THE STANDARD ENVIRONMENT FOR THE ANALYSIS OF LAT DATA

Seth Digel (Stanford University/HEPL)

David Band (GLAST SSC—GSFC/UMBC)

for the SSC-LAT Software Working Group and the Science Tools Working Group

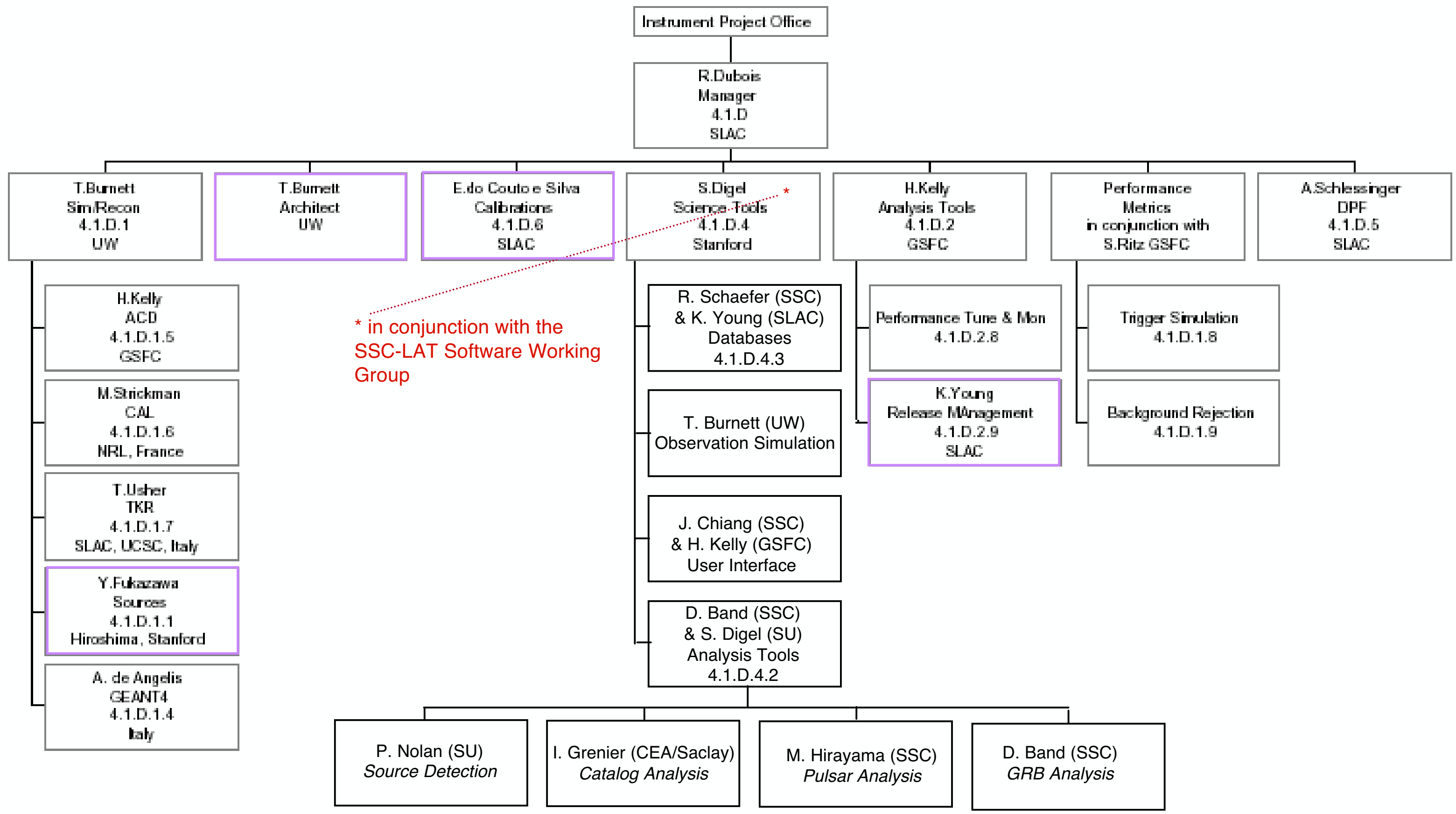
OUTLINE

- Introduction
- Participants
  - LAT Team
  - SSC
- Design Considerations
- Definition Process
- Components of the Analysis Environment
  - Review of Data Release Schedule
- Development Concept
  - Management
  - Software Development
  - Testing
  - Schedule

INTRODUCTION

- The standard analysis environment consists of the tools and databases needed for routine analysis of LAT data.
- This environment will be used by both the LAT team and the general scientific community.
- This environment was defined jointly by the LAT team and the SSC, but will be developed under the LAT team's management with SSC participation.
- The analysis environment does not support all possible analyses of LAT data. Not included, for example:
  - Analysis of multi-gamma events or cosmic rays
  - High-resolution spectroscopy of extended regions
  - Software for developing the point source catalog

SCIENCE TOOLS WITHIN LAT SAS



~~9/5/02~~

- **Science Tools also relates to *Architect, Calibrations, Release Management,* and *Sources* boxes of the LAT instrument simulation effort**
- **Details of organization will be discussed with management plan**

The SSC Within the GLAST Project

- The SSC reports to Mission Operations and Ground Systems (manager: Dennis Small) within the GLAST Project
- On scientific matters the SSC also answers to the Project Scientist (Jonathan Ormes)
- The SSC is a constituent of LHEA’s OGIP
- SSC scientists are generally not members of the LAT or GBM teams
- The SSC is responsible for:
  - the guest investigator program
  - the mission timeline (includes support for TOOs, commands)
  - providing data & analysis software to the scientific community
  - archiving data & software in the HEASARC
  - supplying the Italian mirror site with data & software
  - supporting (logistically & scientifically) the Project Scientist, the Science Working Group, and the Users’ Committee
  - some data processing (e.g., exposure maps)

Members of the SSC

The following are present full and part-time SSC members

- Jay Norris—manager
- David Band—science lead
- Dave Davis—databases
- Yasushi Ikebe—calibrations
- Masaharu Hirayama—LAT scientist
- Dirk Petry—user services
- Jim Chiang—ambassador to LIOC
- Valerie Connaughton—GBM scientist, ambassador to GIOC
- Jerry Bonnell—GRBs/PR
- Bob Schaefer—databases
- Cathie Meetre (part time)—operations
- Robin Corbet (part time)—operations
- Sandhia Bansal—programmer
- Chunhui Pan—programmer
- Sandy Barnes—administrator
- JD Myers (part time)—webmaster

DESIGN CONSIDERATIONS

- The analysis of LAT data will be complex because of:
  - The LAT's large FOV
  - Scanning will be the standard observing mode (photons from a source will be observed with different angles to the detector)
  - A large PSF at low energy
  - A large data space populated by a small number of photons
- The environment will be compatible with HEASARC standards
  - Files will be in FITS format with HEASARC keywords and GLAST extensions
  - Data are extracted by utilities that isolate the analysis tools from databases whose format or architecture may change
  - Data and tools will eventually be transferred to the HEASARC
- Accommodating the user community--the software must be usable by remote investigators with limited institutional knowledge

DEFINITION PROCESS

- The process formally began in January, 2000, with a meeting at SLAC (before the SSC was constituted)
- The data products the analysis environment will use were defined by the GLAST-wide Data Products Working Group (mid-2001 to early 2002). Its report will be the basis of ICDs.
- SSC-LAT Software Working Group established in March, 2002 to define the analysis environment. 3 LAT and 3 SSC representatives, co-chaired by Seth Digel and David Band
- Workshop at SLAC in June, 2002, reviewed the proposed analysis environment (~30 attendees LAT+SSC)
- Definition of analysis environment guided by use cases, anticipation of science possible with LAT data, and expertise analyzing high energy astrophysics data
- A formal review of our plans is in progress; see http://www-glast.slac.stanford.edu/ScienceTools/reviews/sept02/

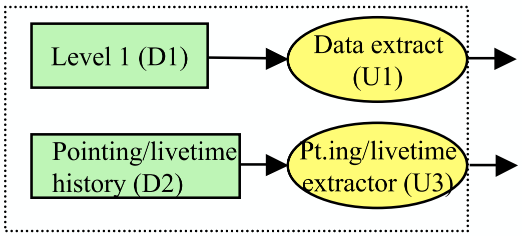
USER INTERFACE

- Users will be able to run the tools from a command line or through an over-arching GUI system. Both interfaces common in high energy astrophysics.
- The command line interface lends itself to scripting by the user to automate certain analyses. Scripting will be possible within and among tools.
- The GUI system will bind together the tools. Users can interact with the tools through GUIs.
- The GUIs will have a common look and feel: common terminology, standard placement of buttons, etc.

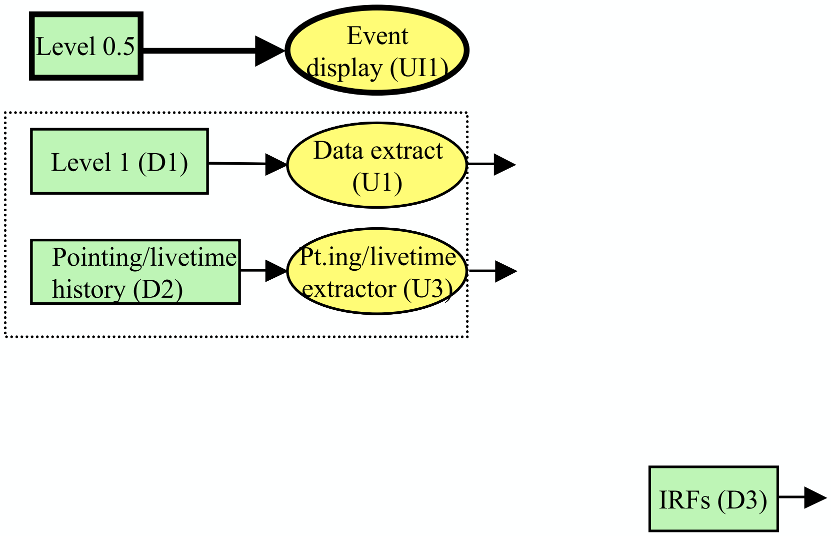
WALKTHROUGH OF THE STANDARD ANALYSIS ENVIRONMENT

- Schematic illustration of the data flow and how the tools relate to each other. Not all inputs (e.g., from user) are explicitly indicated
  - Detailed descriptions of each component are available
- The tool's identification scheme (letter + number) is for convenience; the distinction between U & A can be subtle
  - D – database (in a general sense)
  - U – utility (supporting analyses)
  - A – analysis tool
  - O – observation simulation
  - UI – user interface (common aspects to utilities & analysis tools)
- Common data types that can pass between tools are defined but not included in the diagram
- User Interface aspects of the SAE--such as Image/plot display, Command line interface & scripting, and GUI & Web access--are not shown explicitly in the diagram

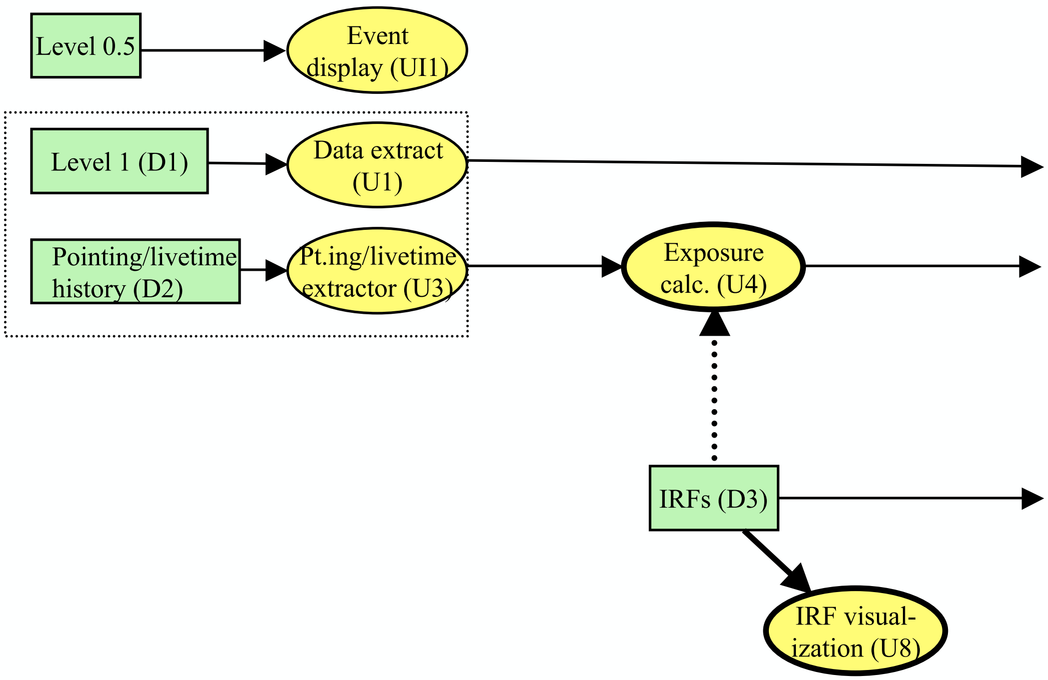
Event data, pointing & livetime history, and response functions



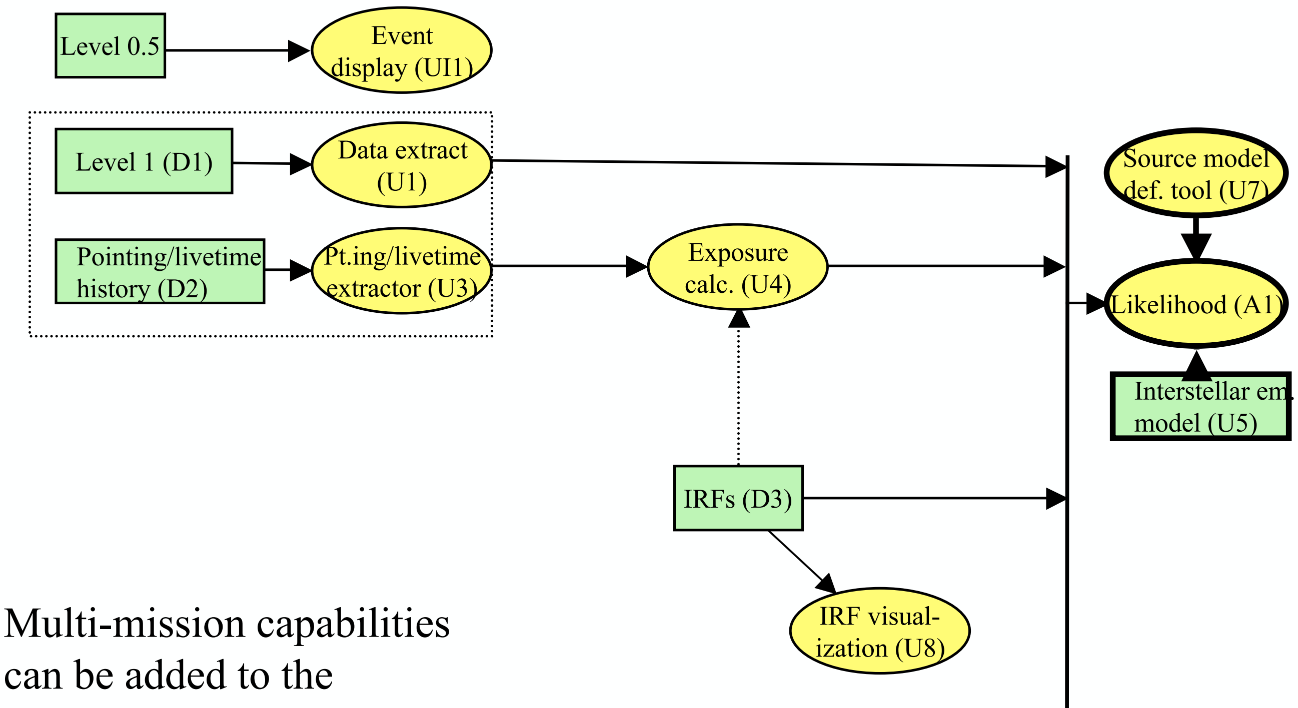
Event display



Exposure

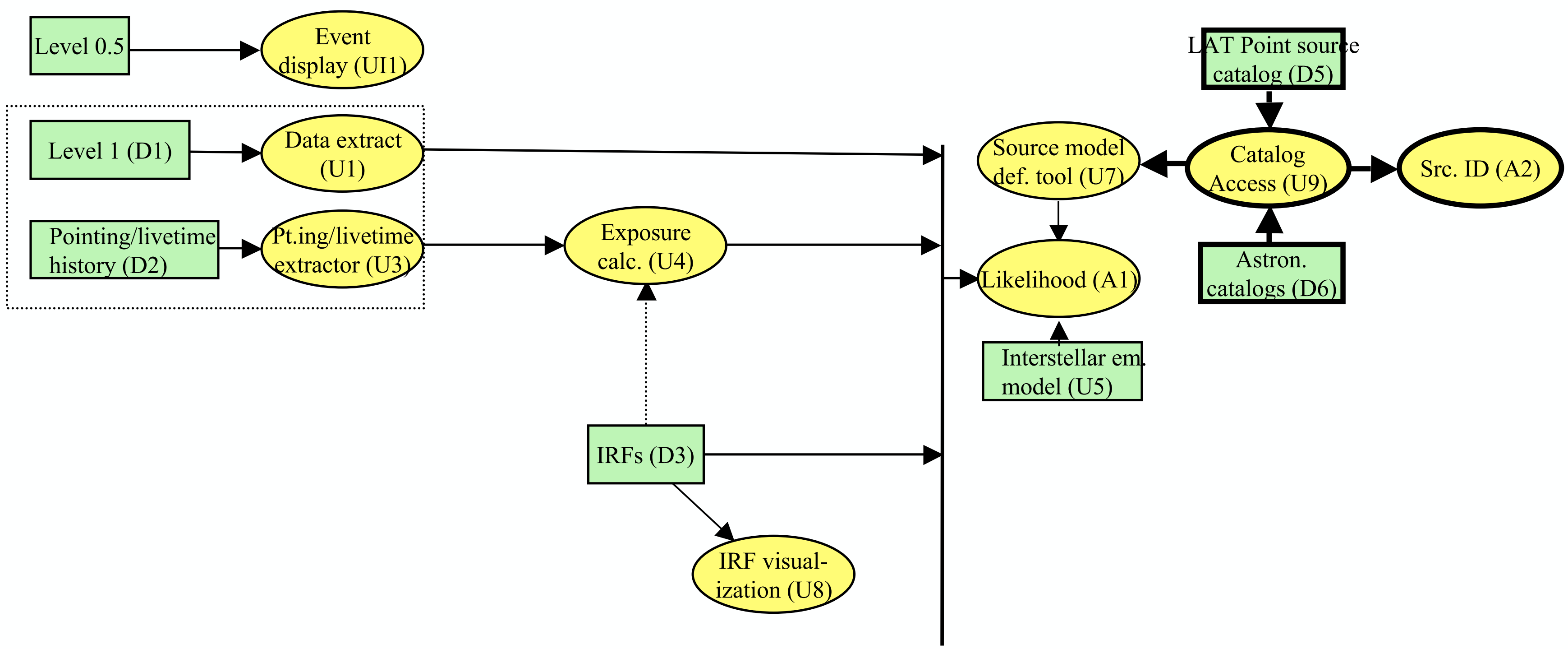

Level 0.5
Event display (UI1)
Level 1 (D1)
Data extract (U1)
Pointing/livetime history (D2)
Pt.ing/livetime extractor (U3)
Exposure calc. (U4)
IRFs (D3)
IRF visual-ization (U8)

Likelihood analysis

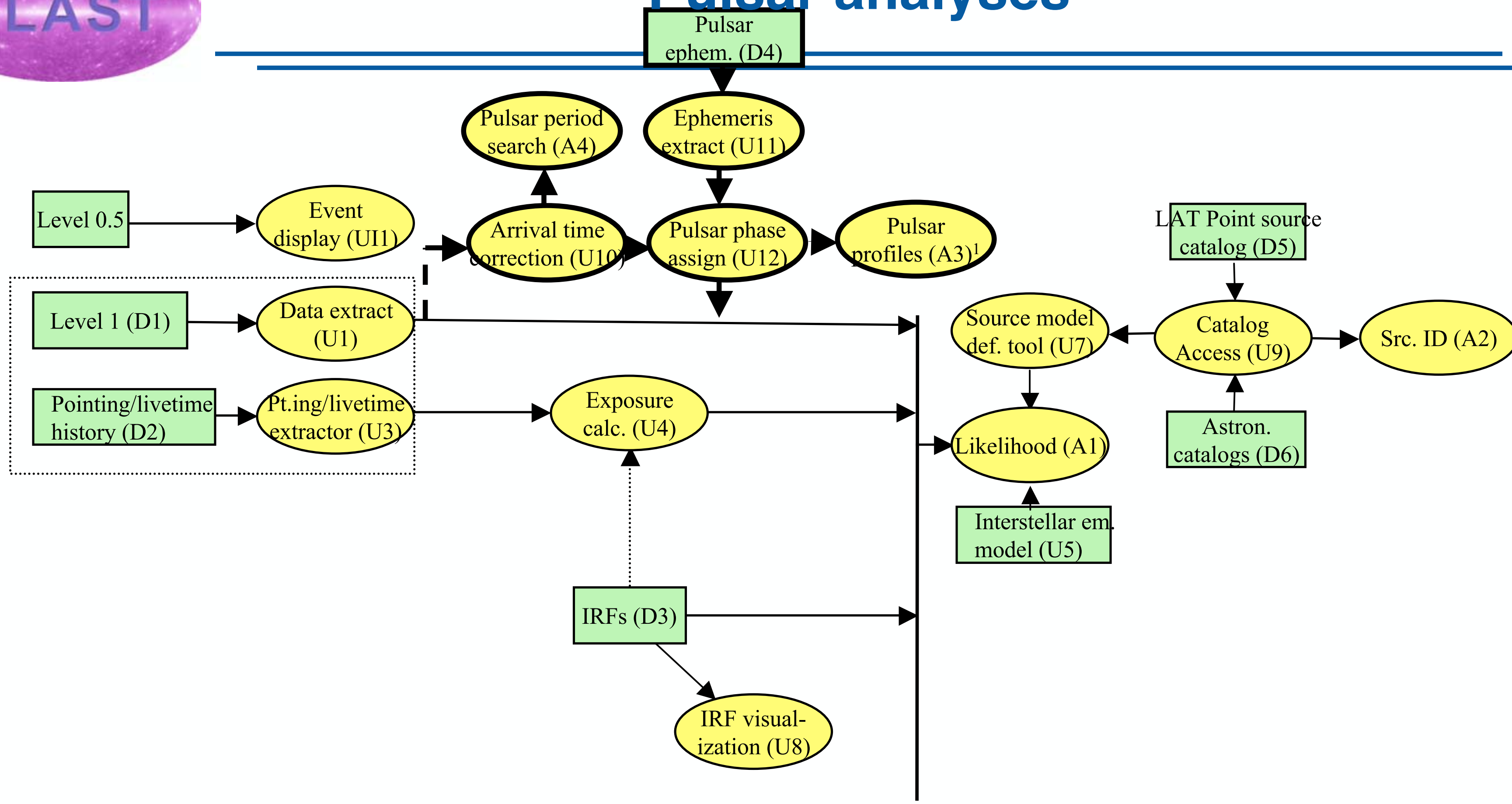


Multi-mission capabilities can be added to the likelihood tool; we will study whether such analysis is computationally feasible.

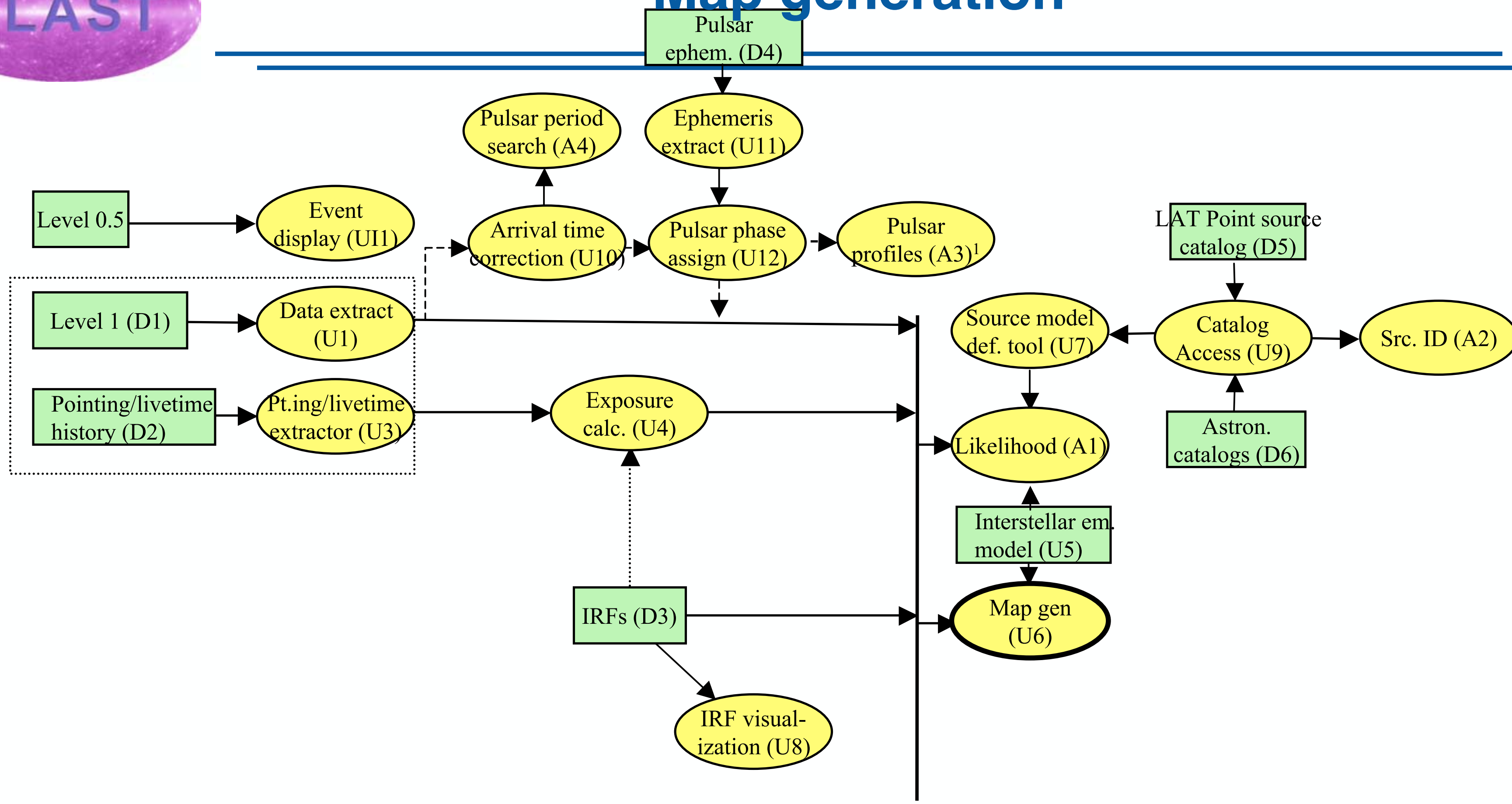
Point source catalog, astronomical catalogs, & source identification



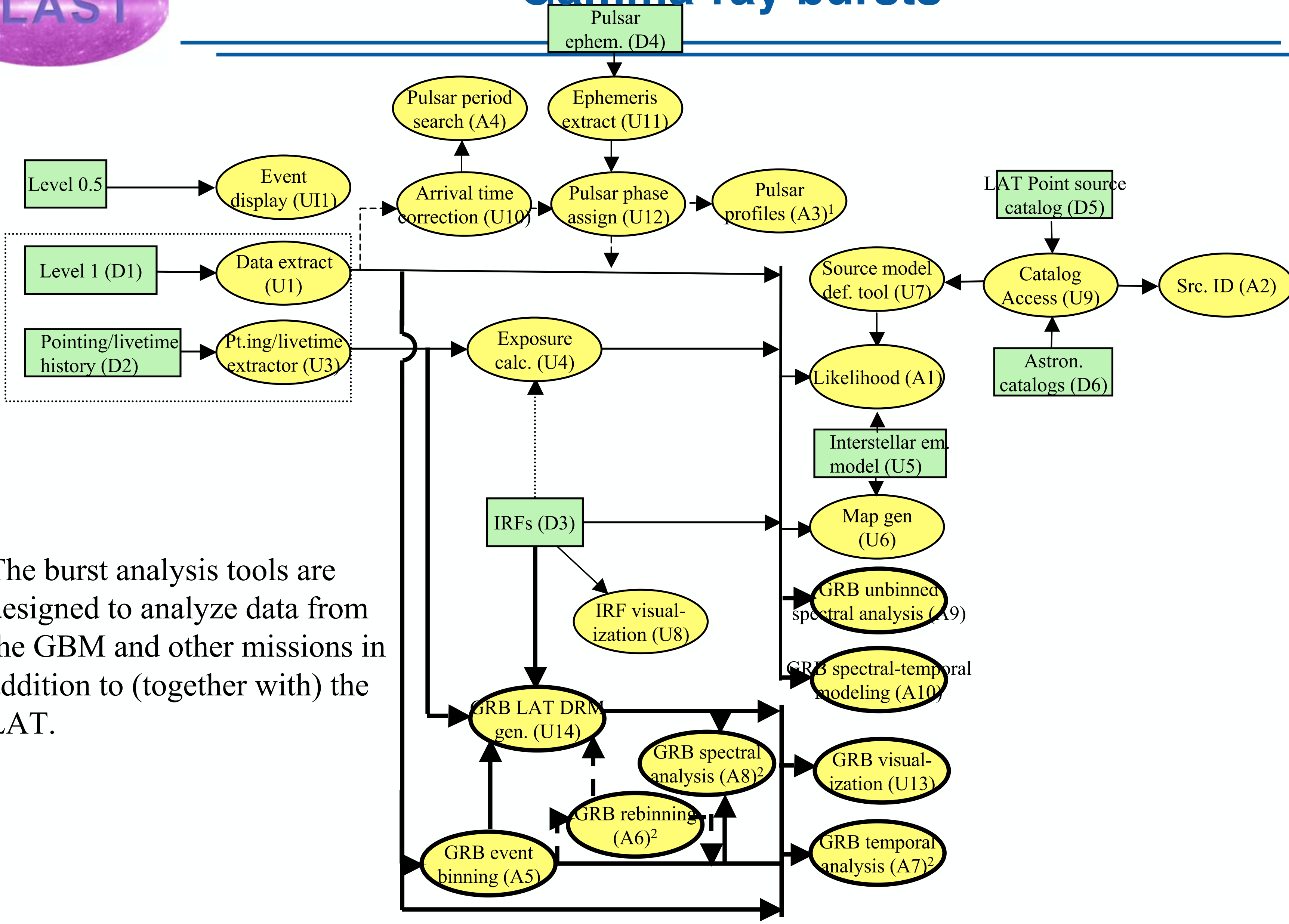
Pulsar analyses



Map generation

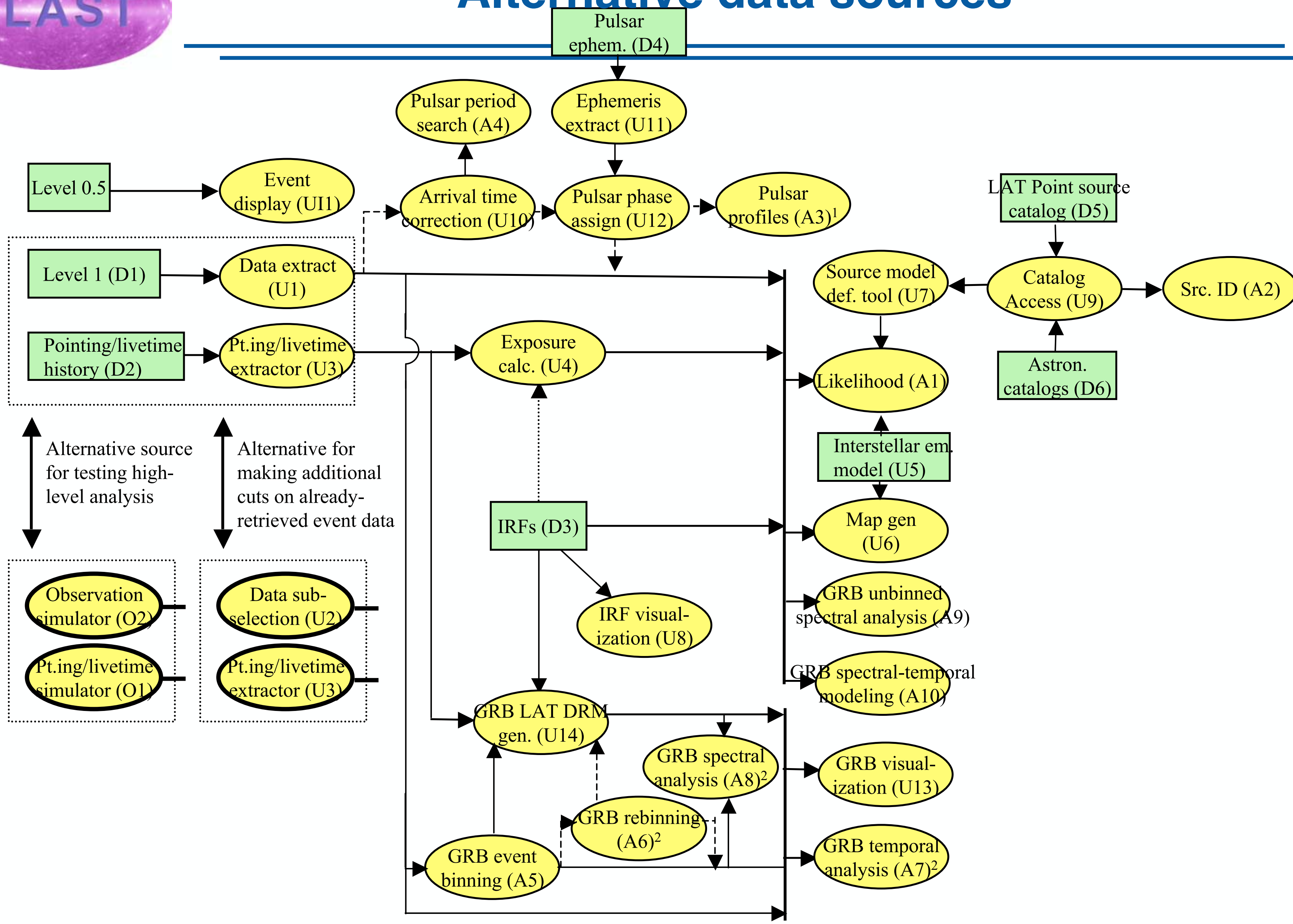


Gamma-ray bursts

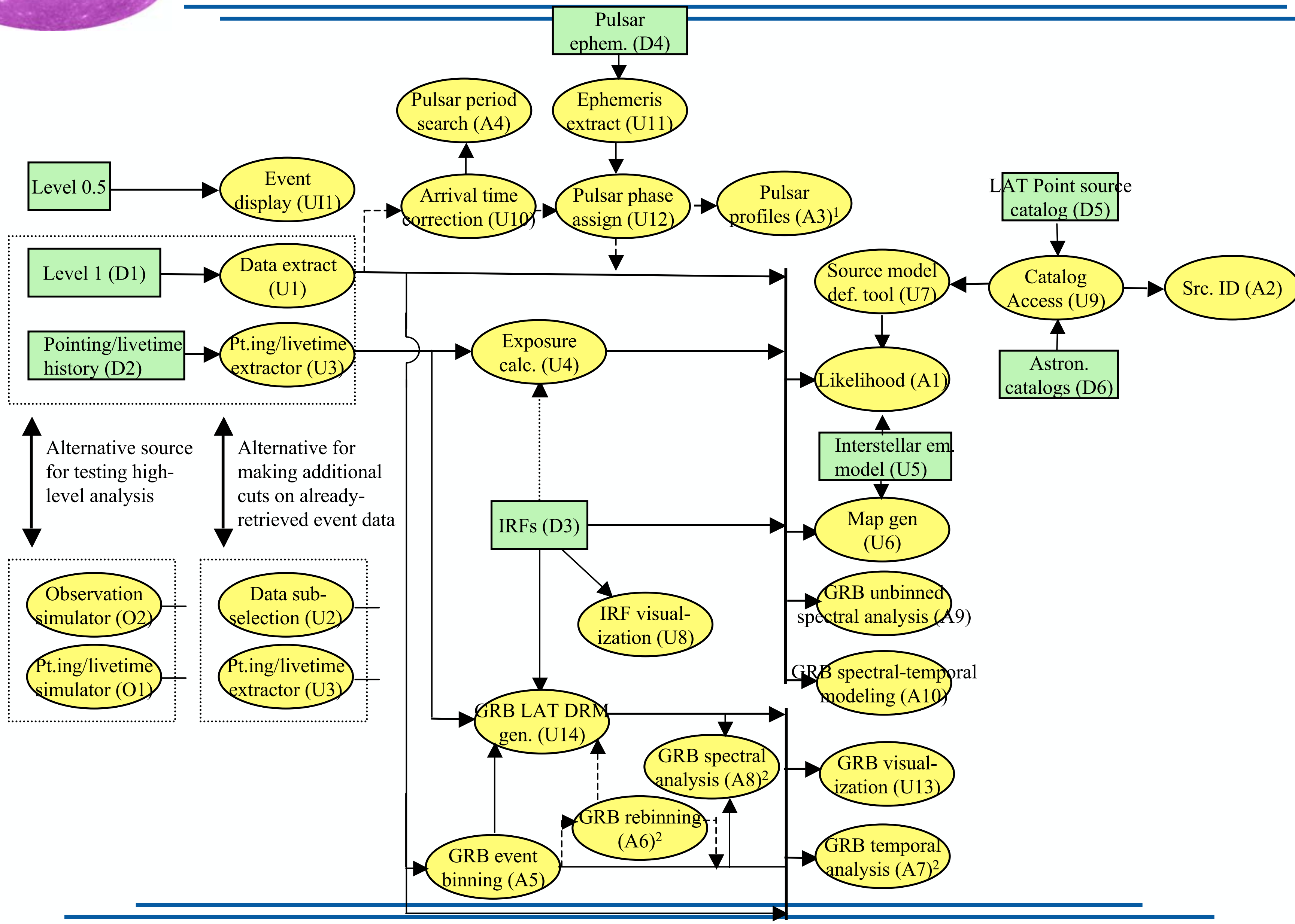


The burst analysis tools are designed to analyze data from the GBM and other missions in addition to (together with) the LAT.

Alternative data sources

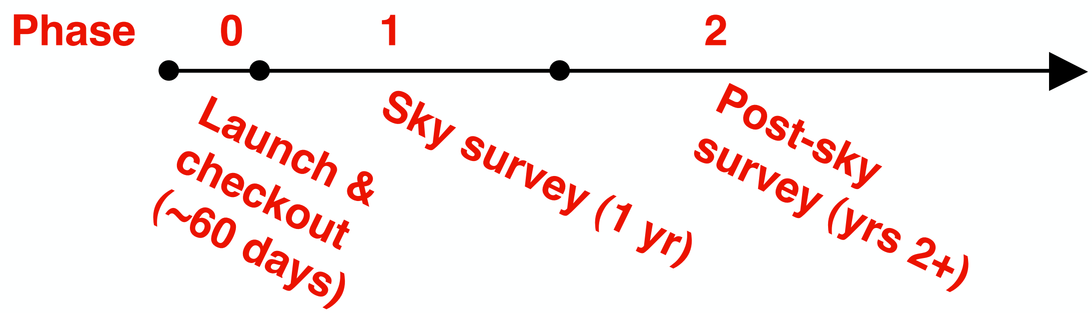


All components together



DATA RELEASE SCHEDULE

- Phases of mission



- Level 1 data (D1)
  - Phase 1 – released by LAT team 30 days after Phase 1 ends; IDSs and ~dozen investigators work with LAT team; data for detected transients released immediately (bursts=transients!)
  - Phase 2 – data released as soon as processed; 90 day proprietary period for ‘ideas’ (accepted guest investigator observing/analysis proposals)
- Pointing and livetime history (D2) released with the Level 1 data D1
- Instrument response functions (D3)
  - Initial versions released with the analysis software; updates to be prepared and released as needed (e.g., based on change of instrument response)
- Pulsar ephemerides (D4) – primarily radio timing information
  - Updated during mission, coordinated by Pulsar Working Group through SSC
  - Schedule TBD
- LAT source catalog (D5) – released at end of Phase 1, updated at end of 2nd and 5th years (tentative); an ‘early release’ version of the catalog may be made available during Phase 1 to aid proposal preparation

DEVELOPMENT CONCEPT--MANAGEMENT

- Difficulty: software will be developed by scientists and programmers at different institutions with varying proficiency in C++. This army will be difficult to command.
- Managers will be established for 7 groups of tools and databases
  - Databases and related utilities: D1, U1, U2, D2, U3, U6 & U9
  - Likelihood analysis: D3, U4, U7, U8 & A1
  - Pulsars: D4, U10, U11, U12, A3 & A4
  - GRBs: U13, A5, A6, A7, U14, A8, A9 & A10
  - Catalog analysis: D5, U5, U6, U9 & A2
  - Observation simulators: O1 & O2
  - User interface: UI1, UI2, UI3, UI4 & UI5
- Managers will be responsible for monitoring the development, maintaining the schedule, tool-level testing, and interfaces with the other development groups.

MANAGEMENT: ORGANIZATION

- **Managers will be responsible to the SSC-LAT Software Working Group.**
- **Current managers for the development areas**

Databases and related utilities	**R. Schaefer (SSC), K. Young (SLAC)**
Analysis tools	
Source detection	**P. Nolan (SU)**
Catalog analysis	**I. Grenier (CEA/Saclay)**
Pulsar analysis	**M. Hirayama (SSC)**
GRB analysis	**D. Band (SSC)**
Observation simulation	**T. Burnett (UW)**
User interface	**J. Chiang (SSC), H. Kelly (GSFC/LAT)**

- **Potentially subject to change**
- **Specific development assignments have not yet been made**
- **Progress in these areas will be reported on webpages linked to http://www-glast.slac.stanford.edu/ScienceTools/**

MANAGEMENT, CONT.

- Tools will be delivered to the Working Group as they are completed.
- We expect the tools to be under configuration control and in a testing regime before being submitted to the Working Group, but formal configuration control and testing will definitely begin at this point.
- The Working Group will oversee functional testing, code reviews and documentation reviews before acceptance.

DEVELOPMENT CONCEPT--SOFTWARE DEVELOPMENT

- The tools will be developed primarily with object-oriented programming using C++.
- Code in other languages will be accepted if robust compilers exist for the supported operating systems and code is supported elsewhere: wrapped into C++ classes
- We will support Windows and Linux.
- We will use the development tools of the LAT instrument simulation:
  - CVS--a code repository with version tracking
  - CMT--build configuration
  - doxygen--documentation
- Class diagrams will be developed for the tools. In addition to structuring the code, these diagrams will identify common classes that can be collected into a GLAST class library.

DEVELOPMENT CONCEPT--TESTING

- The testing plan is based on that implemented for the development of the LAT instrument simulation
- Throughout development, testing of software will be ongoing at several levels
  - 'Package' tests – defined for each package, required to be present, run nightly on then-current development versions
  - Application tests –test procedures for the analysis tools to verify all requirements, with performance benchmarks also logged
  - Code reviews – verify compliance with coding and documentation rules
- The development schedule includes 2 Mock Data Challenges
  - Equivalent to end-to-end tests for the high-level analysis
  - Tests high-level interaction between tools

DEVELOPMENT CONCEPT — SCHEDULE

- Development priority is based on dependencies. For example, simulators are needed to create test data.
- The milestones, after the LAT CDR in April, 2003, are the Mock Data Challenges
  - MDC1: June-October 2004
  - MDC2: June-December 2005

MDCs include development of the challenge datasets, analysis, reconciliation of the results with the input, & repair of the tools

- FTE estimates are conservative educated guesses which include code development, testing, and documentation
- Time for prototyping is scheduled for tools with important open issues, e.g., the likelihood analysis tool (A1)
- Completion of the standard analysis environment is scheduled for the end of MDC2, 9 months before launch.
  - Allow for recovery from unanticipated difficulties with development
  - Also recognizes other software development, LAT team-specific and SSC-specific that needs to take place

SCHEDULE, CONT.

	2002	2003				2004				2005				Total
Development area	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	
Databases and related utilities	6.3	6.3	6.3	3.3	3.3	2	2	0	1.2	1.2	1.2	1.2	1.2	35.5
Analysis tools groups						0	0							
Likelihood analysis	6.7	6.7	7.7	6.7	5.4	4	4	0	0	0	0	0	0	39.2
Pulsar analysis	0	0	0	0	0	0	0	3	3	7	5	0	0	18
GRBs	1.9	1.9	2.8	2.8	1.3	0.9	1.6	0.8	0.8	1	1	0.25	0.25	17.3
Catalog analysis	0	0.5	0.5	0.5	0.5	3.3	3.3	3.8	1.8	1.8	0.5	0	0	16.5
Observation simulation	3	3	3	3	2	0	0	0	0	0	0	0	0	14
User interface	1.7	3	3.8	3.1	2.5	2.5	2.5	1	1	1	1	1	1	25.1
Total required	**19.6**	**21.4**	**24.1**	**19.4**	**15**	**12.7**	**13.4**	**8.6**	**7.8**	**12**	**8.7**	**2.45**	**2.45**	**167.6**
Available FTEs[1]	**11**	**14**	**14**	**15**	**16**	**17**	**17**	**17**	**17**	**17**	**17**	**17**	**17**	**206**

[1] Approximate totals, from survey by institution among participants at the June 2002 Science Tools Workshop at SLAC.

Totals are in person -quarters; divide by 4 to get FTE -years

- **Adequate labor totals are anticipated, although available FTEs doesn't appear to match needs early on.**
- **Shortfall is less than indicated, because development effort for individual tools (assumed to be constant in the above table) can start at a lower-than-average level**

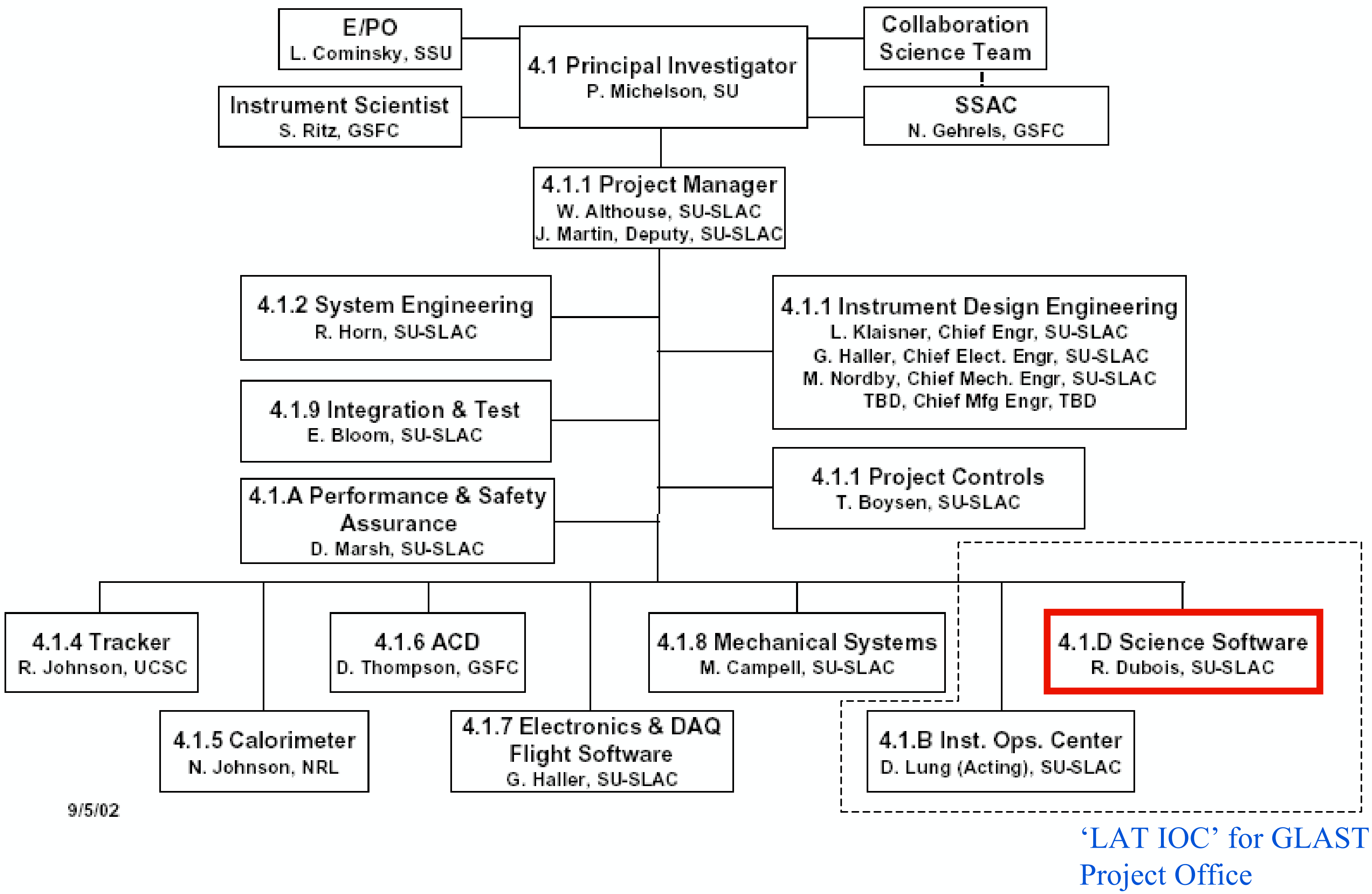
Sources of Science Tools Development Labor

- **Within the LAT team, much of it will be contributed effort from international institutions**
- **Our current understanding of projected FTEs available is based on discussions at the Science Tools workshop held in June at SLAC**

Institution	Standard Analysis Environment Labor
SU	1 FTE now, rising to 2 -3 FTEs by 2004
SLAC	1⁄2 FTE now, rising to 2-3 FTEs by 2004
SSC	3-4 FTEs
CEA	1⁄2 FTE near term, 3+ from 2003
INFN	2-3 FTE
IN2P3	LLR maybe in future, Bordeaux net ~1 FTE
Bochum	1-2 FTE mostly on interstellar emission model
Japan GLAST Collaboration	contribution to standard analysis env. development not clear, current focus is on background simulations, fast MC
Total	~8 FTE near term, ~17 in 2004 time frame

Backup slides follow

GLAST LAT Project Organization



DATABASES

IDNameDescriptionSize after1 yearD1Event Summary(level 1)Summary information (i.e.			

UTILITIES

IDNameDescriptionU1		Basic front end to the event summary database l

UTILITIES, CONT.

U8IRF	visualizationExtracting PSF, effective areas, and energy distributionsfrom the CALD	

ANALYSIS TOOLS

IDNa	meDescriptionA1Like	lihood analysisPoint source detection, characterization (positio

OBSERVATION SIMULATORS

IDNameDescriptionO1Livetime/pointingsimulatorGenerates simulated pointing, livetime, a		

USER INTERFACE

ID	**Name**	**Description**
UI1	Event display	Displays the tracks of an individual event, requires ac

Sequence of an Analysis: Gamma-Ray Point Source

- **Define region of sky, time range, etc. of interest**
  - Typical minimum size (based on PSF sizes) radius ~ 15°
- **Extract gamma-ray data (U1 accessing D1)**
  - Applying selection cuts, including zenith angle
  - Typical data volume (per year): 1 _ 10^6, 10^8 bytes
- **Generate exposure (U3 accessing D2)**
  - May better be called livetime accumulation
  - Matches cuts applied to gamma-ray data
  - Tabulation ~ 700 _ 15 _ 15 _ 15 ~ 2.5 _ 10^6 values ~ 10^7 bytes
- **Define the model to be fit to the data (U7)**
  - Facilitated by candidate source catalog, intensity map and data visualization within U7
  - Models may be considered as a table of parameters, or an XML file, human-readable, including parameters for interstellar emission model
  - Typical region will contain ~dozens of point sources that need to be modelled
- **Fit the model to the data, generating, e.g. 'TS maps' and confidence regions (A1) or spectral fits**
  - Model may need refinement, iteration within A1