Data Challenge 1 Closeout
Lessons Learned Already
Fallout from DC1 (so far)

- SciTools did not run on windows at the last minute
  - Release Manager needs to run Windows builds
    - Now running – glitch with IExternal
    - Awaiting dedicated windows servers from SLAC SCS
    - RM will also provide standardized user facility with all GR, ST builds available, just as for SLAC linux.
- We discovered problems with sources and ACD ribbons late
  - Better unit tests for sources; and system tests for ACD as well as better review process when major changes made to subsystems
  - Need better tabulation of MC Source IDs in the data header
  - Need to revive and update/upgrade system tests!
  - Should check backgrounds in detail (given signal problems)
- Manual handling of the processing
  - No pipeline, manual concatenation etc of files. Error prone.
    - OPUS now beginning to function. Can run DC1 jobs and mostly write results to processing database
    - Wrote tools to handle concatenations
  - No checking of file integrity
• Large failure rate in batch jobs (~10%)
  – Most fixed now; remaining few percent will be tricky to find
    • Need to run a new batch of jobs in debug to see what’s going on

• Tools are not checking inputs much, eg:
  – Easy to get Likelihood into infinite loops with inputs in wrong order etc
    • Input checking has been added
  – GRBobsSpectrum doesn’t check that it can read the ascii photon list

• Code distribution scripts were written manually
  – And we need to rationalize how they run wrt environment variables etc
    • Testing “Release Area” feature of CMT
    • May buy InstallShield installer for the rest

• We lowered the bar in several places:
  – No backgrounds; “approximated” background rejection cuts
  – No onboard filter
  – No pipeline (OPUS) – almost running now
  – No CALDB - implemented
Tools & Data

• Pat has volunteered to install these on SLAC linux
  – FTOOLS
  – XSpec
  – AstroRoot

  – What else do we need?

• We’ll add these to the Windows servers when they arrive

• SLAC Data Server to be revamped
  – Code rewrite in conjunction with upgrades to OPUS, SysTests & Release Manager
  – Need to optimize Root queries & investigate other speed-ups
  – Improvements to user documentation
Dataset Generation

- We have the CPU capacity to generate large datasets
  - viz 150M background events for rejection studies
  - Run at SLAC (Heather) and Lyon (Berrie)
    - (bb)ftp’ed Merit tuples to SLAC
    - Ditto for full tree files for requested events that failed rejection cuts
    - Done “by hand”
  - Starting to test out farm at Perugia
  - ~750 GB per shot – & on tape in Lyon.

- We will want to develop a more coordinated, automatic scheme for multi-farms
  - GRID for GLAST??

- Would be nice to tie code version into data – have it recorded in the output files headers.

R.Dubois
ScienceTools Organization

- After first release of ST, Toby Burnett (UW) and James Peachey (SSC) as co-architects

- Johann Cohen-Tanugi and Jim Chiang acting as “Release Managers” for ScienceTools