Science Tools and DC1

Richard Dubois
Stanford Linear Accelerator Center
richard@slac.stanford.edu

In lieu of Seth Digel

With input & slides from Seth, Jim Chiang, Bob Schaefer, Yasushi Ikebe, Tom Stephens and David Band

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

Outline

- The players
- Introduction to the Full Toolset
- The Subset for DC1
- Outstanding Issues before launch of DC1

What's a science tool?

- Jargon for the analysis software and databases that we'll need for deriving scientific results from the LAT data
- Implicit in the term is that the analysis is high-level, like studying cosmic sources of gamma rays
- Also implicit in the term is that the analysis relies on an abstract characterization of the LAT – via its response functions – and that background rejection will meet the SRD requirements

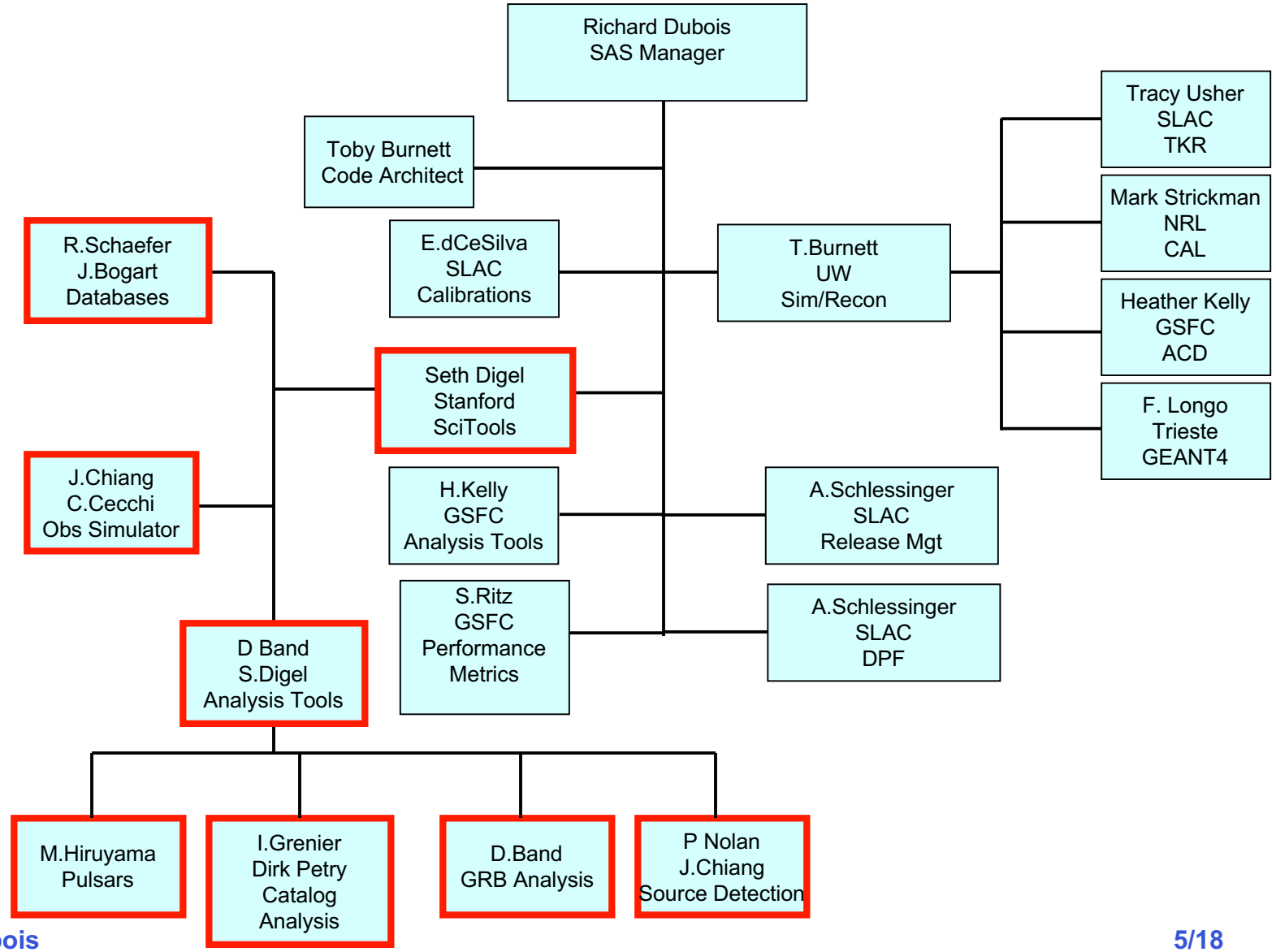
- The Standard Analysis Environment is the group of science tools that we have agreed to develop jointly with the SSC for us and for guest investigators to use
- Within the LAT team we've got additional tools to develop
  - e.g., interstellar emission model, transient source searches, point source catalog generation, in-flight (high-level) calibration monitoring

Who we are

- LAT collab. scientists and software developers & assoc. scientists with free time on our hands + SSC
- Self-reported: 49 people, 18+ FTEs (the majority of whom are available now)
- Cautionary notes: 13 institutions, ~<0.4 FTE/person

SU/HEPL
SU/SLAC
GSFC/SSC
GSFC/UMBC
CEA/Saclay
IN2P3/LLR
IN2P3/Bordeau
INFN/Perugia
INFN/Trieste
INFN/Udine
INFN/Pisa
INFN/Bari
Univ. Washington

SAS Organization



Development of Science Tools

- **Extensive planning on which tools are needed to do science - and their requirements**
  - **One set of tools for all – "astronomy standard"**
  - **Had external review (9/2002) to see if we are on the right track**
    - **No major problems noted**

http://www-glast.slac.stanford.edu/ScienceTools/reviews/sept02/report/review_091602.pdf

- **In progress with the GSSC**
  - **Joint oversight group**
    - **Sorted out technical basis (HEASARC standards; support of community; re-use of LAT developments)**
    - **Has not had to meet for some time – a good sign!**

- **"Core" group watching over infrastructure**

Main Science Tools

Package	Description
Likelihood(*)	Workhorse model fitting for detection & characterization of cosmic gamma-ray sources
Level 1 database access(*)	Extracts desired event data
Exposure calculation(*)	Uses IRFs, pointing, livetime etc. for deriving calibrated source fluxes
Source identification	Identifies gamma-ray sources with cataloged counterparts at other wavelengths
GRB analysis(*)	Temporal(**) and spectral analyses of burst profiles
Pulsar analysis	Phase folding & period searching of gamma-ray pulsars and candidates
Observation simulator	High level simulation of observations of the gamma-ray sky with the LAT

(*) to be included in DC1
(**) not included in DC1

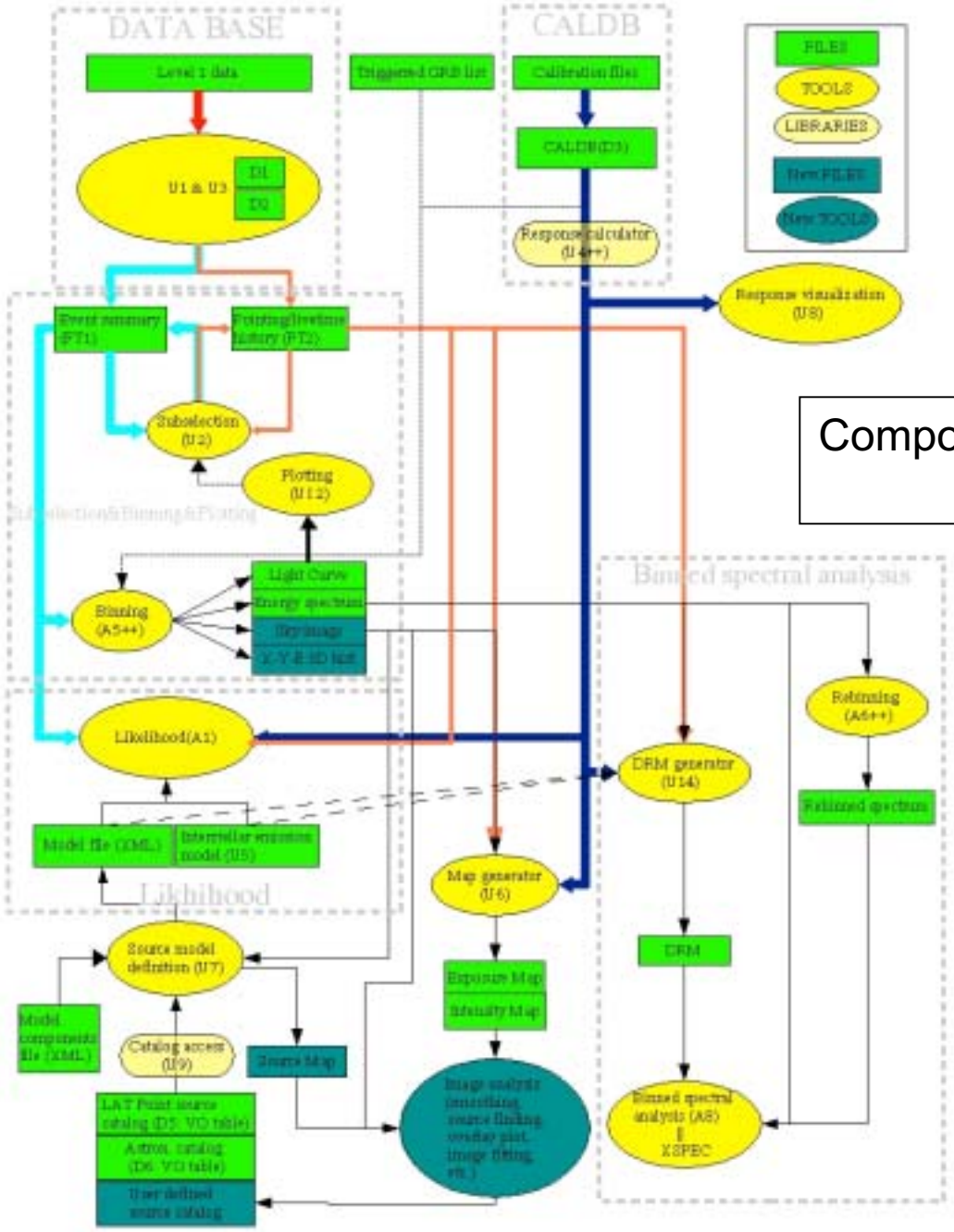
Science Tools Toolkit

Package	Description	Provider	Status
PIL, HOOPS	IRAF parameter access	HEASARC	In use
cfitsio	FITS file manipulation	HEASARC	In use
XSPEC, Sherpa	For GRB spectral modeling	HEA standards	Under consideration
Root	Plotting, gui etc	HEP standard	In preparation
python	Scripting	World standard	In use
doxygen	Code doc tool	World standard	In use
Visual C++/gnu	Development envs	World standards	In use
CMT	Code mgmt tool	HEP standard	In use
cvsweb	cvs web viewer	World standard	In use
cvs	File version mgmt	World standard	In use

Yasushi Ikebe

DC1 Plan:

- ingest data into the databases from Gleam
- extract photons from the databases
- analyze them with the likelihood tool



Components Needed for DC1

Likelihood 1

- Uses unbinned log-likelihood as the fit statistic
- Three different optimizers
  - including Minuit
    - so covariance matrices as well as parameter values can be estimated
  - Model parameters can be fixed or simply bounded as well as scaled
- Status of source modeling
  - Spatial distributions
    - Point sources (as delta-functions in sky location)
    - Diffuse sources whose angular distributions of emission can be described either parametrically (e.g., distributed as a 2D Gaussian) or by using a FITS image as a template (e.g., the EGRET Galactic diffuse model).

Likelihood 2

- **Status of source modeling**
  - **Spectrally:**
    - **Each spatial component is modeled as having a single photon spectrum. Therefore spectral variation across an extended diffuse source must be implemented by dividing the source into smaller components**
    - **Spectra are presently modeled as power-laws, but extending the library of spectral functions will be simple, and a standard menu of spectral models will be provided (e.g., broken power-laws, IR absorption, etc.)**
  - **Spectral components can be combined multiplicatively and additively, just like in XSPEC, Sherpa, ISIS, etc. Fitting for source location is possible, but not yet fully implemented**

Likelihood 3

- **User Interface:**
  - Use Python to glue the pieces together
  - Sources can be defined/added/modified directly through Python, or the source models can be read in from an XML file.
  - Region-of-interest definitions are also accessed via the Python interface or read in from an xml file.
  - Use of HOOPS (interface to IRAF parameters) is planned.

- **Performance:**
  - Fitting a 92 parameter model, comprising 44 point sources, Galactic and extragalactic emission for 70,000 events (4 days sim time) simulated for the Cygnus region takes about 3 hours on a 1.8GHz Pentium 4 box running RH 7.2.

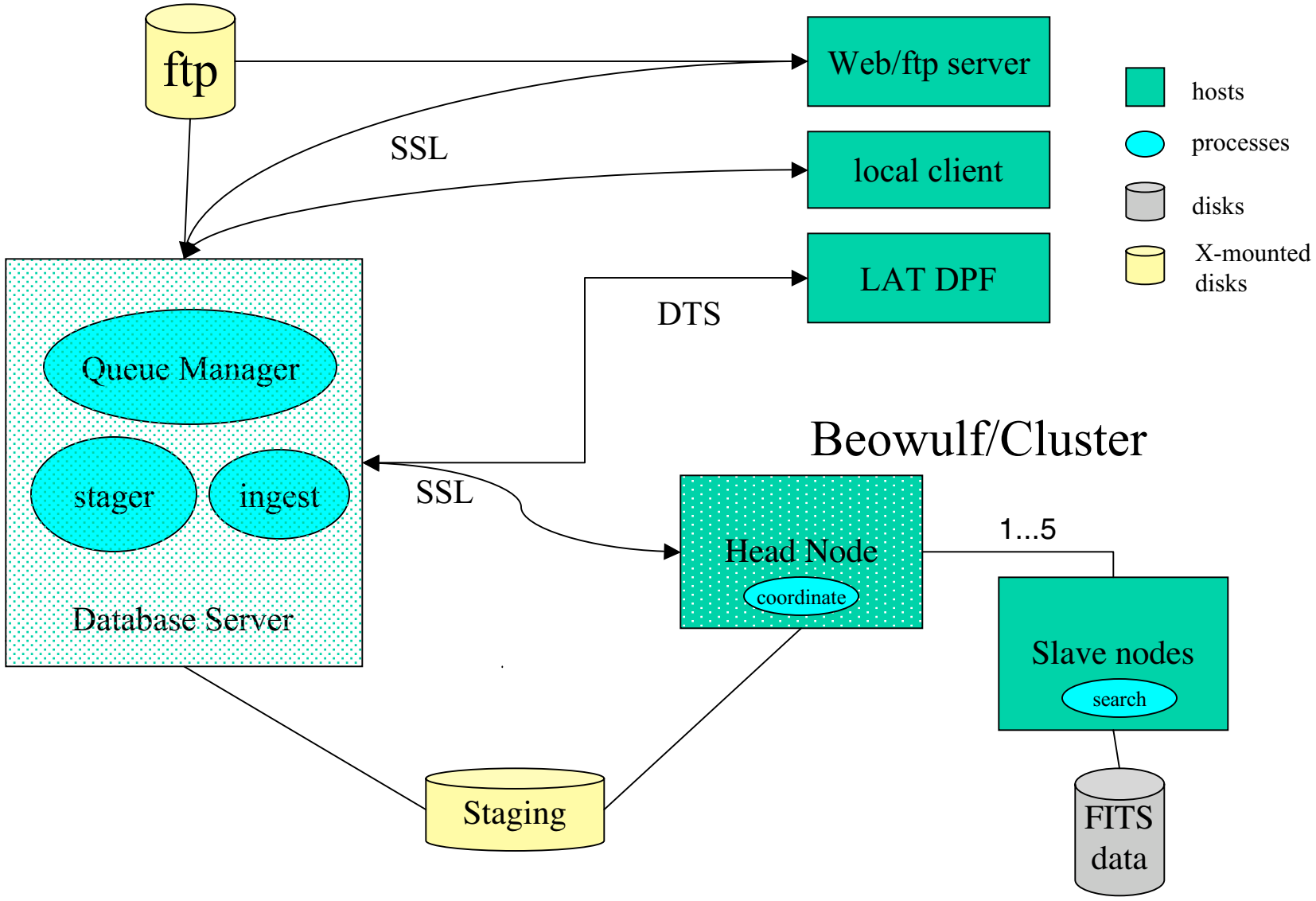
Observation Simulator(s)

- Not needed for DC1 per se, but useful for testing the toolkit
- Three versions exist on CVS:
  - light_sim: Perugia group (Claudia, Francesca, Monica)
  - FastSim: Brian Baughman and Bill Atwood
  - observationSim: uses the flux package and its source definitions.
- Data from the latter version have been analyzed by the Likelihood package. The resulting fits yield source parameters that are consistent with those input to the simulations.
- A month long simulated data set exists for the Cygnus region produced by light_sim and observationSim
  - Non-likelihood analyses using wavelets, independent component analysis, and Bayesian Blocks are proceeding on these data.
- Available types of sources (through the flux package):
  - Point Sources
  - Diffuse Sources
  - Using a FITS image as a template.
  - Uniformly distributed over a specified solid angle.
  - Flaring AGNs

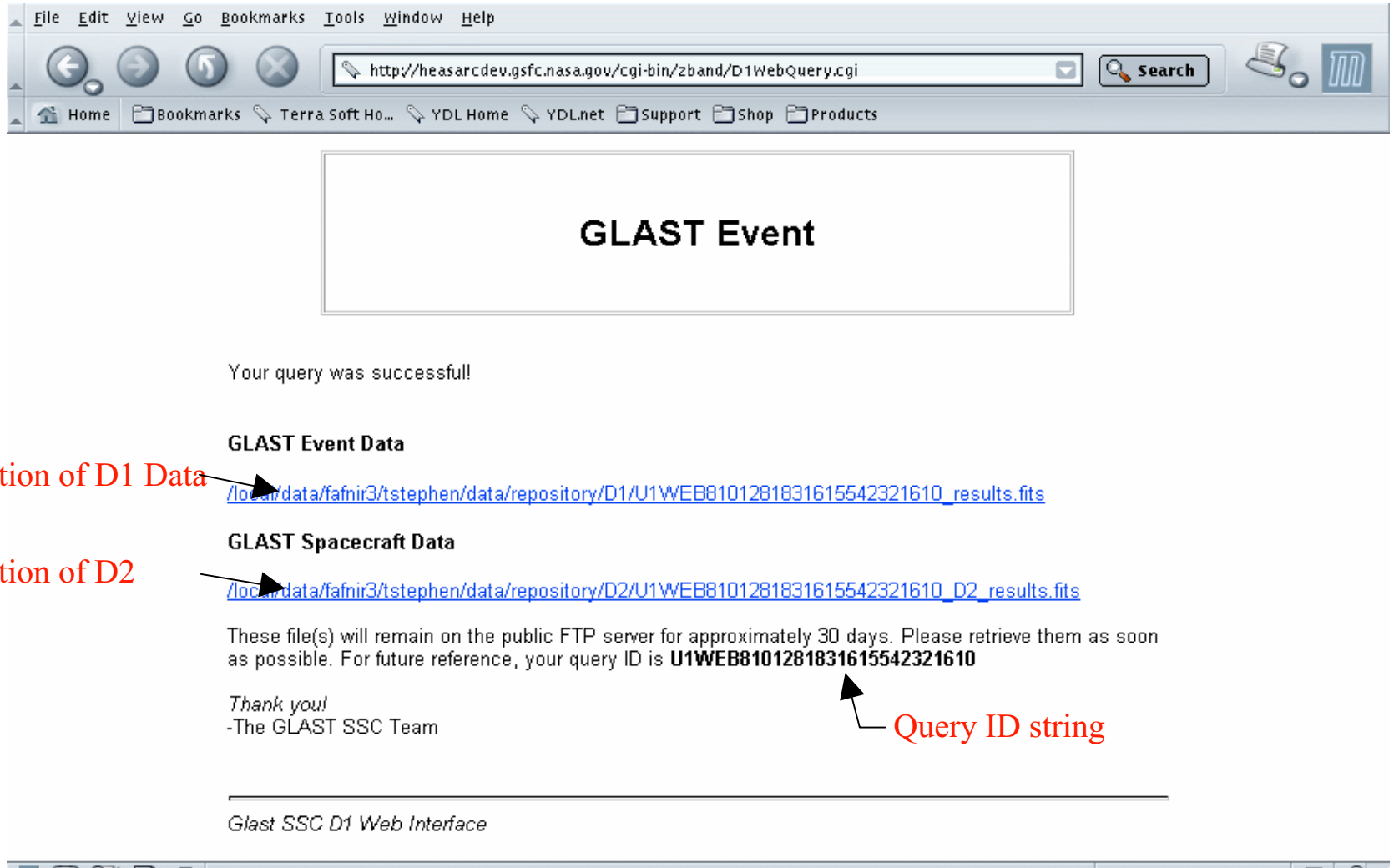
Database definition

- D1 - Event Database:
  - Choice of Database
  - Position – coordinates in RA/Dec
    - Retrieve events various shaped areas – circle, box, ellipse
    - Adjustable size of search area
    - Adjustable rotation of search area
  - Time – start and end time in MJD or UTC (Gregorian)
  - Energy – minimum and maximum energy in GeV
  - Query results in an FT1 file which contains all data matching the selection.
  - Actually two databases - a separate photon database (called photon database) and one containing all events (called event database).
- D2 - Spacecraft Database: LAT pointing, livetime, and mode history data queryable by (most commonly) time range, spacecraft mode, and other spacecraft pointing parameters
  - returns the data in an FT2 file

Photon Database Access



U1/U3 Web Results Page



Location of D1 Data

Location of D2 Data

Query ID string

Databases Status

- Databases being tested at GSSC
- Web Pages being made to look more like other Search pages at HEASARC - plan to add other features (e.g. Simbad/NED lookup of sources)
- Beowulf currently being configured.
- Publicly accessible web page and ftp area being set up
- Should be working well before DC1.

Level1 1 DB contents: http://glast.gsfc.nasa.gov/ssc/dev/fits_def/definitionFT1.html
Exposure DB contents: http://glast.gsfc.nasa.gov/ssc/dev/fits_def/definitionFT2.html

Issues prior to DC1

- First parameterizations of IRFs following Bill's performance work
  - Formatted in CALDB
- Access to L1 photon and exposure data post Gleam
  - Provide FITS files to SSC for ingest into the databases
  - Transfer mechanism to SSC (DTS is initial choice)
- Getting the processing pipeline in shape
  - Planning increasing stress tests starting in October
  - Still have backup scripts in case OPUS not ready for mid-Nov launch
- Otherwise continue to improve Likelihood and L1 DB access tools