ASTRON

Netherlands Institute for Radio Astronomy

The APERTIF Long Term Archive
Or: how to serve a dozen dishes

ALTA, ASTRON, 2017/06/14

Hanno Holties, ASTRON
Roy de Goei, ASTRON
Gijs Noorlander, KxA
Erwin Platen, S[&]T
Nico Vermaas, ASTRON

Menu

APERTIF
APERture Tile In Focus

ASTRON

- Radio Astronomy & APERTIF
- APERTIF Long Term Archive (ALTA)
- ALTA & iRODS
- Summary

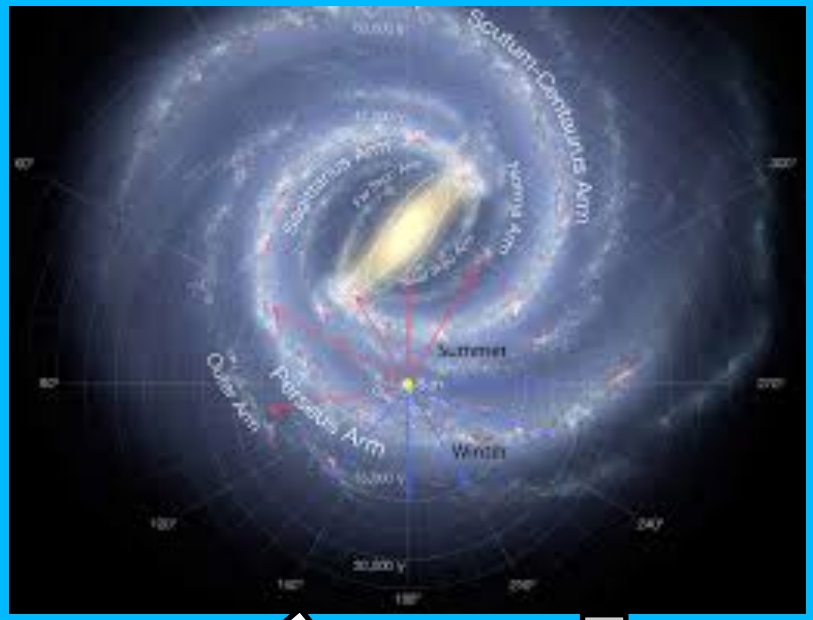
Astronomy I (Optical)

APERTIF
APERture Tile In Focus

ASTRON

Stars

Milky Way (sketch)



Galaxies

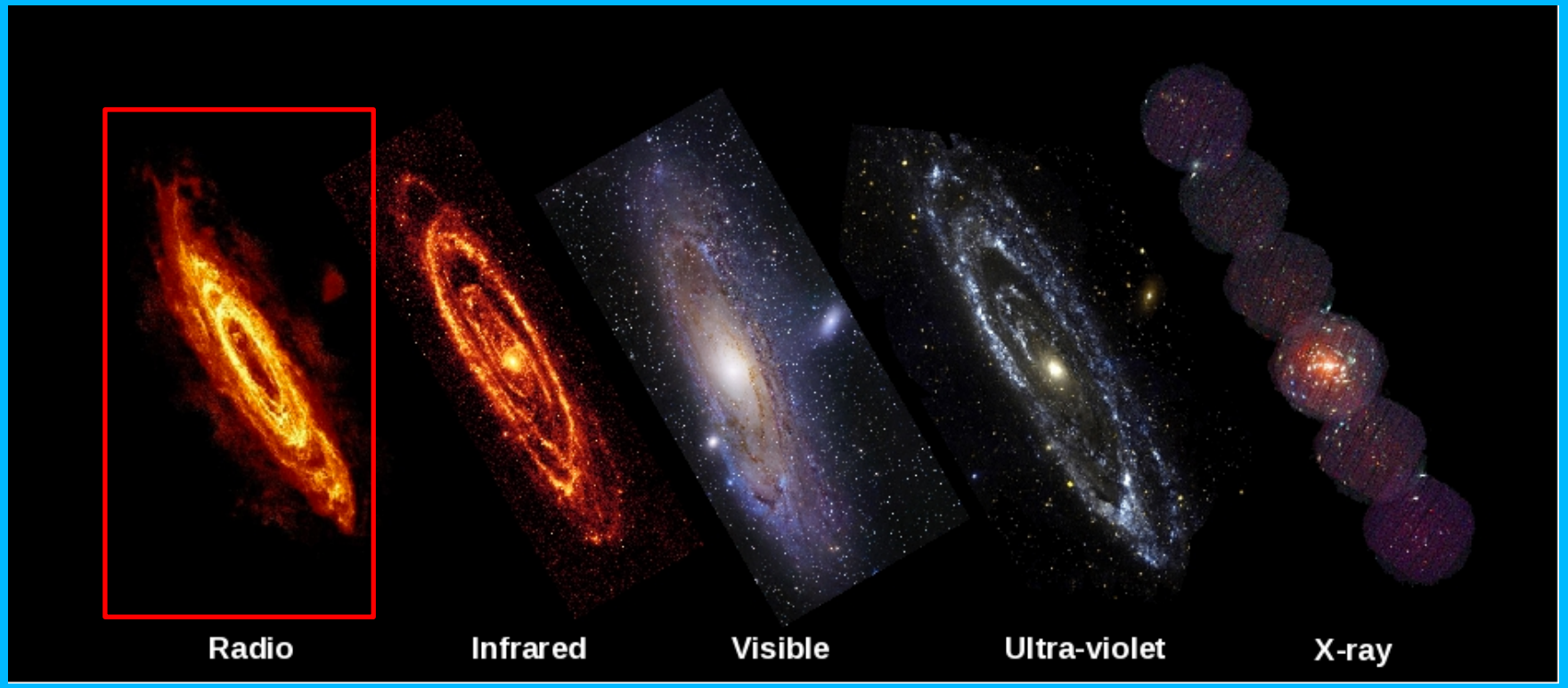
Our sun is one of the many stars in the Milky Way Galaxy

The Milky Way is one of the many galaxies in the Universe

Astronomy II (Radio)

APERTIF
APERture Tile In Focus
ASTRON

Andromeda Galaxy (Multi-wavelength View)



longer ← Electromagnetic Wavelength → *shorter*

Westerbork Synthesis Radio Telescope

APERTIF
APERture Tile In Focus

ASTRON

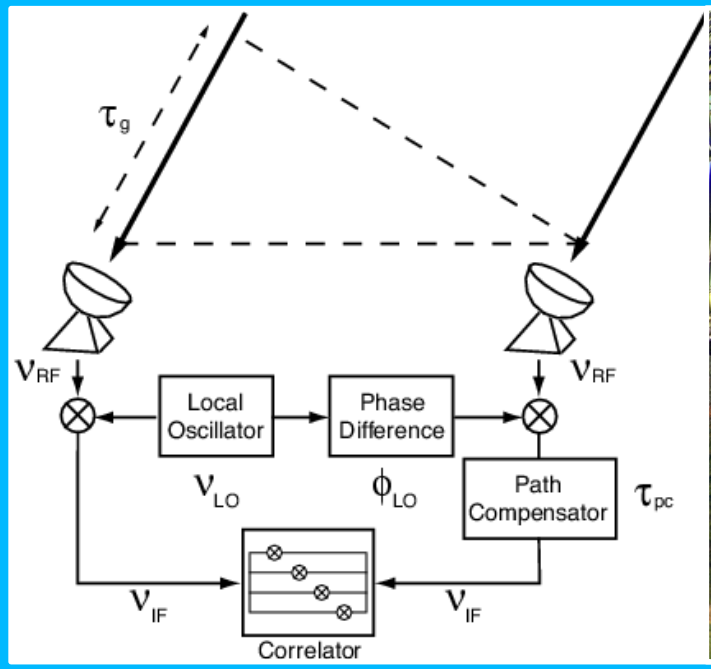
WSRT consists of 14 Radio dishes of 25 meter in Diameter, built in 1970 and operated by ASTRON. It is an East-West array built for radio interferometry

Westerbork Synthesis Radio Telescope -- data production --

APERTIF
APERture Tile In Focus

ASTRON

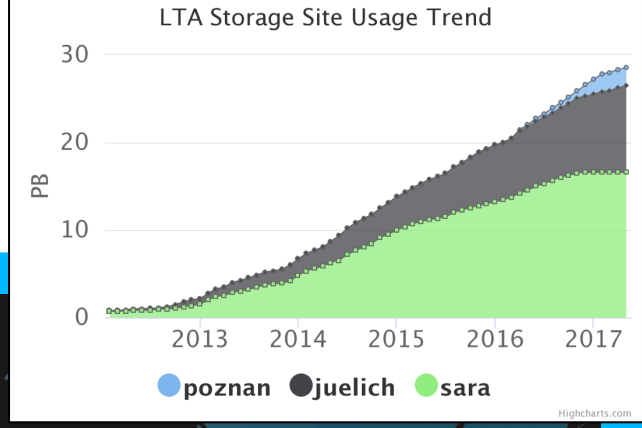
Due to the long wavelength nature of radio astronomy, special techniques have to be used to "image" the sky.

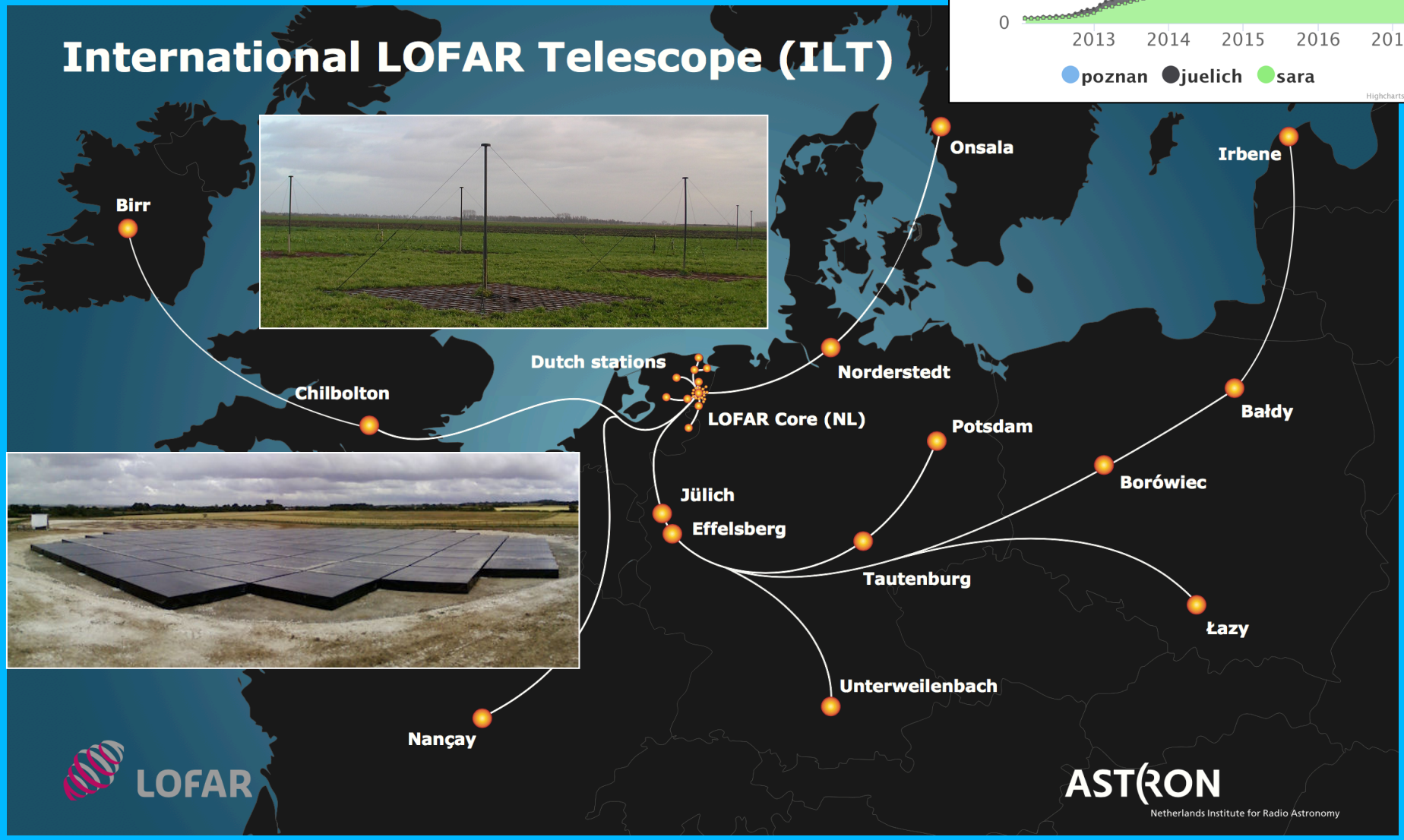


The signal need to be continuously digitized to correlate the data →

Radio Telescopes produce substantial amounts of data, with volumes of "astronomical" proportions

Radio Astronomy at scale: International LOFAR Telescope



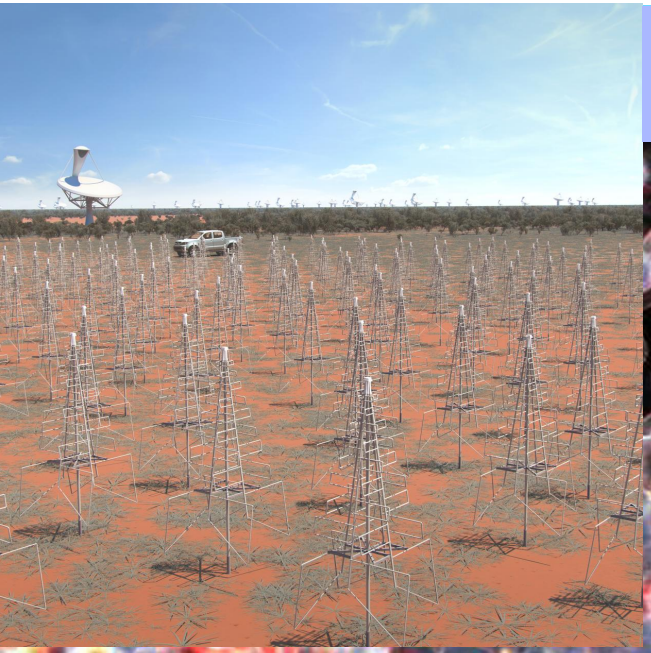


Square Kilometre Array

Taking it to Exa-scale

APERTIF
APERture Tile In Focus

ASTRON



Start of construction 2018

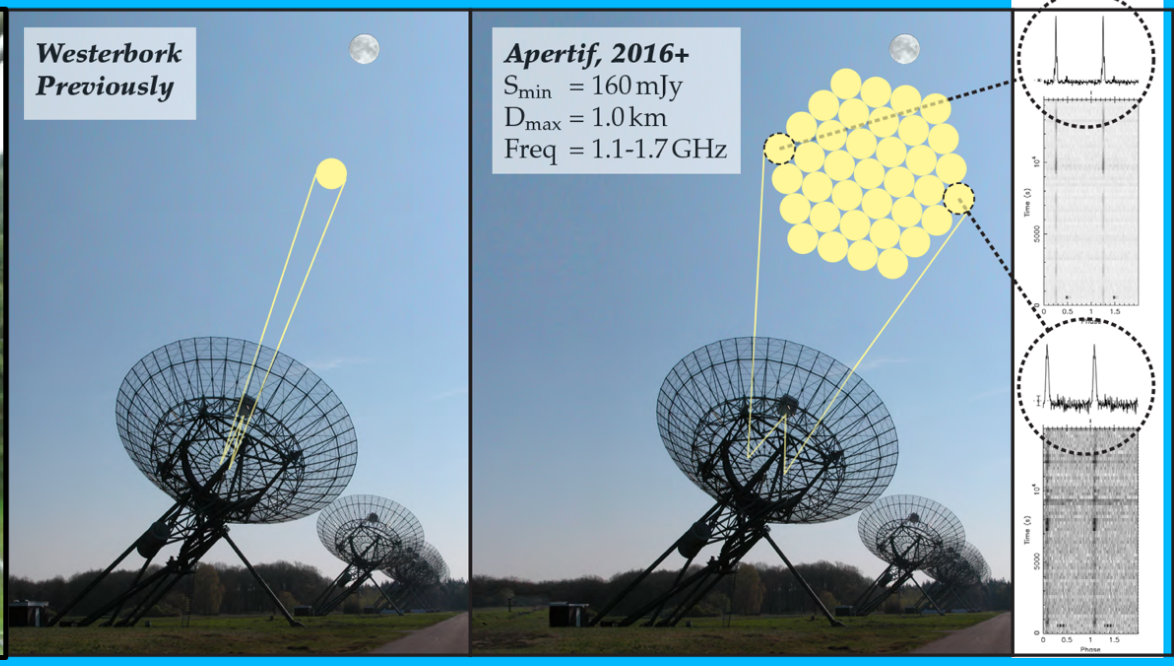
http://skatelescope.org

Westerbork Synthesis Radio Telescope -- APERTIF --

APERTIF
APERture Tile In Focus

ASTRON

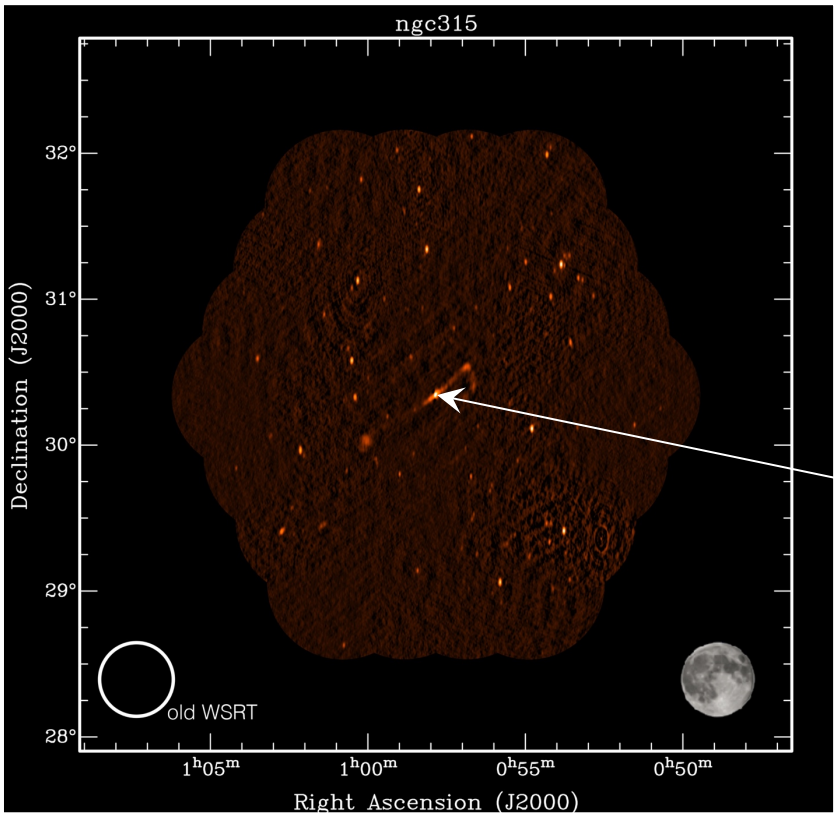
- APERture Tile In Focus:
  APERTIF replaces the ***single*** pixel detectors with an array of ***121 detectors*** forming up to ***40 beams***



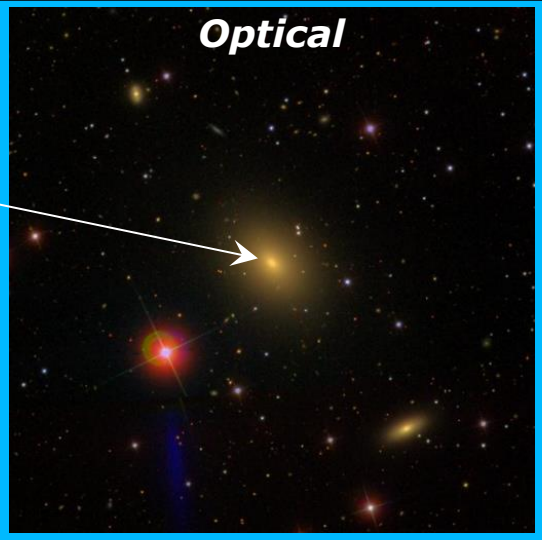
Westerbork Synthesis Radio Telescope
-- APERTIF, First Light! --

APERTIF
APERture Tile In Focus

ASTRON



NGC 315, “active” galaxy, where the central massive black hole, ejects massive amounts of hot gas. Visible as radio jets, which makes it one of the largest single objects in the Universe



10-05-2017

Still in the commissioning phase of the new instrument
http://www.astron.nl/dailyimage/

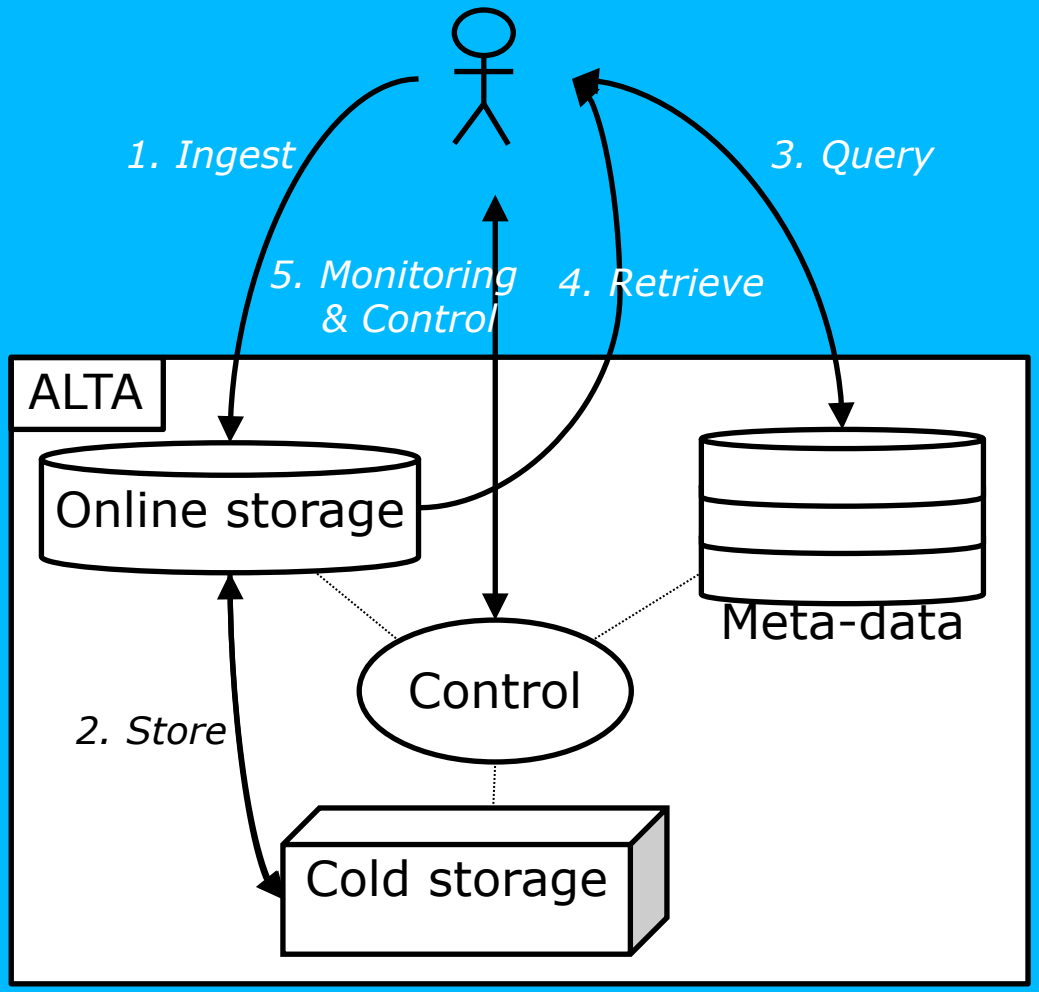
APERTIF and Long Term Archive
-- Purpose & Use Cases --

APERTIF
APERture Tile In Focus

ASTRON

High Level Use-Cases:

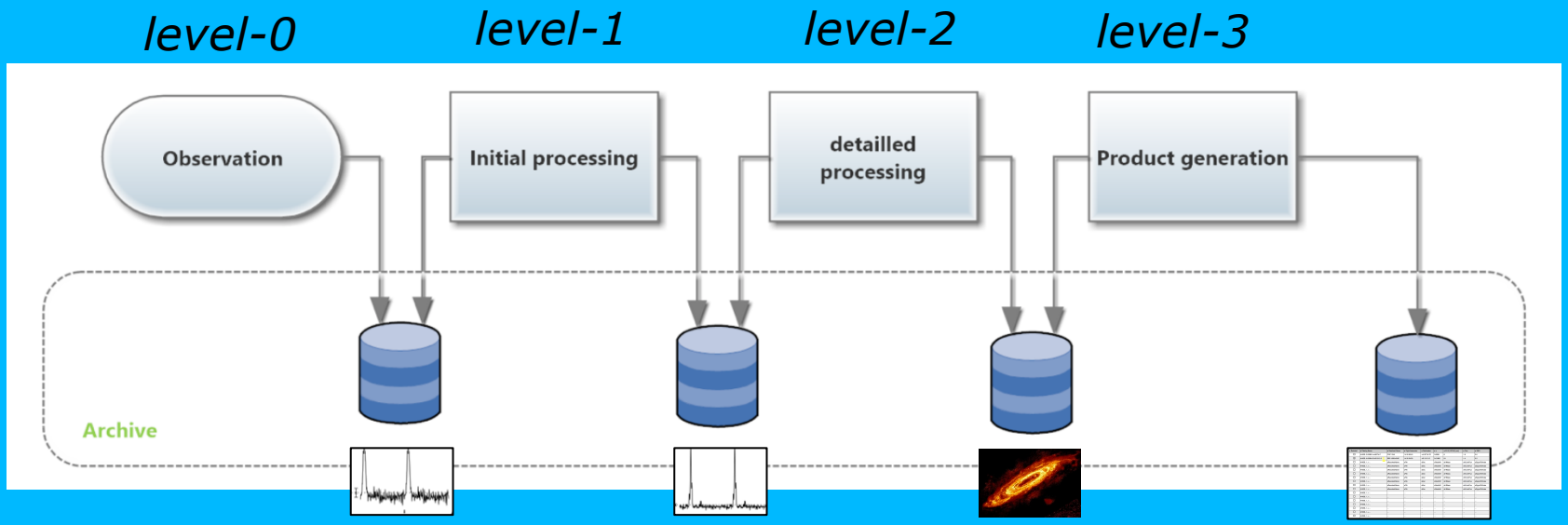
1. Ingest Data
2. Store Data
3. Query Meta-data
4. Retrieve Data
5. Monitoring & Control



APERTIF and Long Term Archive

APERTIF is going to be used as a Survey Instrument:
Standardized configurations and processing pipelines that produce a fixed set of known data-products:

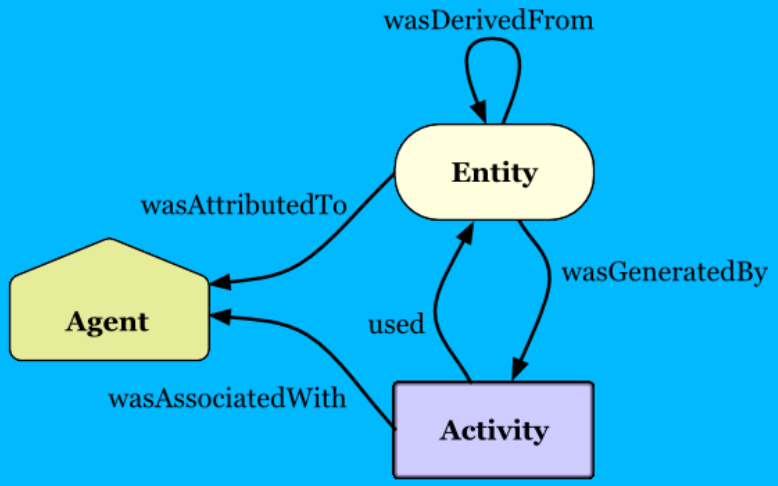
- produce 4 PB per year of data-products, estimated 5yr. → 20 PB
- order 10 to 100 million data-products.
- typical size of a data-product 1 – 60 GB.
- typical data rates: 10 – 20 Gbps
- number of users: hundreds (thousands 'anonymous' users)



APERTIF and Long Term Archive
-- Metadata; Provenance--

APERTIF is going to be used as a Survey Instrument:
Standardized configurations and processing pipelines that produce a fixed set of known data-products:

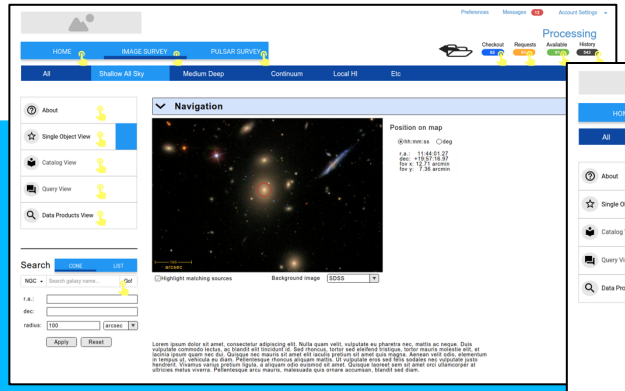
- Each subsequent level that is ingested has metadata that needs to be extracted.
- Processing is done in many different places; this history needs to be recorded → *DATA PROVENANCE*
- Data-model used for ALTA (& *Virtual Observatory*) uses the ***W3C Provenance Model:***

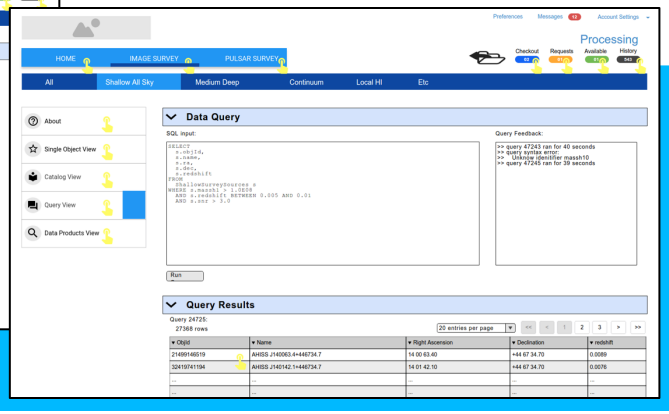


ALTA High level system overview

APERTIF
APERture Tile In Focus

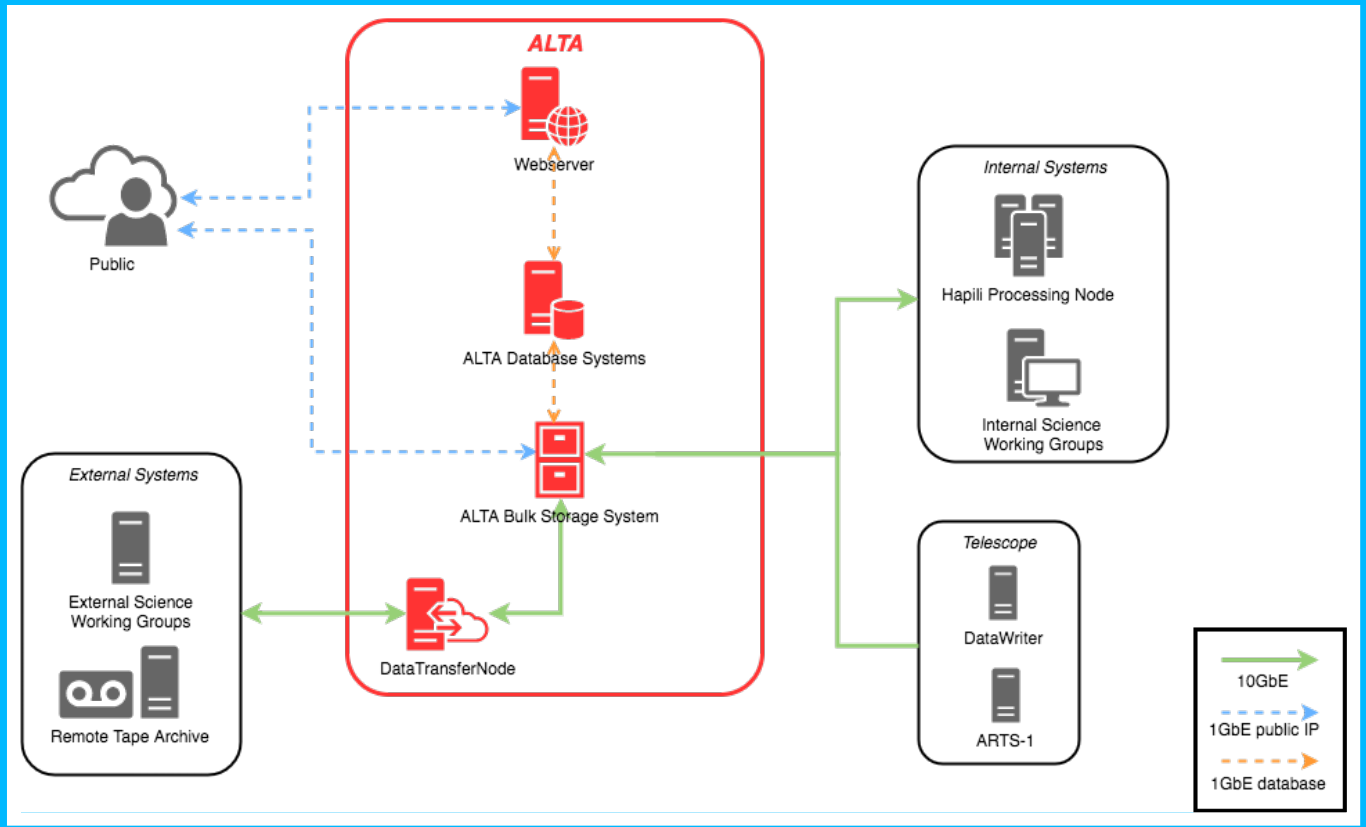
ASTRON



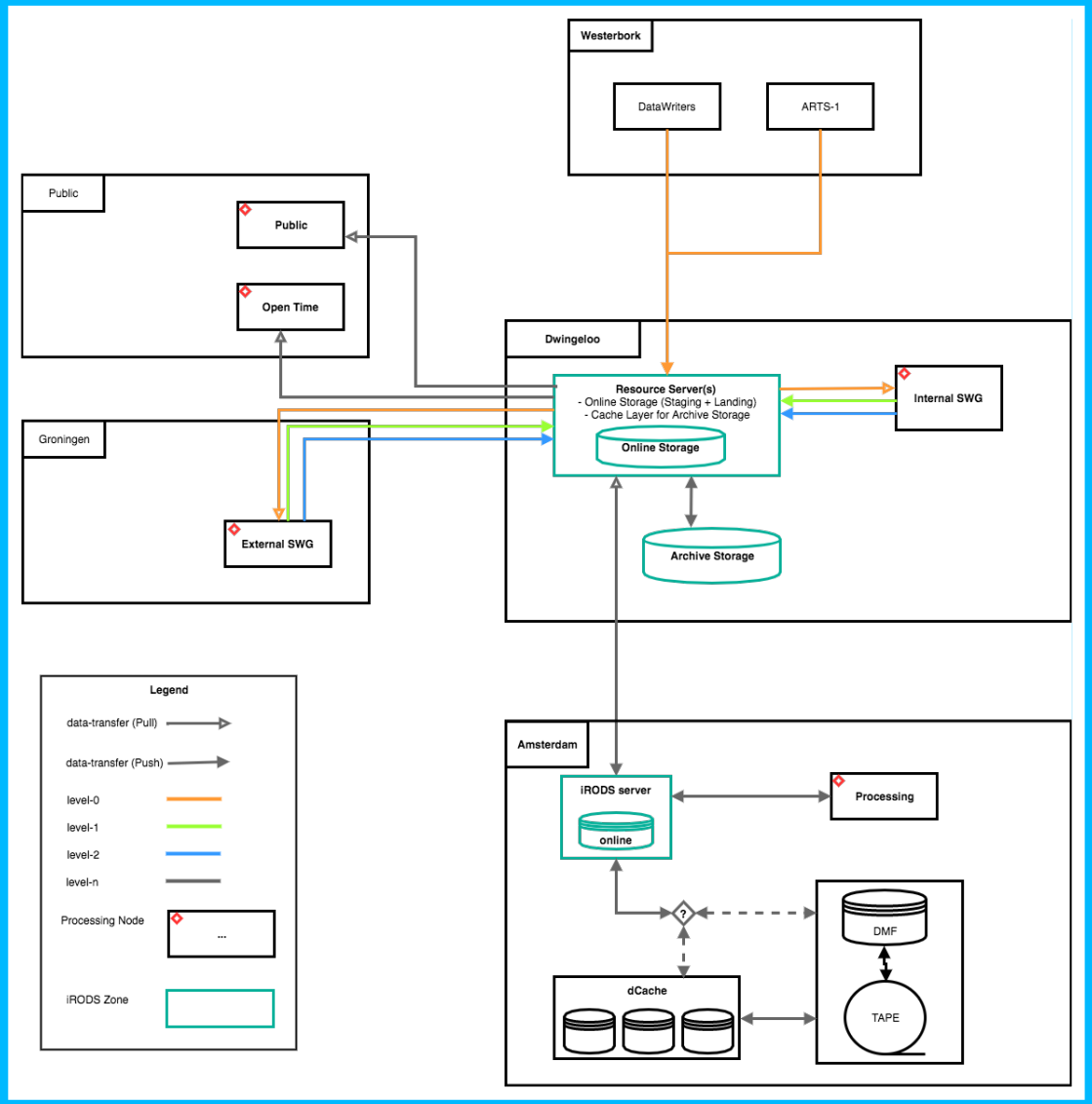


Data analysis processing not in scope of ALTA system

- **Webserver**
  - Main (G)UI
- **Database**
  - iCAT
  - Datamodel
- **Bulk Storage**
  - iRODS
- **DataTransfer**
  - Science DMZ



ALTA data flow diagram



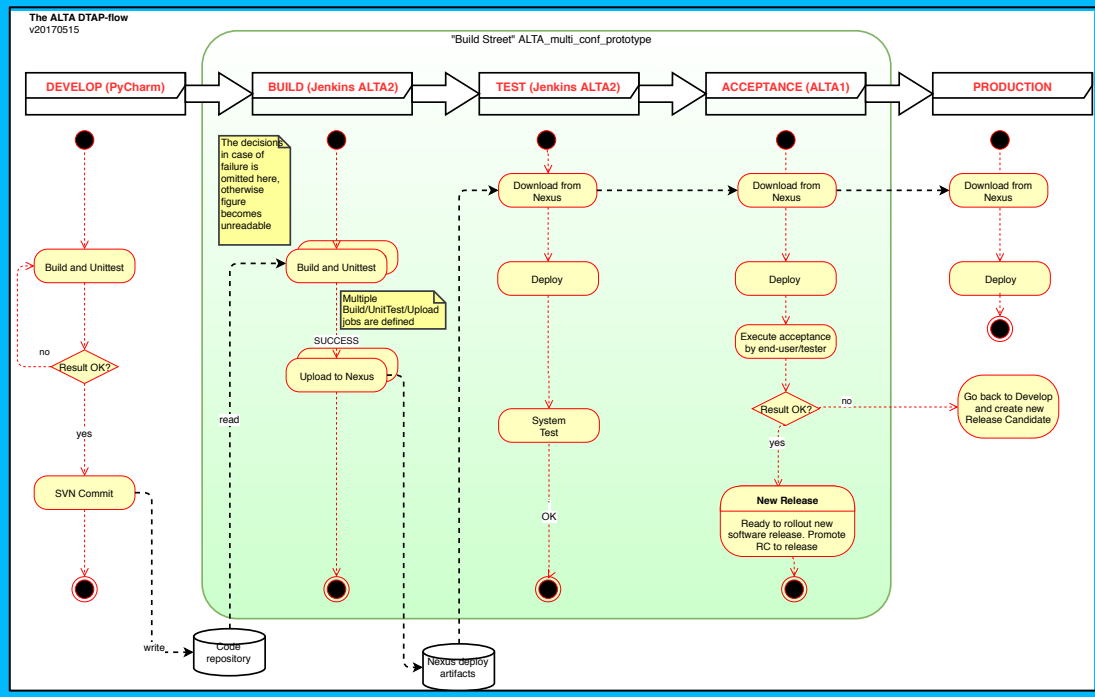
- Dwingeloo & Amsterdam to become **integrated iRODS resources**
- ALTA supports APERTIF **processing data flows**
  - Ingest from instrument & processing clusters
  - Distribution to processing clusters & public
  - Policy based data placement & replication

Dishing out ALTA: Ansible (& Vagrant)

APERTIF
APERture Tile In Focus

ASTRON

- Ansible for deployment
  - Python based
  - YAML configuration
  - Functional installation is defined in ‘roles’
  - Roles are deployed on groups of hosts using a ‘playbook’
  - Hosts are mapped to groups in an ‘inventory’



- Develop/build/test in VM’s (Vagrant based)
  - Complete ALTA environment can be brought up with a single `> vagrant up` command (ask for demo)
- Acceptance/production on dedicated servers (physical + VM)

ALTA & iRODS I

APERTIF
APERture Tile In Focus

ASTRON

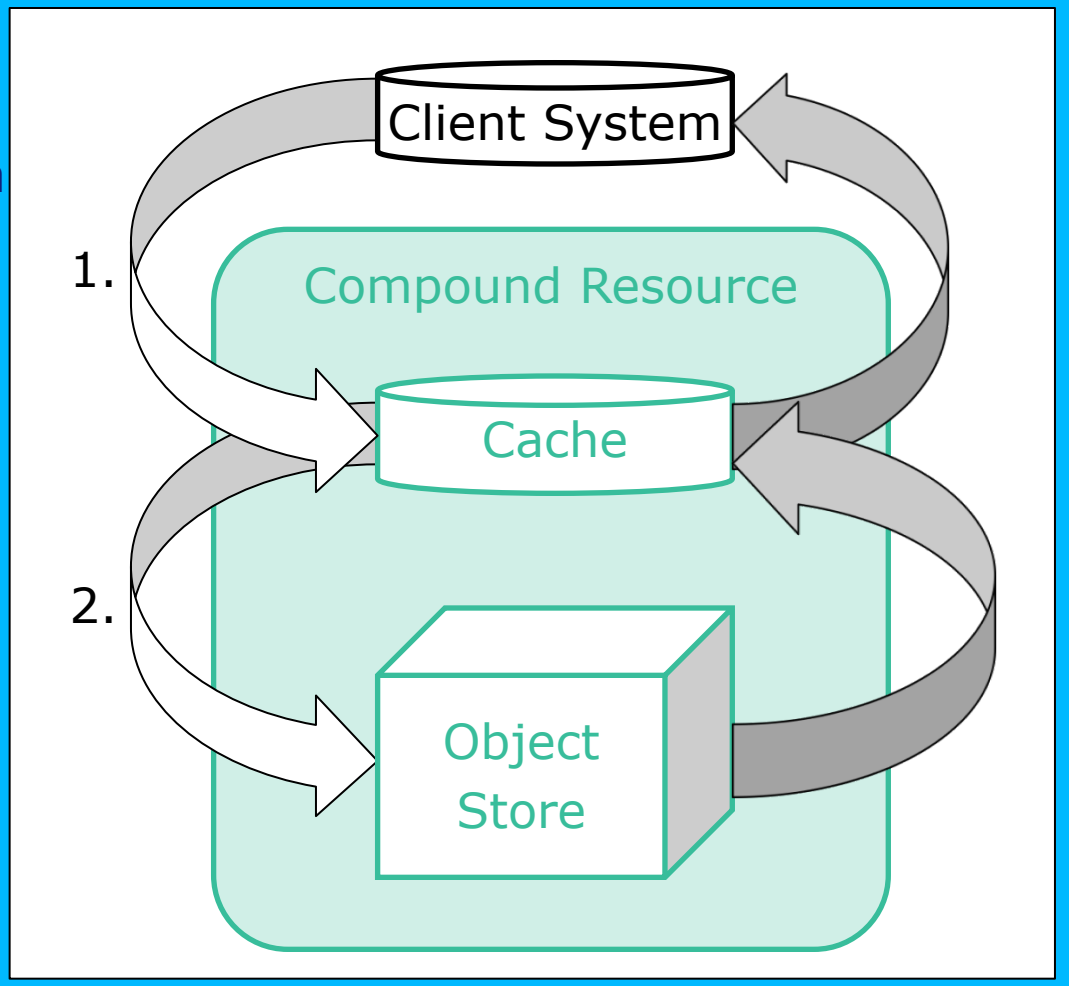
- When comparing similar products we noticed that:
  - iRODS Data management middleware layer supports many of our requirements
    - Abstraction of storage resources
    - Policy based data management
    - Supporting geographically distributed systems
    - Efficient data transfers, proven at scale
  - Active developer & user communities; used/known by most of our partner institutes
  - Documentation & maturity (core functionality)
  - iCAT is single point of failure
  - Flat string-based metadata (performance concern)

ALTA & iRODS II

ALTA needs to support a continuous running survey project, both at peak and average data-transfer rates. **Experimented with object stores.**

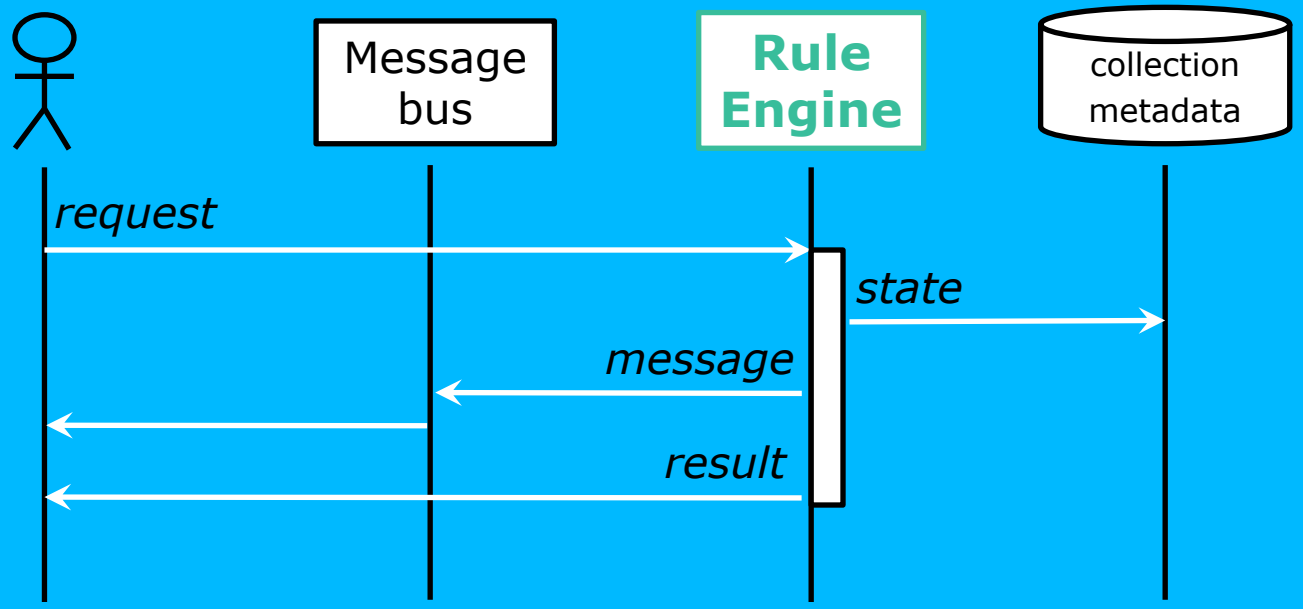
- **Posix cache required** for all puts & gets
- **Scaling out** requires additiona components
  - Load balancer in front of cache servers
  - Distributed file system

'Pure iRODS' solution (multiple compound resources with single object store backend) not attractive as objects are only retrievable through cache node used for storing the data



ALTA & iRODS III

- Current implementing ingest rules within iRODS 4.2;
    1. server side controlled (on the iRODS resource server)
    2. clients make requests for a (bulk) transfer, via collections created in a landing storage area (using iRule iCommand).
    3. communication between client – server will be done via AMQP/Stomp message queues on a Message Broker.



Summary

APERTIF
APERture Tile In Focus
ASTRON

- **iRODS** is a promising technology for Radio Astronomical Archives
- There is a **vibrant developer and user community**
- We develop with a 10+ years horizon: **maturity & stability** essential
- Nevertheless, new capabilities are of **interest & important**
  - Relax requirement on cache in front of object stores (support high throughput) - **MultiPart**
  - Support for (integrating) elaborate meta-data DBs - **QueryArrow**
  - Mature, feature complete, **Python client & server support**
- **ALTA** planned to go **live this year**; APERTIF Surveys will commence in 2018; **First survey release expected in 2019**