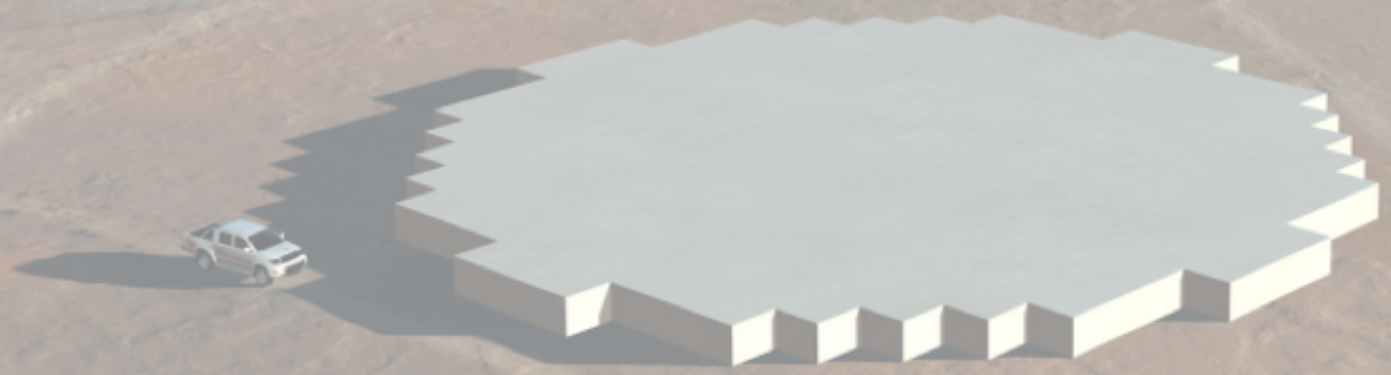
LOFAR, SKA
ASTRON contribution to Obelics WP3.3

Tammo Jan Dijkema & Sebastiaan van der Tol

Obelics F2F, Rome, 26 January 2016



ASTRON is part of the Netherlands Organisation for Scientific Research (NWO)

LOFAR stations

Processing overview

ASTRON

Antennas / Stations

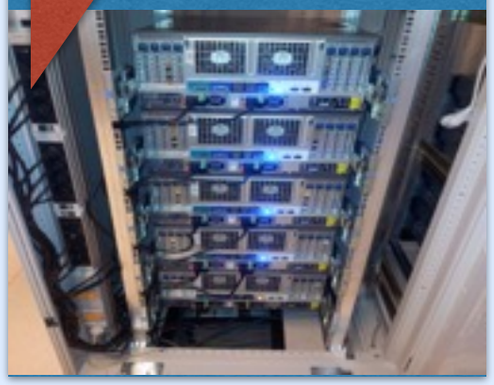
on-site reduction

200 Gbit/s

Realtime System

8 node GPU correlator

80 Gbit/s



Offline Processing

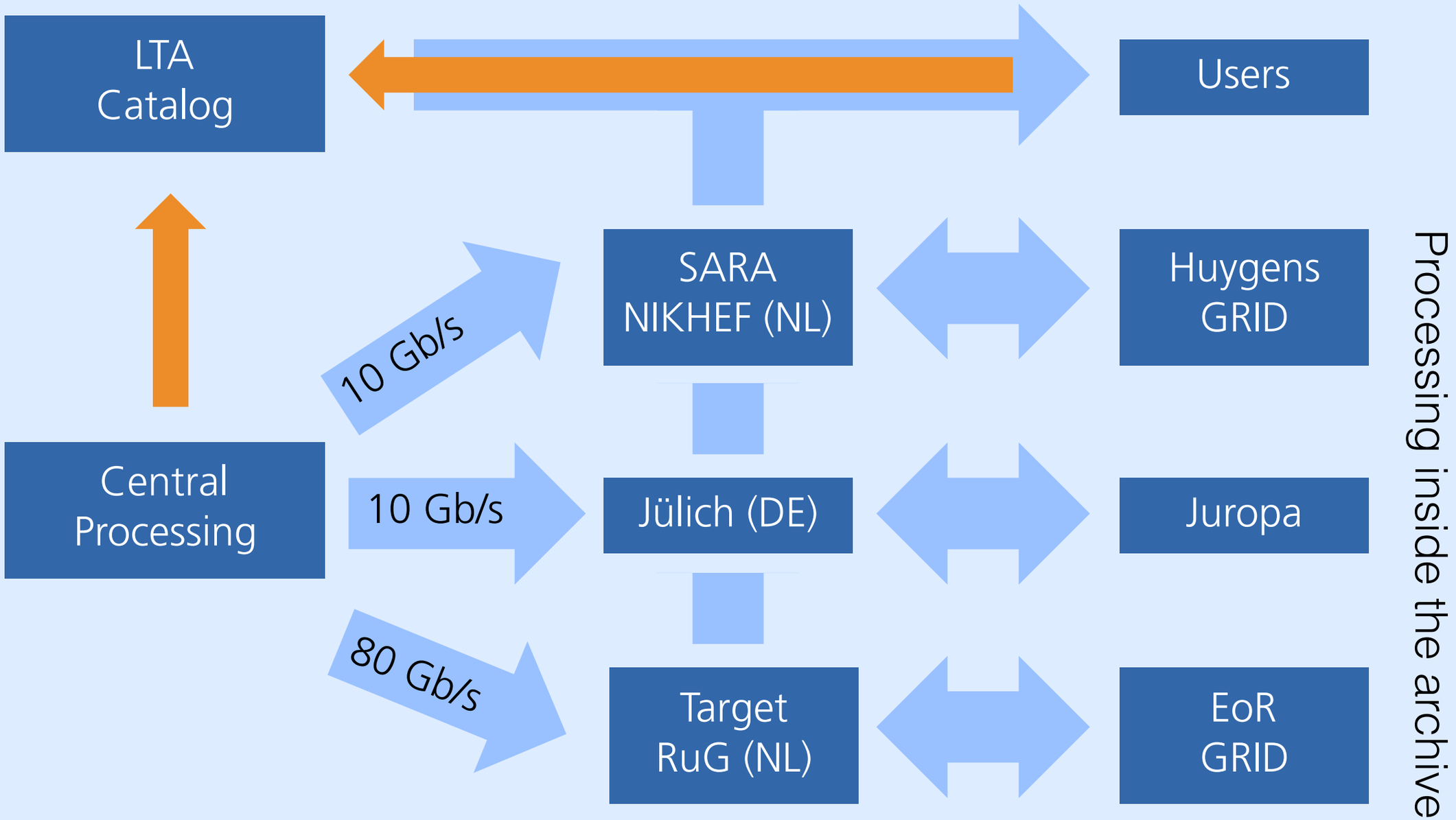
~ 100 node CPU cluster

2 Gbit/s

Long Term Archive

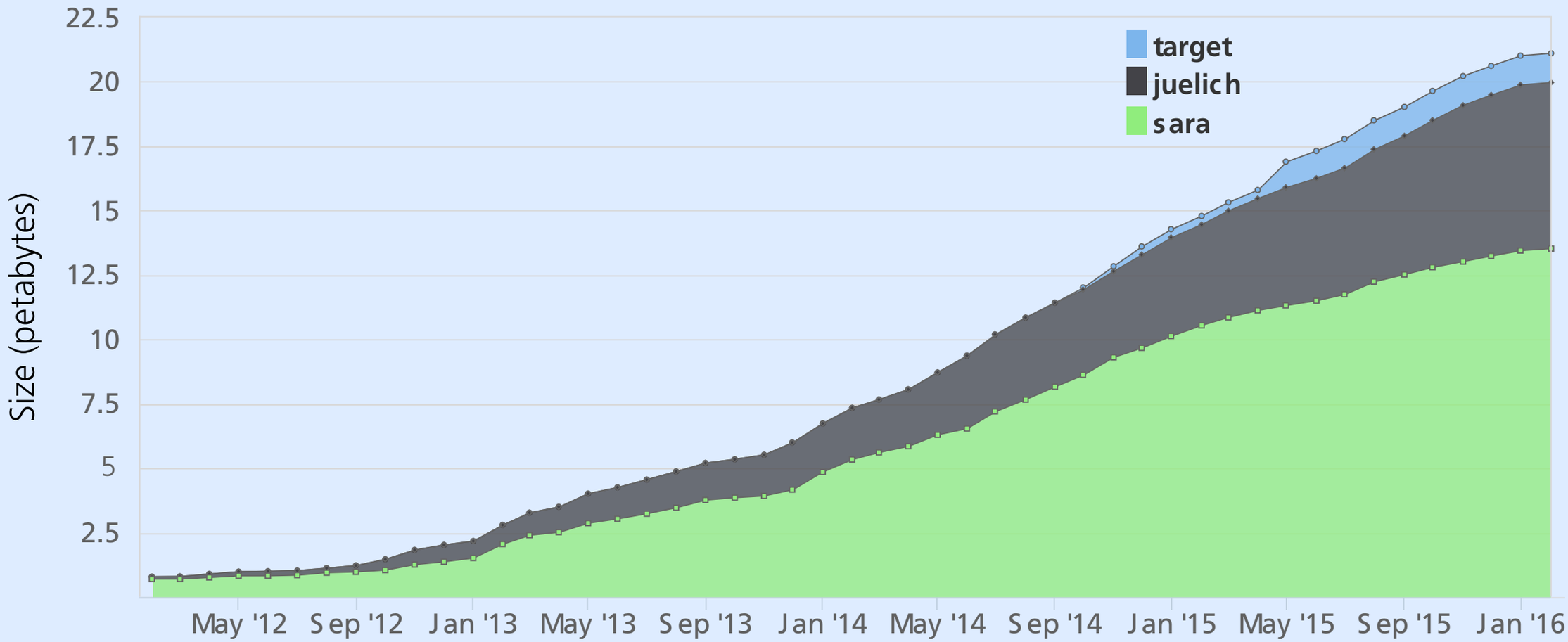
4 grid sites ~20 Petabyte

LOFAR Long Term Archive



Current LTA status

- 5 million dataproducts
- 800 million files
- 400 TB per month archived
- 60 TB per month retrieved



Challenges: LOFAR archive

- Design Complexity
- Data Model, provenance
- Procedures
- Databases, queries
- Network
- Authentication, ownership
- User interfaces, documentation

Challenges: infrastructure

- Network connections
- Security on high bandwidth
- Storage/Retrieval parallelization, previews
- Robustness, redundancy
- Multi node service scaling
- Inter-institute coordination
- Monitoring, alarms
- VM’s, cluster file systems, RAID

LOFAR LTA experiences

Operations:

- Even simple becomes elaborate: **`ls`**, **`cp`** break
- Limits of technology: RAID, disk/network speed
- Not all problems can be corrected
- Risk of scope creep due to evolving user requirements

User:

- It might take a month downloading data
- It might take a day to get a file from tape
- The files might not fit on their computer
- Using a (GRID) cluster: jobs, scratch, debugging, ui
- Using parallel file transfer
- Sharing bandwidth, computing systems

LOFAR Pipeline software

ASTRON

- Now: custom pipeline software
  - Tailored to our (previous) cluster (distributed file system)
  - Distributes jobs over ssh

- Plan 2016: move to Unicore (or Pegasus or Taverna)
  - Good abstraction between pipeline, jobs and specs
  - Grid/cloud support built in
  - Writing + visualizing pipelines becomes easier

Pipeline challenges

- User-contributed tools (CASA, python scripts, …)
- Environments
  - Grid sites have different linux distributions
  - Distributed file systems
- Specification
  - User friendliness
  - GUI / myexperiment?
- Characterization
- Scheduling
- Monitoring

Work package 3, Task 3.3
Data systems Integration

Asterics
Astronomy ESFRI & Research Infrastructure Cluster

WP3.3 organization

- Please register for an account on asterics2020.eu: send mail to cimo@jive.eu
- Track progress (meeting notes, etc.) on internal wiki: https://www.asterics2020.eu/dokuwiki/doku.php?id=intra:wp3:task3.3
- Discuss technical matters on forum (open to public!): https://www.asterics2020.eu/phpbb/

Task 3.3 D-INT: Data systems INTegration

Partner	**LAPP**	**ASTRON**	INAF	UCAM	UCM	IFAE	FAU	INFN
Effort (PM)	72	48	48	48	8	14	36	6

WP3.3 deliverables

Deliverables

Nr	Description	Task	Month	
D3.1	Detailed WP3 Project plan	3.1	4	
D3.2, 3.6, 3.10	Annual user engagement forum, workshops and training events	3.1	12, 24, 36	
D3.3	Analysis Report on Standards and Libraries	3.2	12	
D3.4, 3.17	Release of Software Libraries	3.4	12, 48	
D3.5	Analysis Report on Resource Requirements	3.3	18	**Nov 2016**
D3.7, 3.15	Processing Platform Technology Benchmark Report	3.2	24, 48	
D3.8, 3.16	Database Technology Benchmark Report	3.3	24, 48	**May 2017**
D3.9	Statistical Solvers Technology Benchmark Report	3.4	24	
D3.11	Analysis Report on Frameworks and Architectures	3.2	36	
D3.12	Repository of Services	3.3	36	**May 2018**
D3.13	Repository of WMS Services	3.4	36	
D3.14	Final Integral WP3 Report	3.1	48	