Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Data Preservation at Fermilab and the Tevatron

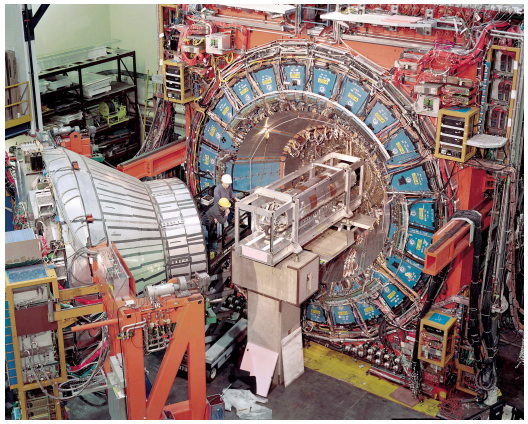
Ken Herner

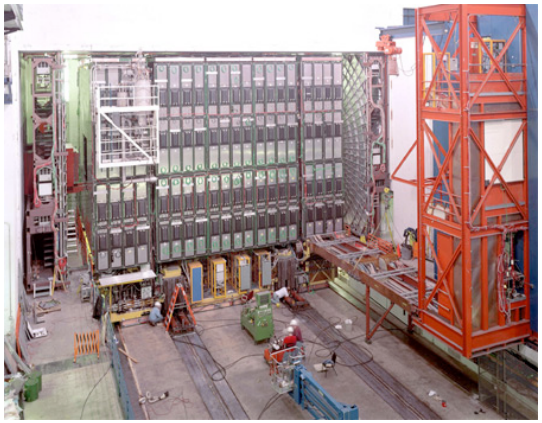
DPHEP Workshop 2015

9 June 2015

Run II Data Preservation Overview

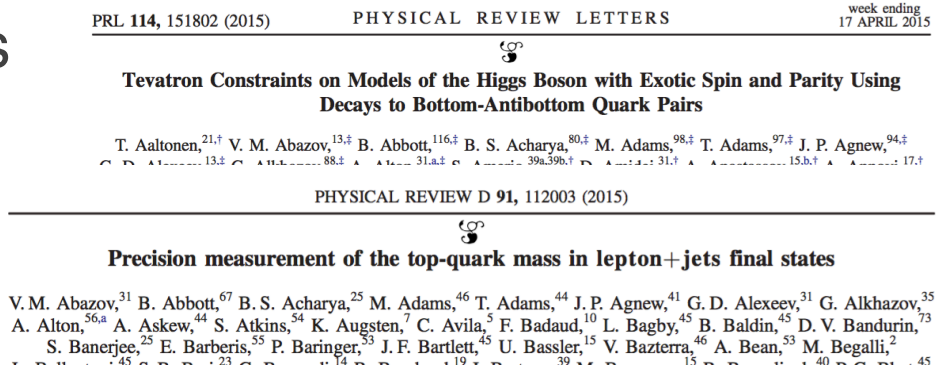
- Overall goal is DPHEP Level 4 preservation of both CDF and D0 through at least 2020
  - Efforts began in 2011 within each experiment
  - Fermilab SCD established R2DP project in 2012 to accomplish specific pieces of the experiments' programs; ran 2013-2015
    - Dedicated experts from CDF and D0, Fermilab SCD technical lead, and project manager
- Technical work complete; now educating users
- Efforts require minimal disruption to ongoing work





D0 Physics Overview

- Presently 30-50 collaborators active in analysis
  - Mostly top, QCD, and electroweak analysis now
  - last Higgs paper published
- 22 papers in 2014; 9 so far in 2015 (PRD and PRL Editor's suggestions); 2 in review
  - 108 in print since Tevatron shutdown on 30 Sep 2011
- About 10 Ph.D.s in 2014; **2 in 2015 so far**



PRL 114, 151802 (2015) PHYSICAL REVIEW LETTERS week ending 17 APRIL 2015

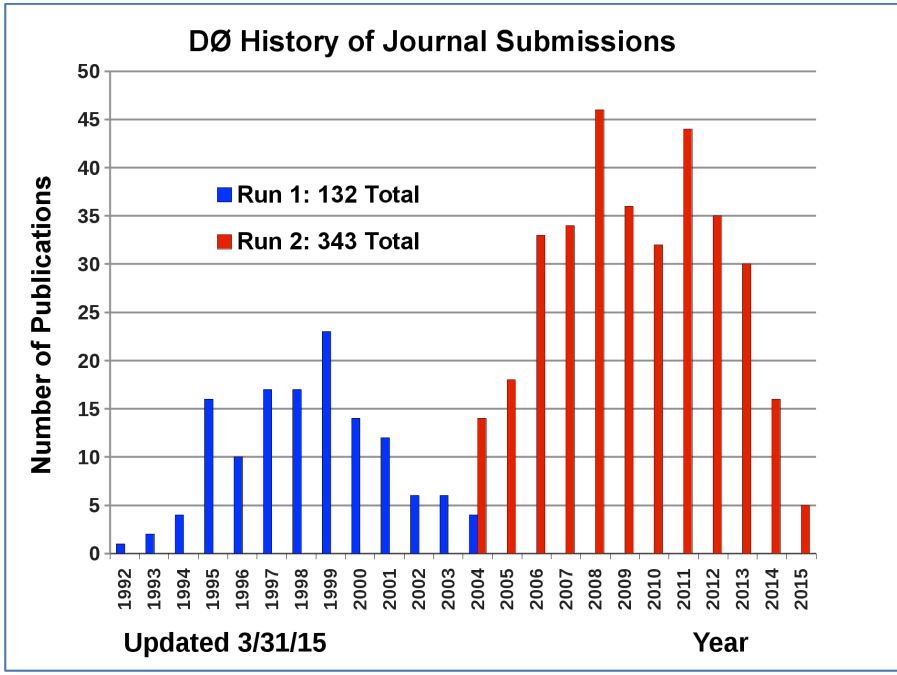
Tevatron Constraints on Models of the Higgs Boson with Exotic Spin and Parity Using Decays to Bottom-Antibottom Quark Pairs

T. Aaltonen,[21,†] V. M. Abazov,[13,‡] B. Abbott,[116,‡] B. S. Acharya,[80,‡] M. Adams,[98,‡] T. Adams,[97,‡] J. P. Agnew,[94,‡]

PHYSICAL REVIEW D 91, 112003 (2015)

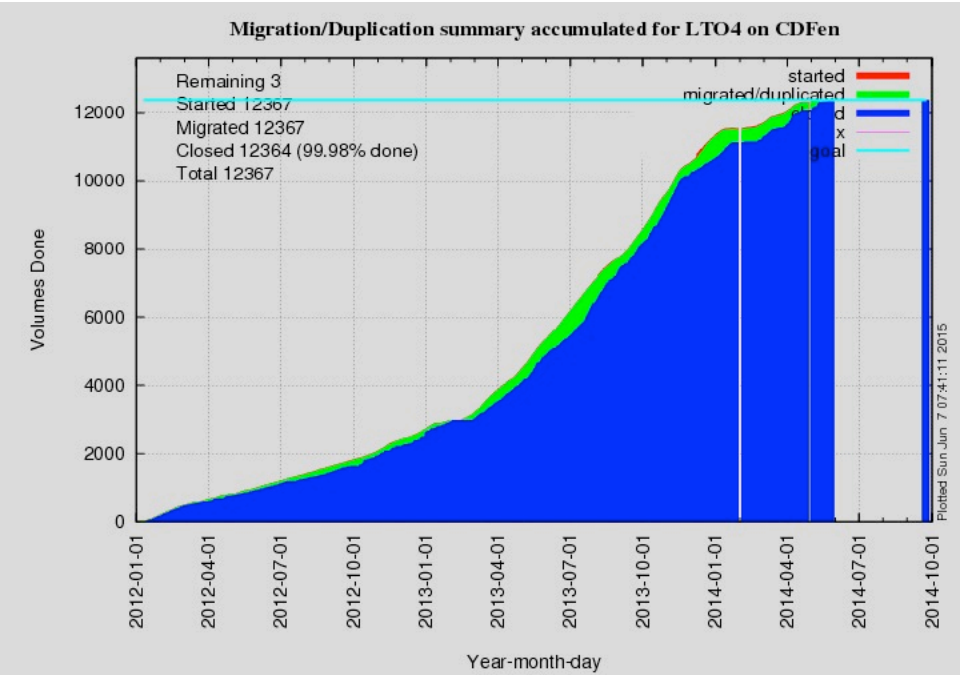
Precision measurement of the top-quark mass in lepton+jets final states

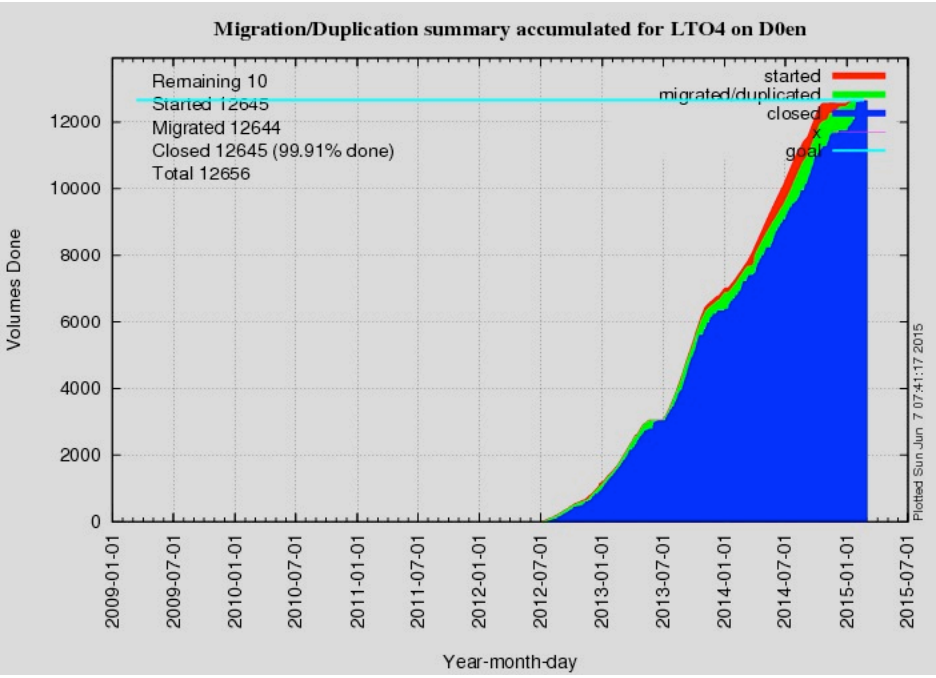
V. M. Abazov,[31] B. Abbott,[67] B. S. Acharya,[25] M. Adams,[46] T. Adams,[44] J. P. Agnew,[41] G. D. Alexeev,[31] G. Alkhazov,[35] A. Alton,[56,a] A. Askew,[44] S. Atkins,[54] K. Augsten,[7] C. Avila,[5] F. Badaud,[10] L. Bagby,[45] B. Baldin,[45] D. V. Bandurin,[73] S. Banerjee,[25] E. Barberis,[55] P. Baringer,[53] J. F. Bartlett,[45] U. Bassler,[15] V. Bazterra,[46] A. Bean,[53] M. Begalli,[2]



Dataset Preservation

- Migrated all raw data, reconstructed data, simulation to T10K tapes-- expected to be readable through 2020
  - roughly 10 PB/experiment
- Additional migration may happen; not set in stone as of now



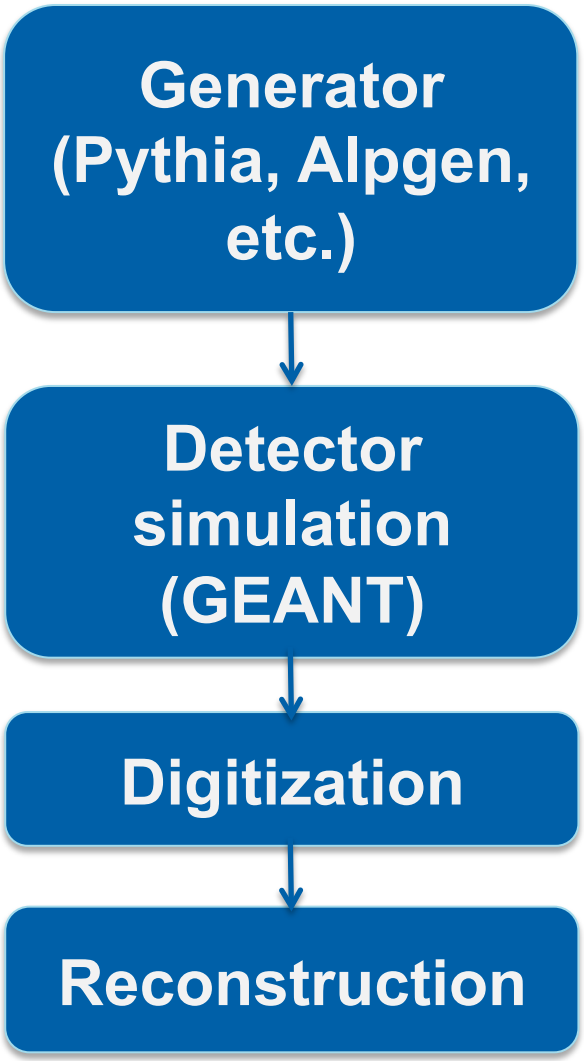


Software preservation: D0

- Current software release (32-bit) built in SL6
- D0's plan is to bring along any needed compatibility libraries within software release (rewriting everything for native SL6 compilation deemed too large an effort)
- **Have verified that there are no issues with building and running release software and common analysis tools within SL6**
- Software releases now published in CVMFS repository
  - At the moment only works at FNAL
  - May transition to a d0.opensciencegrid.org repo (other FNAL experiments have done this; easier to maintain one system)
  - Student on DASPOS funding very helpful setting it up
- Code/products in CVS; commitment from FNAL to support repos

Software Preservation- D0 Monte Carlo

- Important to retain ability to generate new MC (including with modern PDF libraries)
- D0 MC chain is LHA-format compatible (i.e. takes any generator input in LHA format)
  - Done recently with MadGraph for Higgs spin/parity analysis
- Standard generators (typically CTEQ6L1 PDFs), GEANT, digitization, reconstruction code available in CVMFS
  - User can run generator/PDF set of choice, then feed into rest of chain
- *Will maintain this ability for life of project*



Software Validation

- Need to periodically poke things: if it isn't tested, it's broken
- Have selected a small data subsample and some MC samples (top pair production) to serve as validation samples
  - Compare to published outputs
- Small scripts will run selected samples through R2DP infrastructure, compare to reference plots
  - Can be easily run by one person

Job submission at D0

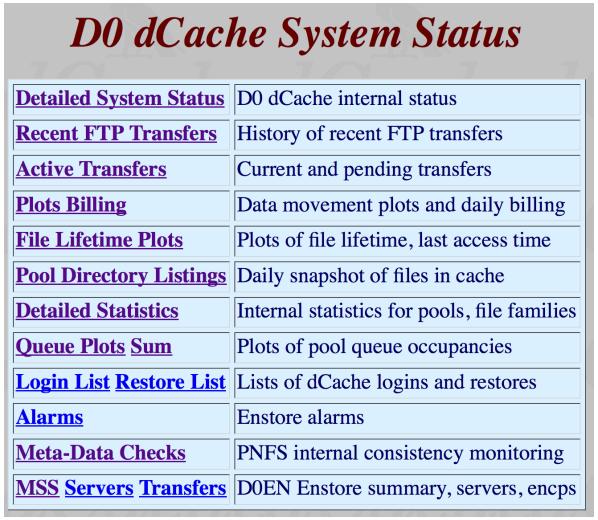
- Historically D0 has a PBS-based job submission system
  - Almost all user analysis jobs to go a Central Backend (CAB) managed by FNAL or Linux cluster at D0 (contrib. from all institutions)
  - Some MC production runs on the Grid
- As machines are retired and resources dwindle, we must find an alternative
  - D0-specific/custom systems no longer an option
- Solution: make use of existing job submission infrastructure used by neutrino/muon/astro experiments at FNAL; run jobs on Fermilab's GPGrid
  - Most experiments use "jobsub", Python/command-line abstraction layer to do condor_submits (jobsub_submit ... -> condor_submit ... behind the scenes.) Adapted by D0 as well

Job submission-- easing transition

- Users want familiar tools (especially if coming back later)
  - Will not make significant effort to learn new system
- Incorporate into standard D0 job submission tools
  - e.g. usually one might do runcafe -cabsrv1 ...
  - for GPGrid submission: runcafe -fermigrid ...
- All the details of building the jobsub_submit command done within the tool
- Exists side-by-side with traditional method, so tool never has to change again
- The -fermigrid option enforces certain other options (new SAM interface, etc.)

File Delivery (1)

- During Run II D0 used SAM and cache areas (about 1 PB) for file delivery to from tape libraries to analysis jobs
- Current SAM architecture moving away from SAM cache to interaction with dCache
  - Preferred solution for D0 DP; much better long-term support. Also don't need to mount PNFS tree on worker nodes via NFS 4.1
- Currently have 100 TB dCache instance set up for D0



D0 dCache System Status

Detailed System Status	D0 dCache internal status
Recent FTP Transfers	History of recent FTP transfers
Active Transfers	Current and pending transfers
Plots Billing	Data movement plots and daily billing
File Lifetime Plots	Plots of file lifetime, last access time
Pool Directory Listings	Daily snapshot of files in cache
Detailed Statistics	Internal statistics for pools, file families
Queue Plots Sum	Plots of pool queue occupancies
Login List Restore List	Lists of dCache logins and restores
Alarms	Enstore alarms
Meta-Data Checks	PNFS internal consistency monitoring
MSS Servers Transfers	D0EN Enstore summary, servers, encps

http://d0dca.fnal.gov

File Delivery (2)

- Originally considered mounting dCache via NFS 4.1 on worker nodes; nixed due to scalability and security concerns
- Solution: Fermilab's IFDHC package (IF Data Handling Client) communicates with SAM
  - SAM can return result in form of gridftp URI
  - IFDHC has C++ API; will perform gftp transfer itself
- Required minimal code modification on D0 (just two packages)
  - Calls for next file from SAM via IFDH interface; returns local copy of file after gftp transfer (automatic cleanup)
  - Exists side-by-side with legacy setup, controlled by config option. No modifications required to end user code!
- **All standard workflows successful with it; now in production**

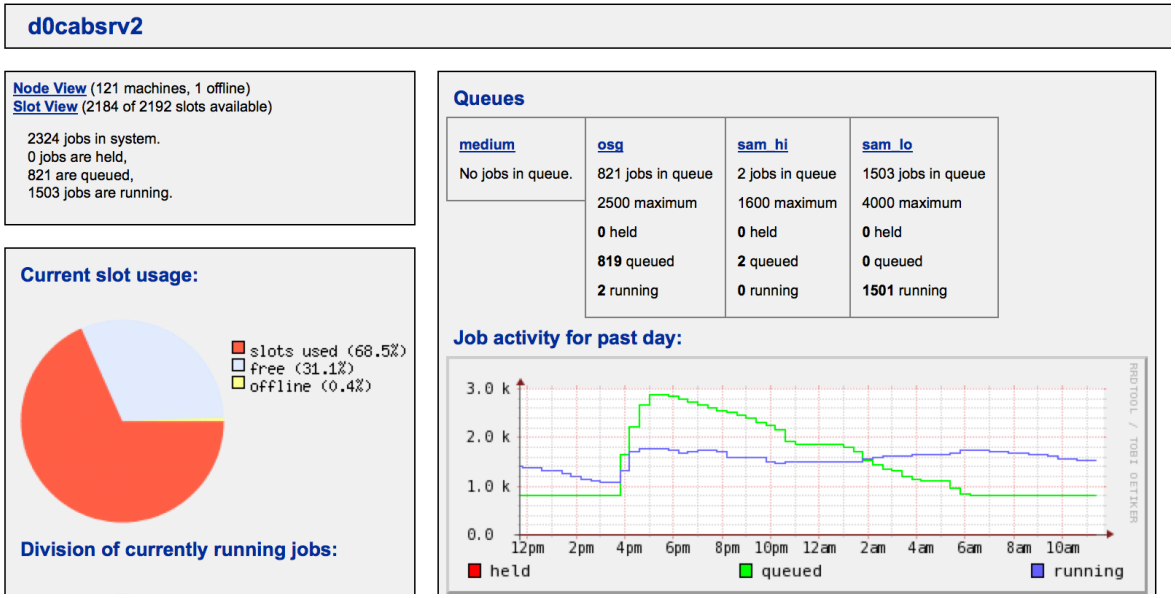
File Delivery (3)

- On CAB cluster, output back to D0 storage via scp or rsync
- Cannot be done directly from GPGrid (SEs not mounted and job doesn't have user krb5 ticket by default)
- Would have required users to copy output to single shared disk from job, then copy to D0 SEs by hand
- Solution: work with jobsub and ifdh developers to pass a krb5 principal to the job (jobsub) and add special "D0:" hook (ifdh) to trigger scp using said principal

- e.g. ifdh cp D0:foo:/bar ... copies /bar from node foo using scp and the krb5 principal; transparent to the user
  - user just adds special principal to .k5login on D0 SEs
- Now **everything** works exactly as users are used to

CAB migration to GPGrid

- As CAB usual slowly decreases, worker nodes being absorbed to GPGrid
- 50% of worker nodes transitioned earlier this year
  - Remainder expected to follow in late summer
- At that point R2DP infrastructure will be the primary means of running



Documentation

- Internal Notes, Agenda server: moves completed
- Detector/online info: Migrated logbooks and DBs to supported software (read-only in some cases)
- Infrstructure documentation
  - **How-to manual for new job submission infrastructure completed**
    - Discusses both common submission tools and some instructions on how to run custom executables and do file delivery
- Mailing lists/discussions: catalog everything to be saved, work with FNAL listserv admins to make sure everything is ported to any future system (probably read-only)
- Wiki: convert to static pages once need for write access is gone

Internal Notes and Agenda Server

- Over 6,000 Internal D0 notes and technical memos
  - Worked with INSPIRE technicians on login authentication system
  - Most will eventually be made public
  - More than 2,000 older notes did not exist electronically; large effort to scan them
  - All notes now migrated
- D0 agenda server was CDS-based
  - All items (18,000) moved to Fermilab Indico
  - Includes meeting records, internal presentations
  - Challenges to convert some event records to suitable format (due to handling of special characters in record names)

Databases

- Start of project: calibration database hardware would not survive full period
- Move some databases to virtual machines, others to supported hardware
- Calibration and luminosity DBs migrated in April 2014
  - No interruption to any efforts
- Software side: DBs are Oracle; various middleware products (Corba, OmniORB, etc.) what if access breaks?
  - Will not fix the access; *write new interface to use http or other in vogue protocol instead*
  - Databases not available during rewrite period
- D0 accepted risk and agreed to rewrite plan (tradeoff to save effort for other areas)

R2DP Project Effort and Cost

- Budgeted 4 FTE in FY 2013; 3 FTE in 2014; 0.3 FY 2015
- Actual numbers were 3; 2.1; 0.4 (CDF+D0)
- Those are undercounts: only included some effort of non-project personnel working on Tevatron things (e.g. dCache and ifdhc, DASPOS student)
  - Also did not count effort not strictly in scope: notes in INSPIRE, moving some DBs, etc.), CNAF effort at CDF (see Silvia's talk)
- My own ~~guess~~ estimate: undercounts at least 0.5 FTE, maybe even full FTE, for first 2 years (plus a similar amount in 2012), plus CNAF efforts
- Vast majority of project expenses were salary

Other Fermilab experiments

- Fermilab is USCMS HQ; also hosting many other experiments in multiple HEP/astrophysics areas
- Range from long-established (Tevatron, Minos, Minerva) to not even built yet (Mu2e, LBNE/DUNE)

ArgoNeuT

μBooNE

NOvA

Mu2e

THE DARK ENERGY SURVEY

g-2

DUNE

Deep Underground Neutrino Experiment

Other Fermilab experiments (2)

- US DOE taking interest now in DP efforts in proposals
  - Most new/current experiments at the 1-page level now. Typically discuss software (version control) and maybe bit preservation
- Newer experiments seem to be learning lessons
  - Most experiments using CVMFS now; makes things much more portable
  - Also generally keeping up with OS changes (c.f. SL5 → SL6.)
  - Fermilab using the lessons of completed experiments to guide the current crop
  - In a few cases, lack of knowledge of older code now causing problems
- Important to establish good practices from Day One
  - Work is vital, but it is typically not very visible. This is a concern.
  - LarSoft project creating "librarian" positions; responsible for code maintenance and validation against new HW/OS/SW products
  - Smaller experiments don't/won't have resources of Tevatron/LHC. Important to come up with a common approach

Lessons Learned

- Do enforce **common** coding practices and file formats wherever possible across the experiment, and don't rely on a specific version of a 3rd-party product if possible
- Do regularly update and validate 3rd-party products
  - Take advantage of natural ebbs in analysis cycle
- Users do not want to deviate from what they know
  - Do hide changes if it is possible (e.g. jobsub). A little extra work there goes a long way to getting people to adopt new tools. If they don't need to know, don't tell them
  - Do carefully document any changes to the usual procedures and provide a HOWTO document
- Don't be afraid of planning for the future now. It is never too early to start thinking about these things!
  - Goes for workflow/data infrastructure design too. Make it easy to change things under the covers (FNAL's ifdhc package good example)

Lessons Learned (2)

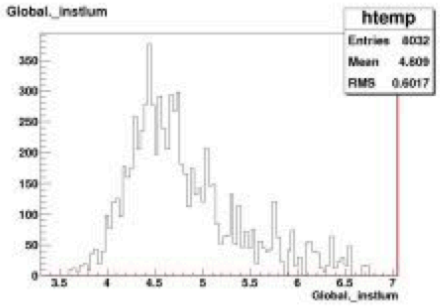
- Don't be afraid of partnering with the host lab's IT division
  - FNAL experience was that they wanted to help
  - All Tevatron problems solved by working together (krb5 copyback good example)
  - Understand where you need things a certain way and where compromises can be made
- Can be hard to get effort from the experiments for testing/ validation, *especially from those closest to the analyses*
  - Back to visibility problem
  - **What can we do to increase recognition for this work in the field? Especially efforts to keep things up to date *before* the experiment shuts down?**

Summary

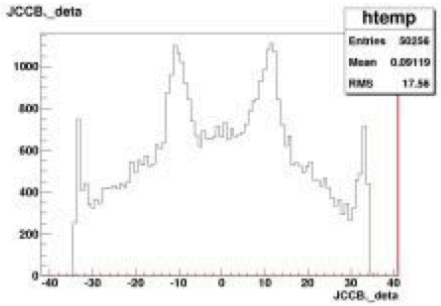
- Run II Data Preservation Project technical work is complete; **accomplished all major goals**
- Transitioned now to deployment and user education phase
- D0 technical work included modifying file delivery system, adapting job submission tools, migrating databases
- Both CDF and D0 had significant effort in documentation preservation and modifying job submission systems
- Not much DP literature out there now
  - Working on a paper describing efforts and results-- probably going to NIM
  - Serve as a guide for future experiments
- Effort will aid other Fermilab experiments as they move into more mature states

BACKUP

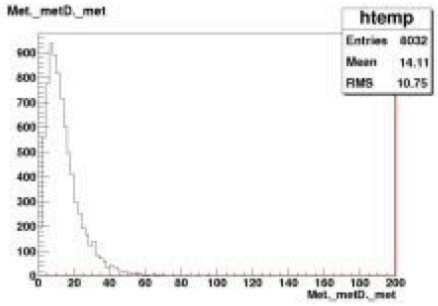
Local cluster vs. Grid comparison



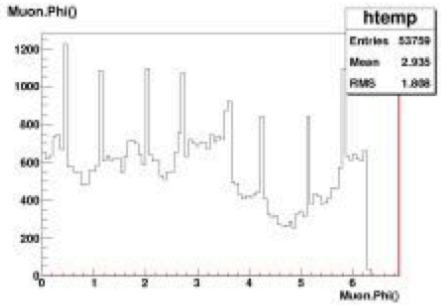
clued0_test_instLum.jpg, [eps]



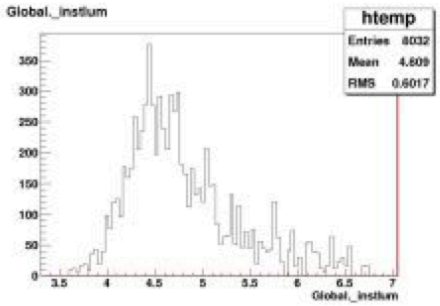
clued0_test_JCCB_deta.jpg, [eps]



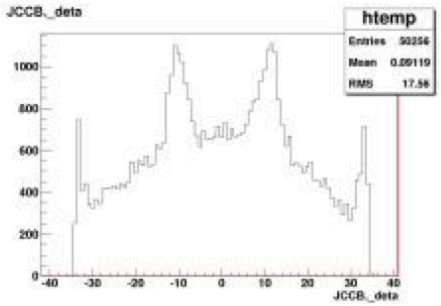
clued0_test_metD.jpg, [eps]



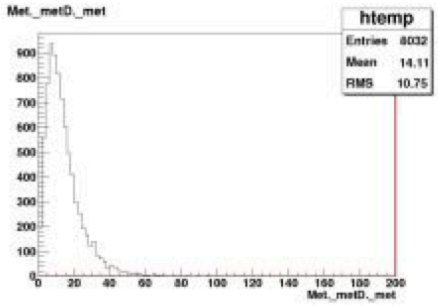
clued0_test_muon_phi.jpg, [eps]



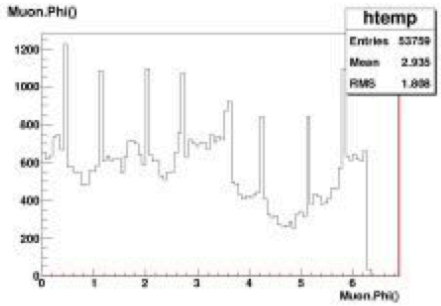
grid_test_instLum.jpg, [eps]



grid_test_JCCB_deta.jpg, [eps]



grid_test_metD.jpg, [eps]



grid_test_muon_phi.jpg, [eps]

Identical!