
ASTRON
Netherlands Institute for Radio Astronomy

LTA Development Meeting

Welcome

ASTRON
Netherlands Institute for Radio Astronomy

Logistics

- Wifi: eduroam, or wiguest.
- Teams: remote participants, and presenting. Details on Indico page (https://indico.astron.nl/event/380/).
- Lunches, coffee breaks, etc — right outside this room.
- WC, coffee machine, etc — round the corner
- Is everybody ok with photos?
- Who to contact if you need something: me, or is-secretariat@astron.nl.

Hotel

- Suggested hotel (Fletcher Landhotel De Borken) is ~20 minutes walk away.
- If you came direct to ASTRON by taxi and you need help with transport to the hotel, please let me know and we'll arrange a taxi and/or carpooling.

Dinner (Thursday evening)

- 18:00 at Onder de Eiken (https://onderdeeiken.nl/, Drift 20a, 7991AB Dwingeloo)
  - 30 minute walk from the hotel
- There will be a fixed price menu, and you can order your own drinks on our tab. Please be reasonable and don't bankrupt us! 🍻
- We can help arrange taxis / carpooling — please let me know if you need transport.
- *Show of hands so we can confirm numbers.*

Why are we here?

- *Social* and *technical* reasons.
- *Technical*:
  - The LOFAR LTA is already among the largest and most complex astronomical data archives. And it's getting larger and more complex, and user needs are growing. We need to start converging on how to manage those challenges.
- *Social*:
  - The best way to manage a challenge of this sort is by building a mutual understanding and rapport. We're here to get to know each other and understand who we are working with in the future.

Structure & Agenda

- https://indico.astron.nl/event/380/
- Present & future of LOFAR as an instrument.
- Current status of LTA sites.
- Ideas about what comes next.
- Learning from other facilities — Rubin, SKA.
- Informal; lots of time for discussion.
- *Note that this is an open (ish) meeting.*

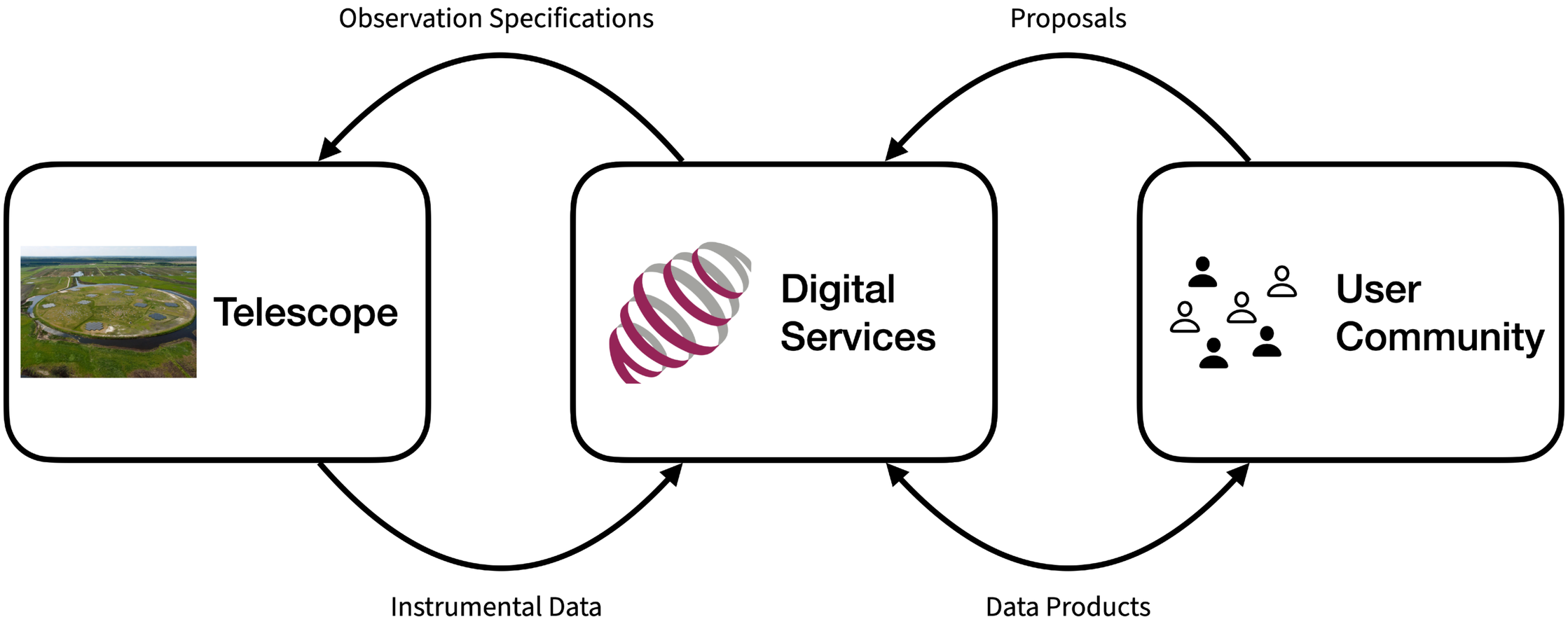
Who are we?

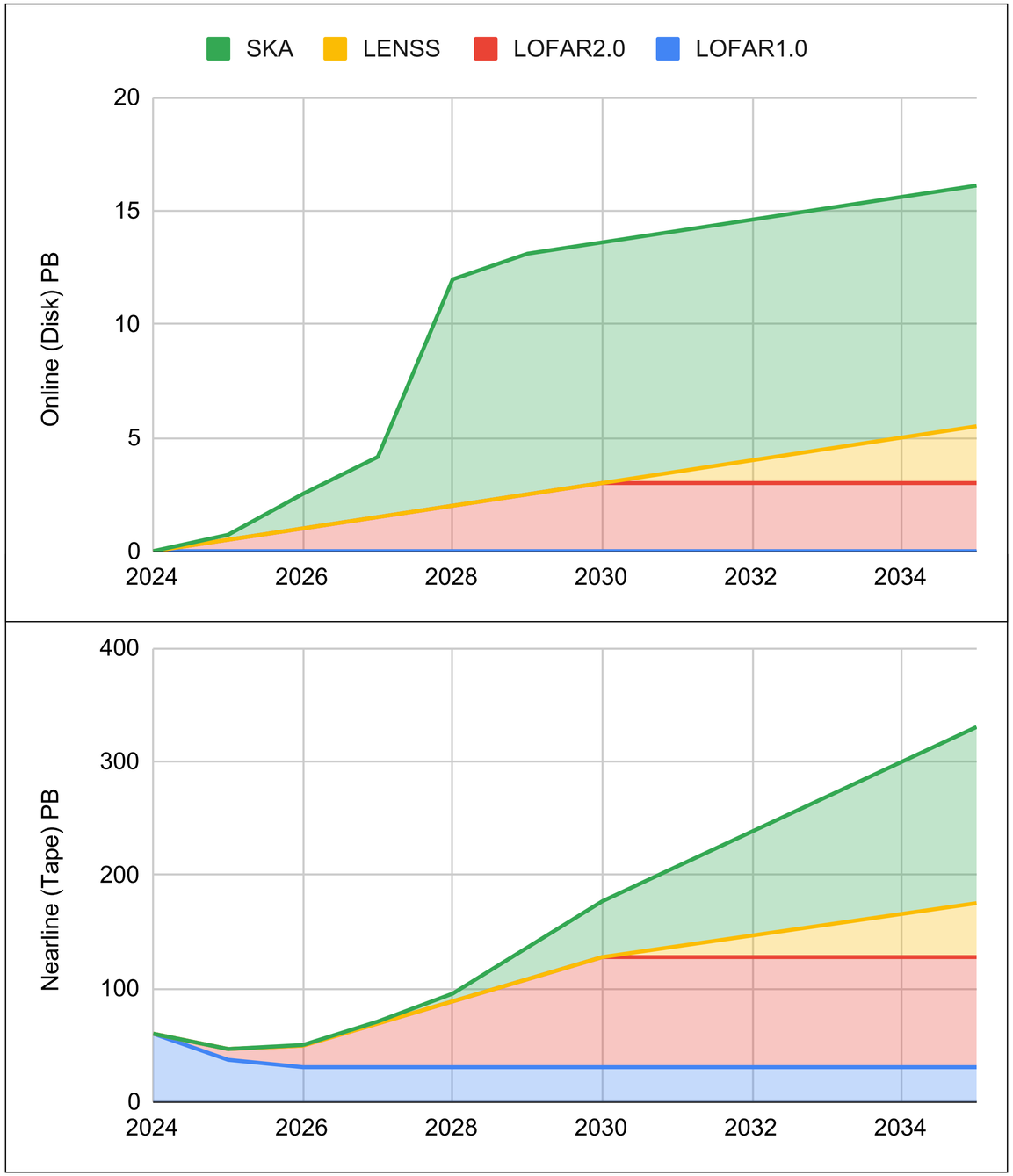
- Table round of brief (~30 second) introductions.

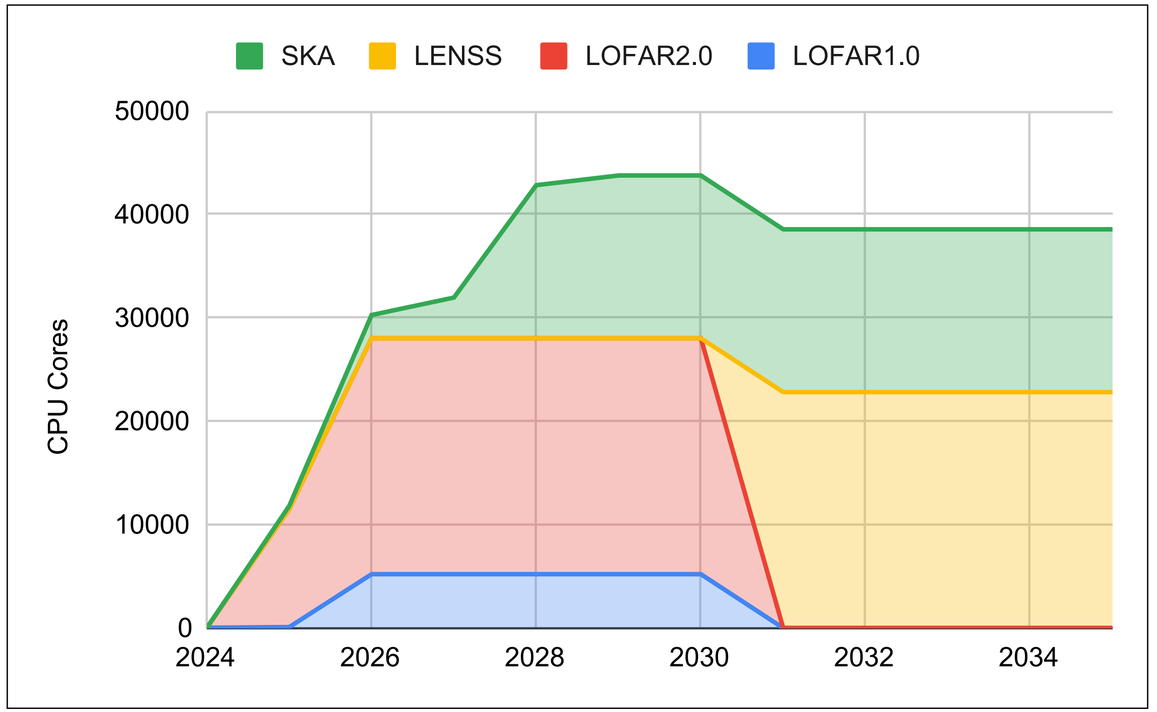
A Vision for LOFAR Digital Services

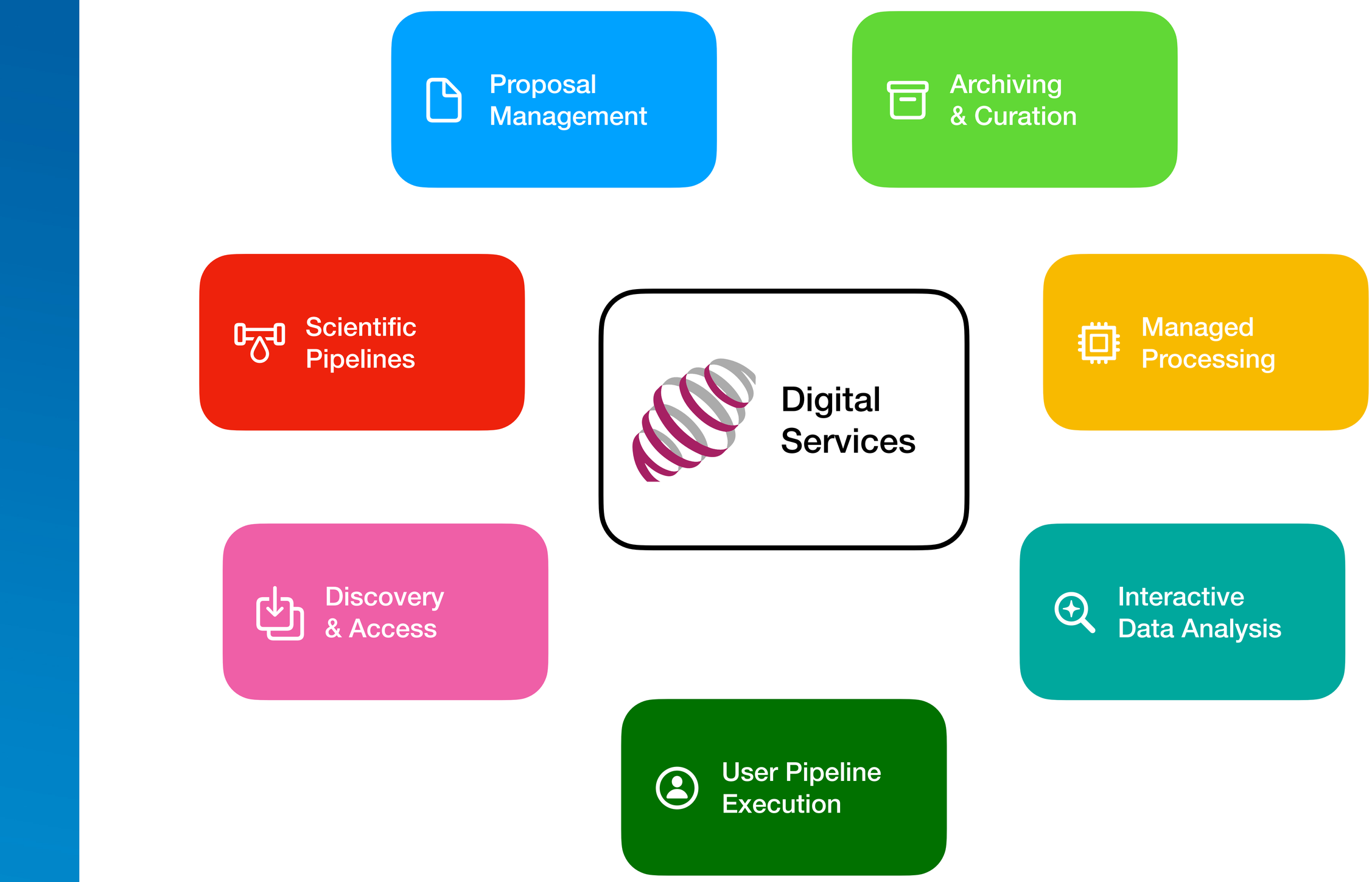
John D. Swinbank
swinbank@astron.nl

ASTRON
Netherlands Institute for Radio Astronomy


Observation Specifications
Proposals
Telescope
Digital Services
User Community
Instrumental Data
Data Products


SKA
LENSS
LOFAR2.0
LOFAR1.0
Online (Disk) PB
20
15
10
5
0
2024
2026
2028
2030
2032
2034
Nearline (Tape) PB
400
300
200
100
0
2024
2026
2028
2030
2032
2034


SKA
LENSS
LOFAR2.0
LOFAR1.0
CPU Cores
50000
40000
30000
20000
10000
0
2024
2026
2028
2030
2032
2034


Proposal Management
Archiving & Curation
Scientific Pipelines
Digital Services
Managed Processing
Discovery & Access
Interactive Data Analysis
User Pipeline Execution

Proposal Management

Archiving & Curation

Scientific Pipelines

Digital Services

Managed Processing

Discovery & Access

Interactive Data Analysis

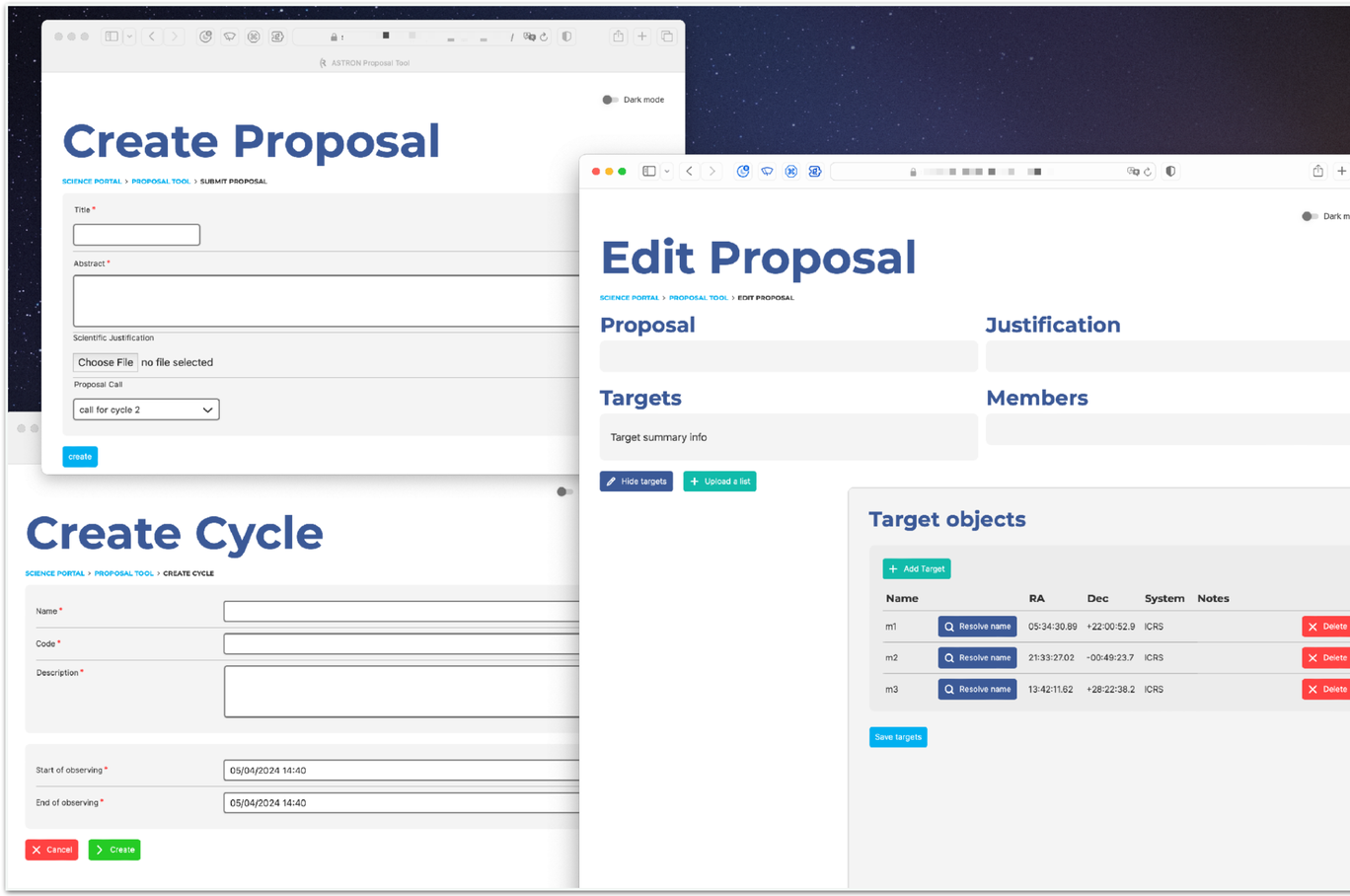
COMING LATER

User Pipeline Execution

COMING LATER

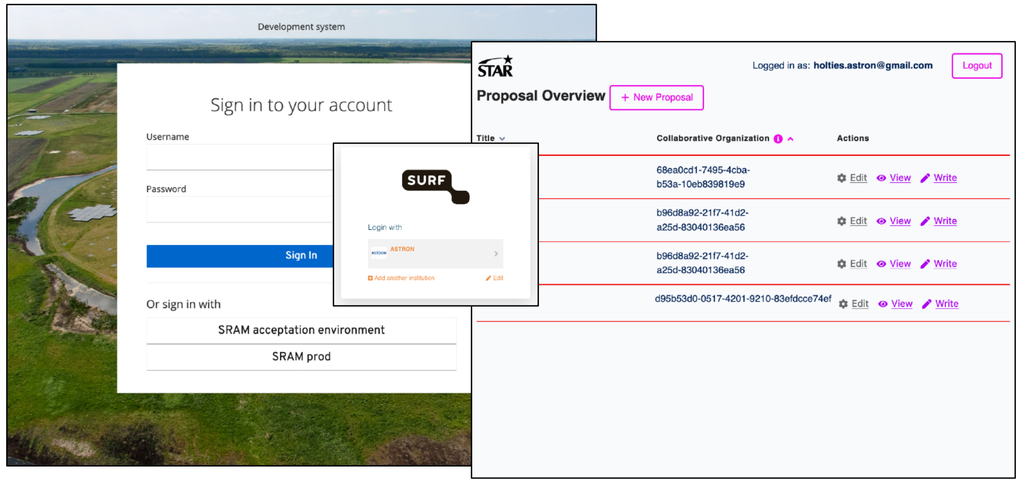
Proposal Management

- Replacement for the Northstar tool used in LOFAR1, which is obsolete & unmaintainable.
- This also includes new user/group management systems: “federated authentication & authorization”.
- Usable / testable / commissionable version ~this summer.

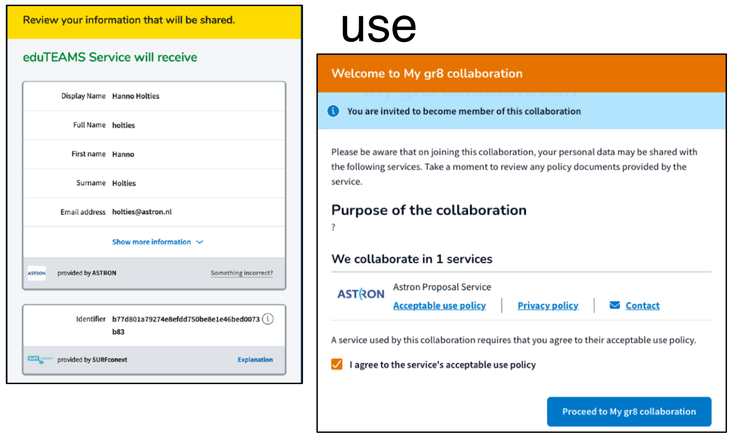


Federated AAI: SURF SRAM

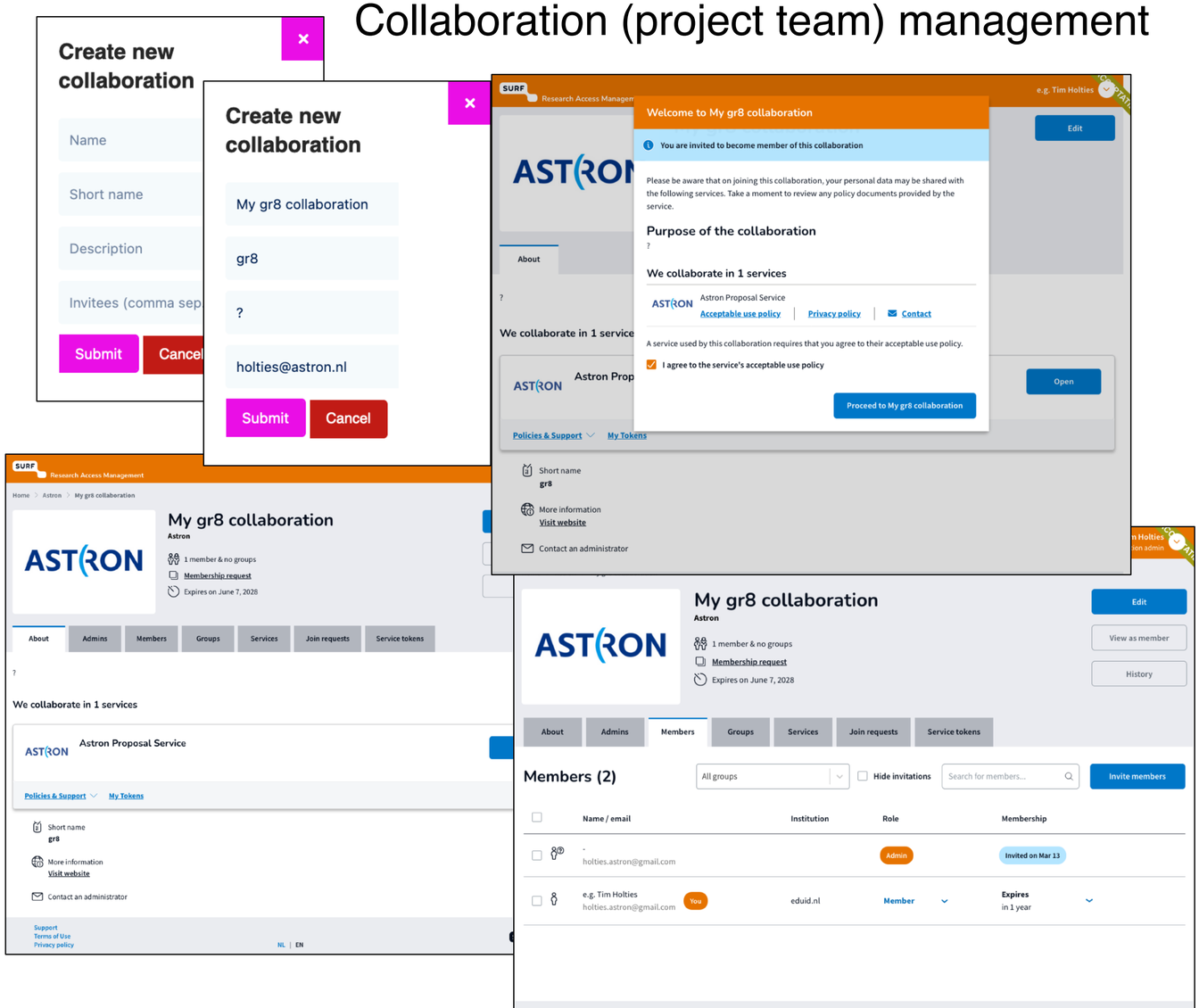
Logging in with your own institutional account



Consent for GDPR & Acceptable use



Collaboration (project team) management



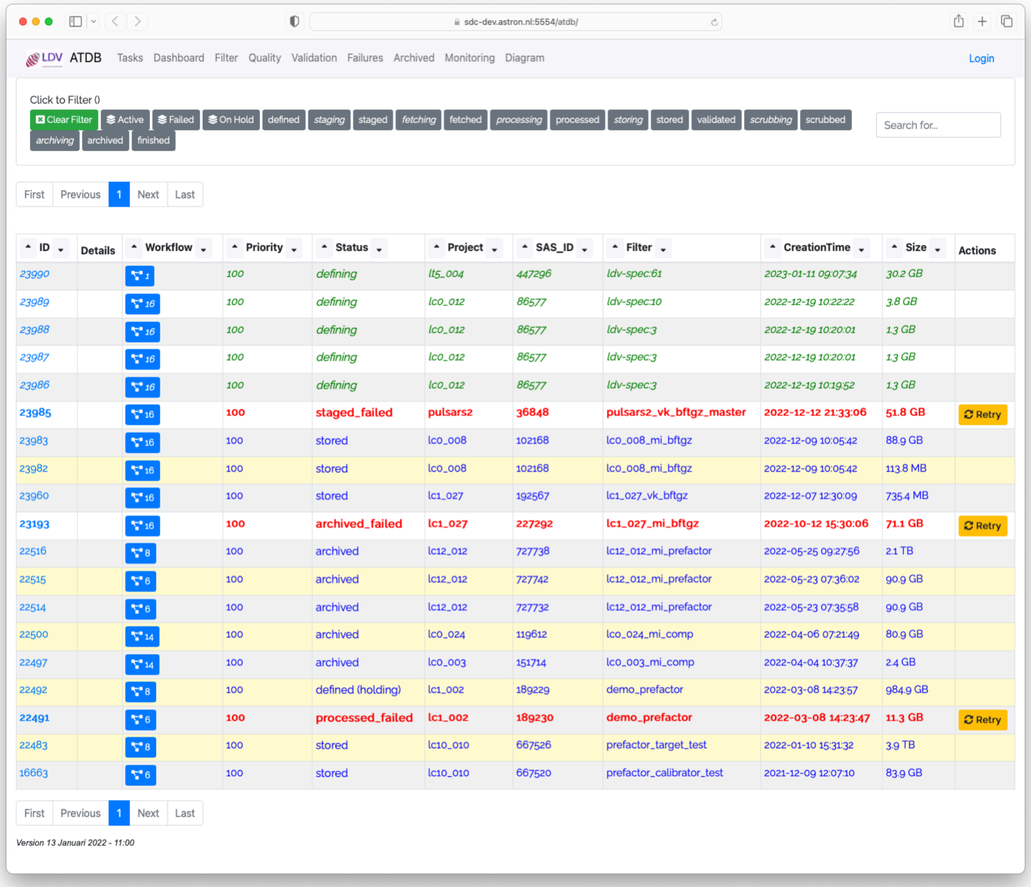
Archiving & Curation

Product Type	Example	Retention Period
Raw	Unprocessed visiblities	Not retained
Instrumental	Flagged & compressed visibilities	O(18 months)
Intermediate	Direction-independent calibrated visibilities	O(18 months)
Advanced	Image cubes	Indefinite
Special Cases	Unique observations that cannot be repeated	For discussion

- LTA Support for LOFAR ERIC agreed data policy.
- LTA support for “advanced” data products (e.g. images).
- Goal: the ability to ingest advanced / science-ready data products generated by the wider community.
  - Including management of data rights.
  - Become a “hub” for access to LOFAR data, wherever it is generated.

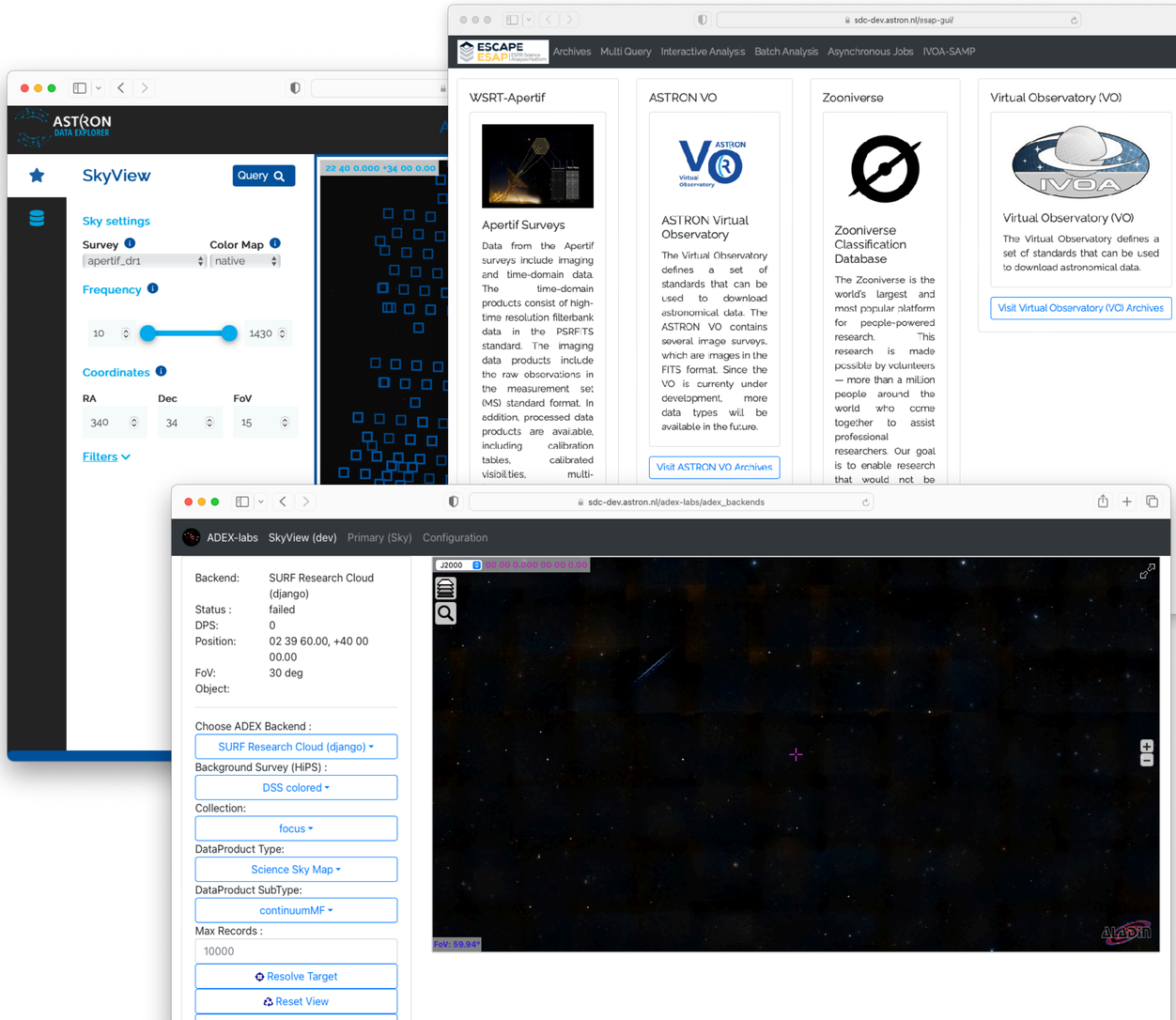
Managed Processing

- Execute predefined pipelines “at scale” against data in the LTA.
- Capability developed in the context of the LOFAR Data Valorization effort, currently running at SURF.
- Future work:
  - Scale to other LTA sites (Jülich, Poznań, maybe more).
  - Increased automation.
  - Incorporate more pipelines.
  - Polish & user enhancements.



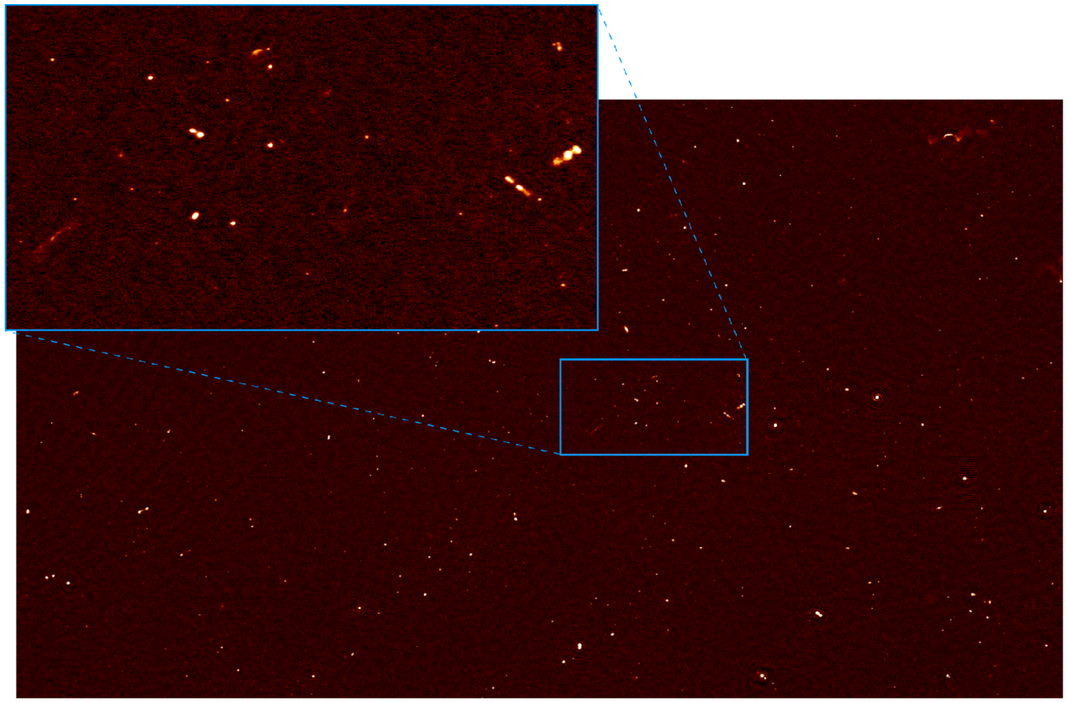
Discovery & Access

- Software solutions to make archive access as robust & reliable as possible...
- ...in tandem with LOFAR ERIC service level agreements with data centres.
- Upgraded archive interface: “ADEX”.
- Pervasive use of Virtual Observatory interfaces for publishing data.
- Aiming for a fully “FAIR” compliant archive:
  - Findable, Accessible, Interoperable, Reusable
  - https://force11.org/info/the-fair-data-principles/



Scientific Pipelines

- The Observatory-supported portfolio for LOFAR2.0:
  - Pre-Processing
  - LINC (direction-independent calibration)
  - Rapthor (direction-dependent calibration)
  - VLBI (postage stamps & wide field)
  - PULP (known pulsars)
  - TraP (image plane transients; *stretch goal*)

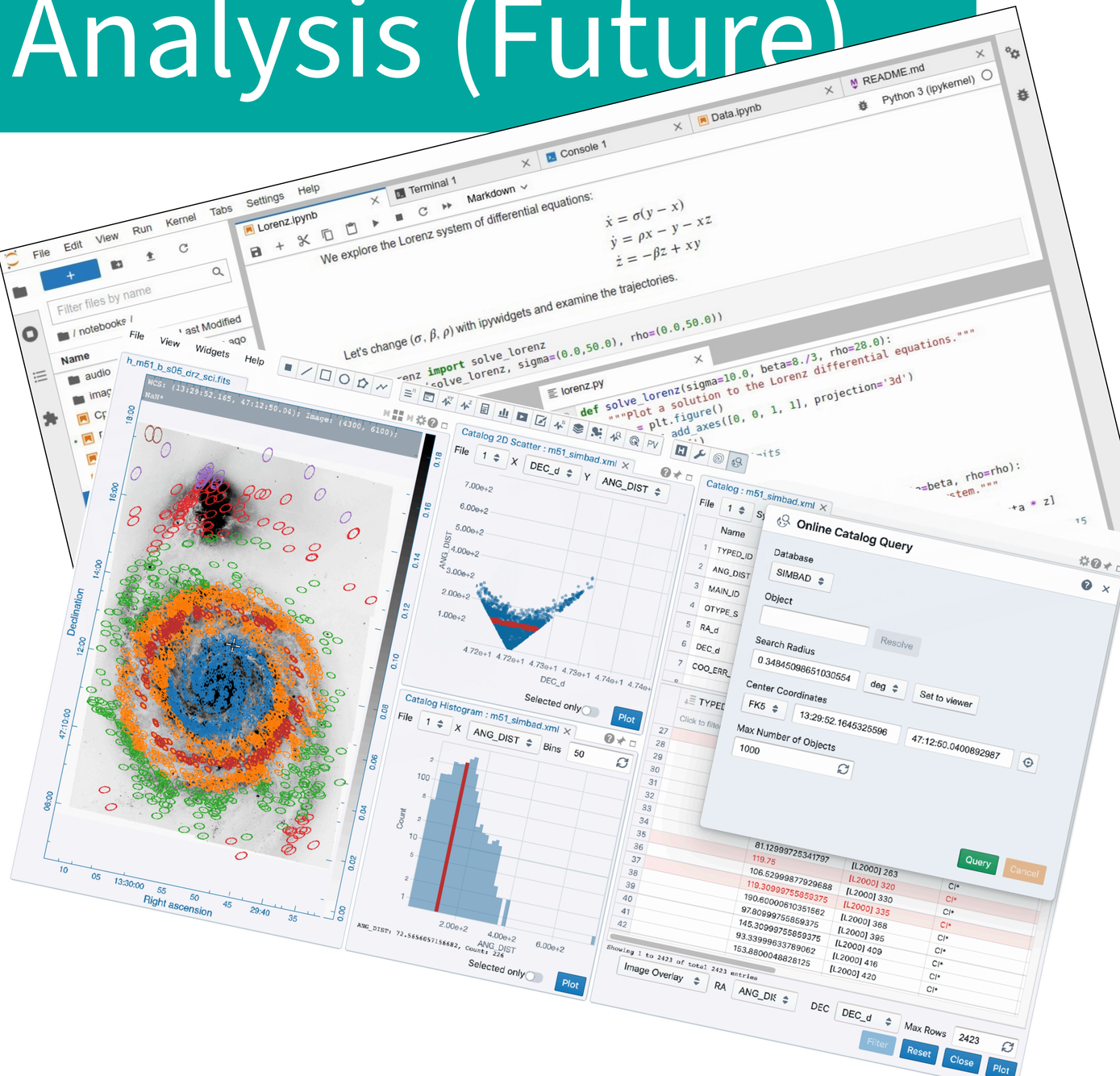


- For cost & science productivity reasons, processing plans/ pipelines must be in place before observations start.

- Not “ASTRON's pipelines”, but *community* pipelines; we work together to make them effective.

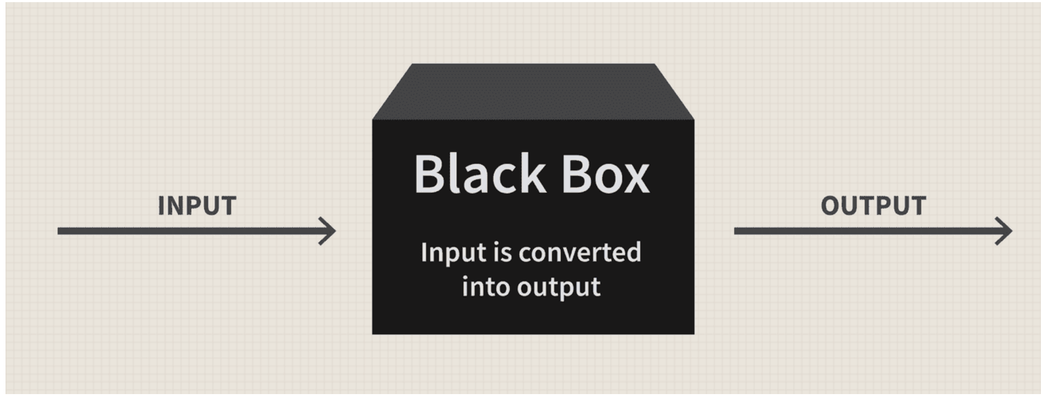
Interactive Data Analysis (Future)

- Common expectation: Jupyter notebooks running next to the data.
  - Jupyter notebook: interactive browser-based environment including live code, text, figures, etc.
- Also: “legacy” graphical applications (CASA, TOPCAT, ...), though e.g. VNC/ remote desktop connections.
- Also also: command line applications, through e.g. SSH.
- Implies the existence of shared, persistent storage for work in progress, output products, etc.



User Pipeline Execution (Future)

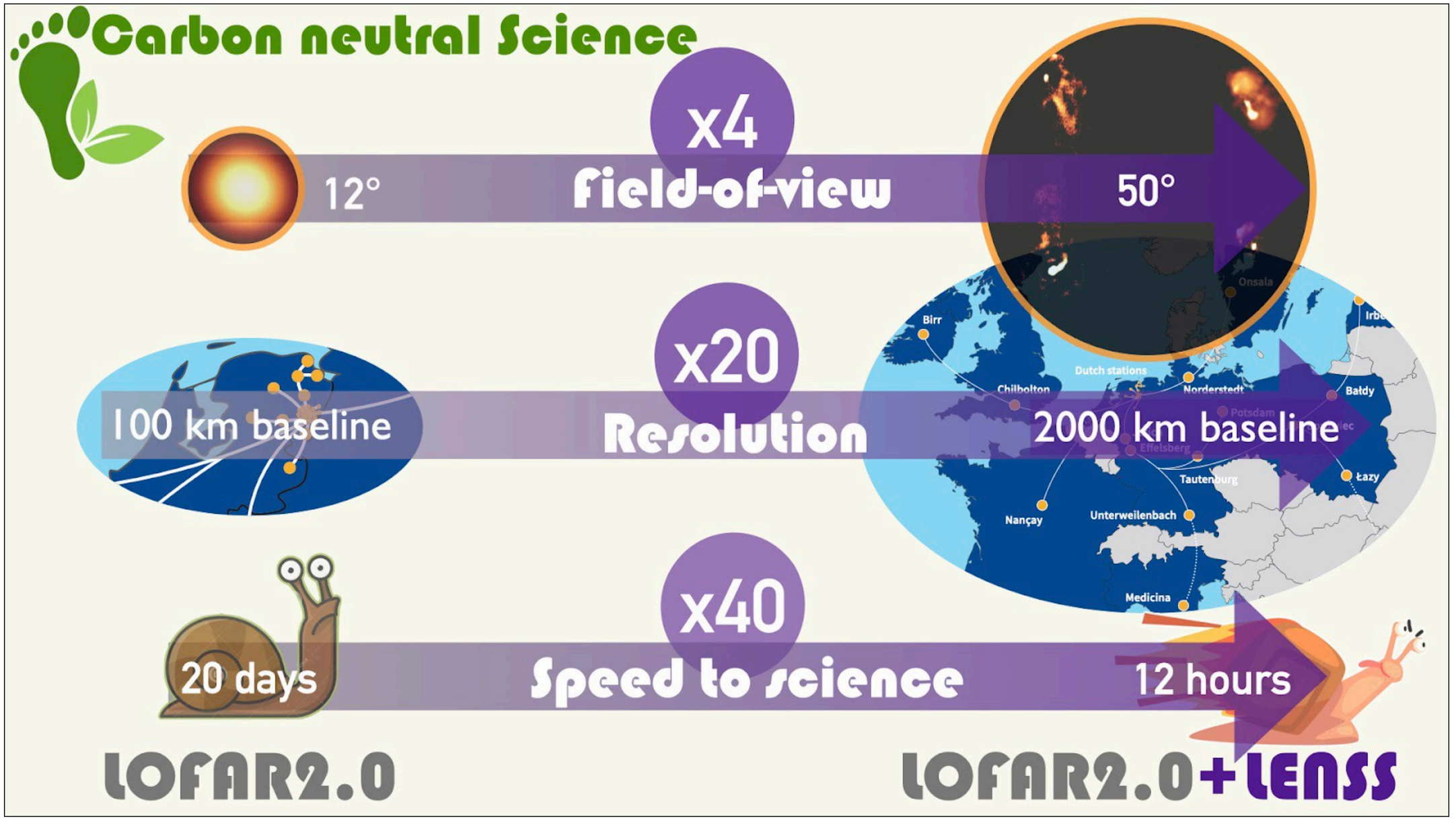
- Whatever observatory-supported pipelines are available , there will always be new science cases that aren't supported.
- Enable them, while minimizing risk to our operational system.
- Provide an API against which pipelines can be developed.
- Provide a “black box” system with appropriate quotas etc for executing untrusted payloads.

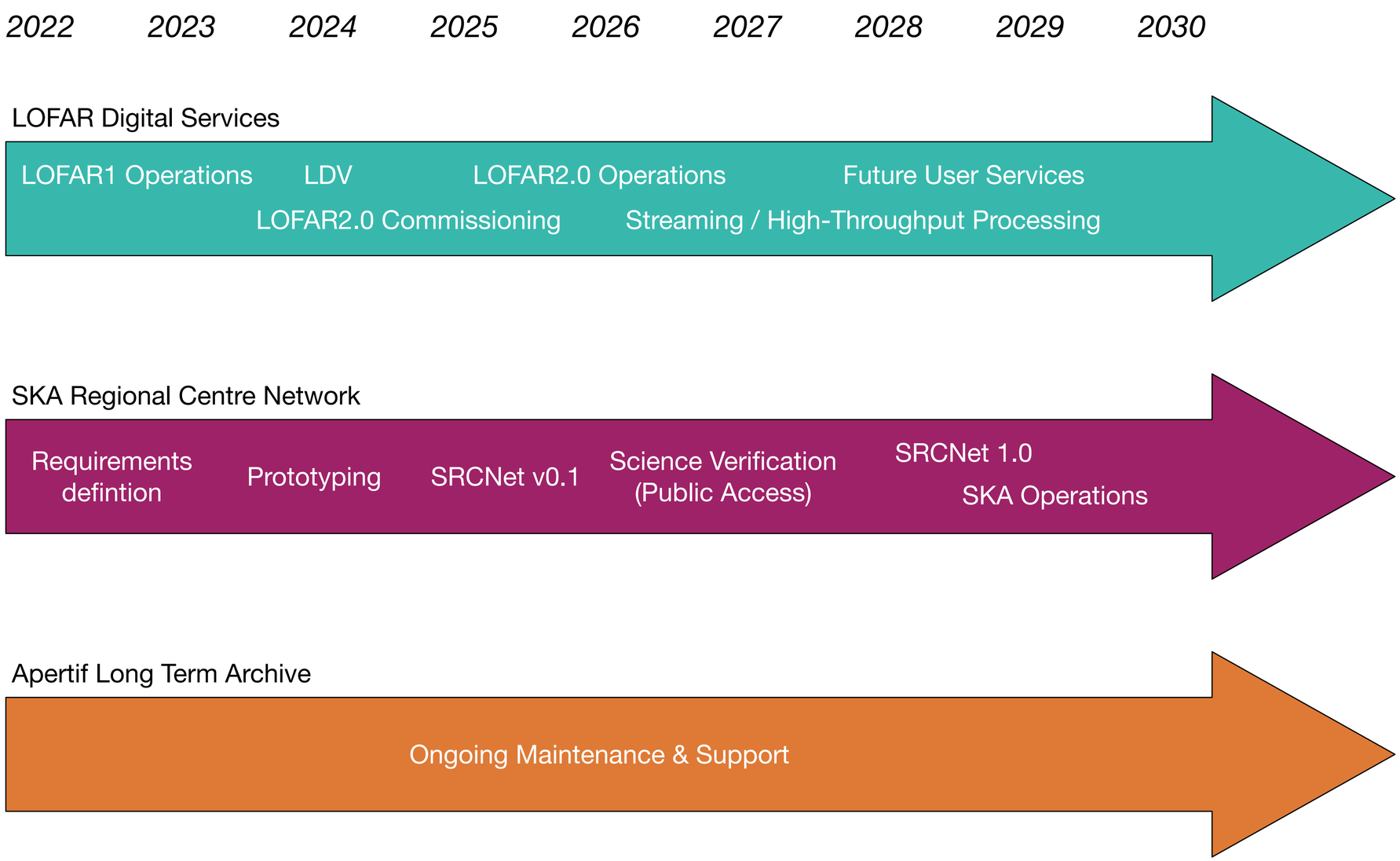


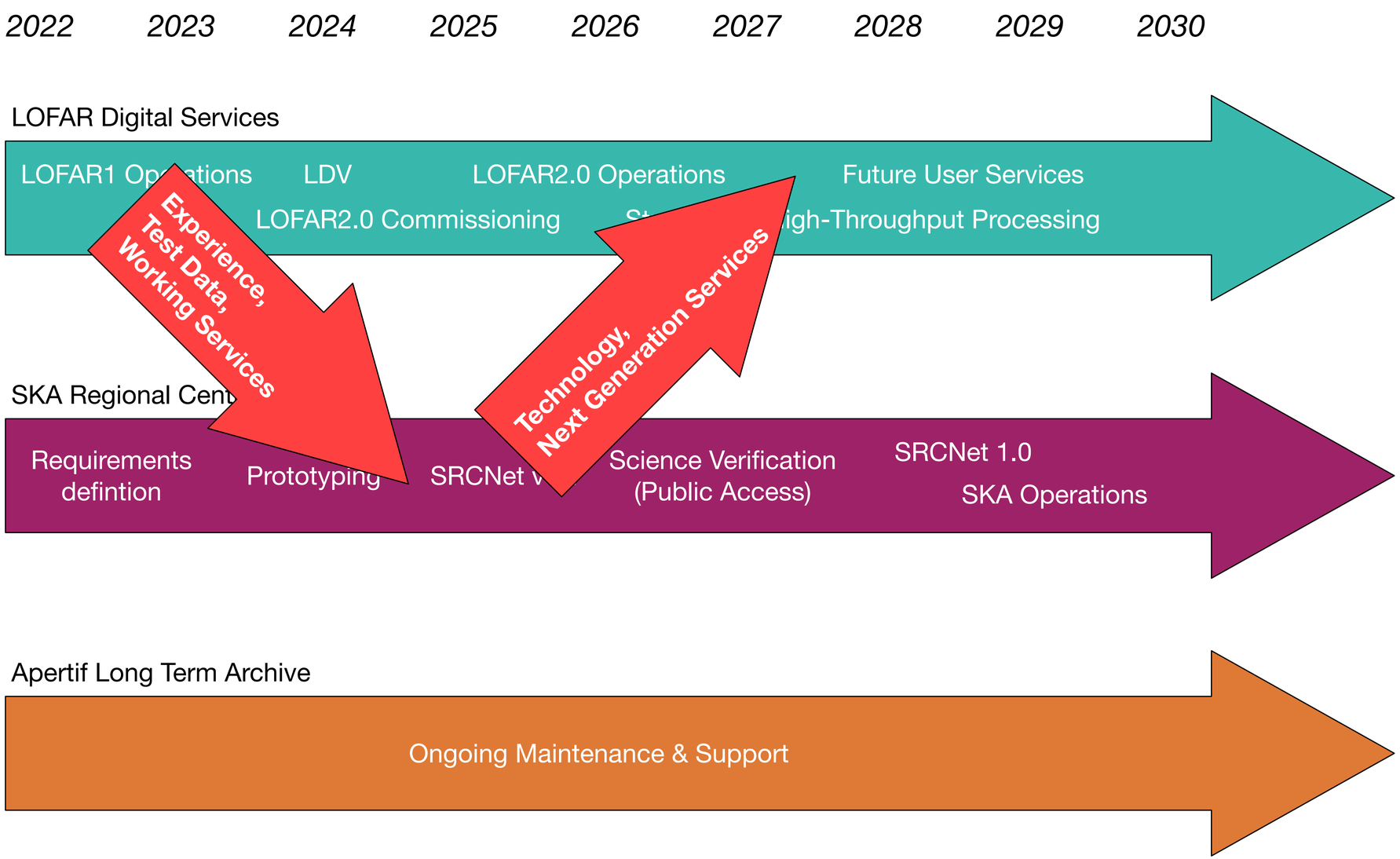
COMMON WORKFLOW LANGUAGE

nextflow

dask


Carbon neutral Science
x4
12°
Field-of-view
50°
x20
100 km baseline
Resolution
2000 km baseline
Birr
Onsala
Dutch stations
Chilbolton
Norderstedt
Potsdam
Effelsberg
Tautenburg
Nançay
Unterweilenbach
Medicina
Łazy
x40
20 days
Speed to science
12 hours
LOFAR2.0
LOFAR2.0+LENSS


2022
2023
2024
2025
2026
2027
2028
2029
2030
LOFAR Digital Services
LOFAR1 Operations
LDV
LOFAR2.0 Operations
Future User Services
LOFAR2.0 Commissioning
Streaming / High-Throughput Processing
SKA Regional Centre Network
Requirements defintion
Prototyping
SRCNet v0.1
Science Verification (Public Access)
SRCNet 1.0
SKA Operations
Apertif Long Term Archive
Ongoing Maintenance & Support


2022
2023
2024
2025
2026
2027
2028
2029
2030
LOFAR Digital Services
LOFAR1 Op ations
LDV
LOFAR2.0 Operations
Future User Services
LOFAR2.0 Commissioning
St
igh-Throughput Processing
Experience,
Test Data,
Working Services
Technology,
Next Generation Services
SKA Regional Cent
Requirements
defintion
Prototyping
SRCNet v
Science Verification
(Public Access)
SRCNet 1.0
SKA Operations
Apertif Long Term Archive
Ongoing Maintenance & Support

Conclusions

- The volume of data already produced by LOFAR is enormous...
- ...and it'll be even bigger in the LOFAR2.0 future.
- That presents us with challenges, both in terms of how we manage the data volume, but also in terms of the range of services we need to provide to end users.
- Addressing those challenges is something we can best address together, and we can also look to learn from parallel efforts in SKA.