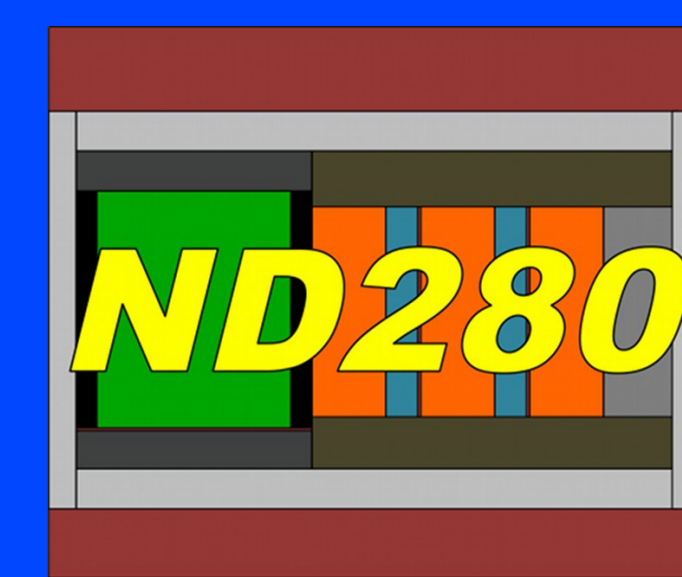




Evolution of the T2K-ND280 Computing Model

Thomas Lindner (TRIUMF) for ND280 Computing Group

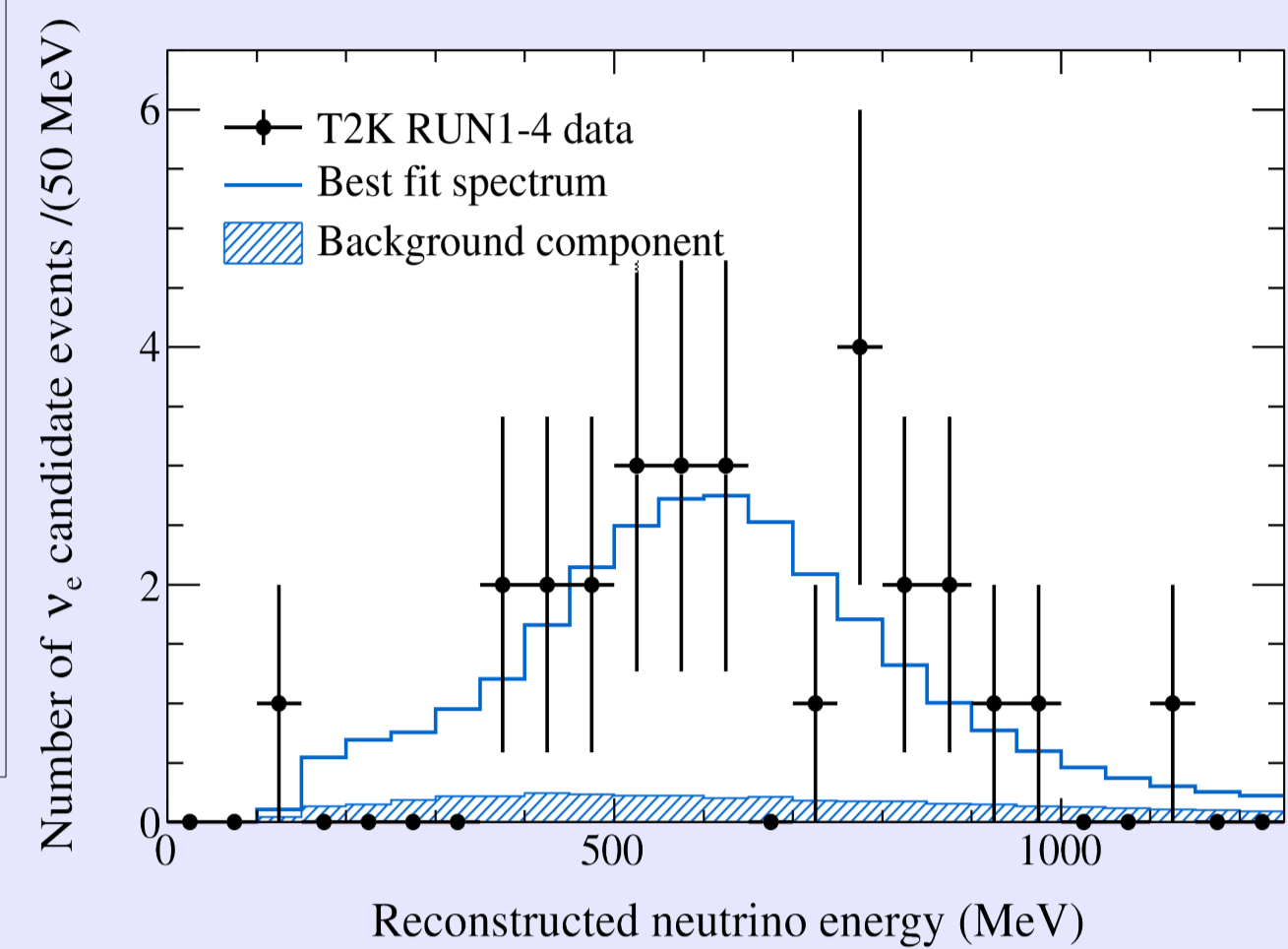


T2K Experiment

- Long baseline off-axis neutrino oscillation experiment.
- High purity beam of ν_μ produced at JPARC proton accelerator in Tokai, Japan.
- Measure the rate of ν_μ and ν_e 295 km away at the Super-Kamiokande water Cherenkov detector.

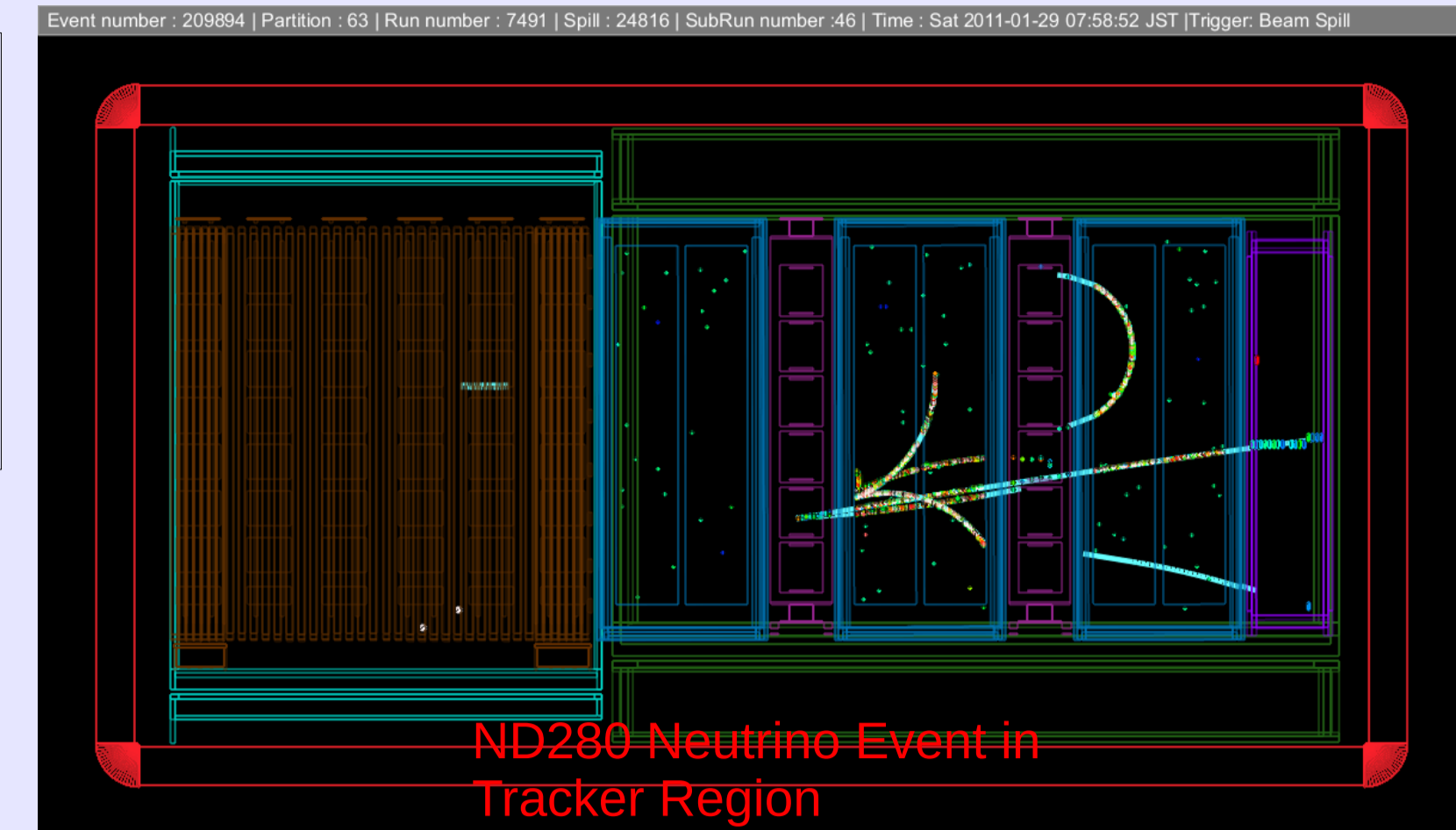
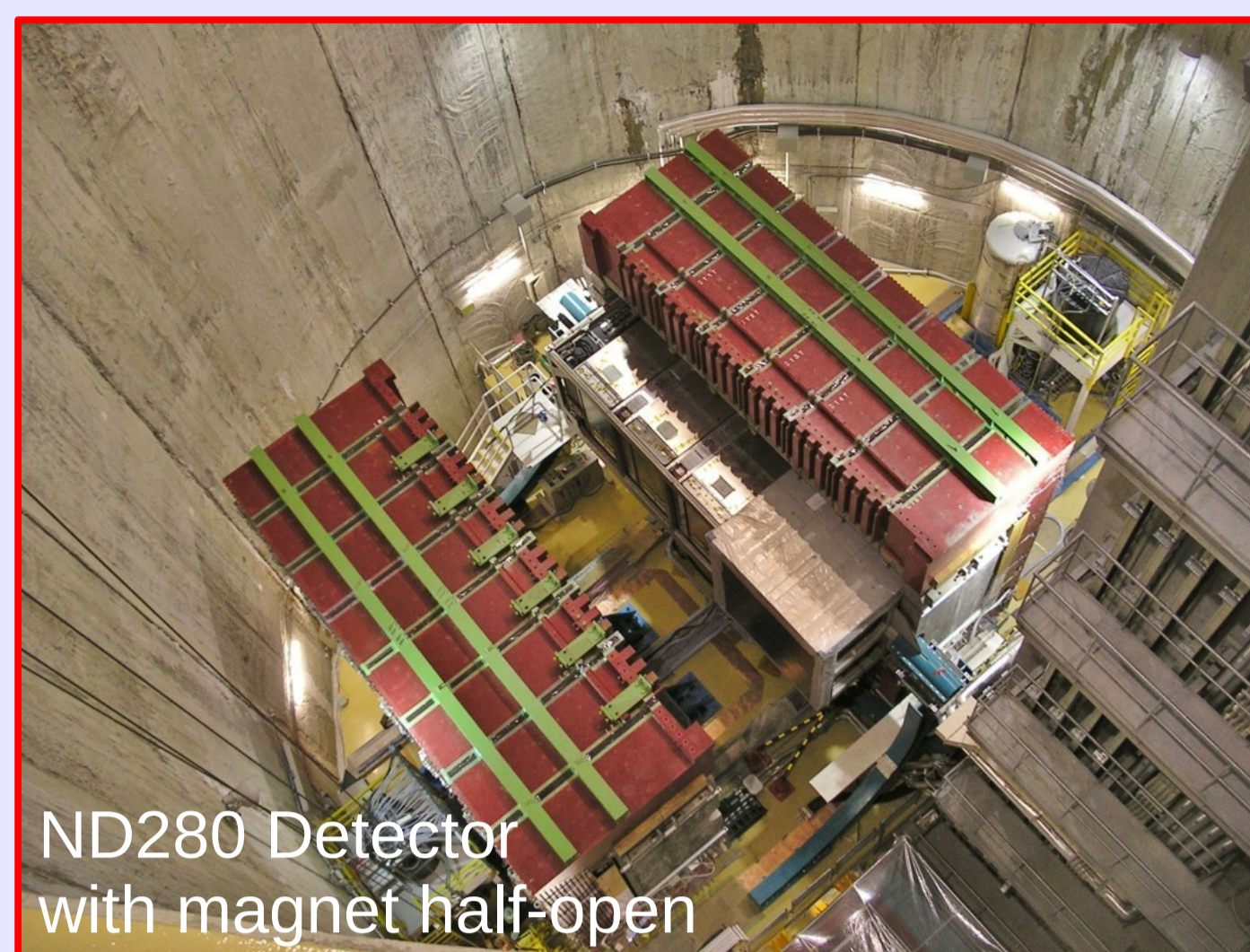
• Allows extraction of neutrino oscillation parameters:

- $\theta_{13}, \theta_{23}, \Delta m_{32}^2, \delta_{CP}$
- Recently presented first observation of ν_e appearance in ν_μ beam (7.3σ effect).



ND280 Detector At JPARC

- Magnetized neutrino detector, 280m from nu target.
- Critical for characterizing neutrino beam flux and composition before oscillations.
- Technology: uMEGAS TPC + scintillator with MPPC.



ND280 Software

- Standard HEP software suite using Geant4 for simulation and ROOT for analysis and I/O.
- NEUT and GENIE for neutrino interaction simulation.
- Custom reconstruction, with RecPack toolkit.
- CVS, CMT and bugzilla for code management and development.

Data Processing and MC Production

Data processing and MC production occurs around world:

Europe:

- 6 sites in UK (with 50% at RAL-LCG2)
- Standard EMI/LCG tools: WMS job submission to ARC-CEs
- output to local SE, FTS replication back to RAL-SE.

North America:

- Bugaboo and Scinet clusters in Canada (batch queues)
- Colorado CMS Tier-3 Cluster (batch queue)
- Files transferred back to TRIUMF-SE with lcg-cp and globus-url-copy.
- Summer 2013 production used ~2500 core-years (HEPSPEC06 normalized).
- Typically have 1 production per year.

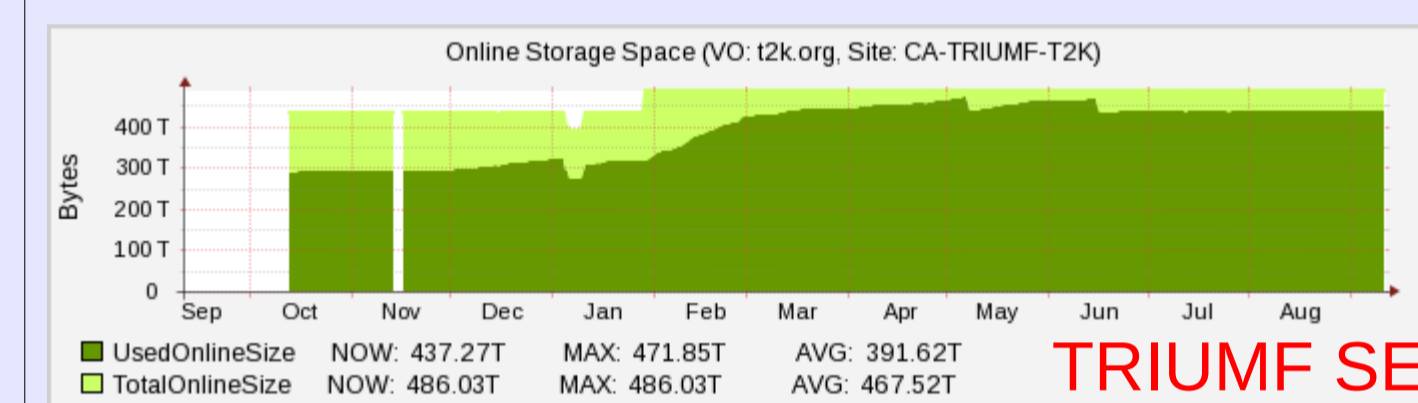
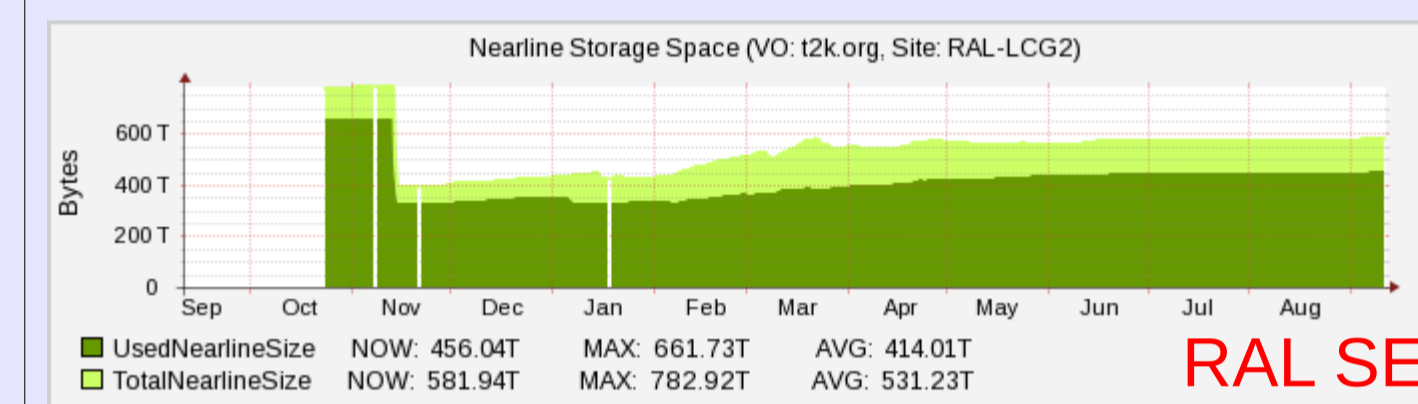
Key Production Challenges:

- Production work is relatively manpower intensive, because need different skill sets for work at different sites.
- Standard LCG toolset has presented many difficulties, in particular with WMS and Proxy Management.
- North American production is often limited by the rate at which output files can be moved back to TRIUMF SE.
- Limited grid support in many institutions and minimal dedicated computing manpower.
- No unified view of production status.

Data Distribution & Management

Principal Data Storage

- A pair of large data storage elements are core of ND280 data management:
- RAL-LCG2: ~700TB allocation at RAL Tier-1 (GridPP-supported)
- CA-TRIUMF-T2K: ~700TB SE, provided/maintained solely by T2K-Canada group.
- These SE provide interface between LCG-Europe and non-LCG-America.



Raw Data Distribution

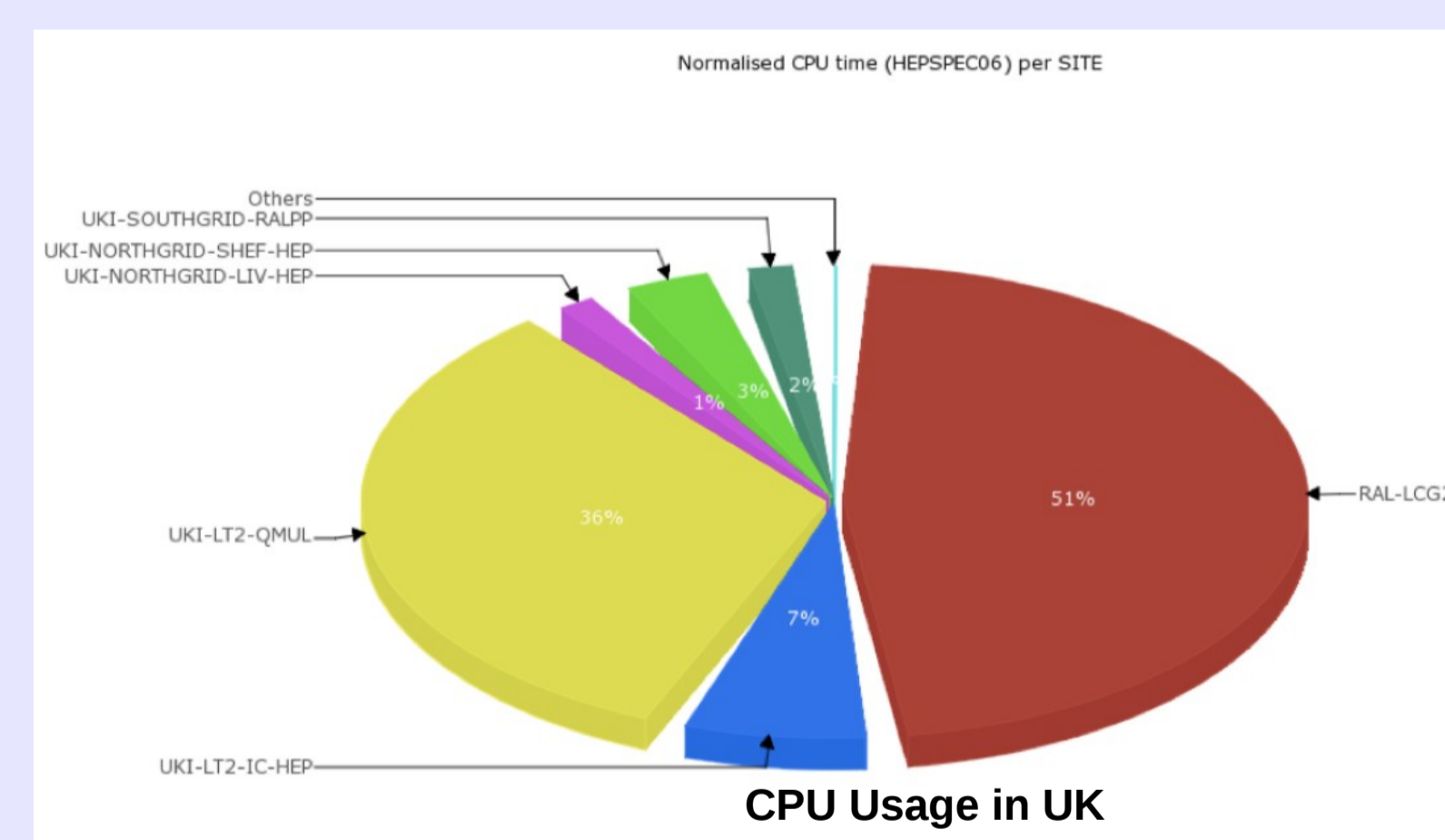
- ~140TB (including replicas)
- Data copied from JPARC to primary storage element at KEK Computing Center.
- Replicated with FTS to RAL and TRIUMF storage elements.

Production Data Distribution

- ~1.5PB (including replicas)
- Data distribution decoupled from production jobs.
- Data from production sites copied to RAL or TRIUMF
- Then fully synchronized between RAL/TRIUMF with FTS
- Users download the summarized data set (~10TB) for analysis.

Key Data Challenges:

- Large data set to manage: ~4% of ATLAS data set in Canada.
- FTS transfers don't provide LFC registering → can cause "dark data"
- Summary data set are still very large; difficult for analyzers.

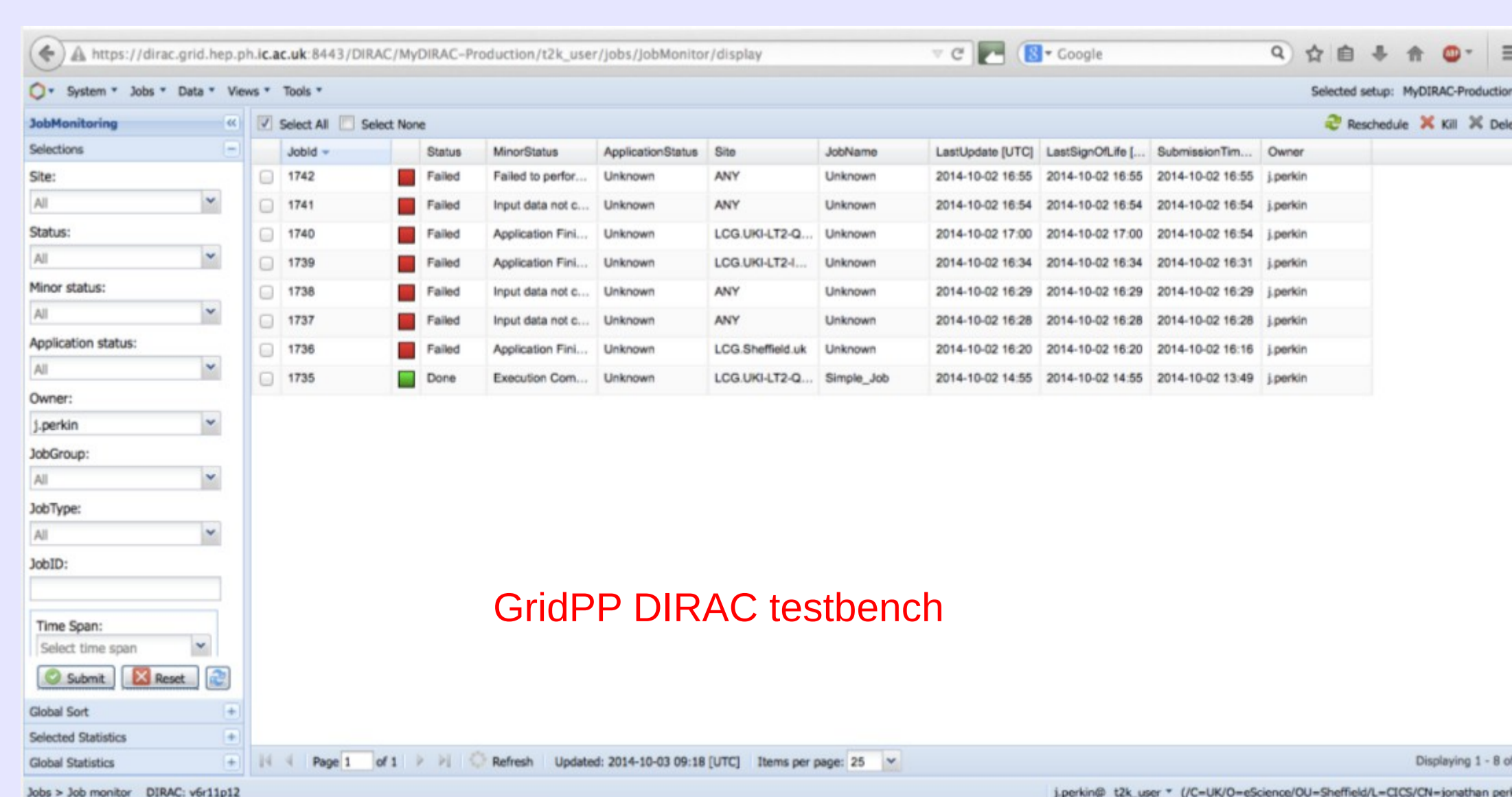


Ganga

- Production work is substantial burden for personnel.
- Have therefore started using Ganga to unify various types of ND280 computing task.
- Wrote several Ganga plugins to wrap ND280 software and data sets.
- Allows for easy transition between different computing environments (ie desktop PC to batch queue).
- Have transitioned over main data processing and control sample creation to Ganga scripts.
- Working on transitioning several other MC production tasks next.

DIRAC

- Default WMS are undergoing decommissioning: So have been working with GridPP colleagues on testing DIRAC as a replacement workload management system.
- May also help unify North American and European production work.
- Have gotten basic ND280 jobs running through DIRAC.
- Some issues with handling LFC naming of existing T2K datasets.
- Next steps will be large scale tests.



CVMFS

- Software installation for grid processing has moved to CVMFS.
- Working now on deploying CVMFS at North American processing sites.

Future Directions

- T2K has started collecting data in anti-neutrino mode.
- Reduces somewhat pressure on computing, since event multiplicity is lower in anti-neutrino mode.
- On other hand, need to continue to be able to adapt to changes in grid tool ecosystem, amid declining computing manpowers.