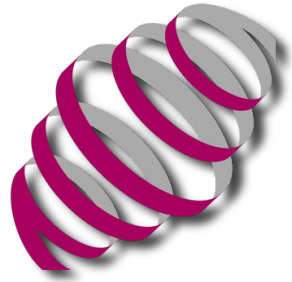




ASTRON

Netherlands Institute for Radio Astronomy

ASTRON



LOFAR

Technical Overview of the LTA

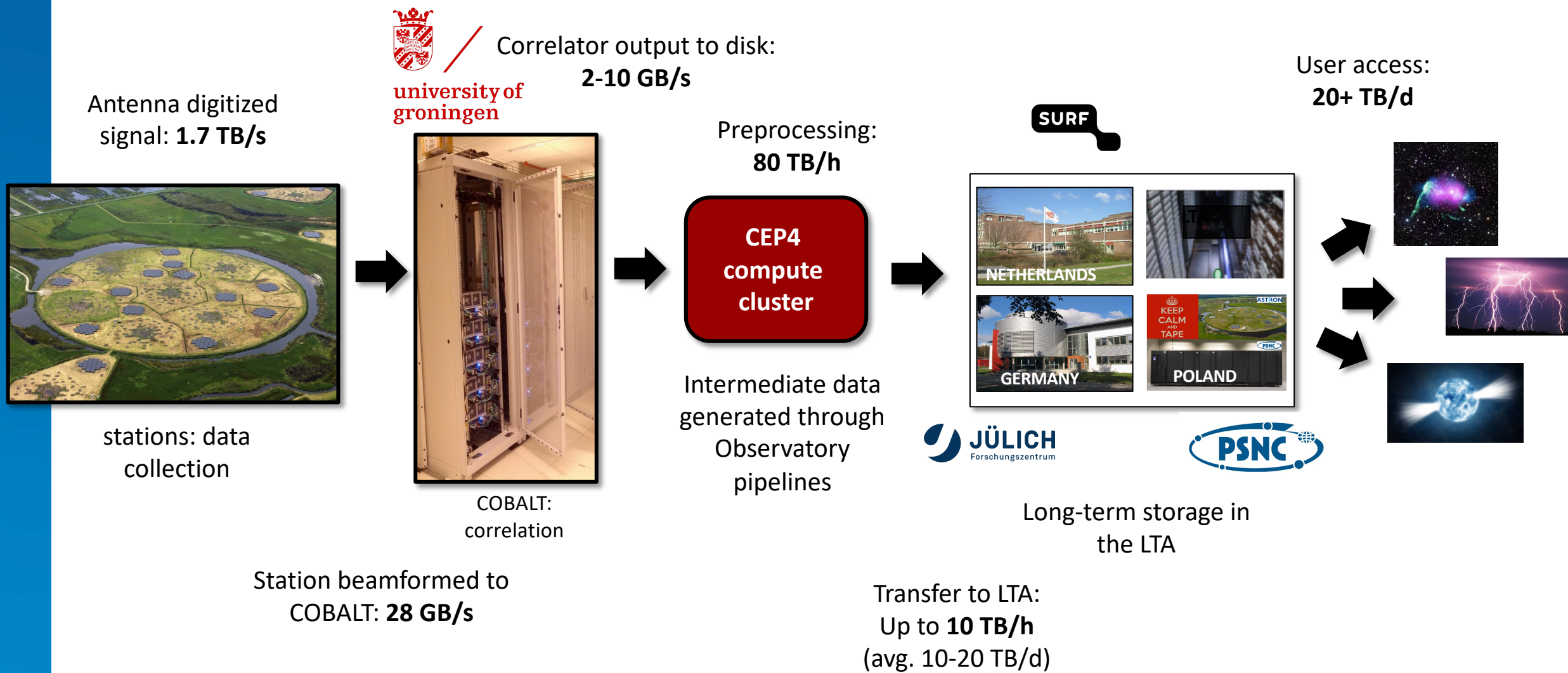
Hanno Holties

ASTRON

Netherlands Institute for Radio Astronomy



LOFAR1 Data Rate Reduction in stages



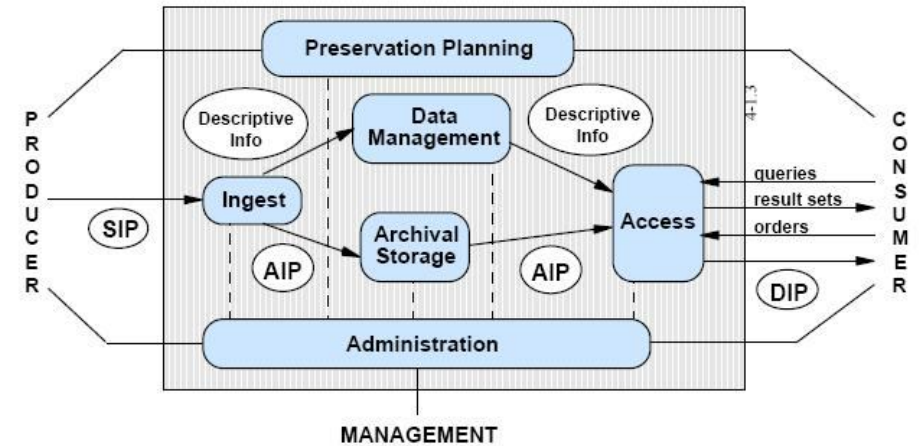
Distributed Research Infrastructure

- Central Processing
 - RUG (Groningen, NL)
 - (Near) real-time processing
 - 3 PB temporary storage (being replaced)

- Long-term archive
 - PSNC (Poznań, PL)
 - FZJ (Jülich, DE)
 - SURF (A'dam, NL)
 - 60+ PB nearline storage



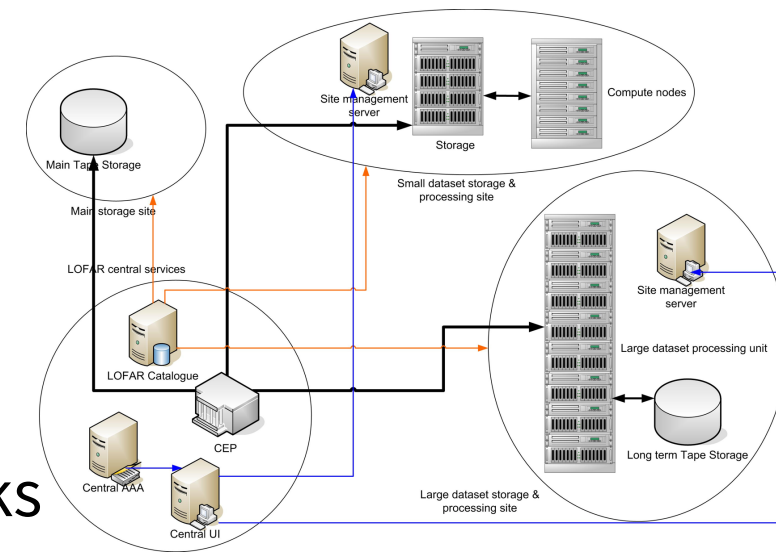
Early LTA design decisions



- **Follow OAIS** (Open Archival Information System) model
 - a.o. ingest data as a self-contained SIP (Submission Information Package)
- **Build on WLCG** (Worldwide LHC Computing Grid) technology for storage and compute
 - Use URLs (SRM-URLs) as ‘global unique identifiers’
- **Additional components:**
 - Single catalogue of LTA content based on AstroWise technology
 - HTTP based data access such that users are not required to have X509 certificates
 - Require users to stage data before retrieval through a central stageer service

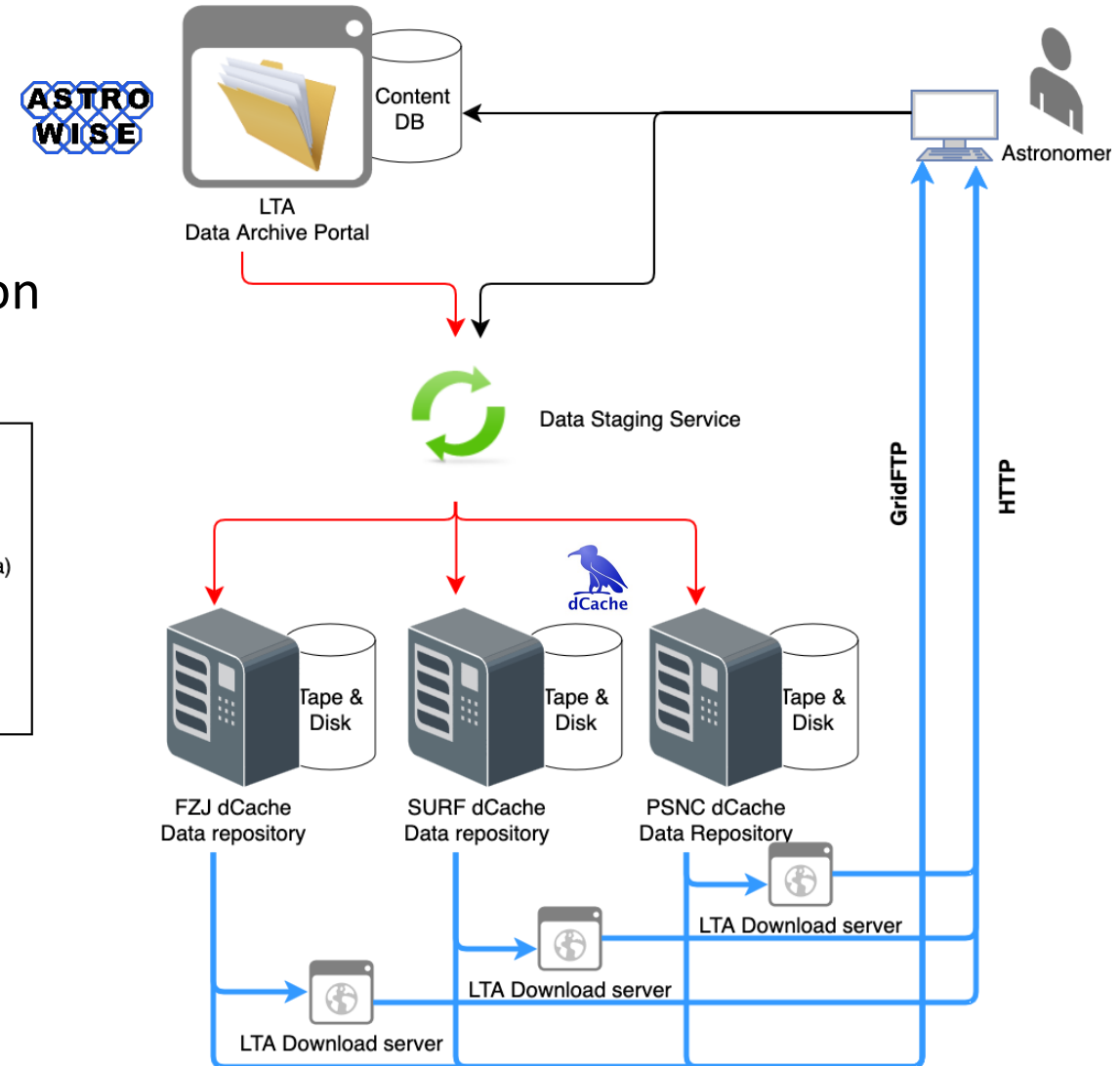
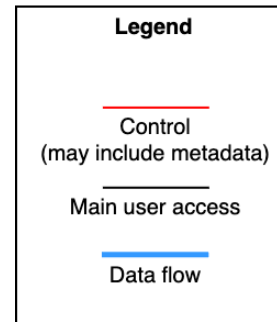
Early LTA design decisions

- **CEP – LTA** site connections as **exclusive networks** over dedicated (guaranteed bandwidth) 10 Gbps links
- Support for the **LOFAR data access policy**, specifically
 - Access restrictions to be applied in accordance with LOFAR **proprietary period**
 - Access control (authentication & authorization) based on LOFAR user administration (**users, projects, roles**)
 - Allow users to share access to data products
 - Implemented by providing ‘**security through obscurity**’ – SURLs include a random hash and regular users are not allowed to browse storage content
- **Data placement policy**
 - **Single replica** on tape (deviation from blueprint archive, which stated redundancy to be required)
 - Disk capacity for **caching ca 3 months** of LOFAR data



LTA – technology

- **dCache** (& StoRM/GPFS – retired 2016)
 - X509 + VOMS Authentication & Authorization
 - SRM & GridFTP protocol
 - *Single replica* on tape (no replication)
 - Disk for cache
 - Permanent disk - not used for archiving
- **Supporting services**
 - Catalogue
 - Ingest from CEP
 - Staging
 - Download servers



LTA user interfaces

- **LTA Portal** (discovery)
 - Supports requests to stage
- **LOFAR LTA Python client** (discovery)
 - Interact directly with catalogue (AstroWise)
- **Stager client** library (retrieval)
 - Wrapper around stager/Stagelt API
- **HTTP Download** servers (retrieval)
 - To be phased out
- **dCache** supported protocols (retrieval)
 - Legacy: SRM + GridFTP (X509)
 - Recent: WebDAV (X509 + Macaroons)
 - REST API – SURF ADA client

LOFAR Long Term Archive

HOME SEARCH DATA BROWSE PROJECTS HELP

LC1_027

Observation 1 to 100 (showing 100 of total 387) -

Averaging Pipeline (total 0) -

Calibration Pipeline (total 0) -

Imaging Pipeline (total 0) -

Long Baseline Pipeline (total 0) -

Pulsar Pipeline 1 to 100 (showing 100 of total 387) -

#	Project	Release Date	Pipeline Name	Pipeline Version	SAS Id	Pulsar Selection	doSinglePulseAnalysis	Strategy Name	convertRawTo8bit [s]	subintegrationLength	Source DataProduct	All Dataproducts	Quality	Pulsars
1	LC1_027	2015-05-15	31544+4937/PULP	n/a	1027091	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	High	0
2	LC1_027	2015-05-15	B1237+25/PULP	n/a	1027069	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	High	0
3	LC1_027	2015-05-15	B1133+16/PULP	n/a	1027047	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	Moderate	0
4	LC1_027											show		0

Staging Service

The following 1 file(s) will be requested for download, with a total size of 6.4 GB.

Cancel Submit

Size	MD5 (checksum)	Filename
6.4 GB	19528471109c3305a6f199875	L10807_35040_uv300.ms_0194373r
6.4 GB		Your choice

```
# Python3 code
from pprint import pprint
from awlofar.main.aweimports import Observation, Pointing, SubArrayPointing
from common.database.Context import context
result = {}
for project in sorted(context.get_projects()):
    print("Project %(project)s" % vars())
    ok = context.set_project(project)
    # do your query
    obs_ids = set()
    query = (Pointing.rightAscension > 95) & \
            (Pointing.rightAscension < 105) & \
            (Pointing.declination > 20) & \
            (Pointing.declination < 30)
    print("Total Pointings %d" % len(query))
    for pointing in query:
        print("Pointing found RA %f DEC %f" % (pointing.rightAscension, pointing.declination))
        query_subarr = SubArrayPointing.pointing == pointing
        for subarr in query_subarr:
            query_obs = Observation.subArrayPointings.contains(subarr)
            for obs in query_obs:
                obs_ids.add(obs.observationId)
    result[project] = sorted(list(obs_ids))
print(result[project])
```

```
wget -ci html.txt
```

https://webdav.grid.surf.sara.nl:2882/pnfs/grid.sara.nl/data/lofar/ops/projects?authz=MDAxY2xvY2F0aW9uE9wdGlvbmFslUwVtChR5CjAwM

lc4_003	Wed Oct 14 19:49:43 CEST 2015
lc2_026	Fri Mar 10 13:04:57 CET 2023
lc3_034	Wed Mar 04 11:50:43 CET 2015
lc2_006	Wed Jul 16 12:10:30 CEST 2014
lc4_011	Wed Jun 24 11:39:53 CET 2015
lc5_010	Mon Nov 16 22:18:25 CET 2015
lc5_001	Fri Jan 29 20:31:12 CET 2016

LOFAR1 LTA – the cold numbers

- **LTA** - LOFAR Long-Term Archive

- Instrument data archive

- **Content**

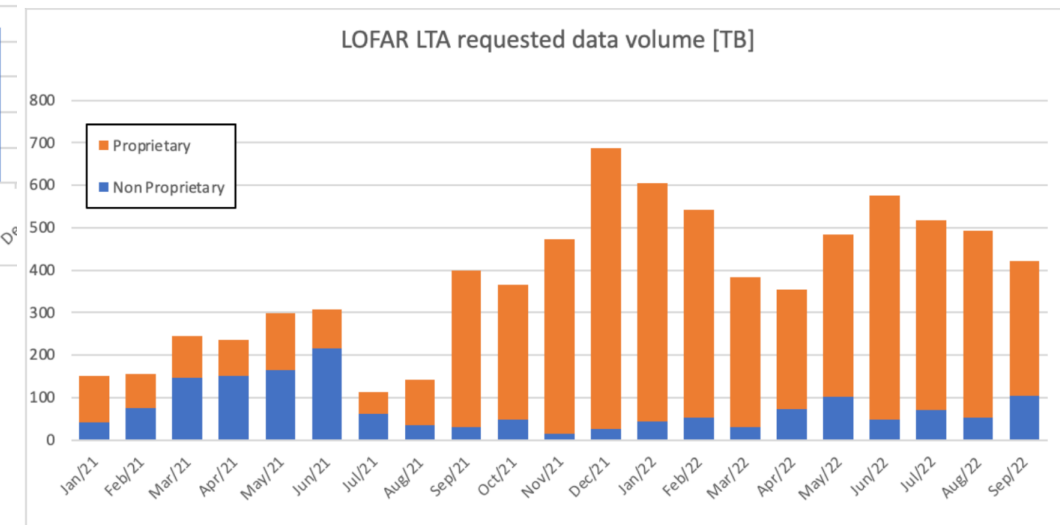
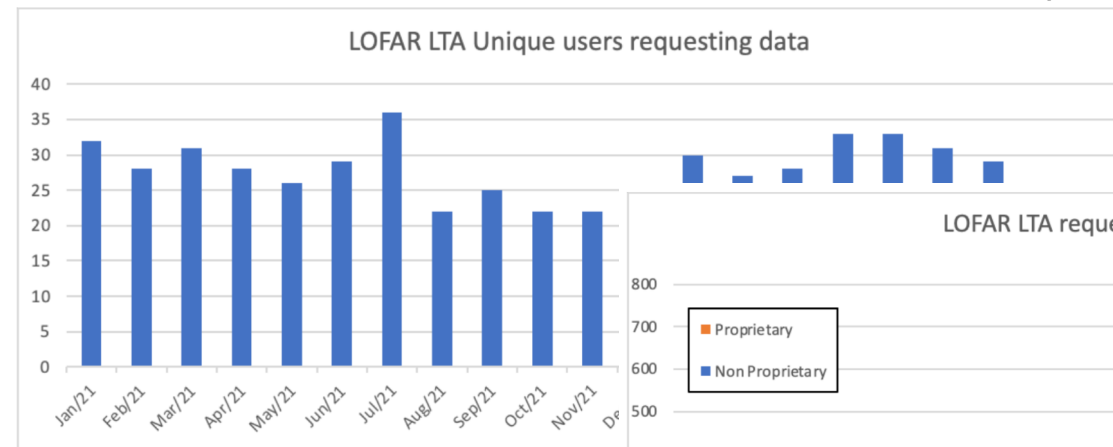
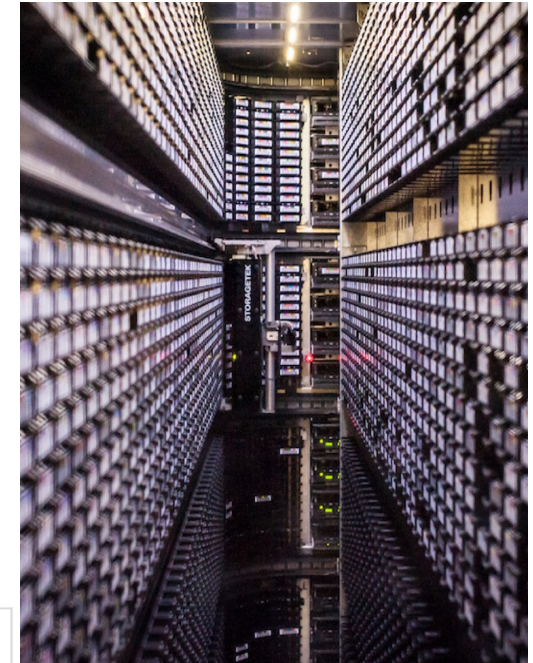
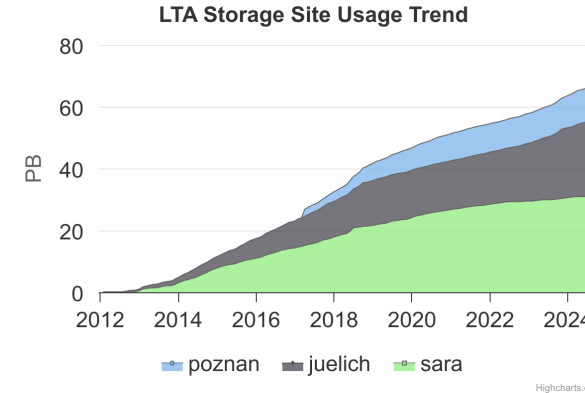
- 60+ PB
- 61k Data collections
- 15.7M Data Products

- **Distribution**

- SURF (2010)
- RUG (2012-2016)
- FZ-Jülich (2013)
- PSNC (2016)

- **Access statistics**

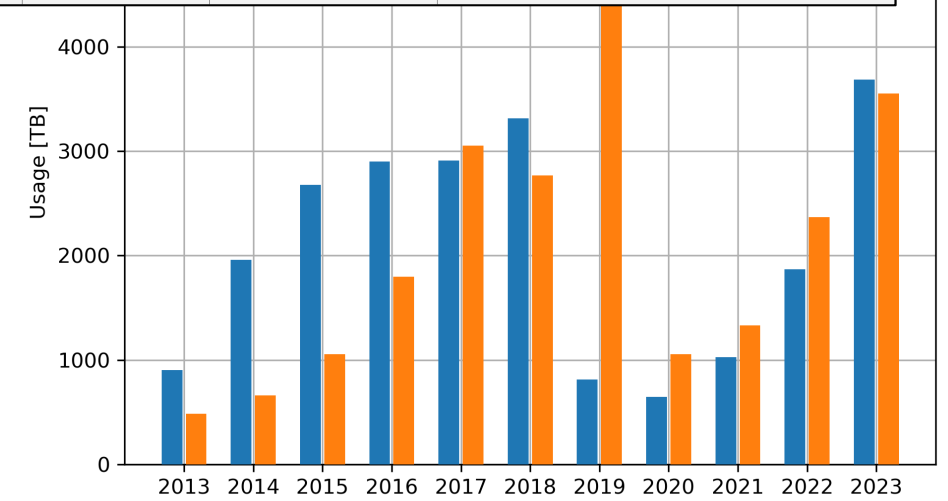
- Ca 30 Unique users/month requesting data
 - Often representing project teams
- Ca 400 TB/month requested



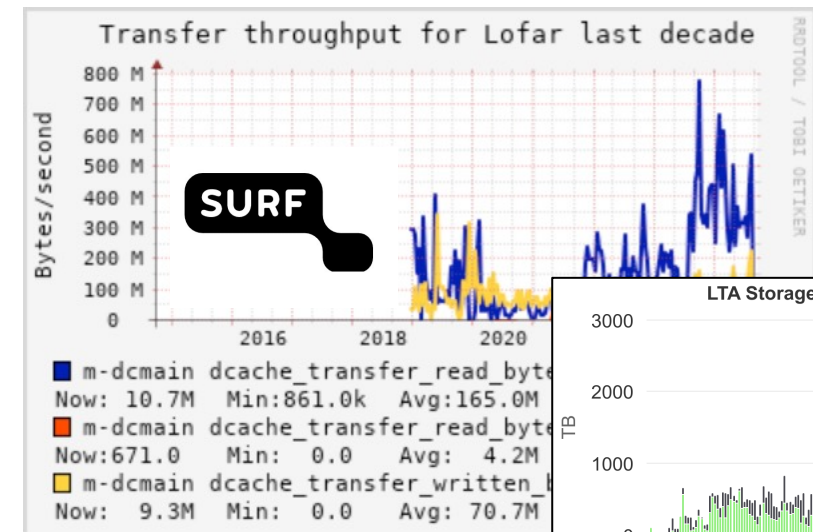
LTA monitoring

- Storage and access varies strongly
- For the Observatory, LTA site behaviour is mostly a **'black box'**
 - CEP-LTA network Bandwidth/Volumes
 - Sites expose some dCache metrics
 - No common API (?) - SURF through Ganglia
 - Access metrics through staging requests
- Most common **issues not easy to identify**
 - Expiring certificates
 - Problems fetching data from tape
 - Download server – dCache load issues
- **Content 'scraping'** for verification
 - Slow

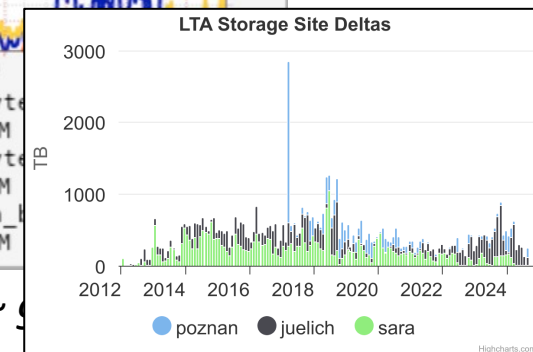
CellName	DomainName	Total Space/MiB	Free Space/MiB	Precious Space/MiB	Layout (preious/sticky/cached/free)
pool01	poolDomain01	94371840	13720	0	
pool02	poolDomain02	94371840	2195	0	
pool03	poolDomain03	94371840	2573	0	
pool04	poolDomain04	94371840	1748	0	



C. Manzano et al., Learning from the present for the future... (2024)



300 MB/s ~ 26 TB/day ~ 9



LTA Some Lessons Learned

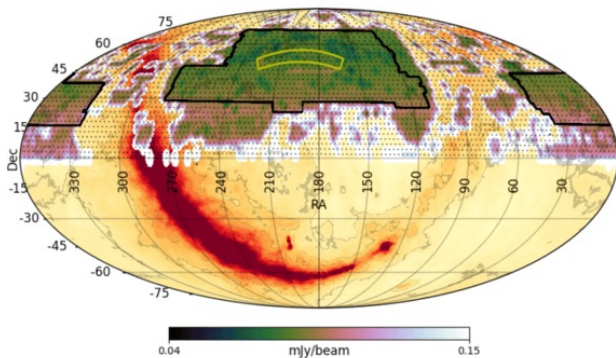
- **Tape** still is a cost effective medium for large data
- Keep **technology in sync** across sites
- Need for **full chain monitoring**
 - Identification of issues
 - Understand usage and relevance
 - Ability to respond to (changing) patterns
- **Instrumental data** is **write once, read once** (or few)
 - Future focus on storage & hosting of advanced data products
 - Will see patterns & data protection needs & data retirement
- Setting up **operating large scale data processing is not trivial**
 - Significant effort and time

See talk Manu later this afternoon

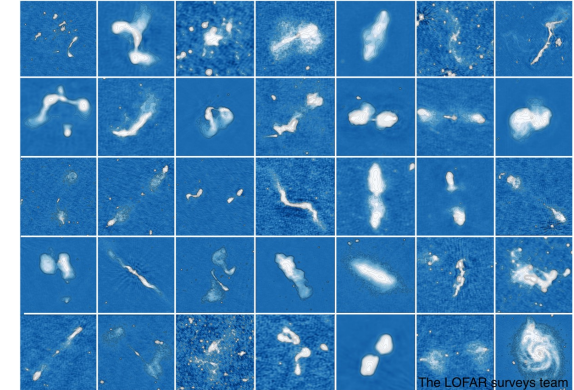
LTA Current developments

- Evolving technology

- All sites now on **dCache 9.2**
- **Stagelt** (Python) replacing old stager (JAVA)
 - Support dCache **WebDAV** protocol – in time phase out SRM/GsiFTP?
 - Support dCache **Macaroon tokens** – in time phase out X509?
 - Based on gfal instead of SRM client - option to move to dCache REST API
- Decommissioning the LOFAR download servers
 - In time phase out LDAP (and custom user administration) in favour of **FAAI/SRAM**
- Enhanced support for **advanced data products** (ADPs)
 - **IVOA** standards based access (**GAVO DaCHS** plus **ADEX**)
 - (Expect) more *intensive data access* patterns
 - Consider *data placement policies* and *new/other data repository* types



LOFAR1 Advanced data: LoTSS DR2



[Shimwell, T.W., et al. "The LOFAR Two-metre Sky Survey. V. Second data release", 2022, A&A, 659, 27](#)

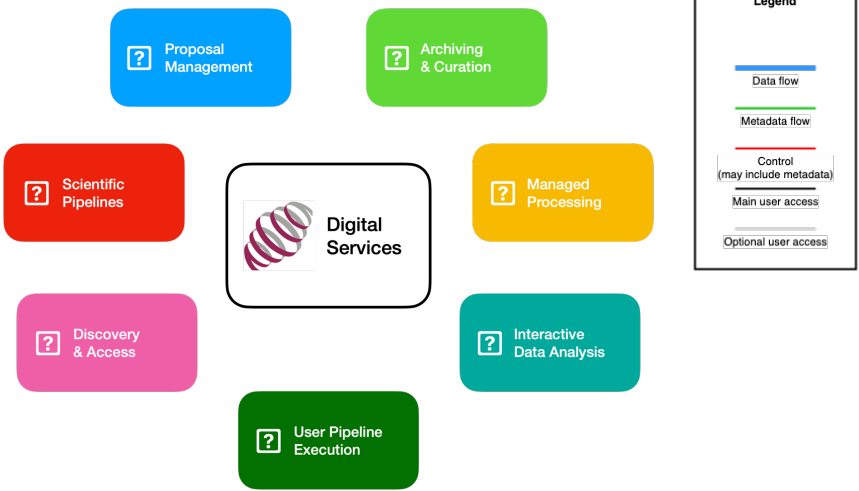
Type	Total size
Mosaiced astrometric corrected Stokes I 6'' and 20'' resolution images	1.5 TB
Individual Stokes I 6'' and 20'' (with and without astrometric correction) resolution restored images with associated DDFacet model, residual and mask images	7.5 TB
Catalogue of the 4,395,448 radio sources and the 5,120,353 Gaussian components that describe them	2 GB
Stokes I 6'' and 20'' HIPS* files	2 TB
Individual observation Stokes QU 20' and 3' resolution undeconvolved 480-plane image cubes	34 TB
Individual observation Stokes V 20'' resolution 120-168MHZ continuum dirty images	400 GB
DI calibrated visibilities and DD calibration solution, facet layout and astrometric corrections	100 TB
Raw data in the LTA	12 PB

ADP's hosted on SURF dCache, accessible via SURF Data Repository and vo.astron.nl

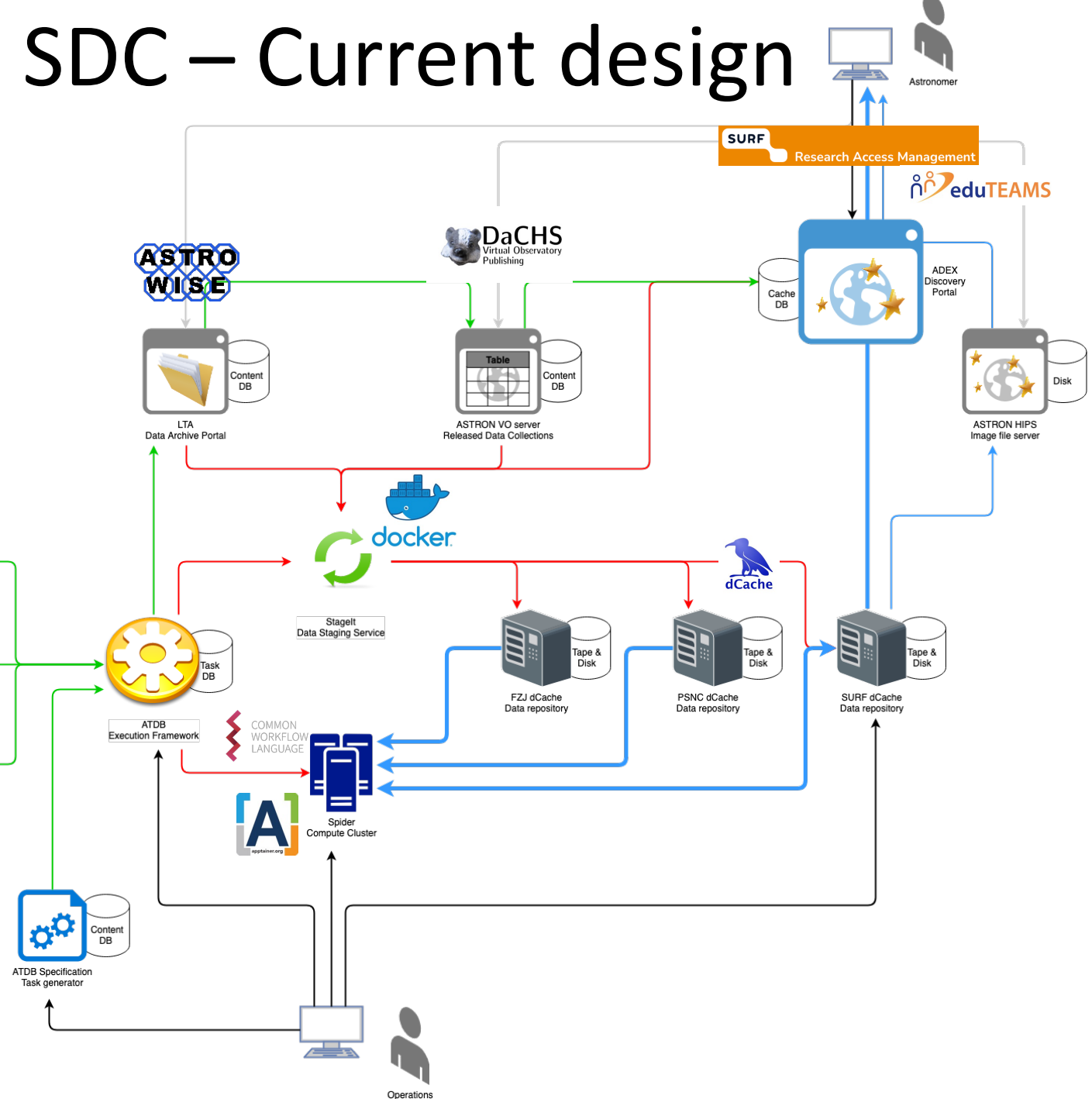
LTA Current developments

- Integrated/managed data processing

- Common Workflow Language (**CWL**) pipeline executors
- **Apptainer/Singularity** based software deployment
 - TBD: caching vs distribution
- Microservice based **framework**
 - **Centralized**
 - Central task database (process 'management')
 - Task specification
 - Stager (Prepare input data for access)
 - Archiver (Ingest output in LTA: catalogue update & move to permanent storage)
 - **Distributed** (running in data center – communicating with central task database)
 - Executor (job submission & monitoring)
 - Datamanager (move data between storage infrastructures)
- *More details tomorrow*



SDC – Current design

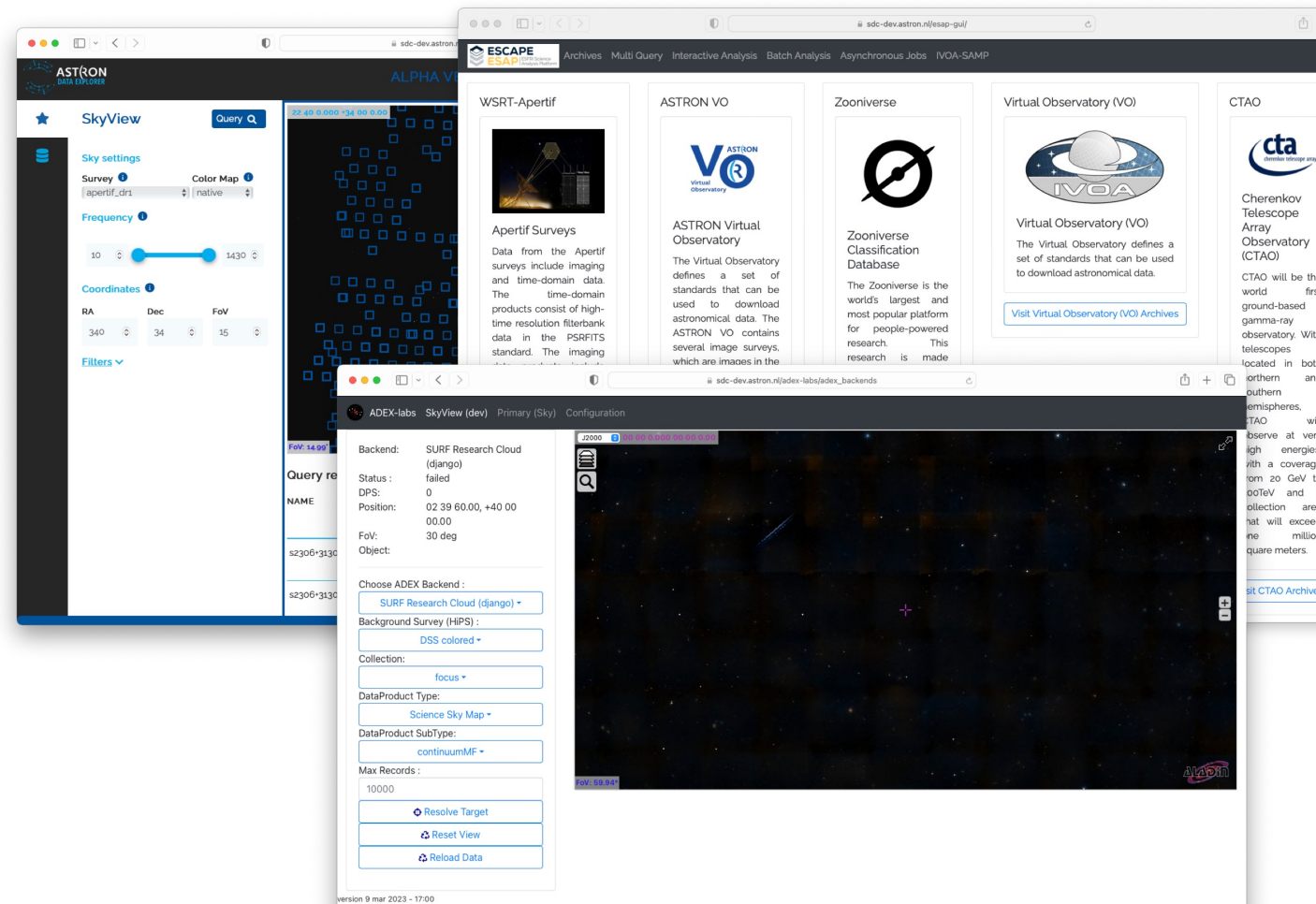


• Core technology

- **Gitlab** code repository & CI/CD
- SURF **SRAM** / eduTEAMS FAAL provider
- **CWL** (Common Workflow Language) Pipeline definitions
- **Apptainer** (formerly Singularity) & **Docker** container technology
- **dCache** tiered storage data management
- **AstroWise** LTA Catalogue & Portal
- **GAVO DaCHS** VO standards service

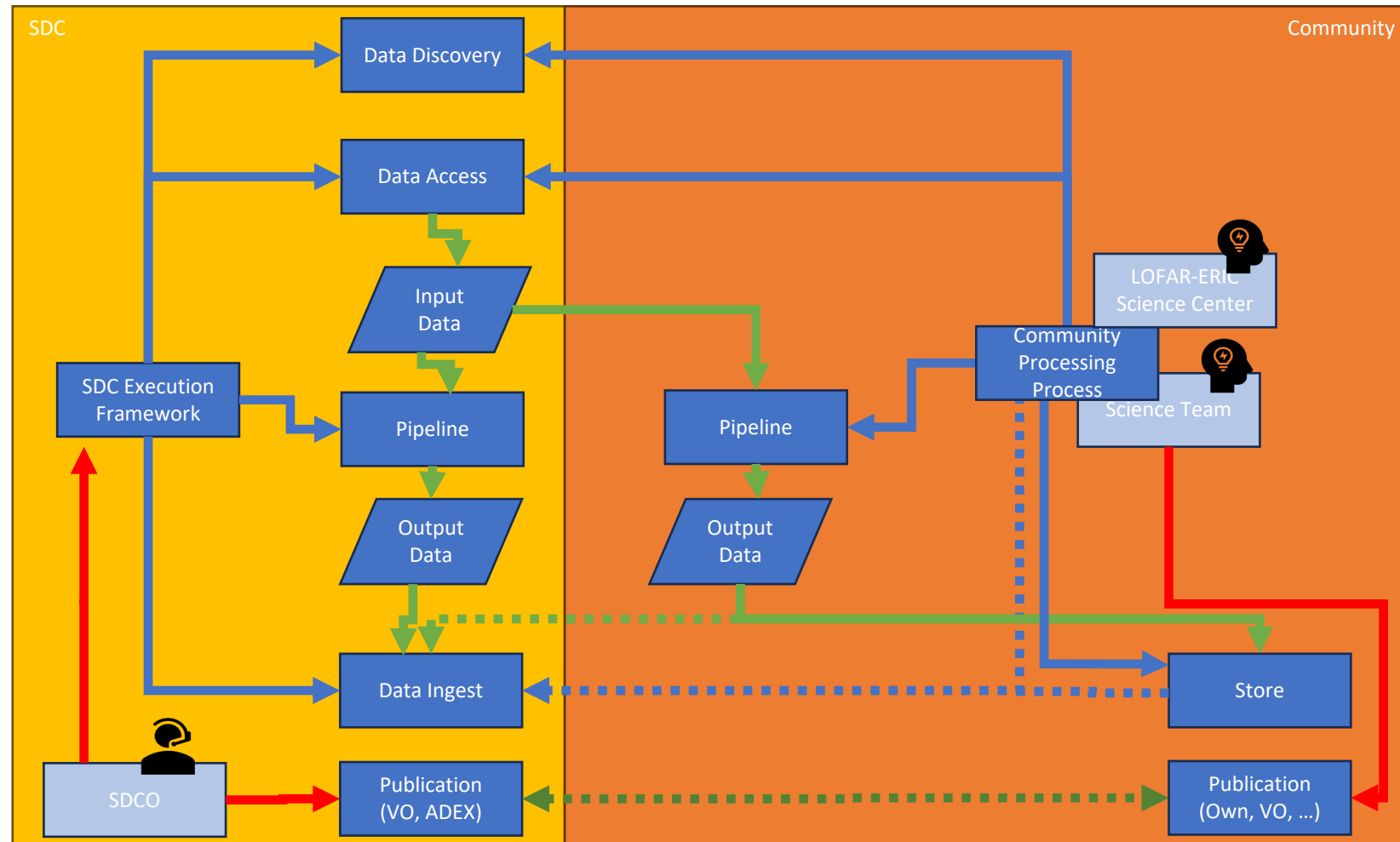
Discovery & Access - new user interfaces

- Shifting focus to **advanced data products** (both 'standard' and community generated)
- All data releases available through **Virtual Observatory** interfaces
- New, friendly, accessible, modernized, **data discovery portal**
- Ambition is for the archive to be fully **FAIR-compliant**
- Move to FAAI & Token based access



User data processing

- There will be **science cases that are not supported by standard pipelines**
- **Provide API** to develop against
 - Discovery
 - Access/retrieval
 - User ingest
 - ICDs
 - Data description
 - Ingest service
 - Quality assurance



LOFAR2.0 LTA Preliminary Prognosis Data Traffic

Prognosis ¹	Intermediate (temporary)	Advanced (permanent)	Calibrated (permanent)	Total Scenario A (optimistic; ~ILT policy)	Total Scenario B (pessimistic)
Write [PB/yr]	21	1.5	15	22.5 PB/yr	37.5 PB/yr
Distribution factor	1	2	1	24 PB/yr	39 PB/yr
Read factor	1.1	5	2	30.6 PB/yr	60.6 PB/yr
Total Write [Gbps] ²	5.3	0.8	3.8	6.1 Gbps	9.9 Gbps
Total Read [Gbps] ²	5.9	1.9	7.6	7.8 Gbps	15.4 Gbps
<i>L2LP Requested [PB/yr]³</i>	<i>14.3</i>	<i>6.3</i>	-	<i>6.8 Gbps (W)</i>	-

¹ Indicative only; volumes and read/write factors for LOFAR2 are not yet known/fixed

² Full year average sustained; each LTA site should support the overall av. sust. rate for prolonged periods (days)

³ Total requested in the call for LOFAR2 Large Programs, averaged over 5 years; not yet allocated and excluding open calls

The End (for now)

