ACTRIS Central Facility: overview of key services

IRWG meeting Spa June 2023

BIRA-IASB
with special thanks to our IT team (Yves Geunes and Tim Ooms)

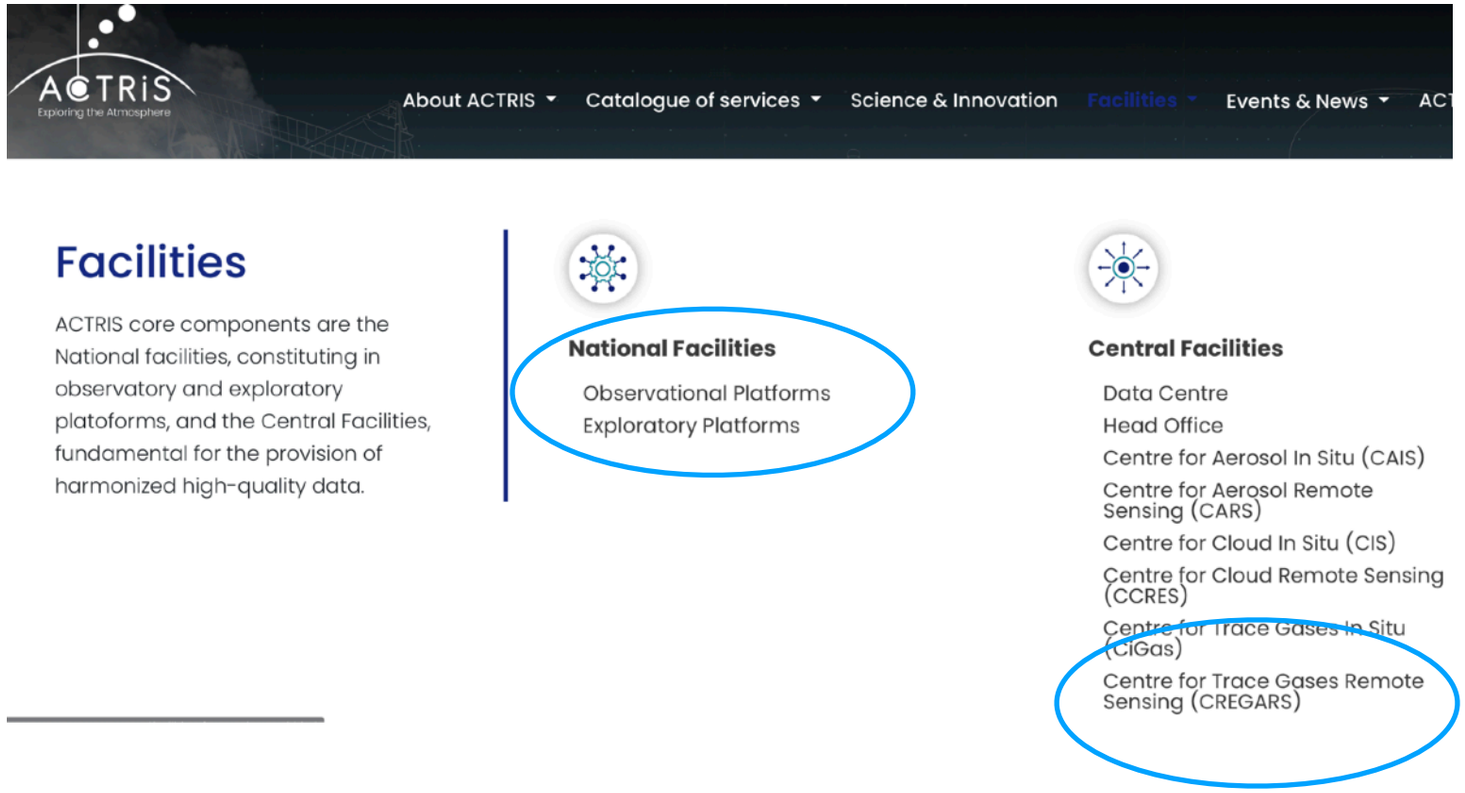
UBremen and ULg

ACTRIS Central Facility: overview of key services

- ACTRIS is a EU “research infrastructure” (as ICOS); it is a legal entity (ERIC) since 25/4/2023 → e.g., it can apply in R&D calls as partner

- Several EU IRWG PIs joined ACTRIS to have a long term stable funding commitment from their national governments and to benefit from ACTRIS services

- ACTRIS has “national facilities (NF)” and “central facilities (CF)”



actris.eu

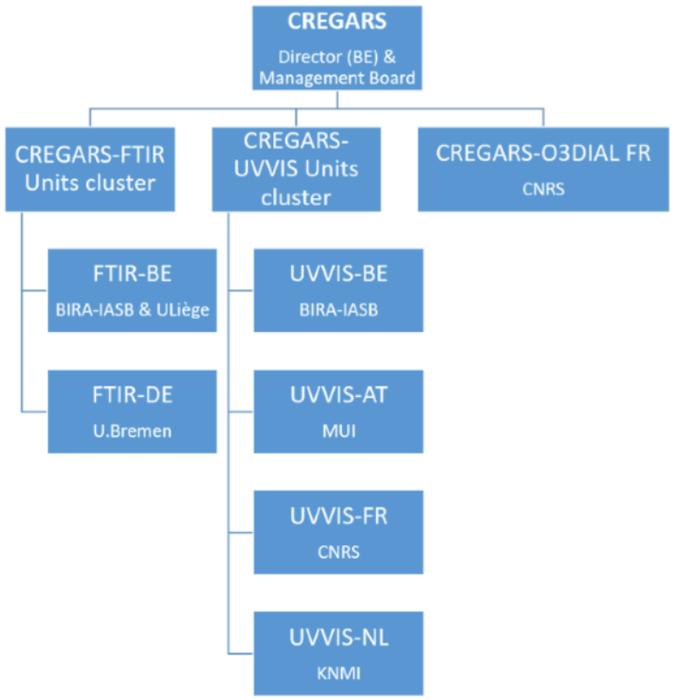
CREGARS: who?

- The mission of the **Centre for Reactive Trace Gases Remote Sensing** (CREGARS) is to facilitate the generation of highest-standard reference data of several key trace gases using ground-based remote sensing techniques

Our Units & Teams

The CREGARS topical TC is divided into 3 main clusters (FTIR, UV-VIS and O3 DIAL), based on the employed measurement technique, but within each cluster the work is shared between different units.

The hierarchical structure is shown in the below organigram:



- Royal Belgian Institute for Space Aeronomy (BIRA-IASB)
- University of Bremen (U.Bremen)
- University of Liège (ULiège)
- Centre National de Recherche Scientifique (CNRS)
- Royal Netherlands Meteorological Institute (KNMI)
- Medical University Innsbruck (MUI)

FTIR Central Facility: key features

- Central processing @ BIRA-IASB
- Targets: O3, HCHO, C2H6, NO2 and NH3
- Current IR related NF: Bremen, NyAlesund, Maido, Garmisch, Jungfrau?, Sodankyla?, Potenza???, …
  -> eg not all Belgian instruments are part of a NF
- ACTRIS Data hosting @ AERIS: GRES https://gres.aeris-data.fr/
- Provide measurement data according to FAIR principles (https://www.go-fair.org/fair-principles/):
  - Findable: rich metadata in an ACTRIS central data catalogue
  - Accessible: open protocols are used to retrieve data
  - Interoperable: hdf/netCDF, GEOMS/CF vocabulary
  - Reusable: data license (CCBY4: share and adapt data)

  (relates to question of Manu and Jim on storage of L0 data on NDACC)

FTIR Central Facility: key features

- Processing is traceable:
  - Processor: SFIT4 versions have DOI (managed by AERIS): DOI links to NCAR GitHub and IRWG pages (spectroscopy db, WACCM prior db)
  - Retrieval strategies (station granularity) have DOI (AERIS): ctl templates + specification of ancillary data used (eg spectroscopy, WACCM, NCEP Reanalysis)
  - Retrieved data (annual granularity and time series) have DOI (AERIS)
  - Instruments have a PID (AERIS or BIRA-IASB?) + logging website (BIRA-IASB)

- Develop harmonized retrieval template per target (fine-tuning per instrument in collaboration with PI): BIRA-IASB+ULg

- Update cycle for retrieval strategies

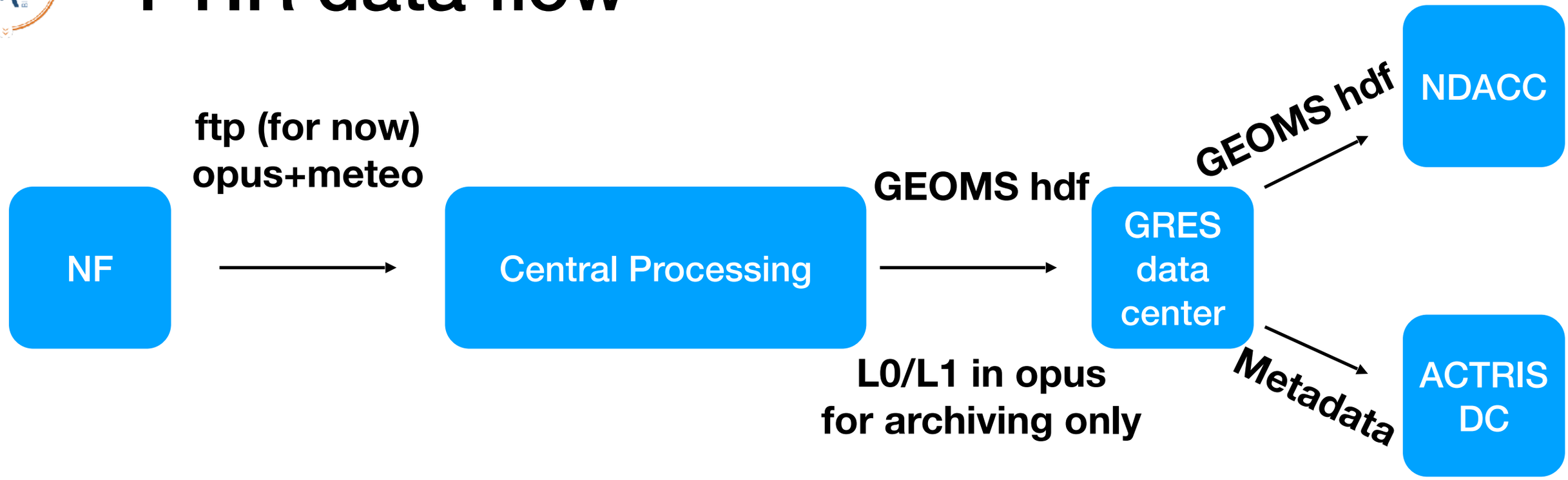
- Maintenance and development of SFIT4 (UBremen, BIRA-IASB)

- Cell calibration service and organization of training workshops: UBremen

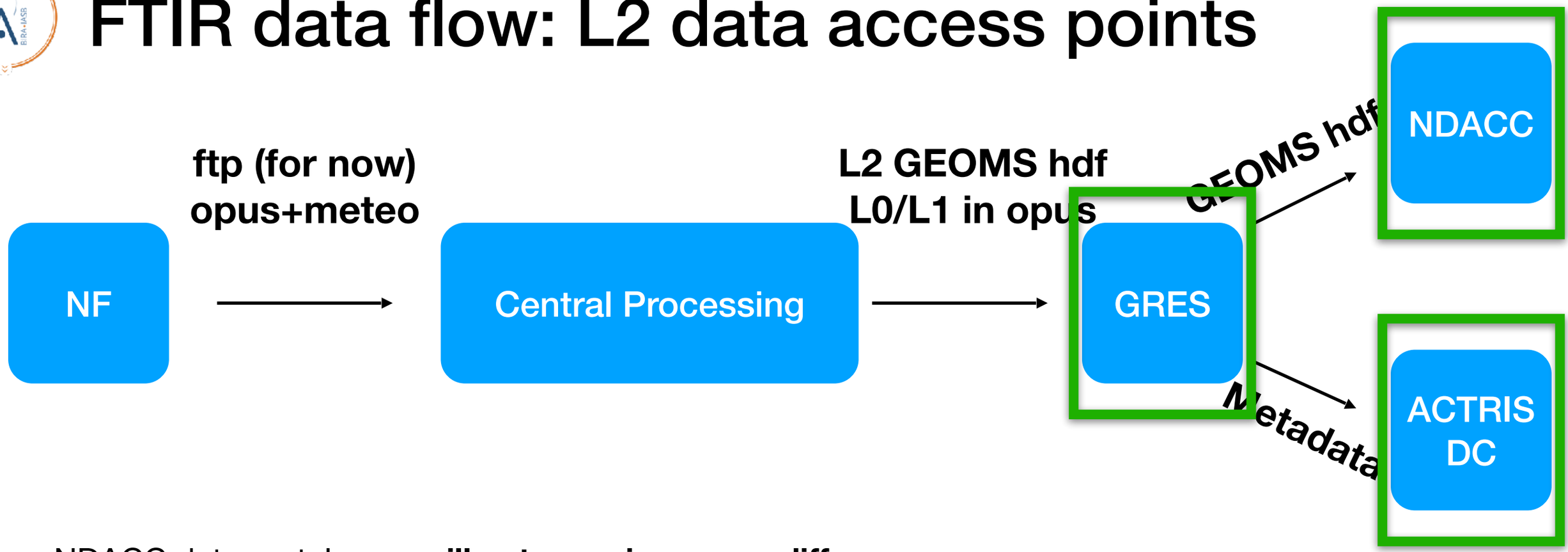
- Retrieval strategy development for low resolution instruments (Invenio/vertex)

- Strategic stock of spare parts: being investigated

FTIR data flow



FTIR data flow: L2 data access points



- NDACC data portal user: **will not experience any difference**
  - GEOMS files from ACTRIS will have a FILE_DOI and “ACTRIS” in FILE_ACCESS and attributes will have links to the relevant DOI’s (retrieval strategy, instrument, processor, time series)

- ACTRIS DC user:
  - will be able to search for ACTRIS data (web and api) and retrieve the data from GRES
  - ACTRIS DC catalogue will use “ACTRIS vocabulary” (next slide)
  - will be able to search for “ACTRIS associated” data and download data from … not decided

- ACTRIS GRES will provide access to data in GEOMS hdf and an alternative version GEOMS + CF in netCDF (next slides)

https://vocabulary.actris.nilu.no/skosmos/actris_vocab

ACTRIS Vocabulary

Content language English Search

Alphabetical Hierarchy

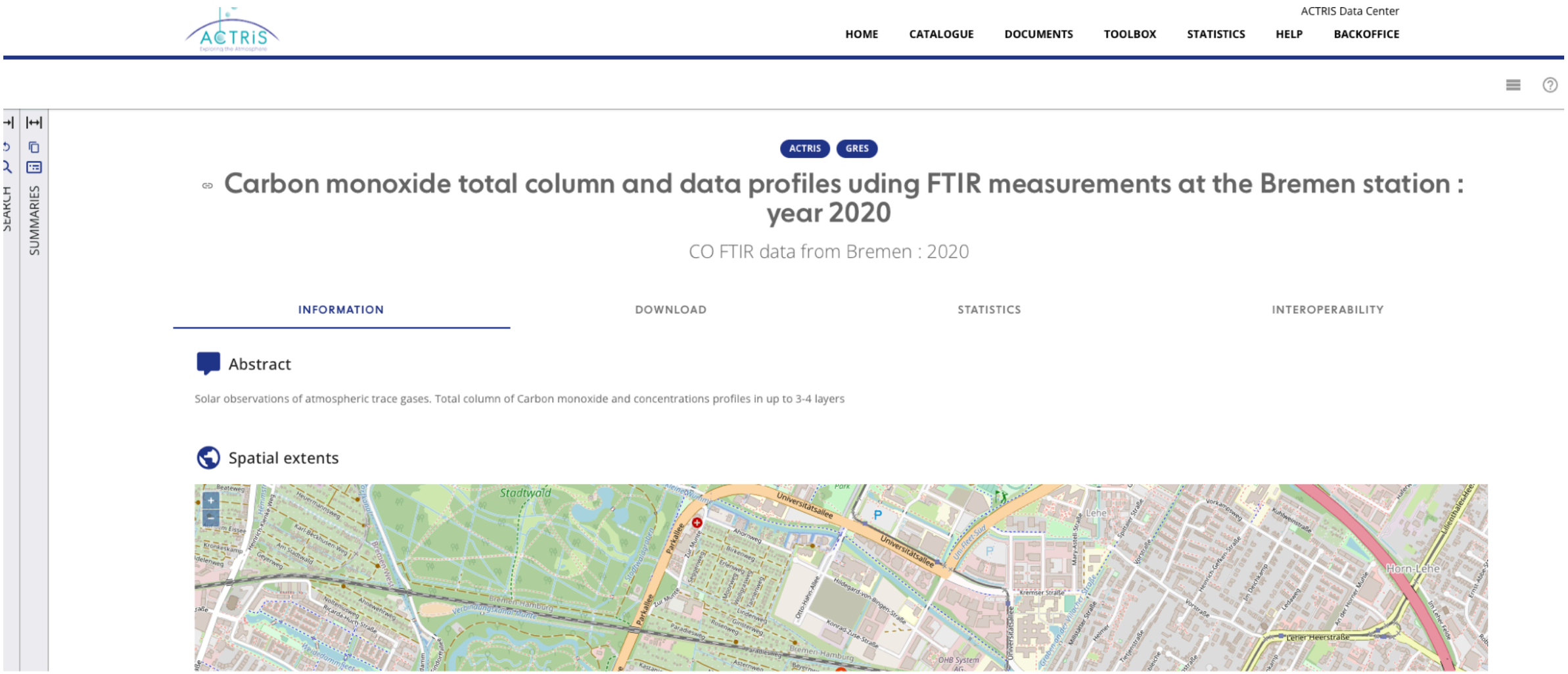
- carbon monoxide number concentration
- caronaldehyde amount fraction
- caronaldehyde mass concentration
- caronaldehyde number concentration
- cyclohexane amount fraction
- cyclohexane mass concentration
- cyclohexane number concentration
- cyclohexene amount fraction
- cyclohexene mass concentration
- cyclohexene number concentration
- di(2-ethylhexyl) phthalate amount fraction
- di(2-ethylhexyl) phthalate mass concentration
- di(2-ethylhexyl) phthalate number concentration
- dichlorobenzene amount fraction
- dichlorobenzene mass concentration
- dichlorobenzene number concentration
- diisopropyl ether amount fraction
- diisopropyl ether mass concentration
- diisopropyl ether number concentration
- dimethyl sulfide amount fraction
- dimethyl sulfide mass concentration
- dimethyl sulfide number concentration
- dimethyl sulfoxide amount fraction
- dimethyl sulfoxide mass concentration
- dimethyl sulfoxide number concentration
- dimethyl-cyclohexane amount fraction
- dimethyl-cyclohexane mass concentration
- dimethyl-cyclohexane number concentration
- dinitrogen pentoxide amount fraction
- dinitrogen pentoxide mass concentration
- dinitrogen pentoxide number concentration
- **ethane amount fraction**
- ethane mass concentration

variable group > trace gas variables > ethane amount fraction

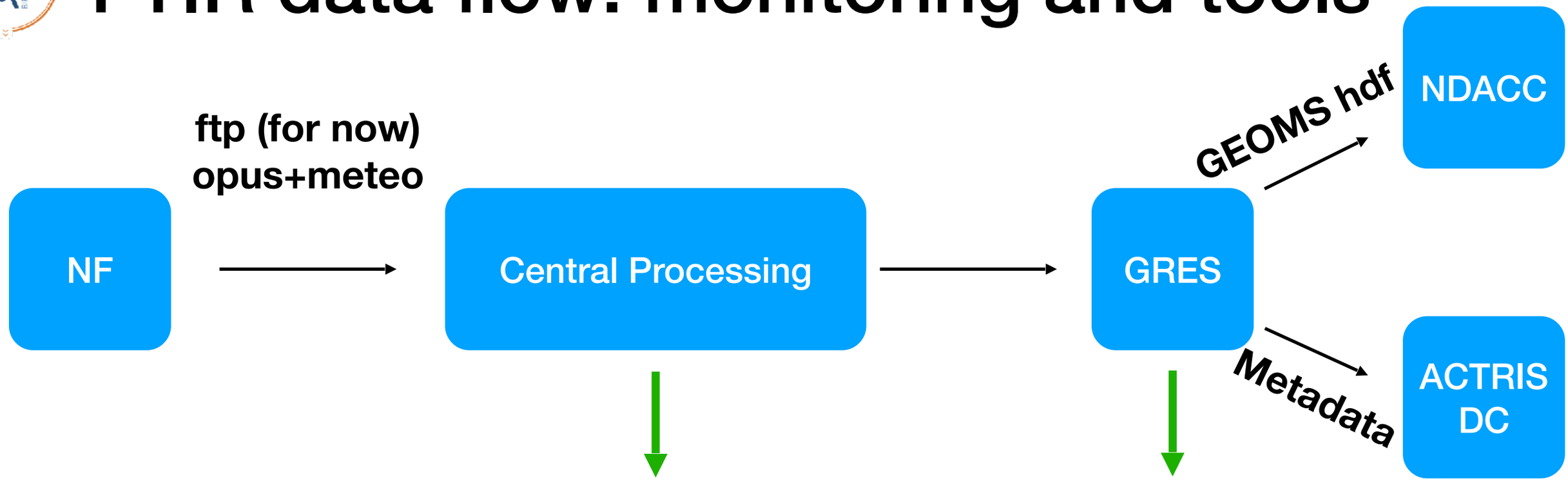
PREFERRED TERM **ethane amount fraction**

DEFINITION	Amount fraction is used in the construction mole_fraction_of_X_in_Y, where X is a material constituent of Y. The chemical formula for ethane is C2H6. Ethane is a member of the group of hydrocarbons known as alkanes. The IUPAC name for ethane is ethane.
BROADER CONCEPT	trace gas variables
CREATOR	https://orcid.org/0000-0003-2972-2851
OBSERVATION MATRIX	gas phase
OBJECT OF INTEREST	ethane
OBSERVED PROPERTY	amount fraction
URI	https://vocabulary.actris.nilu.no/actris_vocab/ethaneamountfraction
Download this concept:	RDF/XML TURTLE JSON-LD

https://gres.aeris-data.fr/catalogue/



FTIR data flow: monitoring and tools



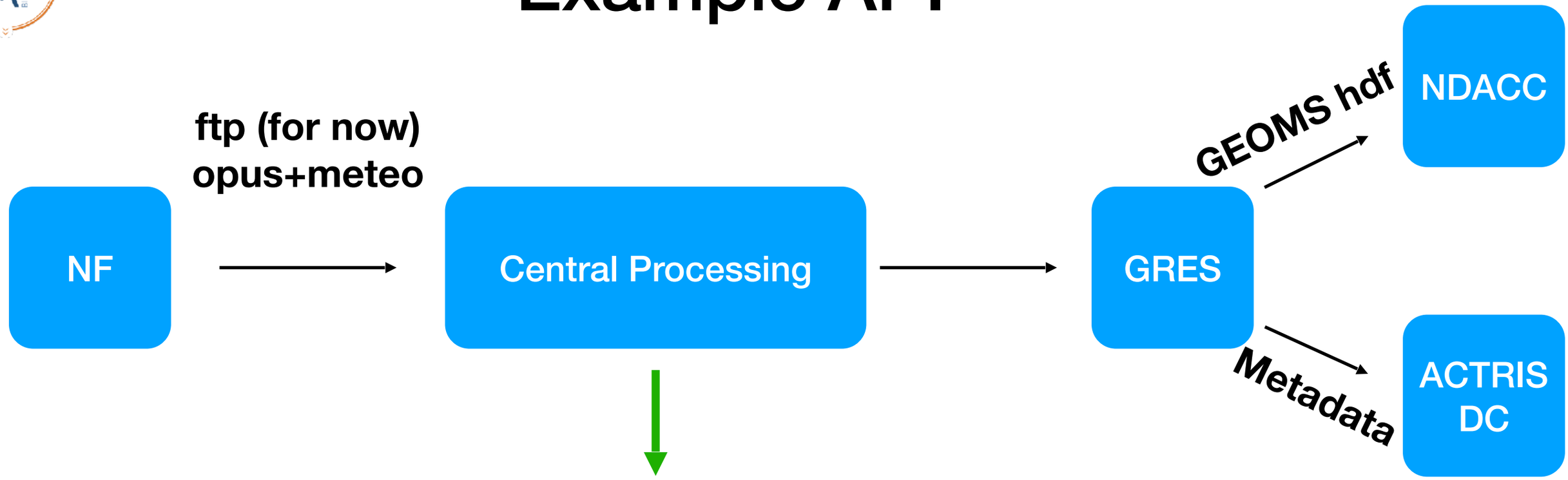
For registered PIs (* = implementation done)

- L0/L1/L2 monitoring tool*
- HDF2netCDF API*
- QAQC API*
- OPeNDAP access to retrieval details: GEOMS like file with all retrieval gases+other quality indicators (see next slide)*
- OPeNDAP access to ZPT prior data @ site: netCDF file
- GEOMS creation API for target + co-retrieved gases
- Instrument database + interface for logging events

GRES

- Perform QAQC on files
- Extract metadata for ACTRIS DC (with vocab)
- Forward GEOMS hdf to NDACC
- Provide public access to L2 data (GEOMS HDF or GEOMS+netCDF-CF)
- DOI management
- Long term archiving of L0/L1/L2 data

Example API



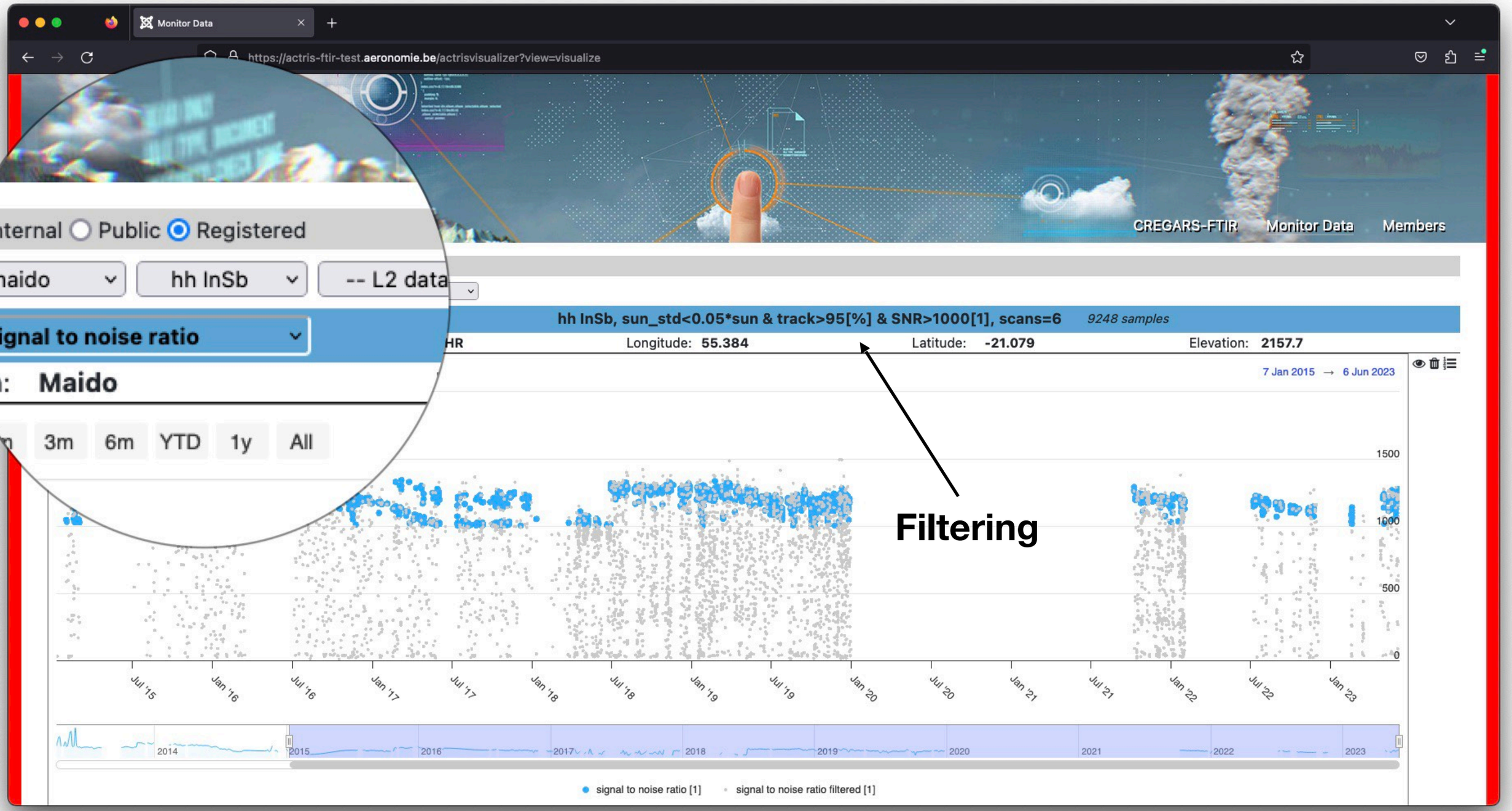
- curl -s --form testfile=@"groundbased_ftir.o3_iup001_bremen_20220106t095100z_20220106t124748z_007.hdf" -H "Accept: text/html" https://api.aeronomie.be/geoms_qaqc > test.html

- or use … -H "Accept: application/json" https://api.aeronomie.be/geoms_qaqc


```
    {"StartTime": "20220106T095100Z", "EndTime": "20220106T124748Z", "Md5Check":
"44056ff3f995985a0628261ff817a43d", "SubmissionTime": "", "SourceFile": "/tmp/tmpv7kl8yed",
"FileGenerationTime": "20220121T155024Z", "GeomsFilename":
"groundbased_ftir.o3_iup001_bremen_20220106t095100z_20220106t124748z_007.hdf", "Station": "bremen",
"Instrument": "FTIR", "Target": "O3", "PIEmail": "notholt@iup.physik.uni-bremen.de", "DataSource":
"FTIR.O3_IUP001", "ErrorCode": 0, "QCversion": "5.37.20", "ContactEmail": "", "QualityCheck": 0, "RDCheck":
-999}
```


- curl -s --form testfile=@"groundbased_ftir.o3_iup001_bremen_20220106t095100z_20220106t124748z_007.hdf" -H "Accept: application/x-netcdf4" https://api.aeronomie.be/geoms2netcdf -O -J

Central processor monitoring tool

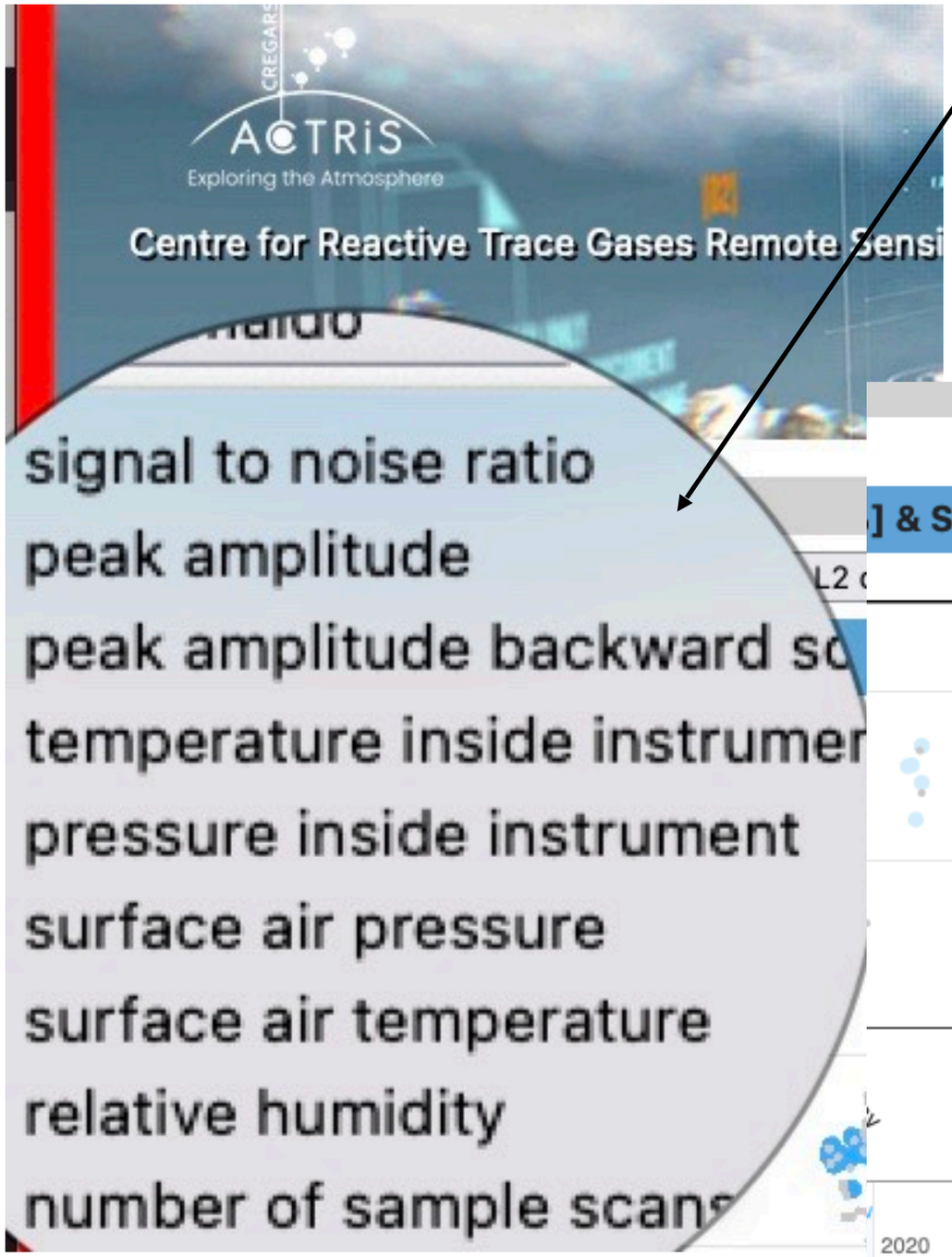


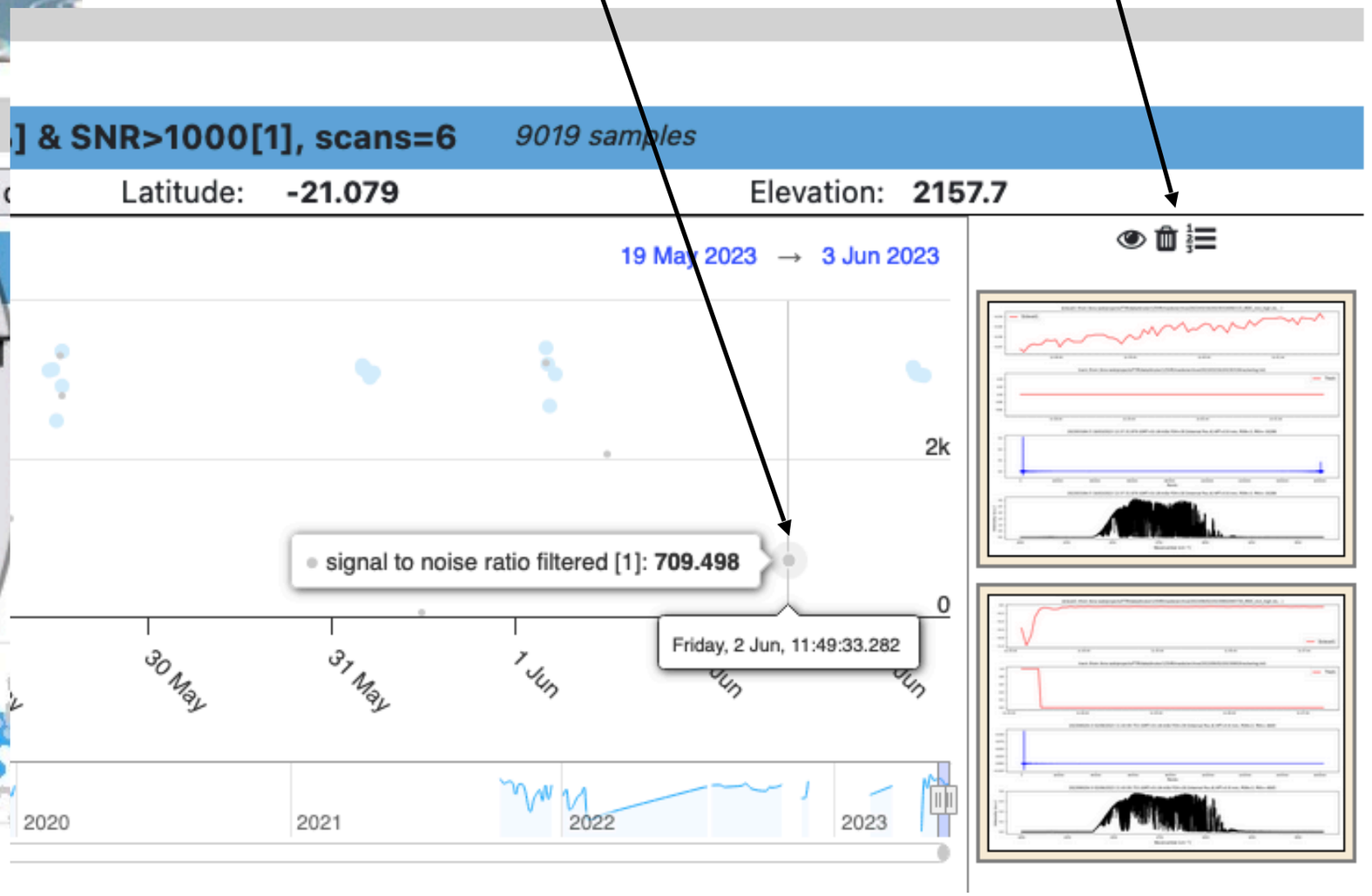
Central processor monitoring tool

Spectral Metadata

Interactivity: shows the IFG/SPC

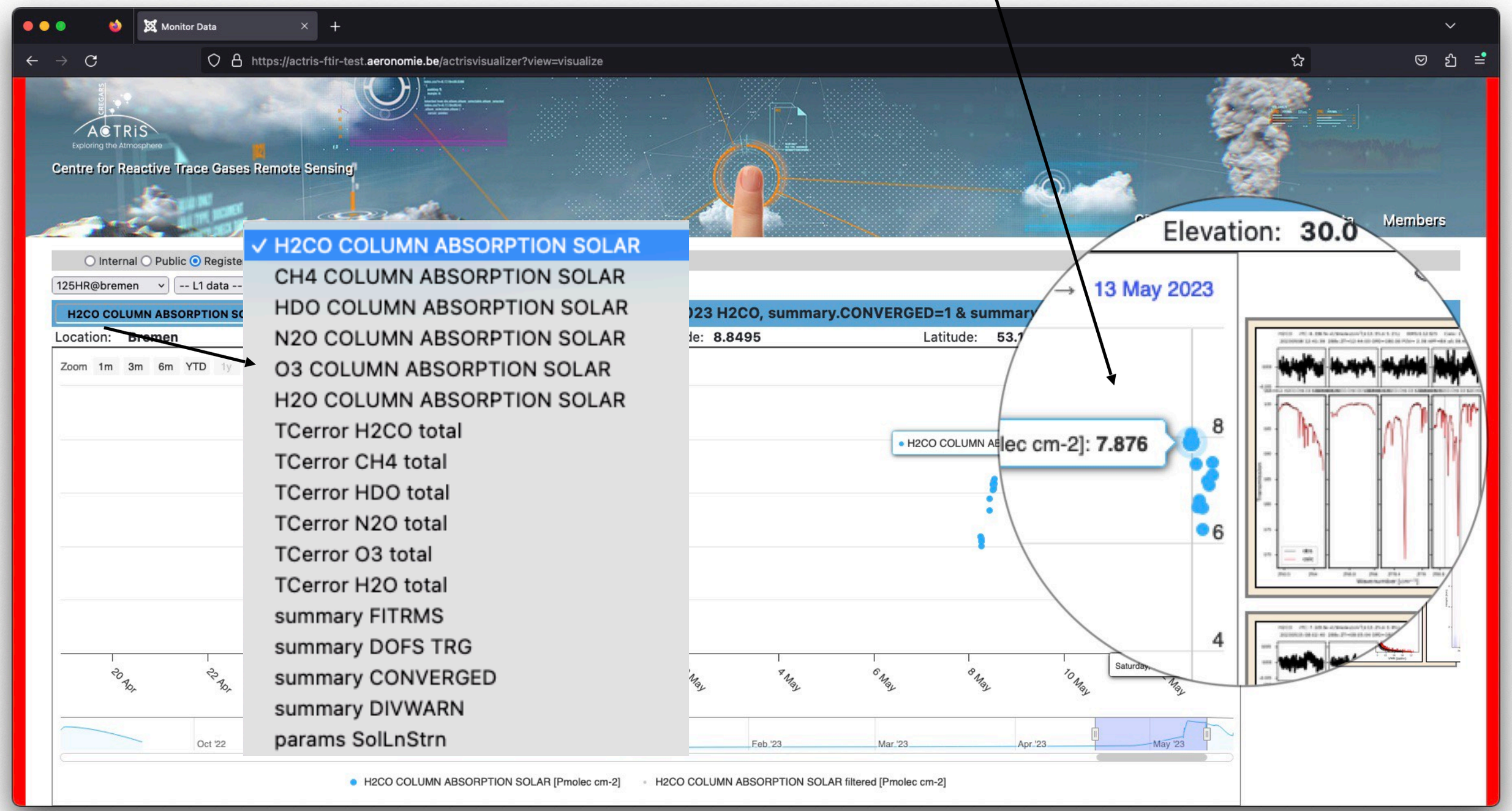
Prints UID … for tagging





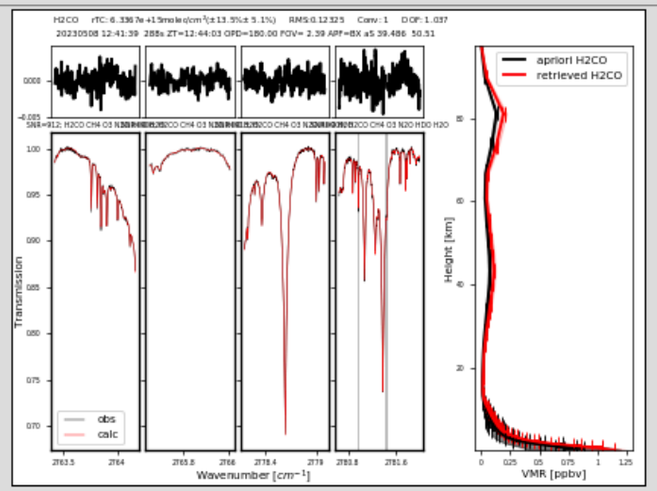
Central processor monitoring tool

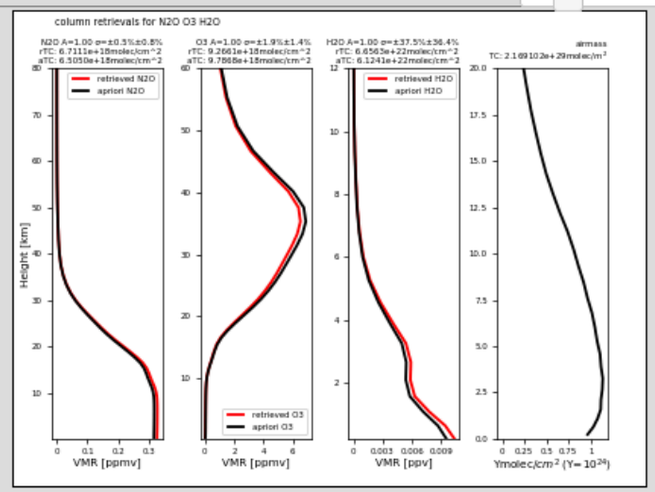
Interactivity for l2 data gives retrieval details

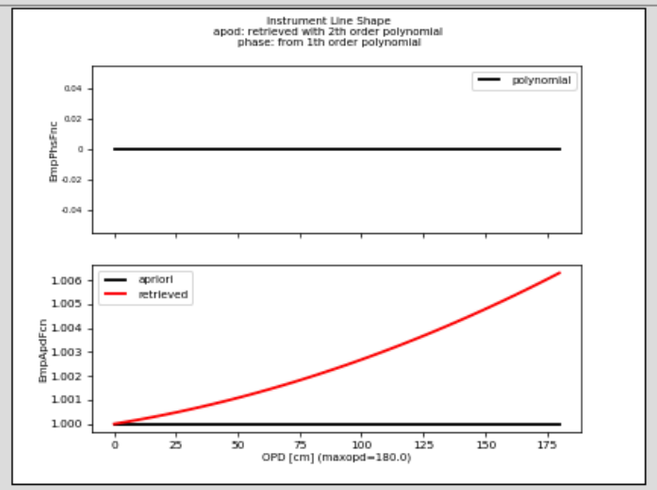


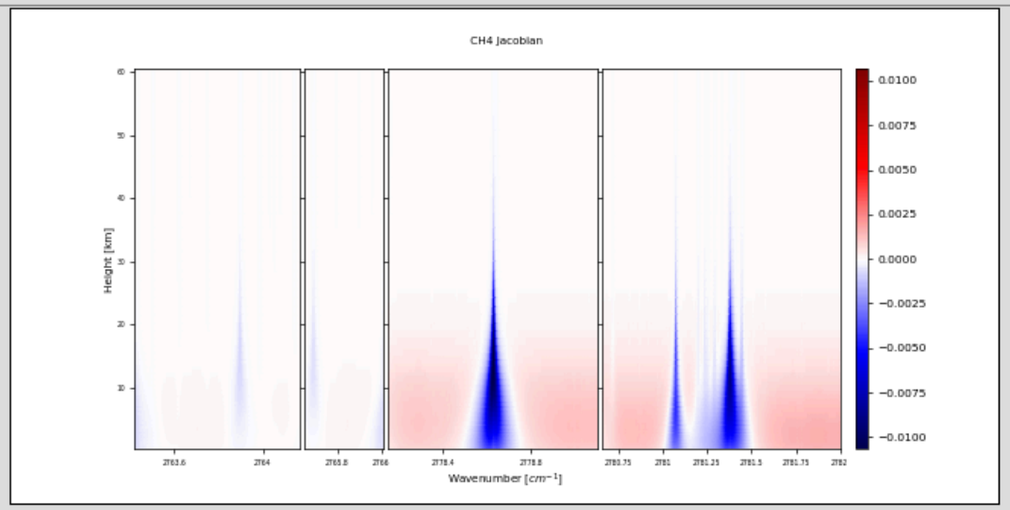
Central processor monitoring tool

Interactivity for l2 data gives retrieval details:
Profiles, column scaling, ILS, Jacobian, …









br20230508s0ehsc_op.017.0

Central processor monitoring tool

Structure of retrieval data netCDF file: GEOMS for all retrieved molecules + extra

H2CO.nc4

- ALTITUDE
- ALTITUDE.BOUNDARIES
- ALTITUDE.INSTRUMENT
- ANGLE.SOLAR_AZIMUTH
- ANGLE.SOLAR_ZENITH.ASTRONOMICAL

CH4

- CH4.COLUMN.PARTIAL_ABSORPTION.SOLAR
- CH4.COLUMN.PARTIAL_APRIORI
- CH4.COLUMN_ABSORPTION.SOLAR
- CH4.COLUMN_ABSORPTION.SOLAR_AVK
- CH4.COLUMN_ABSORPTION.SOLAR_UNCERTAINTY.RANDOM.STANDA
- CH4.COLUMN_ABSORPTION.SOLAR_UNCERTAINTY.SYSTEMATIC.STAI
- CH4.COLUMN_APRIORI
- CH4.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR
- CH4.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR_AVK
- CH4.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR_UNCERTAINT
- CH4.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR_UNCERTAINT
- CH4.MIXING.RATIO.VOLUME.DRY_APRIORI
- CH4.MIXING.RATIO.VOLUME.DRY_APRIORI.SOURCE
- DATETIME
- DRY.AIR.COLUMN.PARTIAL_INDEPENDENT
- DRY.AIR.COLUMN.PARTIAL_INDEPENDENT_SOURCE

H2CO

- H2CO.COLUMN.PARTIAL_ABSORPTION.SOLAR
- H2CO.COLUMN.PARTIAL_APRIORI
- H2CO.COLUMN_ABSORPTION.SOLAR
- H2CO.COLUMN_ABSORPTION.SOLAR_AVK
- H2CO.COLUMN_ABSORPTION.SOLAR_UNCERTAINTY.RANDOM.STAND
- H2CO.COLUMN_ABSORPTION.SOLAR_UNCERTAINTY.SYSTEMATIC.ST
- H2CO.COLUMN_APRIORI
- H2CO.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR
- H2CO.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR_AVK
- H2CO.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR_UNCERTAIN
- H2CO.MIXING.RATIO.VOLUME.DRY_ABSORPTION.SOLAR_UNCERTAIN

- params.BckGrdSlp_1
- params.BckGrdSlp_2
- params.BckGrdSlp_3
- params.BckGrdSlp_4
- params.EmpApdFcn_1
- params.EmpApdFcn_2
- params.IWNumShft_1
- params.IWNumShft_2
- params.IWNumShft_3

FTIR Central Facility

- Advantages:
  - harmonised retrieval/increased traceability
  - increased IRWG response to updates in spectroscopy/prior/processor/ GEOMS template changes/…
  - implement updates of retrieval strategies more quickly (from decade to …)
  - RD service
  - decrease burden on PIs for routine tasks
  - shared detailed quality monitoring tools
  - traceability of the instruments: PID + logging
  - …

- Disadvantages: ACTRIS supports “access to its services”: this means that the central processing functionality could be used by others…. but …

  access to central processing may require a fee … not clear

- L0/L1 are supposed to be public when delivered FAIR