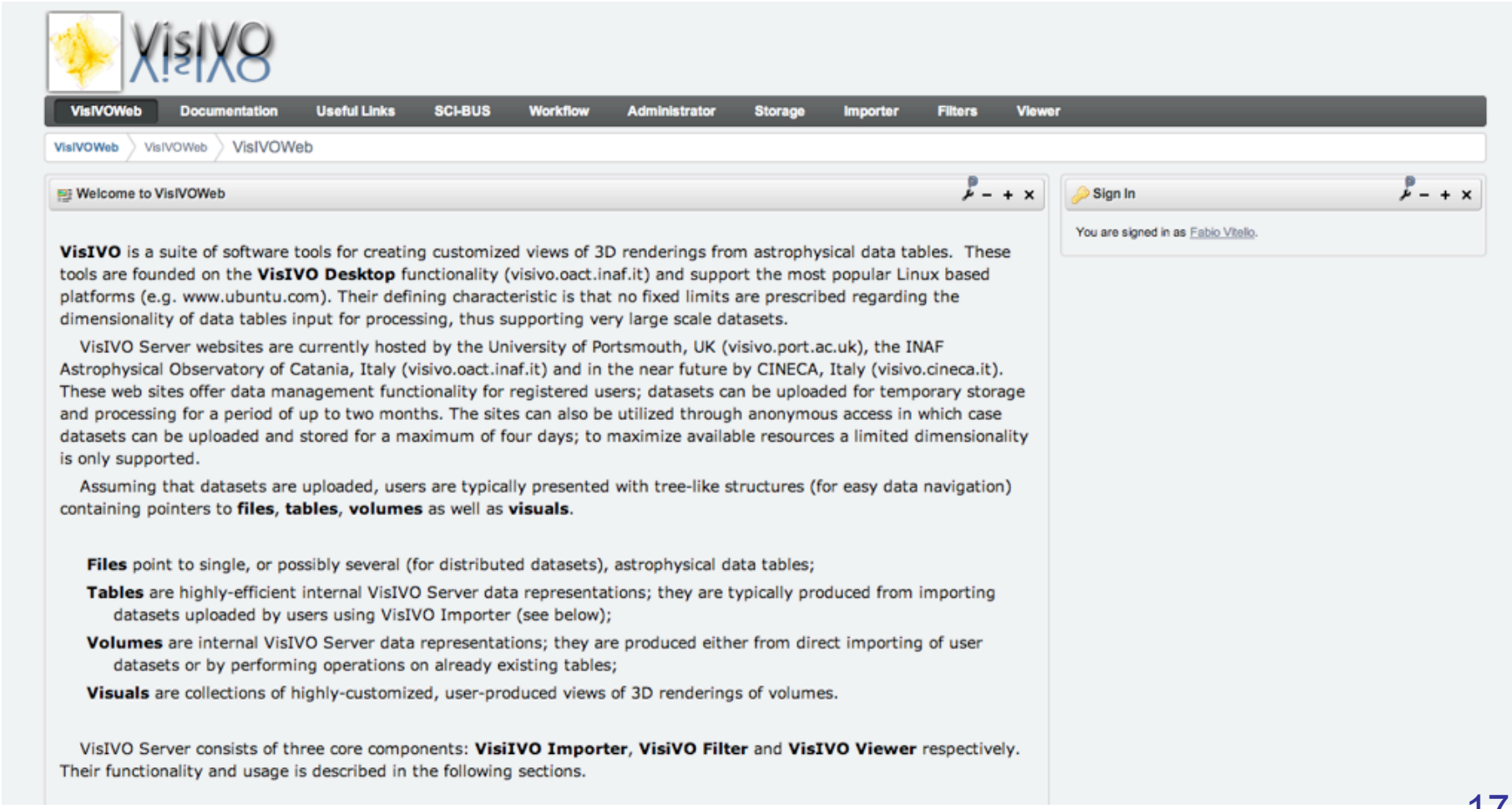
Intuitive usage

Movies can be easily created, following the options

VisIVO @ SCI-BUS

- Creation of a gateway with Liferay and gUse (WS-Pgrade) to extend the VisIVO portal, and the capability to use DCI (Grid, HPC, cloud computing)
- Creation of workflows to explore the dataset and to create movies
- VisIVO iPhone application development

VisIVO


VisIVO
VisIVOWeb
Documentation
Useful Links
SCI-BUS
Workflow
Administrator
Storage
Importer
Filters
Viewer
VisIVOWeb
VisIVOWeb
VisIVOWeb
Welcome to VisIVOWeb
Sign In
You are signed in as Fabio Vitello.
VisIVO is a suite of software tools for creating customized views of 3D renderings from astrophysical data tables. These tools are founded on the VisIVO Desktop functionality (visivo.oact.inaf.it) and support the most popular Linux based platforms (e.g. www.ubuntu.com). Their defining characteristic is that no fixed limits are prescribed regarding the dimensionality of data tables input for processing, thus supporting very large scale datasets.
VisIVO Server websites are currently hosted by the University of Portsmouth, UK (visivo.port.ac.uk), the INAF Astrophysical Observatory of Catania, Italy (visivo.oact.inaf.it) and in the near future by CINECA, Italy (visivo.cineca.it). These web sites offer data management functionality for registered users; datasets can be uploaded for temporary storage and processing for a period of up to two months. The sites can also be utilized through anonymous access in which case datasets can be uploaded and stored for a maximum of four days; to maximize available resources a limited dimensionality is only supported.
Assuming that datasets are uploaded, users are typically presented with tree-like structures (for easy data navigation) containing pointers to files, tables, volumes as well as visuals.
Files point to single, or possibly several (for distributed datasets), astrophysical data tables;
Tables are highly-efficient internal VisIVO Server data representations; they are typically produced from importing datasets uploaded by users using VisIVO Importer (see below);
Volumes are internal VisIVO Server data representations; they are produced either from direct importing of user datasets or by performing operations on already existing tables;
Visuals are collections of highly-customized, user-produced views of 3D renderings of volumes.
VisIVO Server consists of three core components: VisiIVO Importer, VisiVO Filter and VisIVO Viewer respectively. Their functionality and usage is described in the following sections.

VisIVO

Two kind of user:

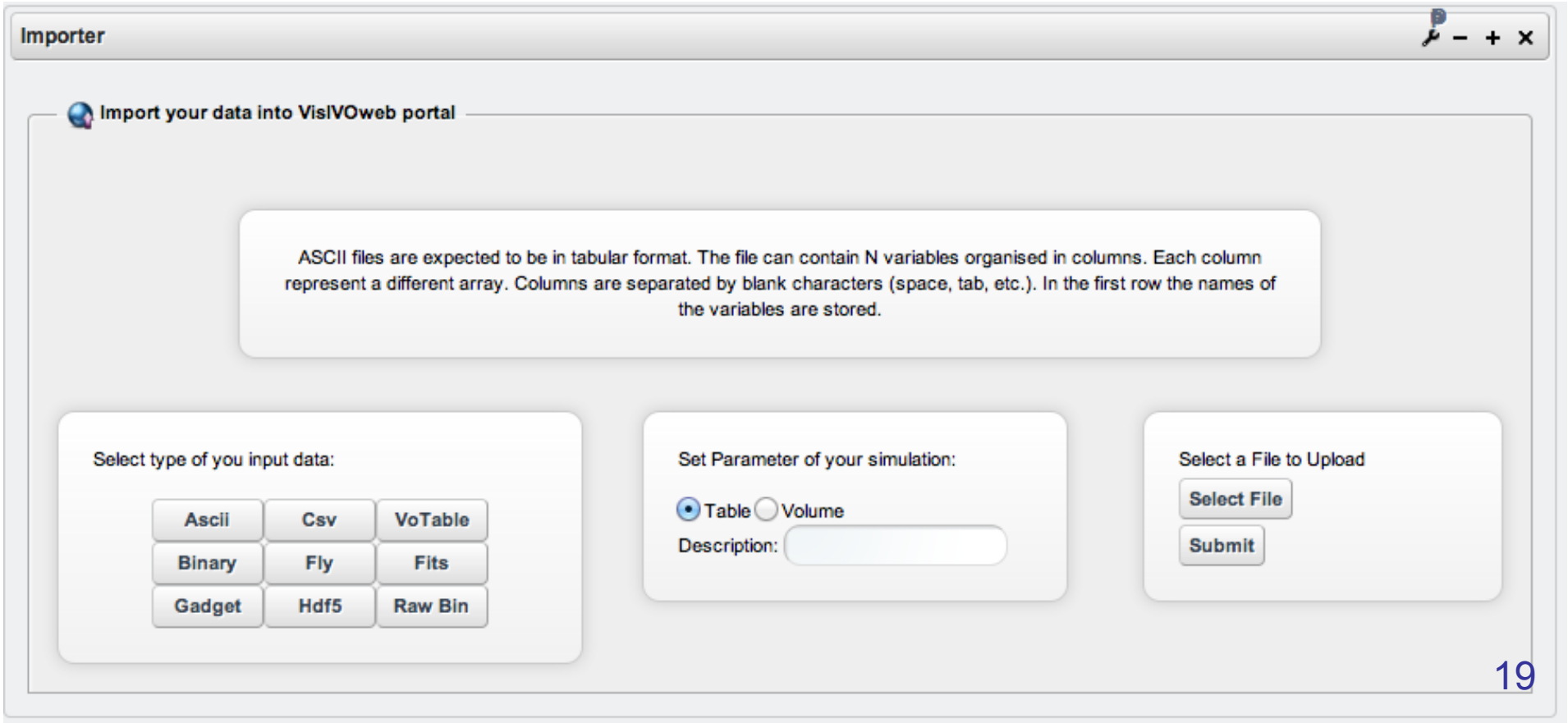
Expert Users. They can develop and configure own workflows.

All the gUse functionalities will be made available.

Standard Users. They use mainly Workflow made available by a “workflow developer” through an ASM.

VisIVO -Standard user

Easy way to use workflow whit an ASM developed by us.



VisIVO – Expert user

Use Graph Editor to create and Concrete to configure their own workflow.