GLAST Large Area Telescope

Science Analysis Software

WBS 4.1.D

Richard Dubois
SAS System Manager

richard@slac.stanford.edu

Outline

- Introduction to SAS Scope and Requirements
- Overall Test Plan
- Data Challenges
- DC1 Summary
- Flight Integration Support
- Network Monitoring
- Outlook

Science Analysis Software Overview

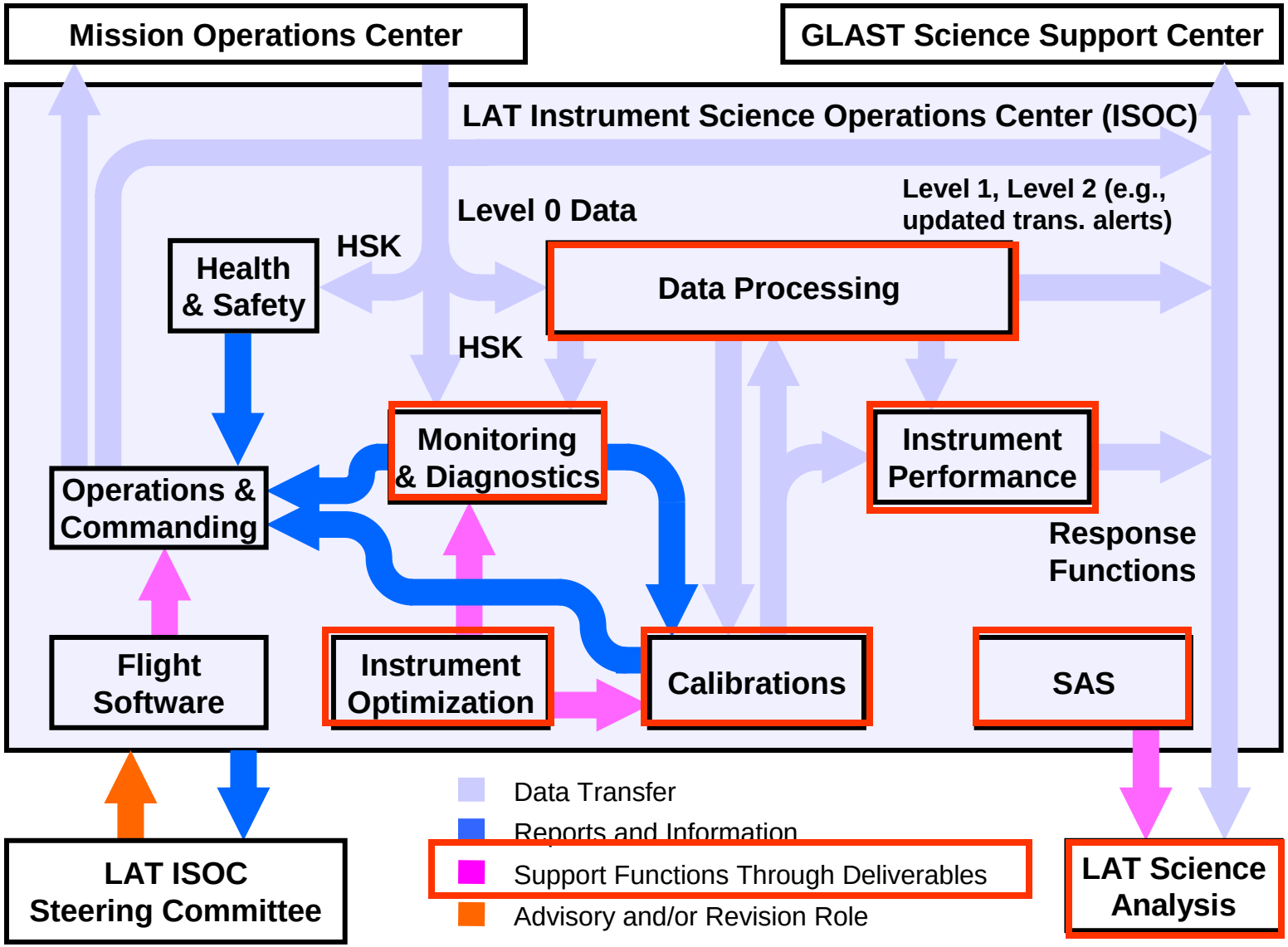
- **Processing Pipelines**
  - Prompt processing of Level 0 data through to Level 1 event quantities
  - Providing near real time monitoring information to the IOC
  - Monitoring and updating instrument calibrations
  - Transients searches (including GRBs)
  - Reprocessing of instrument data
  - Performing bulk production of Monte Carlo simulations
- **Higher Level Analysis**
  - Creating high level science tools
  - Creating high level science products from Level 1
  - Providing access to event and photon data for higher level data analysis
- **Interfacing with other sites (sharing data and analysis tool development)**
  - Mirror PI team site(s)
  - SSC
- **Supporting Engineering Model and Calibration tests**
- **Supporting the collaboration for the use of the tools**

Level III Requirements Summary

Ref: LAT-SS-00020

Function	Requirement	Expected Performance (if applicable)	Verification
Flight Ground Processing	perform prompt processing from Level 0 through Level 1	keep pace with up to 10 GB Level 0 per day and deliver to SSC within 24 hrs	demonstration
	provide near-real time monitoring to IOC	within 6 hrs	demonstration
	maintain state and performance tracking		demonstration
	facilitate monitoring and updating of iinstrument calibrations		demonstration
	archive all data passing through	> 50 TB on disk and tape backup	demonstration
Instrument Design Support	Create simulation tool, based on instrument geometry, that reproduces the interactions of photons and background		system test - comparison to balloon flight and existing data
	Create physics model of expected photons and backgrounds incident upon the instrument		system test - comparison to balloon flight and existing data
	Create algorithms to interpret the data from the instrument to identify the interaction and estimate photon direction and energy		system test - comparison to engineering model tests
	Create algorithms to generate calibration constants for the subsystem components		system test - in conjunction with engineering model tests
High Level Tools	Interface with the SSC and PI mirror sites, sharing selected data and algorithms		demonstration
	Create High-Level Science products. Development of analysis tools	.	demonstration
Mission Support	Support the Software system for the life of the mission		demonstration

SAS in and around the ISOC



Manpower

- Mostly off-project
  - From collaboration and SSC
- Effort divided amongst
  - Infrastructure
    - ~6-8 FTEs
  - Sim/recon
    - ~6 FTEs
  - Science Tools
    - 8-10 FTEs
- Effort ramping up for Flight Integration support
  - From infrastructure and sim/recon areas

Overall Test Plan

- **Combination of Engineering Model tests, Data Challenges and LAT Integration Support**

- **EM tests**
  - **EM1 demonstrated ability to simulate/reconstruct real data from single (non-standard) tower**
    - **All within standard code framework/tools**
    - **Data analyzed with SAS tools**
- **Data Challenges**
  - **End to end tests of sky simulation through astro analysis**
  - **Generate instrument response functions**
  - **Exercise pipeline**

- **LAT Flight Integration**
  - **Combine tools from EM & DC applications**
  - **Sim/recon/analysis & pipeline processing and record keeping**

Purposes of the Data Challenges

S.Ritz

- "End-to-end" testing of analysis software.
- Familiarize team with data content, formats, tools and realistic details of analysis issues (both instrumental and astrophysical).
- If needed, develop additional methods for analyzing LAT data, encouraging alternatives that fit within the existing framework.
- Provide feedback to the SAS group on what works and what is missing from the data formats and tools.
- Uncover systematic effects in reconstruction and analysis.

Support readiness by launch time to do all first-year science.

SAS Checklist

Detailed Simulation

Event Reconstruction

Event Classification

Instrument Calibration

ACD CAL TKR

Processing Pipeline

MC I&T

Re-processing

User Support

Code distribution

Data Distribution

High Level Instr Diags

Quicklook

High Level Analysis

GRBs,
AGN,
Pulsars

Catalogue,
Diffuse

Institutional Mirrors

SSC

LAT Mirrors

DC1
DC2
DC3

I&T EM
I&T Flight

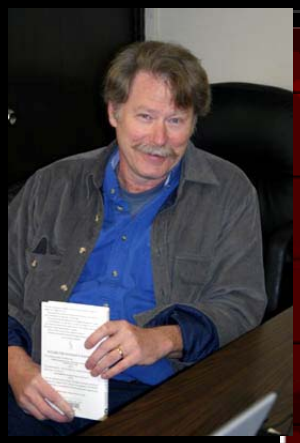
Data Challenge Planning Approach

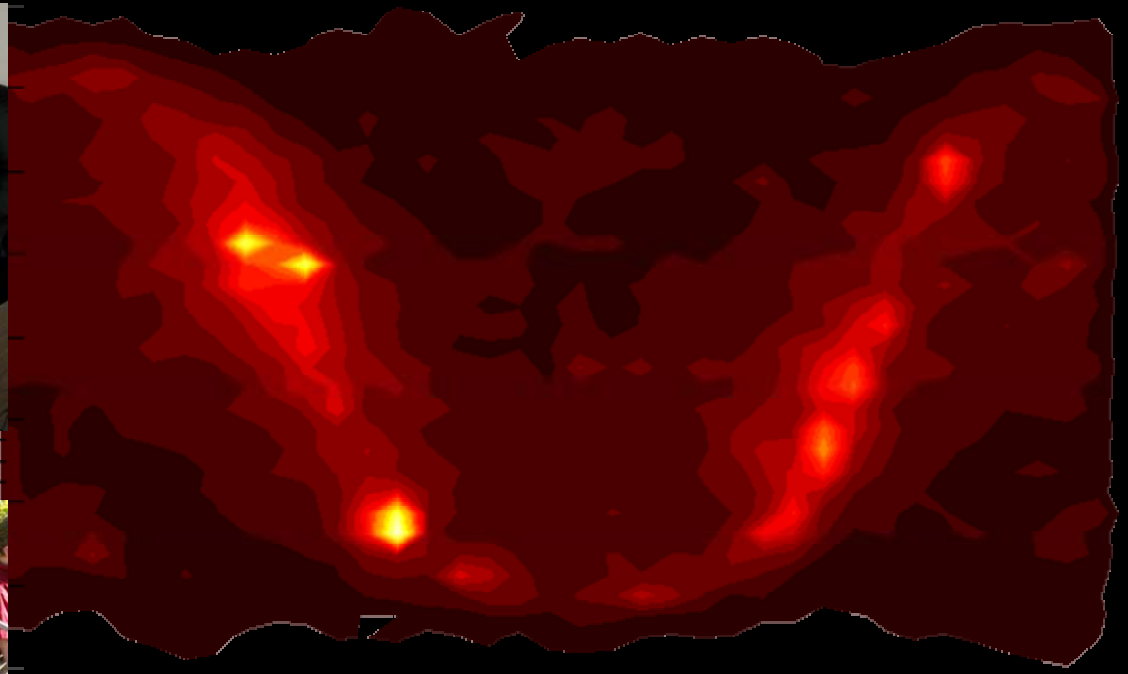
S.Ritz

- Walk before running: design a progression of studies.
- DC1. Modest goals. Contains most essential features of a data challenge. Original plan:
  - 1 simulated day all-sky survey simulation, including backgrounds
  - find flaring AGN, a GRB
  - recognize simple hardware problem(s)
  - a few physics surprises
  - exercise:
    - exposure, orbit/attitude handling, data processing pipeline components, analysis tools
- DC2, start end of CY04. More ambitious goals. Encourage further development, based on lessons from DC1. One simulated month.
- DC3. Support for flight science production.

Data Challenge 1 Closeout 12-13 Feb 2004

http://www-glast.slac.stanford.edu/software/Workshops/Feb04DC1CloseOut/coverpage.htm





DC1 Components

- **Focal point for many threads**
  - **Orbit, rocking, celestial coordinates, pointing history**
  - **Plausible model of the sky**
  - **Background rejection and event selection**
  - **Instrument Response Functions**
  - **Data formats for input to high level tools(*)**
  - **First look at major science tools – Likelihood, Observation Simulator**
  - **Generation of datasets (*)**
  - **Populate and exercise data servers at SSC & LAT (*)**
  - **Code distribution on windows and linux (*)**
- **Involve new users**
- **Teamwork!**

(*) – done – no further comment here

S.Ritz

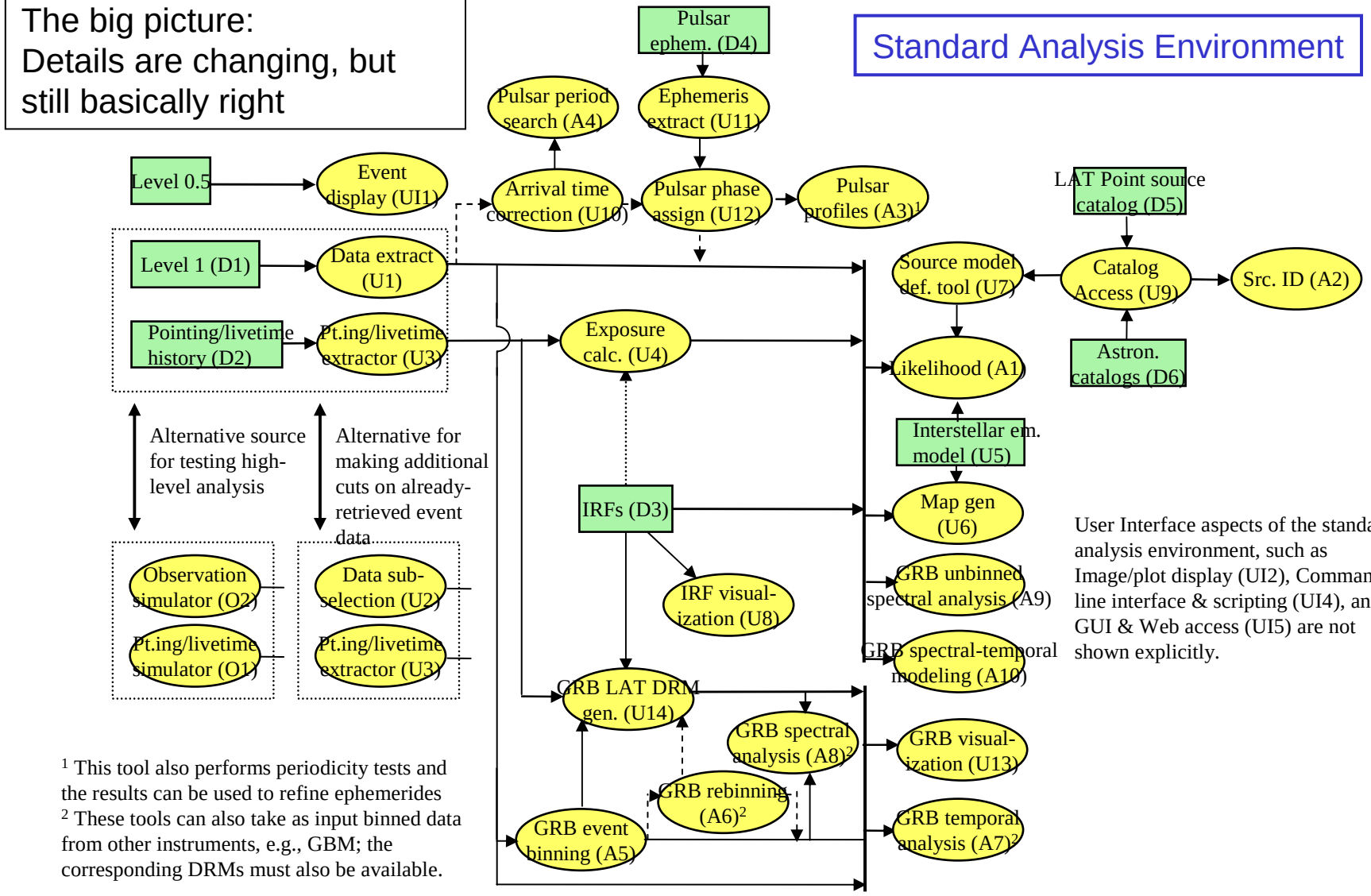
DC1 Minimum Results

- The existence of the data sets and the volume of data generated for background analyses already meets one of the success criteria.
- A minimum set of plots and tables that we must collectively produce:
  - TABLE 1: found sources, ranked by flux (E>100 MeV). Table has the following columns
    - reconstructed location and error circle
    - flux (E>100 MeV) and error
    - significance
    - 3EG identification (yes or no) [note: DON'T assume DC1 sky is the 3EG catalog!]
    - extra credit:
      - » include flux below 100 MeV
      - » spectral indices of brightest sources
      - » comparison of 3EG position and flux characteristics with GLAST analysis
  - FIGURE 1: LogN-logs plot of TABLE1
  - TABLE 2: list of transients detected. Columns are
    - location and error circle
    - flux (E>100 MeV) and error
    - significance
    - duration
  - FIGURE 2: light curve
    - Extra credit: FIGURE 2a: spectra.
  - PLUS: reports of any physics surprises found.

S.Digel and P.Nolan

Science Tools in ~~DC1~~ DC3

The big picture:
Details are changing, but still basically right



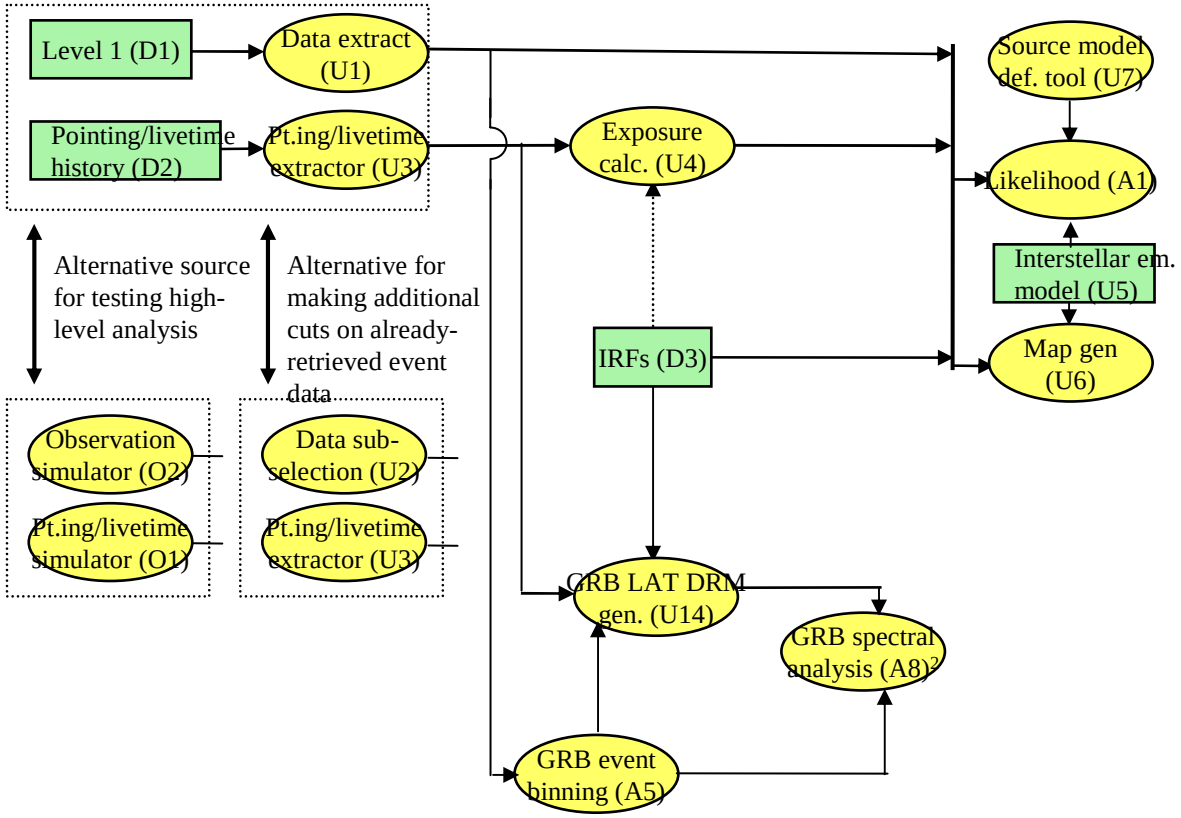
User Interface aspects of the standard analysis environment, such as Image/plot display (UI2), Command line interface & scripting (UI4), and GUI & Web access (UI5) are not shown explicitly.

[1] This tool also performs periodicity tests and the results can be used to refine ephemerides
[2] These tools can also take as input binned data from other instruments, e.g., GBM; the corresponding DRMs must also be available.

S.Digel and P.Nolan

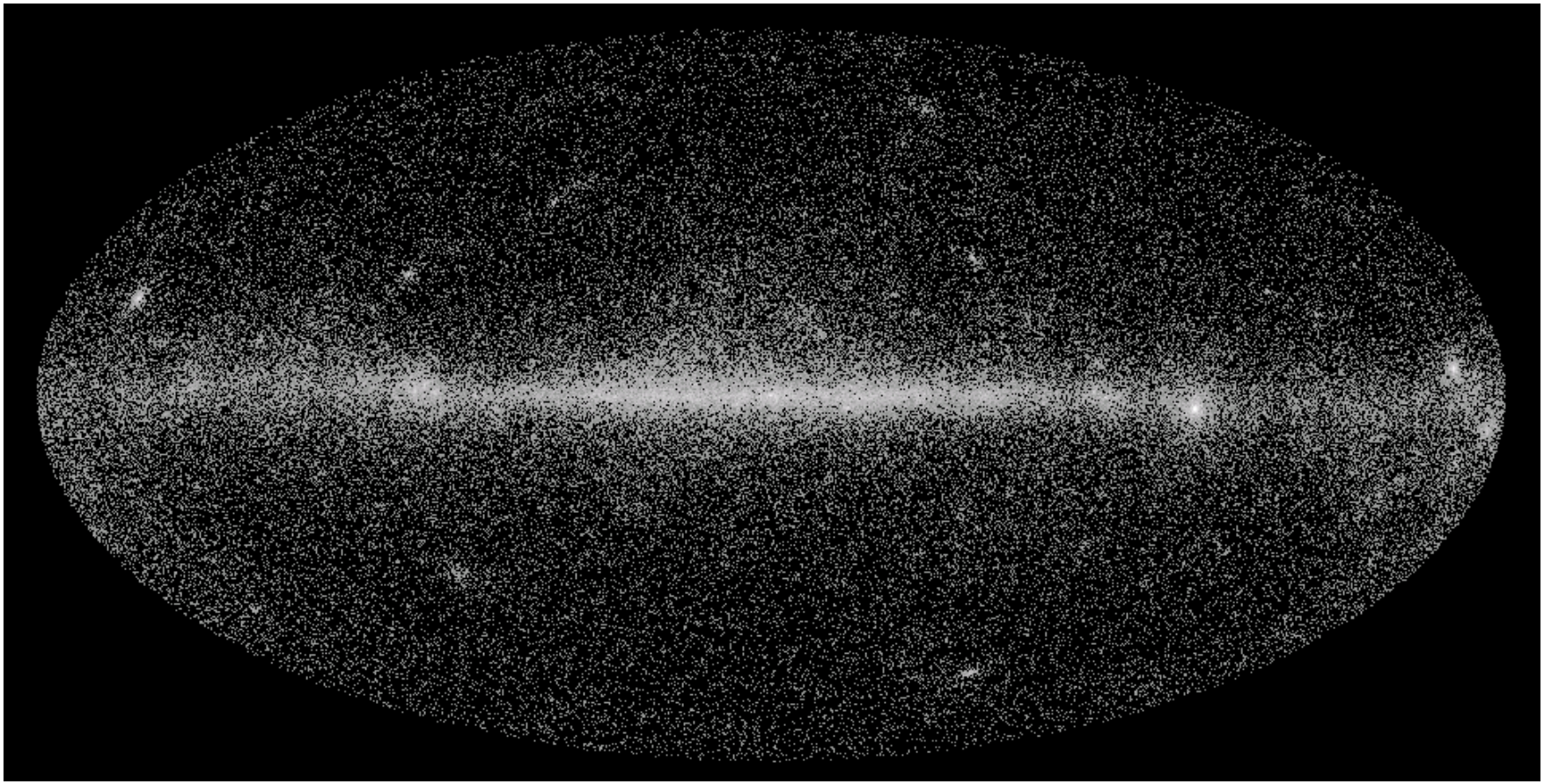
Science Tools in DC1

- **All components are still prototypes**



The DC1 functionality is
Data extraction
Limited visualization
Model definition
Model fitting
Observation simulation

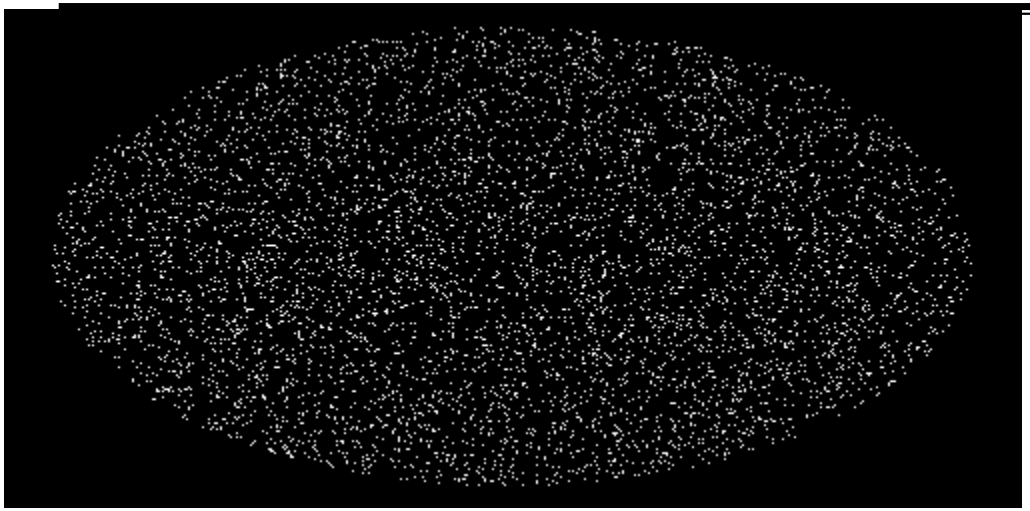
The data



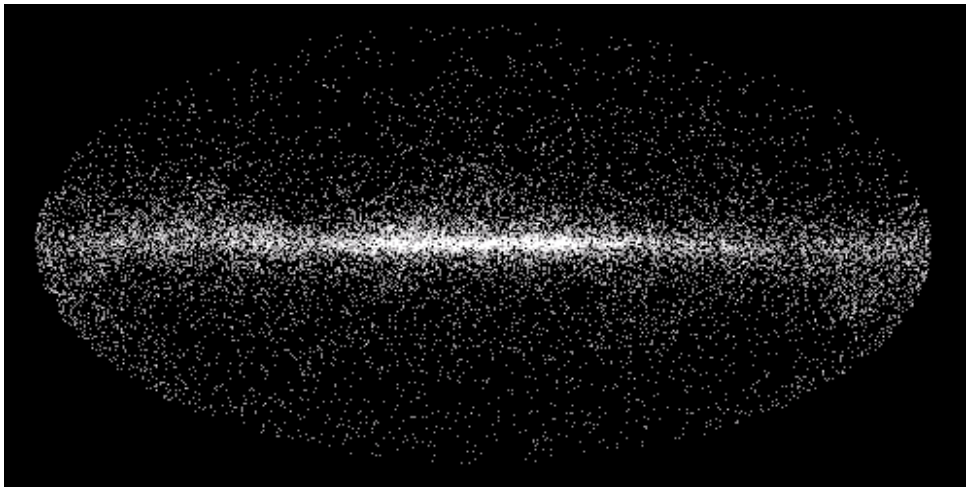
T.Burnett

on to
individual
components!

The Diffuse Truth

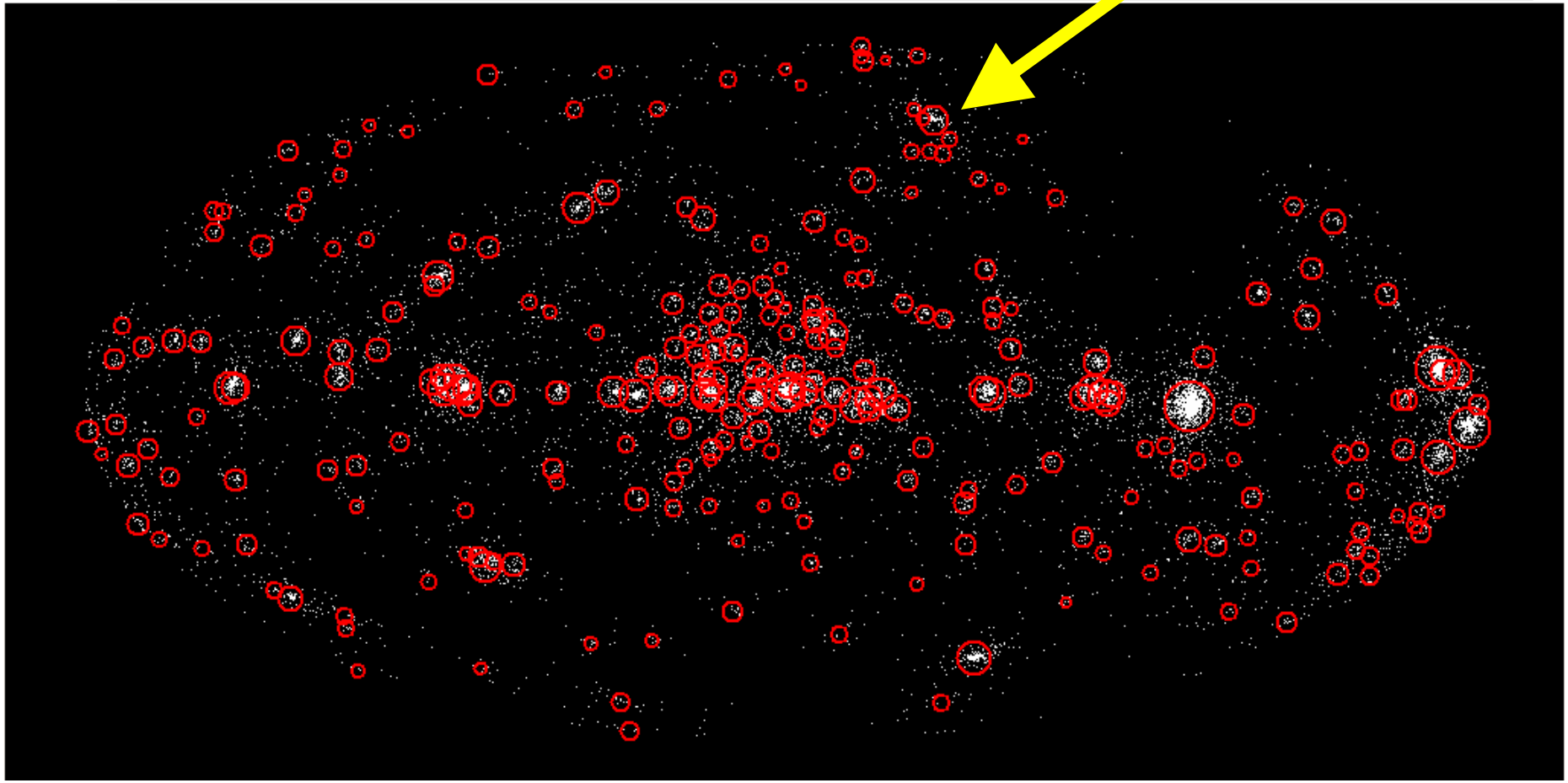


T.Burnett



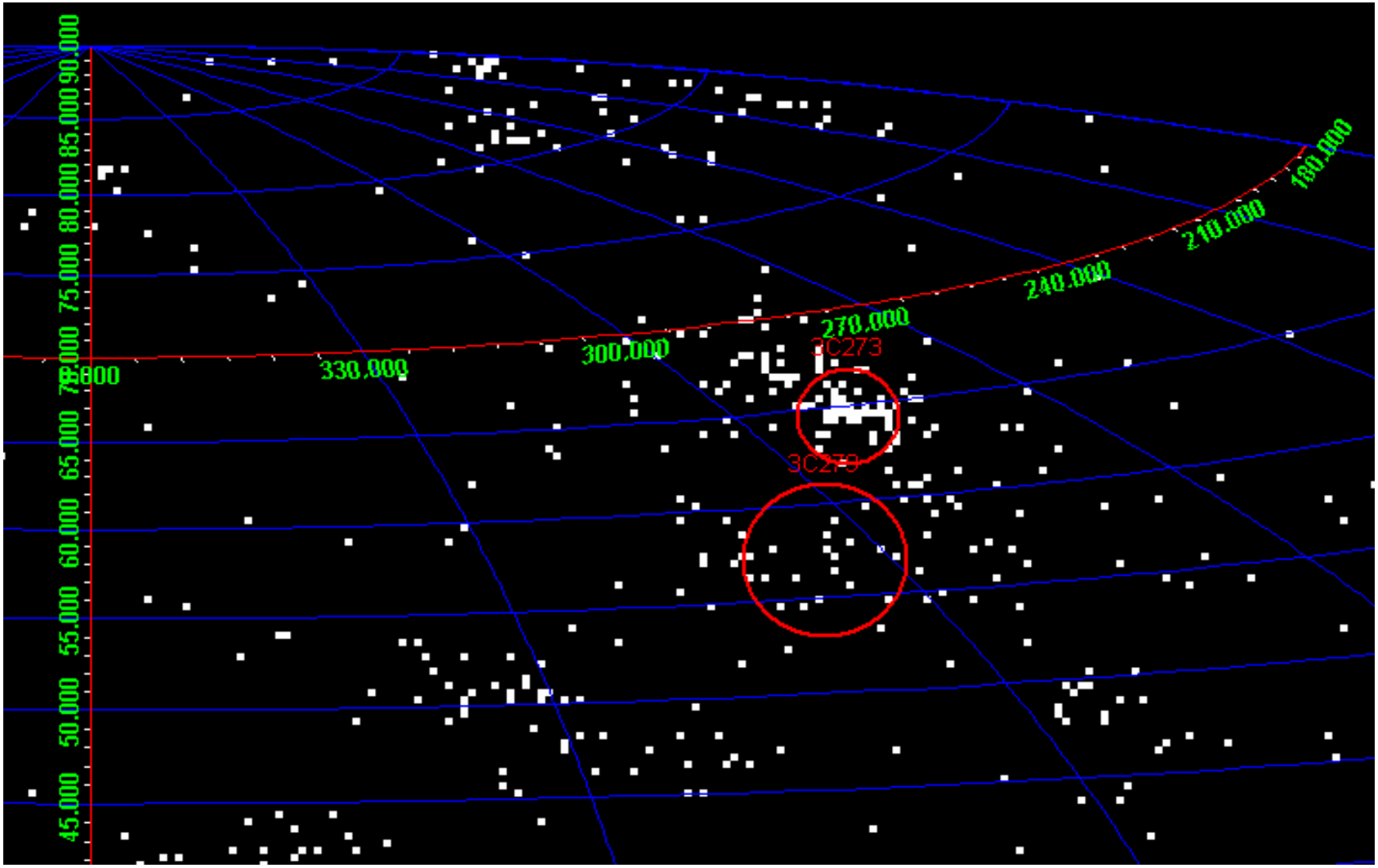
No surprises, excitement

3EG – and a twist



T.Burnett

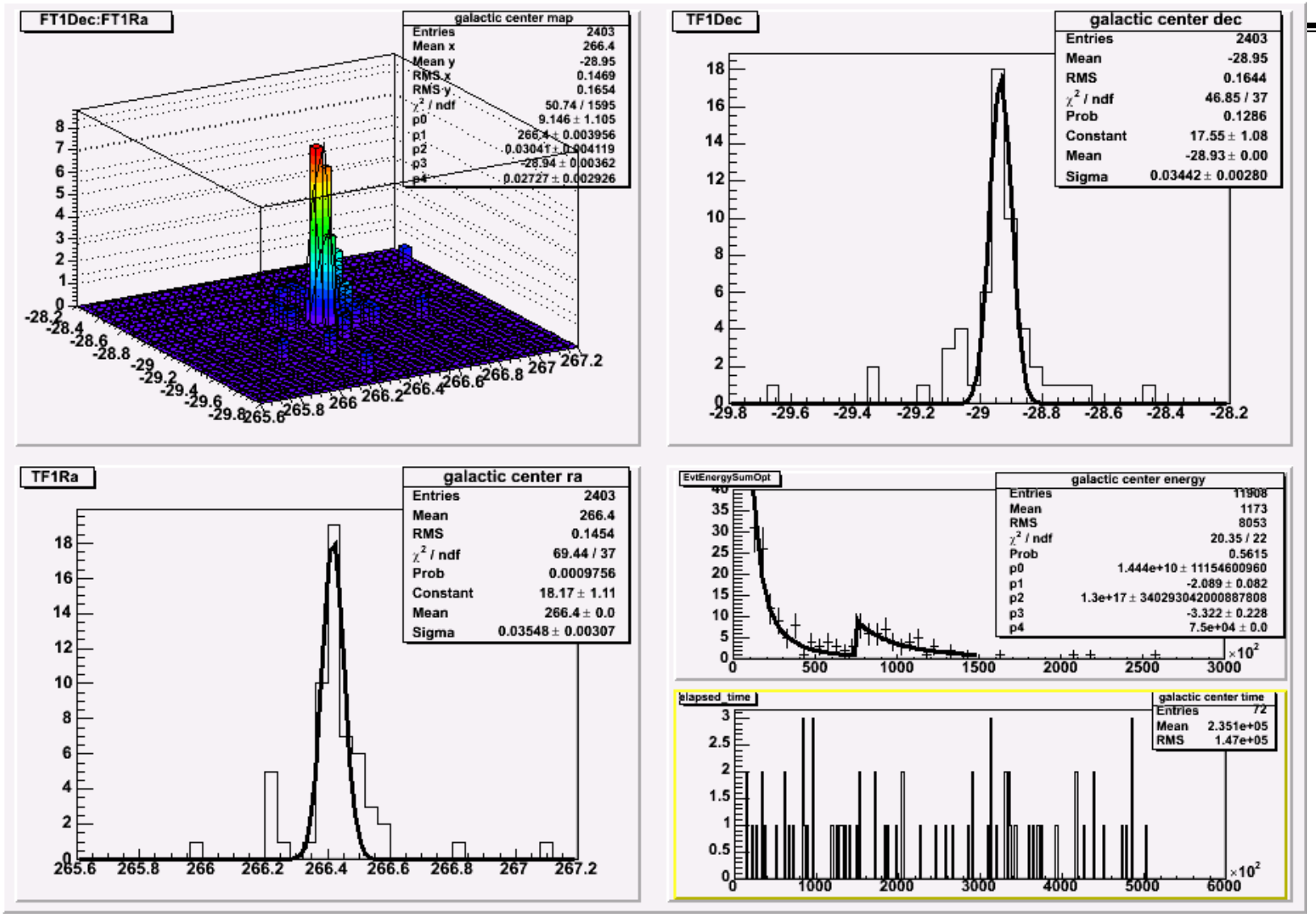
The blow-up



T.Burnett

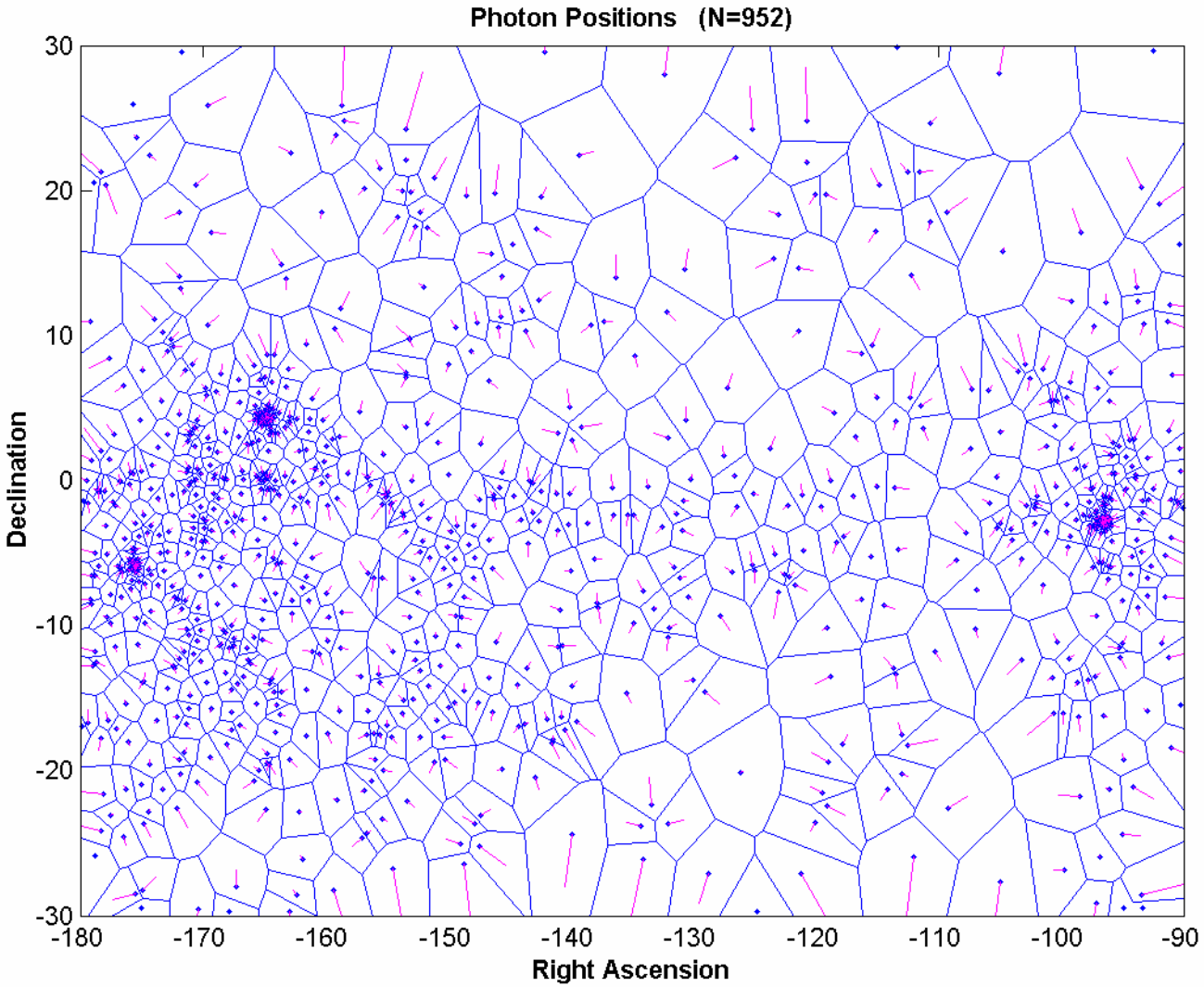
110 GeV WIMP at Galactic Center

Plot of Everything ...



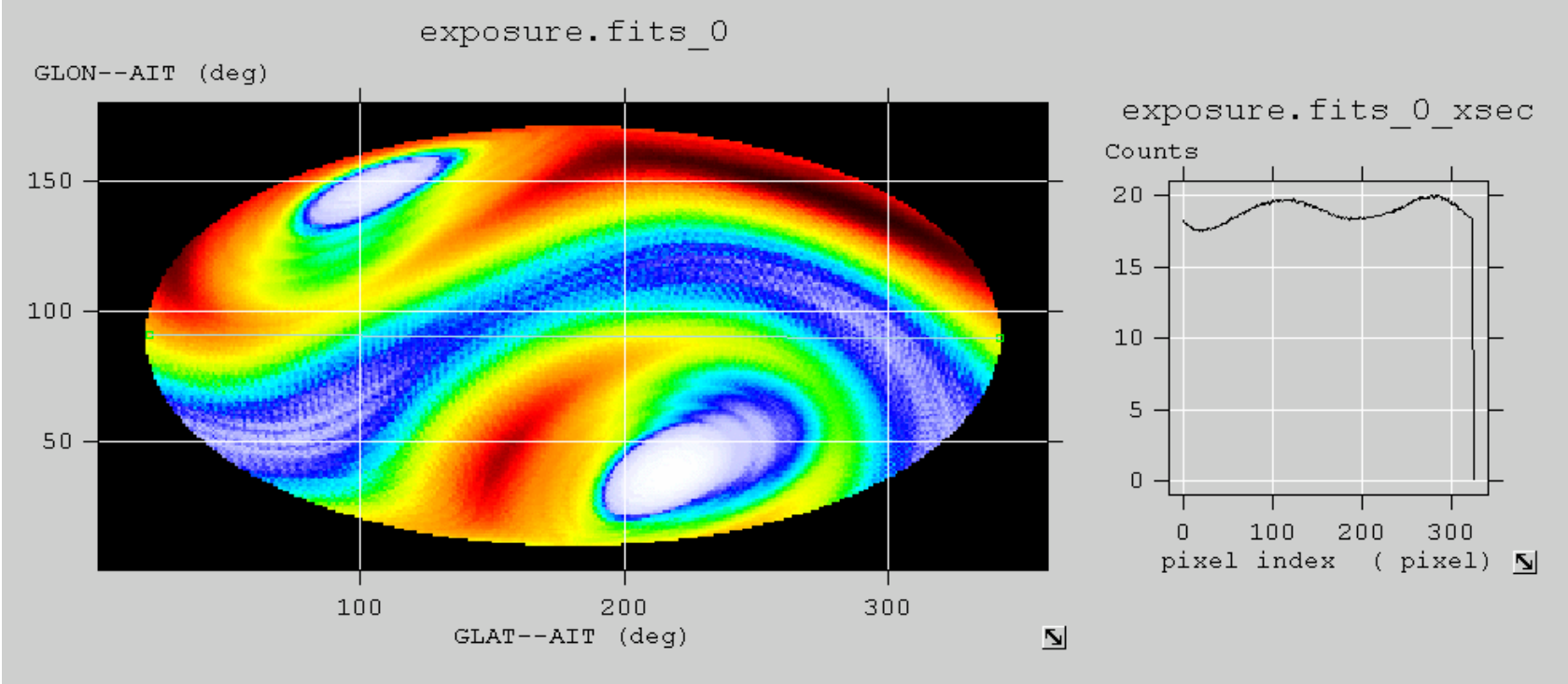
Michael Kuss

Bayesian Block source finding – Voronoi Tesselation



Jeff Scargle

Exposure: the 1-day map



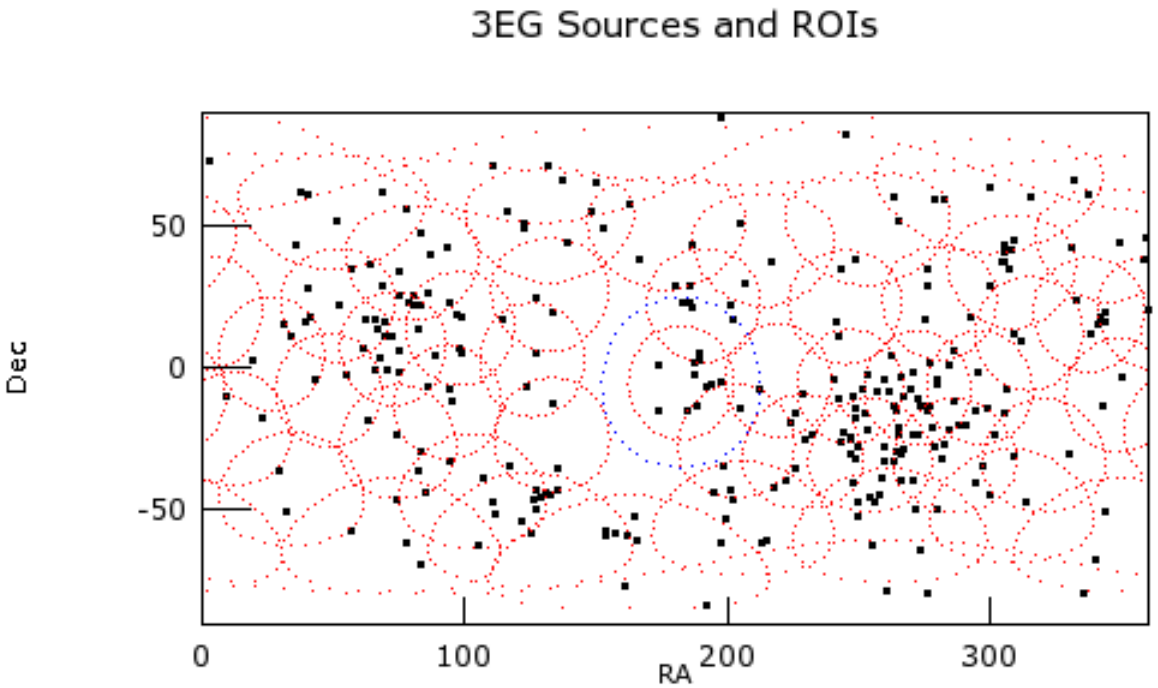
Units are percent of total exposure.

scales wrong: standard AIT projection

profile along the galactic equator

Toby Burnett

Source Finding



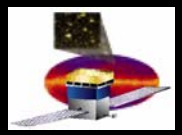
Jim Chiang

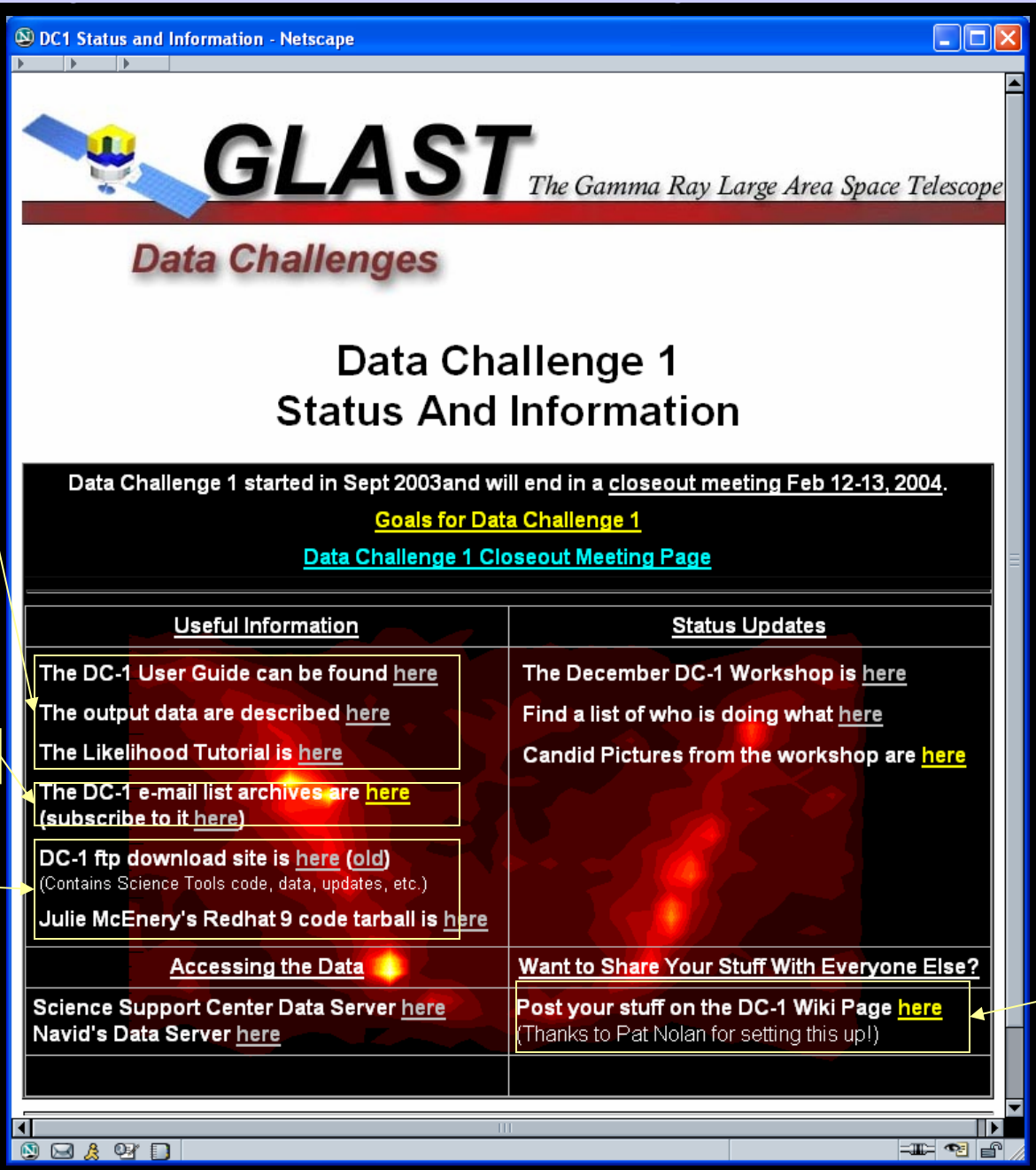
First 8 rows of catalogue

Using 3EG sources as seeds

ID	ROI	ROI dist.	Flux	index	TS	flux	index	catalog ID
0	0	1.82	8.11e-03	1.88	228.95	4.23e-03	1.85	3EG J0010+7309
1	5	11.93	3.42e-03	2.51	35.59	1.20e-03	2.70	3EG J0038-0949
2	4	7.05	1.89e-03	2.61	16.34	5.10e-04	2.63	3EG J0118+0248
3	5	10.44	1.70e-03	3.40	21.07	1.16e-03	2.50	3EG J0130-1758
4	6	7.19	2.78e-03	3.18	37.89	9.80e-04	2.89	3EG J0159-3603
5	4	11.24	1.96e-03	2.67	10.82	8.70e-04	2.23	3EG J0204+1458
6	6	8.50	2.00e-02	2.16	740.77	8.55e-03	1.99	3EG J0210-5055
7	4	10.04	3.06e-03	2.22	49.66	9.30e-04	2.03	3EG J0215+1123

http://www-glast.slac.stanford.edu/software/DataChallenges/DC1/DC1_StatusAndInfo.htm



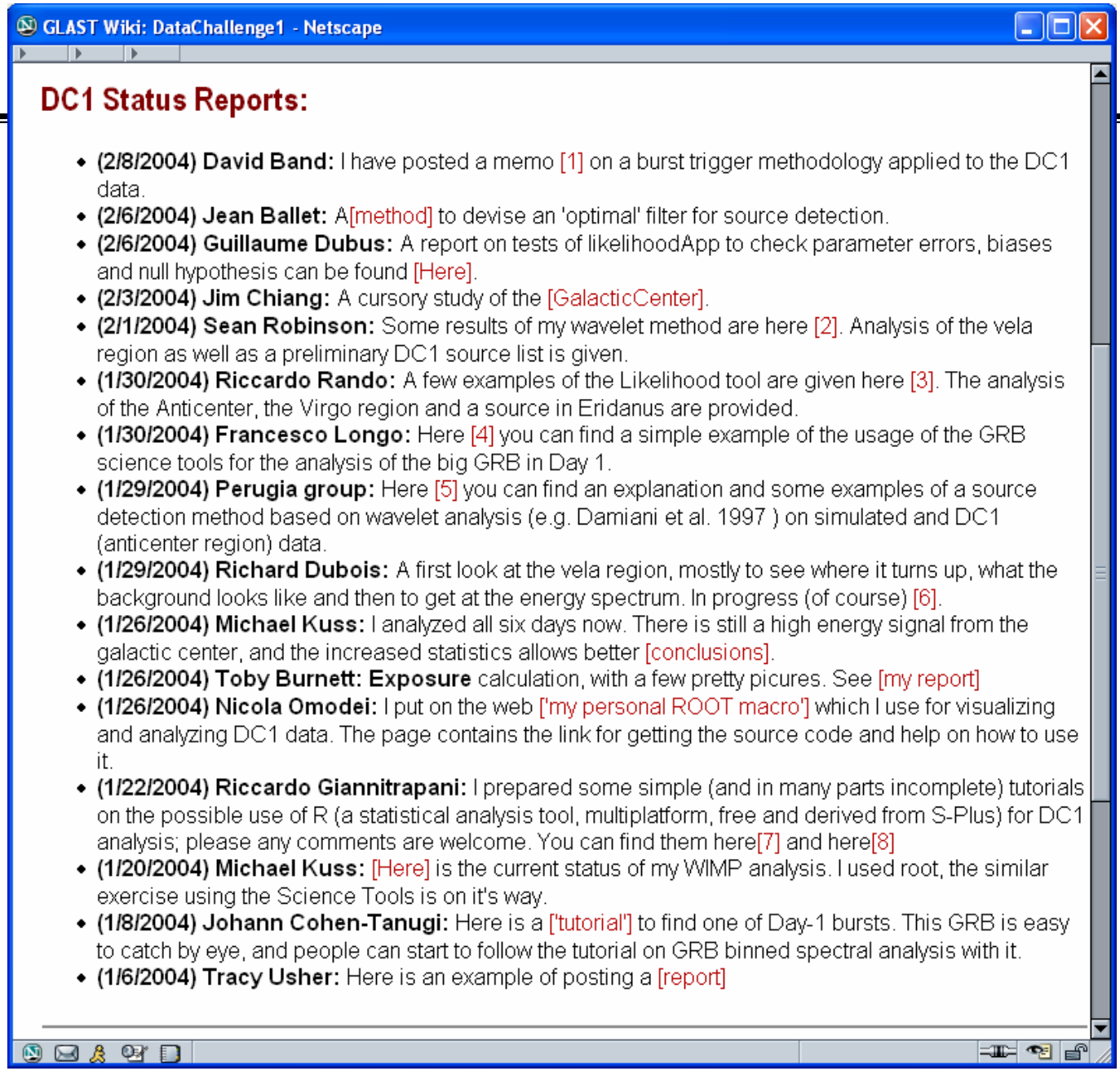


Documentation
- **User's Guide**
- **Data Description**
- **Likelihood Tutorial**

DC-1 Discussion List

Analysis Code download sites

Wiki page for sharing results!


GLAST Wiki: DataChallenge1 - Netscape
DC1 Status Reports:
(2/8/2004) David Band: I have posted a memo [1] on a burst trigger methodology applied to the DC1 data.
(2/6/2004) Jean Ballet: A[method] to devise an 'optimal' filter for source detection.
(2/6/2004) Guillaume Dubus: A report on tests of likelihoodApp to check parameter errors, biases and null hypothesis can be found [Here].
(2/3/2004) Jim Chiang: A cursory study of the [GalacticCenter].
(2/1/2004) Sean Robinson: Some results of my wavelet method are here [2]. Analysis of the vela region as well as a preliminary DC1 source list is given.
(1/30/2004) Riccardo Rando: A few examples of the Likelihood tool are given here [3]. The analysis of the Anticenter, the Virgo region and a source in Eridanus are provided.
(1/30/2004) Francesco Longo: Here [4] you can find a simple example of the usage of the GRB science tools for the analysis of the big GRB in Day 1.
(1/29/2004) Perugia group: Here [5] you can find an explanation and some examples of a source detection method based on wavelet analysis (e.g. Damiani et al. 1997) on simulated and DC1 (anticenter region) data.
(1/29/2004) Richard Dubois: A first look at the vela region, mostly to see where it turns up, what the background looks like and then to get at the energy spectrum. In progress (of course) [6].
(1/26/2004) Michael Kuss: I analyzed all six days now. There is still a high energy signal from the galactic center, and the increased statistics allows better [conclusions].
(1/26/2004) Toby Burnett: Exposure calculation, with a few pretty picures. See [my report]
(1/26/2004) Nicola Omodei: I put on the web ['my personal ROOT macro'] which I use for visualizing and analyzing DC1 data. The page contains the link for getting the source code and help on how to use it.
(1/22/2004) Riccardo Giannitrapani: I prepared some simple (and in many parts incomplete) tutorials on the possible use of R (a statistical analysis tool, multiplatform, free and derived from S-Plus) for DC1 analysis; please any comments are welcome. You can find them here[7] and here[8]
(1/20/2004) Michael Kuss: [Here] is the current status of my WIMP analysis. I used root, the similar exercise using the Science Tools is on it's way.
(1/8/2004) Johann Cohen-Tanugi: Here is a ['tutorial'] to find one of Day-1 bursts. This GRB is easy to catch by eye, and people can start to follow the tutorial on GRB binned spectral analysis with it.
(1/6/2004) Tracy Usher: Here is an example of posting a [report]

Lessons Learned

- **Analysis Issues**
  - Astrophysical data analysis
  - Software usage and reliability
  - Documentation
  - Data access and data server usage
  - UI stuff
  - Software installation and release
  - Software infrastructure & framework
  - Communication and Time frame
- **Infrastructure Issues**
  - SciTools did not run on windows at the last minute
  - We discovered problems with sources and ACD ribbons late
  - Manual handling of the processing
  - No checking of file integrity
  - Large failure rate in batch jobs (~10%)
  - Tools are not checking inputs much
  - Code distribution scripts were written manually

Closeout report will contain the details.

Strawperson Updated Plan for DC2

- **DC2, based on lessons from DC1** S.Ritz
  - 1 simulated month of all-sky survey gammas (backgrounds: see next slide)
  - key sky addition: source variability
    - AGN variability, including bright flares, quiescent periods
    - expand burst variety (and include GBM? see later slides)
    - pulsars, including Gemingas, w/ orbit position effects.
  - more realistic attitude profile
  - background rate varies with orbit position
  - more physics surprises, and add nominal hardware problems (and misalignments?), add deadtime effects and corrections
  - Analysis Goals:
    - produce toy 1-month catalog and transient releases
    - detailed point source sensitivity and localization studies
    - first systematic pulsar searches (timing!); detailed diffuse analyses
    - recognize simple hardware problems (connect with ISOC/SOG)
  - benchmark:
    - processing times, data volume, data transfers.

Flight Ops - Expected Capacity

- We routinely made use of 100-300 processors on the SLAC farm for repeated Monte Carlo simulations, lasting weeks
  - Expanding farm net to France and Italy
  - Unknown yet what our MC needs will be
  - We are very small compared to our SLAC neighbour BABAR – computing center sized for them
    - 2000-3000 CPUS; 300 TB of disk; 6 robotic silos holding ~30000 200 GB tapes total
  - SLAC computing center has guaranteed our needs for CPU and disk, including maintenance for the life of the mission.
  - Data rate less than already demonstrated MC capability
    - ~75 of today's CPUs to handle 5 hrs of data in 1 hour @ 0.15 sec/event
    - Onboard compression may make it 75 of tomorrow's CPUs too

Disk and Archives

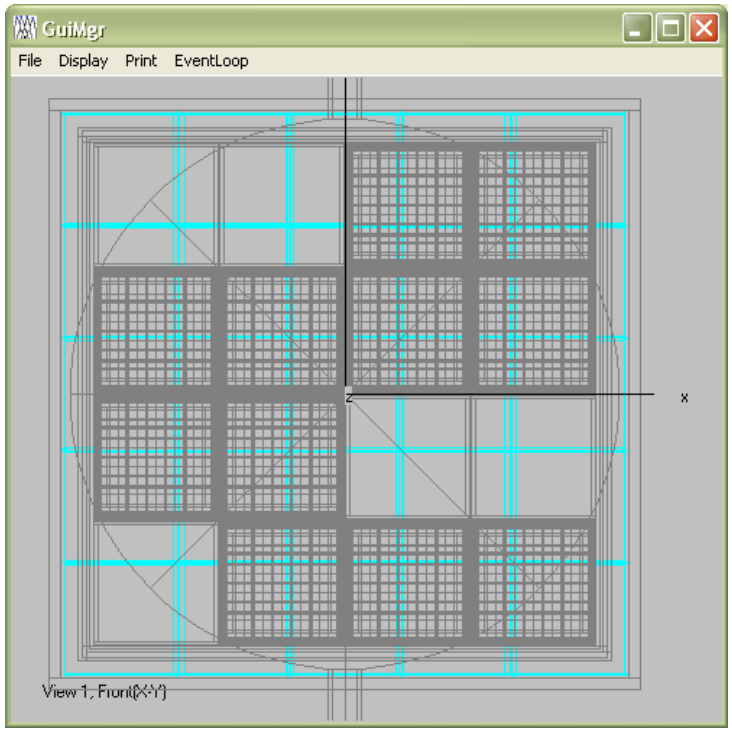
- **We expect ~10 GB raw data per day and assume comparable volume of events for MC**
  - **Leads to ~100-250 TB per year for all data types**
    - **Current filesizes and background rates**
    - **No longer as frightening – keep it all on disk**
  - **Use SLAC's mstore archiving system to keep a copy in the silo**
    - **Already practicing with it and will hook it up to OPUS**
  - **Archive all data we touch; track in dataset catalogue**

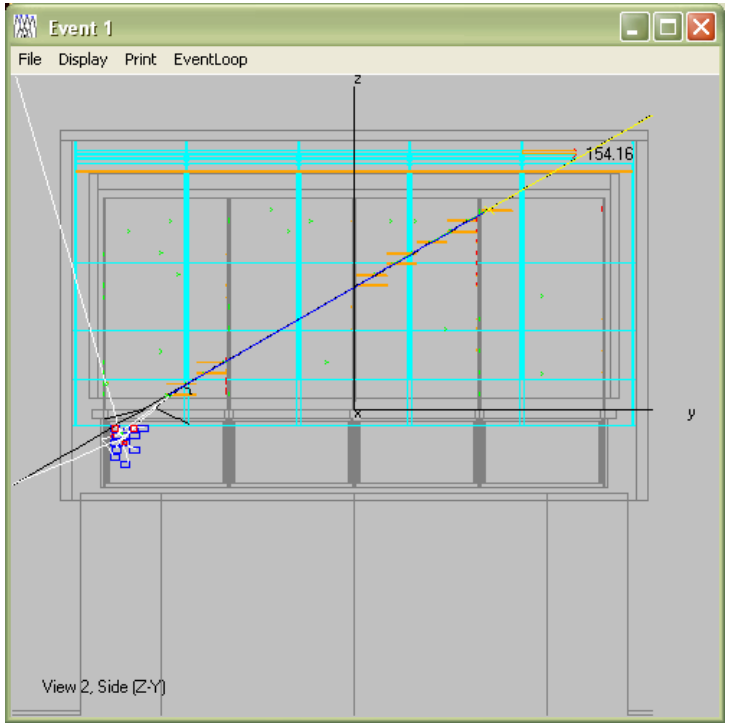
Flight Integration Support

- Simulation/Reconstruction package
  - Running stress tests now
- Calibration algorithms and infrastructure
  - TKR exercising TOT and Splits now
    - Thinking about alignments
  - Negotiating with CAL now
  - User interface for entering parameters into system underway
- Geometry
  - Flexible scheme to describe towers as they are inserted under test now
- High Level Diagnostics
  - Adapt “System Tests” to this purpose
  - Tracked in database etc
  - New version under construction
- Processing Pipeline
  - Due end April with tests demonstrating EM MC & Data handling
- Strategy is to use the same systems for Flight Integration as we expect to use for flight – databases; diagnostics system; pipeline; reconstruction, etc.

Simulating/reconstructing tower data

- Can run full sim/recon on the incremental configurations during installation.
- Uses same code as for EM1 and full 16 towers





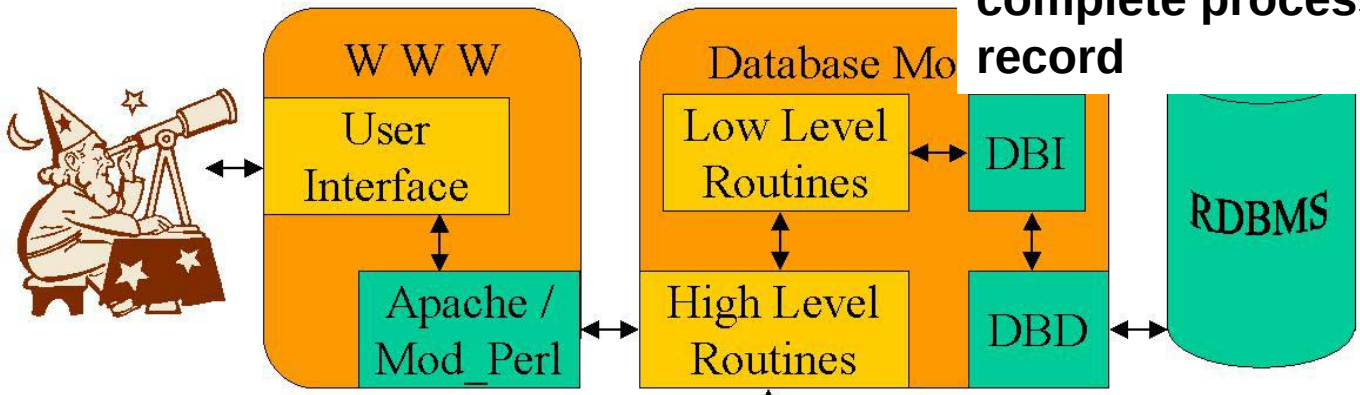
Pipeline Spec

- Function
  - The Pipeline facility has five major functions
    - automatically process Level 0 data through reconstruction (Level 1)
    - provide near real-time feedback to IOC
    - facilitate the verification and generation of new calibration constants
    - produce bulk Monte Carlo simulations
    - backup all data that passes through
- Must be able to perform these functions in parallel
- Fully configurable, parallel task chains allow great flexibility for use online as well as offline
  - Will test the online capabilities during Flight Integration

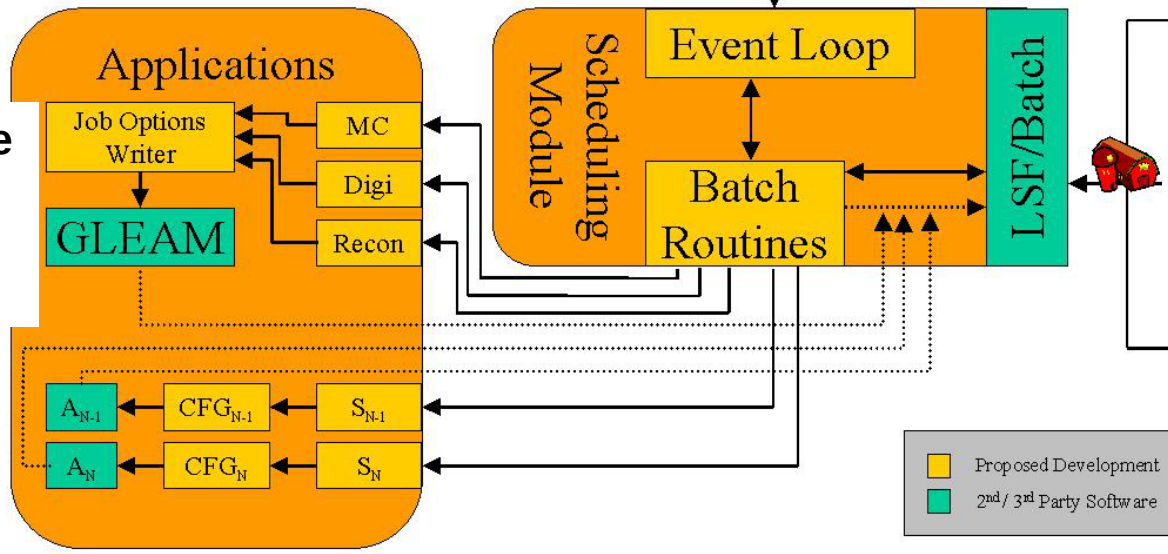
- The pipeline database and server, and diagnostics database have been specified (will need revision after prototype experience!)
  - database: LAT-TD-00553
  - server: LAT-TD-00773
  - diagnostics: LAT-TD-00876

Pipeline in Pictures

State machine + complete processing record

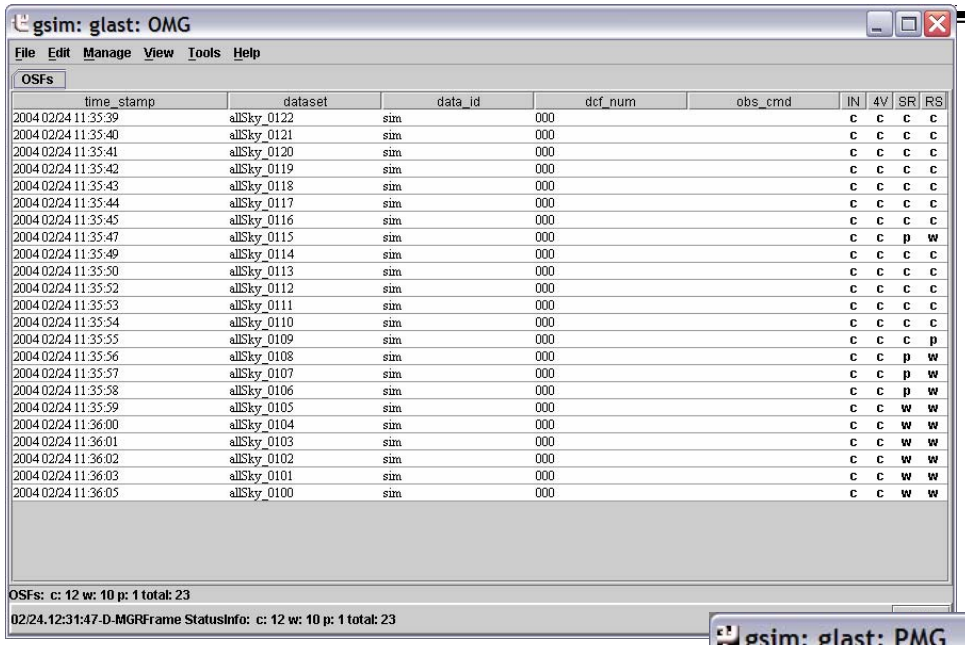


Configurable linked list of applications to run



Expandable and configurable set of processing nodes

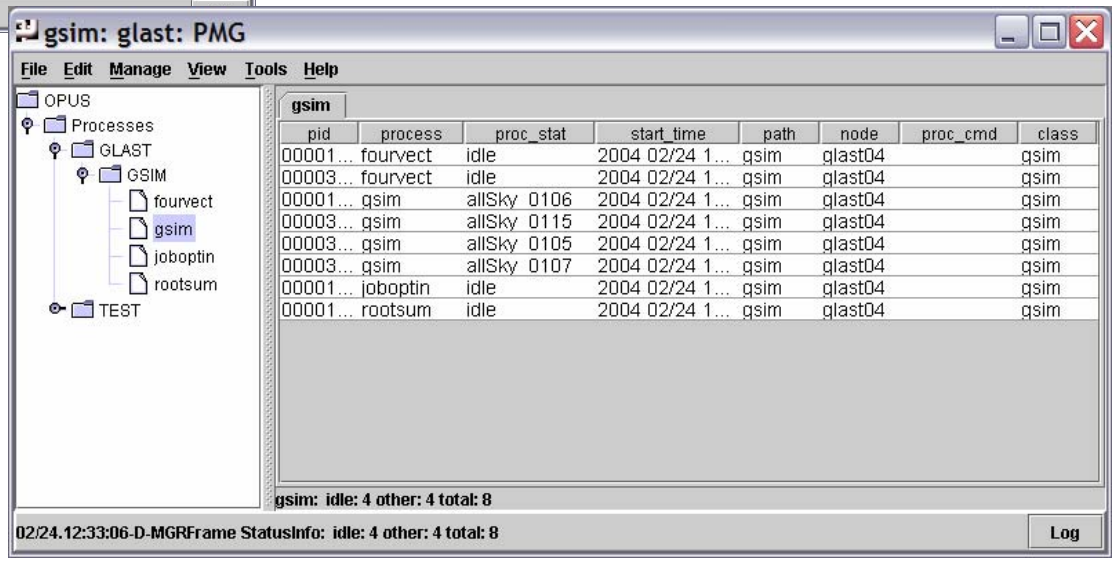
First Prototype - OPUS



Open source project from STScI

In use by several missions

Now outfitted to run DC1 dataset

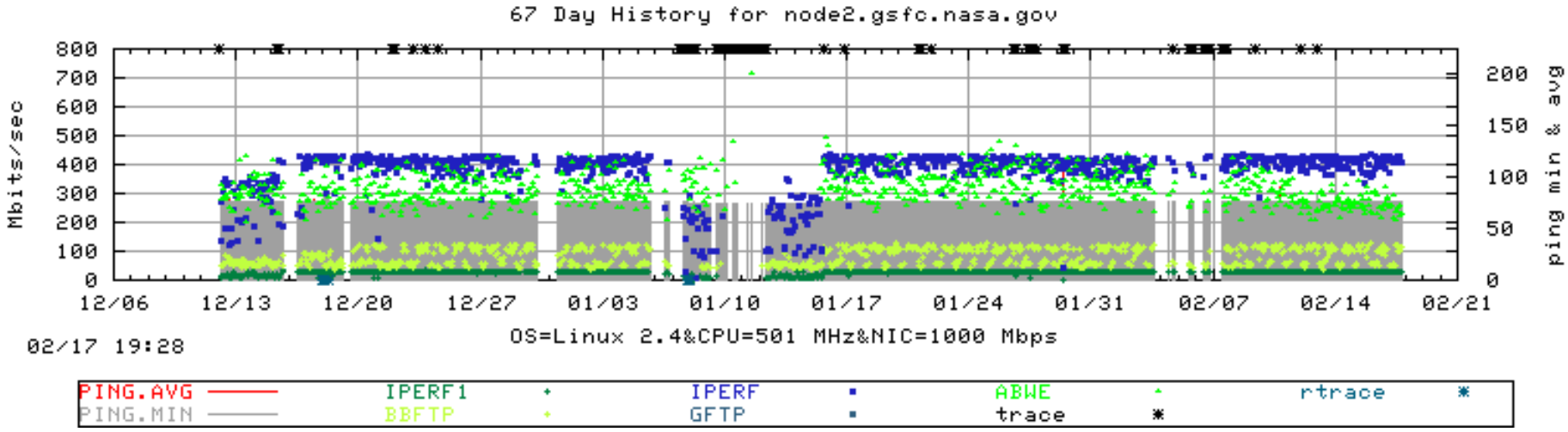


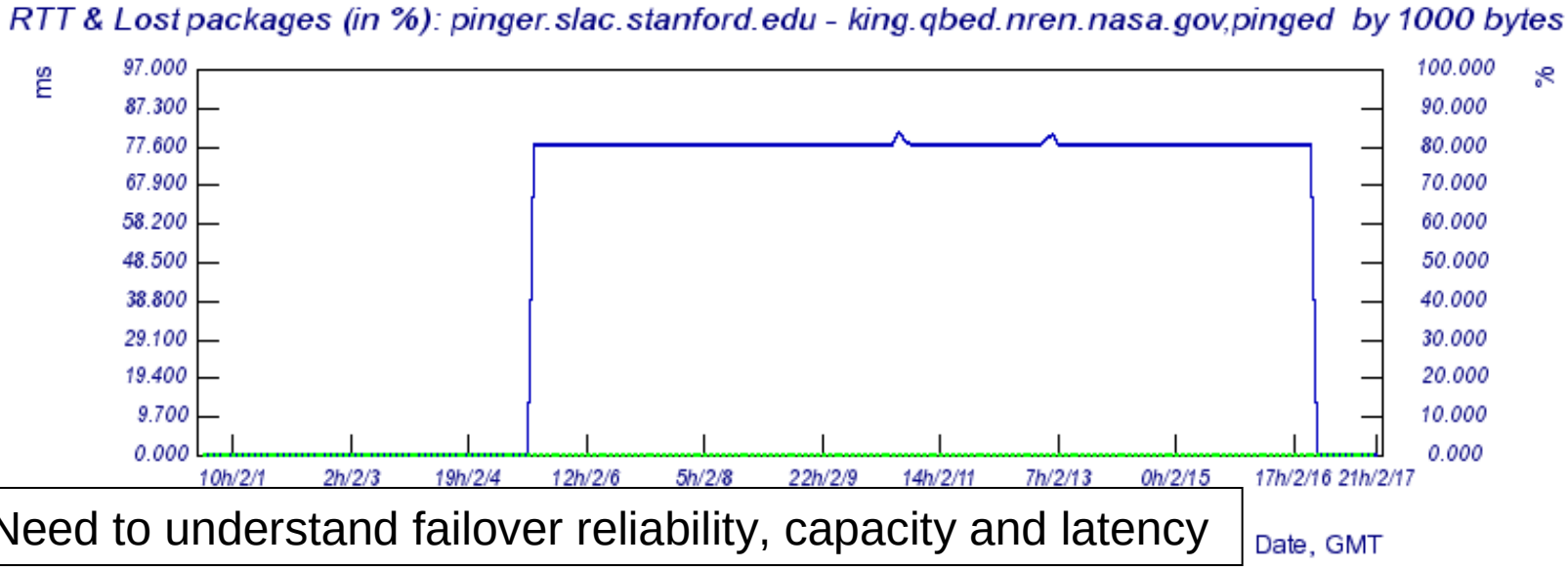
OPUS Java mangers for pipelines

ISOC Stanford/SLAC Network

- SLAC Computing Center
  - OC48 connection to outside world
  - provides data connections to MOC and SSC
  - hosts the data and processing pipeline
  - Transfers MUCH larger datasets around the world for BABAR
  - World renowned for network monitoring expertise
    - Will leverage this to understand our open internet model
  - Sadly, a great deal of expertise with enterprise security as well
- Part of ISOC expected to be in new Kavli Institute building on campus
  - Connected by fiber (~2 ms ping)
  - Mostly monitoring and communicating with processes/data at SLAC

Network Monitoring





Need to understand failover reliability, capacity and latency

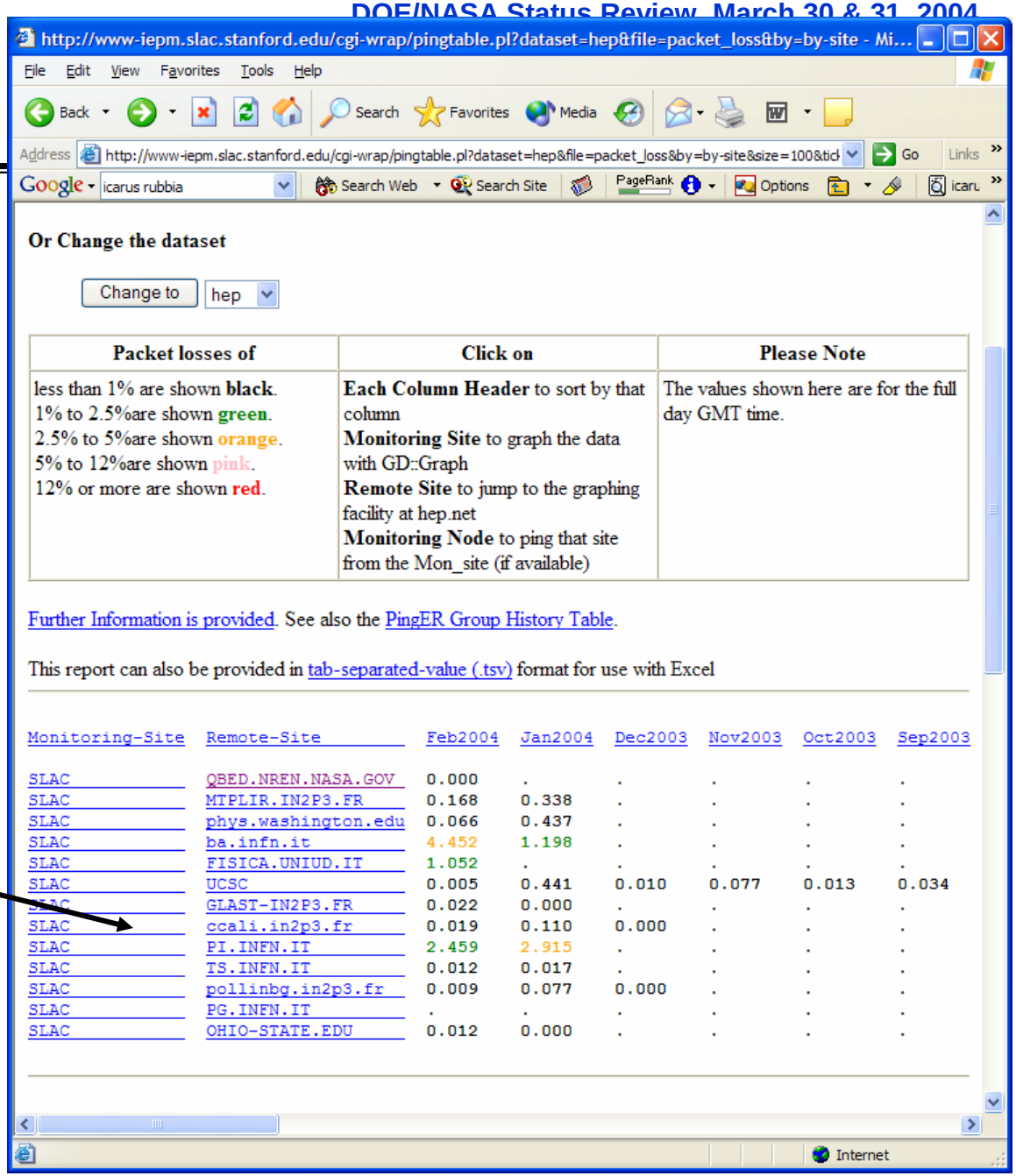
LAT Monitoring

Keep track of connections to collaboration sites

Alerts if they go down

Fodder for complaints if poor connectivity

Monitoring nodes at most LAT collaborating institutions



Outlook for next 12 Months

- **Flight Integration support**
  - **Subsystem calibration algs; Analysis; Pipeline processing**
  - **Getting priority now**
- **DC2 prep**
  - **2nd iteration of Science Tools**
    - **Apply lessons learned from DC1 + new functionality**
  - **Improve CAL digitization/reconstruction based on EM and flight hardware data**
- **Continue infrastructure improvements**
  - **Release Manager upgrades**
  - **Code distribution**
  - **Institute an issues tracker**
  - **An endless list of small improvements**

Summary

- We believe that EMs, DCs and Flight Integration will leave us ready for flight
- EM1 worked with our tools
- DC1 worked well, showing very good capabilities from sky modeling through astronomical analysis
  - Plenty of work still to do, but reasonably understood
- Will be demonstrated in DC2, 3 and LAT Integration, 16-tower cosmic ray tests and the beam test prior to launch

LAT Flight Integration in 5 months

DC2 in 9 months