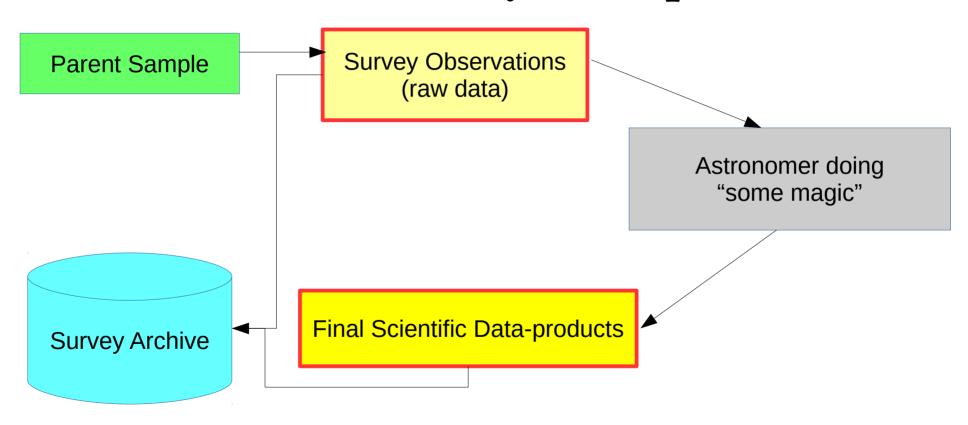
Designing tools to reduce complexity the spectroscopic surveys example

Marco Scodeggio Paolo Franzetti, Marco Fumana & Bianca Garilli

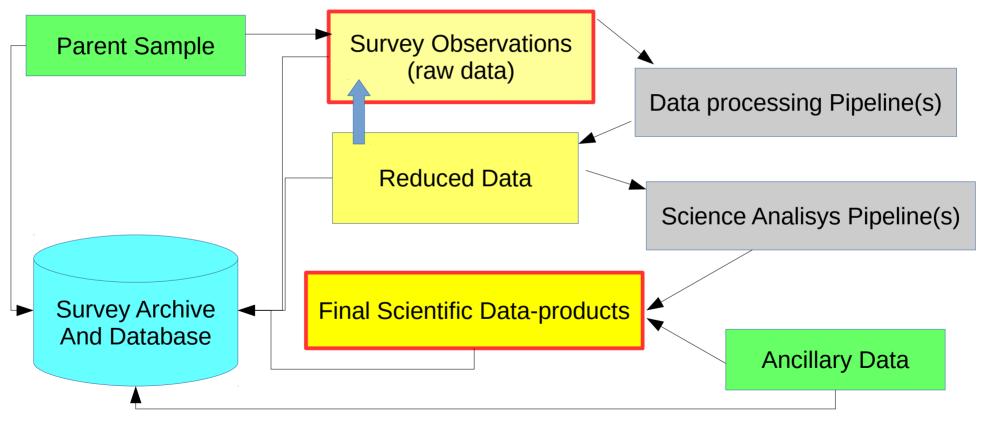
INAF IASF-Milano



The old Survey concept...



A Survey is now a complex "thing"...



June 18, 2019

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

3

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.

INAF Science Archives & Big Data

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....

-					Terminal - root@ti	tan: ~		$- + + \times$
File	Edit Viev	v Terminal	Tabs	Help				
rootā	titan:~# l	s /data/LBT	/RAW/MO	DS/20190201				
		.0001.fits		.20190201.0006.fits	mods1r.20190201.0037.fits	mods2b.20190201.0018.fits	mods2r.20190201.0022.fits	
		.0002.fits		.20190201.0007.fits	mods1r.20190201.0038.fits	mods2b.20190201.0019.fits	mods2r.20190201.0023.fits	
		.0003.fits		.20190201.0008.fits	mods1r.20190201.0039.fits	mods2b.20190201.0020.fits	mods2r.20190201.0024.fits	
		.0004.fits		.20190201.0009.fits	mods1r.20190201.0040.fits	mods2b.20190201.0021.fits	mods2r.20190201.0025.fits	
		.0005.fits		.20190201.0010.fits	mods1r.20190201.0041.fits mods1r.20190201.0042.fits	mods2b.20190201.0022.fits mods2b.20190201.0023.fits	<pre>mods2r.20190201.0026.fits mods2r.20190201.0027.fits</pre>	
		.0007.fits		.20190201.0012.fits	mods1r.20190201.0042.11ts mods1r.20190201.0043.fits	mods2b.20190201.0023.11ts mods2b.20190201.0024.fits	mods2r.20190201.0027.1115 mods2r.20190201.0028.fits	
		.0008.fits		.20190201.0012.fits	mods1r.20190201.0044.fits	mods2b.20190201.0024.11ts	mods2r.20190201.0028.11ts	
		.0009.fits		.20190201.0014.fits	mods1r.20190201.0045.fits	mods2b.20190201.0026.fits	mods2r.20190201.0030.fits	
nods1	b.20190201	.0010.fits	mods1r	.20190201.0015.fits	mods1r.20190201.0046.fits	mods2b.20190201.0027.fits	mods2r.20190201.0031.fits	
nods1ł	b.20190201	.0011.fits	mods1r	.20190201.0016.fits	mods1r.20190201.0047.fits	mods2r.20190201.0001.fits	mods2r.20190201.0032.fits	
		.0012.fits		.20190201.0017.fits	mods1r.20190201.0048.fits	mods2r.20190201.0002.fits	mods2r.20190201.0033.fits	
		.0013.fits		.20190201.0018.fits	mods1r.20190201.0049.fits	mods2r.20190201.0003.fits	mods2r.20190201.0034.fits	
		.0014.fits		.20190201.0019.fits	mods1r.20190201.0050.fits	mods2r.20190201.0004.fits	mods2r.20190201.0035.fits	
		.0015.fits		.20190201.0020.fits .20190201.0021.fits	mods2b.20190201.0001.fits mods2b.20190201.0002.fits	<pre>mods2r.20190201.0005.fits mods2r.20190201.0006.fits</pre>	<pre>mods2r.20190201.0036.fits mods2r.20190201.0037.fits</pre>	
		.0017.fits		.20190201.0021.11ts	mods2b.20190201.0002.1115 mods2b.20190201.0003.fits	mods2r.20190201.00007.fits	mods2r.20190201.0037.11ts	T
		.0018.fits		.20190201.0023.fits	mods2b.20190201.0004.fits	mods2r.20190201.0008.fits	mods2r.20190201.0030.fits	Ц
		.0019.fits		.20190201.0024.fits	mods2b.20190201.0005.fits	mods2r.20190201.0009.fits	mods2r.20190201.0040.fits	
nods1	b.20190201	.0020.fits	mods1r	.20190201.0025.fits	mods2b.20190201.0006.fits	mods2r.20190201.0010.fits	mods2r.20190201.0041.fits	An exam
		.0021.fits		.20190201.0026.fits	mods2b.20190201.0007.fits	mods2r.20190201.0011.fits	mods2r.20190201.0042.fits	
		.0022.fits		.20190201.0027.fits	mods2b.20190201.0008.fits	mods2r.20190201.0012.fits	mods2r.20190201.0043.fits	directory
		.0023.fits		.20190201.0028.fits	mods2b.20190201.0009.fits	mods2r.20190201.0013.fits	mods2r.20190201.0044.fits	directory
		.0024.fits		.20190201.0029.fits	mods2b.20190201.0010.fits mods2b.20190201.0011.fits	mods2r.20190201.0014.fits mods2r.20190201.0015.fits	mods2r.20190201.0045.fits mods2r.20190201.0046.fits	
		.0025.fits		.20190201.0030.fits	mods2b.20190201.0011.11ts mods2b.20190201.0012.fits	mods2r.20190201.0015.fits	mods2r.20190201.0046.11ts mods2r.20190201.0047.fits	the LBT c
		.0001.fits		.20190201.0031.fits	mods2b.20190201.0012.11ts	mods2r.20190201.0010.11ts	mods2r.20190201.0048.fits	
		.0002.fits		.20190201.0033.fits	mods2b.20190201.0014.fits	mods2r.20190201.0018.fits	mods2r.20190201.0049.fits	roduction
mods1	r.20190201	.0003.fits	mods1r	.20190201.0034.fits	mods2b.20190201.0015.fits	mods2r.20190201.0019.fits	mods2r.20190201.0050.fits	reduction

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

Ins. Mode: MOS =	r Data	
MOS	Raw Data Q1 Q2 Q3 Q4 All Quadrants	
Obs. Date: ALL 🗕		
Actions	Show Raw Data - Filter	
Organizer Reduction Browsing Plotting		
Rename Archive Files		
University Collected Films		
Unpack Selected Files		
Unpack All Files		
Rename Selected File		
Delete Selected Files		
Clear Selection		
Reload Organizer Table		
Science2Lamp		
Initialize KBRED		
Run KBRED blindly	Disk Usage: All Data (Gb)	
	0.0 7.5 15.0 22.5 30.0 37.5 45.0 52.5 60.0	

INAF Science Archives & Big Data

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

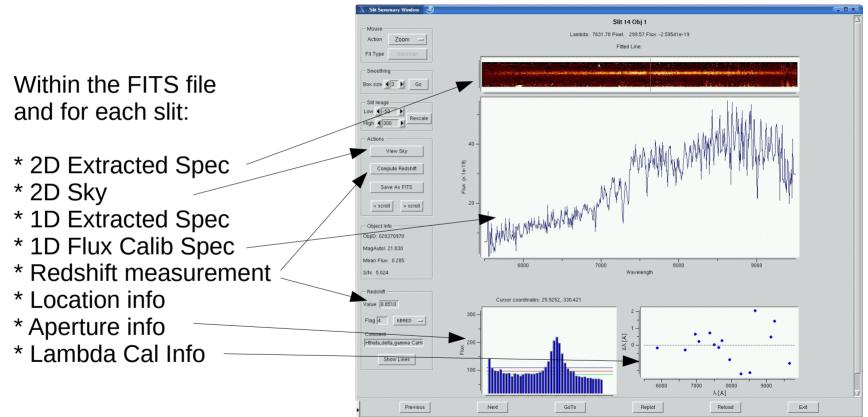
> VIPERS: 288 masks x 4 quadrants x 6 tasks x 20 min = 2400 hours = 300 working days

Ins. Mode: MOS Obs. Date: ALL Actions F02P023_vmO1M1 LR_red Organizer Feduction Browsing Plotting Target: Single Set Create Master Dias Create Master Lamp Preliminary Reduction Create Master Lamp Preliminary Reduction Create Spectra Compute Spectrophotometry Reduce Single Observations Reduce Sequence of Obs. Data Categories Apply Atm. Correction Disk Usage: All Data (3b)				
Actions Single Frames List Organizer Reduction Browsing Plotting F02P041_wm01M1 LR_red Frame Type: science Data Cat: F02P052_vm01M1 LR_red FozP052_vm01M1 LR_red Data Cat: F02P052_vm01M1 LR_red F02P053_vm01M1 LR_red FozP052_vm01M1 LR_red Data Cat: F02P052_vm01M1 LR_red F02P052_vm01M1 LR_red FozP052_vm01M1 LR_red Data Cat: F02P052_vm01M1 LR_red F02P052_vm01M1 LR_red FozP052_vm01M1 LR_red Data Cat: F02P052_vm01M1 LR_red F02P052_vm01M1 LR_red FozP052_vm01M1 LR_red FozP052_vm01M1 LR_red <	Rav		14 All Guadrants	
F02P053_vmO1M1 LR_red sc_F02P052_O1_sa.fts sc_F02P052_O1_aBFCS.fts sc_F02P052_O1_aBFCS.fts sc_F02P052_o1_aBFCS.fts sc_F02P052_vmO1M1 LR_red_O1.fts Master Calibration Frames Data Cat: F02P052_vmO1M1 LR_red_O1.fts msFlat_F02P052_vmO1M1_LR_red_O1.fts msLamp_F02P052_vmO1M1_LR_red_O1.fts msLamp_F02P052_vmO1M1_LR_red_O1.fts msLamp_f02P052_vmO1M1_LR_red_O1.fts msLamp_f02P052_vmO1M1_LR_red_O1.fts msLamp_f02P052_vmO1M1_LR_red_O1.fts msLamp_f02P052_vmO1M1_LR_red_O1.fts msLamp_f02P052_vmO1M1_LR_red_O1.fts msLamp_f02P052_vmO1M1_LR_red_O1.fts msLam	tion Browsing Plotting F02 F02 F02 F02 F02	P029_vmQ1M1 LR_red P040_vmQ1M1 LR_red P041_vmQ1M1 LR_red P052_vmQ1M1 LR_red	Frame Type: science	-
Sc_F02P052_O1_4_BFCS.fits Sc_F02P052_O1_4_BFCS.fits Sc_F02P052_O1_5_BFCS.fits Sc_F02P052_O1_5_BFCS.fits Sc_F02P052_O1_5_BFCS.fits Sc_F02P052_O1_5_BFCS.fits Sc_F02P052_O1_5_BFCS.fits Sc_F02P052_VIDIM_LR_red_O1.fits MsIar_Spec_O1.fits Reduce Single Observations Apply Atm. Correction Disk Usage: All Data [Gb]	Single Set -	PO53_vmQ1M1 LR_red	sc_F02P052_01_5a.ftts sc_F02P052_01_1_BFCS.ftts sc_F02P052_01_1_BFCS.ftts sc_F02P052_01_2_BFCS.ftts sc_F02P052_01_2a_BFCS.ftts sc_F02P052_vmM1_LR_red_01_seq.ftts sc_F02P052_01_3_BFCS.ftts	2
Compute Spectrophotometry Data Cat: F02P052_vmQ1M1 LR_red Reduce Single Observations msFlat_F02P052_vmM1_LR_red_O1.fits Reduce Sequence of Obs. Tab_st_LTT3218_LR_red_O1.fits Apply Atm. Correction Disk Usage: All Data [Gb]	Master Lamp		sc_F02P052_Q1_4a_BFCS.fits sc_F02P052_Q1_5_BFCS.fits sc_F02P052_Q1_5a_BFCS.fits	<u> </u>
Apply Atm. Correction Disk Usage: All Data [Gb]	Spectrophotometry		Data Cat.: F02P052_vmQ1M1 LR_red msFlat_F02P052_vmM1_LR_red_Q1.fits	
Disk Usage: All Data (Gb)	<u> </u>			
	Atm. Correction	Disk Usage: All Data [Gb]		
Batch List		5 15.0 22.5 30.0 37.5 45.0 52.5 (

INAF Science Archives & Big Data

June 18, 2019

Integrated data browsing / plotting tools



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

*		Vipgi				- + ×				
Project Parameters Files										
🗋 🔍 🍝 🔳 🚺										
Reduction Browsing Plot	Projects: MODS 🗢									
Create Bad Pixels Image										
Append Bad Pixels Image	Data Sets 🗢									
Adjust First Guess		TAGET				_				
Auto Adjust		TARGET								
	CALIB test_Slidess_Dual_VFlat_5_ND1.5									
Measure Seeing	CALIBRATION	Sieve_Mask								
Create Master Dark	Chiara_Spiniello	KIDS1453								
Create Master Bias	Chiara_Spiniello	KIDS1453_Acquisition								
Create Image Master Flat	Eliana_Palazzi	PS19qp								
Locate Spectra	Eliana_Palazzi	PS19qp_Acquisition								
Create Master Lamp	Luigi_Mancini	H47wid								
Preliminary Reduction		Mancini H48wid								
	Michele_Fabrizio	Tri2Fab								
Create Sensitivity		NOTSPECIFIED FKJ1344+13								
Create Sensitivity One Click		NOTSPECIFIED PIVOT_Acquisition								
Reduce Single Observations	Roberto_Gilli	Gilli18								
Reduce Sequence of Obs.	Roberto_Gilli	Gilli18_Acquisition								
Apply Atm. Correction	Simone_Antoniucci	D75b_EM_05-104_Acquisition								
Create Summary File	Simone_Antoniucci	IRAS_02276+7225								
	Simone_Antoniucci	V1057_Cyg								
Split 1D Spectra	Simone_Antoniucci	V1057_Cyg_Acquisition								
Measure Redshift	Simone_Antoniucci	V733_cep								
	Simone_Antoniucci	VES_263								
	Simone_Antoniucci	VES_691				=				
	Simone_Antoniucci	VES_691_Acquisition								
	Vincenzo_Testa	J1935+1726				-				
			Import	Rescan	Unclassified	Calib. Files				
					. Change and a solution					

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

						Vi	pgi						- +
oject Parameters Files													
] 🔍 🚱 🗉 🗵		I L L	2 😰 🖉										
duction Browsing	Projects: MODS	•											
Create Bad Pixels Image	Projects. MOD3	•											
Append Bad Pixels Image)	.)				
		li18-ID501202-G		:o_Gilli-Gilli18-ID50 KPTIME DICH	1202-G670L_Dual INAME QUADR4		lli18-ID531947-G400L_Dual	Roberto_Gilli-Gilli18-ID531947-G670	L_Dual Roberto_Gilli	-Gilli18-ID533915-G	400L_Dual R	oberto_Gilli-Gilli	i18-ID53 4
Adjust Hrist Guless													
Auto Adjust	FILENAME ^	FILETYPE	FILTER NAME	DICHNAME	QUADRANT	EXPTIME			DATE OBS				
Measure Seeing	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	05/11/2018 03:38:03		DATE ODS				
Create Master Dark	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 07:16:24						
Create Master Bias	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 07:16:24						
reate Image Master Flat	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 07:38:15						
	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 07:38:15						_
Locate Spectra	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 08:00:33						_
Create Master Lamp	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 08:00:33						
Preliminary Reduction	sc_Gilli18_Non	SCIENCE	Clear	Dual	MODS1B	1200.0	03/01/2019 08:22:24						
Create Sensitivity	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 08:22:24						_
eate Sensitivity One Click	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 08:44:32						_
duce Single Observations	sc_Gilli18_Non		Clear	Dual	MODS1B	1200.0	03/01/2019 08:44:32						
educe Sequence of Obs.	FILENAME A							QUADRANT					
	TILLI VIIIL		1					QUADITIT					
Apply Atm. Correction													
Create Summary File													
Split 1D Spectra	FILETYPE	INS MO	DE GRISM	QUADRANT									
Measure Redshift	**ALL**	\$ **ALL*	* 💠 G400L_Dua	I 🗘 **ALL**	•								
	▲ FILEN/	AME FILET	YPE GRISM	CHANNEL	QUADRANT	VALID FROM			VALID UPTO				
	ImMsFlat	Mar FLAT	G400L_Dual	BLUE	MODS2B	01/01/1900	01/01/2500						
	ImMsFlat	Mar FLAT	G400L_Dual	BLUE	MODS1B	01/01/1900	01/01/2500						
Roberto_Gilli-Gilli18-ID501202-G400L_Dual													
	Koberto_dilli-di	1118-10501202-0	5400L_Dual										

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

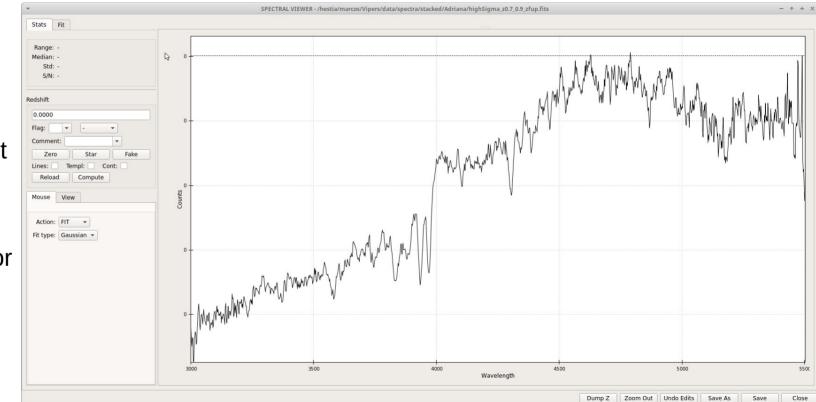
	-	Project	creation	×
	Project name:			
	Instrument: MODS			
	Root dir:			
	Project type: random 💌			
	Dataset ID:		Reduction ID:	
d	CHANNEL FILTER_NAME GRISM INS_MODE MASK_NAME MASK_NUMBER OBSERVING_NIGHT PI_NAME PÕINTING TARGET	>> < < Move Up Move Down	CHANNEL FILTER, NAME GRISM MASK, NAME MASK, NUMBER OBSERVING_NIGHT QUADRANT	>> > < < Move Up Move Down
	Directory tree:		File naming:	
n	CHANNEL FILTER_NAME GRISM INS_MODE MASK_NUMBER OBSERVING_NIGHT PI_NAME POINTING QUADRANT TARGET	>> < < Move Up Move Down	CHANNEL FILTER NAME GRISM INS MODE MASK NAME MASK NUMBER OBSERVING_NIGHT PI_NAME POINTING QUADRANT TARGET	>> > < << Move Up Move Down
			,	OK Cancel

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

June 18, 2019



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

In black: observed spectrum

In yellow; best fitting template

In green: most common spectral lines



INAF Science Archives & Big Data

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) database website public data organization • users data reduction database • parent catalog(s) info ancillary data interna management database • masks • wiki / workgroups masks /OBs preparation team info observations observations mailing lists • measurements monitoring and raw assignments retrival data releases • spectra data distribution for abstracts / papers • measurements derived data data distribution for test analysis

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)

