Data discovery in planetary science

Where it help in data preservation

Pierre Le Sidaner, Cyril Chauvin, Regis Haigron, Chloé Azria, S. Erard, B. Cecconi

Digital data in astronomy

Preservation came 20 years ago with OAIS

Valorisation come with FAIR

(interoperable but also findable and accessible)

- Preservation with metadata
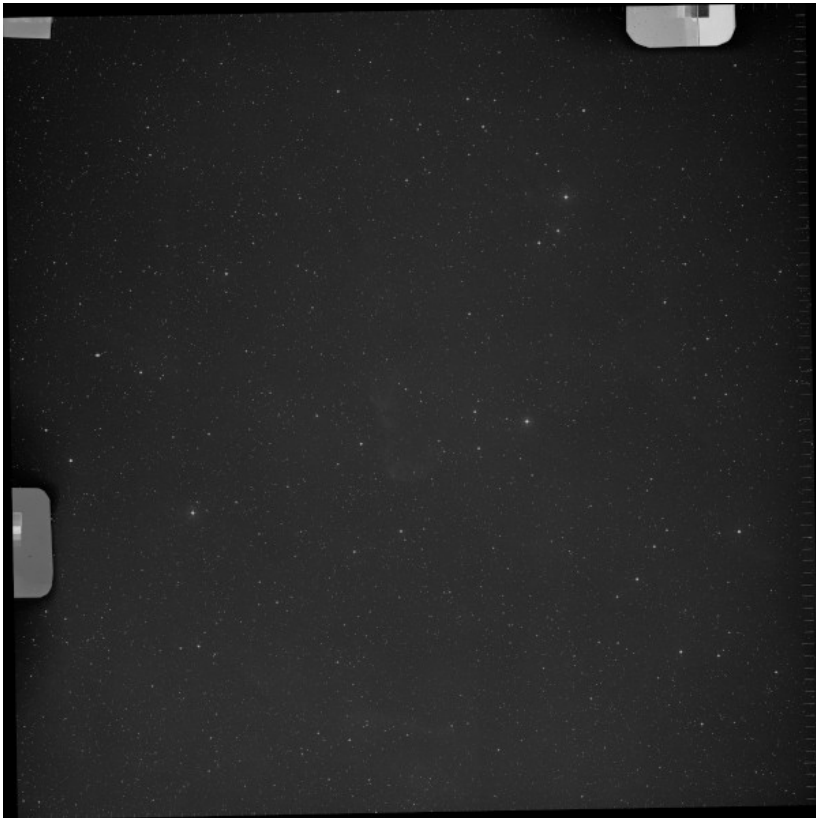

```
SIMPLE  =                    T / This is a FITS file
BITPIX  =                   16 /
NAXIS   =                    2 /
NAXIS1  =                34481 / NUMBER OF ELEMENTS ALONG THIS AXIS
NAXIS2  =                34494 / NUMBER OF ELEMENTS ALONG THIS AXIS
EXTEND  =                    T / This file may contain FITS extensions
EQUINOX =        2000.00000000 / Mean equinox
MJD-OBS =    4.301900000000E+04 / Modified Julian date at start
RADESYS = 'ICRS    '          / Astrometric system
CTYPE1  = 'RA---TAN'          / WCS projection type for this axis
CUNIT1  = 'deg     '          / Axis unit
CRVAL1  =    3.468331946005E+01 / World coordinate on this axis
CRPIX1  =        1.707550E+04 / Reference pixel on this axis
CD1_1   =   -1.865469821496E-04 / Linear projection matrix
CD1_2   =    0.000000000000E+00 / Linear projection matrix
CTYPE2  = 'DEC--TAN'          / WCS projection type for this axis
CUNIT2  = 'deg     '          / Axis unit
CRVAL2  =   -8.446640080320E+01 / World coordinate on this axis
CRPIX2  =        1.890150E+04 / Reference pixel on this axis
CD2_1   =    0.000000000000E+00 / Linear projection matrix
CD2_2   =    1.865469821496E-04 / Linear projection matrix
EXPTIME =    0.000000000000E+00 / Maximum equivalent exposure time (s)
GAIN    =    0.000000000000E+00 / Maximum equivalent gain (e-/ADU)
SATURATE=    4.999782311225E+04 / Saturation Level (ADU)
COMMENT
SOFTNAME= 'SWarp   '          / The software that processed those data
SOFTVERS= '2.38.0  '          / Version of the software
SOFTDATE= '2014-05-27'        / Release date of the software
SOFTAUTH= '2010-2012 IAP/CNRS/UPMC' / Maintainer of the software
SOFTINST= 'IAP  http://www.iap.fr' / Institute
COMMENT
AUTHOR  = 'lesidaner'         / Who ran the software
ORIGIN  = 'CAI-MAMA-VO-PARIS Observatoire de Paris' / Where it was done
DATE    = '2014-06-30T15:35:09' / When it was started (GMT)
COMBINET= 'AVERAGE '          / COMBINE_TYPE config parameter for SWarp
COMMENT
COMMENT  Propagated FITS keywords
VOLTAGE =    8.210000000000E+00 / Lamp voltage of the scanning machine
FSTPAV  = 'SRC003JPB'         / Name of the original pave
INSTRUME= 'MAMA    '          / Microdensitometer
ATLAS   = 'SRC Blue-Atlas '   / Atlas name
FLD     = '003     '          / SRC-J atlas field number
PLTLABEL= 'J2565   '          / Observatory plate label
EMULSION= 'IIIaJ   '          / Photographic emulsion
FILTER  = 'GG395   '          / Filter
ALPHA_50= '02:24.0 '          / Approximate 1950 R.A. for the plate centre
DELTA_50= '-85:00  '          / Approximate 1950 Dec for the plate centre
DATE-OBS= '1976-08-29'        / UT date of Observation
JDAY    = '2443020.27321'     / Julian date of observation at mid exposure
LST_SOE = '02:28:00'          / Local sidereal time at start of exposure
PLTDATE = '1976.6626'         / Decimal date of the plate exposure
EXPOSURE= '60.0    '          / Exposure time (minutes)
PLTGRADE= 'AI2     '          / Plate grade
TELNAME = 'Siding Springs'    / Telescope location
```

What to do with png file without metadata

Digital data in astronomy

- Preservation deal with FAIR

Preservation <=> futur usage of data
    Reusable mean no preservation without metadata

Standardized metadata comes with IVOA
 (interoperable but also findable and accessible)

Planetary science & VESPA

How to make Planetary science data available without reinventing the wheel

It come from several communities :

- **Solar physics**

    surface, activity, corona, wind

- **Planets**

    Interior, surface (OGC), atmosphere

- **Plasma**

    interplanetary medium, solar wind, planetary atmosphere

- **Small bodies & comets**

- **Exoplanets**

- **Minerals and samples**

Standards data format

- Technical issues

- **ASCII tables**
- **Images not only fits**
- **Spectrum (not only incident light, minerals ...).**
- **Dynamic spectrum**
- **OGC data and services**
- **Orbitographic data (spice kernel)**
- **Maps**
- **Events (VO)**
- **.....**

Vespa ⊆ IVOA

Long time maintenance => use existing eco-system

- **adoption of IVOA with its eco-system (registry, TAP, Applications)**
- **deviation of obscore into EPN-CORE**
- **list of mandatory parameters**
- **optional parameter list to satisfy sub-communities**
- **no more RA DEC access => time, objects, mission ....**

- **Define a flat data model**

- **Define a web client to access all data**

- **Double complexity**
  - **Each community have their usage**
  - **Vespa started with only few services**

Vespa client

Already more than 60 services

- **Each service appear indepently in the portal**
- **NASA (PPI) come with 170 services**
- **Vespa succes increase with 20 services / year**

http://vespa.obspm.fr

devlopment version
http://voparis-europlanet-dev.obspm.fr/

- ## **How web portail can handle that?**
  - **Thematic sort ?**
  - **Other grouping of services ?**
  - **What is the response time to the query ?**
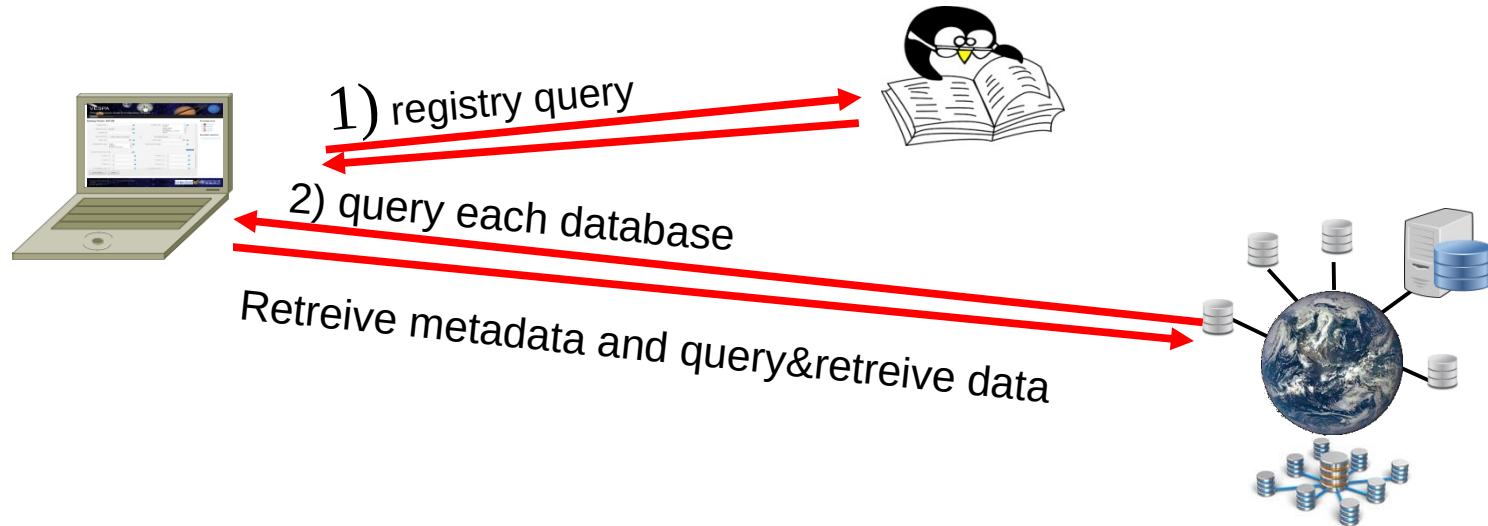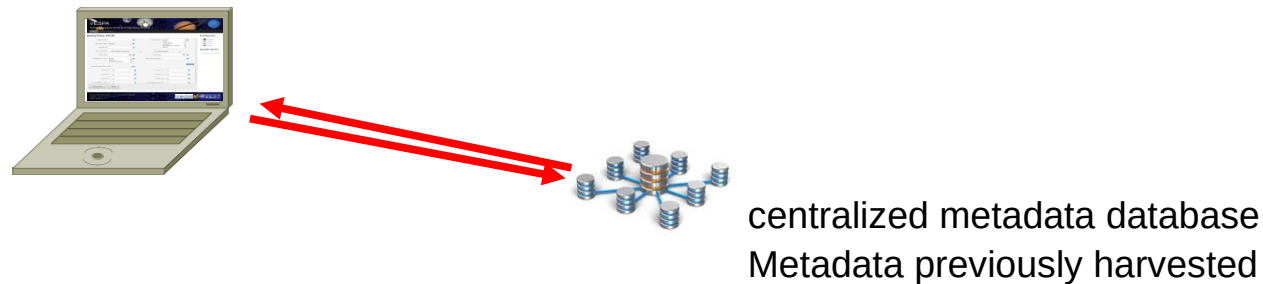 - **When does the dynamic display break ?**

User feedback
Complains

- You can't find what you're looking for ?

- It's not intuitive !

- We don't work like that !

- Everyone has a point of view on ergonomy !

Local aproach

Distributed architecture IVOA

1) registry query

2) query each database

Retreive metadata and query&retreive data

Local metadata database

centralized metadata database
Metadata previously harvested

Local aproach

- retrieval of the 34Millions metadata set
- ingestion in a search engine.

- Creation of spatial range zone (time, frequency) to be treated separately

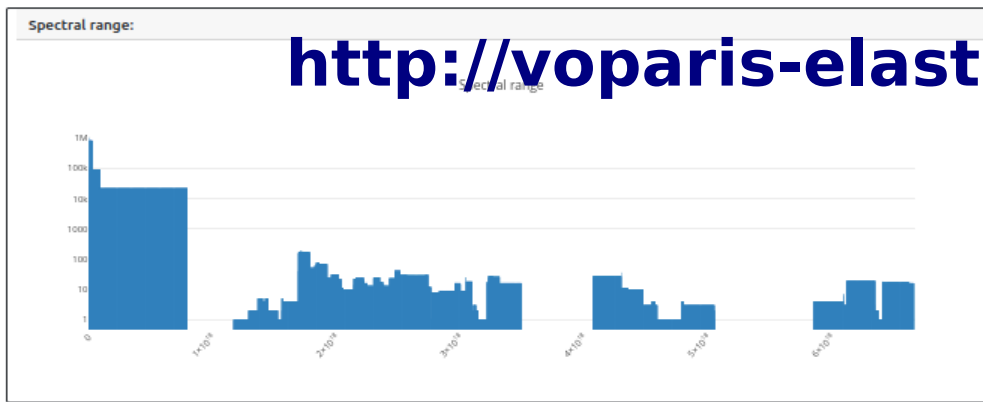   Polygons management problem
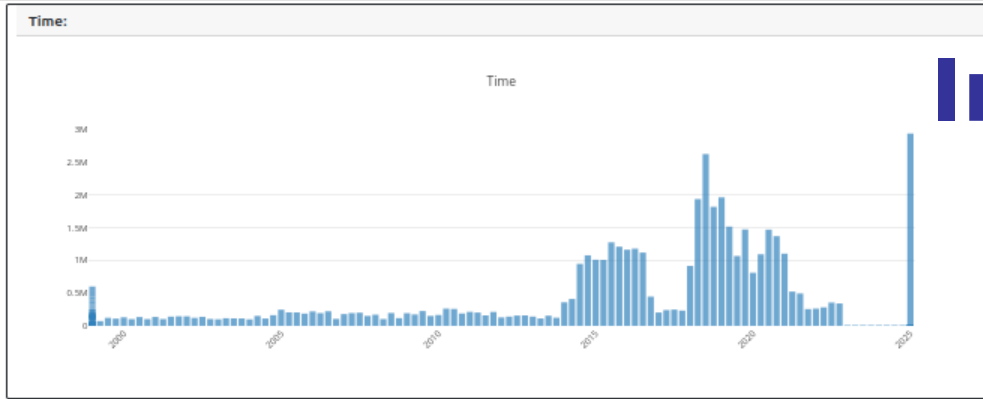
   Range long/lat

   Moc ?

NoSQL facilities/difficulties

- **Easy way to scan meta datas**

- **Facets search make it intuitive**
    - **Search in large heterogeneous datasets**

- **Many difficulties on diplaying hetrogeneous outputs (mixing carot and califlower) !**

- **range filed type are usefull but weighs down the queries.**

- **We have to optimized before clustering.**

Interface

http://voparis-elasticsearch.obspm.fr/dev3

15

Interface

- real progres to do

- Complementary to the calassical VO portal.

- Main question remain how to display results

- Take advantage of the Nasa portal ?