LIGO I Data Usage

Lee Samuel Finn

LIGO Scientific Collaboration

The Pennsylvania State University

Overview: Data User Perspective

- Analysis proposals
  - » Proposal contents
  - » Review criteria & process
- Data and analysis resources
  - » Standard data products
  - » Computing resource use model
- Publications & publications policy
  - » Initiating an analysis paper
  - » Manuscript review

Data Access Proposal Example: Associating Grav., γ-ray Bursts

- Activity Character …
  - » "Proof of principle" or prototype development
  - » LDAS/LAL S/W development and/or (sub)system test
  - » Detector characterization
  - » *Data analysis activity*
- Scientific, Technical Rationale …
  - » Including how activity is related to LIGO (Lab & LSC) goals
  - » *"γ-ray bursts are associated with violent formation of BH … "*
- Technical approach …
  - » *"Analyze on, off-source data sets using lock-in techniques…"*
- Deliverables …
  - » All real products, such as
    - – Software
    - – Reduced data sets or databases
    - – Detector, data characterization
    - – *Expected publications*, technical reports, documentation, etc.
- Required resources …
  - » Human
    - – *Identified LSC, Lab FTEs*
    - – Required travel, site support
  - » Facility
    - – *Computing,* network, storage
    - – Software, *data*

Data Access Proposal Example: Associating Grav., γ-ray Bursts

- Work Plan
  - » Schedule with milestones
    - – Detail sufficient to permit monitoring progress towards goals if proposal accepted
  - » Management plan
    - – Address specifically coordination multiple LSC institutions or team members at different institutions

Proposal Review: Criteria & Process

- Criteria:
  - » Scientific, Technical Merit
    - – Impact, importance for LIGO: cf. projects of similar scope, effort
    - – Contributions toward LIGO schedules & milestones
    - – Value of deliverables to LIGO effort
    - – Long-term potential for LIGO science
  - » Feasibility
    - – Team skills
    - – Schedule definition, reality
    - – Resource availability
    - – Resource requirements relative to efforts of similar priority
- Process:
  - » Evolving model
  - » Recommendations
    - – Software Coordination Committee
    - – Development group chairs
  - » Decision
    - – Lab Director & LSC Spokesperson
- If approved …
  - » Proposal circulated throughout LSC
  - » All interested LSC LIGO I team members may “join” proposal

Data Resources: Data Set Hierarchy

- Level 0: Full IFO Data Stream
  - » Content: Full Data Stream
  - » Rate: 15 MB/s
  - » Lifetime: ≥16h spinning media; ~1 month tape
    - – N.B.: Brief epochs archived for diagnostic purposes
  - » Availability:
    - – Available only on-site and in FRAME format
  - » Anticipated Use:
    - – detector diagnostic studies
    - – On-site science analyses
- Level 1: Archived Data Stream
  - » Content
    - – Important IFO and PEM channels
    - – GDS/LDAS regression, whitening, calibration & instrument state data
  - » Rate: 3 MB/s or 200 TB in 2 yr science run
  - » Lifetime: Five years on-line
  - » Availability:
    - – FRAME or LLW format from LIGO Data Archive ~ 1 wk after acquisition
  - » Anticipated Use:
    - – Extended detector diagnostics
    - – Deep look-back at interesting epochs

Data Set Hierarchy, cont'd

- Level 2: Strain and Data Quality Channels
  - » Content
    - – Strain Channel
    - – Regression & whitening coefficients
    - – "Data Quality" channels: Summary judgments formed from PEM & IFO channels
  - » Rate: 300 KB/s or 20 TB in 2 yr science run
  - » Availability: derived from Level 1 data
    - – FRAME or LLW format from archive
  - » Anticipated Use
    - – Off-site science analyses
    - – Data sharing/exchange with other detector projects
- Level 3: Whitened Strain
  - » Content: (Whitened) best estimate of the GW strain
    - – 512 Hz bandwidth
    - – All known instrumental artifacts removed
  - » Rate: 6 KB/s or 400 GB in 2 yr science run
  - » Lifetime: in perpetuity
  - » Availability:
    - – Derived from level 1, 2 data
    - – Several weeks following acquisition
    - – LIGO Data Archive, tapes
  - » Anticipated Use:
    - – Off-site science analyses
    - – Data sharing/exchange with other detector projects

Data Resources: Metadata

- Information about detector, data, or environment …
  - » Acquired or accumulated from non-LIGO sources
  - » Determined through analysis
- Detector characterization
  - » IFO channel summary statistics, long-term trends, regression coefficients, final calibrations, detector sensitivity estimators, etc.
- Non-LIGO PEM data
  - » Remote seismic sensing, EM storm information, …
  - » Cosmic ray, neutrino detector data, UVOIR astrophysics, …
  - » Other g-wave detectors …
- Data Characterization
  - » Strain channel “anomalies” (which may be GW signals) and everything that has been learned about them through analysis and correlation with other channels

Computing Resources

- On-site
  - » DMT, LDAS Beowulf, Data Conditioning Unit
    - – Production, time-critical or real-time analyses
    - – Analyses requiring access to level 0 data
- Off-site
  - » LDAS Beowulf
    - – Production analyses requiring close proximity data archive or relational database (metadata)
    - – Multi-detector analyses
    - – Data mining
- LSC Member Institution Based
  - » High-end & low-end
    - – E.g., high end: UT Brownsville, UWM Beowulfs, CACR
    - – E.g., low end: LSC Minimum Workstation (cf. Appendix)
  - » Low to medium volume “custom” data subsets
    - – Subsets based on metadata selection criteria
- Non-Lab/LSC
  - » E.g., SDSC, NCSA
  - » Medium to high volume “custom” data subsets
    - – Subsets based on metadata selection criteria
    - – Exploit high network bandwidth to Centers

Publication Preparation & Review

- Initiation
  - » General outline, proposed author list
    - – Presentation at Collaboration meeting may be required
  - » Author list
    - – Previously unpublished data: full LIGO I Collaboration, alpha. order
    - – Conference proceedings: speaker + LIGO I Collaboration in alpha order
    - – Technical results: responsible scientists & engineers only
- Preparation
  - » Initiating author, co-authors
- Review
  - » Evolving model
  - » Draft(s) to LSC Review Committee
    - – Three members, including member of Lab Directorate
    - – Current Members: Gonzales (PSU), Reitze (UFL), Sanders (Lab Directorate)
  - » Committee
    - – Confirms results, comments on content, authorship
    - – Iterates with initiating author(s)
  - » Spokesperson
    - – Adjudicates disputes
    - – Approves final manuscript prior to submission
  - » Final version circulated to entire LSC prior to submission

Overview: Data User Perspective

- Analysis proposals
  - » Proposal contents
  - » Review criteria & process
- Data and analysis resources
  - » Standard data products
  - » Computing resource use model
- Publications & publications policy
  - » Initiating an analysis paper
  - » Manuscript review

Appendix: LSC Minimum Workstation

- Purpose
  - » Support local science (exploratory detector & data characterization, analysis, visualization) at LSC Member Home Institutions
- Hardware Configuration
  - » 0.5 Gflop/s processor speed
  - » 50 Gb disk
  - » TBD WAN access
  - » AIT-2 Tape drive
- Software Configuration
  - » Linux
  - » LDAS & LAL Tools
  - » DB2 client (database access)