ASTRON

Netherlands Institute for Radio Astronomy

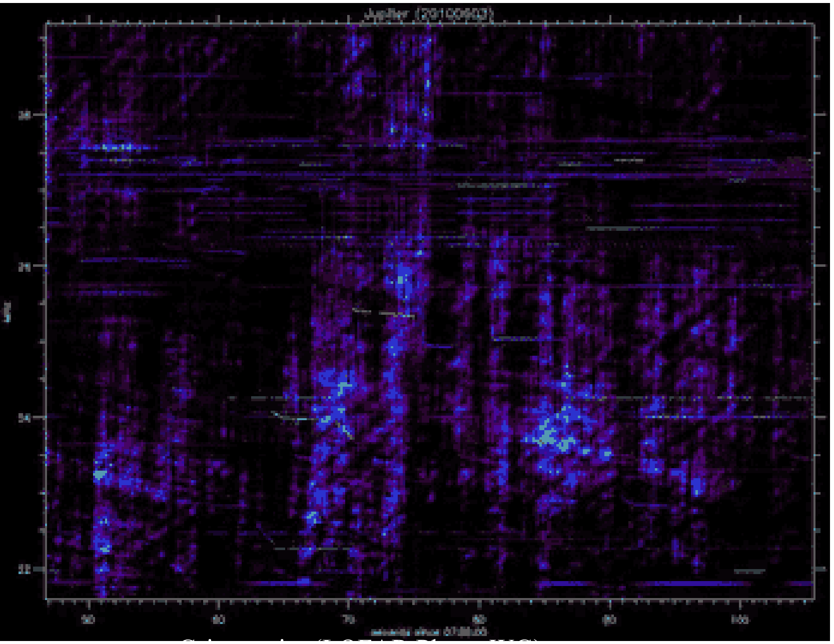
Distributed Data centres for the SKA
Building on LOFAR experiences

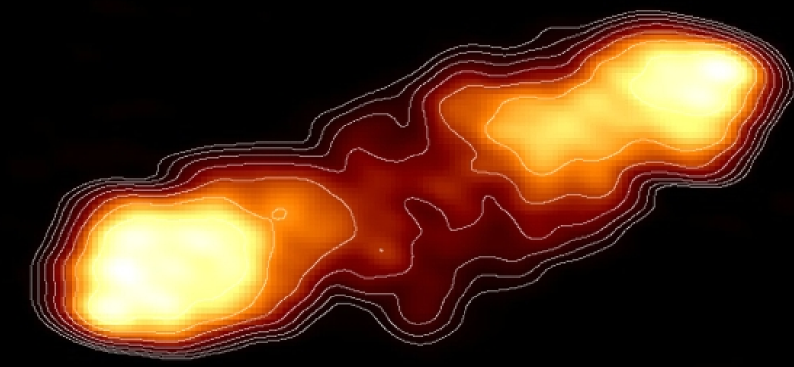
Dr. Marco de Vos (devos@astron.nl)


LOFAR Configuration
International Stations
Onsala, Sweden
Groningen
LOFAR Core
Amsterdam
Potsdam, Germany
Chilbolton, UK
Jülich, Germany
Tautenburg, Germany
Effelsberg, Germany
Unterweilenbach, Germany
Nançay, France

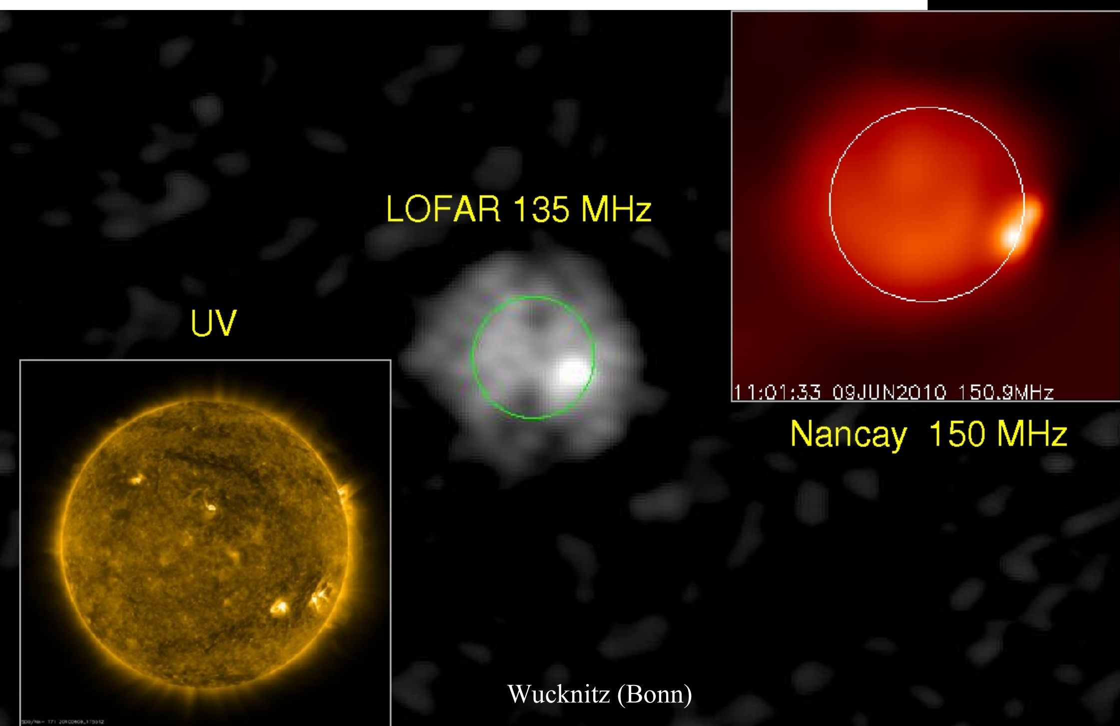
LOFAR Stations Across Europe

Recent LOFAR results

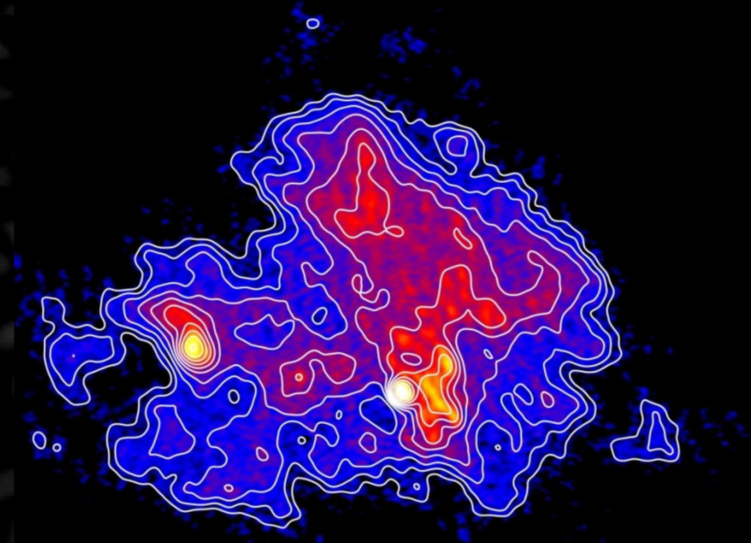




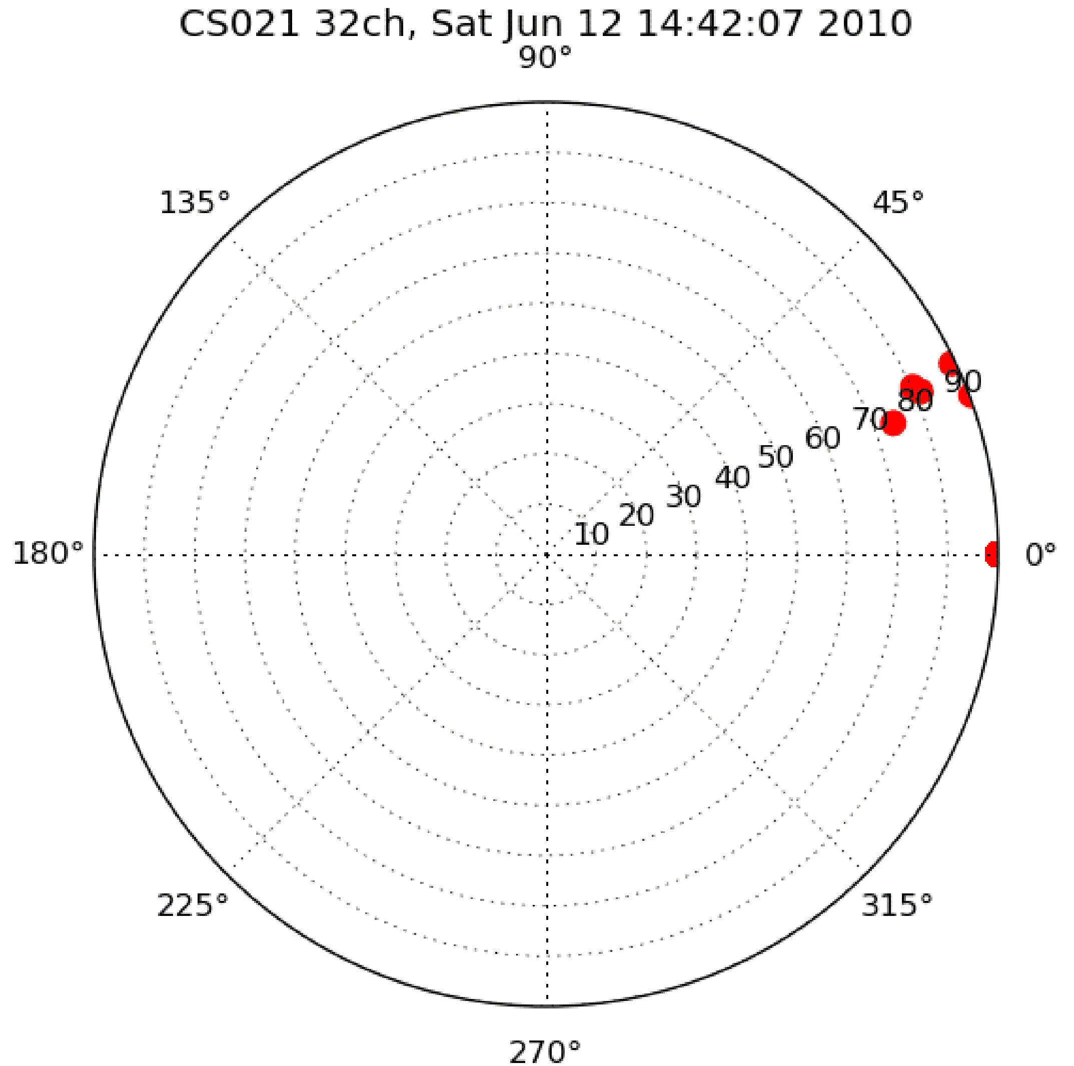
McKean (ASTRON)

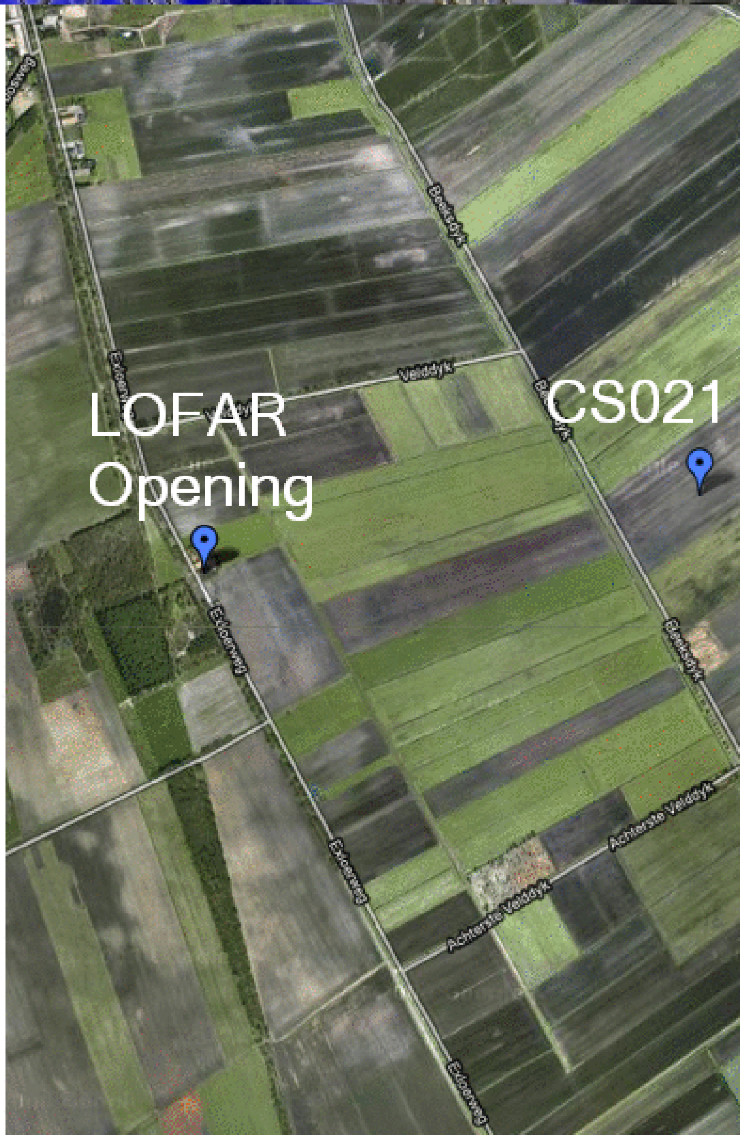


Wucknitz (Bonn)



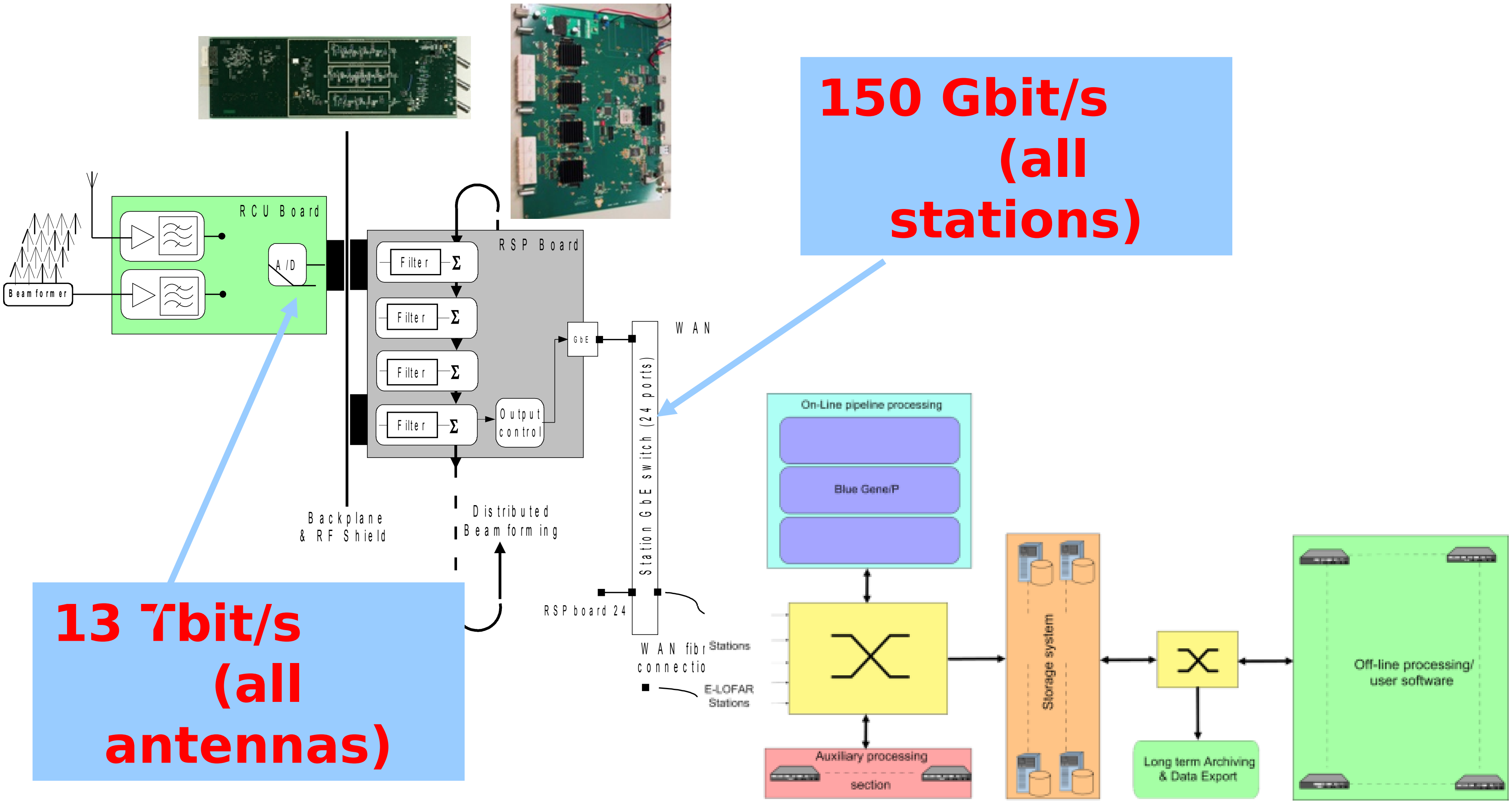
van Weeren, Bonafede, Ferrari, Orrù, Pizzo, Shulevski, van der Tol, Macario

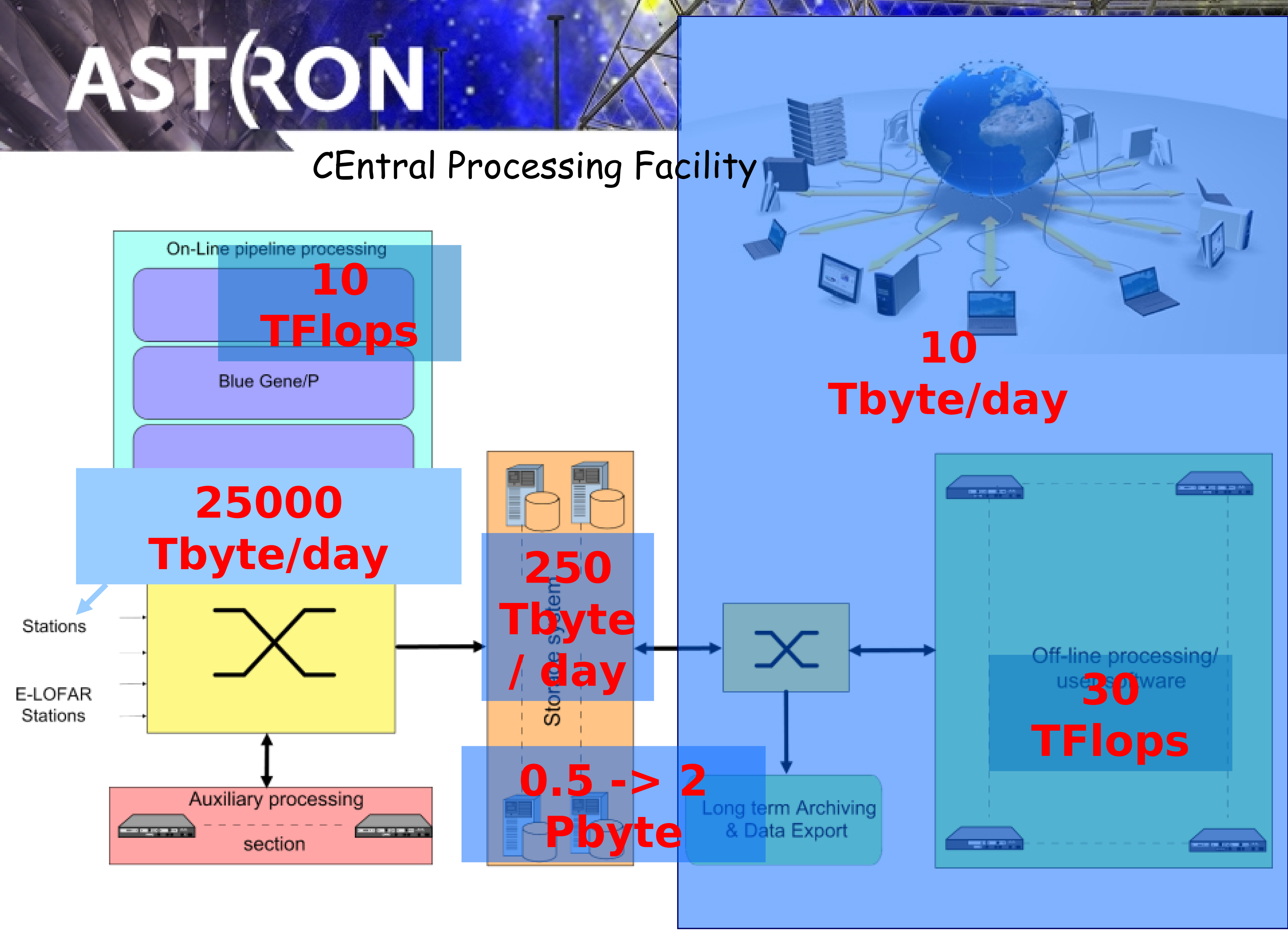

CS021 32ch, Sat Jun 12 14:42:07 2010
90°
45°
135°
180°
0°
225°
315°
270°
10
20
30
40
50
60
70
80
90


LOFAR Opening
CS021

ASTRON Netherlands Institute for Radio Astronomy

LOFAR top level architecture





ASTRON
CEntral Processing Facility
On-Line pipeline processing
10 TFlops
Blue Gene/P
25000 Tbyte/day
Stations
E-LOFAR Stations
250 Tbyte / day
Storage system
0.5 -> 2 Pbyte
Auxiliary processing section
10 Tbyte/day
Long term Archiving & Data Export
Off-line processing/ user software
30 TFlops

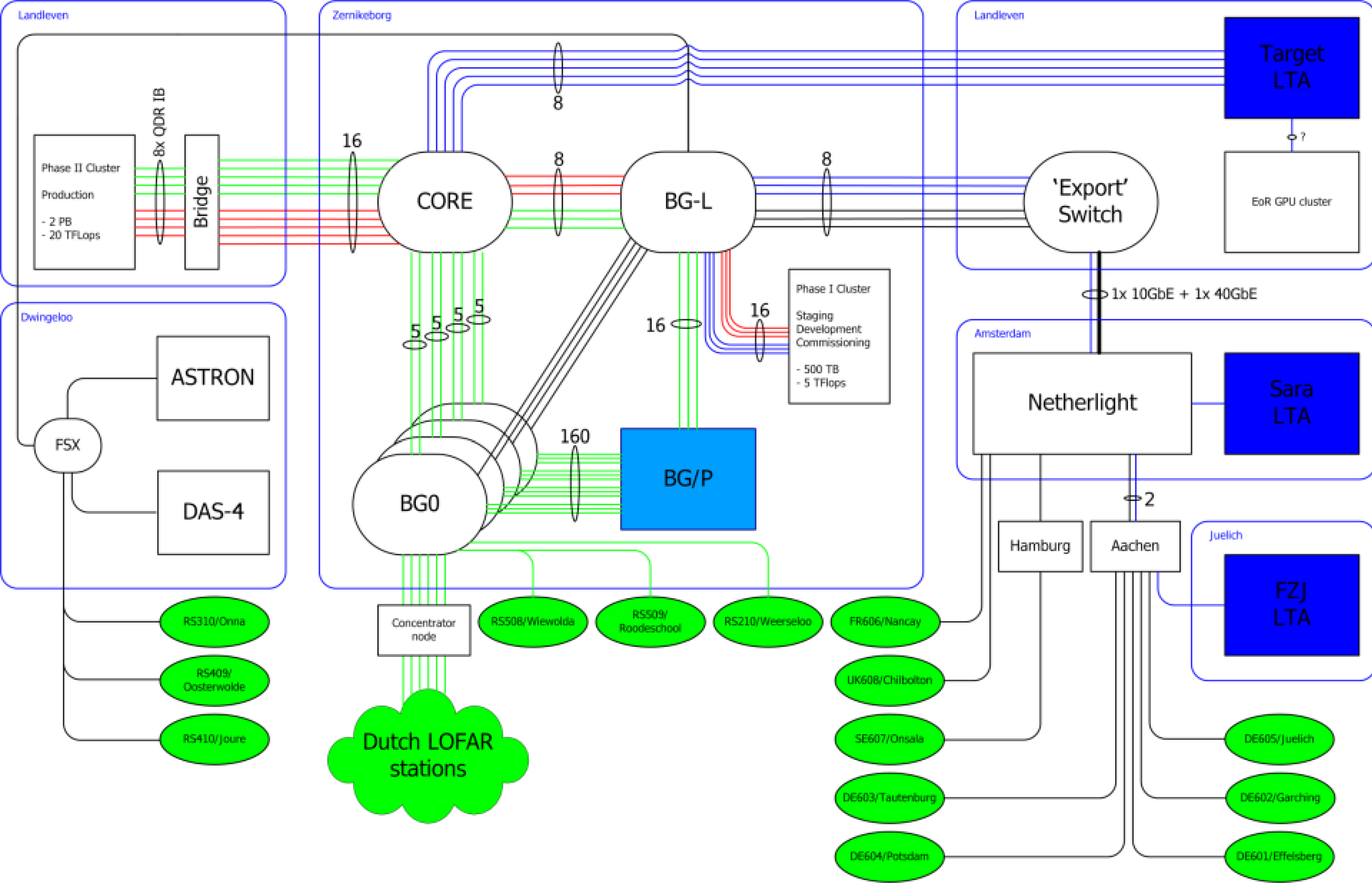
Multi-tier models for LHC & LOFAR

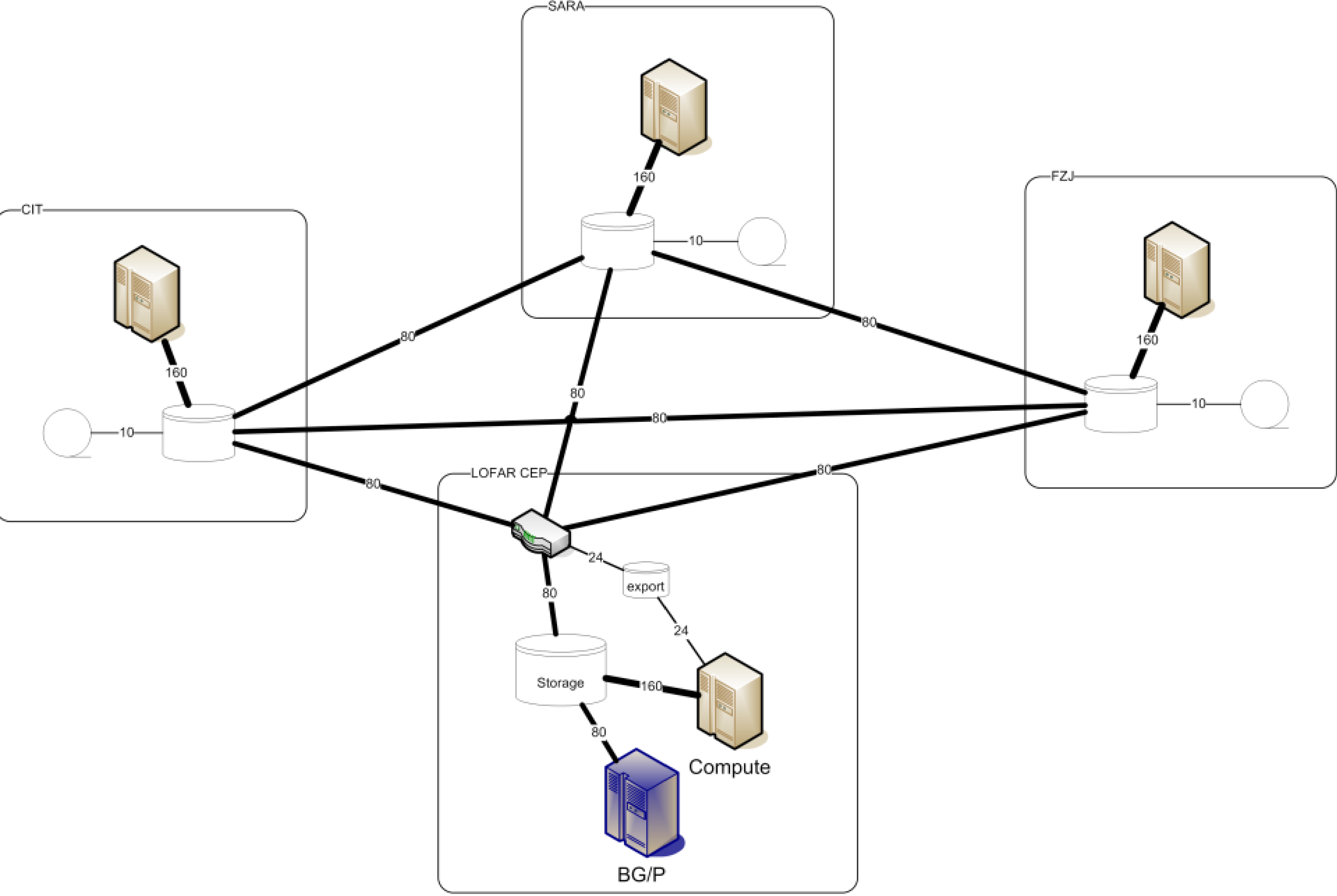
Large Hadron Collider

- Tier-0 (CERN):
  - copy of all raw data
- Tier-1 (~8-10 centers)
  - archive 1/n fraction of raw data & reconstructed data
  - regular re-processing of the raw data
  - archiving data from Tier-2 centers
  - provide central grid services: grid accessible computing and data resources
  - support coordination
- Tier-2 (~100 centers):
  - data analyses
  - no data archiving

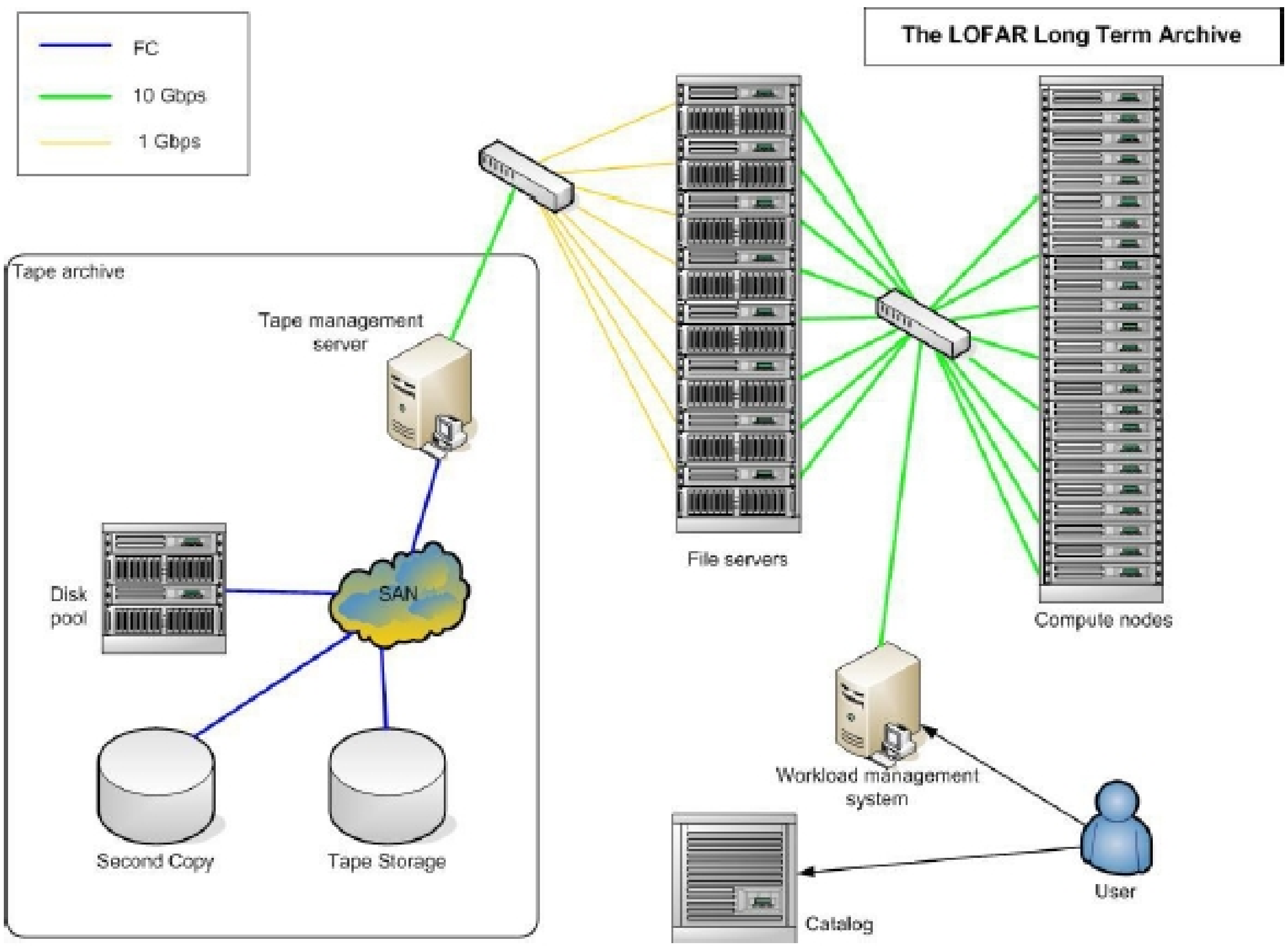
LOFAR

- Tier-0 (ASTRON/RuG): stations, central processor
  - lots of processing, limited archive
- Tier-1 (~8-10 centers worldwide)
  - specialized science support
  - advanced processing of (large) data sets
  - dedicated archive, also for Tier-2 centers
  - provide central grid services: grid accessible computing and data resources
  - support coordination
- Tier-2 (>100 centers):
  - individual scientists
  - data analyses
  - no data archiving

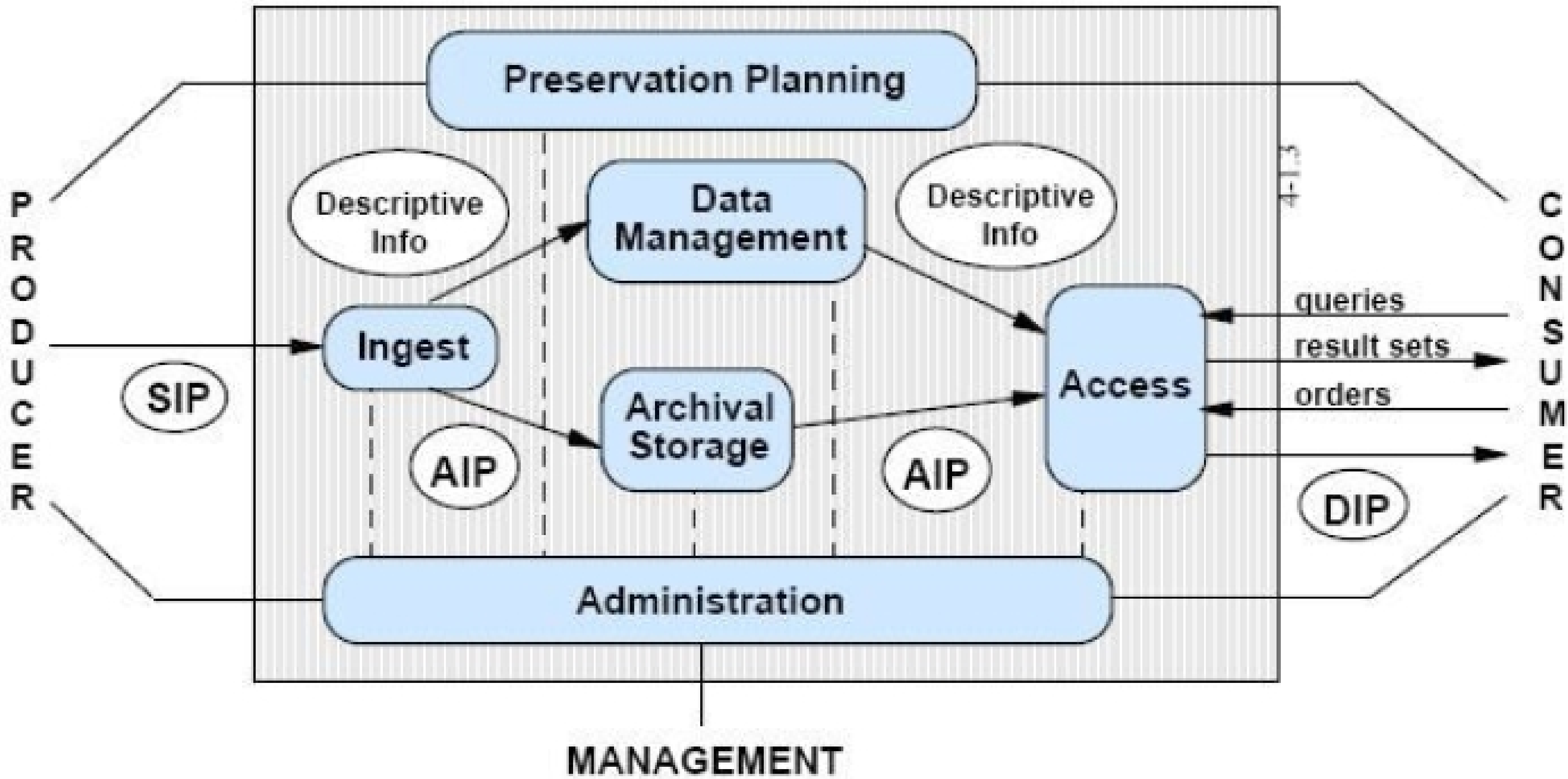

Landleven
Zernikeborg
Landleven
Target LTA
8
16
8x QDR IB
Phase II Cluster
Production
- 2 PB
- 20 TFLops
Bridge
CORE
8
BG-L
8
'Export' Switch
?
EoR GPU cluster
1x 10GbE + 1x 40GbE
Dwingeloo
ASTRON
5
5
5
5
16
16
Phase I Cluster
Staging
Development
Commissioning
- 500 TB
- 5 TFlops
Amsterdam
Netherlight
Sara LTA
FSX
160
BG/P
DAS-4
BG0
2
Hamburg
Aachen
Juelich
FZJ LTA
RS310/Onna
Concentrator node
RS508/Wiewolda
RS509/ Roodeschool
RS210/Weerseloo
FR606/Nancay
RS409/ Oosterwolde
UK608/Chilbolton
RS410/Joure
Dutch LOFAR stations
SE607/Onsala
DE605/Juelich
DE603/Tautenburg
DE602/Garching
DE604/Potsdam
DE601/Effelsberg


SARA
160
10
80
80
80
FZJ
160
10
CIT
160
10
80
80
LOFAR CEP
80
24
export
80
24
Storage
160
80
Compute
BG/P

Preservation Planning

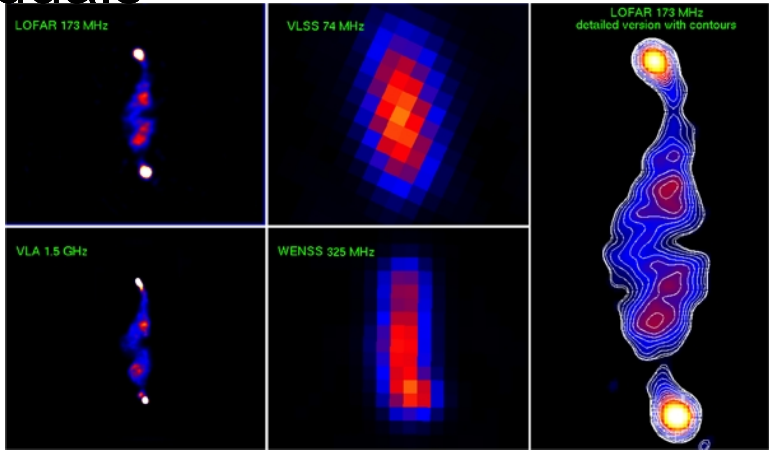


Open Archival Information System



The future of astronomical calibration

- Calibration
  - 1G: apply calibrator corrections
    - Rely on stability of telescope
  - 2G: solve/correct for direction independent effects
    - Assume corrections are the same over the field
  - 3G: correcting for direction dependent effects
    - Assume smoothness (also in changing beams)
  - 4G: statistical analysis of residuals
    - Reduce processing time
    - Assume bright sources can be subtracted



Processor designs for ExaFlop computing (1/3)

- LOFAR: extensive experience with BG/L and BG/P
  - ~10 TFlops correlator; ~30 TFlops postprocessing
- Evaluated several other architectures
  - GPUs by ATI and Nvidia, Intel i7, Cell BE
  - New research cluster (DAS-4)
  - In house FPGA experience (Uniboard)
- Trend: I/O challenge is going to dominate
  - Low computational intensity : I/O bound
  - High computational intensity : memory b/w bound
- Amdahl's law
  - Many-core architectures are here to stay
  - Example: our new OctaContraCota-core machine

Processor designs for ExaFlop computing (2/3)

So what would make an ideal processor for us?

- Excellent complex number support (like BG/P)
- High I/O per Flop ratio
- Integer arithmetic
- Reconfigurable FPU (DP vs. SP)
- Vector manipulation instructions (like on Cell)
- Relatively simple cores

Processor designs for ExaFlop computing (3/3)

System considerations:

- I/O
  - External (getting data into / out of the system)
  - Internal (reordering of data, MPI_Alltoall())
- Power efficiency (Flops/W)
- Reliability
- Configurability
- Durability / Maintainability
  - Compare lifecycle hardware correlator vs. BG/L
  - Graceful degradation

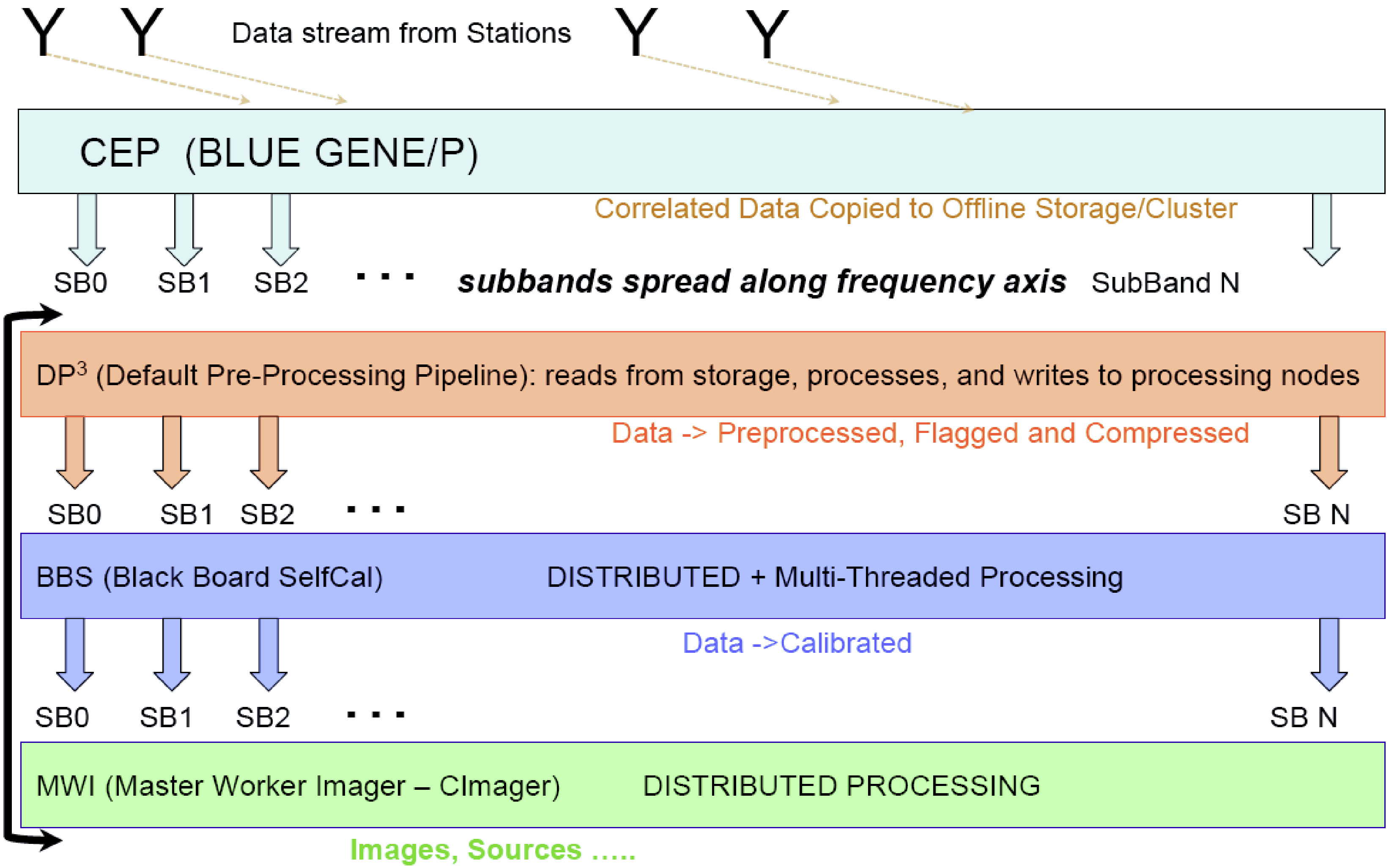
Important high-impact decisions

- Global facilities in a networked world
- Full central archive *versus* purpose-driven distributed archives.
- Full central scheduling ("service mode") *versus* distributed control ("services mode").


South Africa
Lesotho
Maseru
Bloemfontein
Kimberley
Cape Town
Port Elizabeth
East London


Malang
Dili
Northern Territory
Queensland
Australia
Western Australia
South Australia
Perth
Brisbane
New South Wales
Sydney
Adelaide
Canberra
ACT
Victoria
Melbourne

Processing architecture



QPACE@Jülich: The most energy efficient Supercomputer of the world

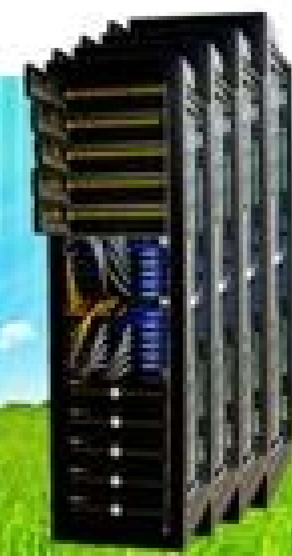
Nr. 1
November 2009

TOP Green500 Top 10
November 2009

- #1 Forschungszentrum Juelich (FZJ)
- #1 Universitaet Regensburg
- #1 Universitaet Wuppertal
- #4 DOE/NNSA/LANL
- #4 IBM Poughkeepsie Benchmarking Center
- #6 DOE/NNSA/LANL
- #7 National Astronomical Observatory of Japan
- #8 National SuperComputer Center in Tianjin/NUDT
- #9 King Abdullah University of Science and Technology
- #9 EDF R&D
- #9 Ecole Polytechnique Federale de Lausanne
- #9 IBM - Rochester
- #9 IBM Thomas J. Watson Research Center
- #9 Max-Planck-Gesellschaft MPI/IPP

QPACE@Jülich 208 TFlops
2048 Processors PowerXCell 8i
240 Kilowatt

THE GREEN 500




SKA
SQUARE KILOMETRE ARRAY
Let’s make it happen...
together!