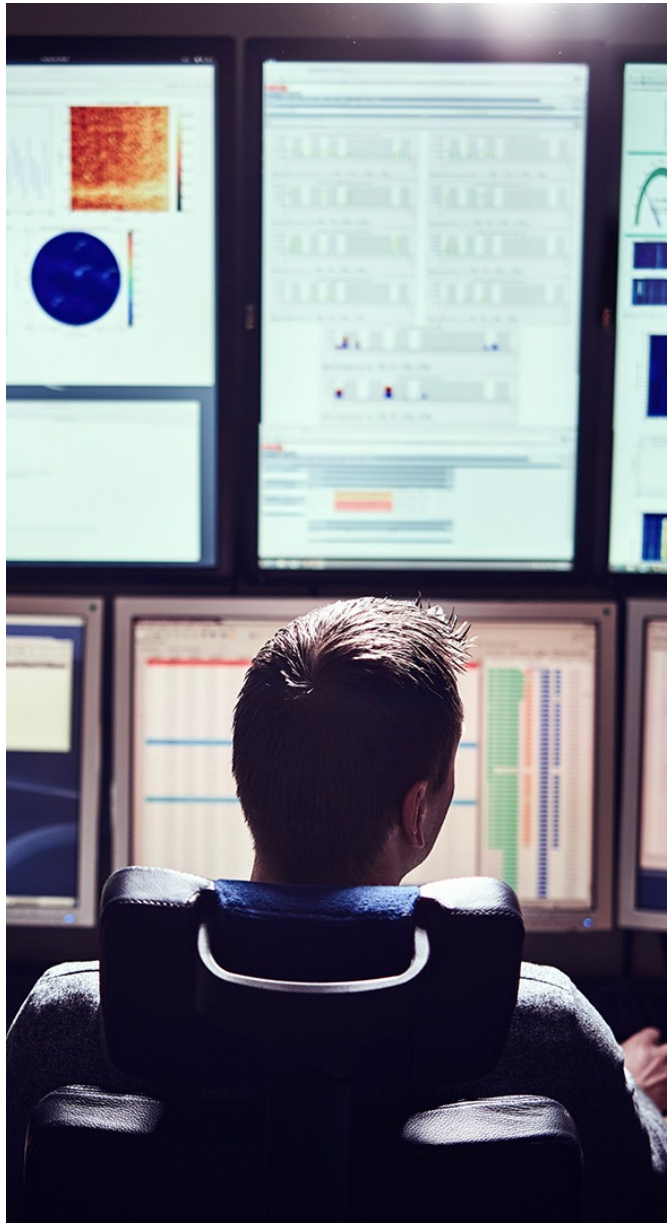
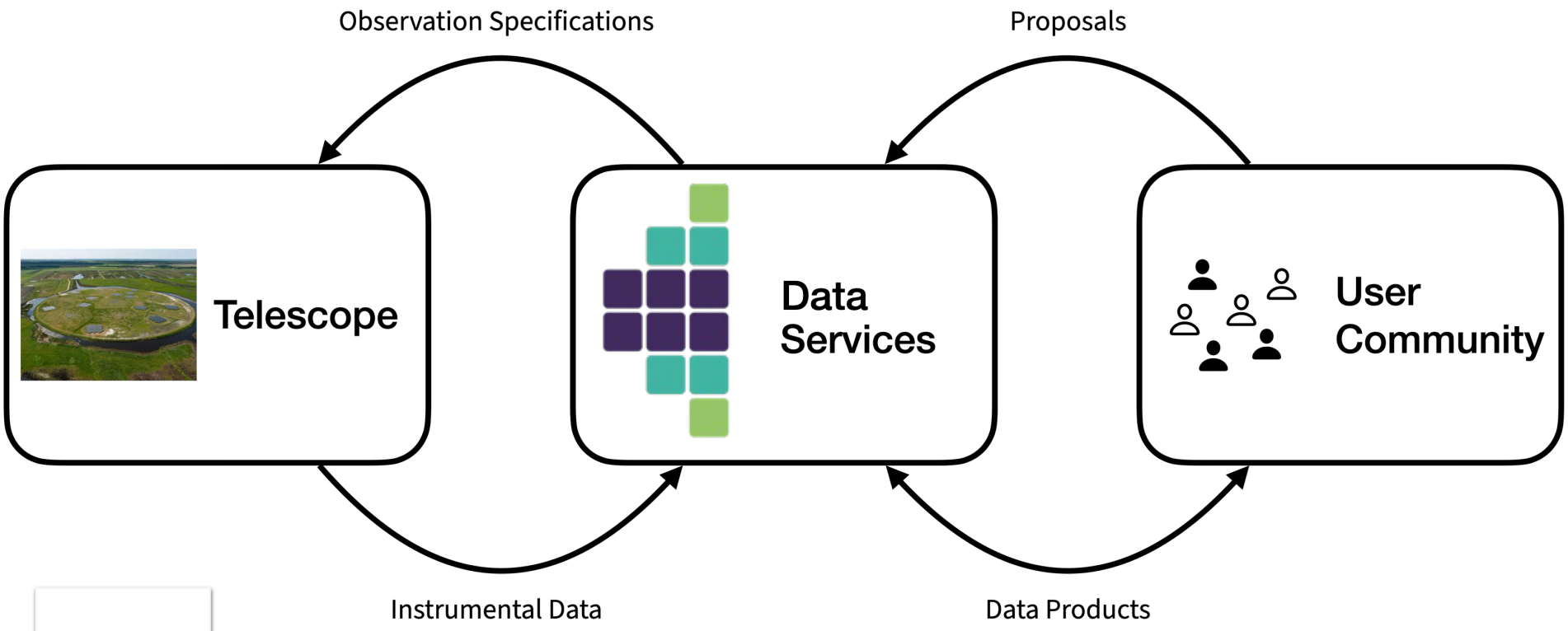
Developing the LOFAR2.0 Data Services

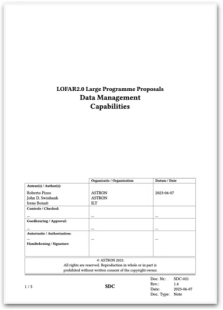
J. Swinbank & R. Pizzo

LOFAR Family Meeting, 23 September 2025



ASTRON
Netherlands Institute for Radio Astronomy

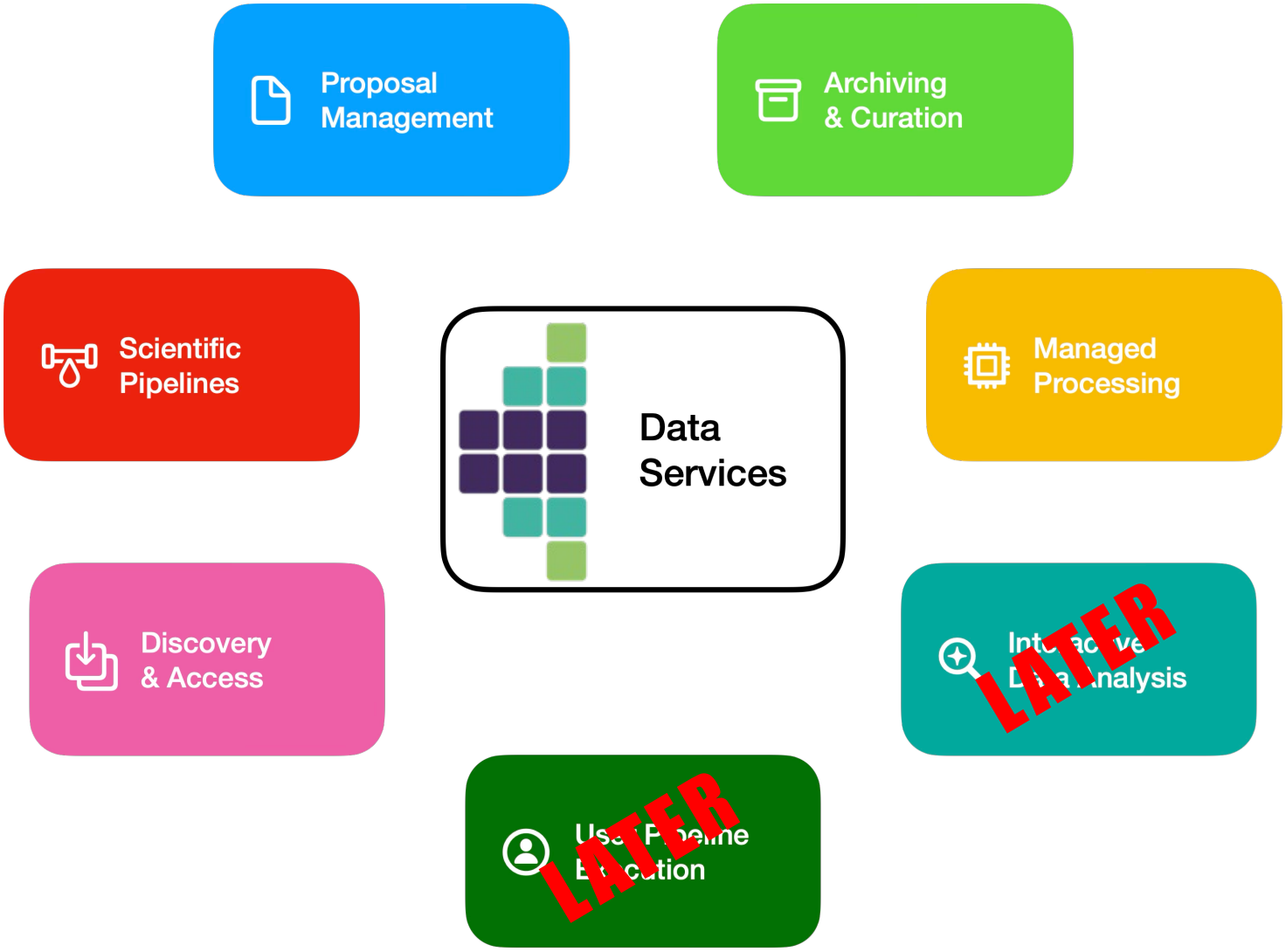




LOFAR2.0 Large Programme Proposals
Data Management Capabilities

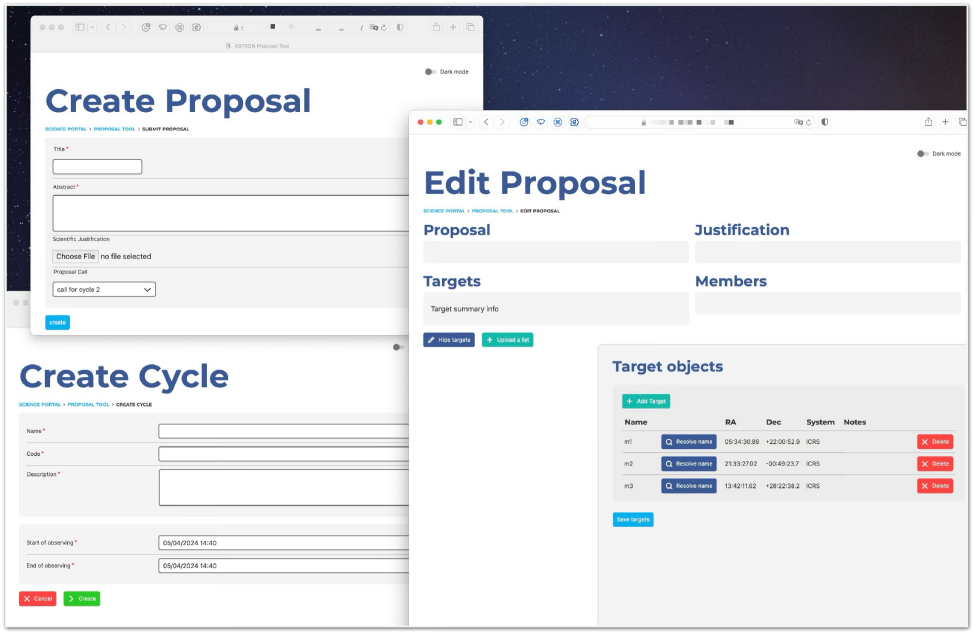
“Services to process, archive, and distribute LOFAR2.0 Data Products. These services, deriving from development effort, operational activities, and infrastructure capacity contributed by various partners, will be provided to end users under the management of the ASTRON Science Data Centre.”

LOFAR2.0 Data Management Capabilities; https://www.lofar.eu/lofar2-0-documentation/


Proposal Management
Archiving & Curation
Scientific Pipelines
Data Services
Managed Processing
Discovery & Access
LATER
LATER

Proposal Management

- A new proposal management tool, TULP, replaces Northstar in the LOFAR2 era.
- **Accessible**, **easy to use**, and **extensible to future observing modes**.
- Includes "**calculators**" for both observing characteristics and resource (storage, compute usage.
- Enables tracking and managing of **resource usage** across the whole science project lifecycle.
- TULP is already used by the LOFAR2 Commissioning Team for organizing **commissioning** observations and will be used for coordinating **Science Verification.**



more on this in next talks

Archiving & Curation

- Initial LOFAR2.0 capabilities:
  - Support the full suite of **advanced** (“science ready”) **data products** that will be delivered in the LOFAR2 era.
  - Ingest *standardized* advanced data products generated externally and make them available through the LTA (“**user data ingest**”).
  - Management of **data rights** per LOFAR ERIC policy.

- Future LTA ambitions include:
  - Support for fully “**FAIR**” data management.
  - Easier and more flexible extensions to new sites.

Discovery & Access

- Initial LOFAR2.0 capabilities:
  - The existing LTA web interface will continue to be supported.
  - A new data staging service (**StageIT**) provides *more robust and reliable capabilities* to download data.

- Future ambitions include:
  - A **new data portal** that will provide modernized, more accessible, easier to use access to LOFAR data collections.
  - Increased **use of Virtual Observatory standards** to improve interoperability with other systems and enable more choice in terms of access tools.

Observatory-Managed Processing

- Initial LOFAR2.0 capabilities:
  - The LOFAR Observatory will be able to **manage processing** (specifying processing jobs, tracking execution, collecting results) **across LTA** sites at SURF, Jülich and (very soon) Poznań.
  - A core set of **Observatory-Supported Pipelines** (see upcoming slides) will be fully tested and integrated on this infrastructure.

- …future plans:
  - Increased range of Observatory-Supported Pipelines.
  - Processing will be rolled out to additional LTA sites as they become available.

Pipeline Components

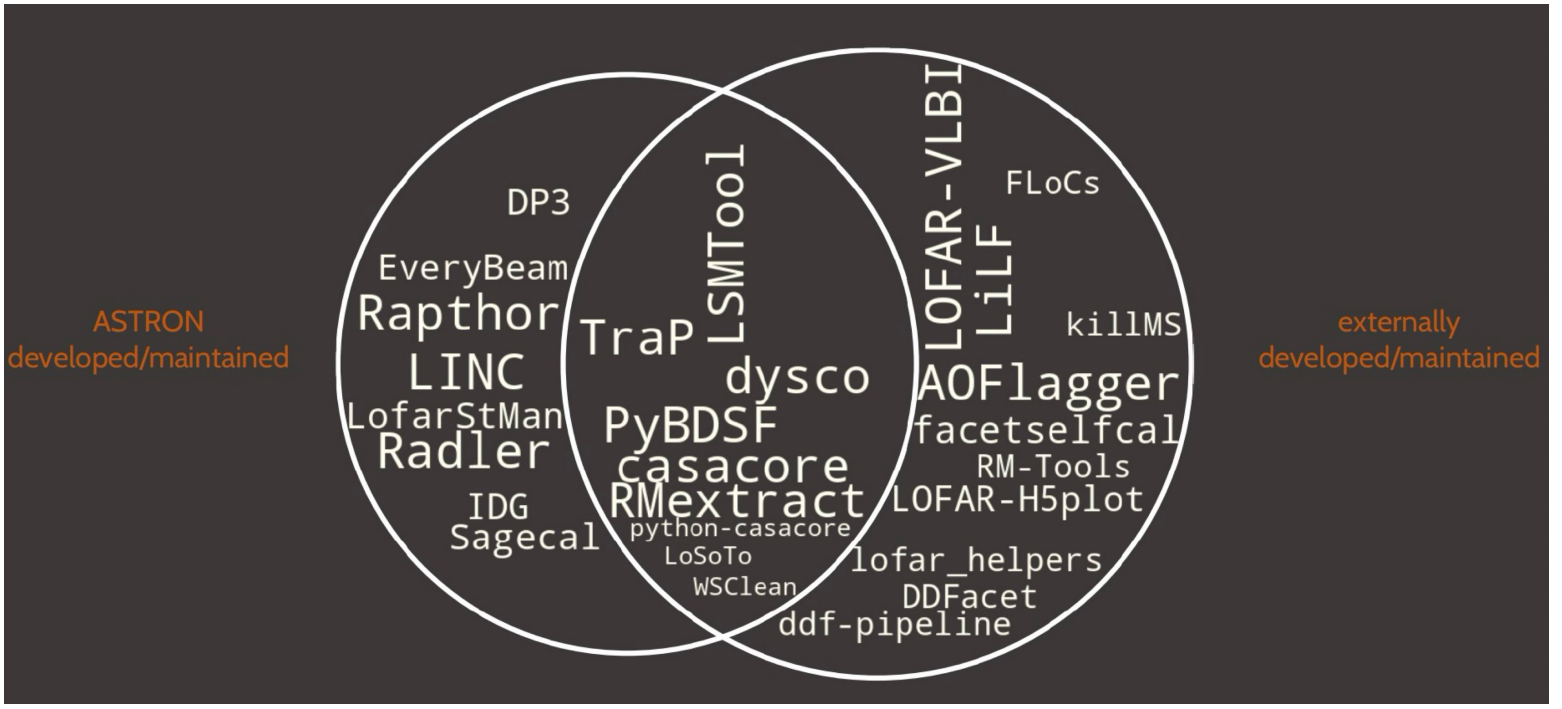


Figure: Sweijen (via Shimwell)

- ASTRON will continue to develop and support a range of critical pipeline components that are used across the LOFAR community and elsewhere. These are fundamental to the success of LOFAR data processing.
- We also acknowledge and thank the others across the community who do likewise!

Pipeline Components: Collaborative Development

- ASTRON has collaboratively developed several requested features, for example:

- **New compression methods**
  For VLBI & archive: model data (Sisco) & metadata compression

- **VLBI support in LINC**

- **New calibration techniques in Dp3, e.g.:**
  → For LBA: Extra constraints, Faraday solver and antenna-dependent solutions
  → For EoR: Robust weighting

- **Speed improvements, e.g.:**
  → Common: Baseline-dependent averaging
  → For VLBI: Data interpolation in Dp3

- **Imaging accuracy improvements**
  → Time and frequency smearing corrections

- **Maintenance tasks**
  → Releases, tests, bugfixes, platform support
  → Casacore, LSMTool, AOFlagger, Dysco, Losoto, Dp3, WSClean, PyBDSF, …

- **Use of GPUs**
  → Rewriting algorithms one by one for GPUs
  → Dp3's calibration and prediction tasks now have a GPU implementation
  → More will follow

Scientific Pipelines

- There are **many**, developed across the full community.
- As a community, we **need to work together** to ensure that the pipelines achieve operational capability.
- **These pipelines will be required to deliver standardized data products to the LTA**.
- The LOFAR Observatory is planning to appoint a person with a dedicated role towards coordinating pipeline development; we hope that will help us all make progress.

- Preprocessing
- PULP (known pulsars)
- LINC (initial calibration, QA)
- DDF-Pipeline (imaging)
- Rapthor (imaging)
- LOFAR-VLBI
- TraP (image-plane transients)
- LiLF (low frequencies)
- Cosmic Rays
- Lightning
- *…. more than conveniently fit in this box …*

Deploying Pipelines

- The Observatory will support two models for processing data:
  - **Centralized**. Directly under the control of the Observatory Operations team, on hardware provisioned at an LTA site (SURF, FZJ, PCSS… potentially UK, IT, others).
  - **Federated**. Managed by an external entity (e.g. a national data centre, a large programme) under agreement with the Observatory/ERIC, *using hardware pledged to and allocated by LOFAR ERIC*. Data products *must be* returned to the LTA.
- To offer the highest-possible standard of service to the widest-possible user community **all data processing will become centralized over time.**
- …but we acknowledge that full centralization from the start is impossible, so **federation is a temporary solution**.

Observatory-Supported Pipelines

- Pipelines that the Observatory takes responsibility for executing in a centralized system, **regardless of origin**.
- Meet standards for performance, reliability, sustainability, openness.
- **Not a transfer to the Observatory**; original developers will stay involved to provide support, maintenance, etc.
- **Draft policy** for Observatory-Supported Pipelines discussed internally at ASTRON and with Michiel as ERIC Director; intend to submit to the LOFAR STC for feedback shortly.

Summary & Conclusions

- The LOFAR Data Services include proposal, project, data, and processing management tools.
  - These are all vital to the success of the Observatory.
  - They represent a significant and ongoing development & maintenance effort.
  - We are on track to deliver usable services for LOFAR2.0, and have ambitions for future extension to create an even more productive system.
- The development of processing pipelines continues apace across the LOFAR community, including inside ASTRON. We will look to deepen collaboration.
- We anticipate two models of processing in the LOFAR2.0 era: ‘centralized’ and ‘federated’.
  - Eventually, all processing will be centralized using Observatory-Supported Pipelines.