



ASTRON

Netherlands Institute for Radio Astronomy

ASTRON data in the VO

Our experiences



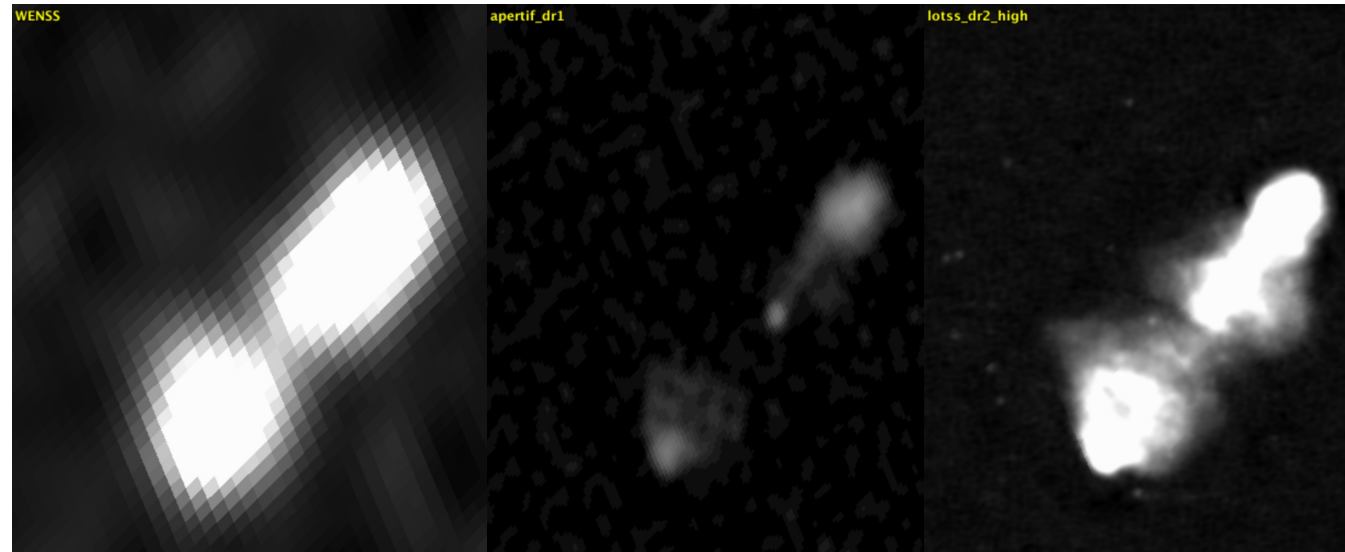
This work is co-funded by the ESCAPE project (Horizon 2020) under Grant number 824064.

Photo: Tom Oosterloo



VO @ ASTRON content

- WSRT, APERTIF, LOFAR
 - Also an ADR of GMRT data
- Most data are images
 - ARTS DR1 coming up -> PSRFITS
 - APERTIF polarisation fields -> PDF
- Source catalogues
- LOFAR archive is mostly visibilities or beamformed data. Historical WSRT data too.
 - Making it available links to having an easy way to run pipelines (science analysis platform)



ASTRON data

- Roughly three data localities
 1. Local to the (dachs) VO service instance
 - The 'smaller' data releases
 2. Data on national infrastructure
 - a) SURFsara (Amsterdam) (LOFAR raw data, LOFAR data releases, APERTIF tape)
 - b) PSNC (Poznań), FzJ (Jülich) (LOFAR raw data)
 3. Data on large storage, at ASTRON
 - APERTIF disk
- Also, some data on disk (ie directly downloadable), some on tape (ie needs to be moved to disk prior to download)





IVOA standards @ ASTRON

- HiPS
 - Ideally all imaging releases have a corresponding HiPS, but some have very limited sky coverage
 - <https://hips.astron.nl>
- SIA, TAP, ObsCore for data access, some cutouts
- DataLink used for linking in ancillary data products
 - Also for visibilities. Ideally relevant standards will exist in the future
- Future: considering UWS for data
- Of course there are a lot of VO standards ‘under the hood’ (UCD, VOTable, SCS, etc.)

VO@ASTRON history

- Disclaimer: the start predates me.
- Basically started by setting up dachs and put data in there in the ASTERICS project, the more complex data releases came in the last few years
- Main issues for (us as) beginners:
 - Found it hard to find a good introduction. The architecture (current version) is a great read; may be turned into a website?
 - You can do a LOT with dachs and sometimes one can get lost in there. Especially if a use case is a bit different from the norm it would be nice to have some guide lines on how things are meant to be used (e.g. UCDs are there to let an application know how to handle the column, does PDF in obscore make sense or not?, How bad are you if you just send 600MB images to users without using cutouts, etc.)

Suggestions based on our experience

- Mailing lists are great but I find them hard to query. Some form of public discussion forum, maybe also actively curated as an FAQ would be useful. I realise the use case is fairly limited.
- Data provider forums are a great way to keep in touch with others with more or less experience
- It may also be a good idea to have some material/information explicitly aimed at the data owner. In our case for instance, I think I somehow understand the context but the person responsible makes it easier to get input for the tables.
- (this is not at all advice for the IVOA but I will make it anyhow): Having a nice interface targeted at data release owners so that they do can do most the work themselves would be fantastic.
- Linting tools are really nice (have now a cron CI/CD pipeline to check HiPS and TAP validity). E.g. our HiPSes disappeared from Aladin because of some mistake on our end.

Discussion on VO standards

- Obviously the standards under development by the radio IG could be considered “missing”.
- Standards that support access to data that is intrinsically remote exist but there are not yet (many) working examples available (remote SODA cutouts would be awesome, data staging using UWS, workflow combining both even nicer).
- Maybe more generally: being prepared for the future means being prepared for larger tables, bigger images and spectra.

In hindsight

- Things for us went pretty smooth. It is absolutely key to engage with the VO community when beginning.
 - Joining working groups, reading the discussions on mailing lists can sometimes be very interesting
 - Knowing what documents to start with as well.
- Going to an IVOA for the first time can be overwhelming (a lot of jargon flying around even in the first session). The beginner sessions should make it a bit smoother.