Data Access and Initial Processing

A.Ederoclite

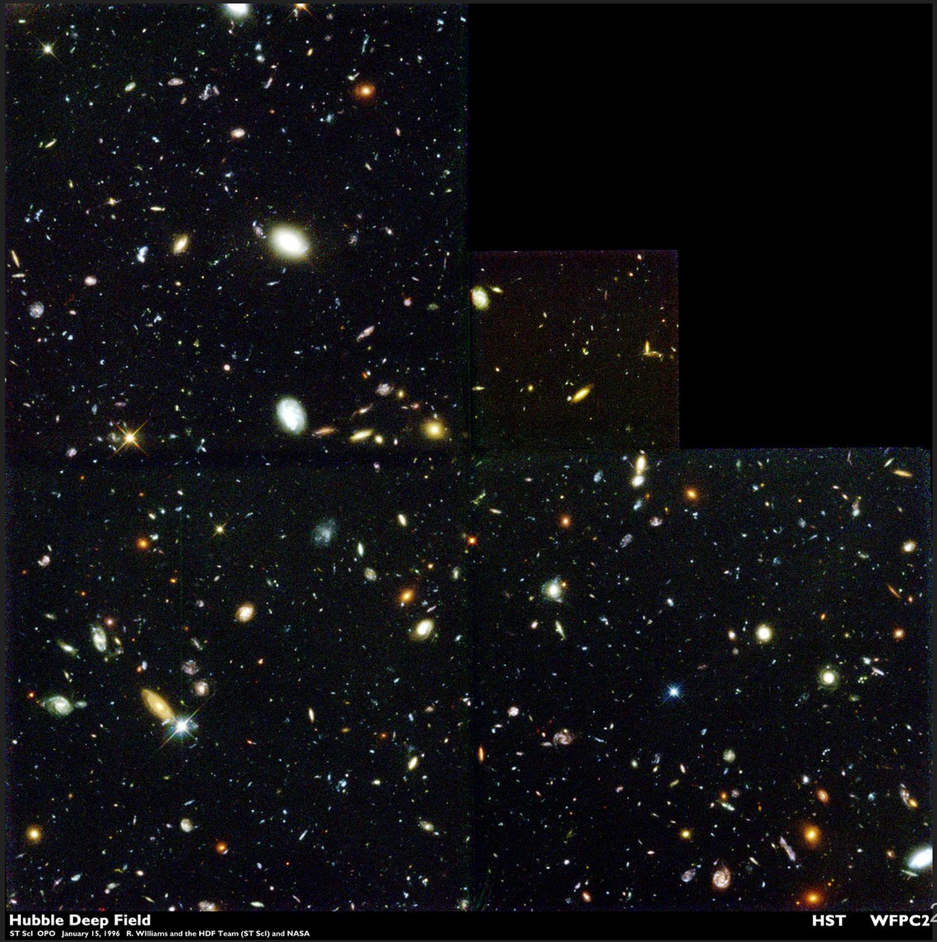
FONDO DE INVERSIONES DE TERUEL
GOBIERNO DE ARAGON
FEDER
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN
AGENCIA ESTATAL DE INVESTIGACIÓN
CEFCA
Centro de Estudios de Física del Cosmos de Aragón
Ayuda ICTS-MRR-2021-03-CEFCA, financiada por:
MINISTERIO DE CIENCIA E INNOVACIÓN
Plan de Recuperación, Transformación y Resiliencia
ESPAÑA PUEDE.
Financiado por la Unión Europea
NextGenerationEU

Images

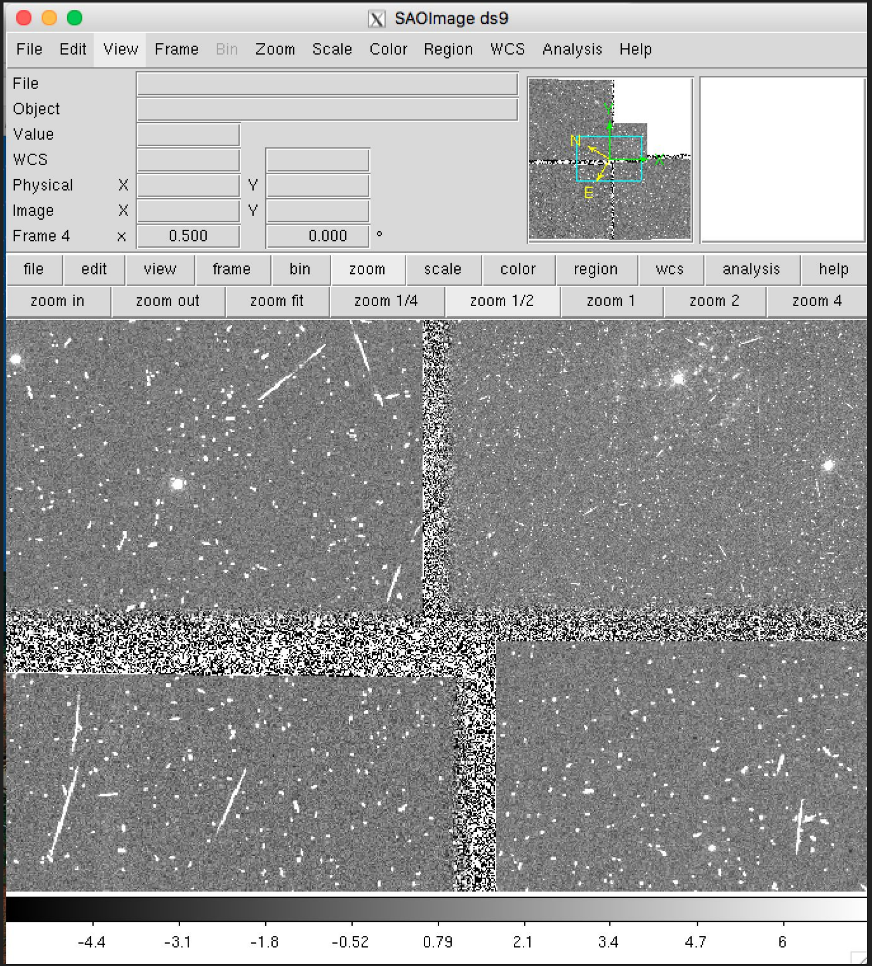
Astronomical images are 2D arrays of "flux".

We tend to see them like this on the newspaper ->

But they look more like this…

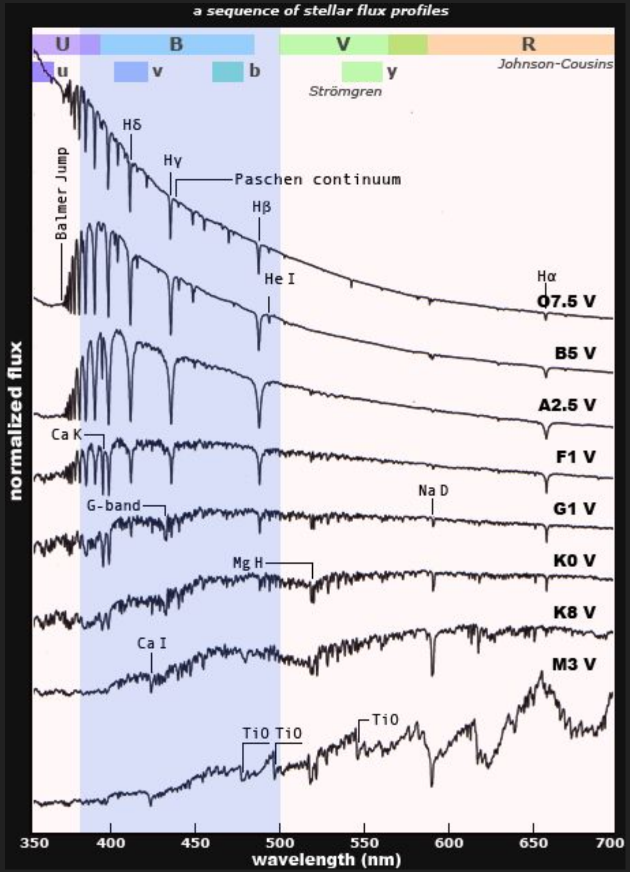


Images from telescopes are in black and white and have all sorts of issues.



Spectra

A spectrum is a 1D array where you have the flux emitted by an object depending on the wavelength (or, which is the same, frequency or energy).

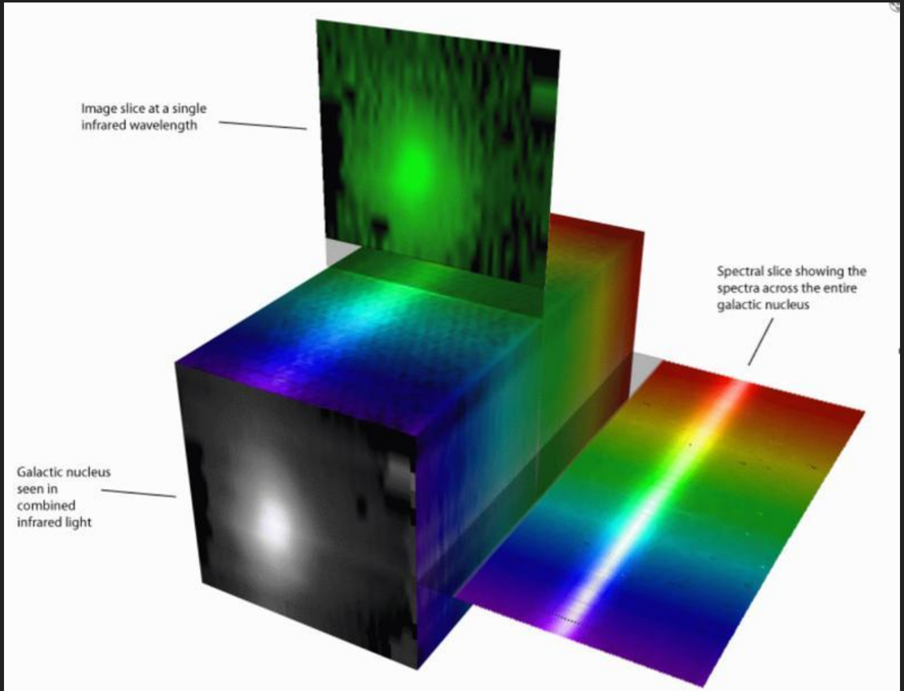


Data cubes

A data cube is a 3D array where (typically), the 3 dimensions are:

- Right ascension
- Declination
- Wavelength

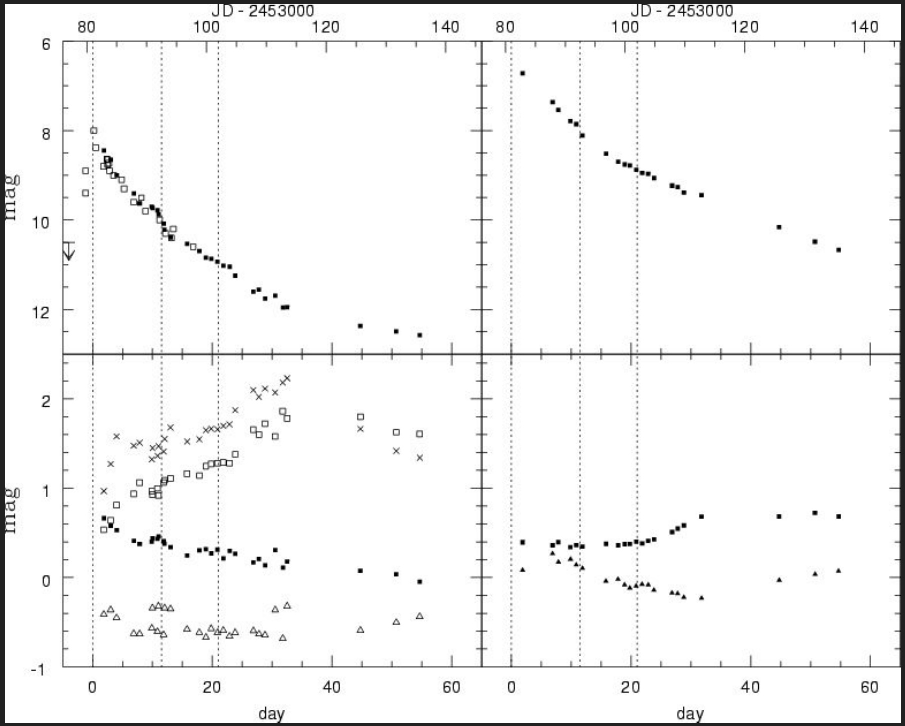
And the flux is displayed in this 3D space.



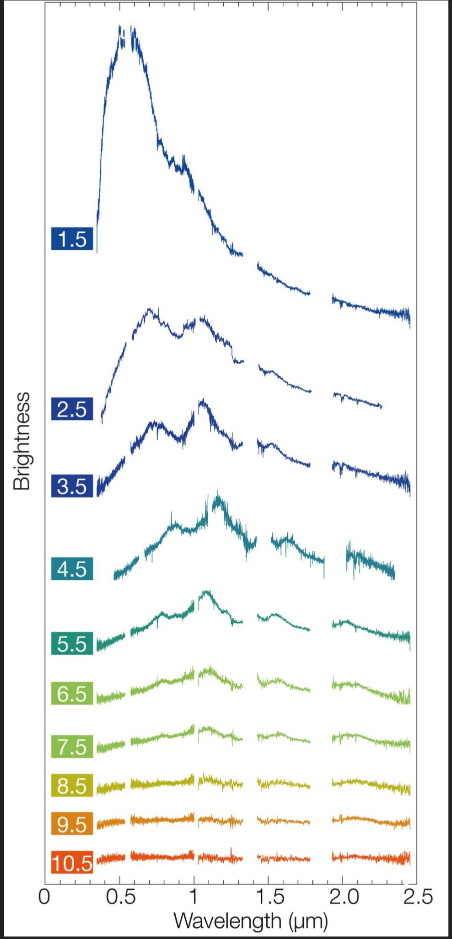
Time series

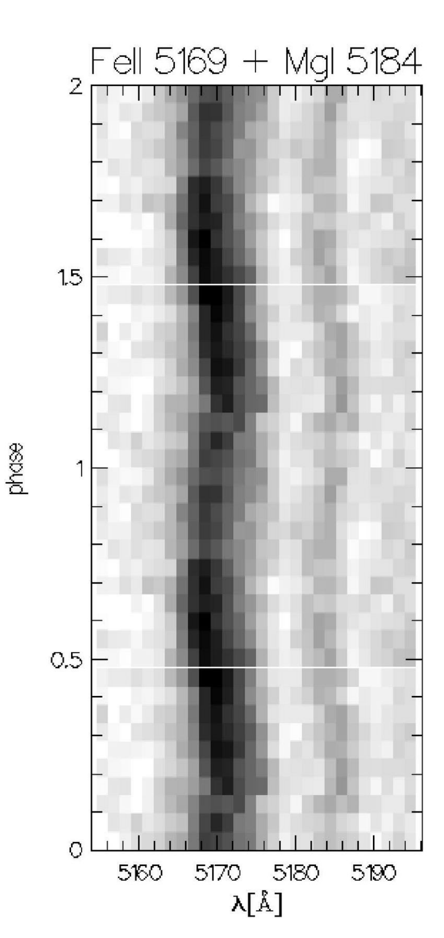
You can plot the magnitude (or the colour) of an object as a function of time. This is a light curve.

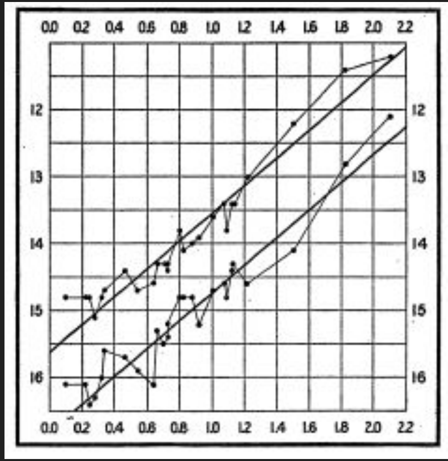
You can also plot a series of spectra.



CEFCA
Centro de Estudios de Física del Cosmos de Aragón







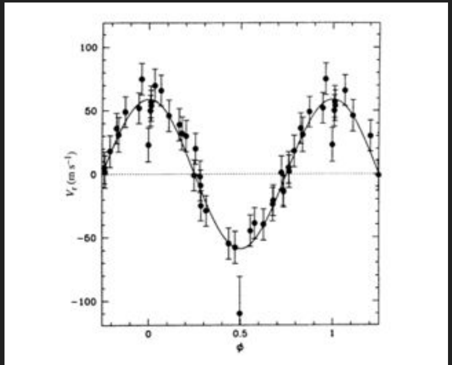


FIG. 4 Orbital motion of 51 Peg corrected from the long-term variation of the γ-velocity. The solid line represents the orbital motion computed from the parameters of Table 1.

Catalogues

A catalogue is (typically) a table with as many columns as needed.

It can be an ASCII, CSV, fits VOTable,...

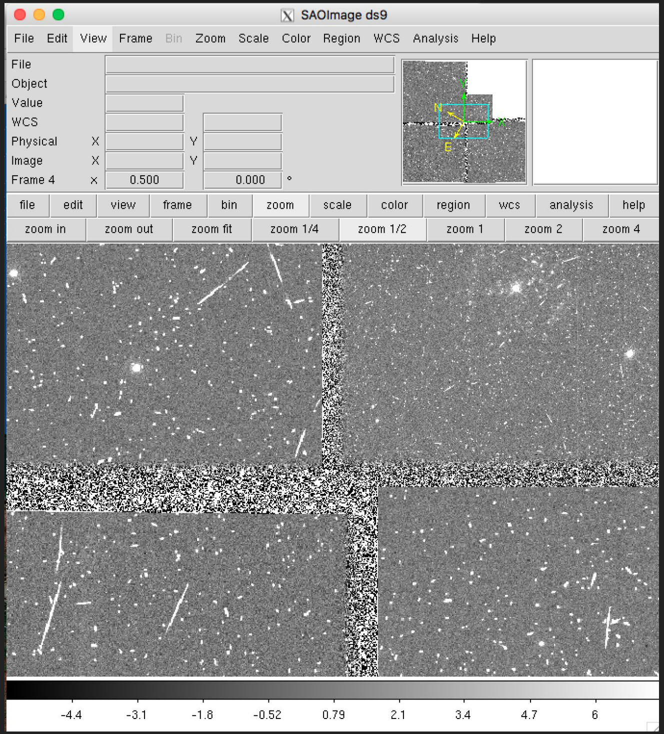
The definition of the columns is **critical**!

Get to know how your catalogue was created before using it!

What do we get out of an astronomical instrument?

- Background noise
- Differences in illumination of the field of view
- “Sky concentration”
- Fringes
- Cosmic rays

(on the right a raw image from HST/WFPC2)



Every time you touch data, you affect them -> noise increases!

Data reduction is painful and everybody makes mistakes.

“Standard” data reduction is a myth.

Images vs spectra

Images: you get position and shape of objects in a given sky region

Spectra: you get the flux emitted at a given energy from a specific sky region

Best of both worlds: integral field spectroscopy - datacubes

Time-series: if you repeat an observation over and over, you get a time-series

“Data reduction”: from raw to science ready

Bias images

(Dark images)

Flat fields

(wavelength calibration - “arcs”)

Extraction of the sources / Extraction of the spectra

Happy science

You want more?

If you participate in the meeting of the Brazilian Astronomical Society, you can attend a workshop on how to reduce images of SPARC4

SPARC4 Pipeline Workshop

26 september, 15h00-18h00

One format to rule them all: the *fits*

"Fits" stays for "Flexible Image Transfer System

This format is endorsed by the International Astronomical Union; here is the webpage of the fits working group https://fits.gsfc.nasa.gov/

Other formats (e.g. theHDF5) are used in astronomy *(but we will focus on fits)*.

In general, a fits has two parts:

- The "header", this is so important that we will discuss it in some detail.
- The data; the data… well, is the data.

In some cases, you can stick together different fits files; each with its header and data section. This is what we call a "multi-extension" fits file (more on this later).

Data in a fits

Fits can host two types of data:

- "Images" - The easiest way to imagine the content of a fits file is an image. The interesting thing that an "image" can have one dimension (a spectrum), two dimensions (an image as you would expect it) or three (a datacube; we will come back to this type of data when speaking about spectroscopy)
- "Tables" - As you would expect, this is a table, with many columns and rows. *Curiosity*: the SDSS spectra are stored as fits tables.

The header

The header carries the information about your data: the “metadata” (*fashion!*).

If you have a collection of files, the first thing you (should) always do is to check what they are. You have two options: either check every object one by one (good luck in telling if you are looking at an image in the *r* filter or the *g* filter) or check what the header tells you.

Note: if the header is not properly filled, you are in **big** troubles.

The World Coordinate System

You hardly ever mess around with the WCS yourself and it is often nicely provided by the observatory.

How do we go from “pixel coordinates” to RA,Dec?

In fact, there are two types of “pixel coordinates”:

- Physical coordinates: the coordinates in terms of pixels which physically exist
- Logical coordinates: the coordinates taking into account pixels which do not physically exist (like a pre-scan region, which may or may not be present)

World Coordinate System (contd.)

At first order, what the WCS does is to give a reference, a rotation and a scale to the image.

The reserved keywords are:

CRPIX*i* - the reference pixel in the *i*-th coordinate (x, y, ….)

CRVAL*i* - the value of the reference pixel

CTYPE*i* - the type of projection used

CD*i_j* - this is a rotation matrix; note that the scale of the pixel is embedded here

Note the use of *i* and *j* and not, necessarily 1 and 2; i.e. this can be anything

A Multi-Extension Fits File (MEF)

In a MEF, the limit is your imagination. A couple of good examples:

If you have multiple CCDs on your detector, you will have:

- Extension 0: it is a “global header” with general information about the observation
- Extension *i* : image taken with your *i*-th CCD with its header and image

Another example can be:

- Extension 0: the “global header”
- Extension 1: an image
- Extension 2: a table with the objects detected on the image

SAOImage DS9

Let’s open an image with the SAOImage DS9 software.
https://sites.google.com/cfa.harvard.edu/saoimageds9

Let’s explore its header:
- WCS
- When was it observed?
- Which filter?

Let’s also play with the zoom and the contrast to see the image better

Searching for data: archives

The ESO archive https://archive.eso.org/cms.html

The STScI archive https://archive.stsci.edu/

The ceFca archive https://archive.cefca.es/catalogues/

The Virtual Observatory

CEFCA
Centro de Estudios de Física del Cosmos de Aragón

What the VO is not

- VO is not a big big computer with a big big archive of astronomical data

- VO is not a big big archive of data

- VO is not a website

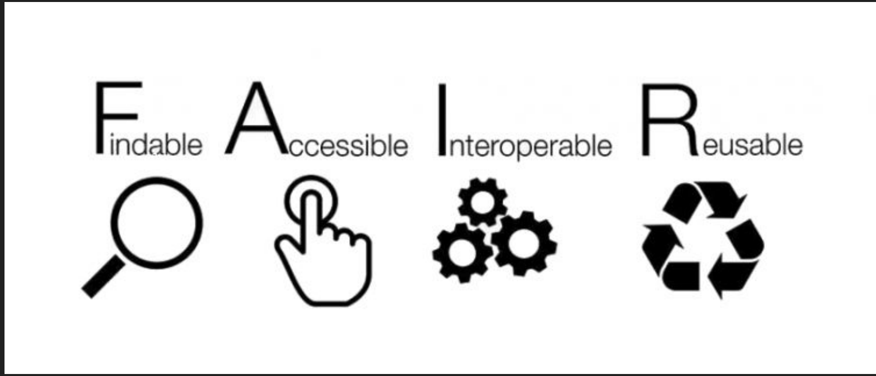
An analogy: internet

- “Internet” exists because of a series of protocols (http, ftp, pop,…) and a series of standards (html,…)

- you use internet through a series of programs (browser, email client,…) which use the standards for you

- Goal: Easy and efficient access and analysis of the information hosted in astronomical archives.

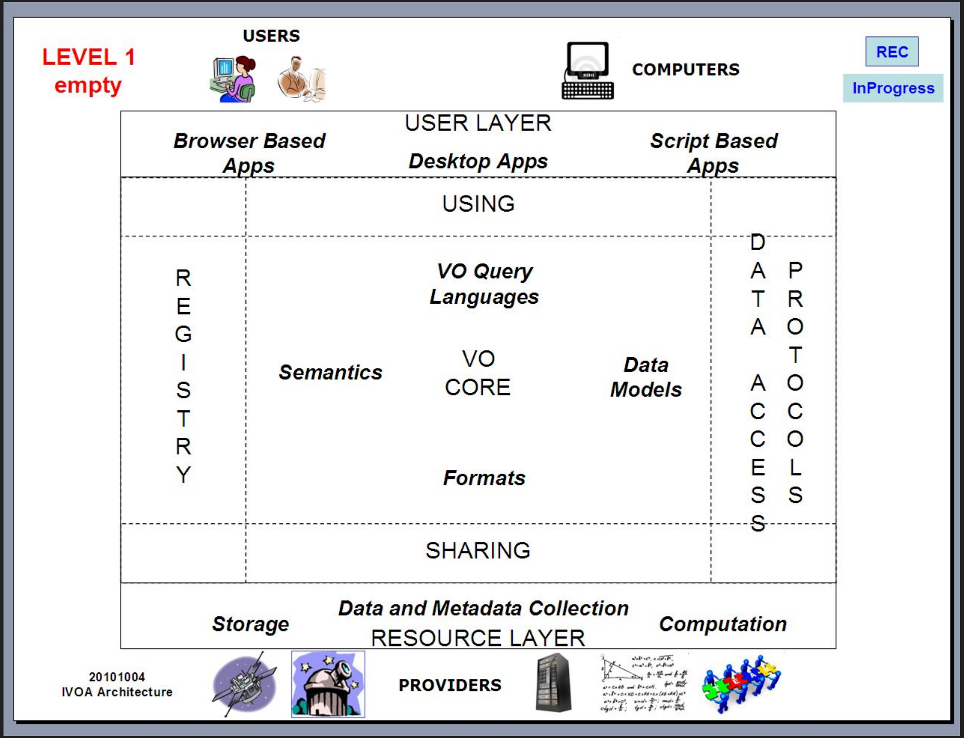
- Making data “FAIR”



Astronomers are up here

VO offers an interface to access the data

The data can be heterogeneous



An example

- Water can be “agua”, “acqua”, “water”, “wasser”,…
- but it is always H_2O
- A user wants “agua”, “acqua”, “water”, “wasser”,...
- H_2O is your resource
- VO offers you a platform so that everybody can share “water” in their preferred language (without the need to learn other languages or chemistry)

A few names

- Basic protocols:
  - SCS - Simple Cone Search - *access a table on position*
  - SIAP - Simple Image Access Protocol - *access a set of images on position*
  - SSAP - Simple Spectrum Access Protocol - *access a set of spectra on position*
  - TAP - Table Access Protocol - *access a database*
  - SAMP - Simple Application Messaging Protocol - *exchange data between VO-tools*

- The registry - *“who’s who”*

CDS: Simbad and VizieR

The Strasbourg Data Centre hosts two services which are vital to most of us:

- Simbad https://simbad.cds.unistra.fr/simbad/sim-fid
  - This is a curated database of astronomical objects
- VizieR http://vizier.cds.unistra.fr/
  - This is a catalogue of catalogues

Let’s look for T CrB in both of them!

ESASky https://sky.esa.int/esasky/

Let’s look for the Hubble Deep Field on ESASky

Images (SIAP), Catalogues (SCS) and Spectra (SSAP).

Even light curves!

Aladin

This is a highly versatile tool for working with images and catalogues.

You can work with images, catalogues, multi-order coverage maps (MOCs),...

You can astrometrise and extract photometry from your images!

It is scriptable!

TOPCAT

This is the go-to program to handle astronomical catalogues.

You can cross match catalogues, explore and plot their contents.

The “engine” running behind it is called `stilts` and this can be run from a command line to make it automatic.

Let’s search for the Hubble Deep Field in Aladin and TOPCAT

Once you have a new object/catalogue

The first thing I do, I check if there is an associated (ideally refereed) paper.

Then I check Simbad.

For catalogues, I use TOPCAT to begin exploring.

Caveat: TOPCAT is not optimal for reproducibility of your results

(more on this tomorrow)