LIGO Laboratory / LIGO Scientific Collaboration

LIGO-M080072-v2 LIGO Laboratory 26 January 2009

A proposal for providing open access of LIGO data to the broader research community

LIGO and LSC Directorate

.

California Institute of Technology
LIGO Project – MS 18-34
1200 E. California Blvd.
Pasadena, CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project – NW22-295
185 Albany St
Cambridge, MA 02139
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

LIGO Hanford Observatory
P.O. Box 1970
Mail Stop S9-02
Richland WA 99352
Phone 509-372-8106
Fax 509-372-8137

LIGO Livingston Observatory
P.O. Box 940
Livingston, LA 70754
Phone 225-686-3100
Fax 225-686-7189

http://www.ligo.caltech.edu/

http://www.ligo.org

Executive Summary

LIGO has prepared this white paper in response to a March 2008 request by the National Science Board to the National Science Foundation on the issue of making future LIGO data open to the broader research community. Our proposal calls for a phased approach to the release of LIGO data moving from the pre-detection era of today to the Advanced LIGO epoch of routine detections.

Of paramount importance in the development of this proposal is furthering the NSF's and the scientific community's primary goal, namely the earliest possible detection of gravitational waves and the emergence of gravitational wave astrophysics as a scientific discipline. The current open structure of the LIGO Scientific Collaboration (LSC) has proven effective at attracting talented scientists and of organizing multidisciplinary teams to ensure the integrity of the results in the exploratory phase; it is important that the long-term data policy not disrupt these structures. Our approach allows these structures to evolve while preparing for broad data release. We have also received advice from astrophysicists in related disciplines who are familiar with LIGO today, who are not members of the LSC, and who have an interest in our field.

We envision two broad phases of the scientific research and their associated data policies.

Detection and discovery phase: In this epoch the collaboration is still bringing the instruments into their most sensitive operation and refining the techniques used to characterize and understand the data and to analyze them for a variety of posited sources. The number of detectable sources is low and the extraction of signals requires understanding of instrumental artifacts. During this initial epoch, which includes the next LIGO science run in 2009-2011 and the early operation of Advanced LIGO, LIGO will make available to the broader research community any science data associated with detections that the Collaboration publishes. In addition, LIGO will remain committed to the continued long-term archiving of LIGO data. During this epoch, the infrastructure for public distribution of broader LIGO data sets will begin development.

Observational phase: As the field transitions past the first few detections, the understanding of the data will improve and the field will move from the novelty of the initial discoveries toward the deeper exploration of the astrophysical content of the gravitational wave signals. There will be the need to establish population statistics and angular distributions and the full power of multi-messenger astronomy will need to be brought to bear. During this epoch the entire body of gravitational wave data, corrected for instrument idiosyncrasies and with environmental perturbations removed, will be released to the broader research community.

We have estimated the additional costs associated with the new scope we describe above. The post-publication release of LIGO data for any detected events in the Discovery Phase is feasible today and has a modest cost of ~ \$0.2M/year for hardware and software for the public interface. The full release of LIGO gravitational wave data during the Observational Phase, which have been processed to remove instrumental or environmental artifacts, is estimated at ~\$4M/year [in FY2008 \$]. This additional cost represents roughly 10% of LIGO and LSC operations costs, and is concordant with the experiences of other major astronomical projects that publicly release their datasets. Given the nature of this work, it is likely that responsibility for the effort to prepare LIGO data for release will need to be taken on by the LIGO Laboratory as a national NSF facility.

1.1 Introduction

On 26 March 2008, the National Science Board passed a resolution that authorized the construction of Advanced LIGO and the continuation of Operations and Maintenance of the LIGO observatories by Caltech and MIT. The resolution further stipulated that, "NSF management shall report back to the National Science Board within 12 months of this award what efforts have been undertaken and what provisions have been implemented to make the data obtained under this award available and useable to the broader research community."

LIGO (consisting of the LIGO Laboratory and the broader LIGO Scientific Collaboration (LSC)) recognizes the importance of developing policies and methods for enabling LIGO data to be accessible to the broader research community. This white paper examines the issue in detail. It briefly reviews the present status of LIGO data access and the history of this data access model. It develops a phased concept for public data release that goes beyond the current model, and examines the implications on LIGO, on NSF, and on the broader scientific community in terms of maximizing the scientific utility of the data, of existing international partnerships with other gravitational wave collaborations, of ensuring the integrity of the scientific analyses and of monetary costs. It then makes a recommendation for a staged data release model that we believe will simultaneously satisfy the needs of the broader research community and LIGO in the short and long terms.

This white paper was prepared by the LIGO Directorate (LIGO Lab Executive Director, Jay Marx; Deputy Director, Albert Lazzarini; and LSC Spokesperson, David Reitze), in consultation with the former collaboration spokesmen (Rainer Weiss and Peter Saulson), the former LIGO Director (Barry Barish) and former Deputy Director (Stanley Whitcomb). Members of the LIGO Program Advisory Committee and astronomers and astrophysicists outside of LIGO were asked to provide input. The leaderships of the Virgo and GEO600 gravitational wave projects were consulted to assess the impact of a LIGO open data policy on existing LIGO-Virgo and LIGO-GEO data-sharing agreements. To assess the costs of implementing an open data policy, input was solicited from the LSC and the LIGO Laboratory. Finally, a number of astronomical observatory and data management projects were consulted to understand the cost implications and to assess the reasonableness of the cost estimates derived from LIGO and the Collaboration. Appendix A lists the individuals and organizations that were consulted.

A draft of this white paper was circulated to the Executive Committee and the Council of the LSC for comment and input. Town-hall discussions were held at the September and December 2008 collaboration meetings. The discussions included Virgo Collaboration members because the existing Memorandum of Understanding (MOU) governing joint data sharing and data analysis between LIGO and Virgo would likely be affected by changes to the current model. This final document represents the consensus view of LIGO Laboratory and the LSC.

1.2 Background

The NSF first addressed the use of LIGO data when it conducted a review of LIGO by the *Panel on the Use of LIGO*, chaired by Boyce McDaniel in June 1996. This watershed review and the Panel's recommendations laid the blueprint that has guided the evolution of LIGO Laboratory and much of the U.S. experimental gravitational wave community for more than a decade. The Panel's report (ref. http://dcc.ligo.org/public/0000/M960078/001/M960078-00-copy.pdf) describes how a self-governing collaboration much larger than the original founding institutions of LIGO, Caltech and MIT, should be nurtured to grow around the then-nascent LIGO Laboratory.

The McDaniel Panel recognized that LIGO was breaking new ground and thus represented an opportunity to develop a paradigm for collaboration and astrophysical exploitation of LIGO data that was best suited to the new field of gravitational wave astronomy.

The Panel report underscores the importance of ensuring the scientific integrity of results derived from the first generation instruments, which were likely to be characterized by many technical idiosyncratic features in the data that would require the community involved in analyzing the data to have deep understanding of the technical issues relating to the performance of the instruments. The report underscores the vital importance of ensuring that the initial discovery would not be a false alarm. This sensitivity to premature and incorrect announcement of a first detection derives from the early history of this field, when it suffered significant and almost irreversible damage due to unsubstantiated detection claims.

To promote the best possible scientific use of LIGO data, a model for the collaboration was proposed by the McDaniel Panel that is, in fact, how the LSC has evolved in the ensuing dozen years. The LSC was to be an *open* collaboration, allowing scientists from the global community to become involved in observations with LIGO data based on their ability and willingness to contribute to LIGO. Any PI who was motivated to utilize LIGO data could do so by becoming a member of the LSC. The openness and success of the model is evidenced by the ensuing growth of the collaboration as well as its breadth. Today it numbers over 50 institutions worldwide and over 600 members, with institutional membership spanning the gamut from major research universities and national laboratories to small liberal arts and teaching colleges. The LSC has established clear international leadership in the field; it has forged cross-collaboration international memoranda of understanding (MOUs) to establish a global network of comparably sensitive interferometric detectors, of which the 3 LIGO instruments are key.

In a second phase of the scientific exploitation of LIGO data foreseen by the Panel, it recommended that in the *post-discovery era*, when detections are being regularly reported and the instruments are understood, that "the possibility of developing distributable data products should be considered. For example, there will be interest in developing independent approaches to searching for various types of GW events. Early work on developing the data processing system should be carried out in such a way as to facilitate the distribution of data products, if it is decided to do so, rather than to preclude them."

In fact, the LSC, working from the very beginning with the Italian-French Virgo collaboration, developed and adopted a standard data format to ensure ease of distribution of data products between collaborations.

1.3 Current status

Now, and until detections come with some regularity, LIGO is in the first phase, the *discovery era*.

All LIGO data are accessible by any member of the LIGO Scientific Collaboration and they are shared with external collaborations with which the LSC has MOUs covering data rights, publication, and authorship rights.

The LSC is an open collaboration: it has an established procedure for considering applications by new members and it regularly accepts new members who propose to work on LIGO and analyze its data. Further, under the LIGO Laboratory visitors program, it encourages prospective collaboration members wishing to become immersed in LIGO and seeking a role within the collaboration. The Laboratory further hosts a large number of summer students at the four LIGO sites who are mentored by LIGO scientific staff and faculty. A number of these projects involve the use of LIGO data for astrophysical analyses.

The LIGO Laboratory is a member of the DOE/NSF "Interactions in Understanding the Universe" (I2U2) national program for high school teachers and students. I2U2 builds on the successful DOE QuarkNet outreach model. LIGO is one of several large projects that support the overall effort by authoring data access and analysis tools, and by piloting the classroom use of LIGO data with a group of local science teachers. The proposal's intent is to bring school students and members of the public into direct contact with authentic data. LIGO archives data from its large suite of (non-astrophysical) *physical environment monitors* (PEM channels). These data are publicly available for use by students and teachers. These channels include infrasound monitors, seismometer arrays, magnetometers, multi-band RF receivers, acoustic microphones, wind anemometers, etc.

The data analysis work of the collaboration is conducted in four major working groups, each one of which targets a certain class of sources (transients, compact binary coalescences, periodic sources, and stochastic backgrounds of gravitational waves). Membership in the data analysis groups is open to all LSC members. The groups are responsible for algorithm development, data analysis, interpretation, and discussion of the results. These groups also provide a forum for teaching new researchers about the idiosyncrasies of gravitational wave data, access to computational resources matched to the data analysis challenges, and internal review of results to ensure validity of results during the time when signals are still at the margins of detectability. The Collaboration has established a Detector Characterization group which cuts across the search groups, and which plays a key role diagnosing instrument and environmental artifacts.

LIGO has entered into joint data sharing agreements with two international partners. In 2001, LIGO and GEO signed an MOU establishing the LIGO-GEO600 interferometer network and giving LIGO and GEO reciprocal rights to all data taken by the LIGO-GEO network. GEO600 is operated by scientists from the UK and Germany and funded by the Max Planck Gessellschaft and the Science and Technologies Facilities Council. GEO scientists are members of the LSC. In 2007, the LSC and the Virgo Collaboration signed an MOU covering joint data analysis and run planning with the goal of maximizing the scientific capabilities of an extended network of detectors including the Virgo detector. Virgo is a French-Italian interferometer jointly funded by the CNRS and the INFN. Virgo members are not members of the LSC, but the international partnership with Virgo has led to the formation of joint data analysis working groups composed of members from both collaborations. The Data Analysis Council (DAC) of the combined collaborations oversees the ongoing activities in all working groups. Results of work in progress are reviewed in multiple weekly meetings at many levels, including bi-weekly meetings of the DAC.

Before consideration for external distribution, presentation, or publication, review teams from both collaborations who are assigned to individual analyses vet all results. These teams, composed of members who are not working closely on the analysis under review, scrutinize the software pipelines, algorithms, results, and interpretation. The review teams make recommendations to the DAC, who then advocates readiness for publication or presentation of specific results.

The complexity of the data analysis process currently requires a timescale of approximately eighteen months from data acquisition to results approval for each paper produced by the collaboration. While this time has decreased since the initial LIGO science run S1, the reason for the extended period of time is that often additional steps are required: (i) data calibration, (ii) data quality assurance and vetoing, (iii) algorithm and review, pipeline efficiency determination via large-scale Monte Carlo simulations, (iv) astrophysics results review and interpretation, (v) manuscript preparation, editing, review, and approval. For example, the most recent science run, S5, ended in October 2007. To date six papers from the first half of S5 (i.e., through January 2007) have completed the process described above: there are still approximately 12 papers in various states of maturity that the different working groups plan to carry through to publication. Based on experience from the prior S4 science run, the last paper to be published from a particular observational science run could take up to four years after the data were acquired.

Summarizing the discussion above:

- The expectation of detections in the current phase of LIGO is still rare, with signal to noise ratios near the limit of detectability. Detection will become routine only with the second generation of instruments presently under design, development, and construction.
- Data analysis with these first-generation LIGO instruments is complex and requires the serious engagement of instrument scientists who intimately understand how the interferometers operate and how the strain measurements are made at almost all steps in the analysis. Data preparation and quality assurance is a time and personnel intensive process that requires the effort of many scientists and detector engineers.
- The LSC is a collaboration that is open to all scientists who are interested in performing astrophysical investigations with LIGO data.
- The LIGO visitors program enables newcomers to the field, who wish to learn more about LIGO and apply for LSC membership, to spend significant time working within LIGO. The NSF REU program and Caltech and MIT undergraduate research programs provide a means of engaging students in LIGO summer research.
- LIGO has developed international agreements for sharing data and performing joint analyses with international partners GEO600 and Virgo.
- Some LIGO data, namely physical environmental monitors, are already shared outside the collaboration. Together with other national laboratories, LIGO has an outreach component, I2U2, that targets K-12 science programs.

1.4 Considerations in developing a LIGO open data policy

A number of considerations naturally arise in thinking about the development of an open data policy. We discuss some of them here.

- *The nature of the data, the size of the data stream, and what analysis tools are needed to manipulate and analyze the data.*

LIGO data differ from most observational astrophysical or physics experimental data in that the data are primarily noise. Any astrophysical signals in the data have low duty cycle if they are transient gravitational wave events and small amplitudes if they arise from continuous gravitational wave emitters. There are correlations with environmental perturbations and correlations with ancillary instrument variables. The data from two detectors at the same site have correlations. The analysis for signals relies on coincidence at different sites in a transient detection and correlation between sites for a continuous detection. Similar, though not identical, analysis problems occur in geophysics in the search for earth normal modes in the continuous case and small amplitude earthquakes or the detection of nuclear tests for transient events.

The first generation LIGO generates 8 TB/year in the primary data channel, the gravitational wave strain channels for three instruments, and an additional 300 TB/year in auxiliary channels. Even when there are no astrophysical events present, the data stream contains many features, such as idiosyncratic instrumental transients or transients produced by the environment. In addition, the data contain simulated ('injected') signals which are used to test the efficiency of the data analysis pipelines. Before a meaningful analysis can take place, the data must be cleaned using vetoes and data quality checks. This cleaning process is computationally and labor intensive, and is imperfect, i.e., it does not completely remove all instrumental artifacts.

The software tools needed for analyzing the data differ widely, depending on the specific class of signals that are being searched for. Searches for binary inspiral signals use matched filtering with a pre-determined template bank consisting of thousands of target waveforms. Searches for unmodeled burst sources use algorithms that look for excess power above the background in narrow time windows and frequency bands. Searches for continuous signals from known pulsars also use matched filtering, integrating over long periods and taking into account amplitude and Doppler shifts induced by the relative motion of the source and the detector. Searches for stochastic gravitational wave signals rely on calculating time-averaged cross-correlations from two physically separated detectors to look for a signal floor above the background.

Storing and analyzing such data requires significant IT infrastructure for data management and processing. The LSC has developed and maintains large computing facilities and a large body of software. LSC software is now in the public domain via the GNU open software license agreement, and includes instructions for compilation and installation on specific platforms. The full scope of what would be provided to support a full data release still needs to be defined, and will take place during the next several years.

- *The broader research community of potential users.*

Astronomers and astrophysicists who study neutron stars, black holes, gamma ray bursts, supernovas, and cosmology are most likely to use LIGO data, since gravitational waves can reveal information about these sources. In addition, physicists specializing in gravitational wave physics, numerical relativity, and strong field gravity would likely benefit from access to LIGO data.

Based on discussions with astronomers and astrophysicists in these fields, it is unlikely that anyone outside the LIGO Scientific Collaboration or other gravitational wave collaborations will want to directly analyze LIGO data on their own in the pre-discovery era. The primary reason for the lack of interest in pre-discovery data is that the external astronomical community recognizes that gravitational wave astronomy, unlike established astronomical observational techniques, is still in its infancy and not yet in a 'normal science' mode. The LSC currently collaborates with a number of astronomers; however, by mutual agreement the analysis of LIGO data is conducted solely by the LSC.

In the post-detection era, i.e. after LIGO has detected many validated gravitational wave signals, interest in independent analyses of LIGO data is likely to grow rapidly from astronomers, astrophysicists, the gravitational wave physics community, and possibly others.

1.5 Staged release of LIGO data

Based on these considerations, our proposal calls for a staged release of LIGO data. In general, we can contemplate two broad stages for LIGO data release:

Detection or Discovery Phase: In the near term, when detections are rare and the understanding of instrumental artifacts and other spurious influences on the data is still evolving, prompt release of *validated* gravitational wave data during and for a suitable time epoch surrounding confirmed events makes the most sense. In this stage, detected events are publicly released, analogous to astronomical post-processed image data. The collaboration would release events for scientific use by the broader research community. Release of events would occur either concurrent with or shortly after publication of these events in the scientific peer-reviewed literature. This stage could begin as early as 2009, when enhanced LIGO goes into operation for the next major LIGO science run.

Detected events would be released in a predetermined format, e.g., the National Virtual Observatory (NVO) model for publication of astronomical data. The LSC engaged the NVO community when the NVO was defining data exchange formats based on XSIL (extensible scientific language). Consequently, a data object, *GWEvent* (analogous to *VOEvent)*, may be defined within the NVO construct in anticipation of future LIGO event data releases[1]. The interface to the broader community would be via web-based tools developed and distributed by the collaboration, in coordination with the NVO. Tools would be developed to manipulate the *GWevent* database to allow the broader research community to process and perform a variety of analyses on the database. Calibrated time series h[t] data frames containing the detections would be published along with associated metadata, which would include detected and reconstructed parameter estimates, along with their uncertainties. The volume of data involved in *GWevent* releases would be very modest, and an entire science run should easily be able to fit on, e.g., a USB memory stick. It should be noted that this style of data release primarily addresses short duration events and possibly very narrow bandlimited continuous wave type signals; however, it would not be suitable for broadband stochastic signals.

Observational Phase: Once detections become common and the understanding of the data characteristics permits removal of instrumental artifacts and environmental influences, release of the full, calibrated time series data will be both practical and essential for maximal scientific utilization. In this stage, LIGO would release continuous calibrated LIGO strain data, h[t], and possibly related products, such as auxiliary channels and time-frequency spectrograms. We would also continue to release *GWEvent* data for detected events. The level of effort to produce full data products will be considerable and significant resources that do not now exist would need to become available for the collaboration and the Laboratory to be able to undertake this effort (ref. the discussion on costs, below). This stage occurs when LIGO is routinely observing signals, and is expected to begin once the routine science operation of Advanced LIGO for astrophysical observations has begun, following its commissioning. Advanced LIGO is scheduled to be completed in the second half of 2014, followed by a one to two year commissioning period to bring the instruments to design sensitivity.

Releasing the continuous h[t] or related products requires a continuous and regular program of data clean up and data quality assurance. Currently, this takes place within LSC working groups. At present, there is

[1] Refer, e.g., to the following www sites: the Caltech event repository and publishing , <http://www.voeventnet.org/>; the NOAO NVO web site, <http://voevent.noao.edu/>; the NVO event standards working group, <http://voevent.org/>, and the following publication on "GCN and VOEvent: A status report", <http://www3.interscience.wiley.com/journal/117927647/abstract>.

no means provided for creating the cleaned up h[t] stream as a final product: it is performed within the software pipelines that have been developed. In fact, the concept of a single "cleaned" h[t] stream is ill defined: it depends on the particular search involved. Moreover, during the course of a science run, generation of h[t] data is an iterative process because new versions of h[t] are generated each time the calibration is refined. Within the LSC, it has been possible to remove old versions of h[t] with less accurate calibrations because the data release and access is controlled. Once we are in an era of open data releases, this will not be possible. The time it takes to converge on a final calibration sets one time scale for the period before data are made public (see discussion below).

Releasing the full data stream would require a dedicated science data product team within LIGO to ensure data integrity. There are a few possible approaches to h[t] release, and it is not clear at this point which will be most effective. The first could be to iteratively remove features that have been identified as artifacts. However, this process is not perfect and some instrumental artifacts will remain in the data. Another approach could be to provide extensive lists of vetoed time intervals; this approach would rely on the external research community to take proper measures to exclude those data tagged with vetoed times from their analyses. Still another approach for consideration is to blank out data epochs containing known artifacts, thereby producing a continuous stream of data segments interspersed by zeros. Moreover, experience has shown that the process of producing a clean h[t] will invariably involve multiple version releases which require some analyses to be redone when subsequent releases are made. In the case of h[t] release, data from a yearlong science run from multiple interferometers would amount to several terabytes of data.

1.6 The cost of making the data publicly available.

The monetary cost of making data publicly available depends on the nature and scope of the data release. Our estimates are based on a survey of the labor and computing needs from each of the four data analysis groups to produce clean and vetted data and from the calibration group, the detector characterization group, the software analysis tools group, and the computer committee.

<u>Phase 1</u> – release of validated gravitational wave events under the *GWevent* model – would cost an estimated \$0.2 million with annual recurring costs on \$0.2 million/year. Initial costs would include development of web-based tools that conform to the NVO formats. In addition, one scientist would be needed from each search group on a part time basis to process and post events.

<u>Phase 2</u> - We estimate that the release of full, calibrated time series from the LIGO interferometers would cost the NSF *an additional \$3-4 million per year*. Approximately 8-10 full time scientists dedicated to data validation, cleanup and production oversight, 8-10 software developers, one system administrator and 2-3 web developers would be needed to complete the necessary tasks laid out above for all data collected from the interferometers. In addition, capital equipment costs for data storage, networking, and serving would add another \$100k. We estimate that a three-year preparatory period will be needed by LIGO to implement a full, open release policy. Given the nature of this work, it is likely that the major effort would have to be provided by LIGO Laboratory as an NSF national facility. This would be in keeping with NASA's model for data preparation and release. An important cost driver will be the need to provide software tools, software consultation, and help-desk functions. This is not something that Laboratory and the Collaboration now typically provide, since the expertise is resident in the scientitifc working groups which analyze the data.

As a basis for comparison, the several NASA projects we approached indicated that typically 10% of their operations budgets go to the support of public data release. The annual NSF budget for LIGO operations and the LSC research program is approximately $42 million per year, so the estimates above are in line with NASA experience.

1.7 Other considerations

Some aspects of the proposed data policy have not been fully explored and must still be resolved.

- *Implications of making LIGO data publicly available which is 'owned' by international collaborations.*

  LIGO has entered into several data sharing and joint analysis MOUs as part of its leadership role to establish an international global network of gravitational wave detectors. These MOUs stipulate how LIGO data are collectively managed and used. A change in policy leading to the public distribution of LIGO data will have a significant impact on these international agreements. LIGO has engaged its international partners on this issue, however there is no official position now.

  We strongly believe that any policy for open release of gravitational wave data should be coordinated internationally. Global gravitational wave detector network analysis has emerged as one of the most powerful means for searching for gravitational wave events. Network analysis requires the use of data from all detectors on the network to be able to better localize events and set lower detection thresholds. An asymmetric policy in which only one or a few collaborations release data will complicate the development of international agreements and possibly compromise the ability to do the best gravitational wave astronomy.

- *Length of time between acquisition of data and their public release.*

  In order to allow sufficient time for the collaboration to calibrate, vet, and clean up the data before release and to allow it to perform its own planned analyses and publish them, there should be a delay between data acquisition and data release. This recommendation is in line with policies currently in place in NASA. Based on experience, the process of cleaning the data (removing instrumental artifacts, tagging hardware injections, providing accurate calibrations) and performing analyses takes from 18 to 24 months. We feel that, initially, a two-year delay before public release is appropriate at this point in our field. For comparison, NASA policy varies, but typically there is a 12-month delay in data releases, during which time mission scientists have access to the data for analysis, vetting, and clean up before they are made available to the broader community. However, it should be noted that, unlike LIGO, NASA has a history of doing public data releases. The first NASA mission to release data was the Cosmic Background Explorer (COBE) in 1990. The delay in release for COBE data was 36 months.

1.8 Summary.

As the McDaniel Panel report made clear, the era prior to and immediately after the first confirmed direct detections of a gravitational wave event is a fundamentally different one from the subsequent era when detections become a regular occurrence – the first detections herald the beginning of gravitational wave astronomy as an observational field in its own right. The first announcement will be a monumental milestone in the field of physics and general relativity – as important as the announcement in 1919 of the observation of the bending of starlight in the sun's gravitational field or the announcement by Hulse and

Taylor of long-term observation of orbital energy loss in a compact binary pulsar system through the radiation of gravitational waves.

Our approach calls for a phased release of LIGO data to the broader research community that takes us from the *status quo* of today to an epoch we envision when Advanced LIGO is making routine detections. Our recommendations derive from the considerations above and responsibly respond to preserve what we feel should be both the NSF's and the scientific community's primary goal, namely the earliest possible detection of gravitational waves and the attendant opening of the emerging field of gravitational wave astrophysics as well as to maximize the science that can come from LIGO. Our phased approach is guided by these principles, recognizing that our field is still in its nascent stage. We identify two epochs:

(i) <u>Detection Phase.</u> This is the current phase of gravitational wave astronomy, and it demands a tight coupling between experimental scientists understanding the instruments and their idiosyncrasies with the scientists who have taken the responsibility for the data analysis. The scientists operate within the structure of the LIGO Scientific Collaboration (LSC). The LSC is a collaboration open to all scientists who are interested in the field of gravitational wave astrophysics and the technology of gravitational wave detection. In this phase of the research, it is critical that the major effort be directed to the initial detection of gravitational waves with scientific reliability. There is not yet sufficient experience and knowledge to replace the vetting and control asserted by the collaboration with automated procedures. Nevertheless, the collaboration has been preparing the groundwork for the future phases with broad data policies by developing uniform data formats, techniques to couple parallel processing across the network and use of the large scale computing grids. During this first epoch, which includes the next LIGO science run in 2009-2011 and the first months of operation of Advanced LIGO, LIGO will make any science data associated with detections that the Collaboration publishes available to the broader research community. At the present time, the Collaboration is developing plans to engage ever more closely other astrophysical disciplines and observatories that promise concomitant signals to LIGO's in order to enhance the likelihood of as early a first detection as possible. Our discussions with the external community of astronomers reveal that it is unlikely that there will be any serious customers for the data in the pre-discovery era.

(ii) <u>Observational phase.</u> This epoch will begin once observations become routine and the detector is operating in a stable and understood fashion. In this epoch, it will be scientifically valuable to involve a larger community of scientists beyond those in the LSC. LIGO will broaden efforts currently underway to engage more closely other astrophysical disciplines and observatories that can provide complementary data in order to enhance the science that is possible combining LIGO and other observational modalities. One thrust during this time will be working to understand the physical significance of the waveforms and beginning a categorization of them -- the so-called "inverse problem." At the same time, other astrophysicists will be working to establish population statistics and catalogs of events, and using gravitational wave observations for other investigations (e.g., cosmological studies)

In this epoch, LIGO is committed to distributing the entire body of gravitational wave data to the broader research community. Demand for LIGO data will grow as the field is established through the reports of validated gravitational wave events, and there may be ways to exploit the full LIGO data set beyond what will be done by the working groups of the LSC. Although it is difficult to predict with certainty, it is likely that the observational era will begin several years after the Advanced LIGO Project is completed. Thus, preparations for implementation should begin in 2012. As discussed earlier, additional funds will be needed from NSF to ensure that this happens in a timely manner.

We have arrived at this approach by carefully considering and balancing the needs of the current organization of the U.S. ground-based gravitational wave astronomy community. We have consulted a number of astrophysicists in related disciplines who are familiar with LIGO today, who are not members of the Collaboration, and who have an abiding interest and appreciation for our field. Last, we also considered the implied additional costs associated with the new scope we describe above. We believe such a phased progression of data release that evolves as the field gravitational wave astronomy matures will maximize the astrophysics that can be achieved with LIGO data by LIGO and the broader research community.

Appendix A

Persons and organizations consulted in developing this open data proposal

Beverly Berger, NSF Program Director for Gravitational Physics

Edmund Bertschinger, Chair, MIT Department of Physics

Karsten Danzmann, Director, Max Planck Institute for Gravitational Physics, GEO600 Principal Investigator.

Francesco Fidecaro, Università di Pisa, INFN Pisa, Spokesperson for the Virgo Collaboration.

Neil Gehrels, PI for SWIFT and Deputy Project Scientist for GLAST, NASA Goddard Space Flight Center, and member of the LIGO PAC (see below).

George Helou, Executive Director for IPAC and Deputy Director for Spitzer Science Center, California Institute of Technology.

James Hough, Professor of Physics, University of Glasgow.

Cole Miller, Professor of Astronomy, University of Maryland and member of the LIGO PAC (see below).

Bernard Schutz, Director, Max Planck Institute for Gravitational Physics.

Max Tegmark, MIT, Large Scale Structure Analysis Team, Sloan Digital Sky Survey

Rainer Weiss, Emeritus Professor of Physics, MIT, Chairman, of the COBE Science Working Group, and Founding Spokesman of the LIGO Scientific Collaboration.

Roy Williams, Caltech, NVO Co-Prinicpal Investigator

The LIGO Program Advisory Committee (PAC): Martin Breidenbach, Deepto Chakrabarty, Lynn R. Cominsky, Neil Gehrels, Giorgio Gratta (Chair), Cole Miller, Fulvio Ricci, Erich Weigold, Benno Willke, Sidney Wolff, Glenn Young.