CALIFORNIA INSTITUTE OF TECHNOLOGY
LIGO Laboratory, M/C 18-34, Pasadena, California 91125
626-395-2129, Fax 626-304-9834

To/Mail Code:	M.Levine, B. Mihalas
From/Mail Code:	A. Lazzarini/18-34 S. Anderson
Phone	626-395-8444/2904
FAX:	626-304-9834
Refer to:	T020162-00-E
Date:	24 October 2002

Mike and Barbara,

The description provides a narrative of the data flow depicted in the attached block diagram.

Description of the revised LIGO data flow model motivated by the choice of SAM-QFS for the LIGO Laboratory mass storage solution:

1) The licensing cost and administrative complexity of SAM-QFS are sufficiently low that it will be run at the Observatories. It was never considered a viable option to run HPSS at the Observatories. Note, QFS has already been used as the interface between CDS and LDAS successfully for both E7 and S1.

2) SAM-QFS running at the Observatories will provide 1yr of on-line data access at each Observatory to its own data via a 700 slot tape robot. This is an increase from the current 3 week disk cache.

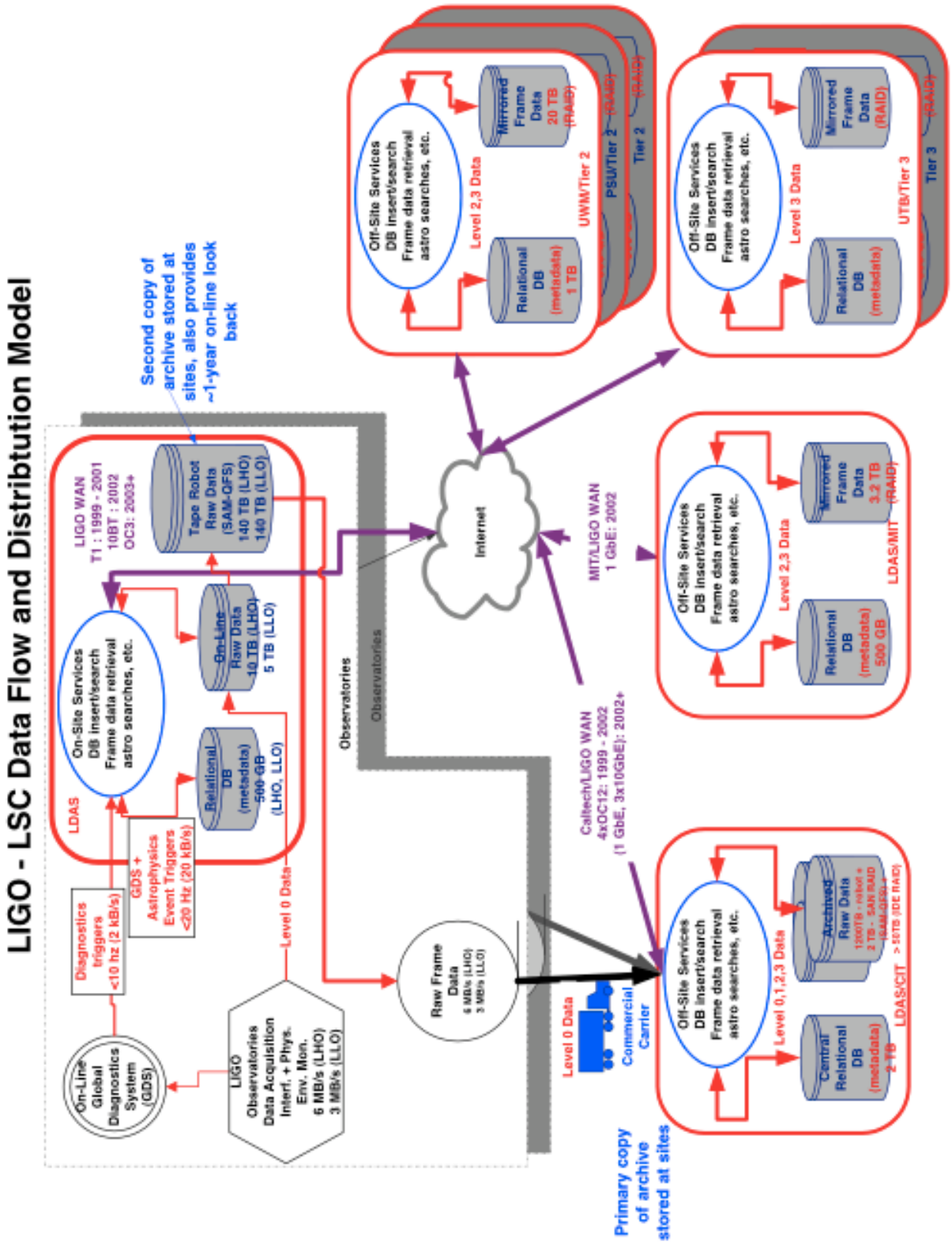
3) As before, two tape copies will be written at the Observatories with one tape copy being physically transported to Caltech. However, unlike the original HPSS model, this transported tape will be loaded directly into the 6000-slot tape robot at Caltech without any need for tape-to-tape data replication, i.e., SAM-QFS allows for direct import of tapes from another archive instance.

4) SAM-QFS allows for the archiving of Metadata to a simple and convenient file. LIGO will dump the metadata for each physical tape at each Observatory to a unique file and permanently archive one copy on local disk as well as transmitting a duplicate copy to Caltech over the Internet with the use of tapes if necessary. At Caltech this second copy of metadata will be used to merge in the data from the corresponding data tape into the archive mass-storage system and also be kept on independent disk storage for redundancy. This leaves the Observatories and Caltech with identical copies of the raw

data and sufficient metadata to describe both their local tape storage as well as each others.

5) When failed tapes are encountered both the Observatory and Caltech instances of SAM-QFS will know the tape labels at both locations so that replication of data from the redundant tape located at the other site will be straight forward.

6) Equally important to archiving the data is the need to efficiently retrieve the data for analysis with the large archive data analysis system at Caltech. The fundamental filesystem interface (QFS) to SAM-QFS solves this problem without writing any additional software as was planned to efficiently stage of data out of HPSS.


LIGO - LSC Data Flow and Distribtution Model
Second copy of archive stored at sites, also provides ~1-year on-line look back
LDAS
LIGO WAN T1 : 1999 - 2001 10BT : 2002 OC3: 2003+
Tape Robot Raw Data (SAM-QFS) 140 TB (LHO) 140 TB (LLO)
On-Site Services DB insert/search Frame data retrieval astro searches, etc.
On-Line Raw Data 10 TB (LHO) 5 TB (LLO)
Relational DB (metadata) 500 GB (LHO, LLO)
Diagnostics triggers <10 hz (2 kB/s)
GDS + Astrophysics Event Triggers <20 Hz (20 kB/s)
Level 0 Data
On-Line Global Diagnostics System (GDS)
LIGO Observatories Data Acquisition Interf. + Phys. Env. Mon. 6 MB/s (LHO) 3 MB/s (LLO)
Raw Frame Data 6 MB/s (LHO) 3 MB/s (LLO)
Observatories
Observatories
Internet
Caltech/LIGO WAN 4xOC12: 1999 - 2002 (1 GbE, 3x10GbE): 2002+
MIT/LIGO WAN 1 GbE: 2002
Off-Site Services DB insert/search Frame data retrieval astro searches, etc.
Level 2,3 Data
Mirrored Frame Data 20 TB (RAID)
Relational DB (metadata) 1 TB
UWM/Tier 2
PSU/Tier 2
Tier 2
Off-Site Services DB insert/search Frame data retrieval astro searches, etc.
Level 3 Data
Mirrored Frame Data (RAID)
Relational DB (metadata)
UTB/Tier 3
Tier 3
Off-Site Services DB insert/search Frame data retrieval astro searches, etc.
Level 2,3 Data
Mirrored Frame Data 3.2 TB (RAID)
Relational DB (metadata) 500 GB
LDAS/MIT
Level 0 Data
Commercial Carrier
Off-Site Services DB insert/search Frame data retrieval astro searches, etc.
Level 0,1,2,3 Data
Archived Raw Data 130TB - robot + 2 TB - SAN RAID (SAM-QFS)
Central Relational DB (metadata) 2 TB
LDAS/CIT > 50TB (IDE RAID)
Primary copy of archive stored at sites