Scientific Data Management in the LHC Era

Fifth Workshop on Data Preservation and Long Term Analysis in HEP, Fermilab

Elizabeth Sexton-Kennedy

Data Management

- Data Storage and Types
- Data Movement
- Data Use
- The Future

GRID = LHC Computing

- LHC experiments started with a fully distributed environment where the vast majority of the resources are located away from CERN.
- LHC Computing Grid was approved by CERN Council Sept. 20 2001
- It evolved into the Worldwide LCG (WLCG) with service support for all 4 LHC experiments

Computing Model

- LHC computing models are based on the MONARC model

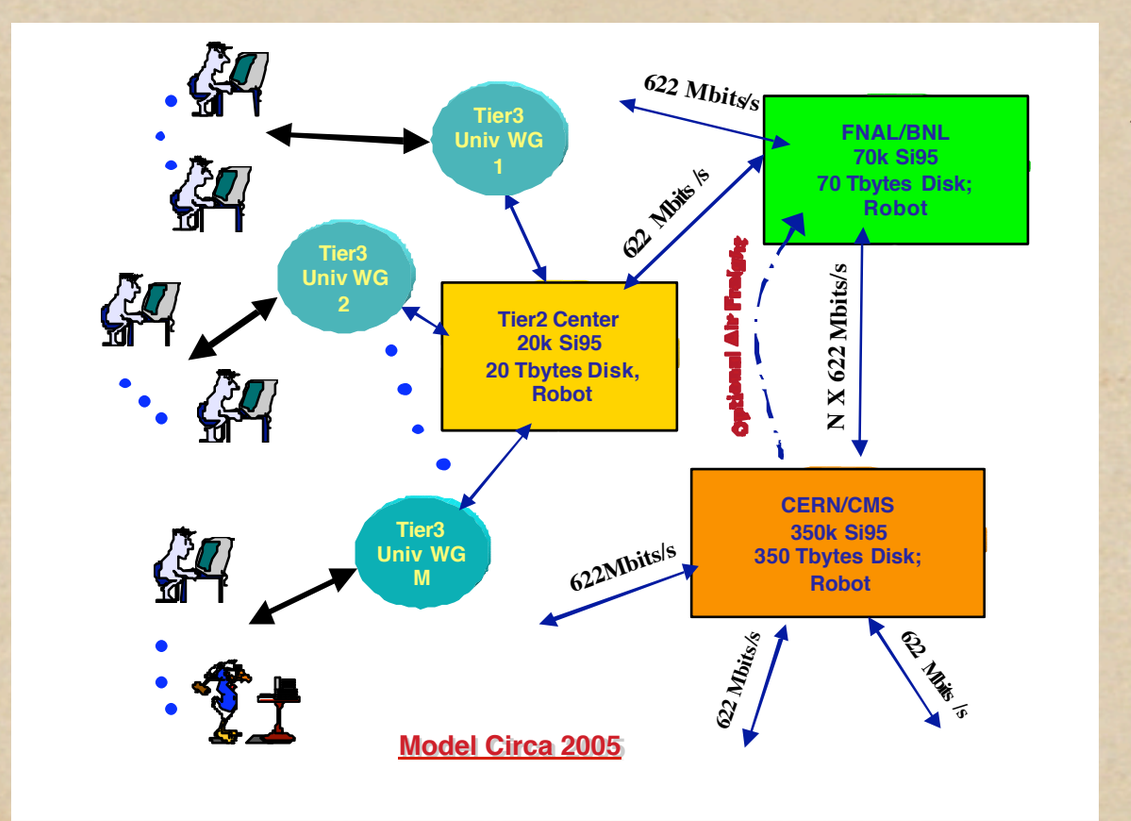
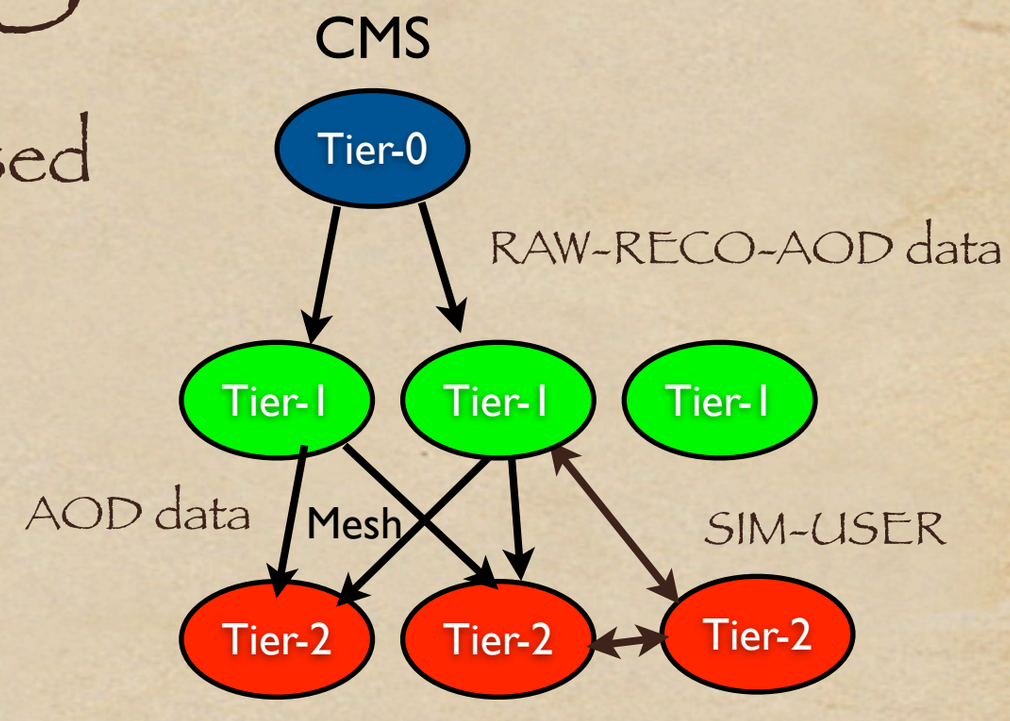
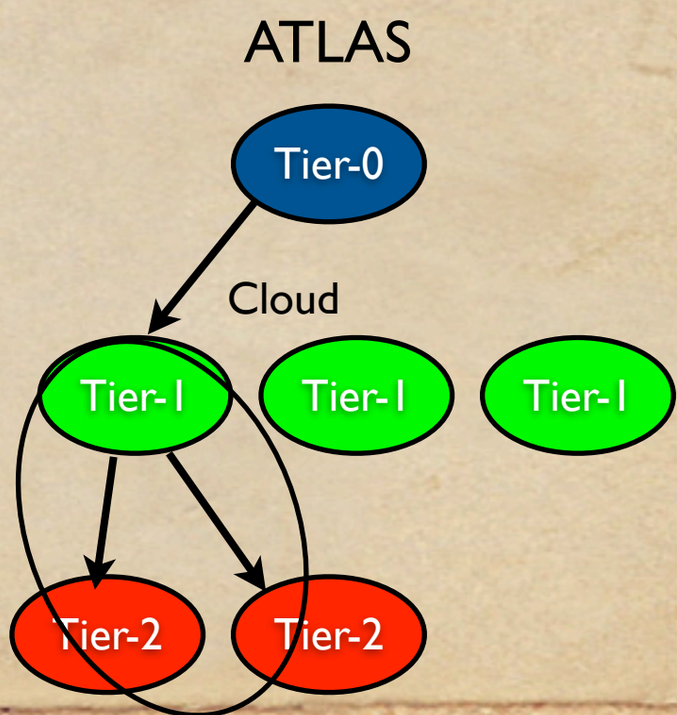
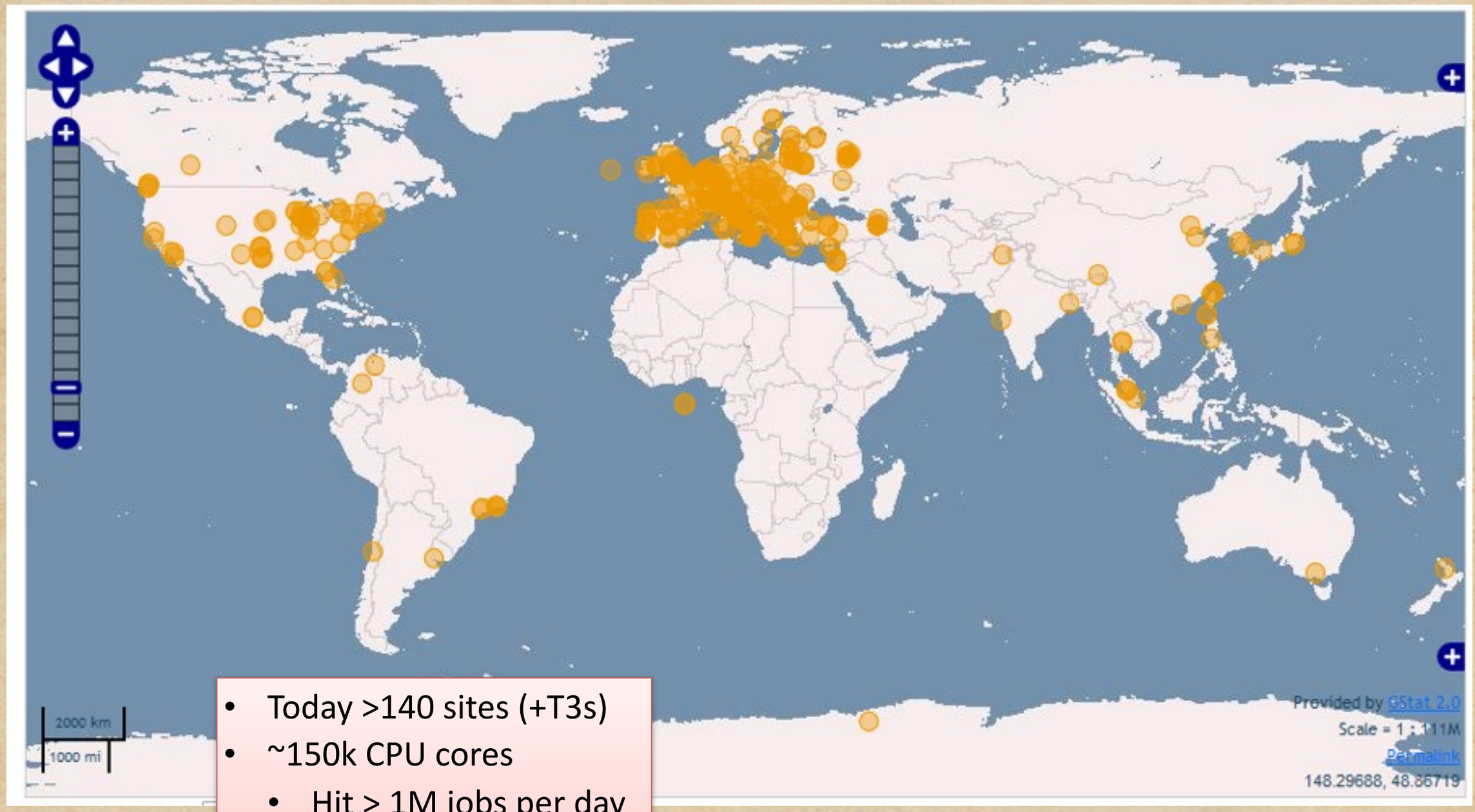


Fig. 4-1 Computing for an LHC Experiment Based on a Hierarchy of Computing Centers. Capacities for CPU and disk are representative and are provided to give an approximate scale).





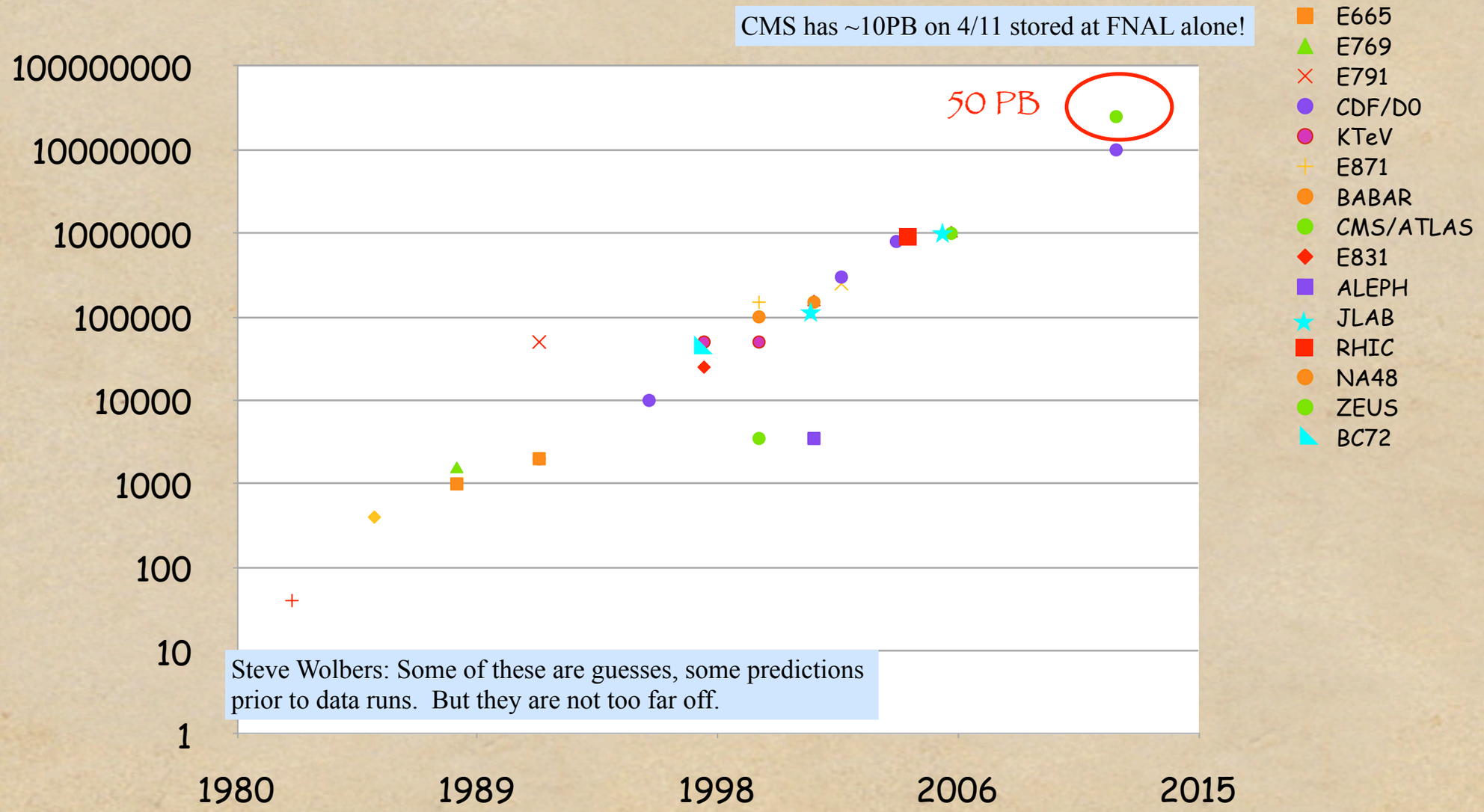
LHC Computing



- Today >140 sites (+T3s)
- ~150k CPU cores
  - Hit > 1M jobs per day
- >50 PB disk

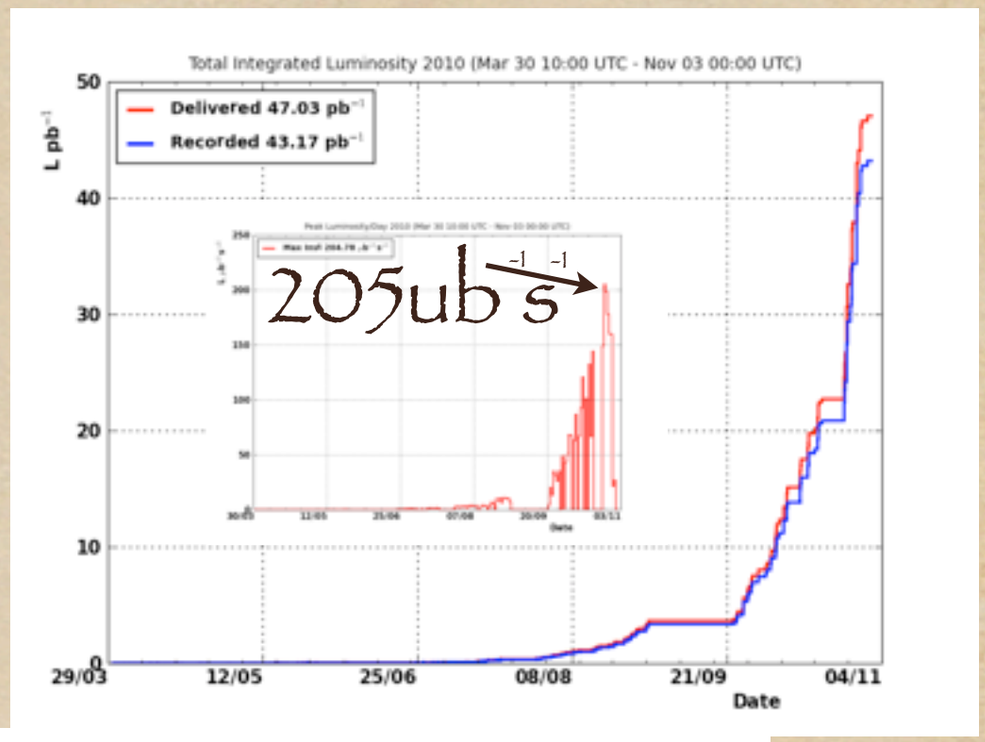
Exponential Growth in Size

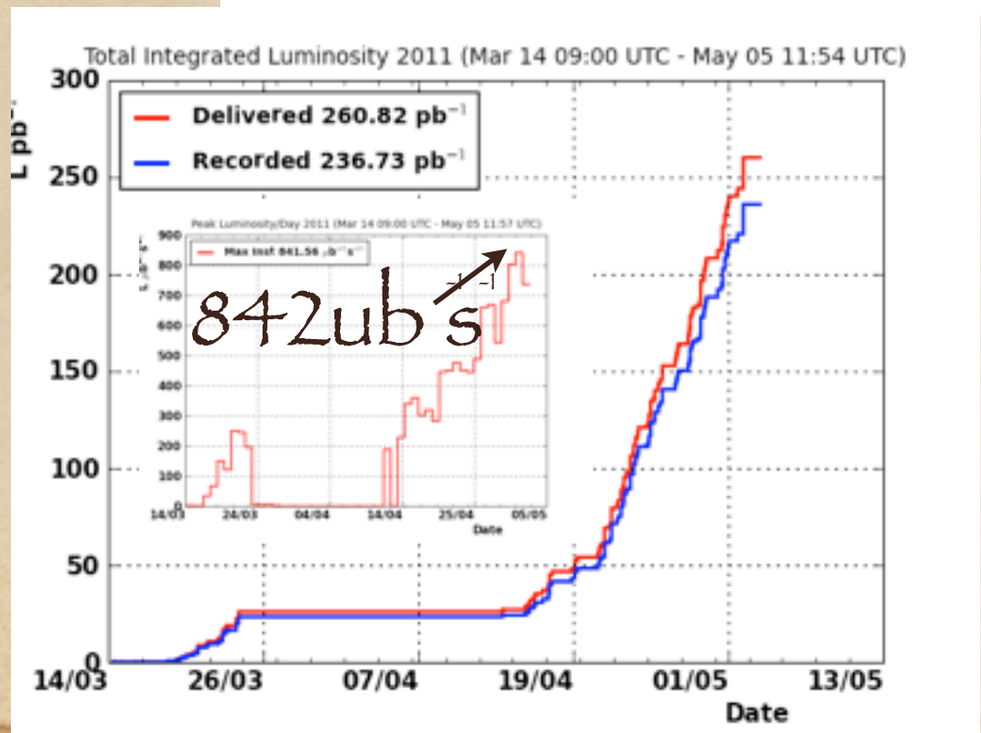
Data Volume by Experiment, in GBytes



The Challenge of Success

- Exponential growth rate of data delivered by LHC is especially challenging
- The full 2010 dataset can be delivered in one store now.





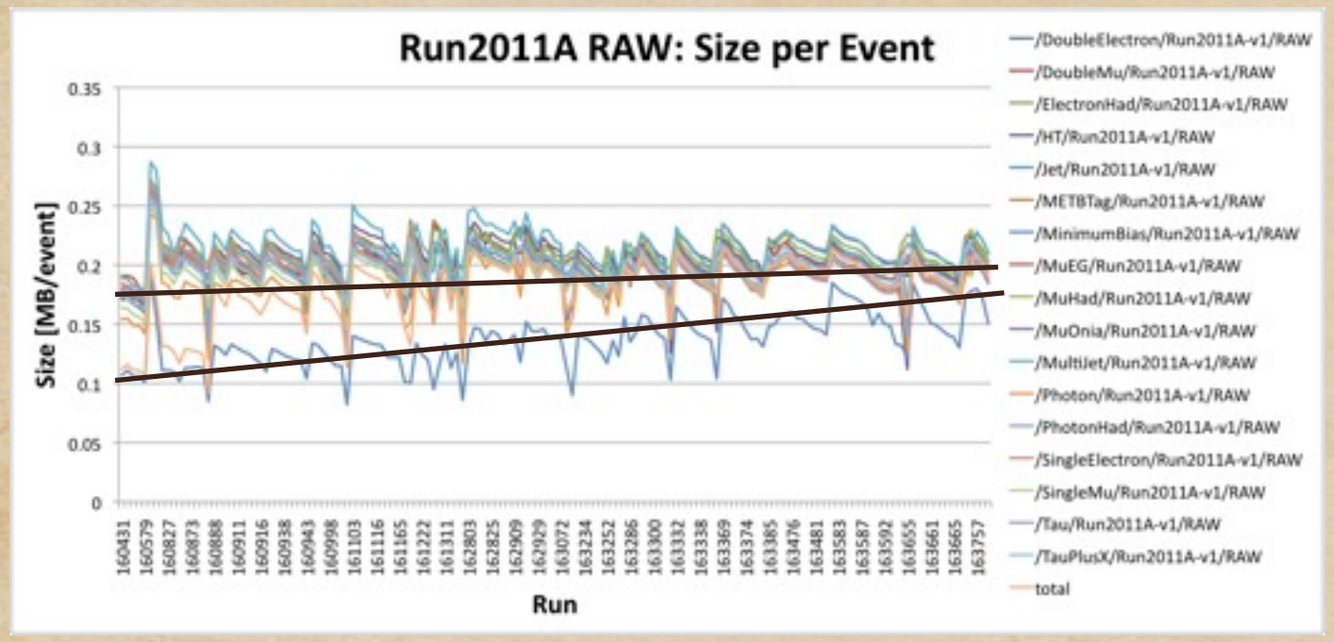
Data Set Growth

- In the absence of pileup, data accumulation is just a function of up time of the accelerator, since we keep the trigger rate constant at 300-400 Hz.
- However pileup effects are significant

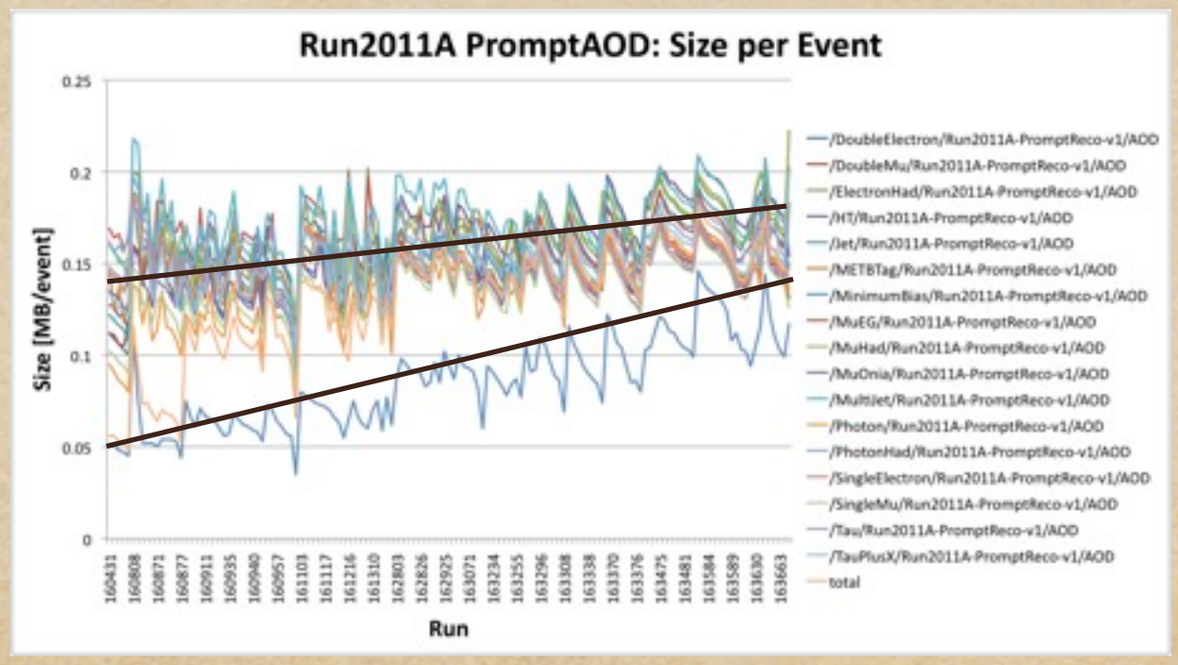
Types of Data (Tiers)

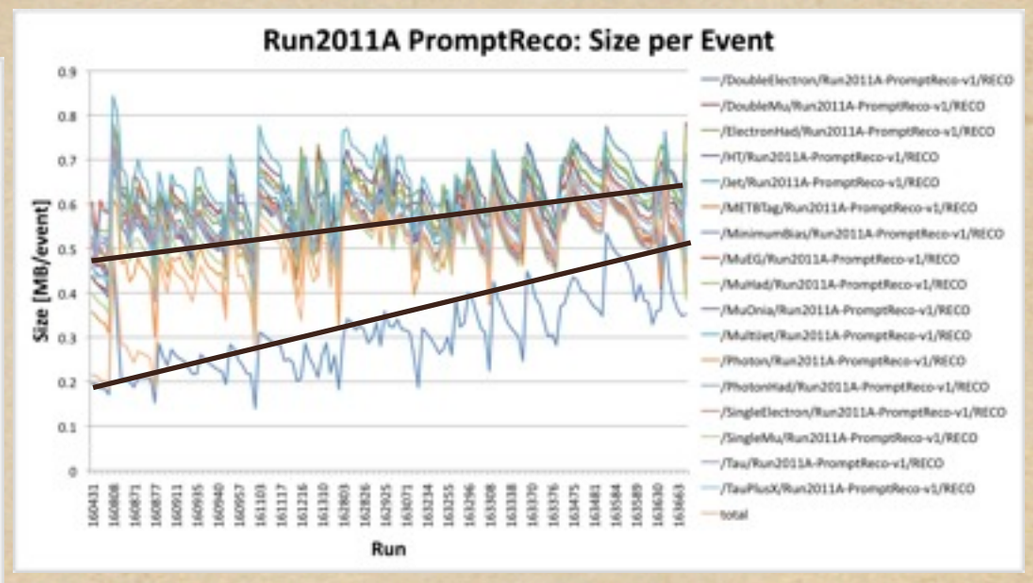
- The online system writes raw data in streams to a temporary root binary format.
- This format is then "repacked" so that the stream is broken out into it's component "Primary Datasets" and reformatted into an simple archival (readable forever) root TTree file. = RAW
- Reconstruction passes create RECO and AOD files, which can be further skimmed, slimmed, and thinned = USER.

Data Size by Tier



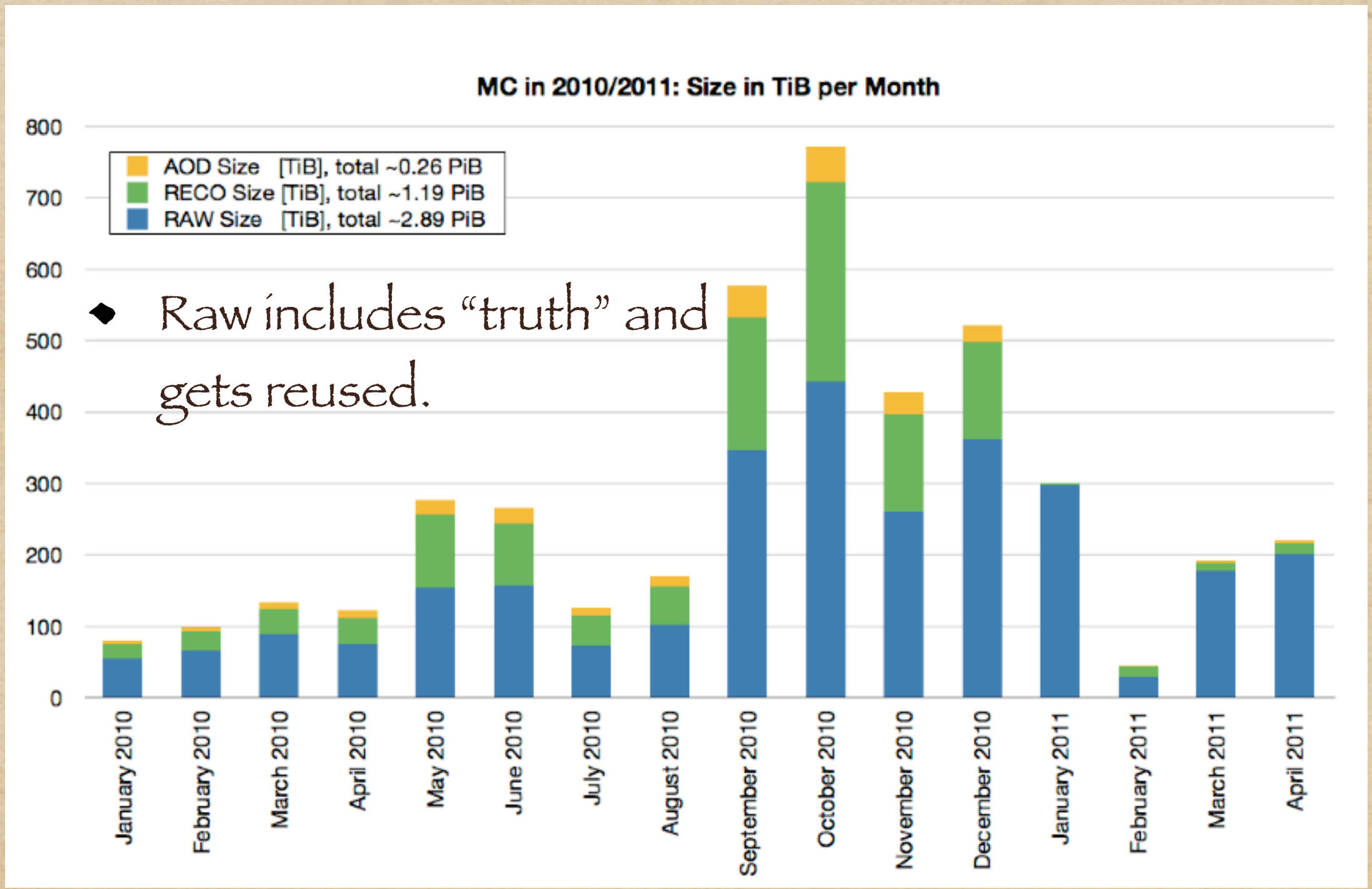
- Data size is a function of primary dataset and instantaneous luminosity





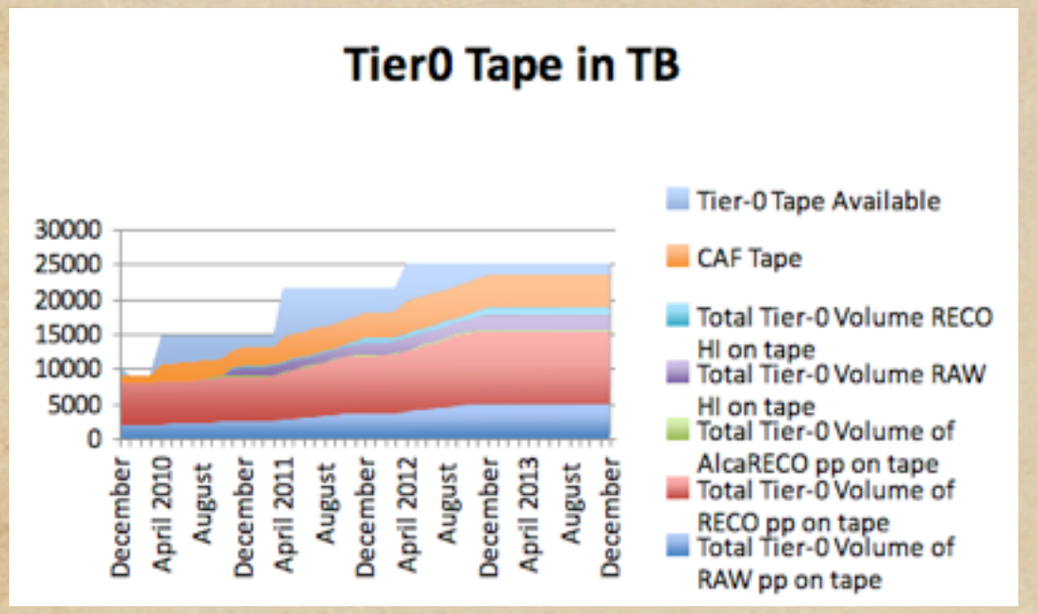
Plots generated by Oliver Gutsche

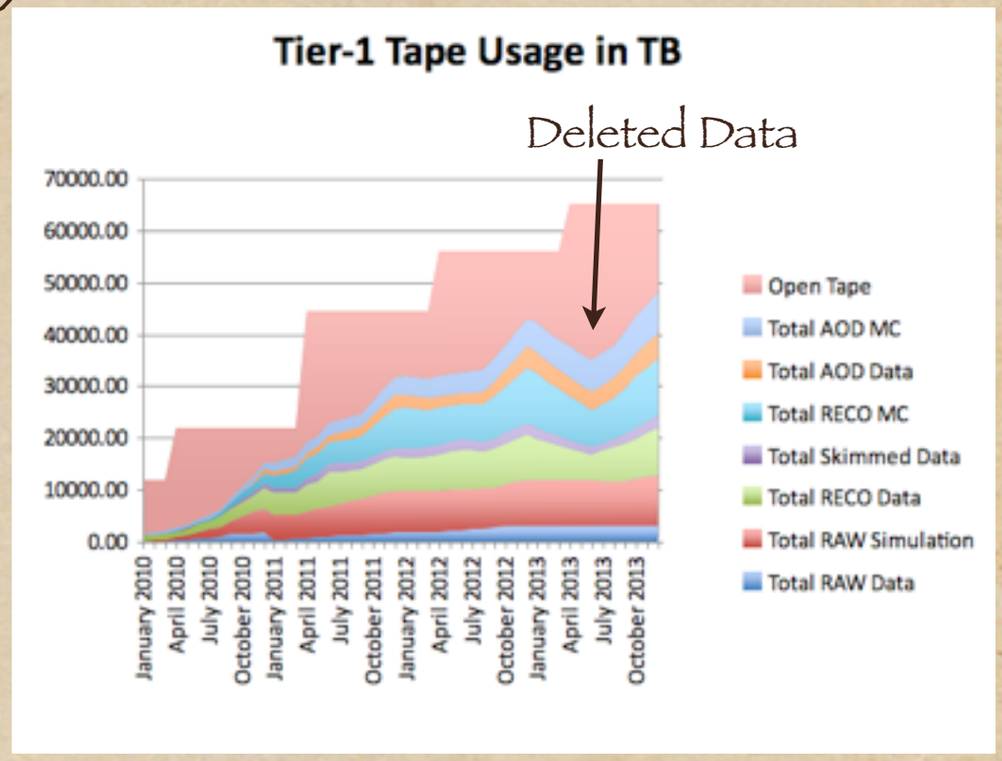
MC Data Size

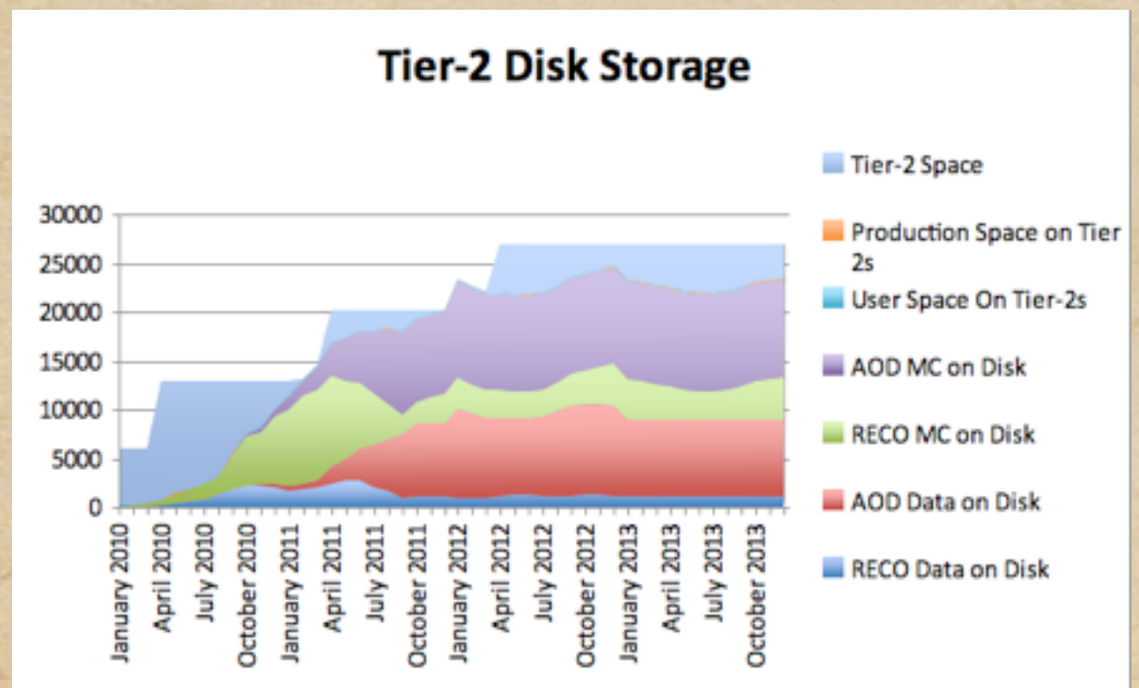


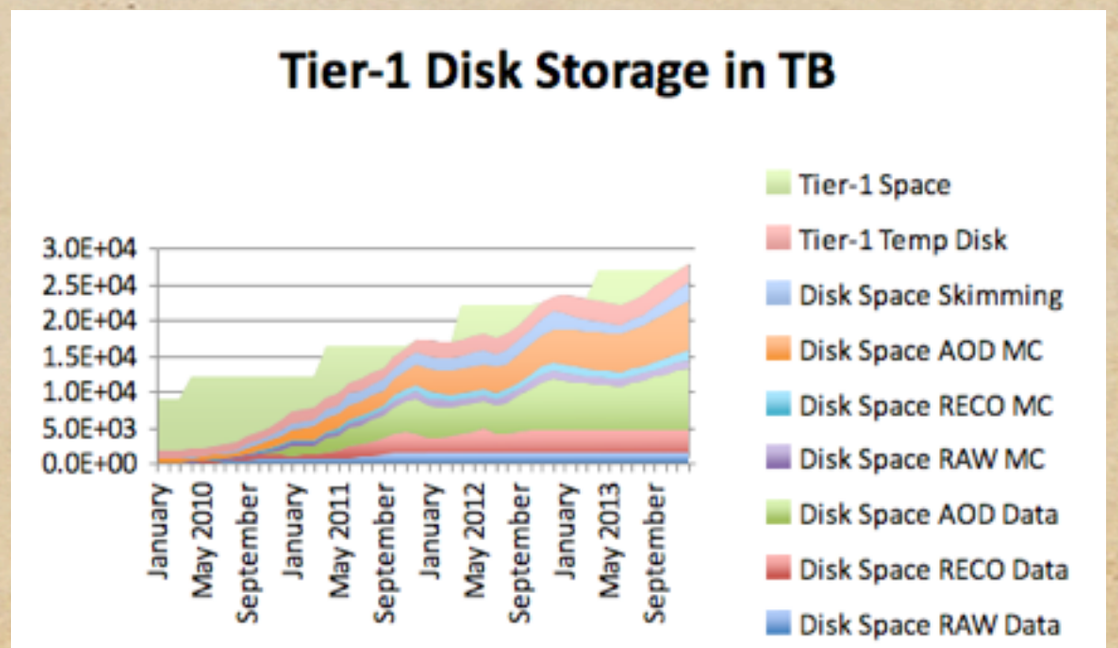
- Raw includes "truth" and gets reused.

Usage Projections



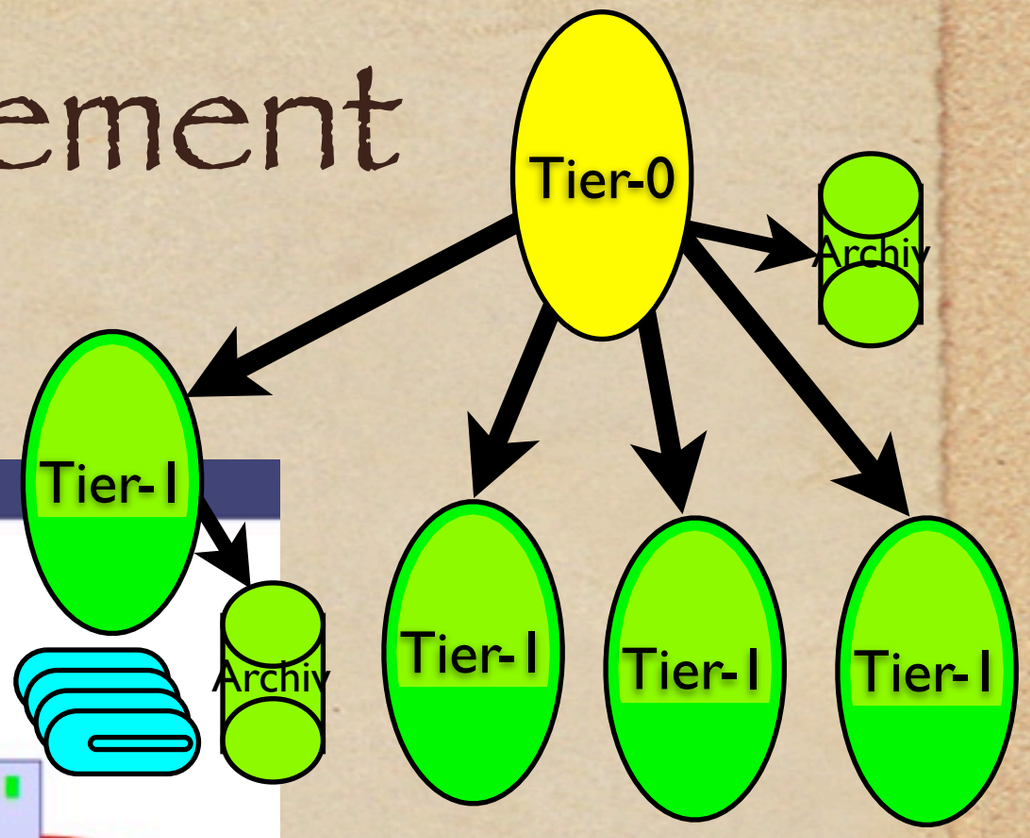


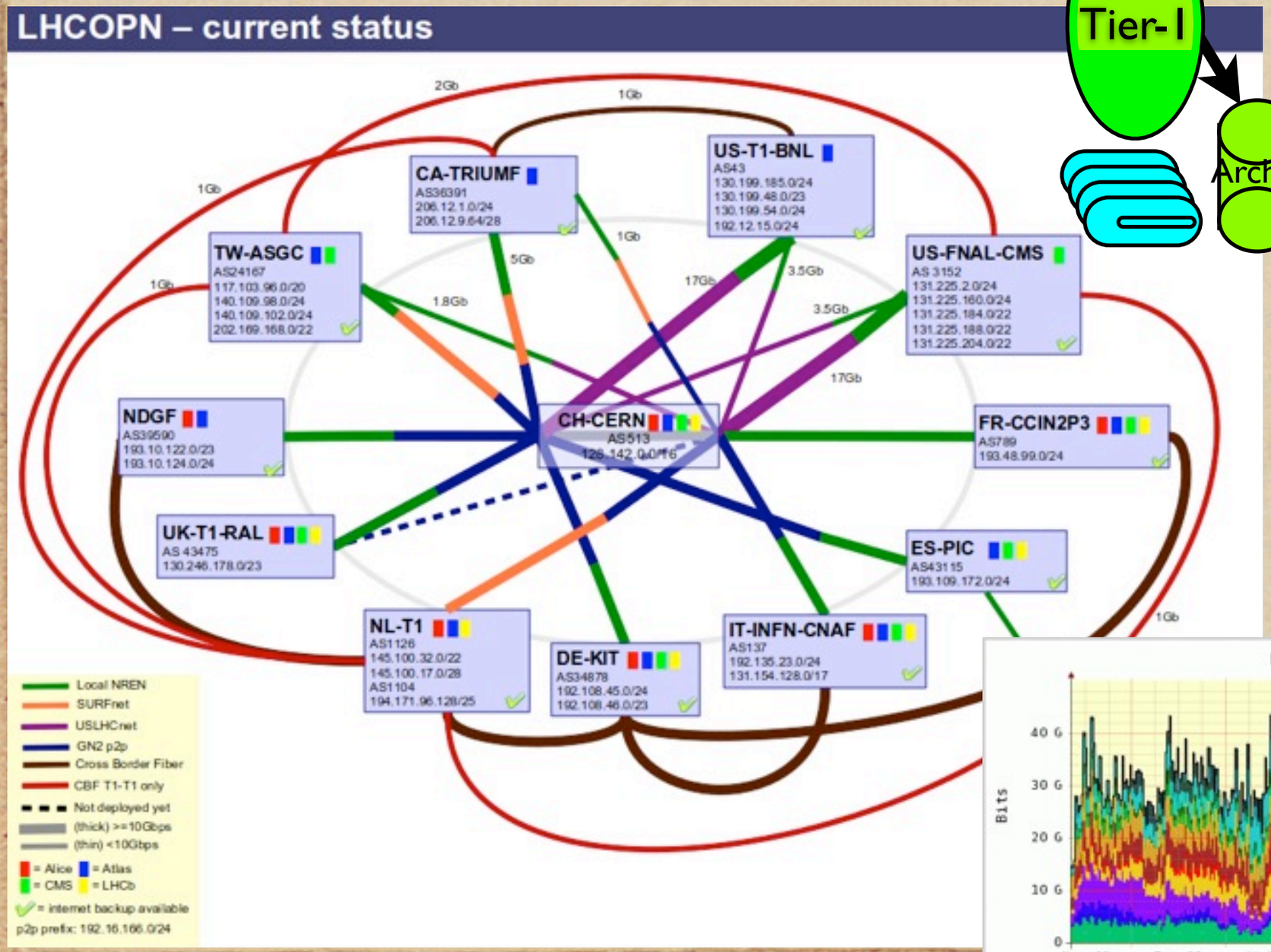




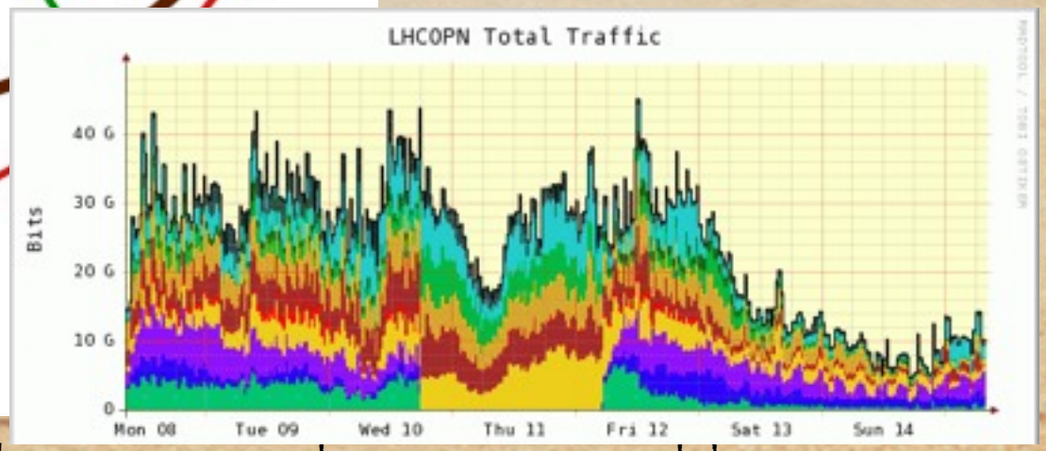
Data Movement

OPN links now fully redundant – last one – RAL – now in production



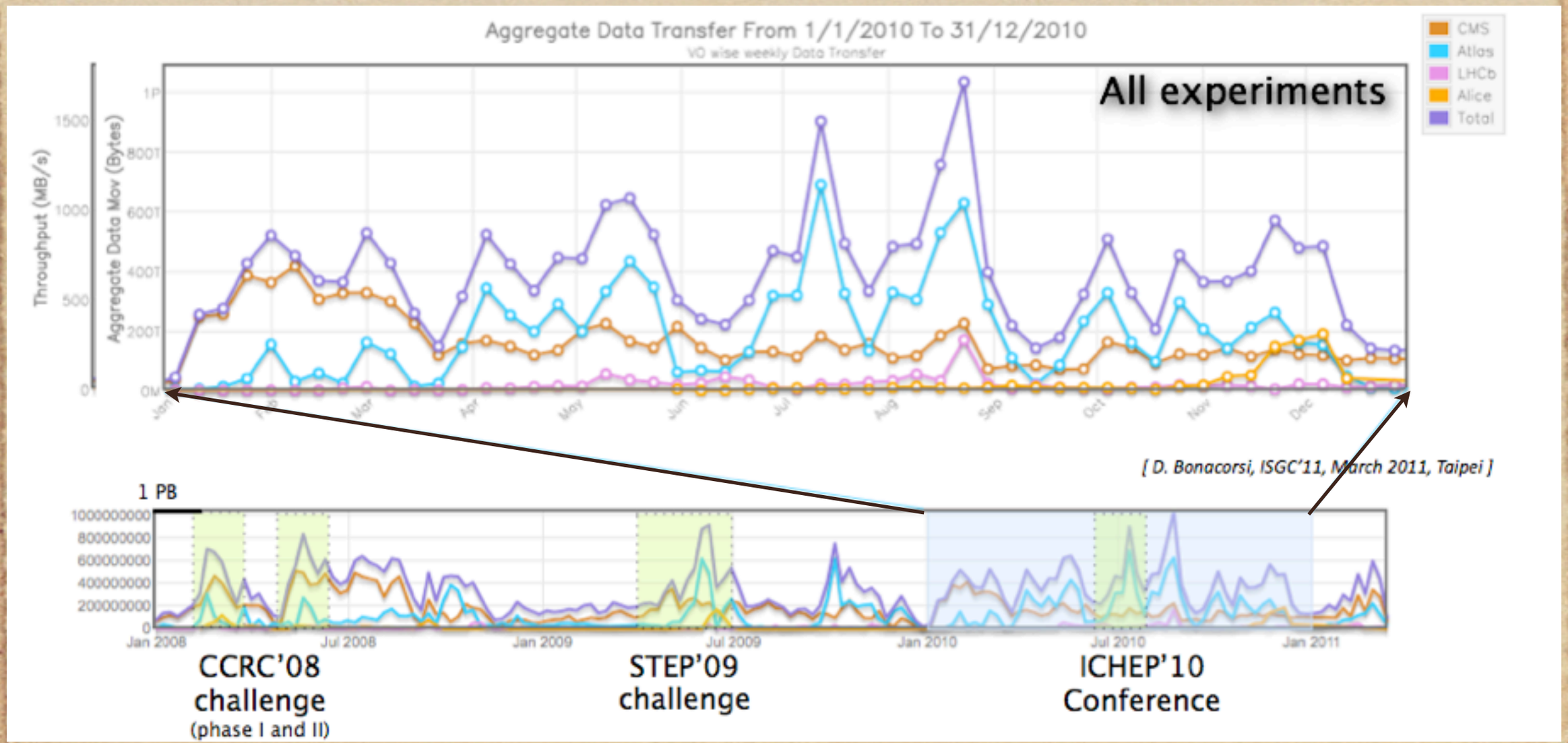


Fibre cut during STEP'09: Redundancy meant no interruption



CERN to T1 Transfers

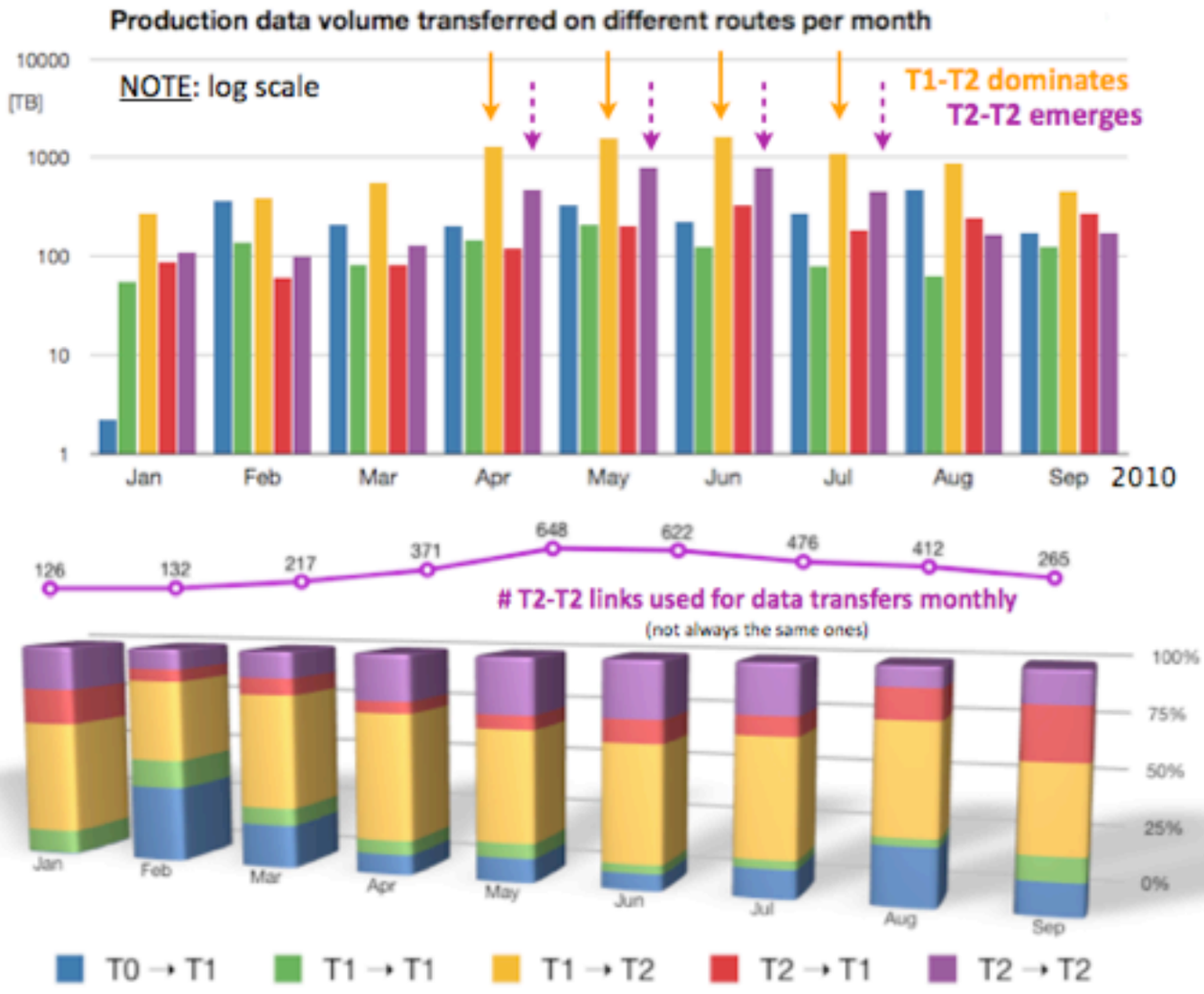
- CERN outbound traffic showed high performance and reliability
- A long and joint LHC commissioning effort



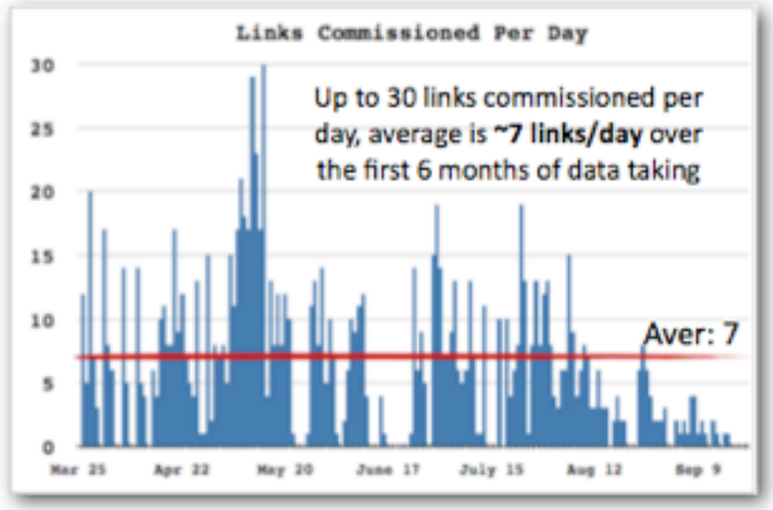
Data Placement for Analysis

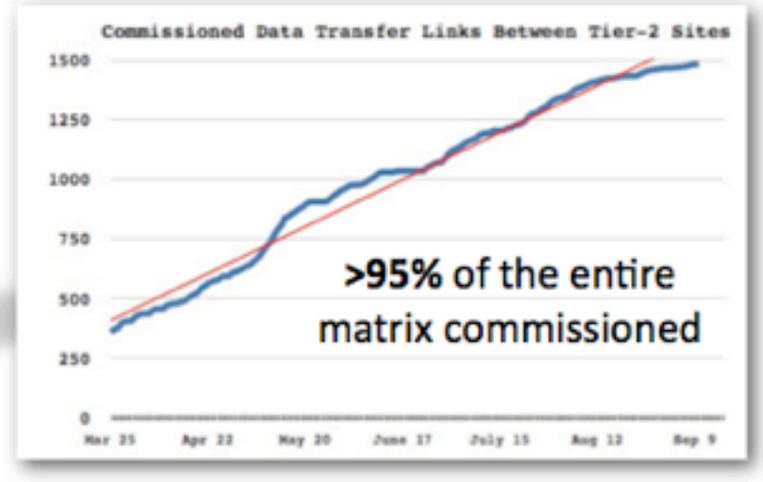
Once data is onto the WLCG, it must be made accessible to analysis applications

- Largest fraction of analysis computing at LHC is at the T2 level
- Flexibility of the transfer model help to reduce the latency seen by the analysis end-users



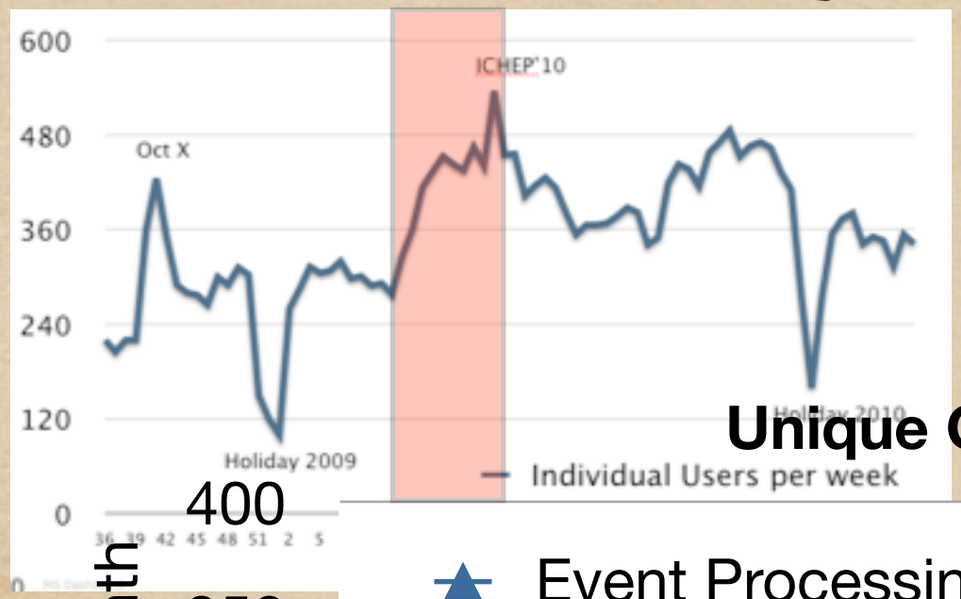
[D. Bonacorsi, CHEP'10 and ISGC'11, Taipei]





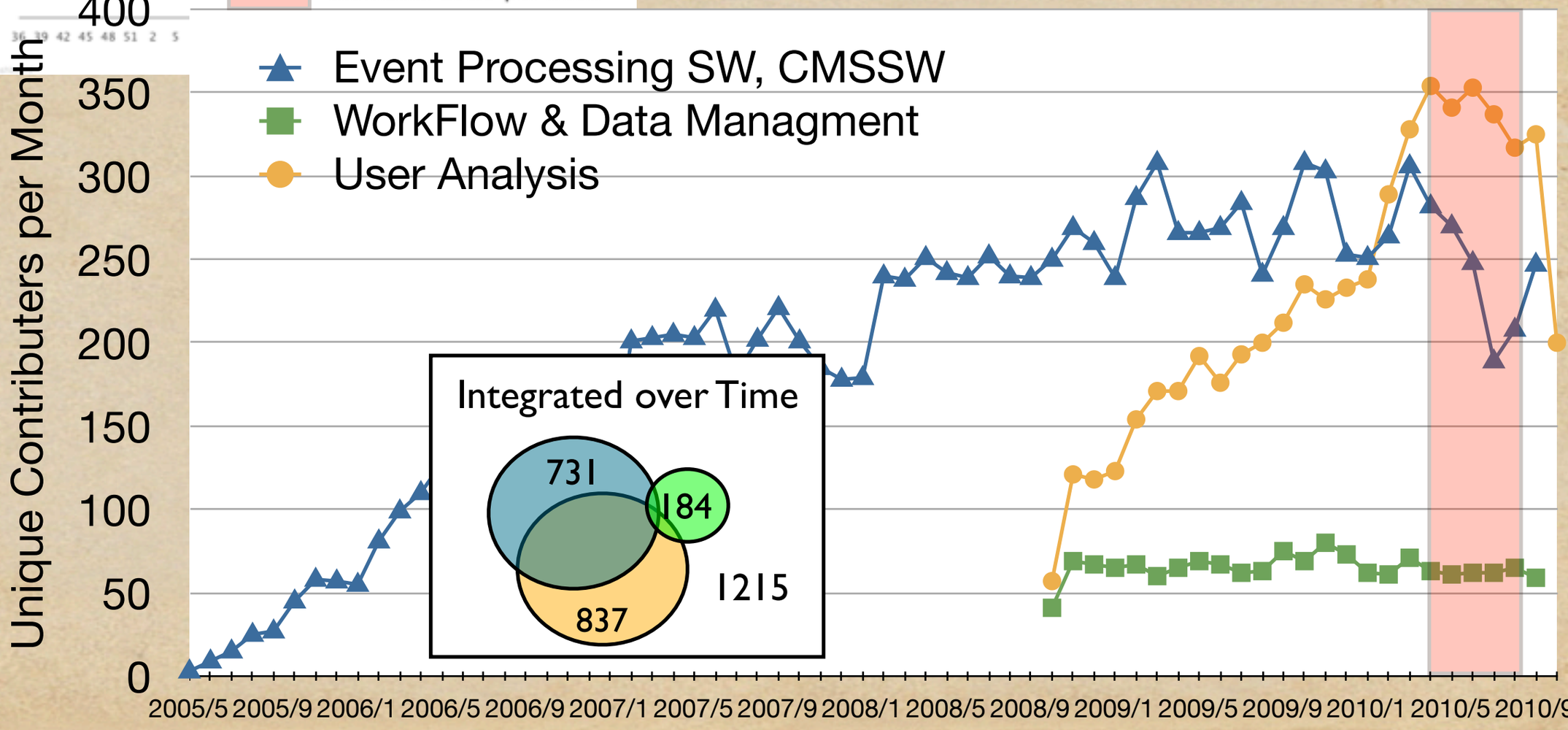
[J. Letts]

Analysis Data Use



- Analysis activity is seasonal
- Scale of the community is large

Unique Contributers to CVS over Time



Future Data Management

- High capacity networking is more available and more reliable than we planned for, so why not have "any data, anywhere, anytime" => NSF proposal based on xrootd technology
- Move to a model where data is requested for analysis at T2s instead of being distributed by agreement. T1 custodial distribution remains the same.
- Improve our monitoring of data popularity and accounting for data access. Even consider automatic removal at T2s, ala Atlas' Victor

Being Provocative...

- CMS's strategy is to have the right data available at the right time in the right place and then deleting it to conserve resources.
- For us the RAW data has to last forever, RAW MC is almost as important.

Special thanks to Ian Fisk, Daniele Bonacorsi, Steve Wolbers and Oliver Gutsche