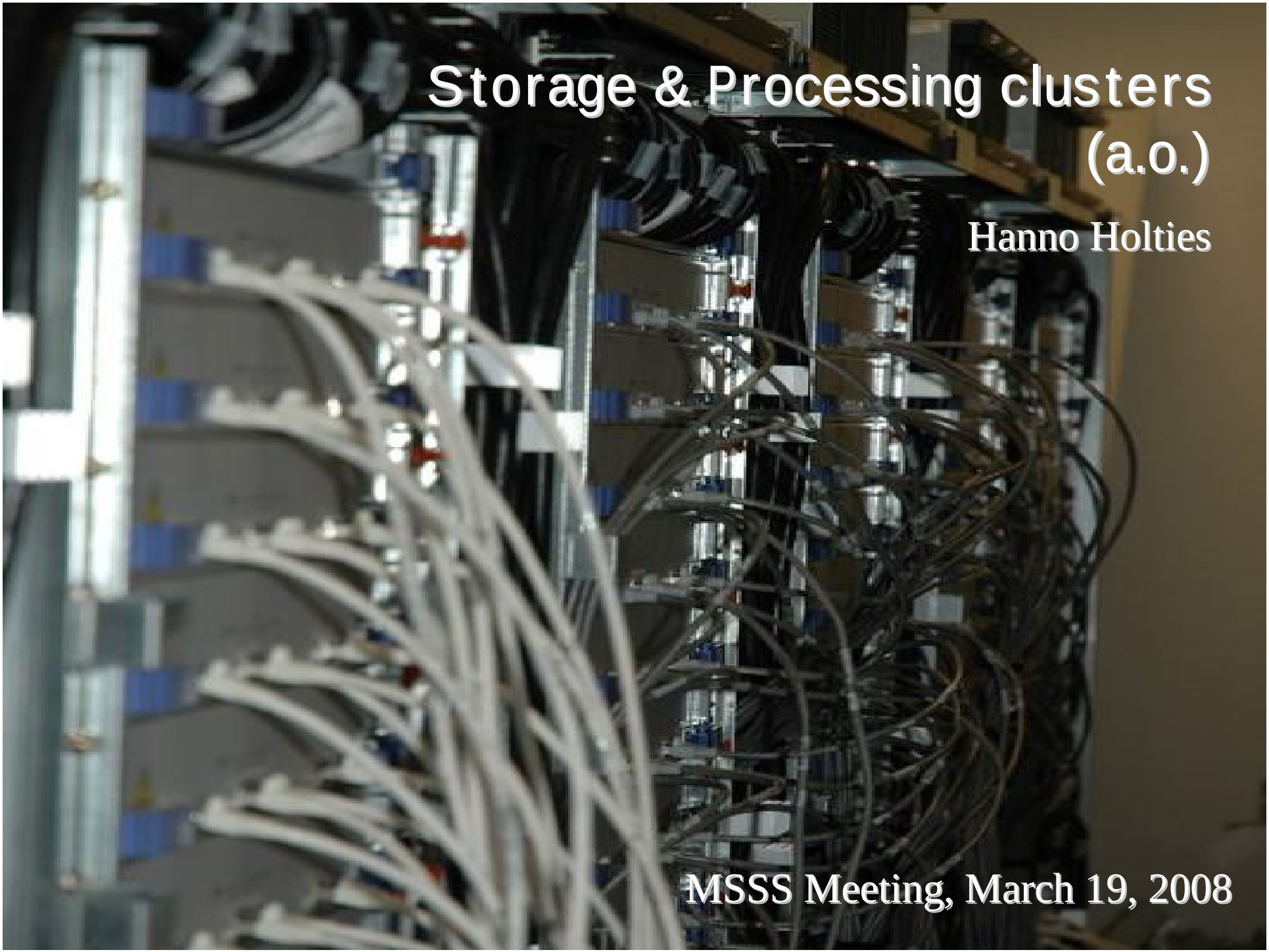
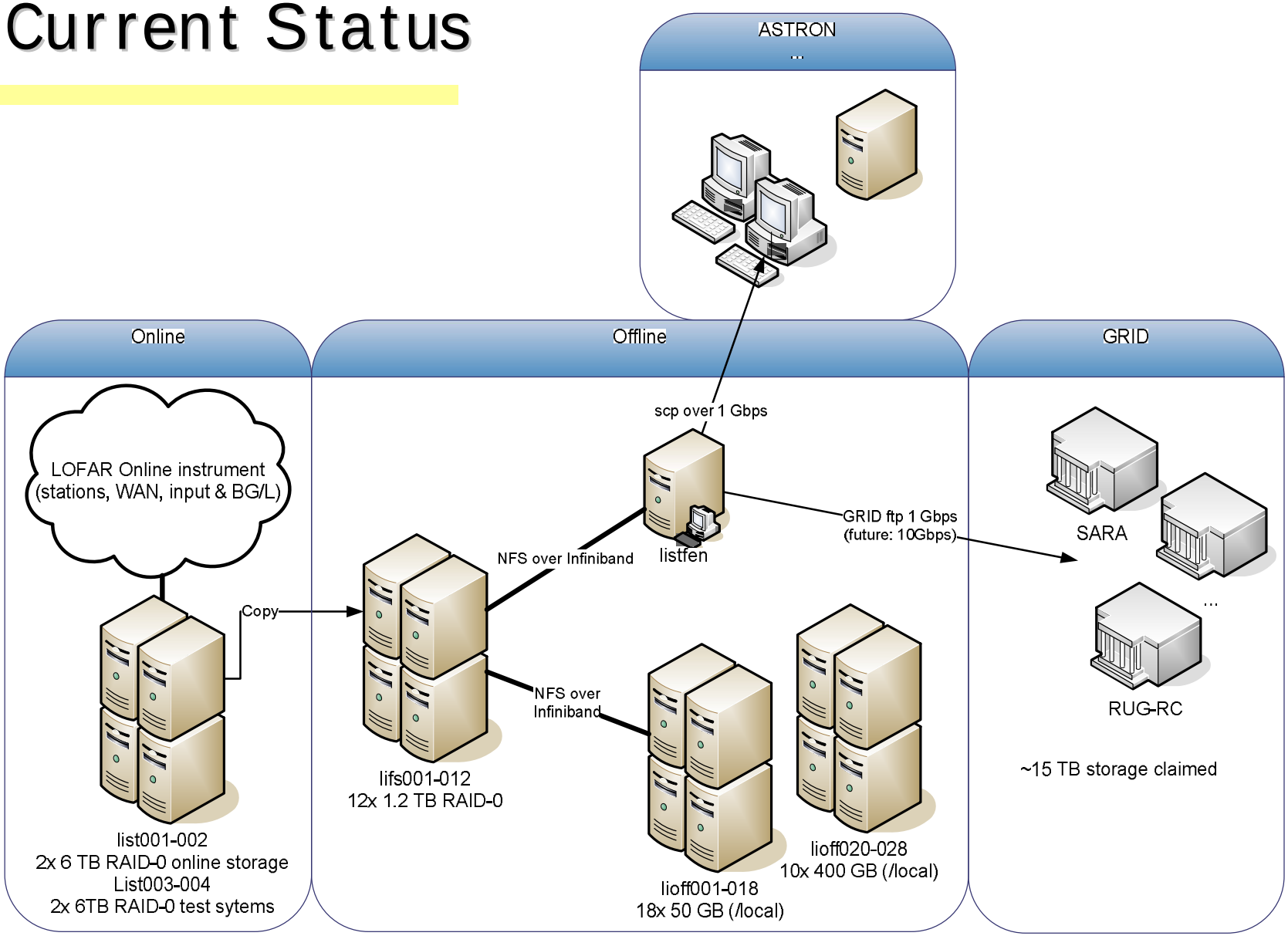


Storage & Processing clusters (a.o.)

Hanno Holties

MSSS Meeting, March 19, 2008

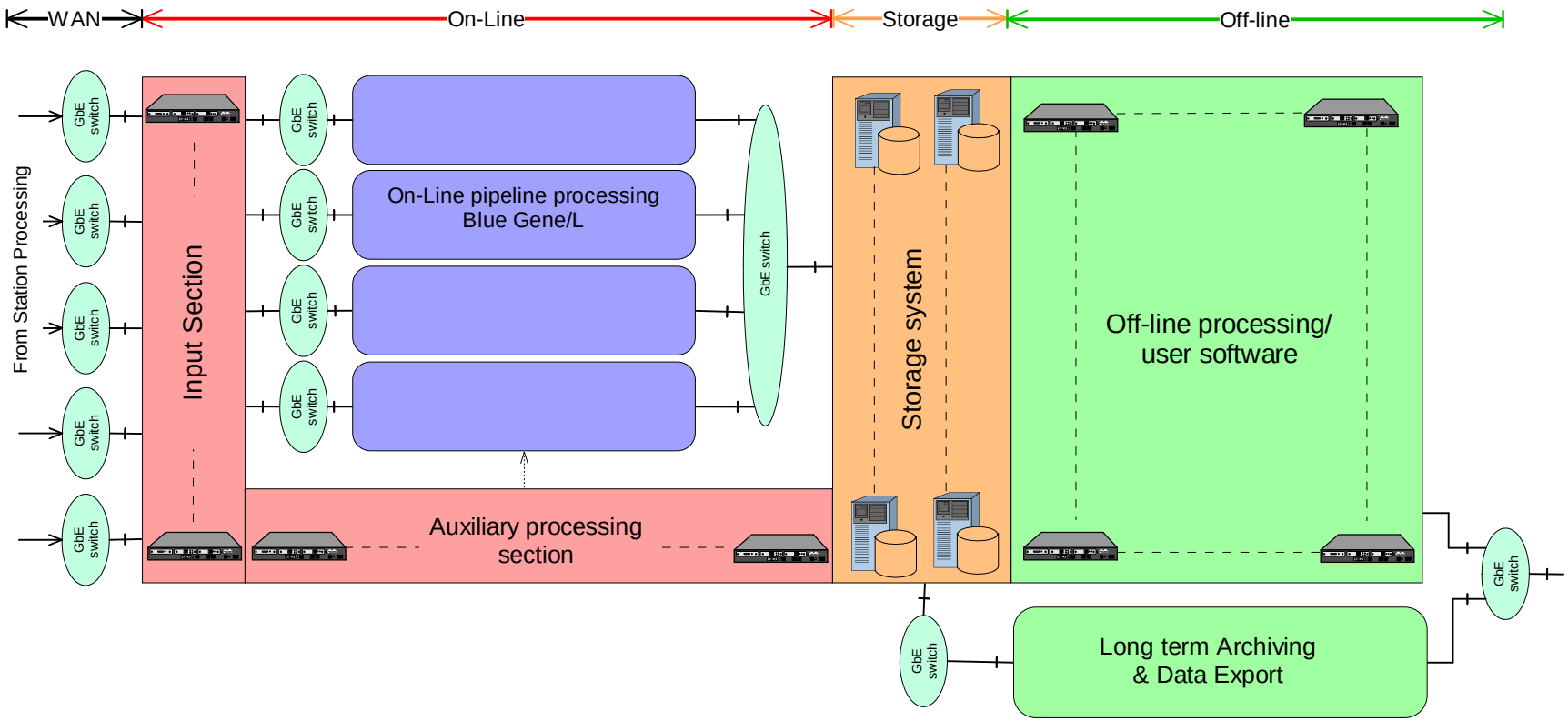
Current Status



Current Status

- Online storage CIT: 24 TB
- Offline storage CIT: 14 TB
- Offline processing:
  - 18x 32 bit Xeon; 50 GB local
  - 10x dual 64bit single core Opterons; ~400 GB local
- GRID based archive
  - According to current policy mostly migrated to tape
  - ~10 TB in use
- GRID based processing
  - SARA Matrix cluster (36x Dual Xeon)
  - CIT GRID cluster (50x 2 Dual Opteron)
- Local inspection
- Manual scheduling
- Manual data handling & inspection

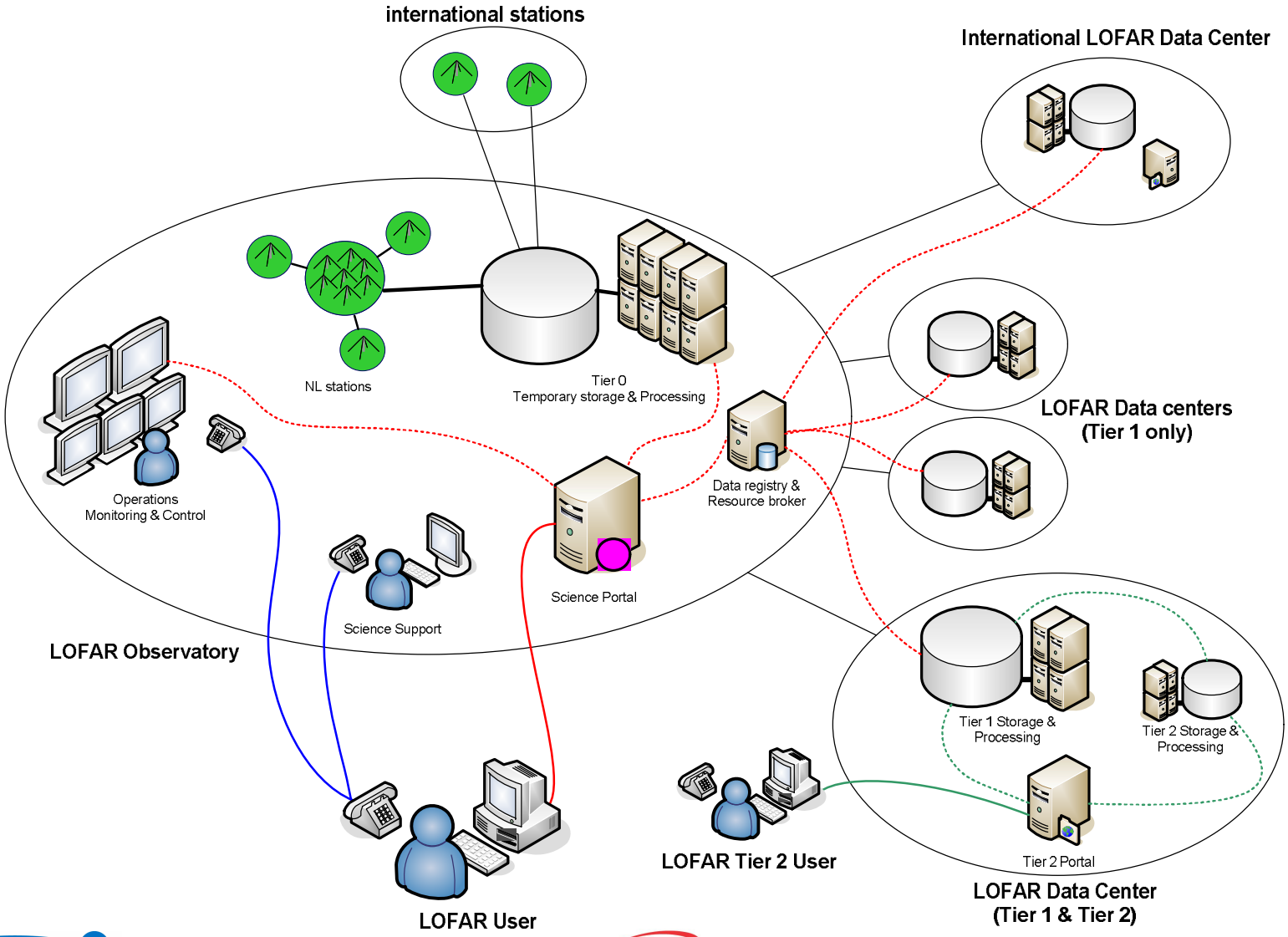
Working towards…



Spring 2009 Status…

- CEP storage & offline processing procured (Phase 1)
  - Specs being defined
  - Design includes 50 Gbps input & 100 Gbps output datarates
  - Multiple 10 Gbps connections to computing centres
- GRID based storage & processing procured
  - BiGGrid second tranch end 2008

Working towards…



Spring 2009: What is needed for MSSS

- MSSS requirements (?)

  Assuming “100% efficiency”

  - Data generation ≤ 6 Gbps (≤ 500 TB/week)
  - Reduced (post DP^3) data ≤ 15TB
  - Offline computing power ???

- Operations:
  - Minimum requirements for system availability?
    - # stations?
    - # antennas/station?

Technical challenges

- Getting the cluster specs right…
- 30,000+ Observations to be scheduled
  - Including offline processing
  - High level of automation required
- Observation catalogue
- 10,000+ Data products to be archived
  - Images
  - Reduced UV data?
  - Raw UV data???
  - Public?
- ~2,500,000 Sources in catalogue

Sleepless nights

- Planning & integration
- System robustness & stability
- Is system “Operations” ready?