



Data Handling for DC1

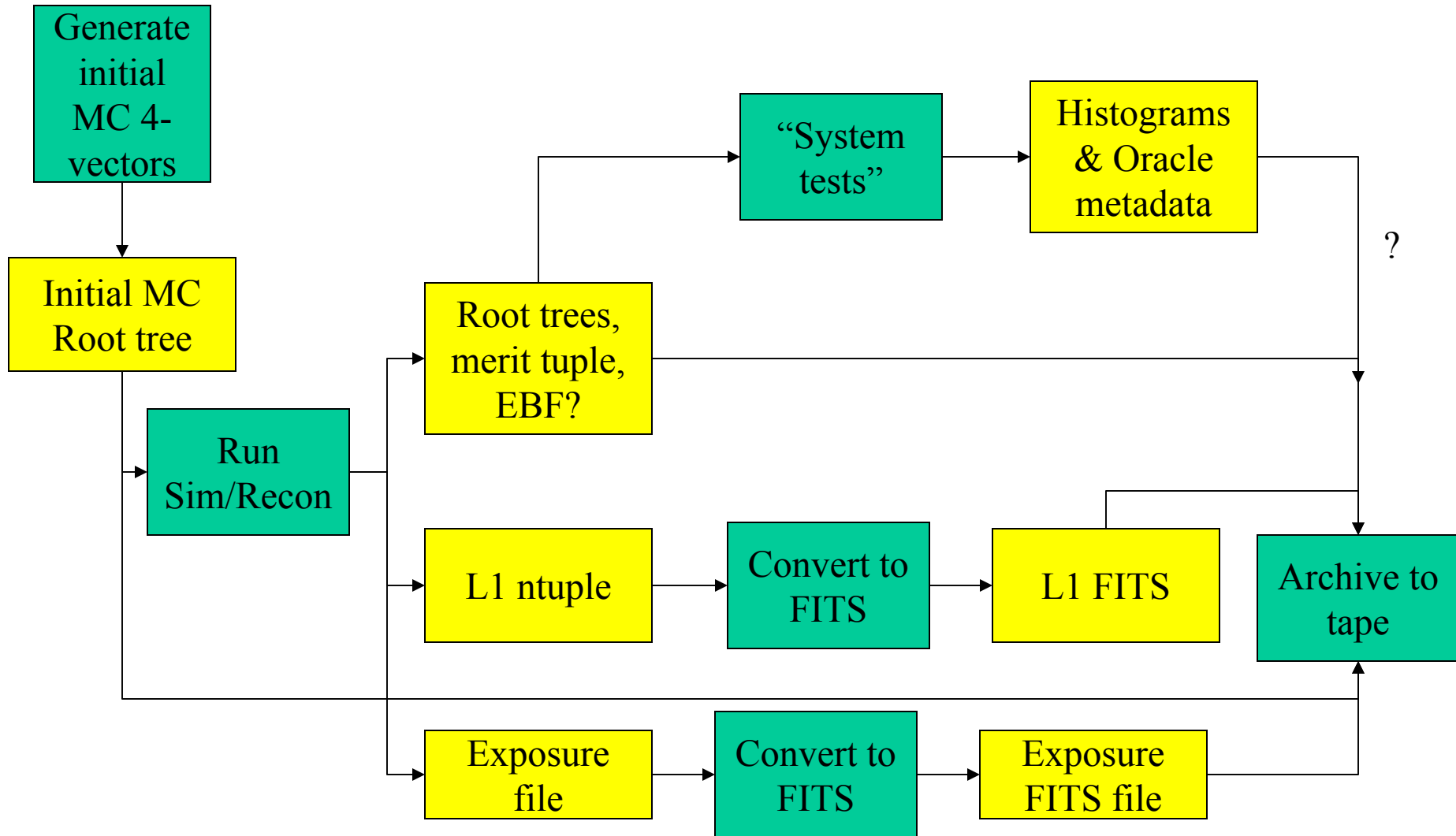
- **GlastRelease strategy for generating the data**
- **Required components in the OPUS process pipeline**
- **Types, form and archive of the output data**
- **Access to the data after the pipeline is complete**
 - **Web catalogue to datasets**
 - **FITS from the LAT**
 - **FITS from the SSC**
 - **Ntuple and tree access**
 - **Web access to “system tests” diagnostics**
 - **Remote FRED/WIRED access to full event display?**
- **Issues from processing at Lyon and Perugia?**
- **Monitoring pipeline progress**



Processing Strategy

- **This was mostly covered in Toby's note**
<http://www-glast.slac.stanford.edu/software/DataChallenges/DC1/>
- **Basic idea is:**
 - **Generate MC “4-vectors” for desired period in orbit**
 - **Run G4, then see if L1/Onboard Filter pass the event**
 - **If yes, run Recon, else toss the event**
 - **Convert ntuple, exposure outputs to FITS**
 - **Archive everything to tape**

Required OPUS processes - Runs





Required OPUS post-processing

- **Assumes database fully updated for all 86k individual runs**
- **Concatenate output of successful runs – one (or small number of) file(s) each for the day**
 - **Ntuples & Root trees**
 - **FITS files**
 - **Transfer to SSC**
 - **Archive all to tape**
- **Update database with pointers to concatenations**
 - **Database must be prepared to associate these concatenations with DC1 task, presumably by a group type**

Resource Issues

- **CPU**
 - Need to build/benchmark optimized version of GR (including G4)
- **Disk**
 - Must acquire sufficient disk to handle nominal plan of storing full output post Onboard Filter.
 - Karen to clean up existing partitions
 - Request in to SCS for 2-3 new TB
- **Tape**
 - Are there any existing utilities to manage tapes? ie so we would not have to keep track of how full the tapes are.
 - Where do the tapes come from?
 - Do we have enough margin on CPU and tape to consider running the full processing chain and write all the fully processed events to tape?



Issues & Miscellany - 1

- **Remote server connection to the Root trees for event display?**
 - **There is no Root tree interface for HepRep**
 - **Would we need for the server to be running Gleam?**
Could it fire up a mini-Gleam on the server?
- **CGI “Peeler”?**
 - **Allow users to specify a list of run/event (ie event id) and have Root files created to contain those events**
 - **Should be fast to just read run/event to find matches?**
 - **Should we even try such a thing to create FITS files on the fly from the tuples? We could easily supply simple query criteria à la SSC.**
 - **We would need some facility to clean up peeler files – garbage collection.**
 - **Do we need a Root tool to query the database to find files?**



Issues & Miscellany - 2

- **System tests**
 - Surely we don't want 86k individual sets of files to monitor?
 - Is there a post-processing step to concatenate all the histogram files?
- **Web catalogue**
 - What do we want to display from the catalogue?
 - Karen has a simple interface set up already
- **Multi-site processing**
 - Do we run OPUS everywhere?
 - If yes, do we have a problem for customizing db and batch use?
 - How does the database get handled?
 - Are there multiple databases that just need coordinating?
 - Should we check with BABAR?
 - How do we share the produced files?
 - Who are the people responsible for processing at each site?