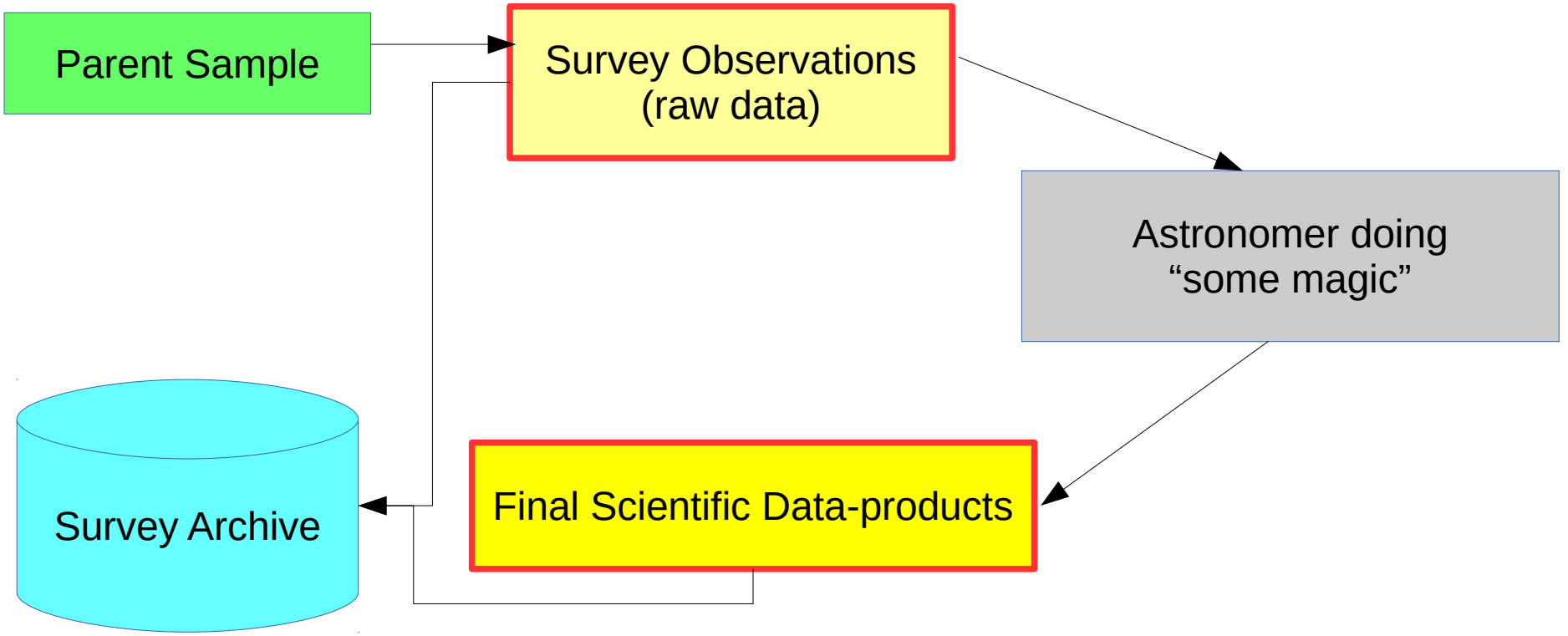
Designing tools to reduce complexity

the spectroscopic surveys example

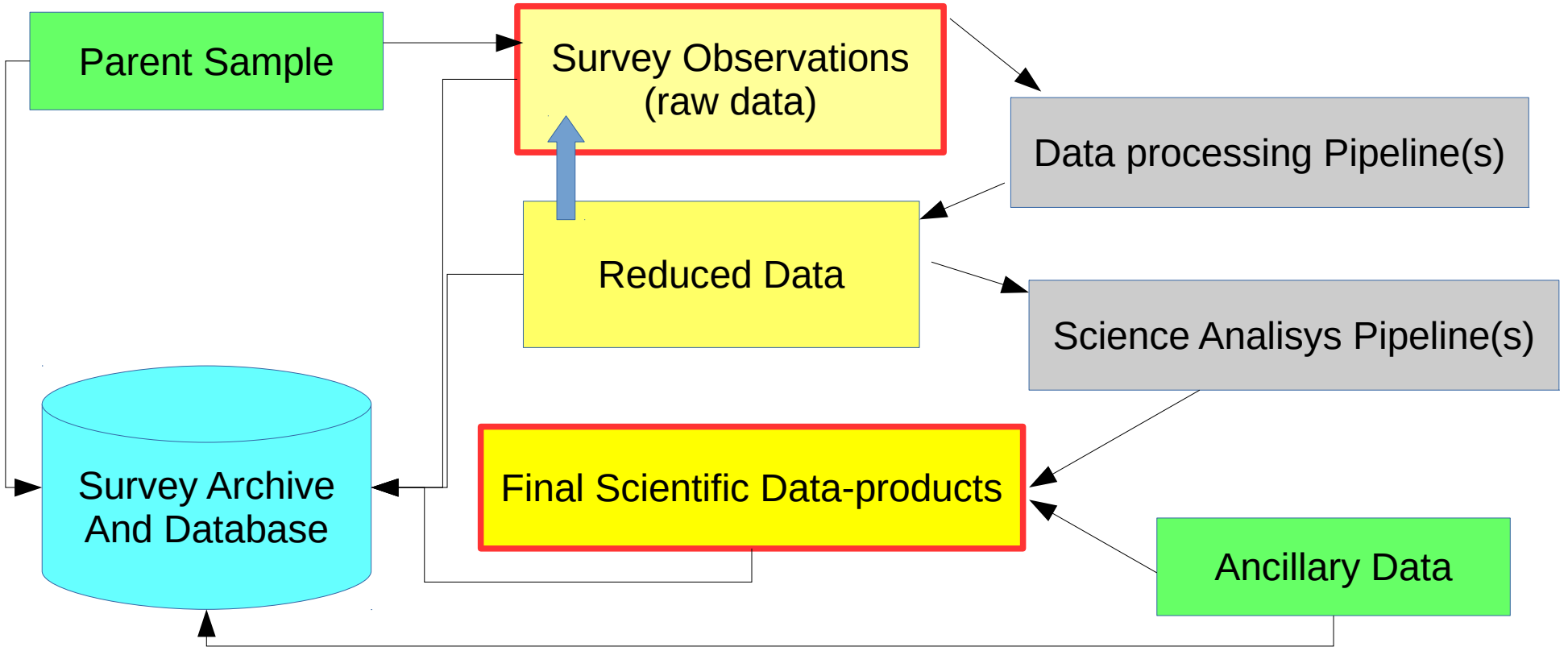
Marco Scodeggio
Paolo Franzetti, Marco Fumana
& Bianca Garilli

INAF IASF-Milano

The old Survey concept...



A Survey is now a complex “thing”...



In the beginning there was IRAF...

The IRAF Data Reduction and Analysis System

Doug Tody

National Optical Astronomy Observatories
P.O. Box 26732, Tucson, Arizona, 85726

ABSTRACT

The Image Reduction and Analysis Facility (IRAF) is a general purpose software system for the reduction and analysis of scientific data. The IRAF system provides a good selection of programs for general image processing and graphics applications, plus a large selection of programs for the reduction and analysis of optical astronomy data. The system also provides a complete modern scientific programming environment, making it straightforward for institutions using IRAF to add their own software to the system. Every effort has been made to make the system as portable and device independent as possible, so that the system may be used on a wide variety of host computers and operating systems with a wide variety of graphics and image display devices.

1. Introduction

The IRAF project began in earnest in the fall of 1981 at Kitt Peak National Observatory (NOAO did not yet exist at that time). The preliminary design of the system was completed early in 1982, and the first versions of the command language (CL) and the applications programming environment were completed during 1982. The NOAO IRAF programming group was formed in 1983. The first internal release of the system occurred at NOAO in 1984, and a beta release of the system to a few outside sites occurred in 1985.

Reducing data with IRAF

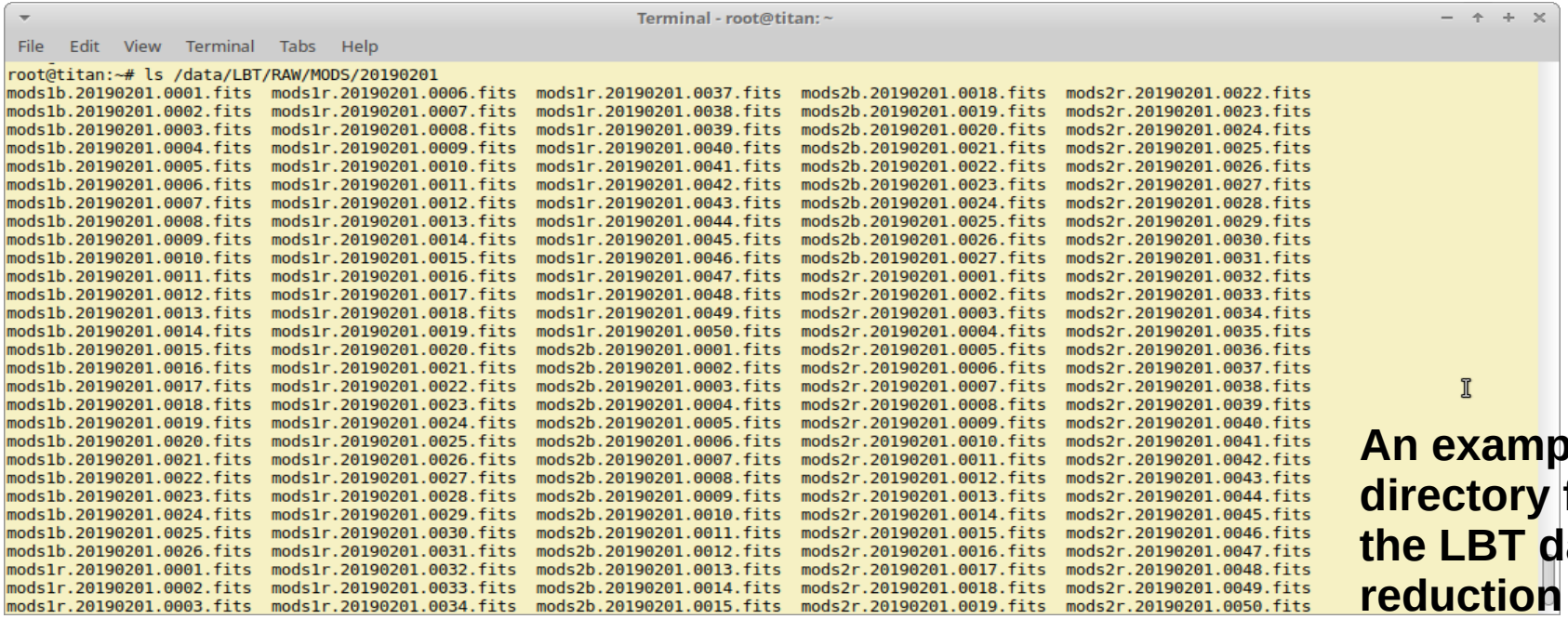
- Wavelength calibration: **identify**, **reidentify**
- Flux calibration: **standard**, **sensfunc**
- Spectrum extraction: **apfind**, **aprecenter**, **apresize**, **apedit**, **aptrace**, **apsum** (they can be run in a sequence using apall)

VIPERS:
100,000 spectra
x 10 tasks
x 30 sec =
8500 hours =
1000 working days

multiple IRAF tasks can be executed via a script, and new IRAF tasks can be coded using the IRAF programming language

The problem with data proliferation

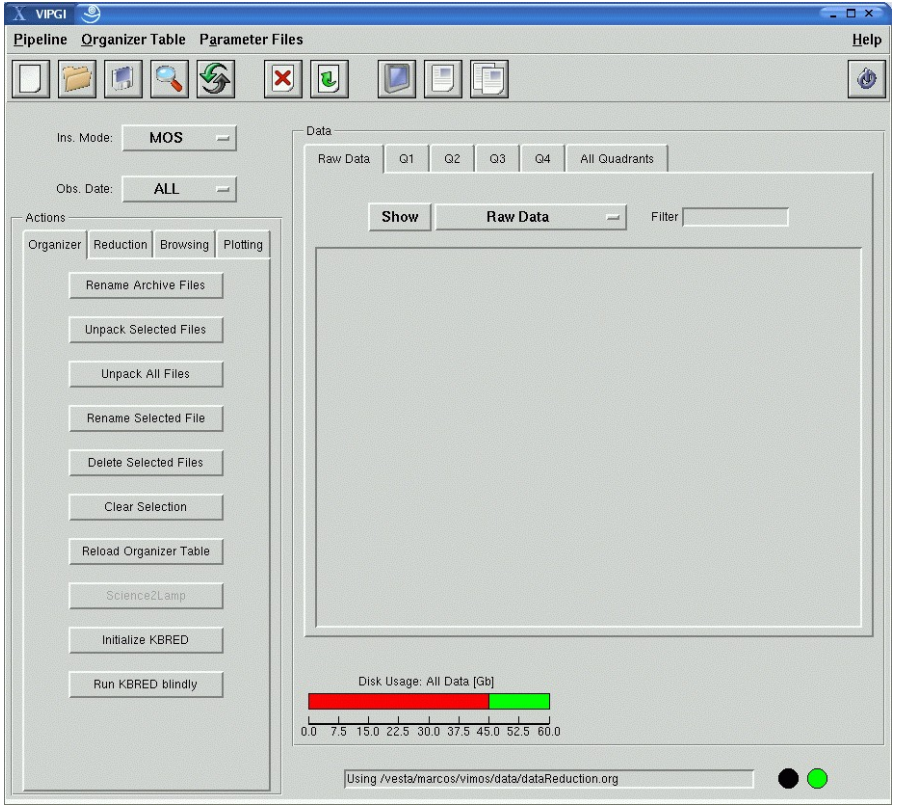
Data archiving is generally observatory-oriented,
and not really astronomer-oriented….



An example directory from the LBT data reduction center

A tool for a project: **VIPGI**
the VIMOS Interactive Pipeline and Graphical Interface

* Data Organizer
* Smart Data Browser
* Interface to Pipeline Tasks
* Data Visualization
* Direct Interface to
  Redshift Measurement Tool



* Data Reduction Recipes:
  150K lines of C code

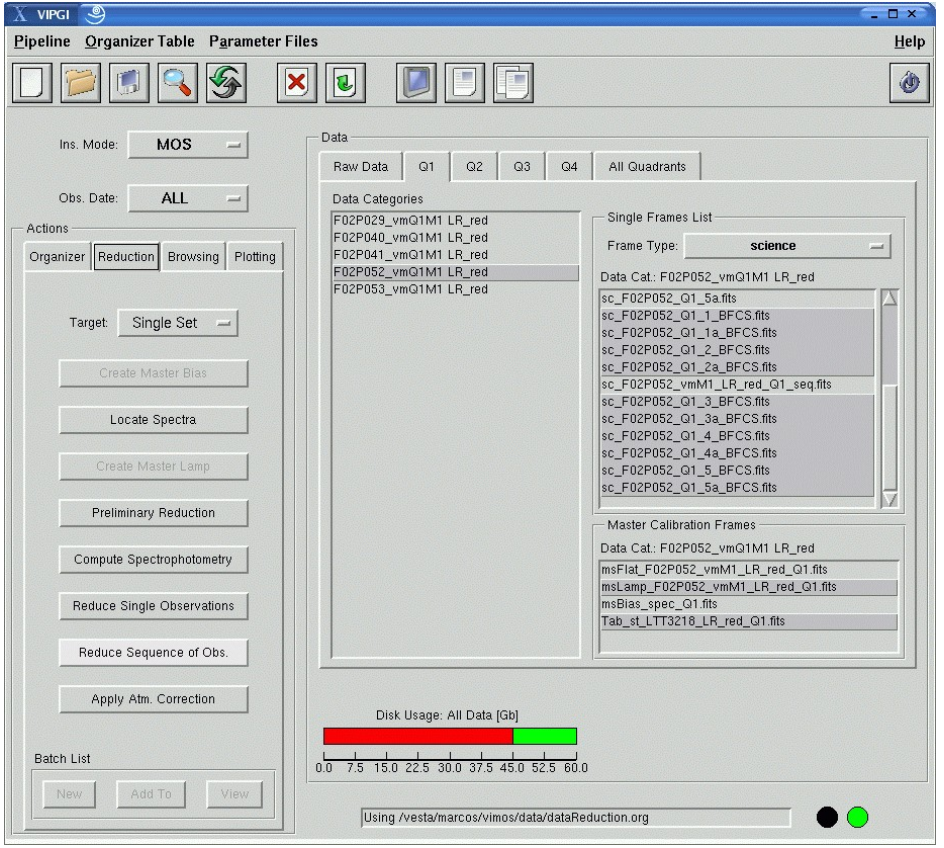
* GUI and plotting/browsing:
  16K lines of Python code

Data reduction with just a few clicks

Smart Browsing;
Data organized by Pointing, Mask, and Data Type

Data Reduction: a few tasks available, depending on the selected data type

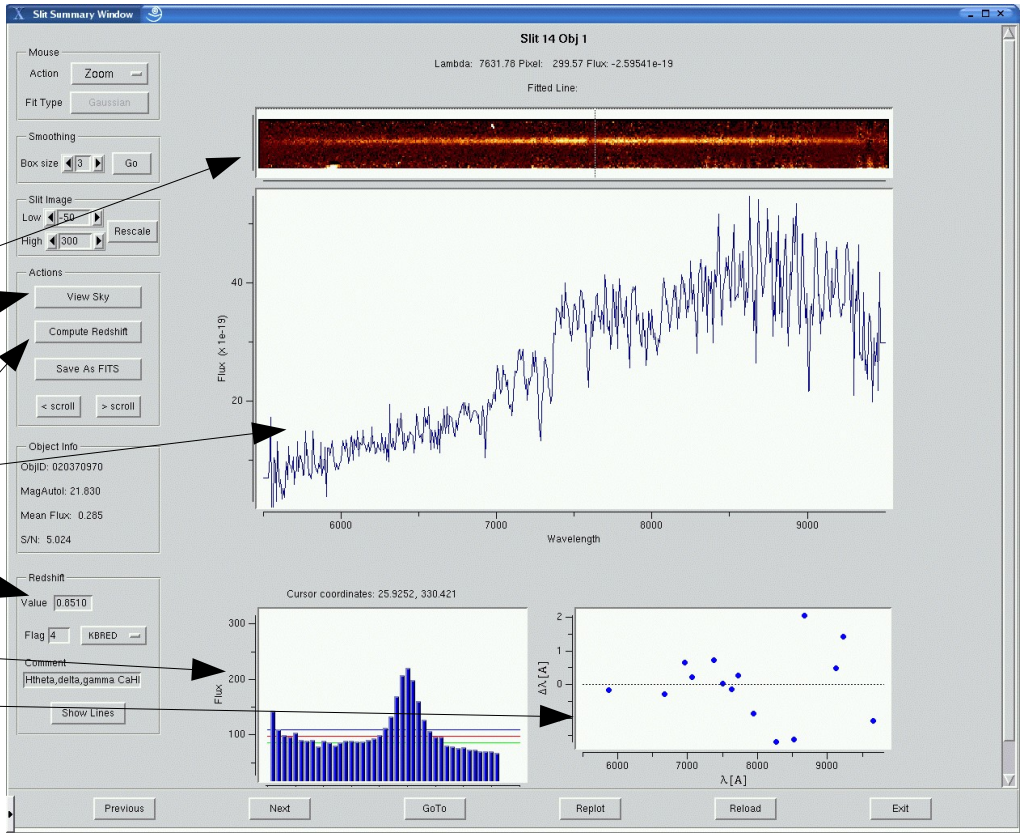
VIPERS:
288 masks
x 4 quadrants
x 6 tasks
x 20 min =
2400 hours =
300 working days



Integrated data browsing / plotting tools

Within the FITS file
and for each slit:

* 2D Extracted Spec
* 2D Sky
* 1D Extracted Spec
* 1D Flux Calib Spec
* Redshift measurement
* Location info
* Aperture info
* Lambda Cal Info



From **VIPGI** to today's tools

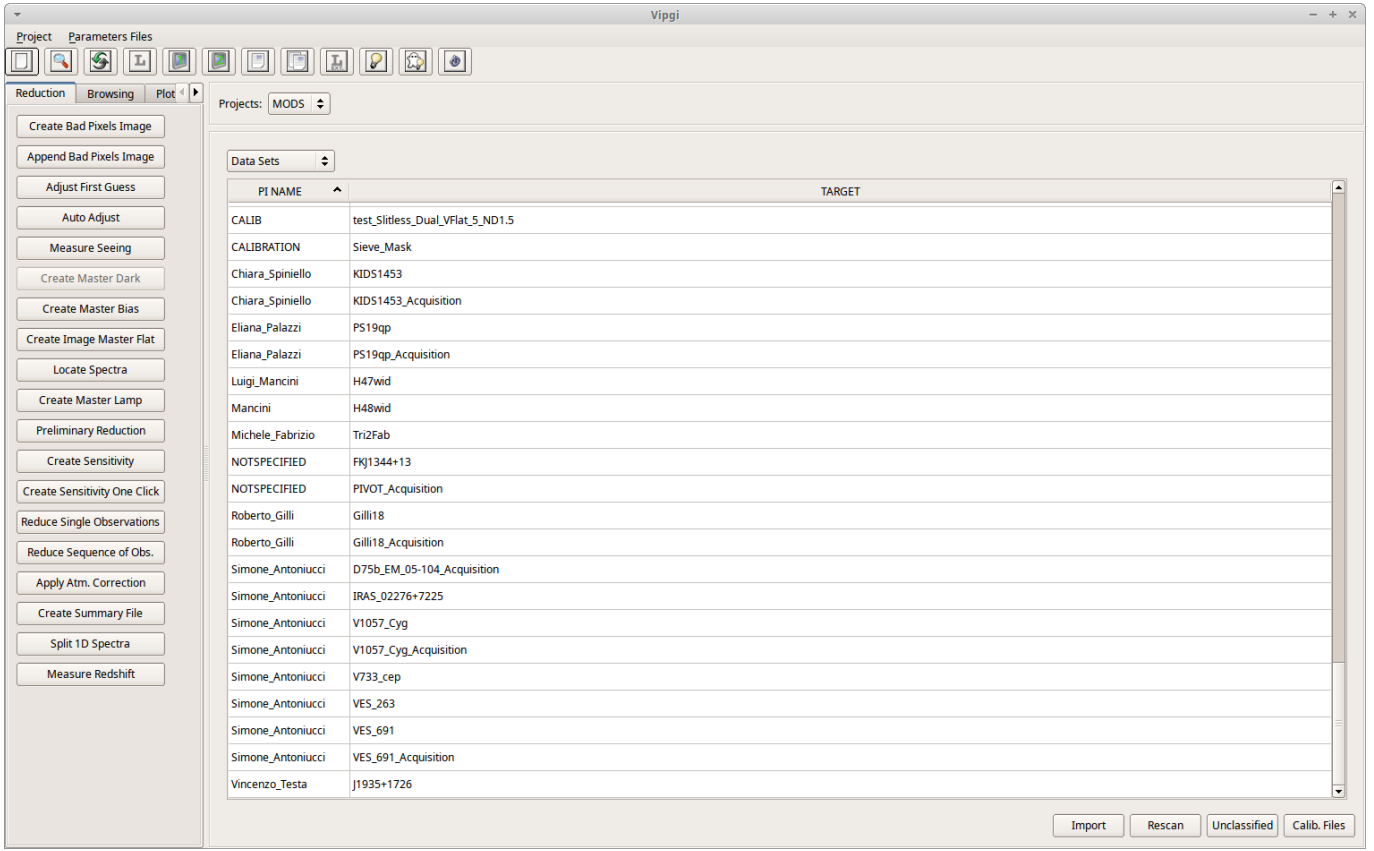
- Two separate paths towards today's spectroscopic data reduction and project management
- The LBT / general purpose **new VIPGI**
- The Spectroscopic Survey tool **Easylife**

The new VIPGI

Like in the original VIPGI:

* Data Organizer
* Smart Data Browser
* Interface to Pipeline Tasks
* Data Visualization

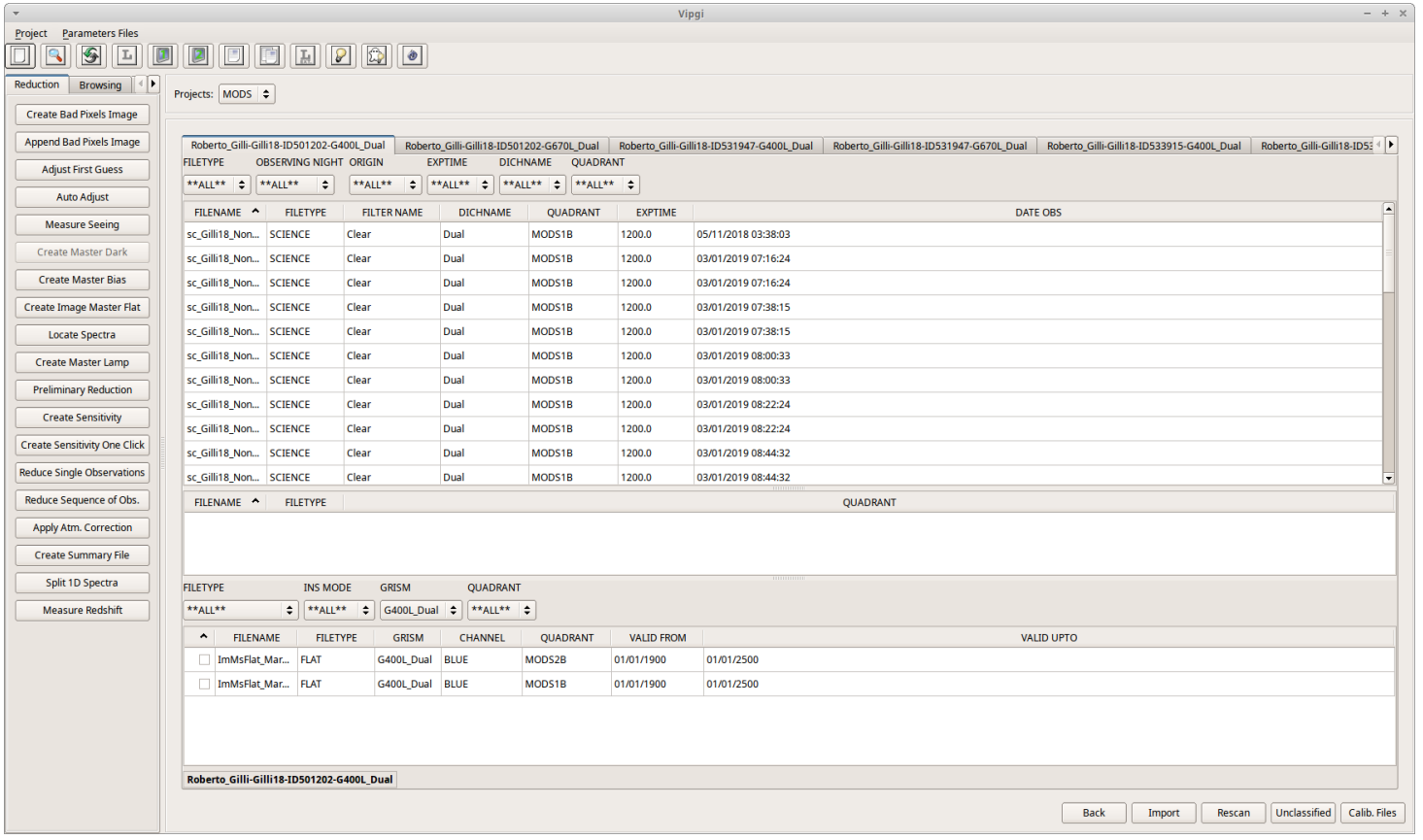
More Python and less C code



The new **VIPGI**

The power of having data organized into smart categories (at the price of needing a special data import facility)

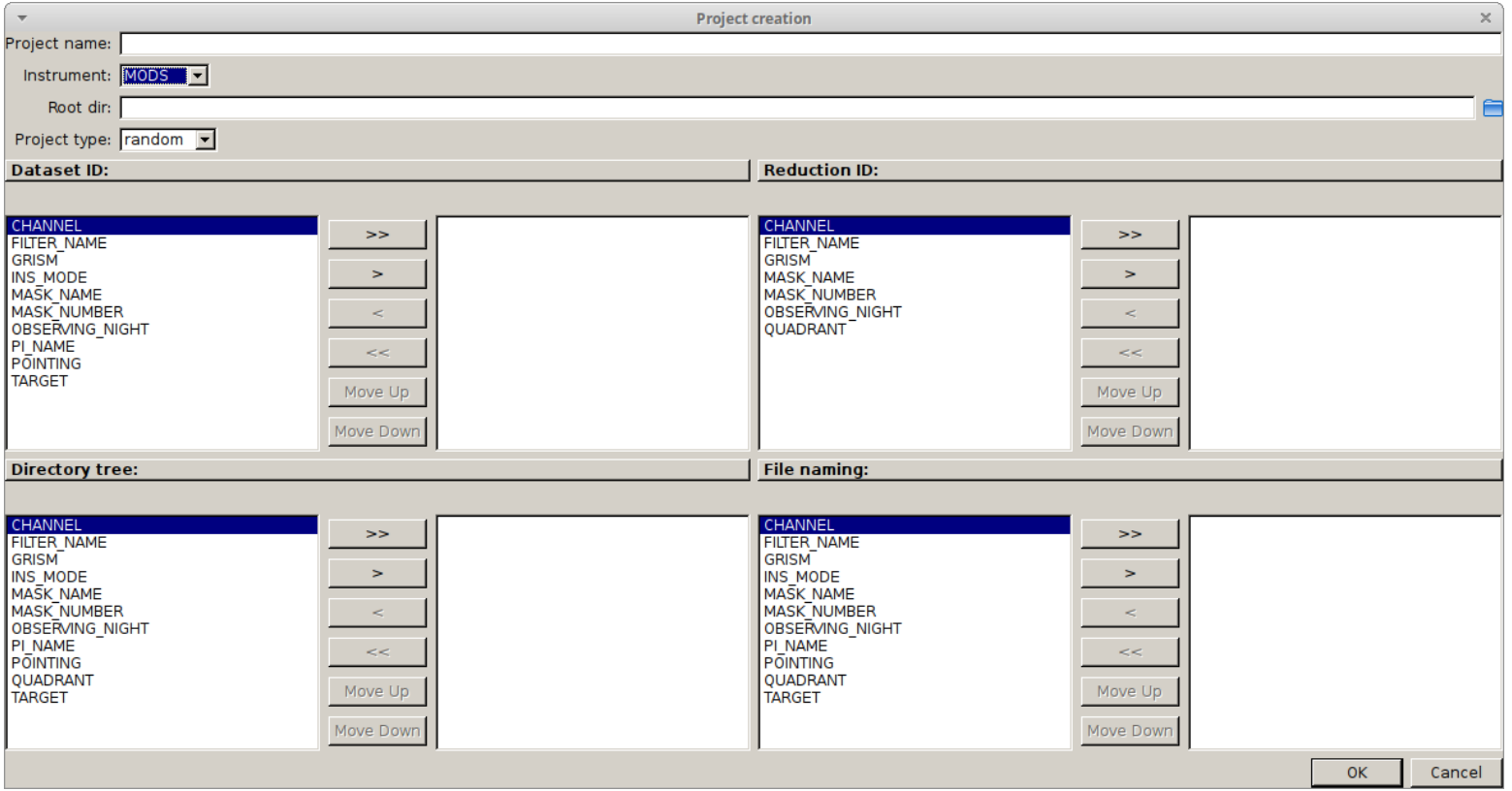
Only homogeneous sets of data are visualized and can be selected to run pipeline taks



The new **VIPGI**

The data organizer can be fully customized by the user according to his/her preferred scheme.

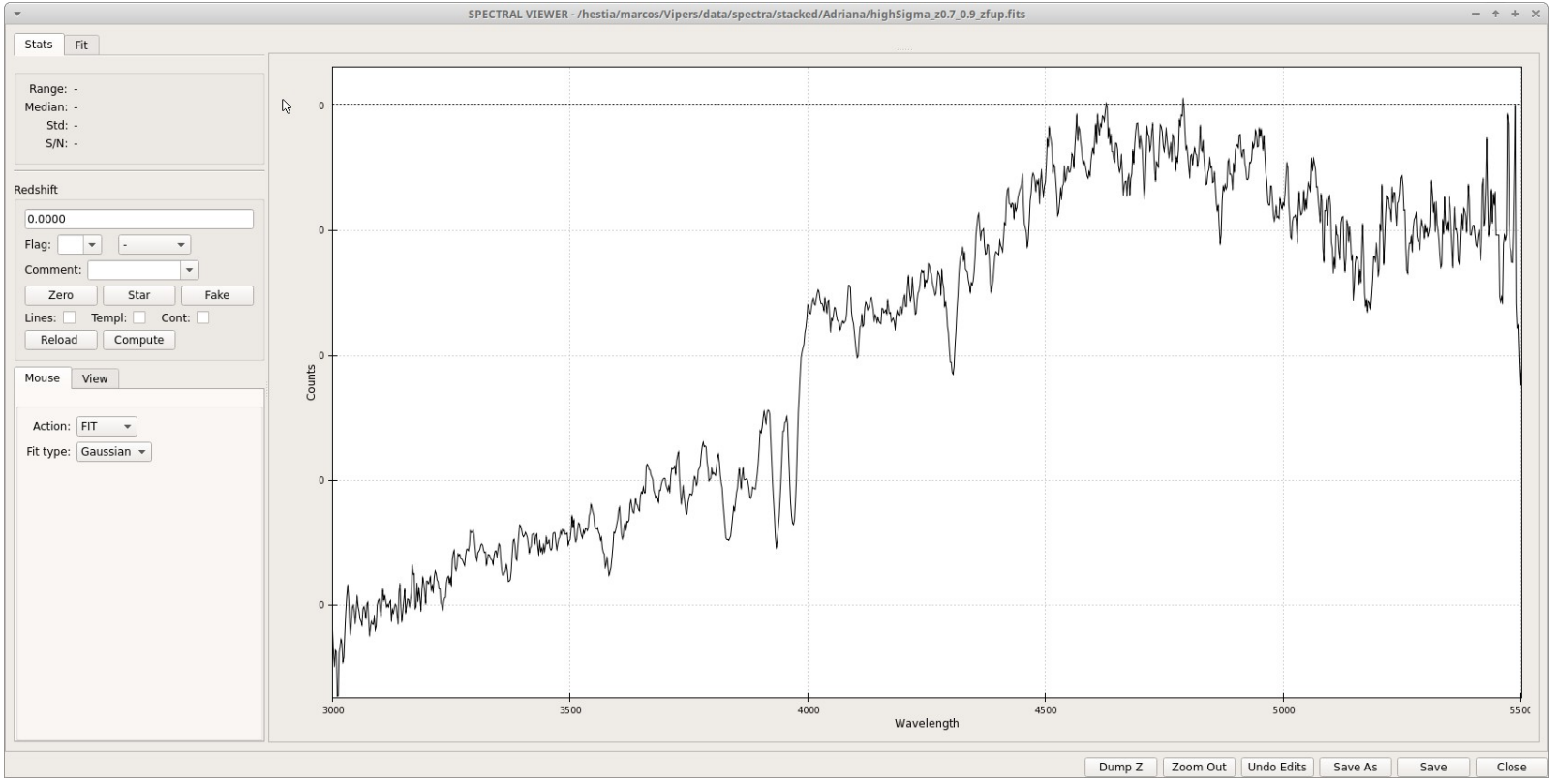
This involves both the location of files on disk and the visualization through VIPGI



Spectra visualization and redshift measurements: **pandora.ez**

Visualization tool, but also redshift and spectral lines measurement tool

Can be used standalone, or integrated within VIPGI



Spectra visualization and redshift measurements: **pandora.ez**

In black: observed spectrum

In yellow; best fitting template

In green: most common spectral lines

Easylife for Spectroscopic Surveys

Easylife's concept is relatively simple

- Store all survey info in a Database
- Join together with a script the various pipeline tasks
- Add quality control steps in between the main tasks
- Add Survey Management tools
- Make it all available via a Web interface

easylife design

pipeline(s)

- data organization
- data reduction

management

- masks /OBs preparation
- observations monitoring and raw retrival
- data distribution for measurements
- data distribution for analysis

database

- users
- parent catalog(s)
- ancillary data
- masks
- observations
- spectra
- derived data

website

public

- database
- info

internal

- database
- wiki / workgroups
- team info
- mailing lists
- measurements assignments
- data releases
- abstracts / papers

test

Easylife main components

- An optimized Database Interface (and not just a data archive) because a Survey is a living thing, and stored info changes daily
- A Web interface open to the whole Survey team (because many people need to manipulate info)
- Smart Survey management scripts, that reduce to a minimum astronomer’s decisions (and the chance to make mistakes!!)
- Smart Pipelines (same concept as in the previous point)

Final Remarks

- Some twenty years of experience with Spectroscopic Surveys and Data Reduction Pipelines at IASF-Milano
- Managed all big ESO redshift surveys, like VVDS, zCOSMOS, CLASH-VLT, VIPERS, VANDELS
- Managed almost 10 years of LBT data reduction center for spectroscopic data

Final Remarks

- Un-ortodox choices in terms of data format often save a lot of time and headaches
- Managing survey projects end-to-end requires integrating a lot of tools, and can be as challenging as dealing with huge amounts of data
- An INAF SpecLab is becoming a reality, and it will require many interactions with the Archives/Data management community represented here
  (see https://indico.ict.inaf.it/event/851)


That's all Folks!
lsberg©