

Using the ObsCore DM to Index VizieR associated data



Gilles Landais – Laurent Michel

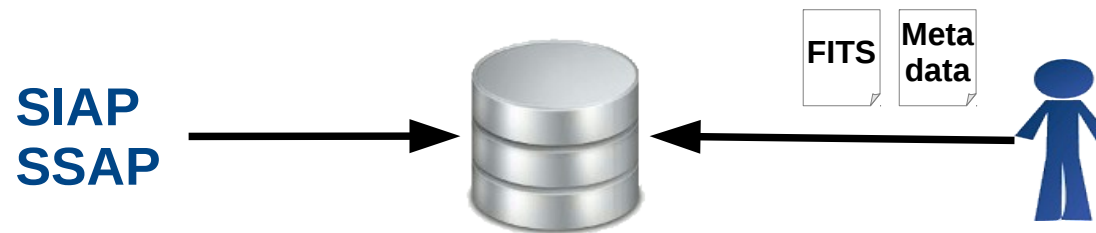


Associated data storage in VizieR: images, spectra, time-series, SED

Current situation

- Stored on file system (FTP, plot & linked in VizieR web pages)
- Different formats: ascii, tabular data, FITS
- Resources not indexed and without global documentation

Goal



Data preservation context

- Scientific data including associated data must be preserved
- Indexation & documentation needed for search, understanding, and data reuse
- Discussion of CDS with AAS about assessment of the usage of **ObsCore** (jan 2014)



Precision concerning the ObsCore choice

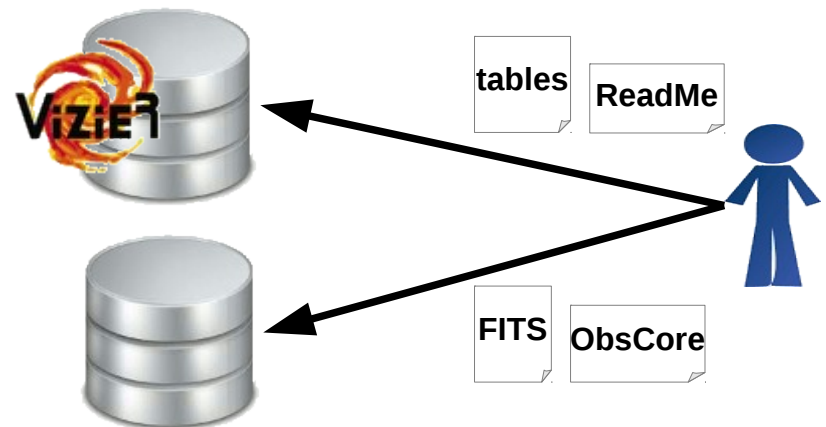
- The ObsCore population is limited by
 - finding ObsCore informations in documents header
 - no standards header (FITS)
- The traps to be avoided
 - incorrect filled informations
 - discouragement owed to a too important number of meta data



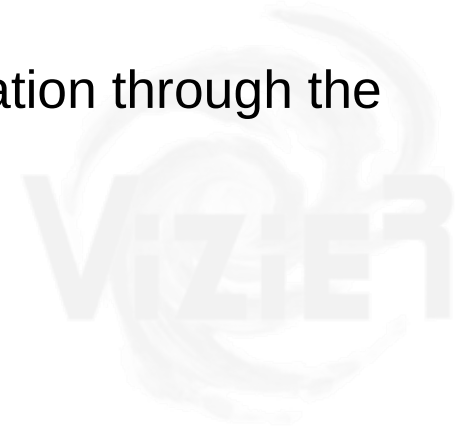
- Limit ObsCore to the **mandatory items**
- Fill ObsCore even if some informations misses
- Every items are encouraged but not needed!

The VizieR options

- Confirm the choice of ObsCore for every catalogs whatever the source : A&A, AAS, MNRAS (Monthly notice), ...
- Demand the author (ex: A&A) contribution to populate the database by describing their data
- Provide tools to help authors : meta data extraction, verification
- Limit (in a first time) to incoming files with the format FITS
(FITS format is not the most important part of associated data in VizieR (~300 catalogues), but CoRoT ~300,000 time-series)
- Provide VO services : SSA, SIA
- Build a new database dedicated for the associated data



Choose **SAADA** as database generator and for the dissemination through the VO



The choice reasons

- SAADA is oriented to the VO (SSA, SIA are available)
- Load test performed successfully (CoRoT: 300,000 times-series)
- A dynamic software maintained and which evolves
- Customizable with adding columns: *VizieR ID, URL link, etc.*
- Adapted to pipelines which enable meta data in input (data in input goes with ant script (XML) in ingestion)
- The proximity with the CDS and a well collaboration with Laurent Michel

- Tables
- Catalogues
- Photometry



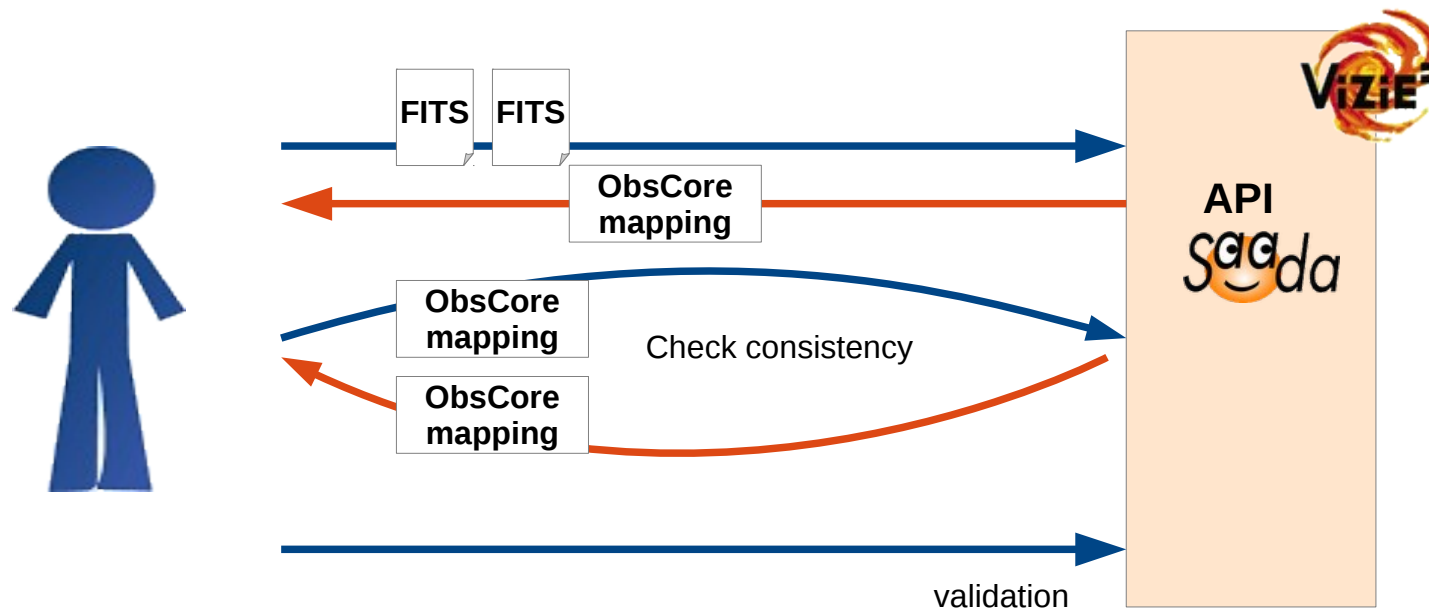
**2 distincts
databases**



- Images
- Spectra, SED
- Time-serie

Helping users to fill Obscore

- Use the SAADA API to infer from FITS headers the values, keywords or computation needed to Obscore
- Propose a pre-filled mapping containing report
- Verify the consistency
- Validate by the author



Encourage authors to fill the meta-data for indexation

Note: The data description won't be required in VizieR, but encouraged with warnings in function of the indexation level.



Level 0: No Positions

Level 1: Positions only
(ObsCore: RA,DEC)

Level 2: Positions + spectral coordinate
(ObsCore: RA, DEC, TARGET_NAME, SPATIAL_RESOLUTION,
EM_MIN, EM_MAX)

....

Actions on SAADA and VizieR progresses in parallel

SAADA

- Use the ObsCore DM as built in the Saada data model
- Smart mapping:
 - Auto-detection of the mapping
 - Preview and report
 - Advanced expression

VizieR

- New VizieR applications for catalogues ingestion and dissemination
- The new VizieR pipeline dedicated for associated data





Integration of ObsCore DM in the SAADA engine.

- Update the meta-data of stored objects with ObsCore
→ requires **important updates** in the deep layer of the SAADA engine
- Update the SAADA admin GUI

Currently, Saada proposes a simple mapping dedicated fo each category

Current version

- Position mapping
- Coord. Syst. mapping
- Position error mapping
- Spectral range mapping
- User meta-data mapping



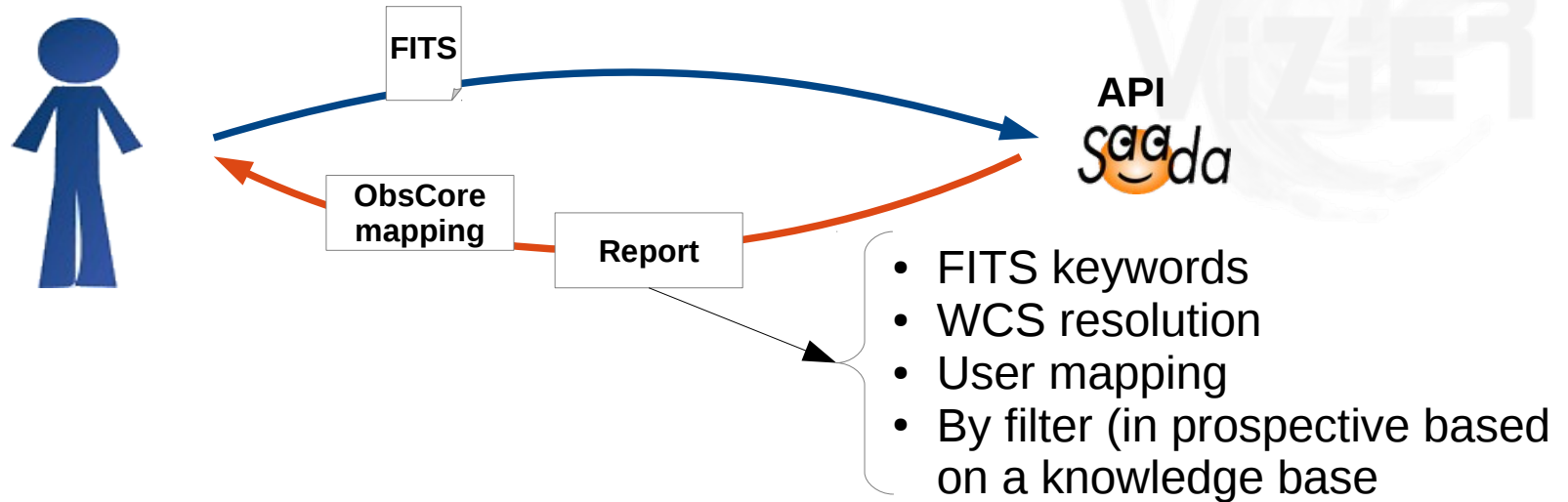
New version (in development)

- ObsCore (mandatory items)
i.e.: 16 items without identifiers
and technicals)
- User meta-data mapping

- Meta-data storage:
 - The ObsCore mapping are stored in a ObsCore Table
 - The FITS keywords mapping stored in SAADA tables

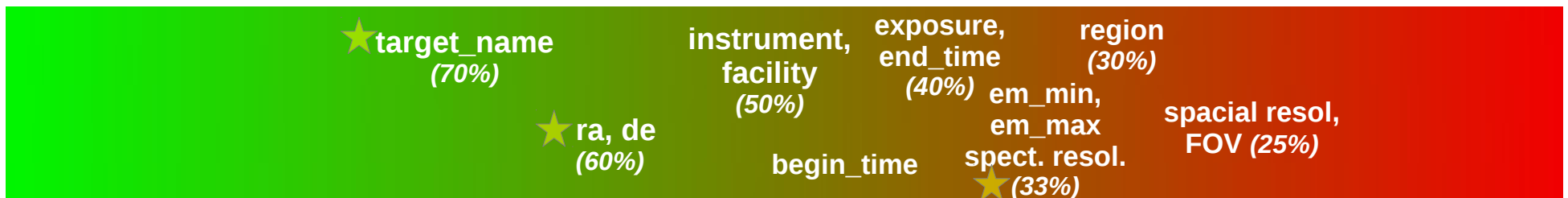
SAADA Mapping updates

- Improve the pipeline capabilities with complementary methods to make the mapping



- Advanced expression are available for the User mapping:
Example: $T_EXPTIME = \text{MJD}(\text{OBS_END}) - \text{MJD}(\text{OBS_START})$
 (ObsCore) (Fits keywords)

- The SAADA mapping score: → limited by the lack or not standardized unit descriptions

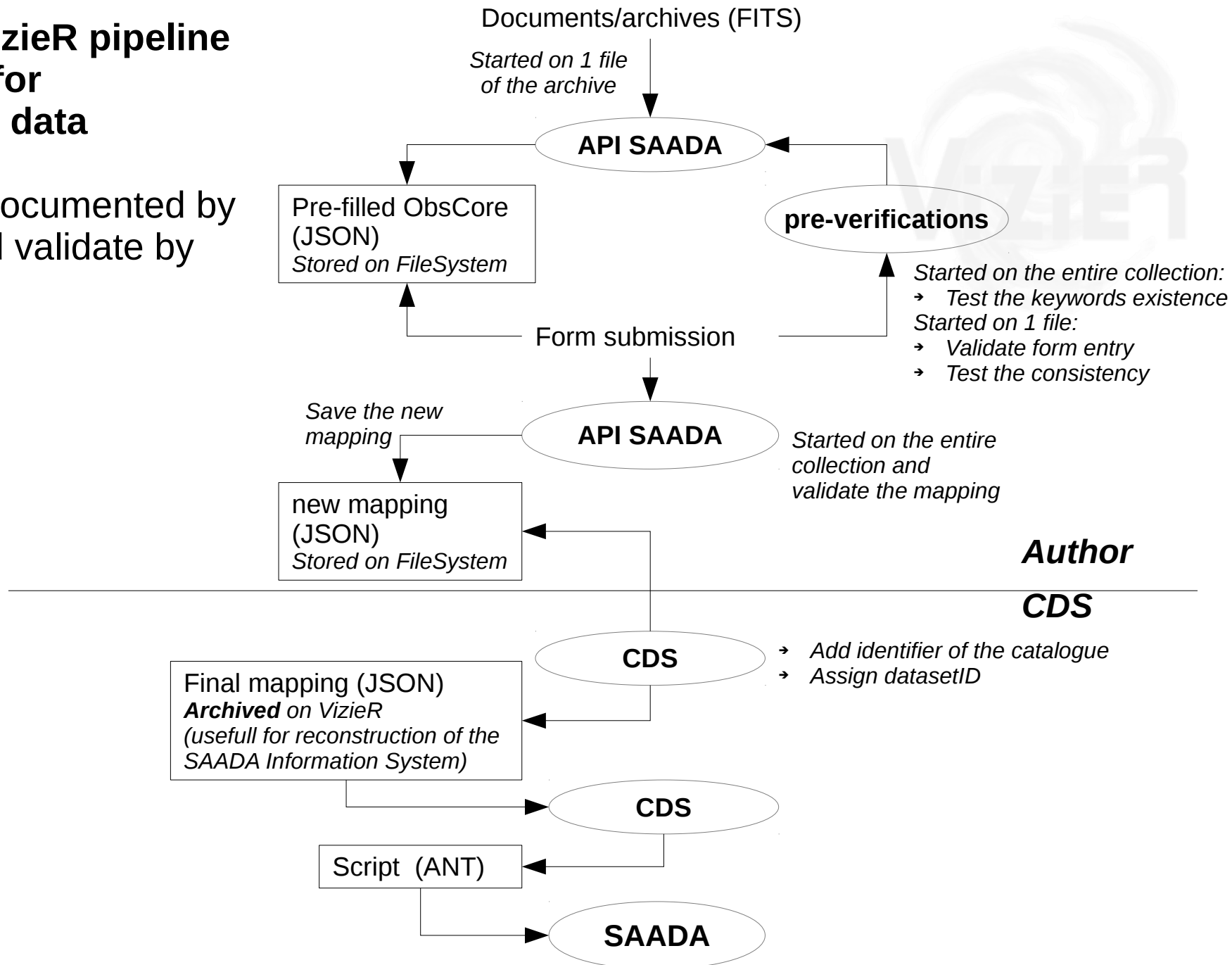


New VizieR applications for catalogues ingestion

- Need adapted tools for documentalists/authors to ingest the associated data
 - Need semiautomatic pipeline to find meta in FITS headers
 - use the SAADA API to propose a mapping
 - Need documentation (in particular for authors) to describe data
- The new web interface for catalogues ingestion
 - Upload tabular data
 - ReadMe skeleton generator
 - Upload images/spectra with ObsCore
- The interfaces capabilities
 - Ingestion step-by-step, with help and with asynchronous jobs (UWS) which enables to continue the ingestion later
 - Check the meta-data consistency:
 - Syntax, existence of keywords, verification of the formula
 - Avoid suspicious values : Size of FoV (in particular for spectra), elapsed time

The new VizieR pipeline dedicated for associated data

A pipeline documented by authors and validate by the CDS



Identify the Objects

- **obs_collection:** the catalogue name
example: SDSS, CoRoT, J/A+A/378/861
- **obs_id:** the filename
example: 10144aa.fit
- **obs_publisher_did:**
example: the spectrum 10144aa.fit from the catalogue J/A+A/378/861
ivo://CDS/J/A+A/378/861/10144aa.fits



The ObsCore Data model filled by authors

Portal Simbad **VizieR** Aladin X-Match Other Help

VizieR upload catalogue

You are logged as landais [report](#)

Upload tabular data **Fill the ReadMe** Upload Spectra (optional) Upload Images (optional)

You can upload associated data as **spectrum/time-series** or **images** in VizieR.

FITS is the most adapted format today. For these documents, a dedicated database indexes Spectra and images and provides the outside.

These documents need descriptions: the VizieR engine will extract some metadata in their that you can update or change.

In this web page you will upload your **Spectra/time-serie** in **FITS** format.

Upload your spectra

You have some spectra

Only FITS format are accepted! Please, upload documents in an other format later.

You can upload your documents one by one by describing them **independently**
OR if you have documents with **similar header** you can upload a **collection** (an archive in tar, zip format) and put a comment.

Add new document(s) No file selected. which is ???

[Archive spectrum/J apj_703_894_collection0/.* \(111 files\)](#)

[Remove all documents](#)

[Fill the ReadMe](#) [Upload Images \(optional\)](#)

[Archive spectrum/J apj_703_894_collection0/.* \(111 files\)](#)

Position

Target name unit:

Right ascencion unit:

Declination unit:

Field of view unit:

Region unit:

Spatial resolution unit:

Time

Begin time unit:

End time unit:

Exporure time unit:

Time resolution unit:

Spectral

Spectral min unit:

Spectral max unit:

Spectral resolution unit:

Others

Polarization unit:

Facility name unit:

Intrument name unit: