LOFAR: CHALLENGES AND SOLUTIONS TO OPERATE THE WORLD'S LARGEST RADIO TELESCOPE

R. F. Pizzo
Head Science Operations & Support
ASTRON Radio Observatory

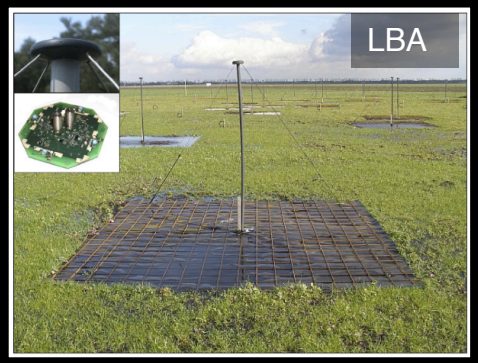
ASTRON
Netherlands Institute for Radio Astronomy

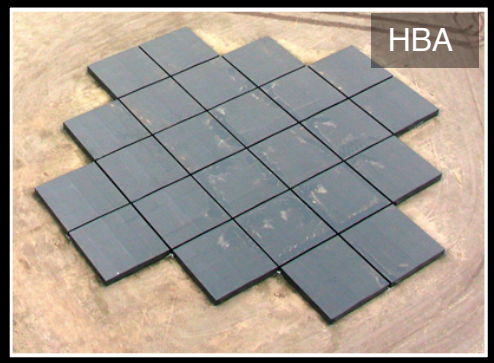
Groningen, October 9th 2019

LOFAR

THE LOW FREQUENCY ARRAY – KEY FACTS

- Array of 52 dipole antenna stations distributed across EU (~2000 km baseline)
- Soon station also in Italy (2022)
- 10-250 MHz
- Low band antenna (LBA; 4800 dipole pairs, 96 LBA per station, Area ~ 75200 m^2; 10-90 MHz)
- High Band Antenna (HBA; 47616 dipole pairs, 48/96 tiles per station in NL/EU, Area ~ 57000 m^2; 110-250 MHz)
- Several observing modes - 96 MHz bandwidth (multi-beam option)
- 7 years of operations - > 40000 hours of science delivered





THE LOFAR SYSTEM: DATA FLOW

stations: data collection

COBALT2: correlation

CEP4 compute cluster

Processing through RO pipelines

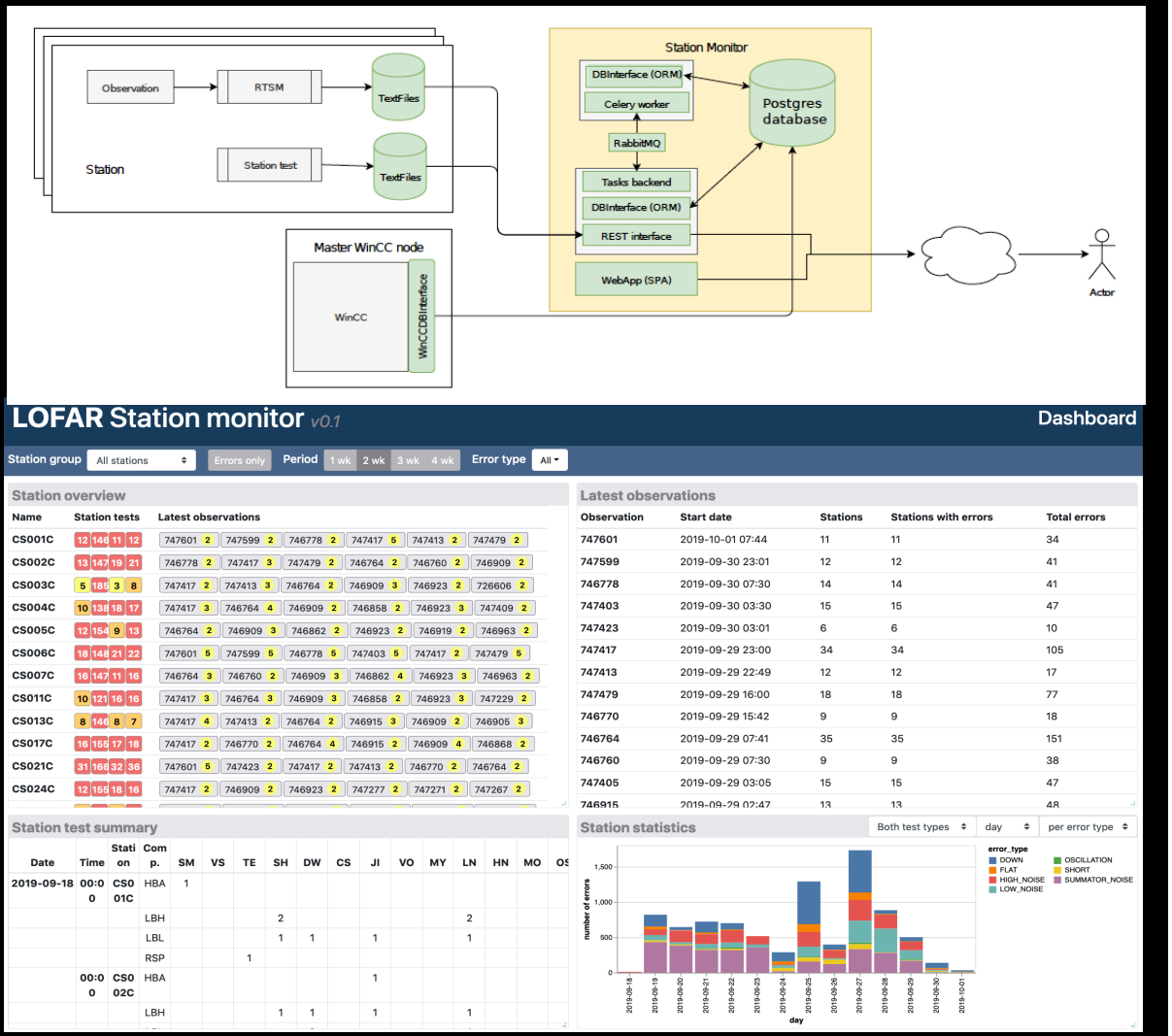
ingest into Long-term archive (3 sites: NL, D, PL)

- Transport, processing and storage of large amounts of data :
  - Data flow from all antennas combined: 1.7 Tbyte/s
  - To COBALT from station after beamforming: 28 Gbyte/s
  - Correlator output to disk: between 2-10 Gbyte/s
  - Data storage challenges: ~ 80 TB/h
- LOFAR: important technological pathfinder for next-gen facilities and data intensive astronomy

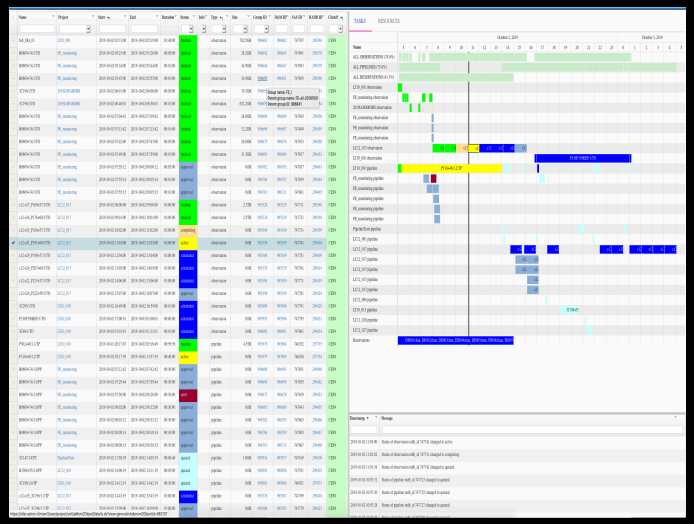
HARDWARE HEALTH

- Quality of the antennas severely reduced by mechanical and electronic problems -> hardware maintenance
- Station tests data spread across the system -> Maintenance Management Information System (MMIS)
- Station monitor: database with middleware to ingest and provide access to data through a REST API and a web user interface
- Challenge: present all data points from hundreds of station components in a comprehensive way
- Further evolution: automatic enabling/disabling of elements

See poster by M. Mancini (P8.5)



RESOURCE MANAGEMENT



SCHEDULING:
RESOURCE ASSIGNER

CEP4 PROCESSING

- Challenge: Keep steady data flow – per Cycle: 7 PB products + ~50 PB raw data
- Scheduling on available resources
- Pipeline queue handled automatically: SLURM
- Automatic data cleanup after ingest
- Manual ingest after data quality assessment

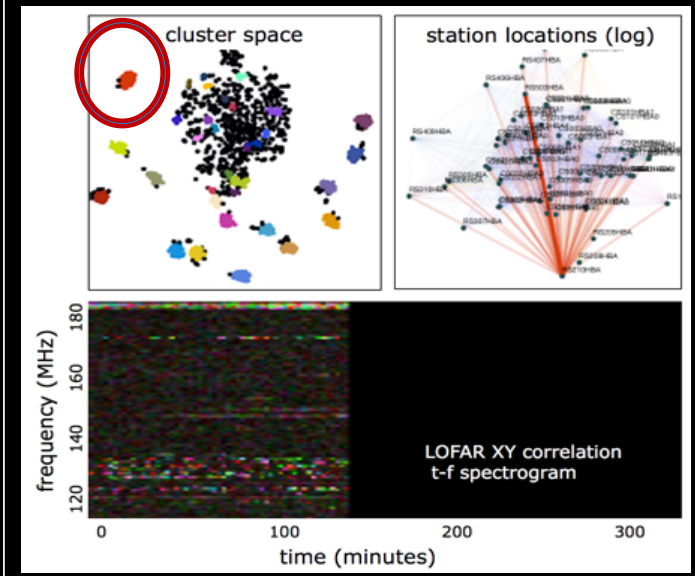
DATA QUALITY ASSESSMENT

- 6000+ validation plots automatically generated -> visual inspection prohibitive
- Next generation inspection plots based on AI are being tested
- Dynamic spectrum plots presented to the user grouped per cluster
- On going research
- Goal: automatic observation reports

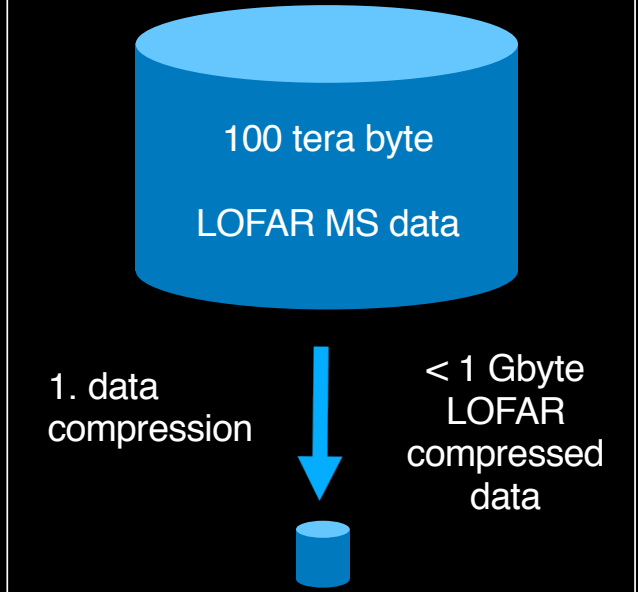
See poster by A.J. Boonstra (P3.5)

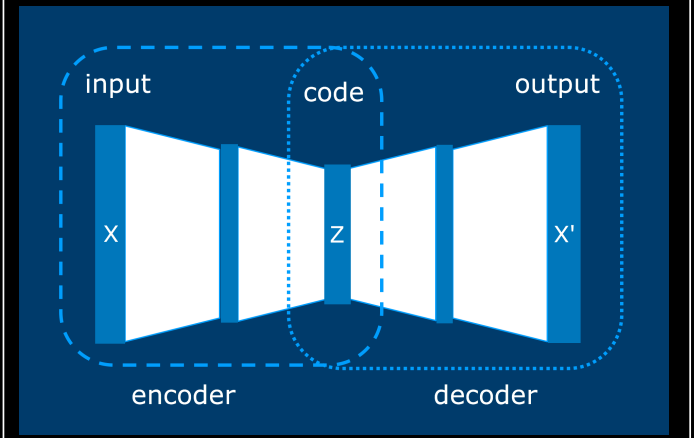
Collaboration with

eScience center



3. Cluster spectrograms

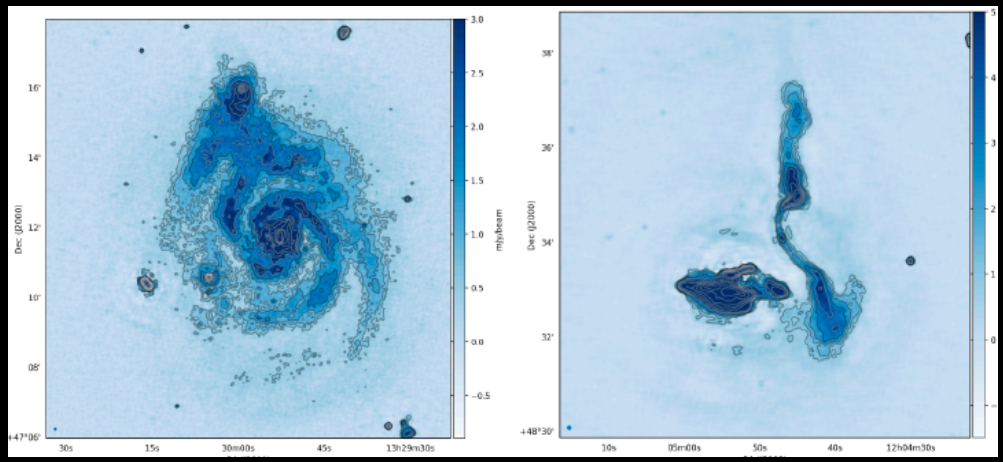




2. Neural network

DELIVERING SCIENCE-READY DATA

- Challenge: deliver automatically high level data products ready for scientific exploitation -> make LOFAR accessible to an even wider community
- Method:
  - Implement in production advanced reduction pipelines
  - Scientific exploitation of data *in loco*



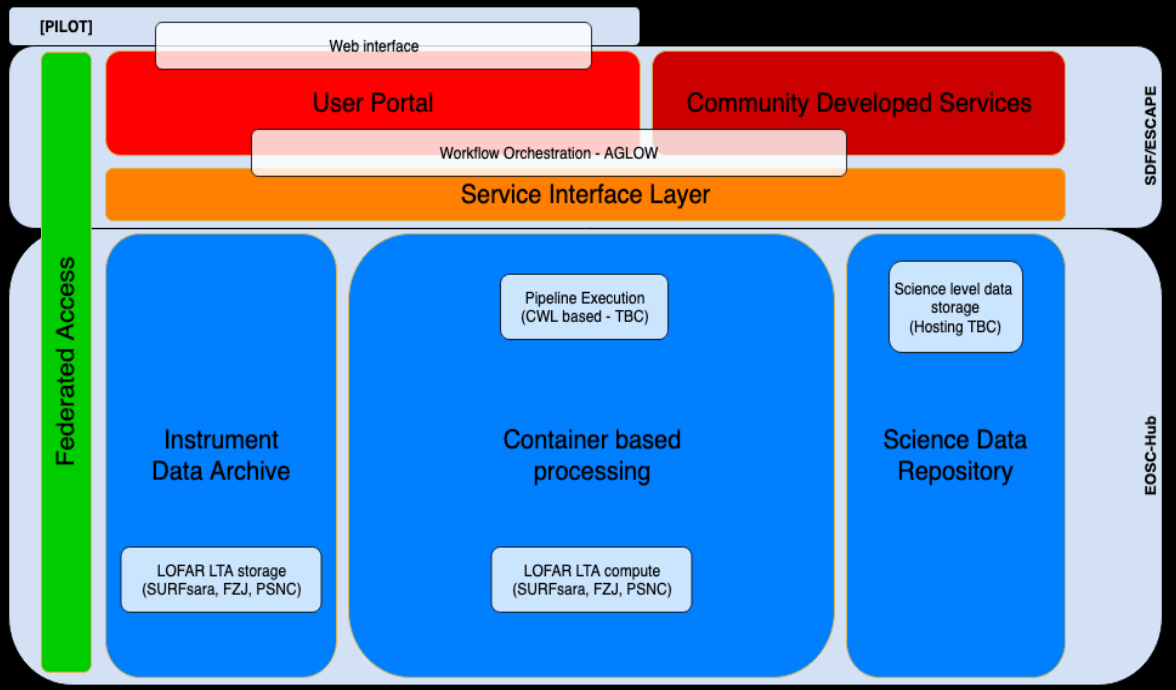
See session 10 tomorrow:
A. Mechev & A. Offringa

DELIVERING SCIENCE-READY DATA

- Concerted efforts of active project
  - SDF: contact: R. Pizzo
  - EOSC: contact: H. Holties
  - ESCAPE - contact: Z. Meyer, Y. Grange

- Robust imaging on demand by Q1 2020

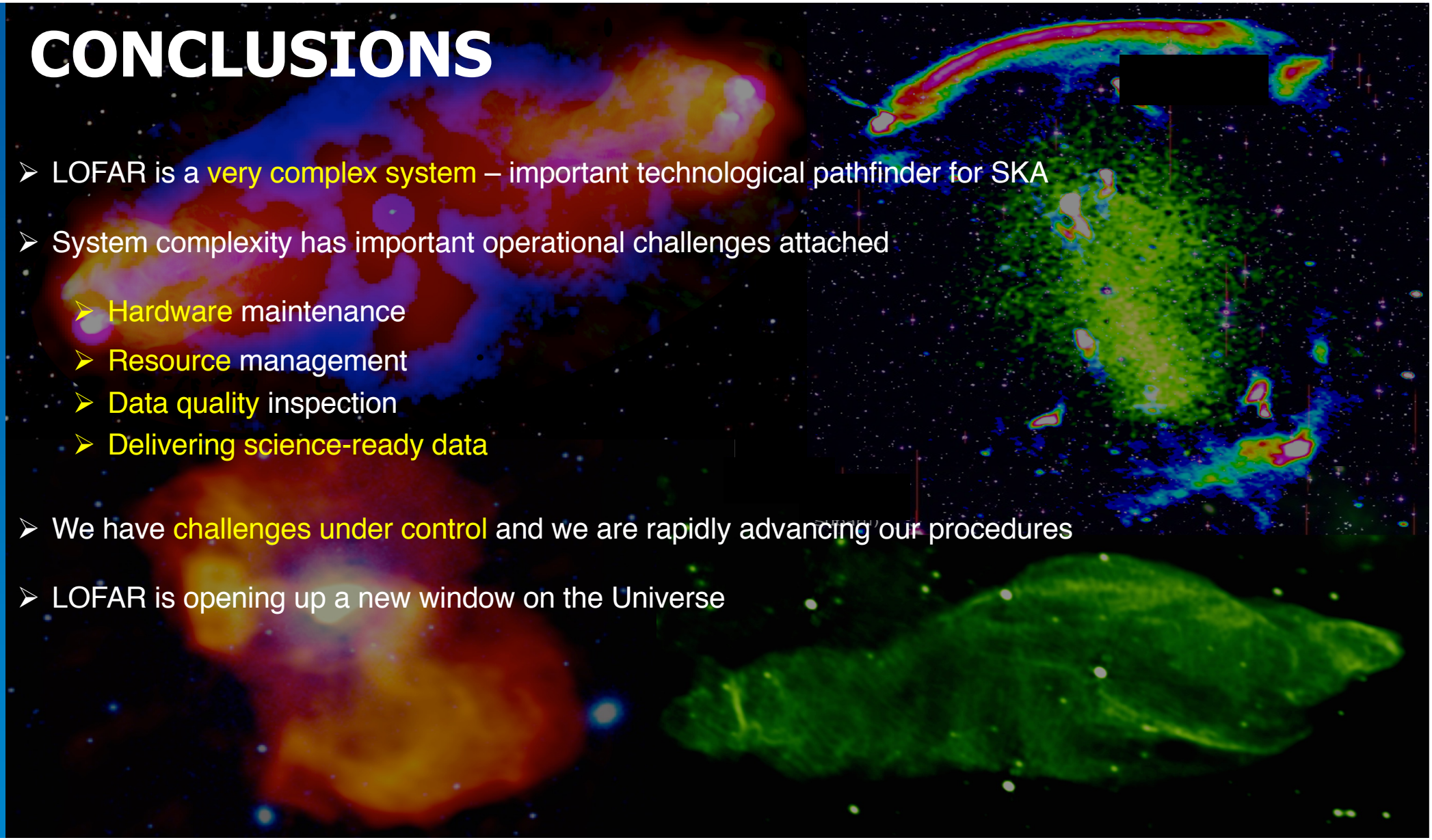
1. Select a pipeline workflow language
2. Develop and maintain a pipeline management tool + web interface for execution (see talk by A. Mechev tomorrow)
3. Generate an advanced data product archive



Courtesy of H. Holties

See session 10 tomorrow

CONCLUSIONS



- LOFAR is a very complex system – important technological pathfinder for SKA
- System complexity has important operational challenges attached
  - Hardware maintenance
  - Resource management
  - Data quality inspection
  - Delivering science-ready data
- We have challenges under control and we are rapidly advancing our procedures
- LOFAR is opening up a new window on the Universe