ARTEMIX

Yaye Awa Ba, Philippe. Salomé, Michel. Caillat
(LERMA)
with credits to : L. Loria, N. Kasradze

Archive and Data Mining

l'Observatoire de Paris | LERMA

Goals

(i) Search by products not by instrumental configuration

(ii) Provide trans-project queries (ie famous sources)

(iii) Have a rapid idea of the data content (fits files)

Means

(i) ALMA observing configuration previews (meta-data)

(ii) ALMA cube previews (science products QA2)

Context (1)

(i) To stand just beyond what is provided by the ALMA Observatory Science Archive (not delivering data, not providing material for data reduction). —> Redirection to the ALMA Science Archive

(ii) To use public meta-data and public fits data cubes

(iii) Not to redo what already exist in the ASA (ie rapid metadata query by multi-filters)

Context (2)

Other similar software tools

- A powerful server-side visualization tool (PI: Erik Rosolowsky) which allows users to browse and manipulate the very large ALMA data cubes without having to download them to disk first : CARTA (Cube Analysis and Rendering Tool for Astronomy)
- Japanese Virtual Observatory (JVO) science-ready ALMA images (JVO portal (http://jvo.nao.ac.jp/index-e.html)
- ESO - dev project (Mansardi) study to determine if re-imaging and delivering the full data cubes is feasible (data reduction / QA / storage) + eventually plug automated analysis tools (like ADMIT) —> long term plan

Context (3)

—> To Provide **a pilot study** of **remotely** operated tools for **quick look** visualisation (regular discussions with F. Stoer to keep informed about the current developments

—> Not developed in ESO coding standards, but with selected techno i.e. Serveur HTTP:NodeJS, Database:MongoDB; FITS server:python…

—> Developed at the Paris Observatory, LERMA, in the framework of the French AA-ANO3 duties coordinated by the OASU (OASU, Obs. Paris, OSUG, IRAM)

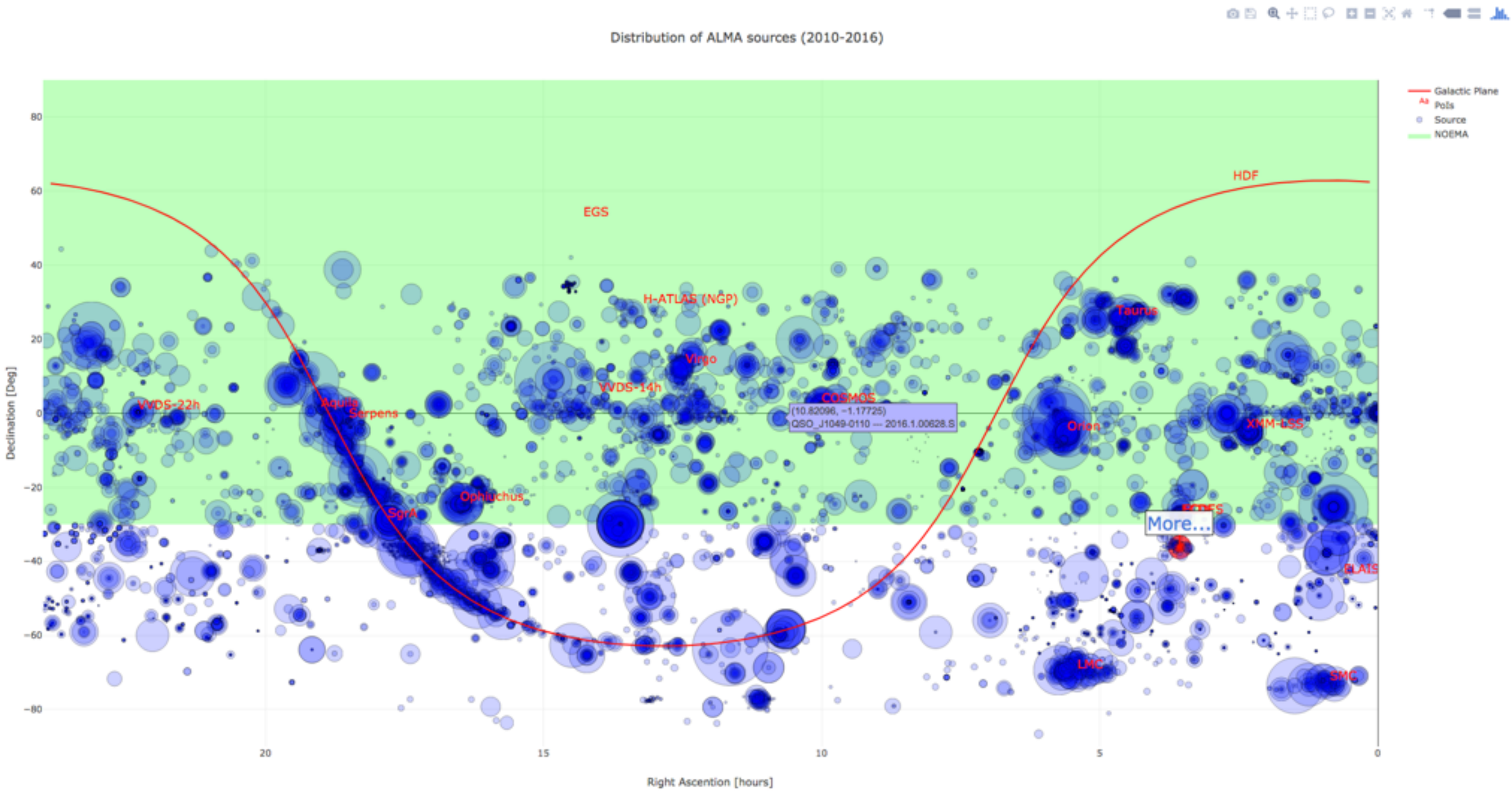
1- Sources Display

—> Very light information. Ingested in a Mongo Data Base on a daily basis

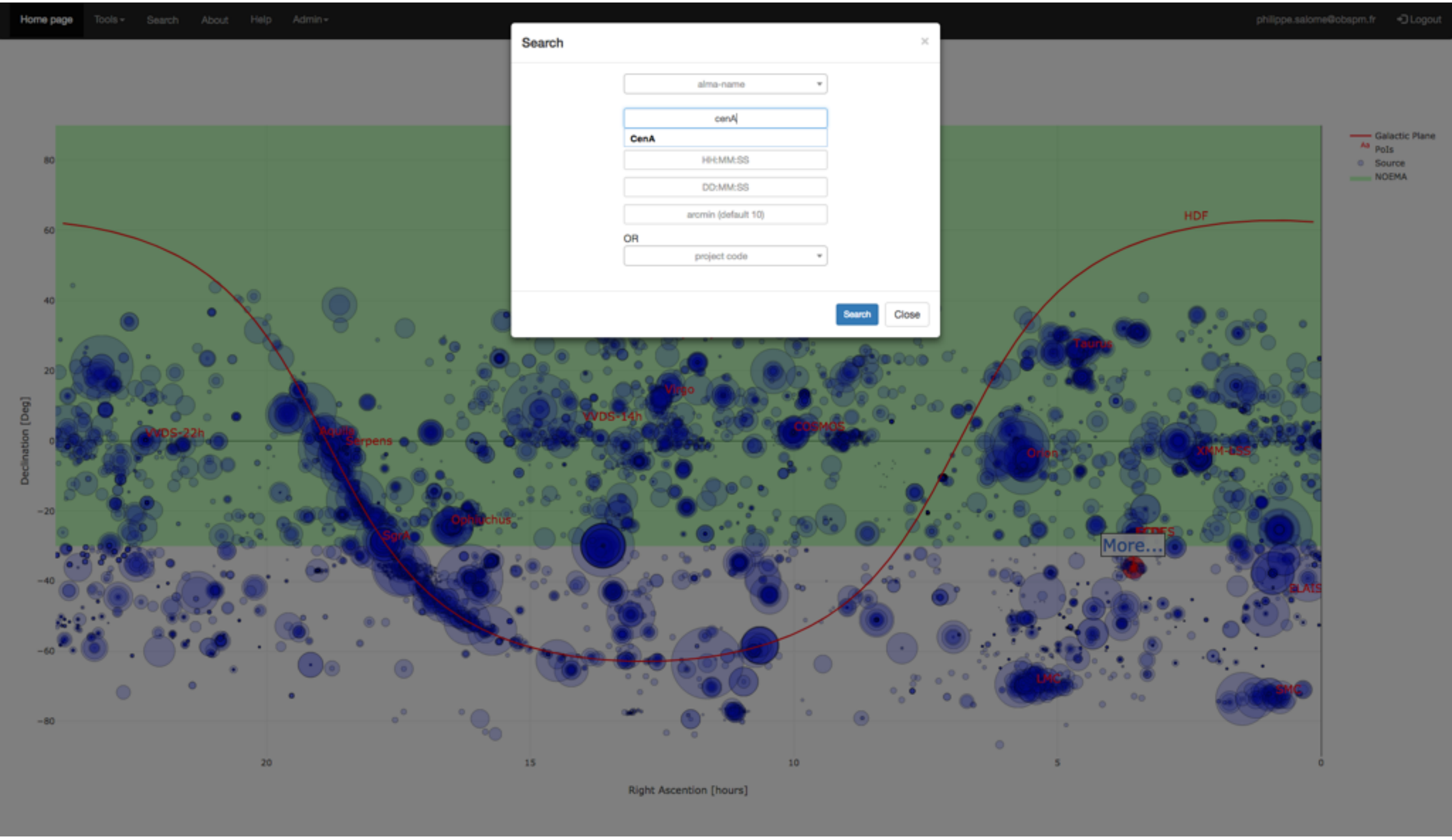
—> Display the sky coverage (RA, DEC) via the plotly library (interactive : zoom, name and position)

—> Basic search method (by sesame-name, alma-name or position inside a given radius) : resolve the source position when necessary, grab pointings (projects) around the object if the source is nearby and extended

Goal : retrieve all the observed projects for a given source (region) of interest (whatever the PI, whatever the project)


Distribution of ALMA sources (2010-2016)
Galactic Plane
PoIs
Source
NOEMA
HDF
EGS
H-ATLAS (NGP)
Virgo
VVDS-14h
COSMOS
VVDS-22h
Aquila
Serpens
Taurus
Orion
XMM-LSS
SgrA
Ophiuchus
More...
LMC
SMC
Declination [Deg]
Right Ascention [hours]

ARTEMIX


Home page
Tools
Search
About
Help
Admin
philippe.salome@obspm.fr
Logout
Search
alma-name
cenA
CenA
HH:MM:SS
DD:MM:SS
arcmin (default 10)
OR
project code
Search
Close
Galactic Plane
PoIs
Source
NOEMA
HDF
Taurus
Virgo
VVDS-14h
COSMOS
VVDS-22h
Aquila
Serpens
Orion
XMM-LSS
Ophiuchus
SgrA
More...
ELAIS
LMC
SMC
Declination [Deg]
Right Ascention [hours]

ARTEMIX

2- Region of Interest

—> Display the **frequency range observed** (basebands) for all the projects that correspond to a given region of interest

—> Display the region of interest into **AladinLite** (help getting velocity / redshift if searched by sesame-name)

—> Display a **table of the header keywords** for all fits files that correspond to this region of interest (different Project codes, different FoV…)

—> (Soon) : display a table with all the metadata for these projects (resolution, t_obs..)

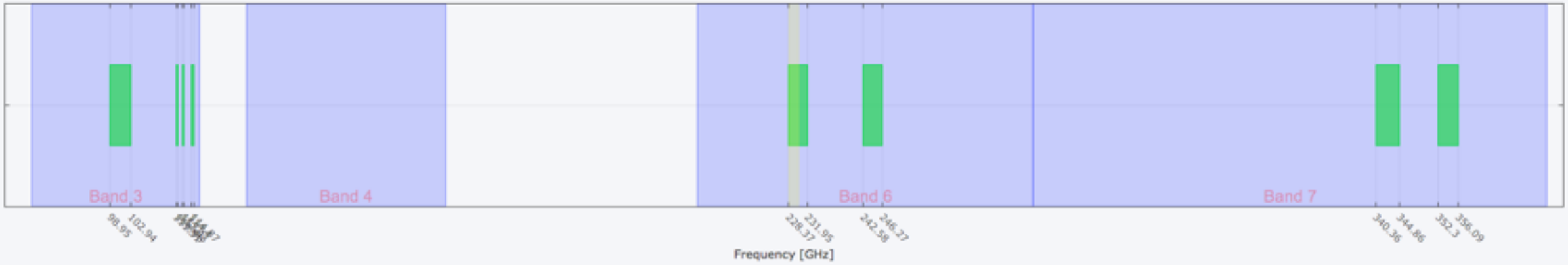
Goal : quick visual inspection of what has been observed .vs. what has been imaged. Provide a link to the data cube (or 2D if in yellow)

—> On click in the table : overlay the frequency slice that has been imaged (**metadata vs fits header**), overlay the FoV (fits header box)

—> Link to **ADS** for publication check

—> Link to **ESO/NRAO ALMA archive** for data retrieval

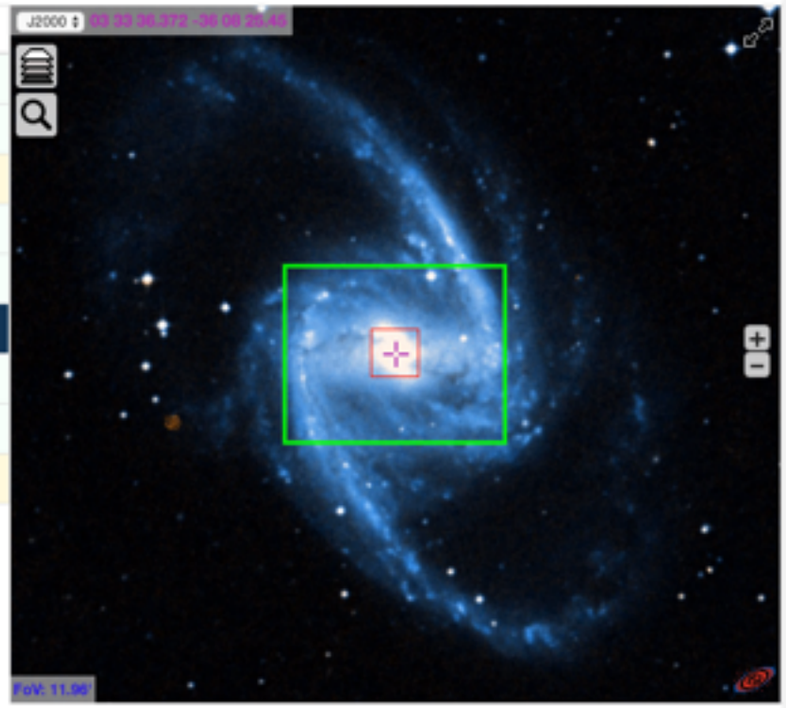
Home page | Tools | About | Help | Admin | philippe.salome@obspm.fr | Logout



Redshift | Velocity (km/s) | All Info

Show all data. Initially shown are the data of which the filename ends with ".pbcor.fits", ".pbcorr.fits", ".image.fits", "line.fits", "cont.fits" or "clean.fits"

#	Fits file	Target	RA	DEC	Cube size	Freq. range	Proj. code	uid
1	Info	NGC1365	03:33:36.38	-36:08:25.70	2916x2560x10	246.011 --- 245.864	2013.1.01161.S	uid://A001/X12f/X321
2	Info	NGC1365	03:33:36.38	-36:08:25.70	2916x2560x30	229.066 --- 229.506	2013.1.01161.S	uid://A001/X12f/X321
3	Info	NGC1365	03:33:36.38	-36:08:25.70	2916x2560x1	228.458 --- 246.218	2013.1.01161.S	uid://A001/X12f/X321
4	Info	NGC1365	03:33:36.38	-36:08:25.70	2916x2560x10	230.572 --- 230.718	2013.1.01161.S	uid://A001/X12f/X321
5	Info	NGC1365	03:33:36.38	-36:08:25.70	2916x2560x11	243.714 --- 243.553	2013.1.01161.S	uid://A001/X12f/X321
6	Info	**NGC1365**	**03:33:36.38**	**-36:08:25.70**	**1344x864x500**	**228.4 --- 230.323**	**2013.1.01161.S**	**uid://A001/X12f/X319**
7	Info	NGC1365	03:33:36.38	-36:08:25.70	216x216x497	228.444 --- 230.355	2013.1.01161.S	uid://A001/X12f/X31b
8	Info	NGC1365	03:33:36.99	-36:08:36.33	2048x1500x121	229.064 --- 229.523	2013.1.01161.S	uid://A001/X12f/X317
9	Info	NGC1365	03:33:36.99	-36:08:36.33	2048x1296x1	229.837 --- 246.192	2013.1.01161.S	uid://A001/X12f/X317



ARTEMIX

3- Quick Look Viewer

—> **Display the data cube** (2 images, 2 spectra) : 1 channel map, 1 moment map, 1 spectra extracted from a pixel, 1 spectra extracted from a spatial region (square). Interactive and self-consistent

—> Based on **GILDAS Mapping « go view »**. Same functionalities implemented (frequency selection, region selection, integrated flux computation)

—> Link to the detailed fits header

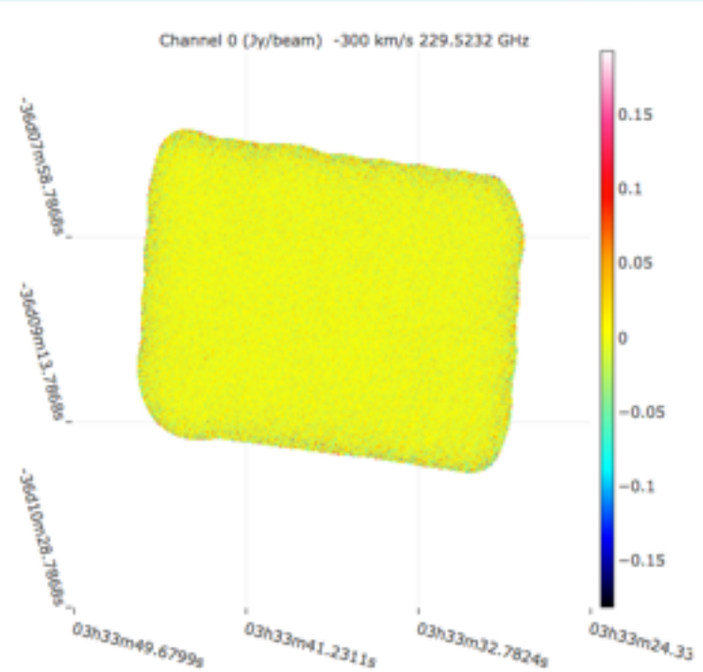
Goal : give a quick look preview of the data cube content

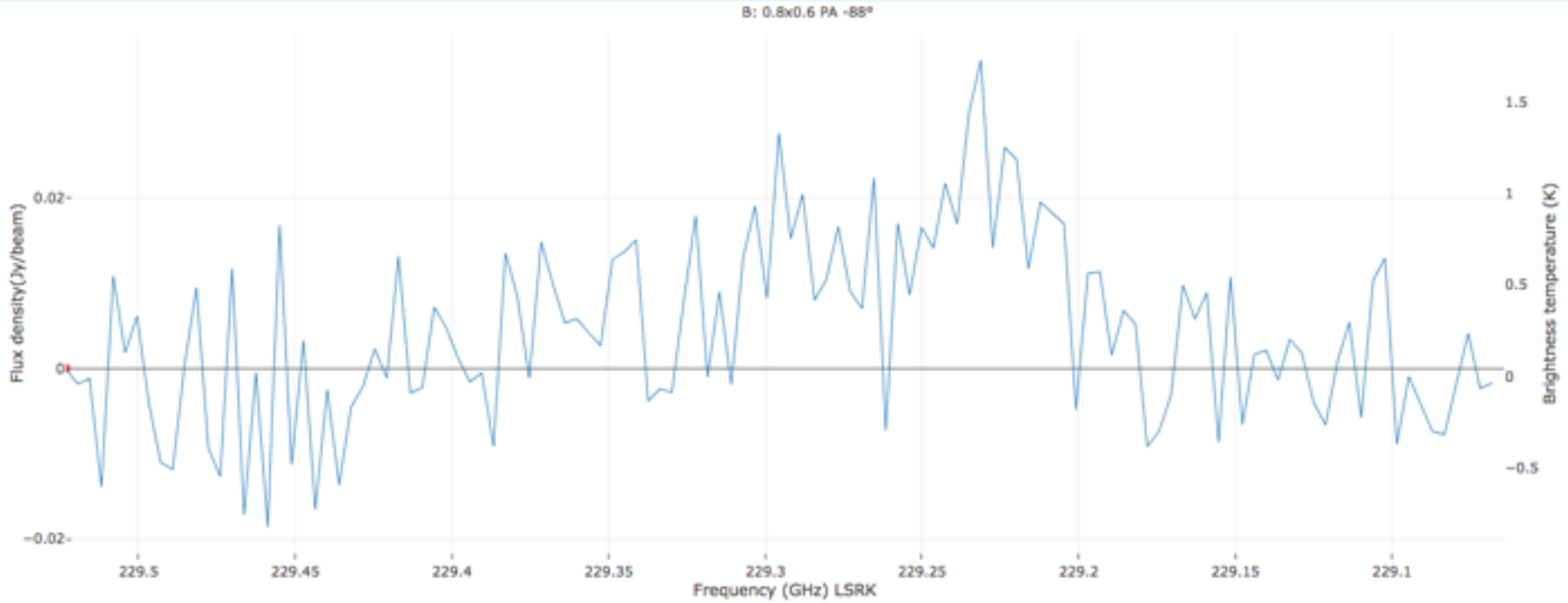
- Sometimes (oldest data) the x-axis reference frame varies (LSR, HELIO, BARY…). No computation done apart Freq —> Vel and Vel —> Freq in the same frame.
- Only slices have been extracted (not always the best cut-out or spectral resolution). No re-calibration applied.

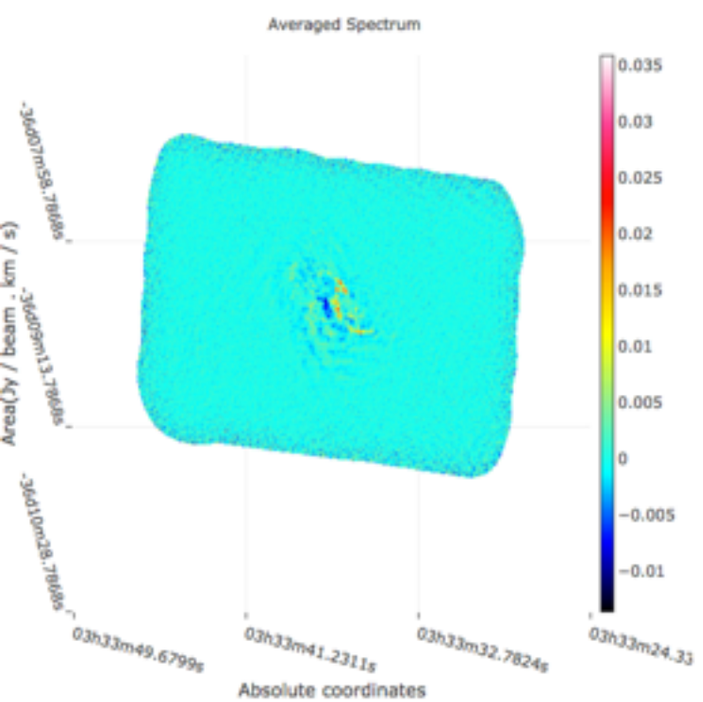
—> PI are encouraged to retrieve the ALMA raw data and use the standard pipeline

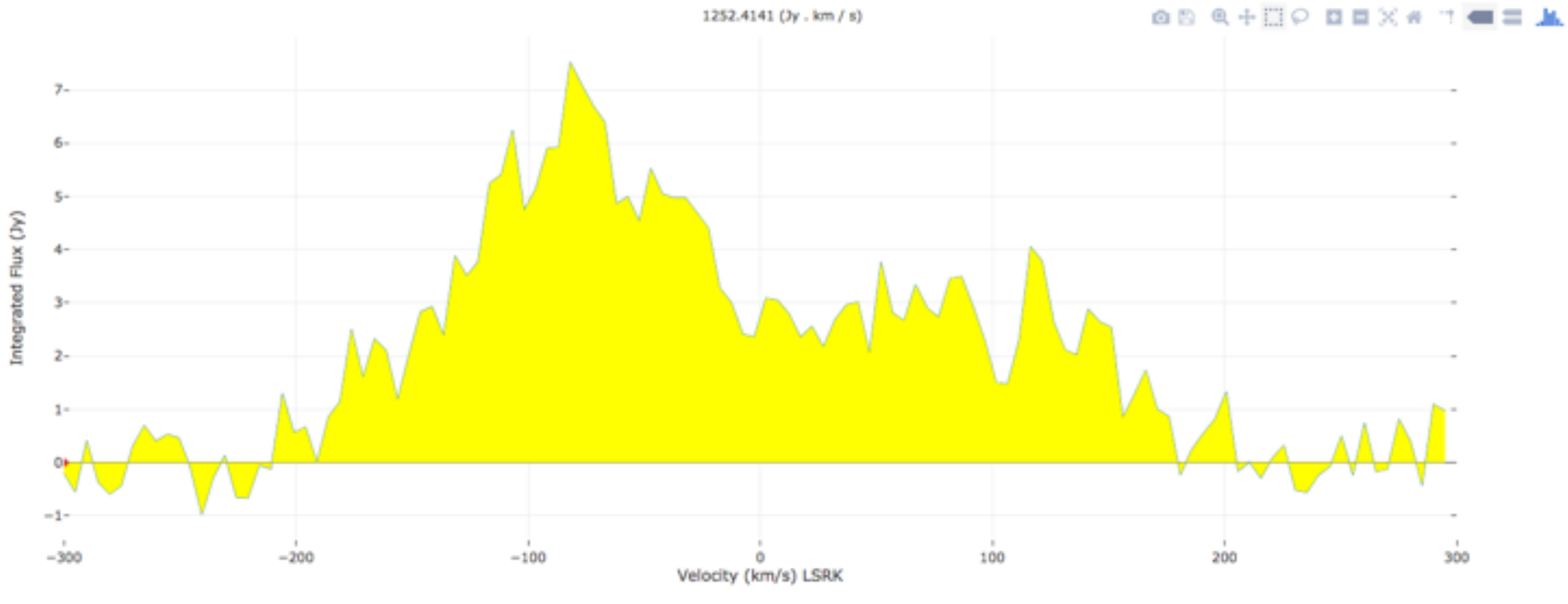
Home page | Tools | 0-10000 | Open | Recent files | Open | About | Help | Admin | philippe.salome@obspm.fr | Logout

uid___A002_Xa16f89_X2a8c.ms.split.cal.image.spw0.co21.image.pbcor.fits - OBJECT = **NGC1365** - NAXIS = **4** - NAXIS1 = **2048** - NAXIS2 = **1500** - NAXIS3 = **121** - NAXIS4 = **1** etc.









Volumetry

Data cubes (ie 512 x 512 x 3000) in (x,y,v) —> 10s of MB to 10s of GB (16-20 GB at max). Difficult to handle and visualise cubes larger than that anyway. This would need other methods (ie cut-out service)

Note : The fits file are produced in QA2 —> only slices around the line of interest (less than ~10% of the observed (x,y,v)). Here again, PIs are advised to retrieve the raw data from the ALMA Science Archive.

Total (all fits, as of last week) : ~ 6 TB (updated on a daily basis : not yet achieved)

- 3.5 TB de fichiers > 1 GB
- 2.6 TB de fichiers > 2 GB
- 2.1 TB de fichiers > 4 GB

Solution for fast reading : SSD PCI Express. Very fast load of the data. Fast computation (means are parallelized). Slow part : the network transfert 2 x (NxN) and 2 x Nchans

Next : a cut-out Service ?

A cut out service to display / analyse only part of the data fits file.

1. Spatially (sub-region in the map) to reduce x-axis and y-axis size
2. In Velocity/frequency (range of channels).

To do so, there is a need for

- tools to help the user to choose which part to look at
- tools to split the files in order to scan all the data cube (or all the selected area if still too big).

Local / Remote

Local

Large range of analysis tools

But

Need to download all the fits files to be checked (even if no detection)

Speed limited by local computer performances and/or software optimization (for display) —> often need a local server

Remote

Optimization on dedicated machines (load fits, calculations)

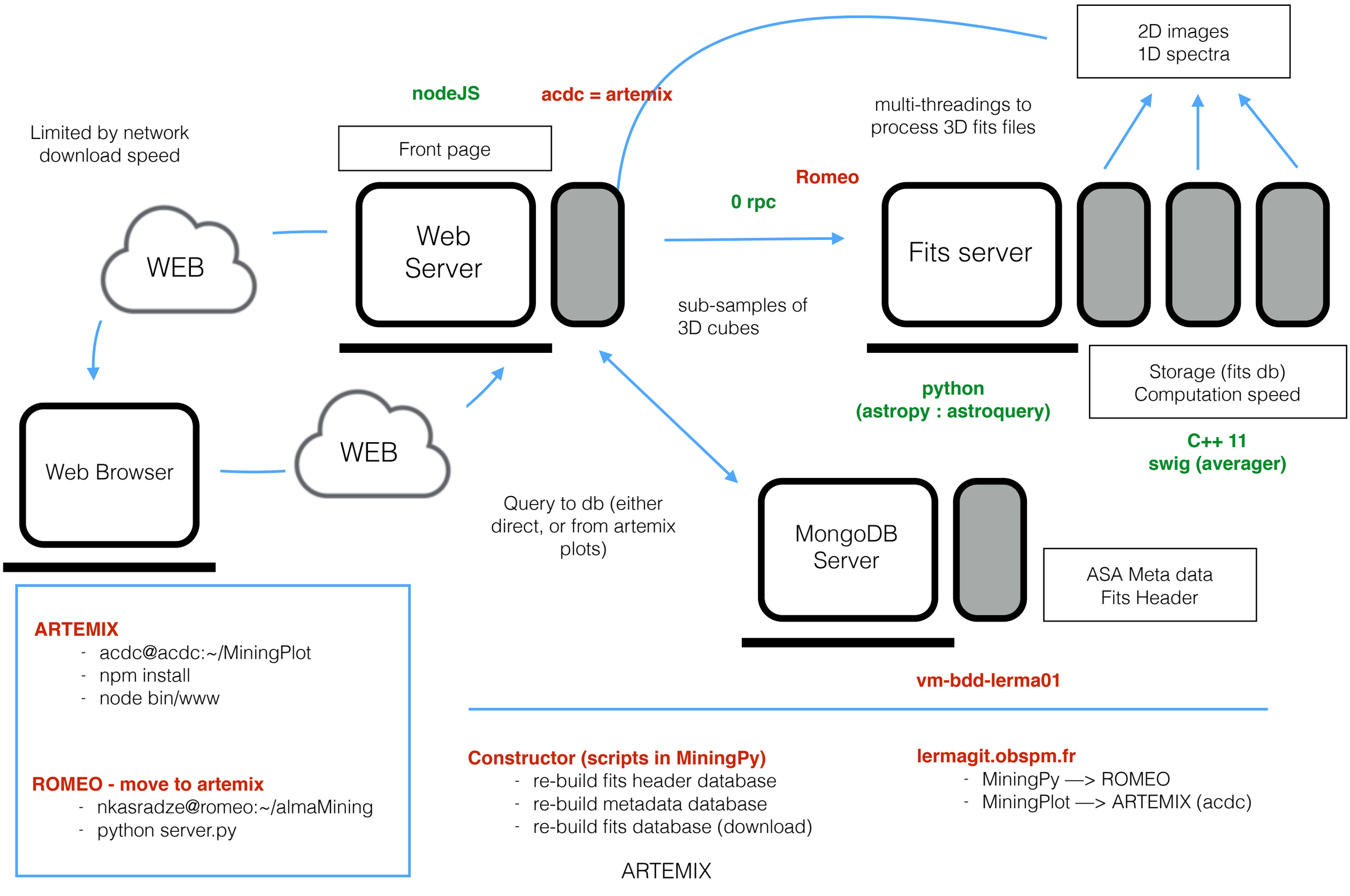
No need to download fits files on local disk (if many and from different projects)

But

Delay for download (11 MB/s at most) and loading large file (> 2GB)

Limited analysis (yet)

Overview



ARTEMIX

- acdc@acdc:~/MiningPlot
- npm install
- node bin/www

ROMEO - move to artemix

- nkasradze@romeo:~/almaMining
- python server.py

Constructor (scripts in MiningPy)

- re-build fits header database
- re-build metadata database
- re-build fits database (download)

ARTEMIX

lermagit.obspm.fr

- MiningPy —> ROMEO
- MiningPlot —> ARTEMIX (acdc)

Thanks !