



***NenuFAR and the
Centre de Données de
Nançay***

***B. Cecconi, J. Girard
E. Thétas***

Nançay Data Centre

Quick Facts

- Started in 2014
- Offers computing and storage facility for Nançay projects
- Data distribution capabilities
- Current storage capacity: 500 TB (on 3 servers)
- 4 multi-core processing servers (nancep1 to 4)

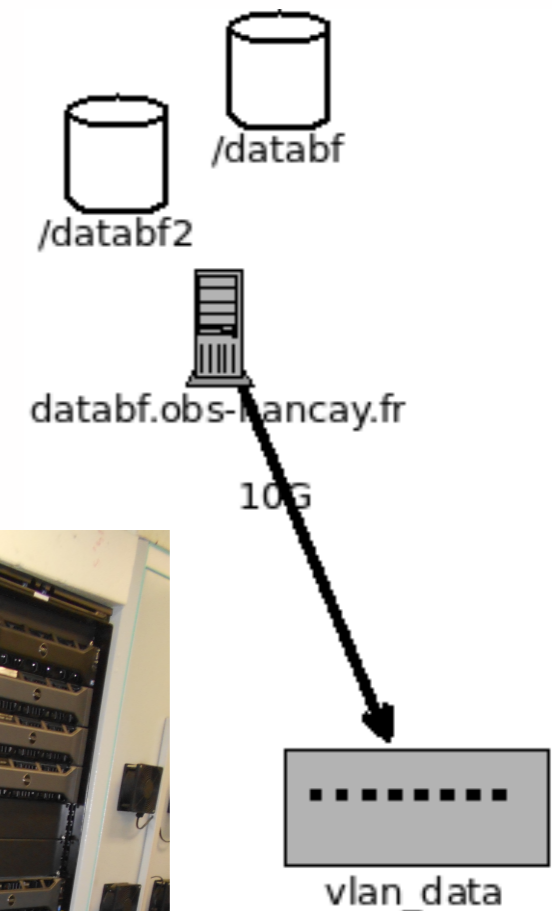
Management team

- Project Manager: Emmanuel Thétas
- End-User Representatives: Baptiste Cecconi & Julien Girard

Current Architecture of the Nançay Data Centre

Data Storage (databf)

- Server node = databf.obs-nancay.fr
 - 10Gbps connection to local data network
- 2 physical disk storage servers:
 - databf (320 TB): 8 % free space
 - databf2 (140 TB, up to 600 TB): 76 % free
→ 100To for NenuFar data
- Projects: LOFAR, NDA, NRT, NSA... and NenuFAR
- NB: Dedicated solar data storage (60TB)



Current Architecture of the Nançay Data Centre

Computing (nancep1/2/3/4)



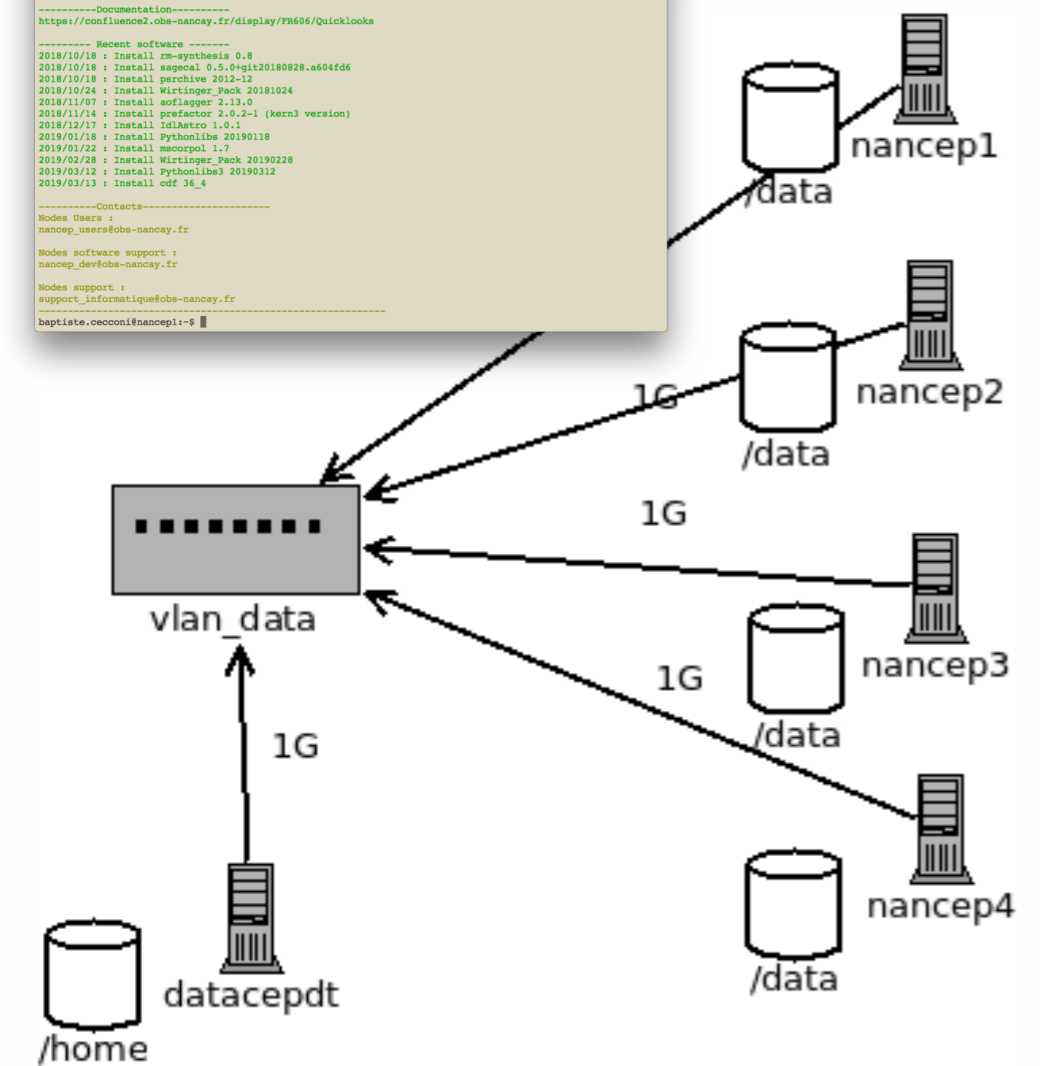
```

macbookbc1~ baptiste$ ssh nancep1
System information as of Mon Mar 18 09:45:53 UTC 2019
System load: 0.45      Memory usage: 2%   Users logged in: 10
Usage of /: 73.7% of 1.80GB  Swap usage: 1%
-> /data is using 87.3% of 18.19TB
Last login: Fri Mar 15 13:55:49 2019 from 145.238.186.5
Starting to work on nancep1@obs-nancy.fr.
Login directory for baptiste.ceconin is /home/baptiste.ceconin
09:45:54 up 293 days, 18:57, 31 users, load average: 0.50, 0.94, 1.40

-----Documentation-----
https://confluence2.obs-nancy.fr/display/FR606/Quicklooks

-----Recent software-----
2018/10/18 : Install rm-synthesis 0.8
2018/10/18 : Install sajpegal 0.5.0+git20180828.a604fd6
2018/10/18 : Install psrchive 2012-12
2018/10/24 : Install Wirtinger_Pack 20181024
2018/11/07 : Install aoflagger 2.13.0
2018/11/14 : Install prefactor 2.0.2-1 (Kern3 version)
2018/12/17 : Install IdlAstro 1.0.1
2019/01/18 : Install PythonLibe 20190118
2019/01/22 : Install macropol 1.7
2019/02/28 : Install Wirtinger_Pack 20190228
2019/03/12 : Install PythonLibe3 20190312
2019/03/13 : Install cdr 36_4

-----Contacts-----
Nodes Users :
nancex_users@obs-nancy.fr
Nodes software support :
nancex_dev@obs-nancy.fr
Nodes support :
support_informatique@obs-nancy.fr
baptiste.ceconin@nancy1~$
    
```

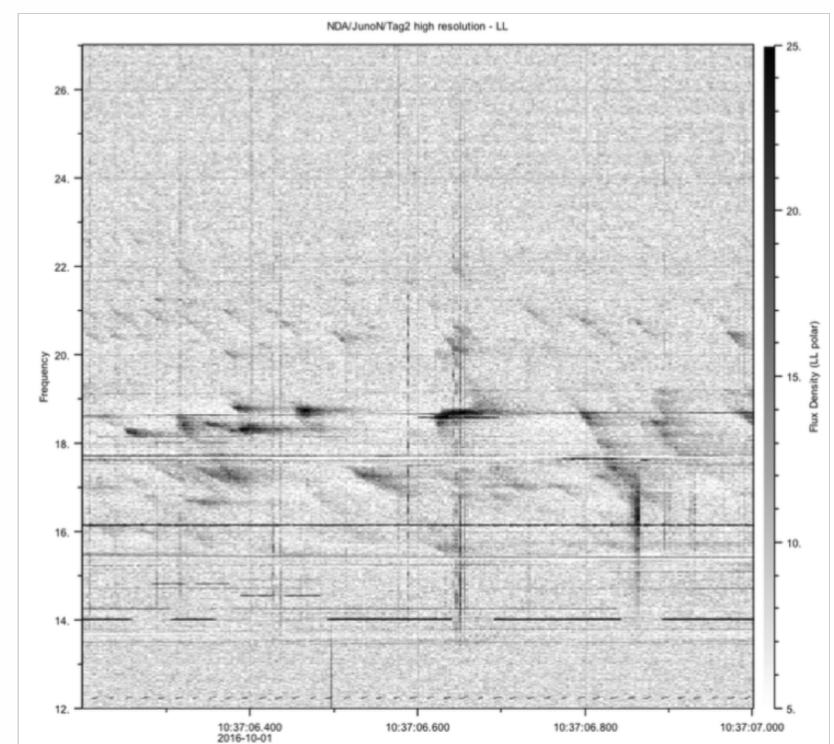
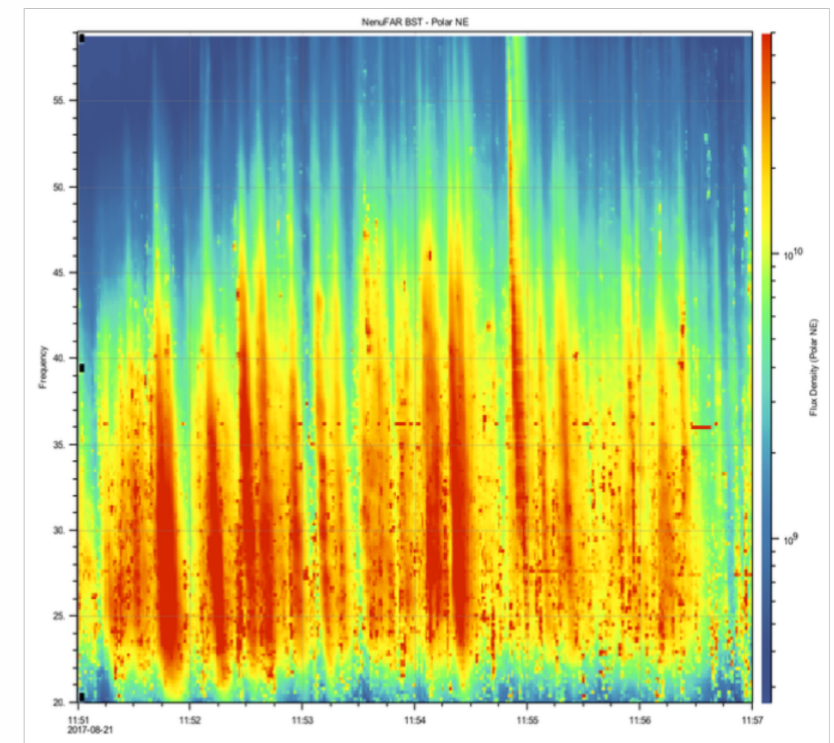


- Server node = nancep1/2/3/4
 - 1Gbps connection to local data network
- 4 physical servers:
 - nancep1/2: 16 CPU, 396 GB RAM
 - nancep3: 20 CPU, 256 GB RAM
 - nancep4: 28 CPU, 512 GB RAM
- Data access:
 - Shared /home directories (quota 50 GB)
 - Data volume on each node (for local data) quota = 1.5 TB
 - nancep1/2: 20 TB
 - nancep3: 30 TB
 - nancep4: 50 TB

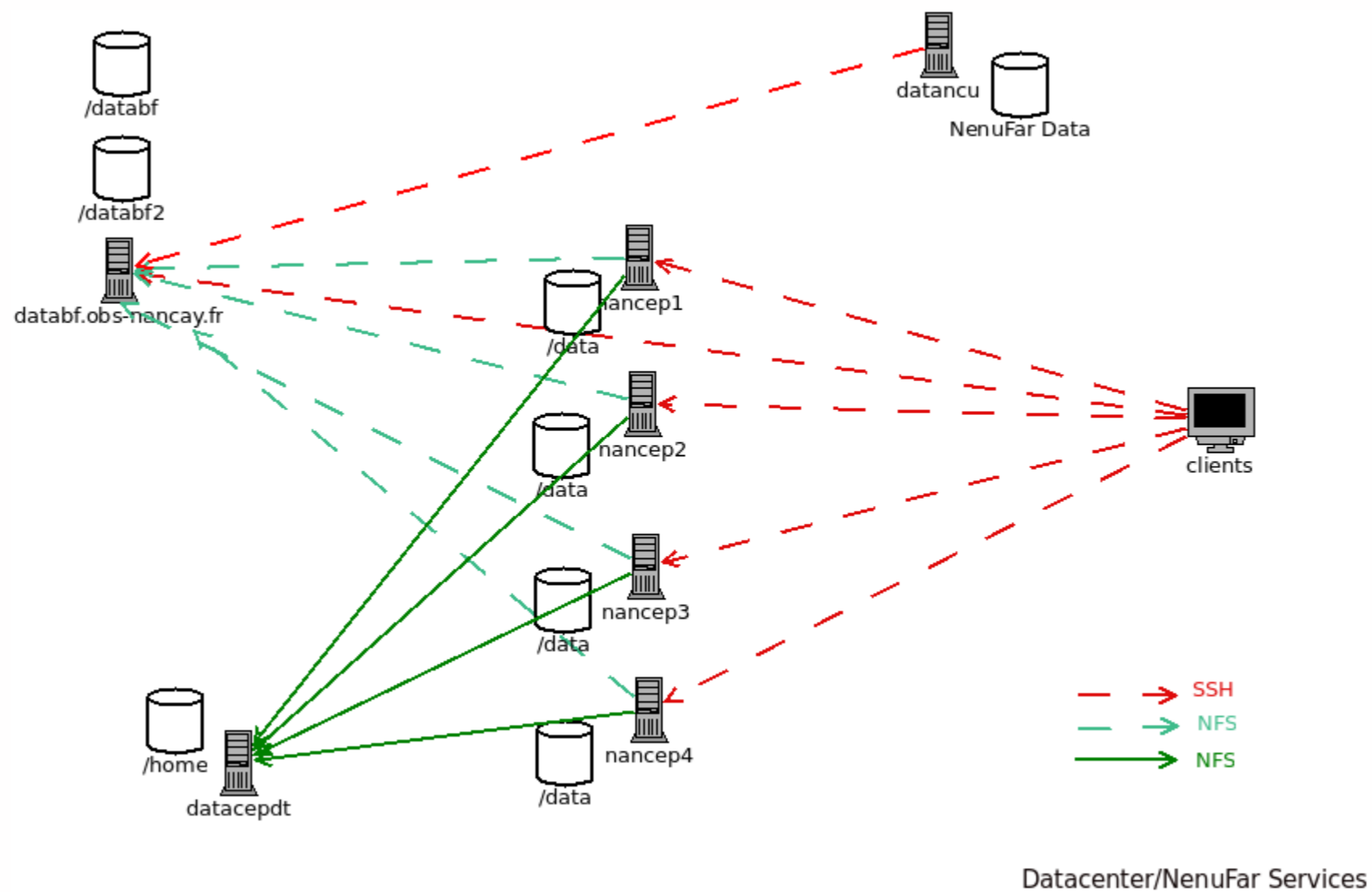
Current Architecture of the Nançay Data Centre

Data distribution (MASER)

- MASER (Measurement, Analysis, Simulation of Emissions in the Radio range) in Nançay:
 - Virtual Observatory server: <http://vogate.obs-nancay.fr>
 - Das2 server: <https://das2server.obs-nancay.fr>
- Virtual Observatory:
 - Server framework = DaCHS (by Uni. Heidelberg, DE)
 - Server implements all VO protocols from IVOA.
 - Currently only TAP services (Table Access Protocol), for NDA (and LOFAR prototype)
- Das2 (<https://das2.org>, by Univ. Iowa, USA)
 - Data streaming interface for time-series and spectrograms
 - Implemented for NDA (and NenuFAR prototype)
 - Visualisation client: Autoplot (<https://autoplot.org>)



Nançay Data Centre and NenuFAR



- NenuFAR data copied on *databf* in /databf2/NenuFAR
- Data available on *nancep* node for computing

Software available for processing

LOFAR Software

- Lofar
- Python-casacore
- Casacore
- « Prefactor » pipeline

NenuFAR

- Nenupy (from Alan Loh)
- Custom IDL library

Other

- Wirtinger Pack (from Cyril Tasse)
- DS9
- IDL
- Python (numpy, scipy, matplotlib, astropy...)
- Casa
- WSClean
- Aoflagger
- PSRchive

NenuFAR pre-processing facility project

Use and extend infrastructure of Nançay Data Centre

- Increase of data storage capacity up to 1 PB
- Add at least 1 (probably 2) nancep servers

NenuFAR pre-processing

- Beamformer : IDL & Python custom pre-processing tools (cf. Hands on) : de-dispersion, calibration, RFI mitigation, integration t-f
- Imager : standard Measurement Sets (nenums.py)
- Data production: ~1 PB / year (beamformer) + 3 to 4 PB / year (imager)

NenuFAR Archive

- Under study: local or distant or cloud (next slides)

NenuFAR post-processing

- Under study: cloud facility (next-next slides)

NenuFAR Archive

Several Technical Options

- Local archive in Nançay (multi-PB tape juke-box):
cost + manpower (maintenance + operation)
- National archive facility (CINES): *cost*
- European long term cloud storage (to be discussed with ASTERICS and ESCAPE projects): *cost + sustainability*

Open Access and FAIR

- Open access searchable metadata catalogue with coverage information (temporal, spectral, target...)
=> Findable
- Open access after embargo period
=> Accessible
- Using standard data formats, metadata and access protocols + Documentation
=> Interoperable and Reusable
- *For next NenuFAR "2.0"? include provenance information*
=> Reproducibility

NenuFAR post-processing online facility project

EOSC (European Open Science Cloud) facilities

- EC-funded (H2020) federation of data storage and data computing facilities (e.g., EGI, CERN, IN2P3...)
- Several projects with different goals: EOSC-Hub, PRACE, EUDAT, OpenAIRE...

Project with EGI (European Grid Infrastructure)

- Discussion initiated last week. Goal:
 - Online storage for computing (post-processing) and long term (archive)
 - Online computing for NenuFAR team members and NenuFAR observers
- Next steps: telecon with EGI (next Friday) and EOSC workshop in May 2019.

NenuFAR Imaging instance

Optimised instances for tasks that require more memory relative to virtual CPUs.

Features:

- High amount of RAM per CPU core.
- Up to 240 GB of RAM in total.
- Reserved instances

Ideal for:

- Running in-memory database
- Running in-memory stores (e.g. redis, memcached)
- In-memory big data processing engines (e.g. Apache Spark).

Configuration options:

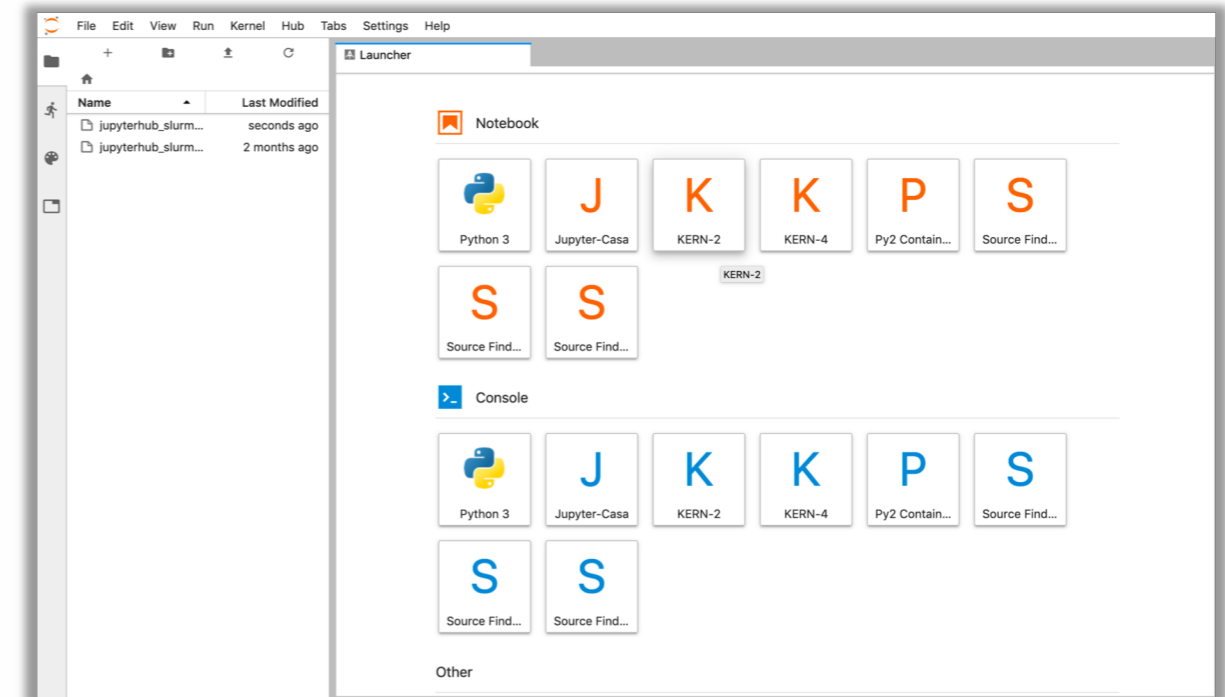
- NUMBER OF CPU CORES: 8
- AMOUNT OF RAM PER CPU CORE (GB): 120 GB
- LOCAL DISK (GB): 40 GB
- NUMBER OF VM INSTANCES: 1
- ACCESS TYPE: Reserved
- START OF SERVICE: 15/03/2019
- NUMBER OF PUBLIC IP: 1
- NUMBER OF DAYS: 365

ADD TO CART

NenuFAR post-processing online facility project

Detailed project discussed with EGI

- Once the NenuFAR pre-processing is done in Nançay:
 - Transfer and store the data (3 to 4 PB per year) in a online facility.
 - ➔ « *NenuFAR online data repository* »
 - Manage access to *NenuFAR online data repository* to NenuFAR team members and selected observers.
 - Possible use of « EduTEAMS » from GEANT
 - Provide VMs (including Jupyter Notebooks) to observers with preconfigured software for post-processing.
 - The NenuFAR team is managing the VMs' configuration.
 - There may be several kinds of VMs, depending on the processing.
 - Run the VM, with read-access to the *NenuFAR online data repository*,
 - The output of the processing is stored on a user's space in the *NenuFAR online data repository*.
- Open questions: cost and test+implementation time line



Example: *ilifu (idia SKA SA)*