

Designing tools to reduce complexity the spectroscopic surveys example

Marco Scodeggio
Paolo Franzetti, Marco Fumana
& Bianca Garilli

INAF



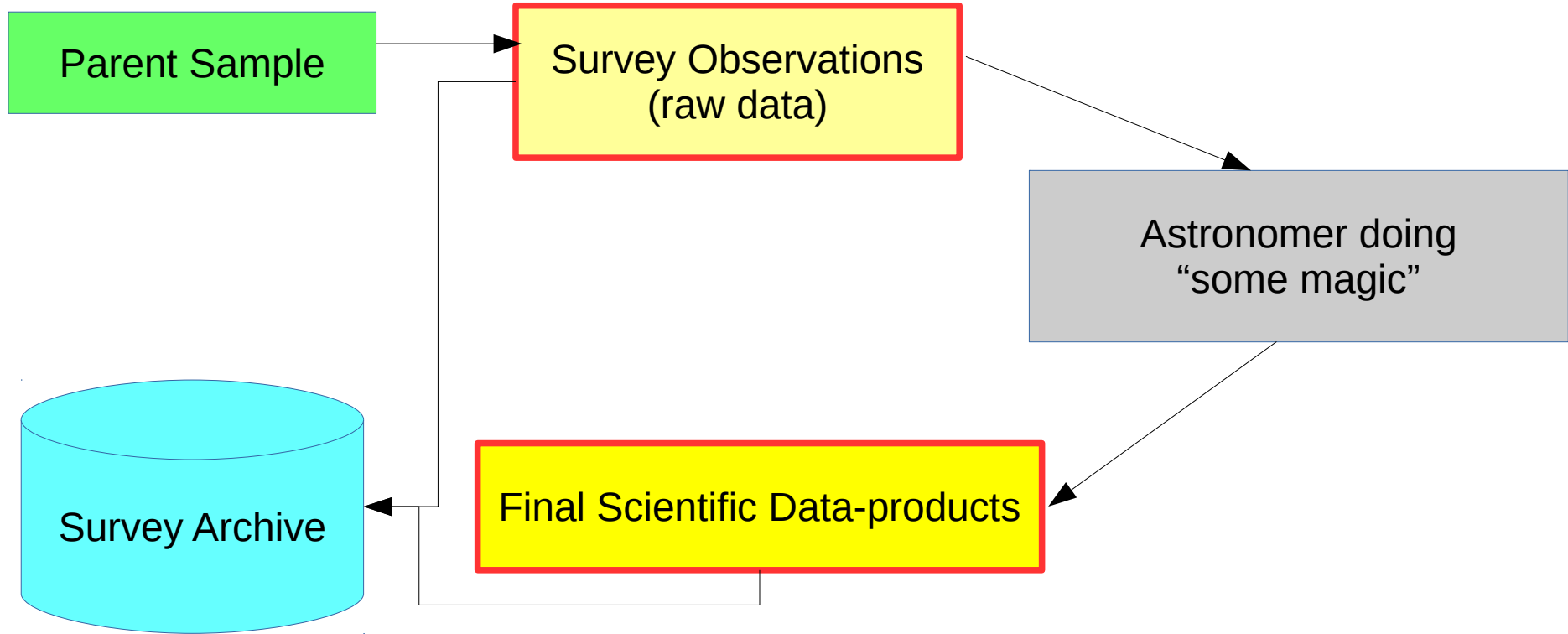
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Istituto di Astrofisica Spaziale e Fisica cosmica di Milano

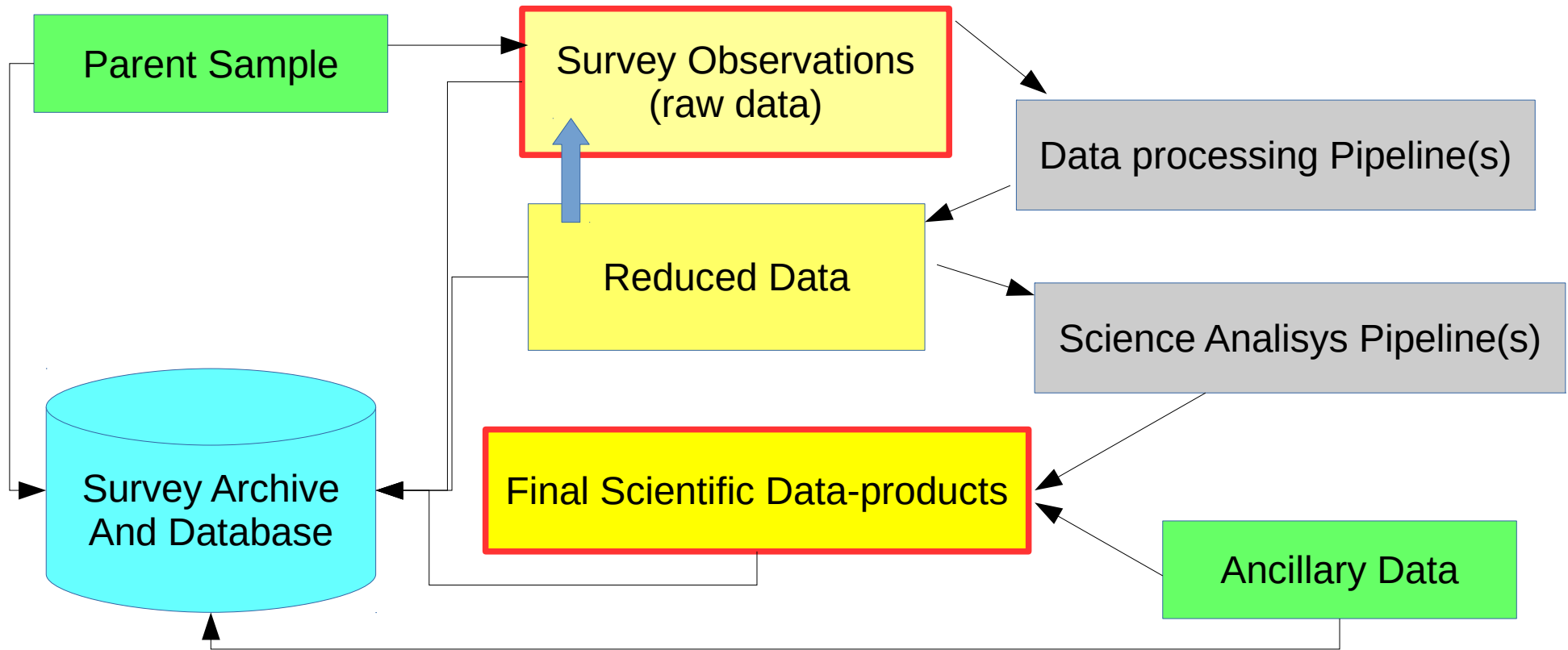


INAF IASF-Milano

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

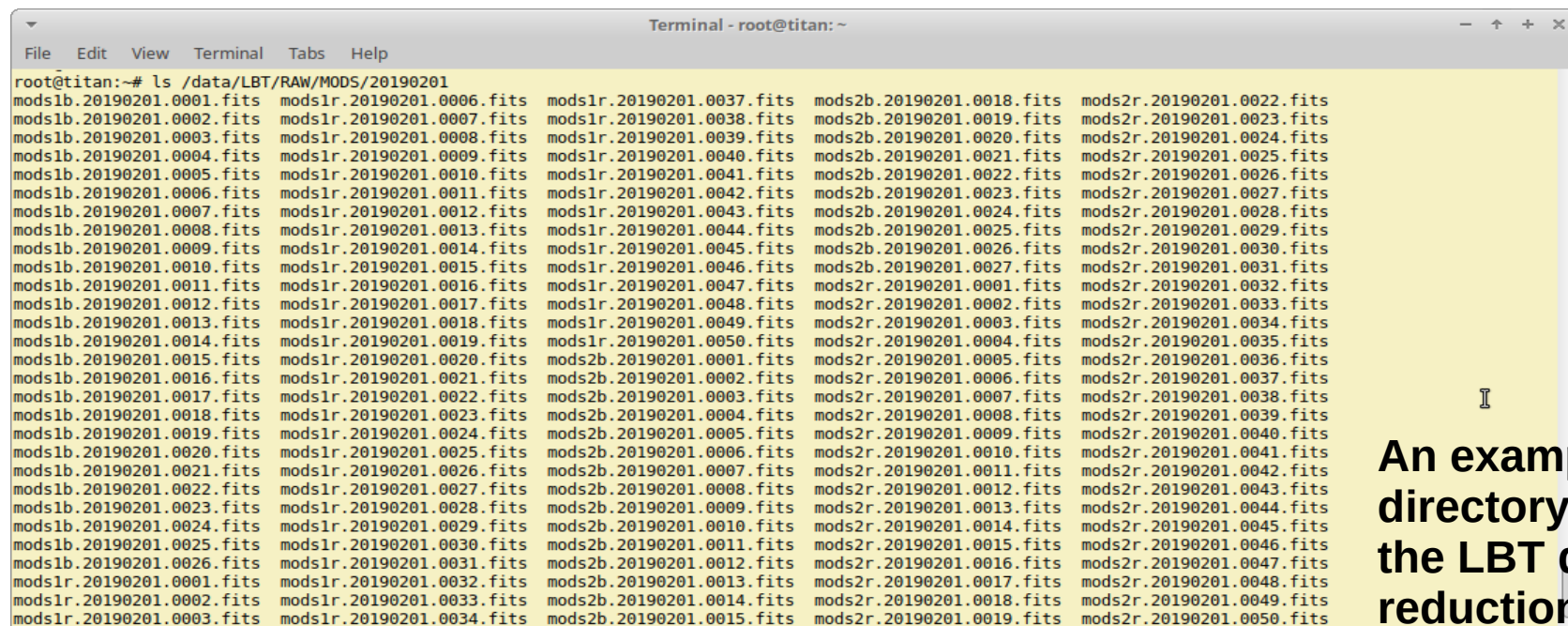
- Wavelength calibration: **identify, reidentify**
- Flux calibration: **standard, sensfunc**
- Spectrum extraction: **apfind, aprecenter, apresize, apedit, aptrace, apsum**
(they can be run in a sequence using apall)

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....



A terminal window titled "Terminal - root@titan: ~" displays the output of the command `ls /data/LBT/RAW/MODS/20190201`. The output is a long list of FITS files, organized into five columns. The files are named with a prefix (e.g., `mods1b.`, `mods1r.`, `mods2b.`, `mods2r.`) followed by a date and time stamp (e.g., `20190201.0001.fits`). The list shows a dense collection of data files, illustrating the volume of data generated.

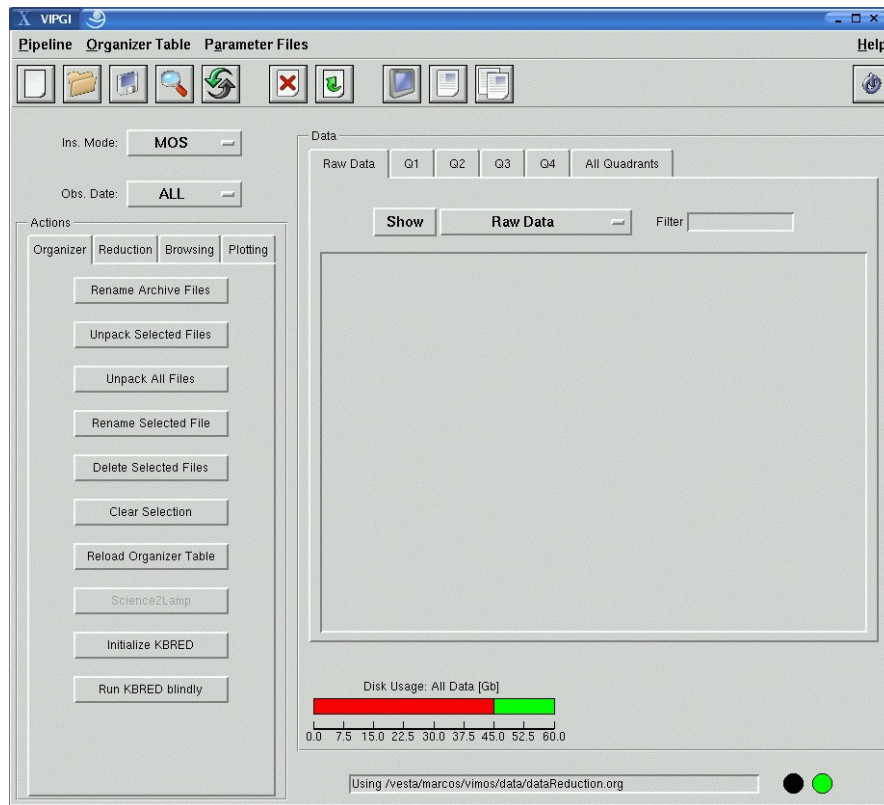
```
root@titan:~# ls /data/LBT/RAW/MODS/20190201
mods1b.20190201.0001.fits  mods1r.20190201.0006.fits  mods1r.20190201.0037.fits  mods2b.20190201.0018.fits  mods2r.20190201.0022.fits
mods1b.20190201.0002.fits  mods1r.20190201.0007.fits  mods1r.20190201.0038.fits  mods2b.20190201.0019.fits  mods2r.20190201.0023.fits
mods1b.20190201.0003.fits  mods1r.20190201.0008.fits  mods1r.20190201.0039.fits  mods2b.20190201.0020.fits  mods2r.20190201.0024.fits
mods1b.20190201.0004.fits  mods1r.20190201.0009.fits  mods1r.20190201.0040.fits  mods2b.20190201.0021.fits  mods2r.20190201.0025.fits
mods1b.20190201.0005.fits  mods1r.20190201.0010.fits  mods1r.20190201.0041.fits  mods2b.20190201.0022.fits  mods2r.20190201.0026.fits
mods1b.20190201.0006.fits  mods1r.20190201.0011.fits  mods1r.20190201.0042.fits  mods2b.20190201.0023.fits  mods2r.20190201.0027.fits
mods1b.20190201.0007.fits  mods1r.20190201.0012.fits  mods1r.20190201.0043.fits  mods2b.20190201.0024.fits  mods2r.20190201.0028.fits
mods1b.20190201.0008.fits  mods1r.20190201.0013.fits  mods1r.20190201.0044.fits  mods2b.20190201.0025.fits  mods2r.20190201.0029.fits
mods1b.20190201.0009.fits  mods1r.20190201.0014.fits  mods1r.20190201.0045.fits  mods2b.20190201.0026.fits  mods2r.20190201.0030.fits
mods1b.20190201.0010.fits  mods1r.20190201.0015.fits  mods1r.20190201.0046.fits  mods2b.20190201.0027.fits  mods2r.20190201.0031.fits
mods1b.20190201.0011.fits  mods1r.20190201.0016.fits  mods1r.20190201.0047.fits  mods2b.20190201.0028.fits  mods2r.20190201.0032.fits
mods1b.20190201.0012.fits  mods1r.20190201.0017.fits  mods1r.20190201.0048.fits  mods2b.20190201.0029.fits  mods2r.20190201.0033.fits
mods1b.20190201.0013.fits  mods1r.20190201.0018.fits  mods1r.20190201.0049.fits  mods2b.20190201.0030.fits  mods2r.20190201.0034.fits
mods1b.20190201.0014.fits  mods1r.20190201.0019.fits  mods1r.20190201.0050.fits  mods2b.20190201.0031.fits  mods2r.20190201.0035.fits
mods1b.20190201.0015.fits  mods1r.20190201.0020.fits  mods2b.20190201.0001.fits  mods2r.20190201.0005.fits  mods2r.20190201.0036.fits
mods1b.20190201.0016.fits  mods1r.20190201.0021.fits  mods2b.20190201.0002.fits  mods2r.20190201.0006.fits  mods2r.20190201.0037.fits
mods1b.20190201.0017.fits  mods1r.20190201.0022.fits  mods2b.20190201.0003.fits  mods2r.20190201.0007.fits  mods2r.20190201.0038.fits
mods1b.20190201.0018.fits  mods1r.20190201.0023.fits  mods2b.20190201.0004.fits  mods2r.20190201.0008.fits  mods2r.20190201.0039.fits
mods1b.20190201.0019.fits  mods1r.20190201.0024.fits  mods2b.20190201.0005.fits  mods2r.20190201.0009.fits  mods2r.20190201.0040.fits
mods1b.20190201.0020.fits  mods1r.20190201.0025.fits  mods2b.20190201.0006.fits  mods2r.20190201.0010.fits  mods2r.20190201.0041.fits
mods1b.20190201.0021.fits  mods1r.20190201.0026.fits  mods2b.20190201.0007.fits  mods2r.20190201.0011.fits  mods2r.20190201.0042.fits
mods1b.20190201.0022.fits  mods1r.20190201.0027.fits  mods2b.20190201.0008.fits  mods2r.20190201.0012.fits  mods2r.20190201.0043.fits
mods1b.20190201.0023.fits  mods1r.20190201.0028.fits  mods2b.20190201.0009.fits  mods2r.20190201.0013.fits  mods2r.20190201.0044.fits
mods1b.20190201.0024.fits  mods1r.20190201.0029.fits  mods2b.20190201.0010.fits  mods2r.20190201.0014.fits  mods2r.20190201.0045.fits
mods1b.20190201.0025.fits  mods1r.20190201.0030.fits  mods2b.20190201.0011.fits  mods2r.20190201.0015.fits  mods2r.20190201.0046.fits
mods1b.20190201.0026.fits  mods1r.20190201.0031.fits  mods2b.20190201.0012.fits  mods2r.20190201.0016.fits  mods2r.20190201.0047.fits
mods1r.20190201.0001.fits  mods1r.20190201.0032.fits  mods2b.20190201.0013.fits  mods2r.20190201.0017.fits  mods2r.20190201.0048.fits
mods1r.20190201.0002.fits  mods1r.20190201.0033.fits  mods2b.20190201.0014.fits  mods2r.20190201.0018.fits  mods2r.20190201.0049.fits
mods1r.20190201.0003.fits  mods1r.20190201.0034.fits  mods2b.20190201.0015.fits  mods2r.20190201.0019.fits  mods2r.20190201.0050.fits
```

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

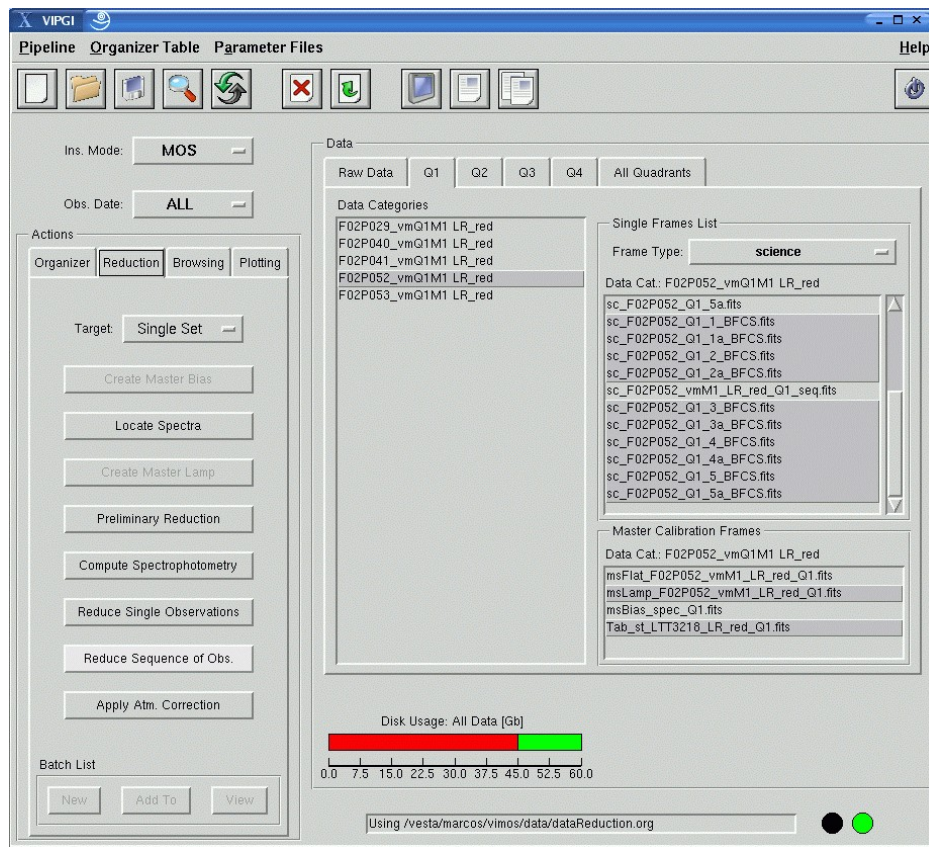


Data reduction with just a few clicks

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

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

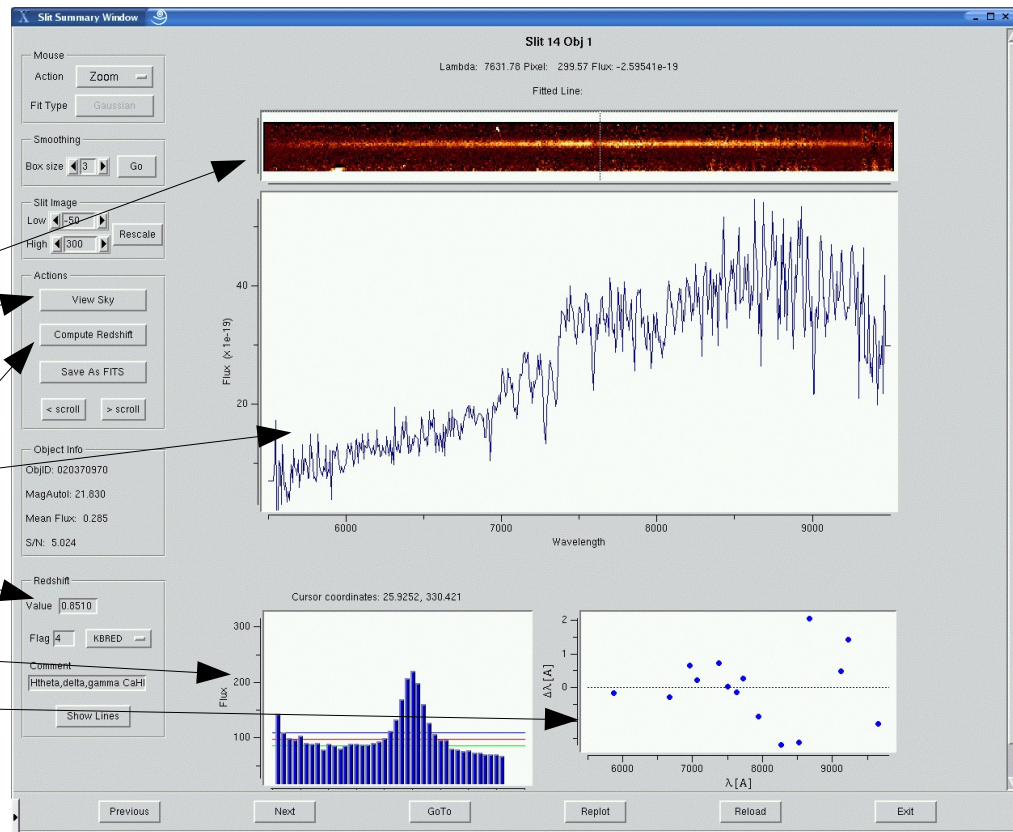
VIPERS:
288 masks
x 4 quadrants
x 6 tasks
x 20 min =
2400 hours =
300 working days



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

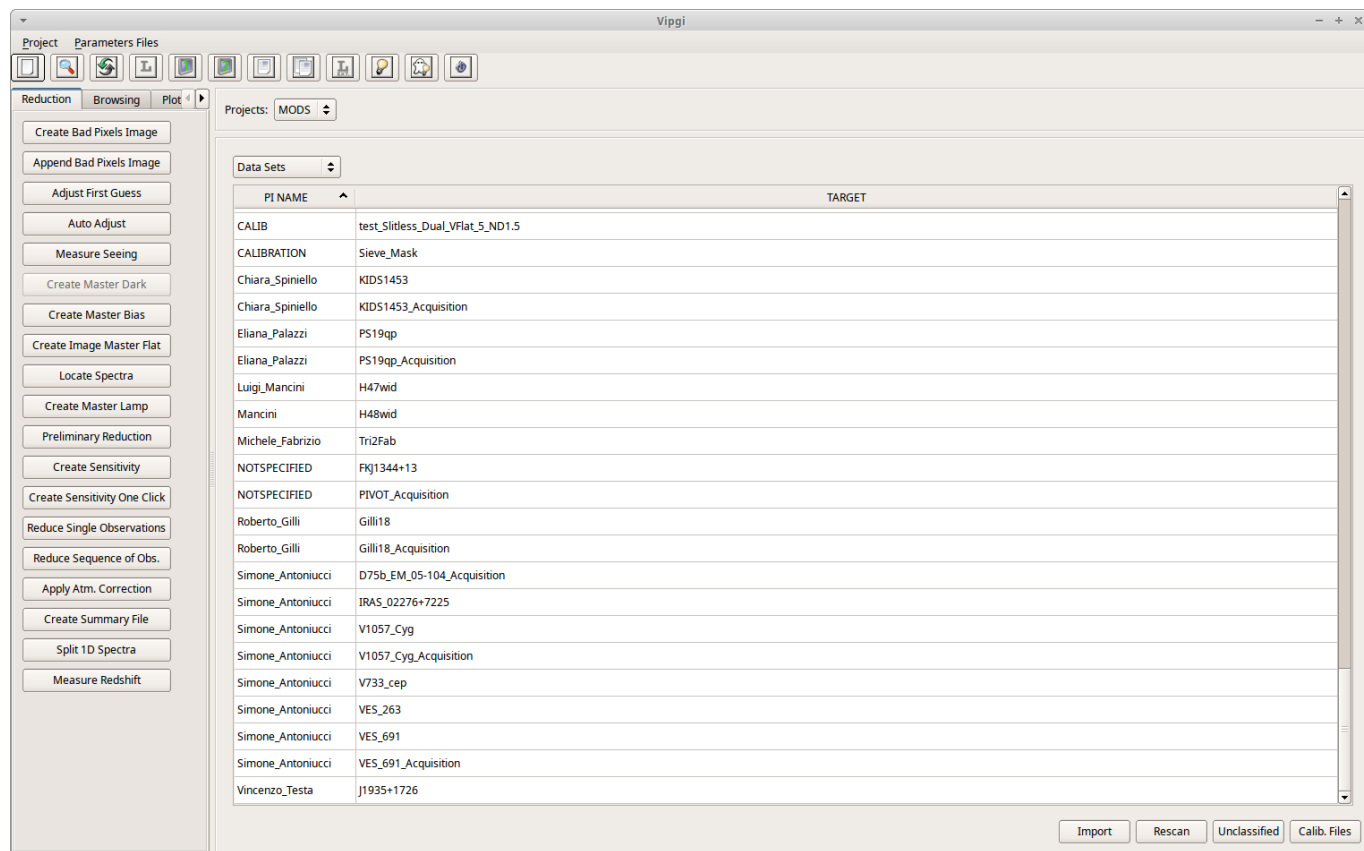
- Two separate paths towards today's spectroscopic data reduction and project management
- The LBT / general purpose **new VIPGI**
- The Spectroscopic Survey tool **Easylife**

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 tasks

The screenshot shows the VIPGI software interface. On the left is a sidebar with various processing buttons like 'Create Bad Pixels Image', 'Append Bad Pixels Image', 'Adjust First Guess', 'Auto Adjust', 'Measure Seeing', 'Create Master Dark', 'Create Master Bias', 'Create Image Master Flat', 'Locate Spectra', 'Create Master Lamp', 'Preliminary Reduction', 'Create Sensitivity', 'Create Sensitivity One Click', 'Reduce Single Observations', 'Reduce Sequence of Obs.', 'Apply Atm. Correction', 'Create Summary File', 'Split 1D Spectra', and 'Measure Redshift'. The main window displays a list of projects and a table of observations.

FILENAME	FILETYPE	FILTER NAME	DICHNAME	QUADRANT	EXPTIME	DATE OBS
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	05/11/2018 03:38:03
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 07:16:24
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 07:16:24
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 07:38:15
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 07:38:15
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:00:33
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:00:33
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:22:24
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:22:24
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:44:32
sc_Gilli18_Non...	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:44:32

Below the table, there is a section for 'QUADRANT' with a table of parameters:

FILETYPE	INS MODE	GRISM	QUADRANT	VALID FROM	VALID UPTO
<input type="checkbox"/> ImMsFlat_Mar...	FLAT	G400L_Dual	BLUE	MODS2B	01/01/1900 01/01/2500
<input type="checkbox"/> ImMsFlat_Mar...	FLAT	G400L_Dual	BLUE	MODS1B	01/01/1900 01/01/2500

At the bottom, there are buttons for 'Back', 'Import', 'Rescan', 'Unclassified', and 'Calib. Files'.

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

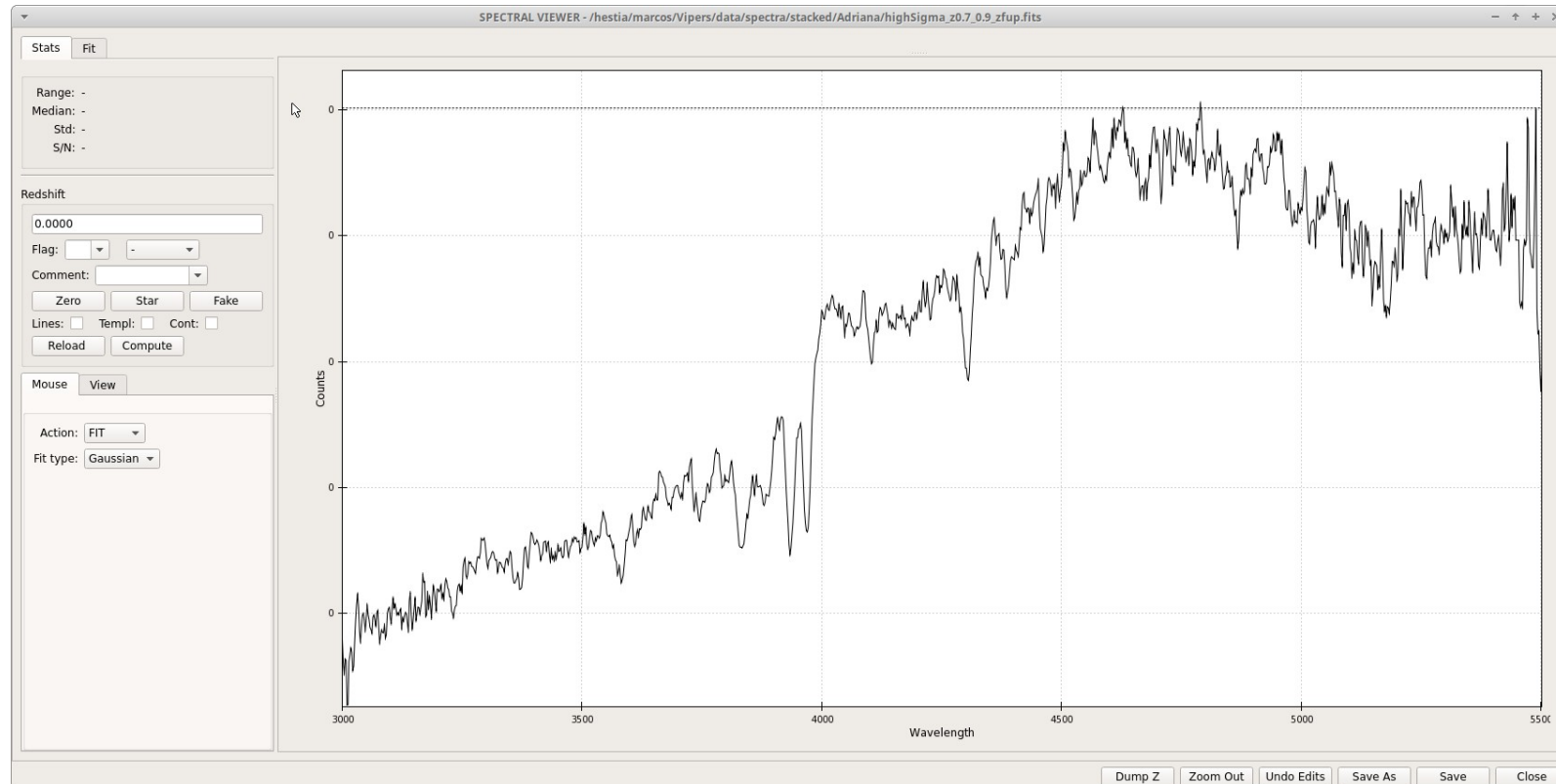
The screenshot shows the 'Project creation' dialog box with the following fields and sections:

- Project name:** A text input field.
- Instrument:** A dropdown menu with 'MODS' selected.
- Root dir:** A text input field.
- Project type:** A dropdown menu with 'random' selected.
- Dataset ID:** A text input field.
- Reduction ID:** A text input field.
- Channel selection:** Two identical panels, each containing a list of channels (CHANNEL, FILTER_NAME, GRISM, INS_MODE, MASK_NAME, MASK_NUMBER, OBSERVING_NIGHT, PI_NAME, POINTING, TARGET) and buttons for moving them (>>, >, <, <<, Move Up, Move Down).
- Directory tree:** Two identical panels, each containing a list of channels (CHANNEL, FILTER_NAME, GRISM, INS_MODE, MASK_NAME, MASK_NUMBER, OBSERVING_NIGHT, PI_NAME, POINTING, QUADRANT, TARGET) and buttons for moving them (>>, >, <, <<, Move Up, Move Down).
- File naming:** A text input field.
- Buttons:** 'OK' and 'Cancel' buttons at the bottom right.

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

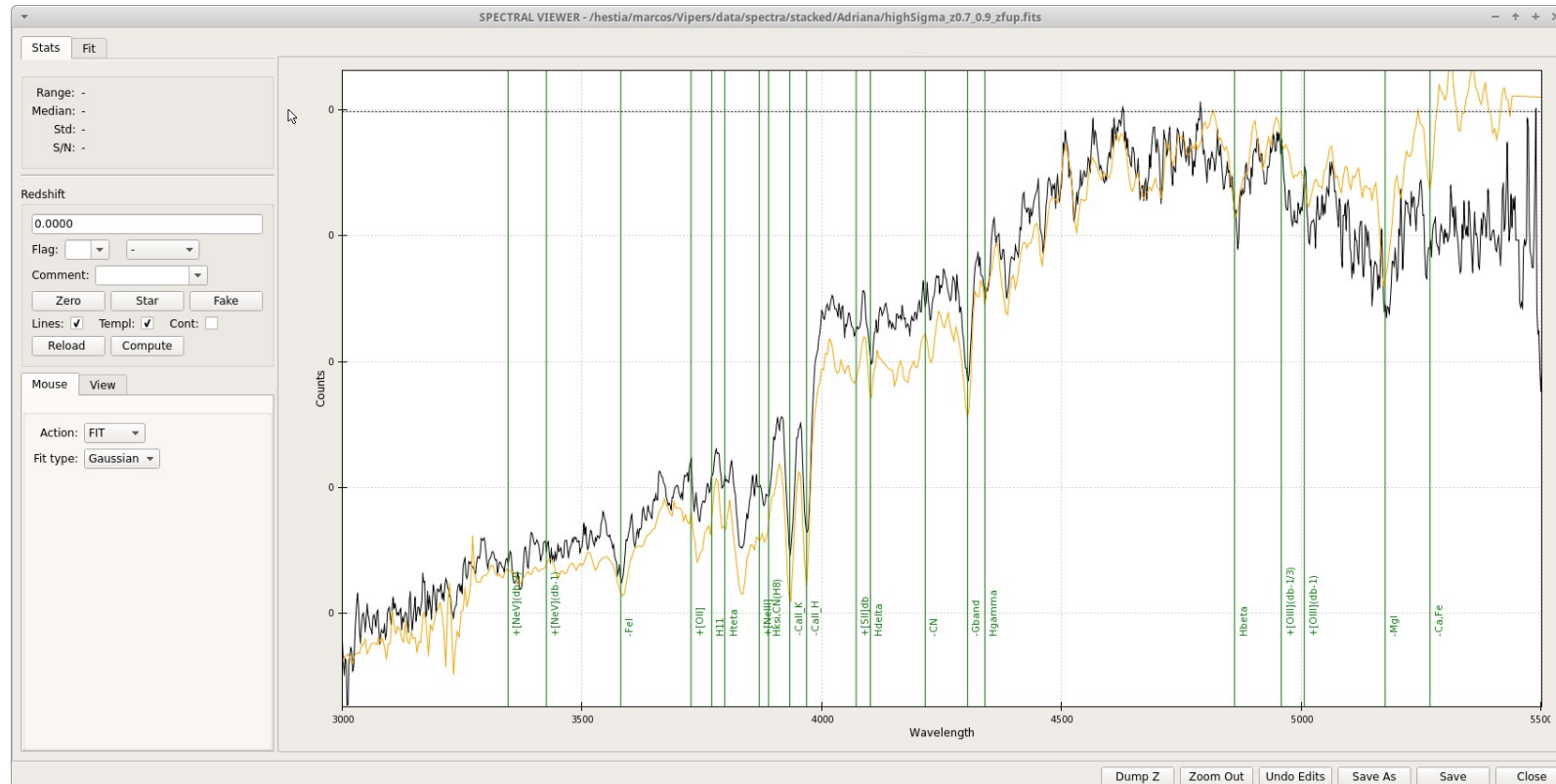


Spectra visualization and redshift measurements: **pandora.ez**

In black:
observed
spectrum

In yellow;
best fitting
template

In green:
most common
spectral lines



Easy**life** for Spectroscopic Surveys

Easy**life**'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

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!

Isberg®