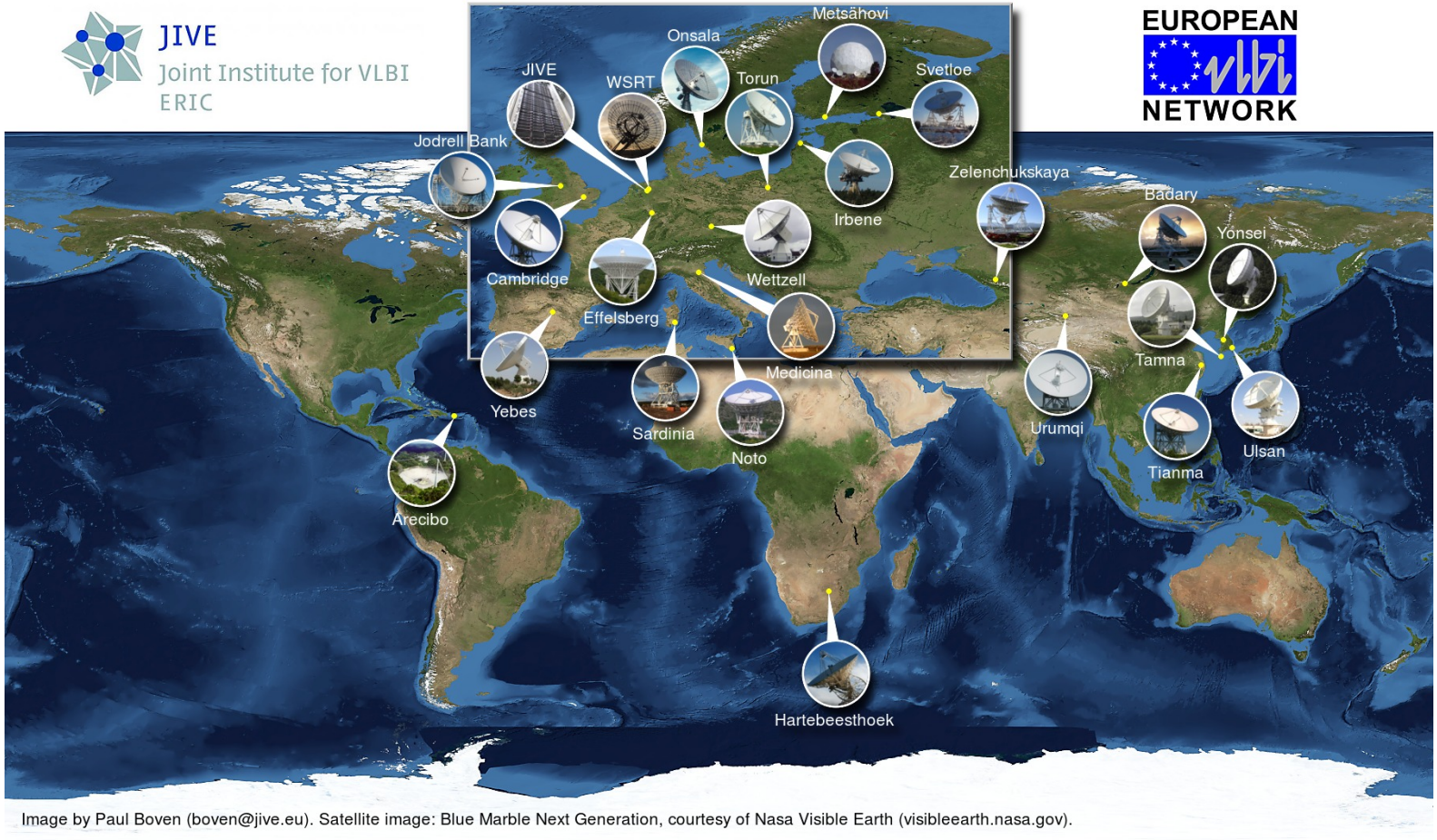
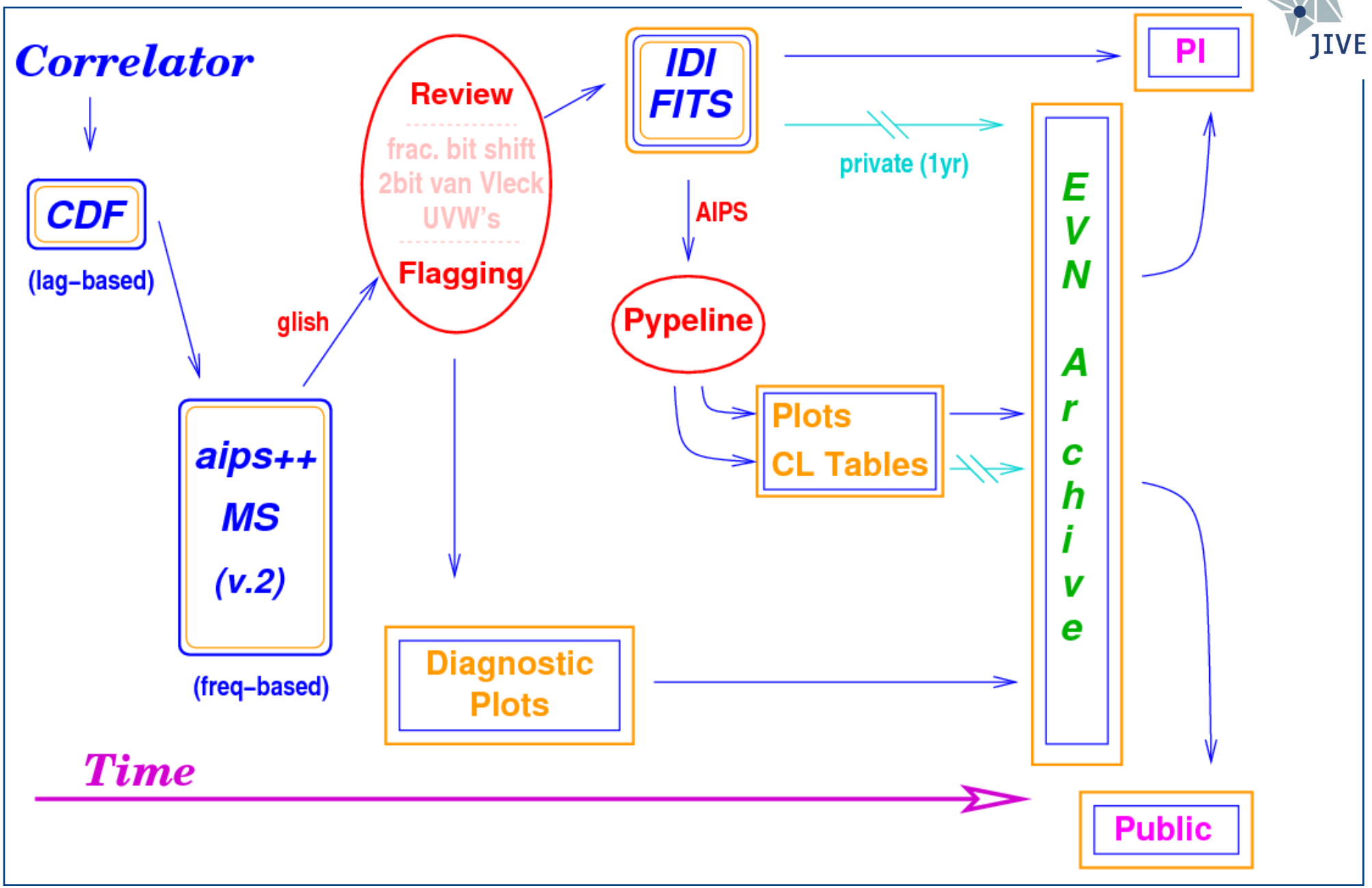
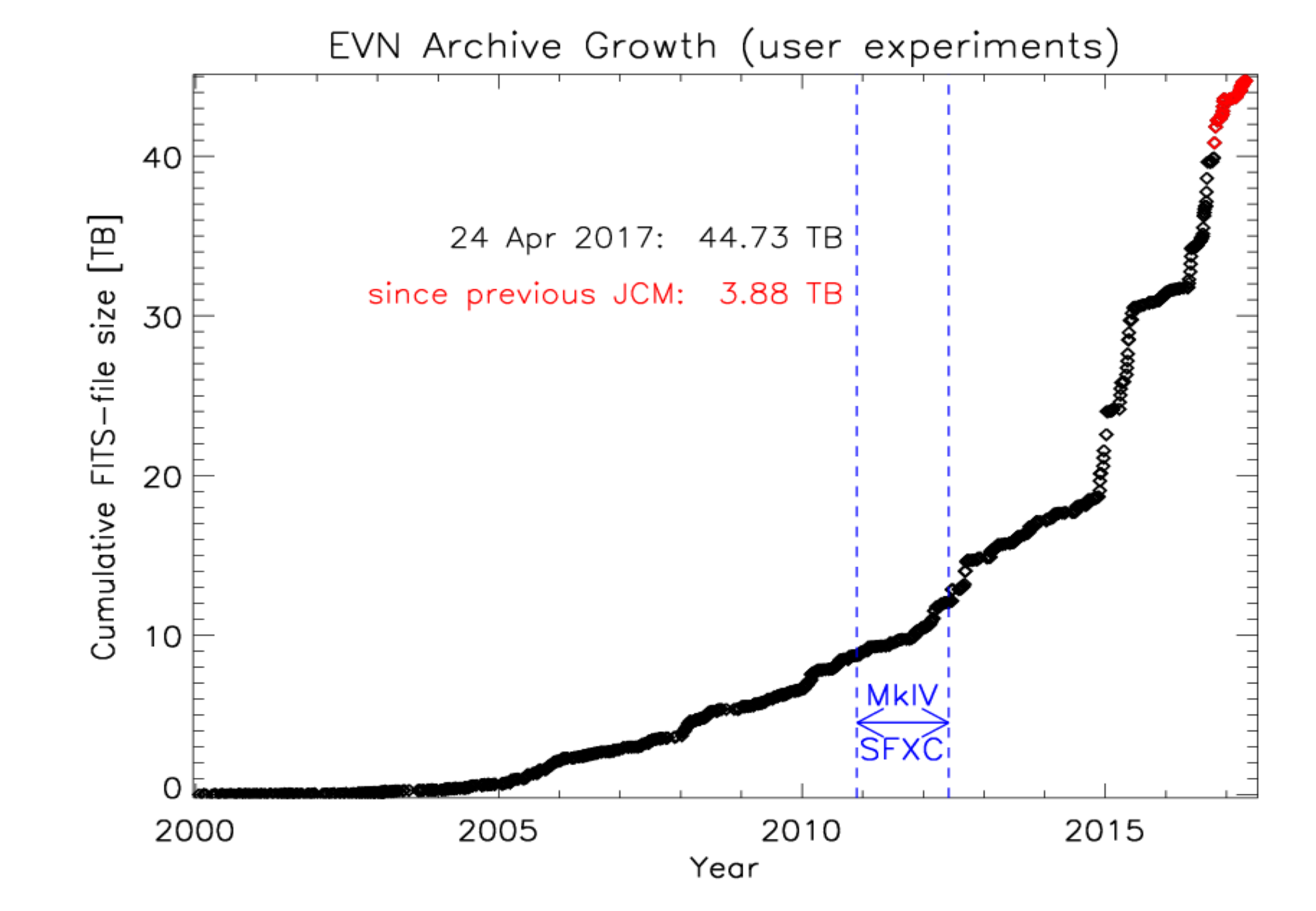
JIVE/EVN use case



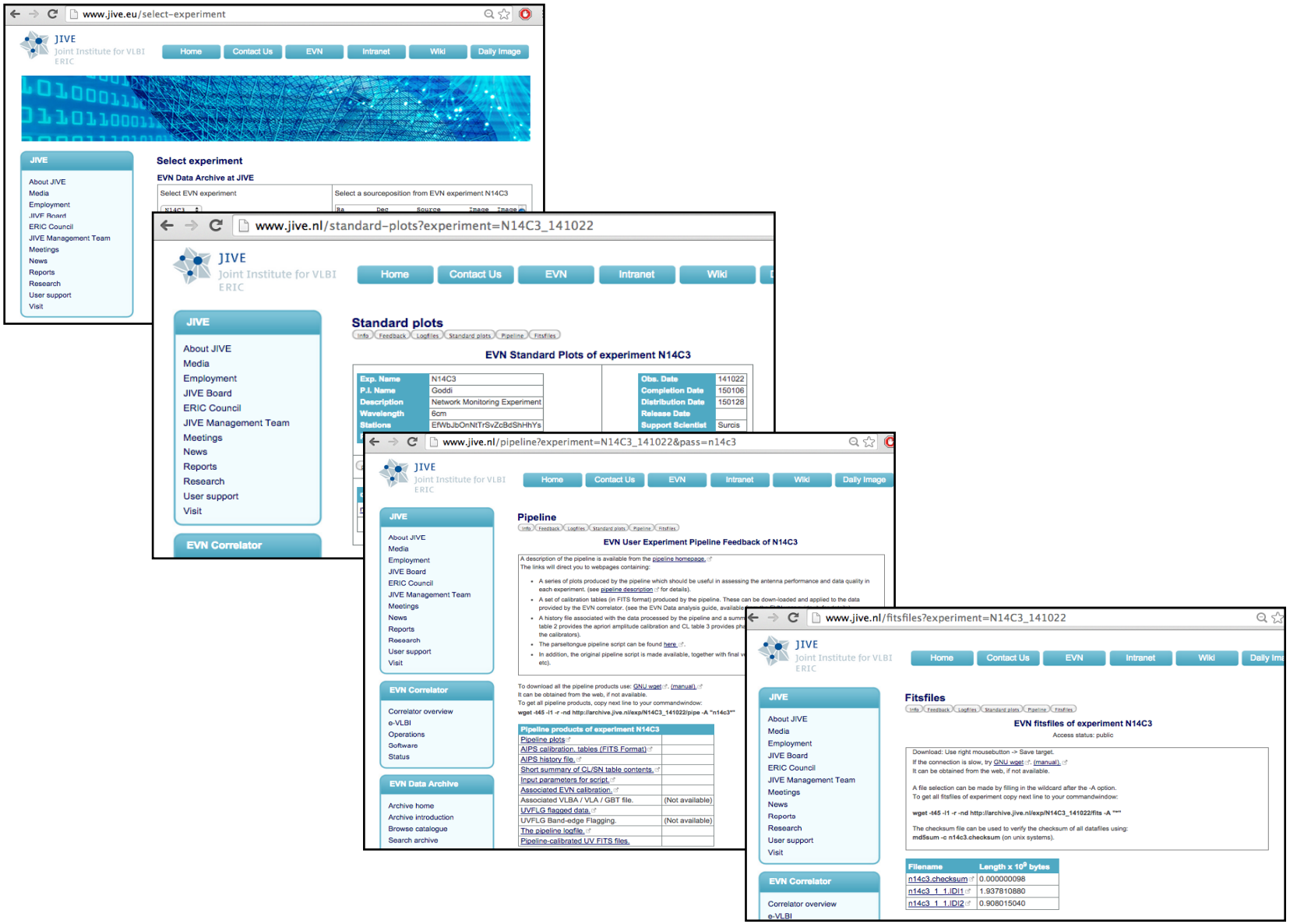
Arpad Szomoru

Post-correlation data flow

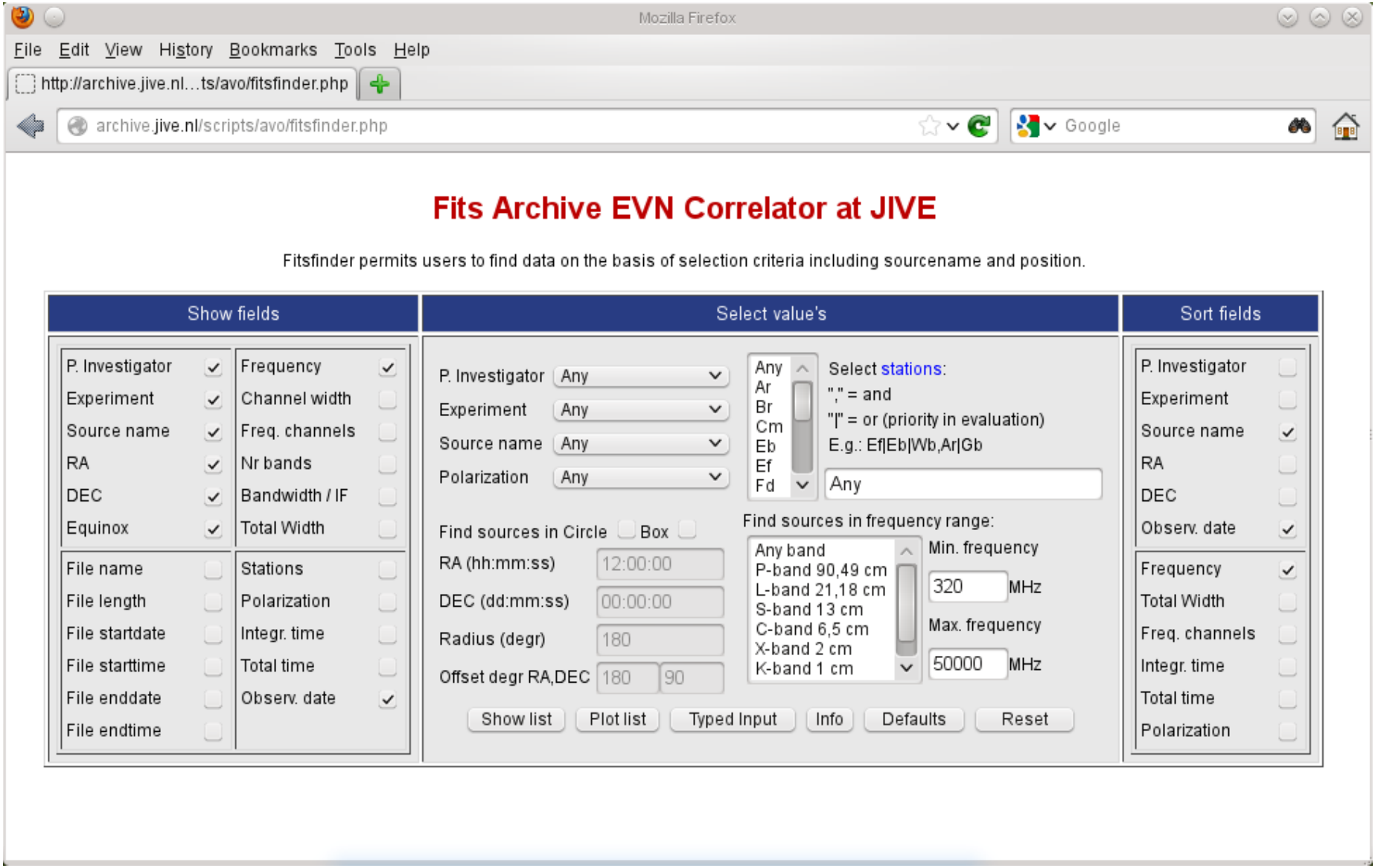



EVN Archive Growth (user experiments)
24 Apr 2017: 44.73 TB
since previous JCM: 3.88 TB
MkIV
SFXC
Cumulative FITS-file size [TB]
0
10
20
30
40
2000
2005
2010
2015
Year

EVN archive interface



Searching in the EVN archive



topics

- 2 types of data
- Correlation: gets thrown away, about 1.5 PB per session, 256 MB files
- Post-correlation: see archive plot, files of 2 GB. UV data in archive, calibration, imaging done by user

- First use case: archive in EOSC, searchable trough VO, tools like CASA in Jupyter workbooks, pipelines
- All addressed in WP3, 4 and 5

- Keep voltage data, re-correlation as service, wide field, multiple phase centers, search for FRBs

Processed data

- Correlation produces UV data
  - UV data are calibrated, not-publication ready images made
  - Stored in archive in FITS format (UV data, only very few images)
  - Files of 2GB
- Proprietary period of 1 year
  - PI can download data for further processing and imaging
  - Password protected
- After one year anyone can access the data
  - Does not happen terribly often though
  - Data is fairly FAIR, however the R could be improved upon

Processed data: future?

- Archive in EOSC
  - Possibility to feed back data
    - New data reductions, publications
  - Modernised, more user-friendly pipelines
  - Minimal re-computing
- Data reduction in EOSC using Jupyter-like notebooks
- Archive integrated in VO

Raw data

- Three sessions of several weeks per year
  - Roughly 1.5 PB raw voltage data is generated per session
  - First stored at stations, automatically e-shipped to JIVE after each separate observation ends
  - Using home-grown transfer software and UDT
    - Although we also still ship disk packs!
  - File size 256 MB, fuse system
- 10 - 20 times per year real-time e-VLBI, without any data recording
  - Straight from telescopes into correlator
  - Same software, but UDP (small amounts of packet loss acceptable)
- After correlation and validation of final product raw data are deleted (typically few months after observations)
- Total amount could easily be doubled (higher sensitivity)
  - Storage (and price of it) only limiting factor

Raw data: future?

- Keep raw data (forever of course)
  - After initial correlation according to original proposal
- In EOSC (of course)
- Offer correlation as a service
  - After proprietary period
  - Multiple phase centers
  - Real wide-field imaging
  - Search for Fast Radio Bursts

- Demonstrator should be feasible

JIVE & ESCAPE: WP3

WP3: mostly provide the tools that will be made available through the EOSC

- Analysis of functionality that is still needed to make CASA a complete VLBI data reduction package
- Implementation of missing functionality
- Integration of CASA6 in Jupyter or similar notebook
- Further work on containerization of software
- Creation of a VLBI data reduction pipeline suitable for the EOSC

JIVE & ESCAPE: WP4

Make radio data (starting with our EVN archive) accessible through the VO

- Investigating the handling of radio astronomical data in the VO
- The definition of a VO interface to the EVN archive
- Determine what metadata will be needed
- The design of a supporting database schema
- The implementation of web services and a database

JIVE & ESCAPE: WP5

Provide the tools and workflows that will make the software developed in WP3 accessible through the EOSC

- Analysis of the functionality of the JIVE archive
- Enable re-running pipelines with different parameters
- Archiving of new processing of data
- Enable feedback from users to archive
- Create a central control of information flows at JIVE