
# The LIGO I Science Run

## NSF Review of the LIGO Operations Proposal for FY2002 - 2006

*LIGO Hanford Observatory*

*Hanford, Washington*

*26 February 2001*

***Albert Lazzarini***

***LIGO Laboratory Caltech***


# LIGO/LSC Data Analysis Model

- Now:
  - » Initial engineering runs starting to set the stage for how science, research is done
  - » Data being archived at Caltech
  - » Access from archive according to LIGO Laboratory MOUs
  - » “Stress testing” of software and hardware systems - LIGO Data Analysis System, Global Diagnostics System, Data Acquisition System, Control and Data System
  - » Initial data analyses focus on
    - – sorting out commissioning issues, improving interferometer performance
    - – understanding the terrestrial and man-made environments,
    - – calibrations, data conditioning, pre-processing
    - – data quality assessment, data compressibility


# LIGO/LSC Data Analysis Model

- Now:
  - » Focus is on integration of LIGO scientific analysis and search algorithm software into the LIGO Data Analysis System (LDAS) environment being developed by LIGO Laboratory
  - » Integration overseen by LSC Software Coordination Committee with LIGO Laboratory
    - – Sequence of incremental mock data challenges (MDCs) aimed at validating key aspects of integrated pipeline analysis process
    - – 2000/08 : Data pre-processing common to a number of searches
    - – 2001/01: Parallelized search over templates using MPI, PC Linux cluster under control of LDAS
    - – 2001/03: Validate insertion/retrieval/mining of relational dabases to be used for storing event/trigger information
    - – 2001/05 and beyond: testing of various other algorithms for key upper limits searches
    - – 2001/10: Test of retrieval of data from main LIGO archive at Caltech for pipeline processing off-site and data distribution

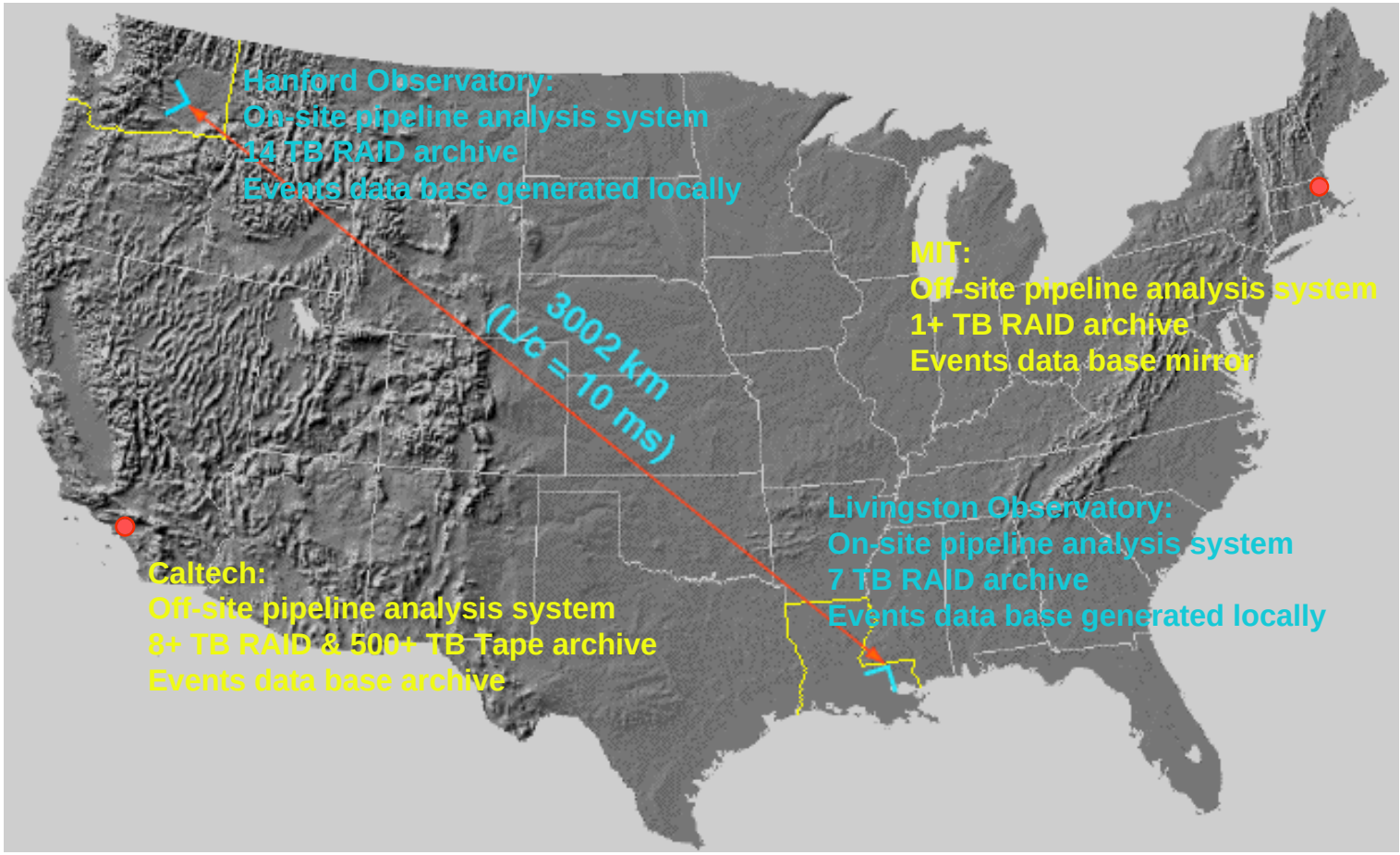

# LIGO/LSC Data Analysis Model

- Near-term (3Q2001):
  - » LIGO science will focus on using engineering runs to extract meaningful first upper limits
  - » Organized around 4 upper limits papers using ~1 week of data in 3Q2001
  - » Opportunity to set current best upper limits on these classes of sources
  - » Provides a basis to "exercise" the LSC data analysis groups
    - – Brings together theorists and experimentalists
  - » Provides a basis for future organization of the LIGO I Science Run search teams
    - – groups will expand as interest grows in LIGO science.


# LIGO/LSC Data Analysis Model

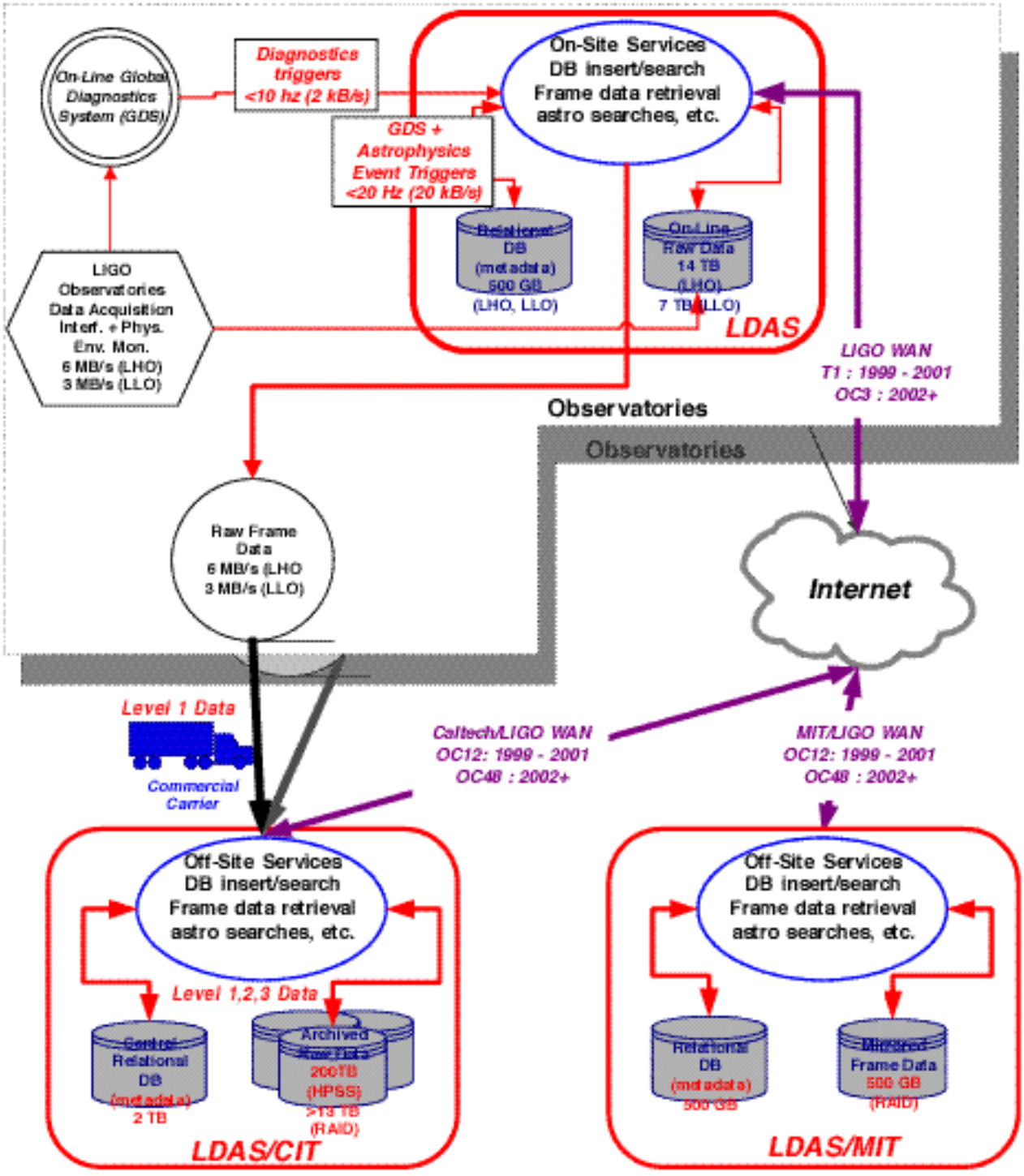
- LIGO I Science Run (2Q2002+ ):
  - » Key astrophysical searches follow the LSC Data Analysis White Paper plan:
  - » Organized around teams, as in the near-term upper limit studies
    - – Open to all who are willing to work
  - » LIGO Laboratory Data Analysis System resources will be used for searches, will be shared among the teams
  - » LSC member institutional resources used by individual researchers
  - » *Longer term: establish up to 5 LIGO/LSC Tier 2 centers ("University Research Centers" or URCs) to provide additional computational, data distribution resources across collaboration*
    - – *This activity is being performed within the GriPhyN Collaboration*
      - *Several LSC institutions are members, as well as LIGO Laboratory*

# LIGO Laboratory Sites




LIGO-G010040-00-E

# LIGO Science Run Data Flow Model




LIGO-G010040-00-E


# What are the LIGO Data?

- Continuous time series: $2^N$ samples/second, 16 bit data
- Data analysis: digital signal processing
  - » 90% of CPU time spent on f ´ t transformations
  - » Analysis performed in both domains
    - – Single channel, over a long time; many channels, over a short time
    - – How to cache, catalog, replicate, this *virtual data*
- Results of analysis: events, spectra, nD representations (“images”)
  - » Environmental, instrumental “events”: vetoes
  - » Astrophysical events
  - » Time stamp, Process ID generating event, Parameters associated with event, ...
  - » Stored in a relational database for later retrieval, reanalysis:
    - – tables, “blobs”, links to data

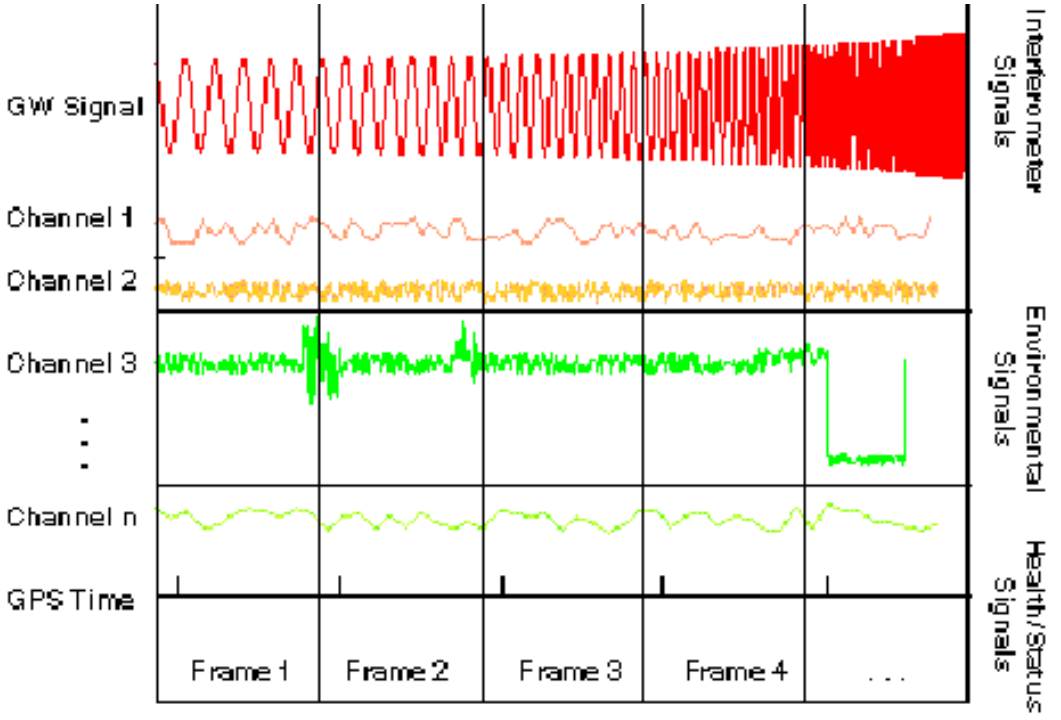
# Raw Data Stream Characteristics


***Science Channel (1%)***

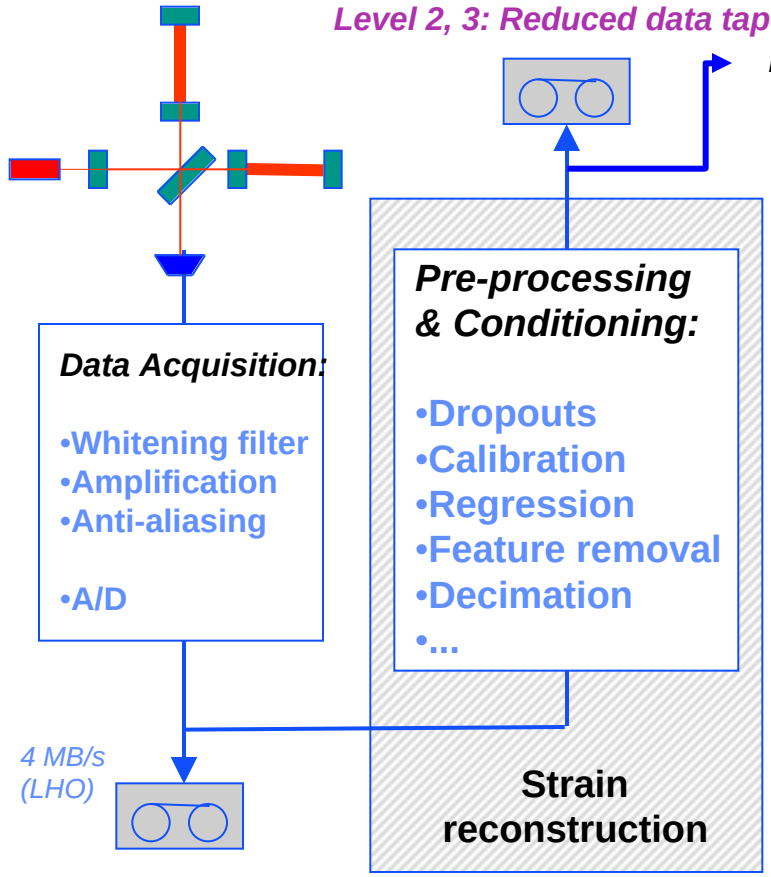
**Interferometer Channels (86%)**

**Environmental Channels (13 %)**

- **All interferometric detector projects have agreed on a standard data format**
- **Anticipates joint data analysis**
- **LIGO frames for 3 interferometer are ~ 9MB/s**
  - **96 kB/s strain (2 Bytes x 3 IFOs x 16 kSample/s)**
  - **~ 7.7 MB/s other interferometer signals**
  - **~ 1.2 MB/s environmental sensors**
  - ***Strain is ~1% of all data !***



# Data pre-processing at observatories




- Master data tapes transported to Caltech for deep archive (HPSS)
- Reduced data tapes provide reduced bandwidth sample of data stream; needed for search algorithms
- Whitening required due to dynamic range of signals
- Regression & feature removal reduces RMS, dynamic range from narrowband line features
- Resampling & decimation matches data rate to search bandwidth
- Calibration provides physical strain

# LIGO Data Products - time series data

| ***Mode*** | Raw and Derived Data for On-line Diagnostics | **Level 1** Full (100%) frame data for archiving | **Level 2** Strain and data summary, QA channels | **Level 3** Strain best estimate |
|---|---|---|---|---|
| Uncompressed Rate (MB/s) | LHO: 9.479<br>LLO: 4.676<br>Total: 14.155 | LHO: 4.698<br>LLO: 2.278<br>Total: 6.975 | Total: 0.300 | Total: 0.006 |
| w / 50% Hardware Compression MB/s onto tape media | - | LHO: 2.349<br>LLO: 1.139<br>Total:3.488 | Total: 0.150 | - |
| Data growth rate, per year of integrated running, ***TB/yr***. | - | LHO: 74<br>LLO: 36<br>Total:110 | Total:9.5 | ***Total: 0.200*** |
| Total including redundant 100% backup, ***TB/yr***. | - | LHO: 148<br>LLO: 72<br>***Total:220*** | ***Total:19*** | - |
| **Purpose** | For on-line monitoring of interferometers | Deep permanent archive | Science analysis, data exchange | Science analysis, data exchange |
| **On-site look-back time** | Must use real-time control and monitoring system (CDS) disk caches | LHO Disk cache: 60 hr<br>LHO Tape robot: 49 d<br>LLO Disk cache: 60 hr<br>LLO Tape robot: 100 d | - | - |
| **Off-site look-back time** | - | As long as required | In perpetuity | In perpetuity |


# Times series data uses

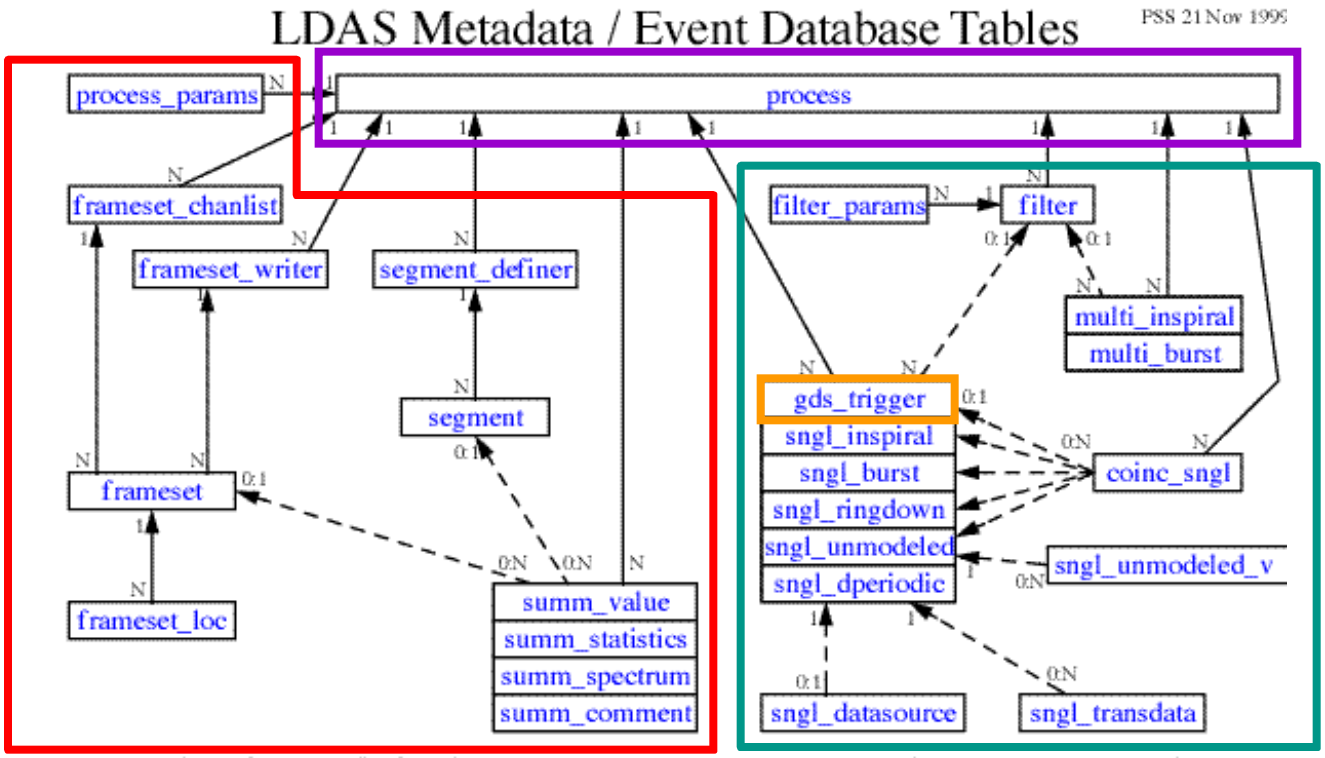
- Collaboration-wide searches (Lab resources)
  - » On-site at observatories: LIGO Data Analysis System (LDAS)
    - – 7x24 pipeline analysis to provide first pass through data
    - – Events (both instrumental vetoes from on-line monitors and astrophysical events from pipeline) registered in database
    - – Single-interferometer detections
    - – Near-real time information (e.g., SNe bursts, … )
  - » Off-site at Caltech: LIGO Data Analysis System (LDAS)
    - – Data ingestion into deep archive; mirroring of site event databases
    - – Pipeline analysis to provide second pass through data
      - Follow-up to on-site first passes
      - Multiple interferometers
    - – Events (both instrumental vetoes from on-line monitors and astrophysical events from pipeline) registered in database


# Times series data use

- Individual exploratory research (institutional resources)
  - » Reduced data sets available from Caltech archive
    - – Binary frame format
    - – LIGO-Lightweight data format (XML)
  - » Download to locally owned,managed resources for exploratory research
    - – Internet (small data sets), ftp, pftp
    - – Tapes (larger data sets)
  - » Analysis environments:
    - – Commercial tools
      - Matlab, Mathematica, IDL, …
    - – Replica installation of LDAS tools, APIs
      - Other LSC institutions
      - *Off-lline LDAS Development & Test systems at Caltech*
    - – Prototype tools, public domain code -- ROOT, GRASP, ...

# Data products: Event database

- Event source (processes or filters)
- Raw data characteristics/location
- Raw data statistics
- Instrumental triggers (vetoes)
- Astrophysical search triggers
  - Single interferometer
  - Multiple interfertometers




Arrows indicate "foreign key" referential integrity constraints. Values near the ends of the arrows (1, N, etc.) indicate the possible multiplicities. Dashed lines indicate optional relationships. Stacked tables (grouped by thick lines) have common relationships with other tables, except for relationship arrows connecting along the right edge. Examples: 1) Each segment is related to one segment_definer, 2) Each segment_definer is (generally) related to many segments; 3) A frameset is related to one frameset_chanlist entry and to one frameset_writer; 3) A summ_value (or summ_statistics, etc.) entry may or may not be related to a segment and/or a frameset; 4) A single-interferometer event (gds_trigger, sngl_inspiral, etc.) entry may be related to up to one sngl_datasource and/or any number of sngl_transdata entries.


# Event data use

- Collaboration research, individual exploratory research
- Event data available from observatories (recent) or Caltech DB archive (long term)
  - LIGO-Lightweight data format (XML)
- Use LDAS resources to query DB.
  - » Download to locally owned,managed resources for exploratory research
    - Internet (small data sets), ftp, pftp
  - » Analysis environments:
    - Commercial tools
      - Matlab, Mathematica, IDL, …
    - Replica installation of LDAS tools, APIs
      - Other LSC institutions
      - *Off-lline LDAS Development & Test systems at Caltech*
    - Prototype tools, public domain code -- ROOT, GRASP, ...


# Summary

- LIGO Laboratory on threshold of transition to scientific observatory:
  - » Participation in astrophysical searches by the Laboratory scientists through the LSC
  - » Data analysis will become a key "business" of the Laboratory across all sites
    - – Database, archive use & maintenance
    - – Pipeline analysis SW and HW maintenance
    - – Continued R&D into algorithms, new technologies
  - » Stimulation & modeling of interferometers to support detector understanding, data analysis