OIDB data format requirements

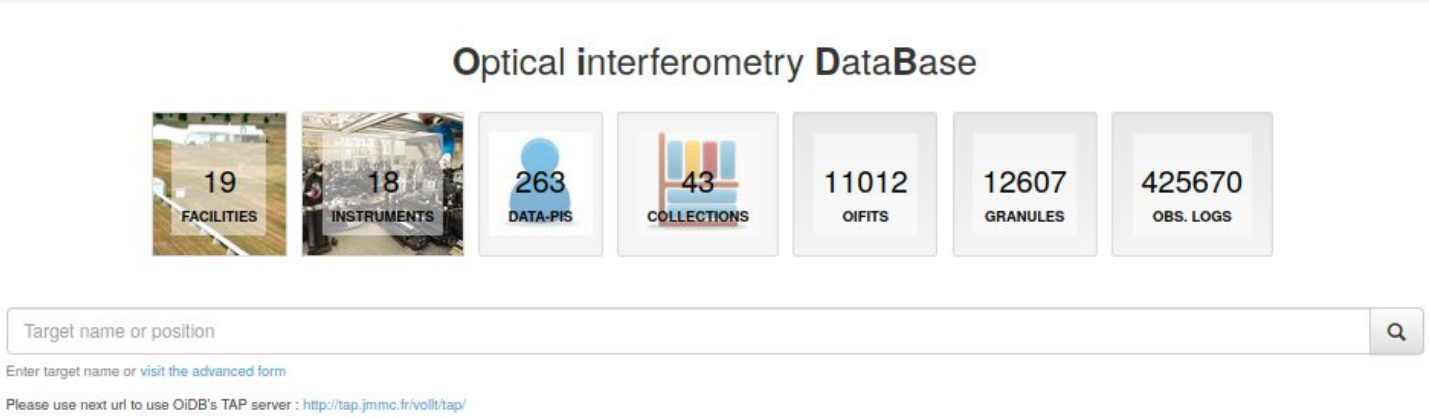
G. Mella, J-P Berger, L. Bourgès, G.Duvert

VLTI reduced data workshop

2022-06-21

Summary

- Short introduction OiDB
- Data content
  - OiFits
  - Metadata
- Requirements
  - Data injestion
  - Validation/curation
  - Protection



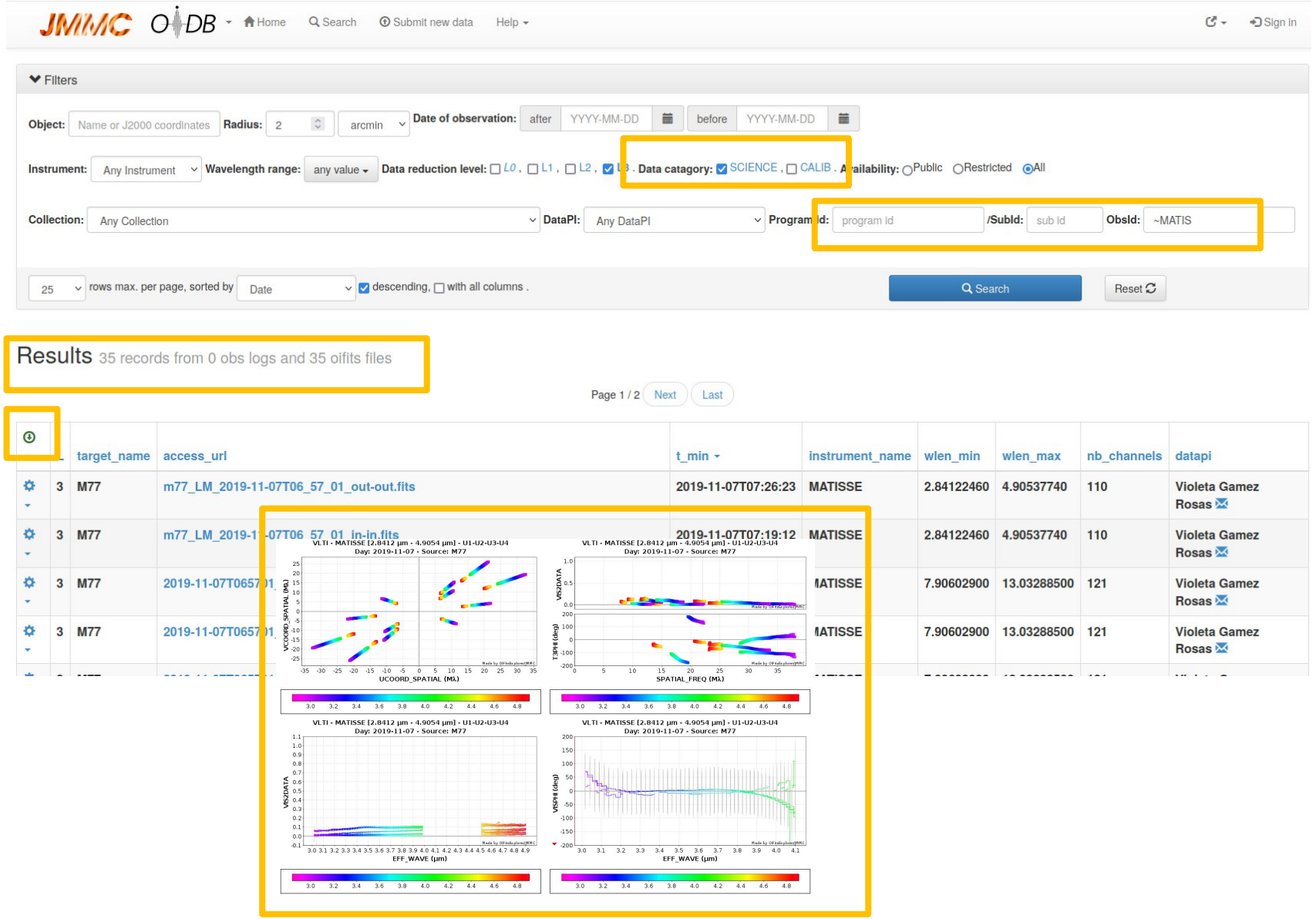
Short introduction

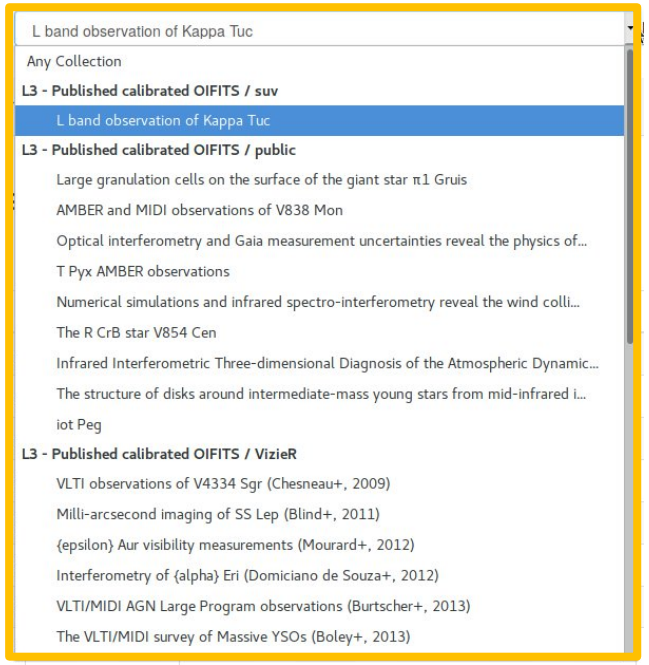
https://oidb.jmmc.fr

- Central access to optical interferometry data
- Web & programmatic interface
  (IVOA Obscore/TAP)
- Multiple feeding
  - harvesting
  - user uploads
  - Survey synchronization (SPICA@CHARA)
- Ready for public or protected data
- Worldwide facilities
  (& space telescope simulated data)

- Data products:
  - Obs records (L0)
  - OIFits files
    - Uncalibrated L1
    - Calibrated L2
    - Calibrated and published L3
- Data categories :
  - SCIENCE, CALIBRATOR
- Additional data (datalinks):
  - Quick views
  - Quality reports
  - Can be anything behind an URL : images,etc,...

OiDB screenshots





OiFits are split by granules, grouped by collections

OiDB data **format** requirements

- OIFits first requirement level
  - **Every file must conform to OIFITS standard V1 or V2**
    - Validation codes rely on JMMC oitools. See also http://oival.jmmc.fr/about.html
  - OiDB validates user files during upload
    - Non oifits files are rejected
    - Published data (L3) are ingested without message
    - Users should try to fix reported errors for L1/L2
  - OiDB trusts metadata submissions (external deposit)

OiDB data **content** requirements => OiDB metadata requirements

- Metadata requirements
  - Everything is optional, but the more information is provided the better
  - Data ingestion pre-fills some user fields looking for some standard keywords (progid, category...)
  - OiDB helps users to register targets with Simbad identifier
    - -> some target names are not resolved by simbad
- Metadata statistics highlight some faulty data
    - E.g. Obs date in the future or before 1900...
    - Target names and coordinates are varying (in OIFITS) so correct cross-identification is difficult (ex : Sirius ...)

OiDB curation

- Standardization is very important :
  - Helped us to code some rules to help the users to go in the right direction
  - save time
- Some rules/logic may be hard to address using simple UI
  - Our approach consists in fixing specific issues by hand and implementing automatization when the need is well defined.
  - We developped an API to manage our catalogs and use it on OiDB to run in a fast and flexible way to fix database content.
    - We used to run manual sql commands but now prefere this declarative approach.

OiDB data **protection** requirements

- Public data are easier to use, but some data must be protected
- Embargo periods follow archive data policy :
    - OiDB uses their release_date as references
  - protected/secured OIFits + datalink may be hosted by OiDB (and remote)
  - JMMC accounts (email) can be linked to one or more datapi name so we can generate (every day) a protection file on remote data storage (PIONIER_L2, SPICA)
- (SUV) expertise center & individual protected data deposits get an additional delay for reduced data release_date

Conclusion

- OIFitsExplorer tool can use the same credentials to retrieve and merge all of your protected data in a glance (+ public ones without any cost)
- Convinced by the interoperability between tools :
  - do more than the sum of what can be done by each individual tool.
- Good data & metadata help to provide some efficient gateways accross data repositories.
  - e.g. : just giving the proper id, your can visualize and explore all data of your PIONIER program

Questions ?