Data Mover / Grid

Henry Nebrensky

Brunel University

MLCR data handling

MICE DAQ data is stored by the Event Builder as a series of files per run on the Online Buffer. The corresponding Online Monitoring, Reconstruction and Data Quality histograms are created and stored on other PCs in the MLCR.

A badly-named file "compactor" process combines all of these into a single tarball per run saved on a large RAID array still in the MLCR.

These tarballs are then saved to the Grid by an automated process – (as will be the historical target data and EPICS Archive).

Grid Transfer Box

The sole formally-agreed route for access to data (DAQ output) is via the Grid.

The Grid Transfer Box (miceacq05) is located in the MLCR. It runs an autonomous agent that reads the data from the RAID system in the MLCR and uploads it to the Grid, in particular to the CASTOR tape system at RAL-Tier1 and the dCache disk store at RAL-PPD.

The Online Review process specifically flagged up issues with data volumes and access conflicts – random scp's off miceraid1 are Not Allowed.

Data

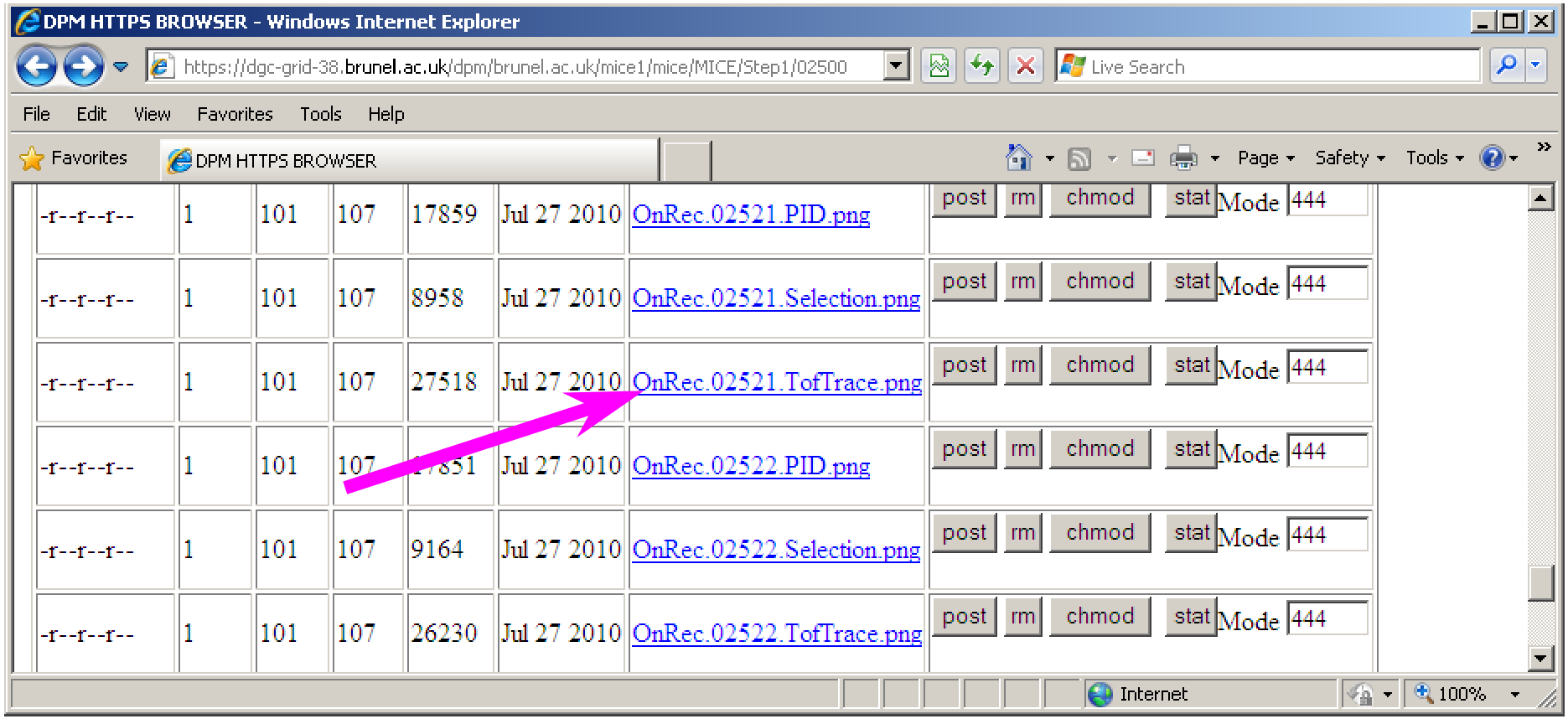
- Data is routinely and automatically being copied from micestore to the CASTOR tape robot and then on to RAL PPD, and beyond… (see CM27 talk)
- All DAQ data tarballs are available via the Grid, with the most recent data becoming available minutes after the shift ended (with a bit of luck).
- Owing to the limited disk space in the current server, only the most recent data will be available via the Grid/web interface at this time.
- Tarballs and plots all live in the /grid/mice/MICE/Step1/<*century*>/ LFC namespace

Online plots…

If you go to the Grid/web interface

https://dgc-grid-38.brunel.ac.uk/dpm/brunel.ac.uk/mice1/mice

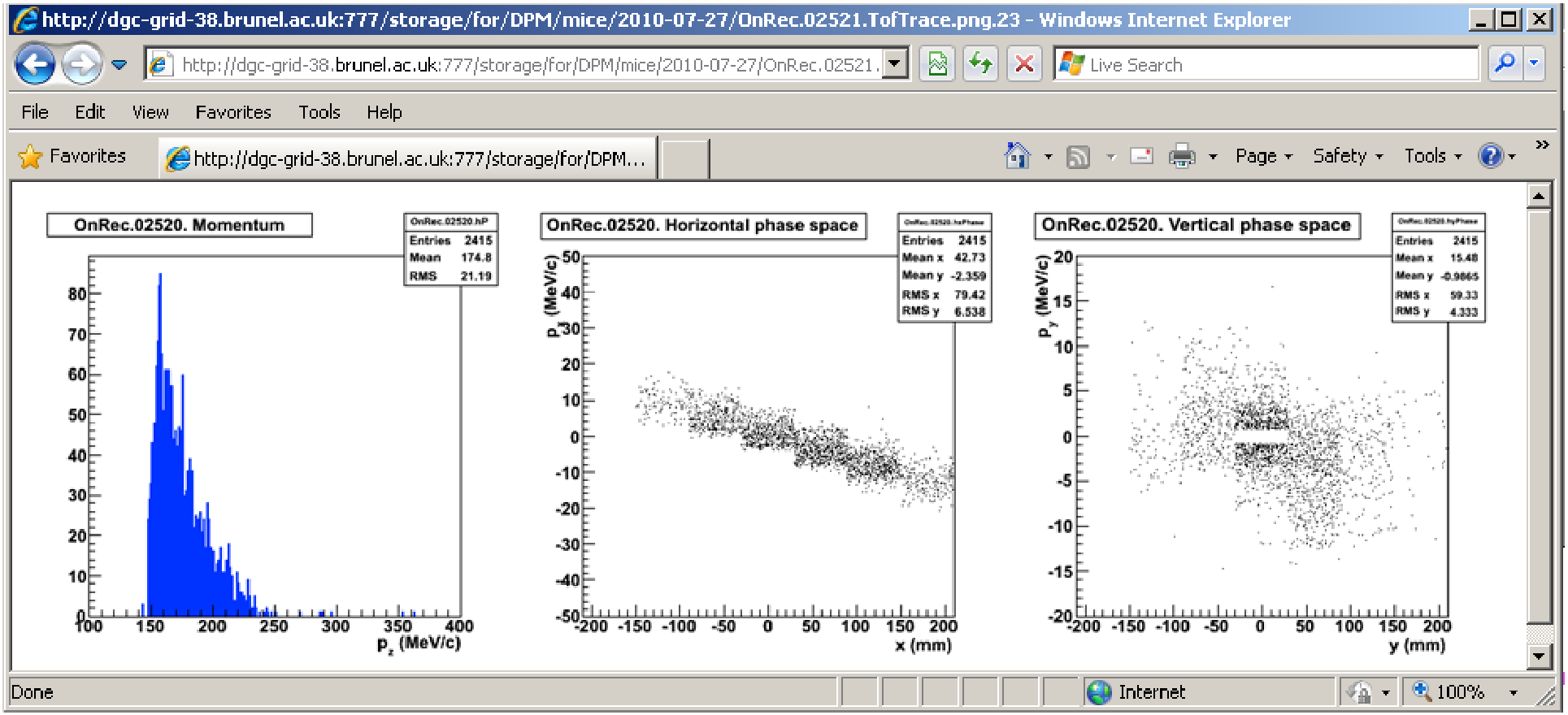
and follow the usual path to get to the most recent data:



… online

you can see the latest plots!

(currently Monitoring, Reconstruction, DataQualityCheck and Scalers)



Grid https interface and data dissemination

As said at CM27, we have a specialised Grid server at Brunel that provides a web-browser compatible interface.

This is still running on ancient (2003) hardware that is now failing.

It should be replaced with one or more off-site data stores, MICE still needs to make a decision about what data it wants kept and how it should be made available.

High-availability data server

Providing useful services at a collaboration level is not a trivial exercise that can be left to some postgrad with a bit of spare time.

Requires proper server-grade hardware (redundancy, hot-swapability) and a vendor support contract.

Data Distribution

- Need people to use it and say it's useful, or to not bother and save the time and resources (£3K) needed to provision the https interface properly.
  - "use it or lose it" (when it breaks)
- If we don't want this route, what do we want?
- We know from the Configuration DB server that from a November go-ahead it may be February before we take delivery.

Grid Computing

- At CM20 (Feb. '08) there were 8 UK sites + Sofia making CPU available to MICE.
- Now 11 UK sites + Sofia + Roma III - many 1000s of CPU

- OTOH, by CM21 most sites had G4MICE 1-9-5 already installed for end-users.
- Now only available on two *clusters*, Sheffield ce0 - O(100) cpus - and Brunel dgc-grid-35 - 4 cpus

- Rest have been lost as sites upgrade hardware and OS...

Reminder

Appearing on MICEmine (MICEnest?)

- Grid certificates and joining VOMS:

  http://micewww.pp.rl.ac.uk:8080/projects/computing-software/wiki/GridCertificate

- Data access coming soon

(Linked from the Software page:)

- Grid certificates and joining VOMS:

  http://people.brunel.ac.uk/~eesrjjn/mice/mice_auth.htm

- Data access:

  http://people.brunel.ac.uk/~eesrjjn/mice/mice_data.htm

- *Occasional* queries:

  micedataman@stfc.ac.uk

Getting at the Data

The big hurdle is getting your certificate in the first place. After that, data is accessible via web browser or Grid clients. Online Monitoring histos are at

`/grid/mice/MICE/Step1/...`

and the DAQ tarballs from run 1300 onward are temporarily at

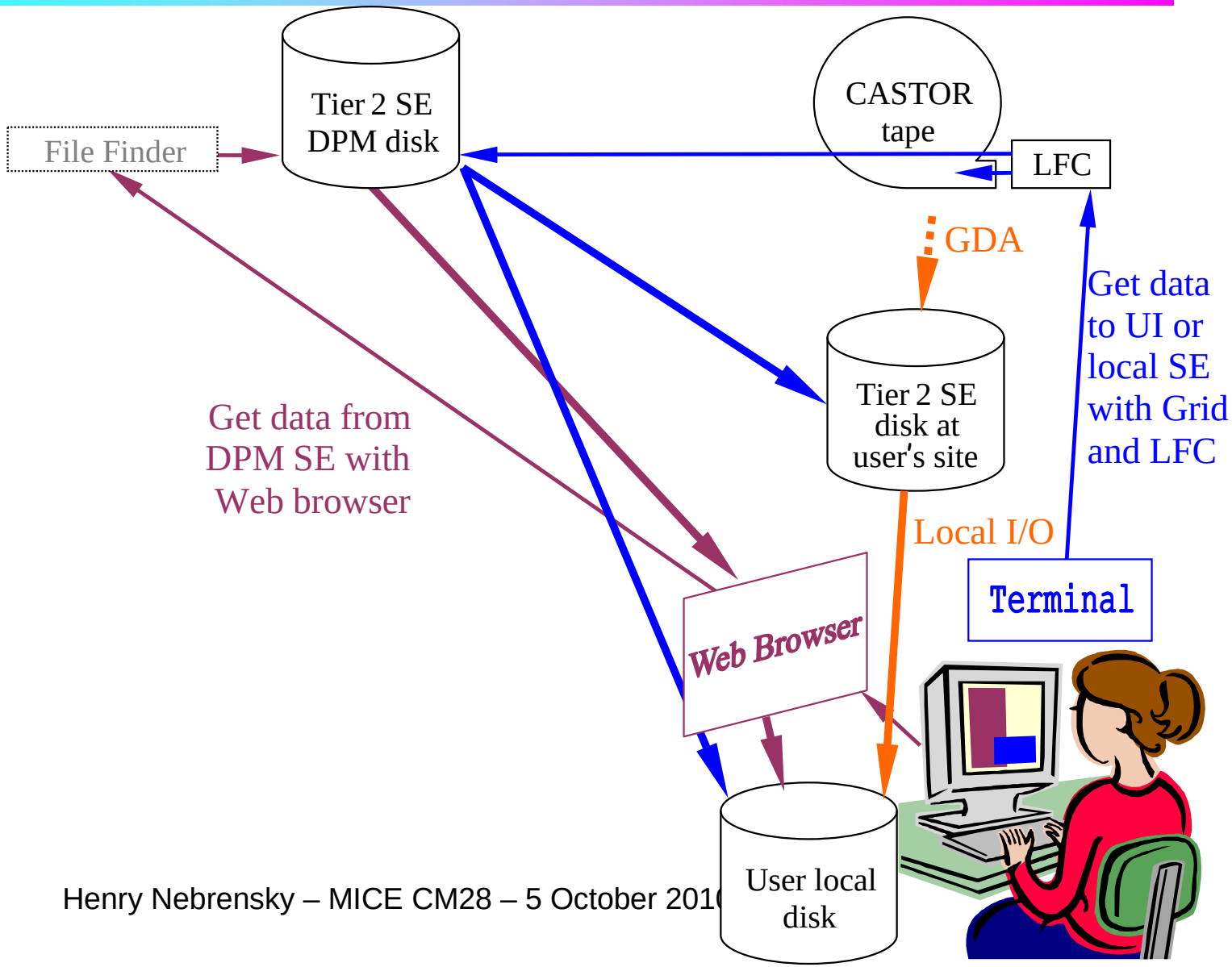
`/grid/mice/users/Nebrensky/MICE/Step1/...`

Instructions are at

`http://people.brunel.ac.uk/~eesrjjn/mice/mice_grid.htm`

At least 7 MICE have managed to get real data off the Grid, and it wasn't as painful as they'd expected!

Getting at the Data (2)



Getting at the Data (3)

Grid UI

- you need gLite UI installed
- that means Scientific Linux or similar

Web Browser interface

- platform independent
- hardware and support needed, else will break. Again.

Local I/O

- research groups need to provide storage
- either Grid-visible storage, or need to run GDA
- GDA will need someone to look after it (and better name)