



Report on
Community Input
 for
LIGO Open Data



version 1.2, January 2012, LIGO-P1100182

edited and compiled by
 Roy Williams , Christian Ott, Steve Groom

Table of Contents

1. Introduction2

1.1. Gravitational Wave Open Data Workshop (GWODW)3

1.2. Open Data Survey3

1.3. Comments on Open Data.....3

1.4. The First Detection is Different.....4

2. Rapid Alerts.....6

2.1. Dissemination6

2.2. External triggers7

2.3. Rapid Alert Community7

 2.3.1. *Getting Follow-up Resources*.....7

 2.3.2. *Joint Observations*9

2.4. Feedback for Follow-up Observations9

2.5. Trigger Repository.....9

2.6. Very Low Latency Alerts10

3. Source Catalog11

4. Full data12

5. Building the LIGO Open Science Center.....13

5.1. Documentation and Curation14

5.2. Mock Data Challenges.....14

5.3. Software.....14

6. Appendix.....16

6.1. Gravitational Wave Open Data Workshop.....16

6.2. Open Data Survey text.....16

7. References19

1. Introduction

In 2011, LIGO began the process known as ‘Open Data’ – ensuring public access to data from the observatories, as mandated by the NSF. When the exciting era of gravitational-wave astronomy begins, in the next few years, LIGO will be ready to extract the maximum scientific return by working with the broader community, taking into account both the unique nature of the first detections, and the expectations of those who have spent their careers building the detectors. The Open Data program is defined in a ‘Data Management Plan’ (DMP), whose first version was released by LIGO in January 2011. One of the objectives listed in the DMP is to ‘understand requirements from the broader community’, which motivated a survey and a workshop to gather opinions from the broader community; this document reports on results from both.

The LIGO Scientific Collaboration (LSC) has over 800 members from over 70 institutions to ‘explore the fundamental physics of gravity, and develop the emerging field of gravitational wave science as a tool of astronomical discovery’. The LIGO Laboratory manages the LIGO Observatories at Hanford, WA and Livingston, LA, and is the key institutional member of the LSC with about 200 of its technical and scientific staff being members of the LSC. Members of the LSC have exclusive access to the data from LIGO and GEO observatories. In 2007 the LSC signed an exclusive data sharing agreement with the Virgo Collaboration, and results from the data from LIGO and Virgo detectors are published with an author list consisting of about 700 LSC and Virgo Collaboration members. A renewal of the LSC-Virgo agreement in 2011 agrees with the provisions in the LIGO Data Management Plan (for LIGO data).

In the following, we use the term ‘broader community’ to be those who are *not* members of the LSC or the Virgo Collaboration, and ‘open data’ to mean data that is accessible by the broader community. When discussing open data, it is critical to understand who is an LSC member and who is not, since the former already have full access and the latter do not.

The LIGO Data Management Plan [1] defines plans for data release by the LIGO Scientific Collaboration (LSC) in the advanced detector era. In the discovery phase (phase 1) these plans foresee limited data releases around detections and around significant non-detections, for example in conjunction with interesting electromagnetic/neutrino observations. In the observational phase (phase 2) a much broader data release is planned, including near-real-time alerts. Furthermore, 24 months after data taking, all LIGO data will become publicly accessible.

The 24 month proprietary period comes from the white paper "A Proposal for Providing Open Access of LIGO Data to the Broader Research Community" [5] (January 2009), where it says "*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.*"

The material below is based on two kinds of feedback from the astronomical community: a workshop attended by 45 people at LIGO Livingston in Louisiana 2011 October 27/28; and a survey that was requested from a large number of astronomers and astrophysicists, which received 80 replies in the month of October 2011. In both exercises, the participants included members of the LIGO Scientific Collaboration and members of the Virgo Collaboration, as well as other astronomers who do not have current access to LIGO data. Much of the content below is quotations from these people, sorted and organized into categories, each with a short summary. Quotations from the workshop transcript have been

edited from the circumlocutions of live speech to a more concise form that preserves the meaning. In the following,

“quotations and near-quotations are in their own font, like this.”

The quotations do not intend to represent any community, but only individual opinions of the scientists who shared their opinions.

The organizers thank the people¹ who have contributed by being directly quoted in the notes below, as well as all those who came to Louisiana for the workshop, and those who filled in the online survey.

1.1. Gravitational Wave Open Data Workshop (GWODW)

The GWODW workshop [2] brought together experts from the gravitational wave community with astronomers and open data experts in a two-day workshop to address the question: ‘How should LIGO proceed in releasing gravitational wave data to maximize scientific return?’ While there were many science talks about GW sources and possibilities of detection, there was also discussion of the LSC current and future data release policy and publication models. The workshop featured ‘breakout groups’, where small roundtable discussion was tasked with specific issues, and the leader of each breakout group presented the results of the discussion to the plenary. The quotations listed below come from a transcript of those presentations, suitably edited for print over speech. The agenda for the workshop is listed in the Appendix.

1.2. Open Data Survey

This web-based survey [3] was initiated to gauge the community requirements for the Open Data release, and was widely distributed among astronomy and astrophysics mailing lists on October 1. This document summarizes the 80 responses obtained 2011 Oct 1 through Nov 9. Of these, 22 identified themselves as LSC members (28%). There were 27 observational astronomers, 24 theorists, 16 members of other gravitational-wave detection efforts (such as Virgo), and 16 numerical relativists (may add to more than 80 because multiple categories were allowed). There were 14 questions on the survey, as defined in the Appendix.

1.3. Comments on Open Data

At the end of the open data survey, participants were asked for their comments on the whole Open Data program. Some responses are below.

“I recommend the data release process be as open and transparent as possible.”

“Release of real time triggers, with directional probability maps, will be most useful to the broader observational astronomy community, rather than the raw/processed LIGO data.”

“Keep up the excellent work. The sooner the better.”

¹ Bruce Allen, Jean-Luc Atteia, Scott Barthelmy, Patrick Brady, Joshua Bloom, Geoff Bower, Nat Butler, Ranga Ram Chary, Sam Finn, Dale A. Frail, Neil Gehrels, Gabriela Gonzalez, Steve Groom, Michael Kramer, Andrea Lommen, Brian Metzger, Cole Miller, Benoit Mours, Christian Ott, M. Alessandra Papa, Tom Prince, Peter Shawhan, Xavi Siemens, Alicia Soderberg, Marcin Sokolowski, Tom Stephens, Chris Stubbs, Rai Weiss, Roy Williams

“Much of what I have written is less associated with an external release and more relevant to operating within the LIGO Collaboration. Part of the reason I have not involved students (nor myself) into the data analysis is because access to the data is damn cumbersome. I am hoping that by preparing data for the outside community we will also make it easier to access the data and also to be able to use the programs that have been developed by the Collaboration.”

“The LIGO Laboratory is a national facility. The data should be generally available to the scientific community. There is no legitimate reason to delay data release beyond what is required for calibration and distribution to repository sites. Restricting access to data is inimical to quality science. It prevents the independent replication or checking of analysis results; more insidiously, it shields disagreements and alternative interpretations of observations from the broader scientific community.”

“When the LSC was formed in the late 90s, LSC management (Barry Barish and Gary Sanders) said, “for now, the doors are open, but once we start to take science data, they will shut”. The philosophy at that time was that access to LIGO data would ONLY be available with collaboration membership. So the idea was that the data was “open” but only in the sense that to get access to LIGO data, one needed to join the LSC and work collaboratively together. Those of us who work mainly on data analysis software agreed in 2005 to put our LIGO data analysis codes under GPL license. Many of us (including myself) did so willingly, because we reasoned that our scientific investment in these codes (which represent many people-years of work) were protected, because to use them for GW science, one would need LIGO data and thus would need to be an LSC member. The current situation leaves us feeling that a “bait-and-switch” has taken place. If LIGO data becomes available to everyone, then our scientific competitors from outside the LSC can benefit from OUR work, but we won’t benefit from THEIR work. In fact the current GPL codes can be used by anyone without even acknowledging our contributions. To simply say “oh, this is all good for the progress of science, so it doesn’t matter, it is just sociology” is ignoring the reality of people’s careers and motivations. This is not a healthy situation, and it is one of the reasons that I am wary of the Open Data movement and its possible consequences for the LSC.”

“24 months is too long to generate much interest in non-LSC astronomers. They will assume that the data has already been mined thoroughly. Furthermore, 24 months is not an incentive on LSC authors to work faster and better. We should push toward shorter lags---perhaps formulating the policy as ‘data will be released as soon as the LSC had done such and such... but not after 24 months’.”

“The detailed plan of data release and overall coordination of joint observations should be developed with the goal of optimizing the scientific output. Realistic simulations of the joint searches are needed to determine optimal observation strategy and data analysis requirements. Special effort should be dedicated to characterizing the background of accidental transients (those not related to GW) for the participating telescopes.”

1.4. The First Detection is Different

The current LIGO Data Management plan does not include public releases of data or triggers until Phase 2, in an era of regular detections. In the workshop, there was discussion of earlier releases of triggers, even before the first gravitational wave detection is confirmed. While observational astronomers are clear on the science benefits of early release (‘Unambiguous detection of an EM counterpart is critical for GW discovery declaration.’), many in the LSC see the first GW detection differently: as Rai Weiss put it: ‘We want to be absolutely sure we’re right.’

“Extracting GW signals from instrumental noise is difficult: the signal is weak and noise is as complex as the instrument, there are a serious risks of misinterpretation. Therefore Virgo plans to release only reliable information.”

“I’m attempting to catch what really distinguishes [first detection] from the steady-state observing. I think that the fundamental fear is that the discovery will, in fact, not be an LSC/Virgo paper, ... rather reporting a counterpart ... observed because of a gravitational-wave trigger... Of course, it comes from the gravitational-wave signal initially, so [they can’t] pretend it wasn’t the LIGO and Virgo that detected [it] originally, but if that other paper comes out first, that’s the thing people will first pay attention to, and that definitely changes the impact.”

“[The LSC has] a privilege, we have the data and we generate the alerts, and we generate these alerts in conjunction with Virgo, so this data will not be public for the first discovery. We are going to generate the alert, and, of course, we are working very hard to make these alerts timely and accurate and as useful as possible.... LSC does not have the expertise to organize an observing campaign. We do need partners, and we don’t have telescopes.”

“It’s certainly not in the plans and it would be very difficult to have a public release of triggers that are happening before the first detection. It’s not impossible, but it’s not the default plan at this time.”

“All the issues around the first detection call for explicit coordination.... We want to be absolutely sure we’re right ... we don’t want to put out something marginal. In fact the electromagnetic observations could be the thing that tells us that we’re absolutely right: ... unambiguous detection of the counterpart could be critical for the discovery declaration. Those are the two sides of things that we’re trying to balance and to take advantage of in an appropriate way.”

“So, possibly paradoxically it was felt that in this first detection era we really had to do a lot of explicit working together and coordination to allow things to proceed with less-explicit coordination in the long term.”

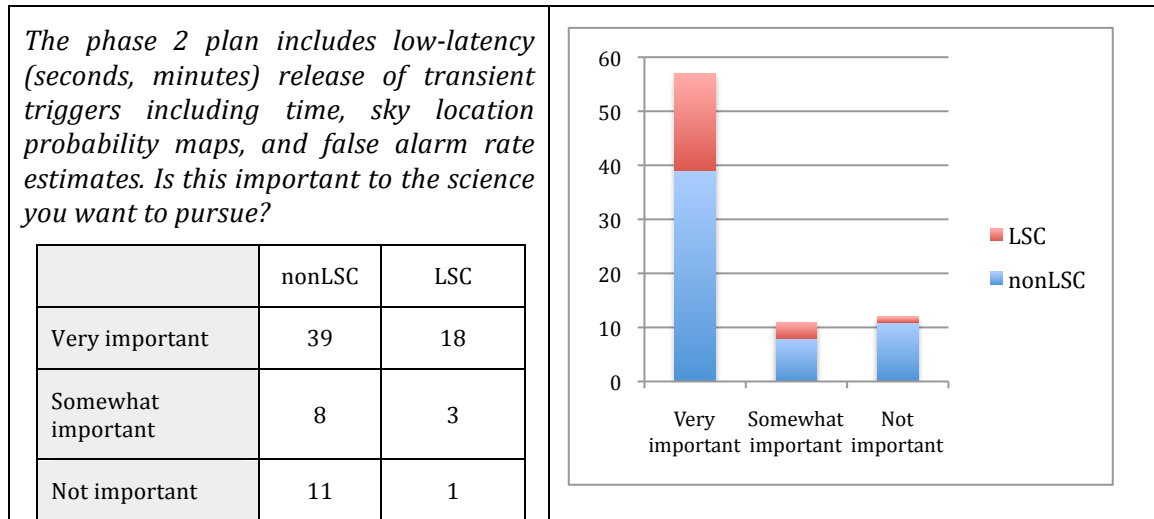
“The fundamental fear: That the ‘discovery paper’ will not be an LSC-Virgo paper, but a paper reporting an EM counterpart that was found thanks to the GW trigger.”

“We could imagine there are MOUs that might be rather basic. They might only touch on a couple of things. For instance, don’t publish until we’ve had a chance to analyze our data and say whether we have an event or not, so that would be something that would address that fundamental fear.”

2. Rapid Alerts

2.1. Dissemination

Question from the survey:



“Release of real time triggers, with directional probability maps, will be most useful to the broader observational astronomy community, rather than the raw/processed LIGO data.”

“We’ll follow-up GW alerts, but not only. As such we may have different policy for data sharing and property, depending on the origin of the alert.”

“I believe integration into the tools and techniques familiar to observational EM astronomers will be key in enabling good cross-force science. E.g., VO, FITS tables, VOEvent, Astronomer’s Telegrams.”

“[Events will have a] probability map, and these are much larger things than a single pair of numbers (an RA and a Dec), so ... the notice has a URL in it, and then ... robotic operations can extract that URL and do a wget on it and download ... and then they can go searching through probability space here, picking out the tallest islands first and tiling outwards.”

“On the longer time scales ... there is Astronomer’s Telegram and GCN Circulars, [where humans] write real sentences and paragraphs about what was found ... with caveats and asterisks.”

“[Use] standard mechanisms like GCN or Skyalert, backed up by detailed information on a web site and information aimed at humans ... we did also recognize that there will be multiple notices ... as information is updated after the initial information.”

“What protocols are there to push selected triggers out to the right place? ... There’s two real-time networks. Roy Williams runs Skyalert and Scott Barthelmy runs GCN, ... and they both have multiple protocols, [including] VOEvents ... GCN has ... sockets and various forms of e-mail, and then they both have pull technologies. You can go to the web sites and everything’s archived there, and then you can download anything that happened earlier in the morning or a month ago.”

“There are two different time scales. You can do the immediate notification, T plus 10 seconds, 100 seconds, whatever the time scale of your event is. Inspirals will obviously be short

events, so it mostly depends not on the event itself, but what your process and detection time is.”

“It’s no longer an information-technology problem but rather a project-management or science-management problem, of deciding how the project wants to proceed and whether it’s a project-driven thing or a community-driven thing. [There are] technologies out there ... many different models, and it’s a matter of deciding which one you want to do.”

“[To begin], enable several prototypes and see what develops, how the community evolves and which ones they’d prefer. Then centralize on those and maybe enhance them.”

“This field is very dynamic. Nobody expects anything to be perfect at T plus 10 seconds, so there’s a whole series of upgrades, improvements, corrections, etc. [I] would encourage the LIGO team: don’t feel you have to get it right the first time at 10 seconds, 100 seconds after the event. Nobody expects that.”

“Sending the triggers out to the world is not the most important thing; it’s the **only** thing.”

2.2. External triggers

These comments consider the idea of rapid LIGO analysis and reporting, following an event in the sky that might be thought to produce gravitational waves. The sky localization of these events assume detection with a network of gravitational wave detectors, including LIGO and Virgo detectors. Could such reports – detections or non-detections – be disseminated rapidly to the broader community?

“[Consider] externally triggered LIGO search, responding to the gamma-ray burst. People ... will see LIGO is taking data, and maybe we can even put out the range or something like that, and this might then be of interest for dedicated follow-up.”

“It does raise one interesting point that wasn’t actually discussed ... the sort of rapid publication of information on non-detections, for example, externally triggered searches. It seems clear that if we want to join in ... following up things like GRBs, it’ll be interesting to ... send a [notice] within a day that says, “We followed up that event, and this was our limit”.

“Maybe it would be nice to have continuous optical monitoring for LIGO. Of course this is already done to some extent, but the real motivation of this was ... [a discussion] that maybe there could be types of exotic supernovae that have extreme asymmetries that would make them detectable. The hope would be that you could guarantee coverage of all the potential nearby galaxies, and I’m not talking about necessarily every night, as amateur astronomers are doing, but maybe every 10 minutes ... a global series of very wide field telescopes. “

“It’s important to send out also the status of the detectors, a web page where everyone can see ... what detector’s on or off, especially in the case of an externally triggered LIGO search, ... and maybe we can even put out the range or something like that, and this might then be of interest for dedicated follow-up.”

2.3. Rapid Alert Community

2.3.1. Getting Follow-up Resources

The requirements for follow-up of events from the global GW network are different from those in optical astronomy: rather than a precise sky position, the GW detectors will produce a ‘skymap’ (probability density), that may have an ‘island chain’ of local probability maxima. Here we consider how to mobilize astronomical resources quickly and deeply to find the EM follow-up.

“An idea that ... met with a lot of positive response was instead of the usual target of opportunity proposal ... [but rather] make direct arrangements with the directors of these facilities – who understand the importance of this – to dedicate the observing time and the analysis of the data in the service of this grand project of the detection and follow-up. This would, then, bypass the politics and the hassle of the competition for who would be the PI on this project or whatever and just achieve that, those follow-ups, without those complicating issues.”

“It would be useful for LIGO ... to speak to all the major observatory directors to motivate them, to say that if we have a significant event in the gravity-wave signal then they commit some target-of-opportunity time on their telescopes. [Otherwise] there will be significant reluctance by whoever is at the telescope to give up their time just to map out tens of square degrees for something which may not [produce] success.”

“A facility that might be useful would be when a telescope does decide to look at this location on the sky that they can announce that they’re [pointing] there. And so other people can decide, ‘Well, I’m gonna look there too,’ or, ‘I’m gonna look someplace where no one is looking,’ so not a top-down direction but rather an information-sharing mechanism.”

It would work well with “humans in the loop”, who are doing something an hour later, half a day later, because they ... can look and see, well, I’m 12 hours late, so to speak, and this whole island has been mapped ... down to 20th magnitude, so then they would go to the next island.”

“We discussed for a while about dividing up the follow-up community: these three robotics go after this island and these [others] go after this other island... Speaking for myself, I don’t really want to arbitrate this – you get it and you get this one and you get that one, and a day later you won and these guys didn’t. I know there’s all these kinds of statements about [agreement] and tiling of the sky, and they’ll all be authors on the paper. But that one guy, he’s gonna be able to write in his grant-renewal proposal that he discovered it, and the other two guys won’t be able to say that.”

“How about a way for follow-up observations to be collected, astrometrically federated, so that the progress of the joint observation can be seen in real time?”

“There may be scientific partnerships formed, even if all the data is out there in the open. And going beyond whatever data we release to everyone, there may still be special arrangements for following up low-threshold events, and that would be by special arrangements with partners, even in the open-data era.”

“The Kepler project right now is identifying candidates and coordinating how follow-ups are executed among the science team. How do you allocate resources, who will observe what? [There’s] a stream of things that need to be followed up ... the same thing comes out of these automated surveys. The PTF, I believe, is generating transients that then have to be followed up somewhere else, and PTF itself is following up on things that’re being generated elsewhere. There are examples of both directed follow-up, as well as community volunteer. .. who’s got a telescope free that you can look in this part of the sky? I do. I’m gonna go try and get it.”

“We talked about a possible role for amateur astronomers, and one idea was ... to parcel out observing assignments, work units of observing ... these people are not gonna publish the results, and we’d have to figure out how to use the results ... but it could potentially improve what we know about the sources.”

2.3.2. Joint Observations

Question from the Survey: *For Phase 1, the LSC seeks partners to carry out broad joint observations. Would you be interested in taking part in these campaigns and what contribution would you be willing to commit?*

“Yes, we could provide telescope time for a limited number of well defined sky locations.”

“Yes, the PTF-2 project is interested. Details to be negotiated.”

“Yes. We are building a telescope for 'Astroparticle physics', that is : GRBs - neutrinos - GWs - VHE gamma-rays. Interaction with GW community is welcomed.”

“Would be interested in coordinating ground-based followup, particularly at the level of developing new instrumentation. This will likely feed back onto observing strategies. “

“Absolutely. We have substantial optical and IR telescope resources that can be brought to bear on this problem. “

“Yes, I would be interested to help generally with theoretical interpretation of EM counterparts; or in the near term, e.g. setting up plans for observing trees based on the expectations of predicted counterparts (e.g. if X observed [or not] then trigger search Y, etc.)“

“I'd love to help out, insofar as a theorist can help. I'm involved in LSST, and may be able to help on that end. “

“MWA (wide angle radio observations) is interested in partnering with LIGO. I am involved in proposals for wide-angle X-ray monitoring that would also seek LIGO partnership, if selected.”

“The detailed plan of data release and overall coordination of joint observations should be developed with the goal of optimizing the scientific output. Realistic simulations of the joint searches are needed to determine optimal observation strategy and data analysis requirements. Special effort should be dedicated to characterizing the background of accidental transients (those not related to GW) for the participating telescopes.”

2.4. Feedback for Follow-up Observations

A topic of discussion at the workshop was the possibility that observers, following up on a LIGO or LIGO-Virgo trigger, would post their results to a forum in real time, with the idea of increasing the science result from the shared observations.

“It was in the original ... scope of the GCN circulars that people could publish [intention]: ‘I’ve got Hubble time in six hours and I’m going to go look at it,’ ... out of the 12,000 circulars that have been published in 15 years or so I think there’s been less than 10 like that.”

“You wouldn’t want to control observing based on what people claim they’re going to observe, because somebody will claim, ‘I’m about do the high-probability [island],’ and they will attempt to do it but they may fail for weather or technical reasons or other reasons ... you could imagine a perverse situation in which someone jumps up and says, ‘I’m gonna get the high-probability region first,’ but they may fail for technical reasons, and so then everyone avoids that high-probability region.”

2.5. Trigger Repository

Each pipeline run on the data stream produces a stream of events, of what it has “found”. By adjusting the threshold significance, the stream can be reduced to any required rate. This

section considers how a repository of triggers, going down to ‘sub-threshold’ (the frequent, low significance triggers that are most often instrument noise), could be used for science.

“It seems like you should be releasing events that are below what would be publication threshold based on the detection alone. I think there are a lot of astronomers that are willing to spend ... their time going after LIGO error boxes, and I certainly think we should enable them to do that if they desire.”

“There will be a set of sub-threshold gravitational-wave candidates, ... the idea was discussed that we should ... expose those to people (via some relationship to be defined) ... for cross-correlation studies with other instruments and for large-scale data studies.”

“Should the access to the sub-threshold triggers require joining the collaboration, even after some proprietary period, and if so, what are the responsibilities and publication policies?”

“We talked about potential new scientific tools ... some type of online searchable database ... if someone ... discovers some strange object, and they don’t wanna go through this whole process of joining the collaboration ... would like to search and just ask, ‘Are there any even sub-threshold events within a given time window or within a given area on the sky?’ and then they should be able to search that relatively quickly. What we were not agreed upon was whether this type of thing should be available to everybody after a proprietary period, so anyone could kind of dig into this, whether just to collaboration members.”

2.6. Very Low Latency Alerts

In this section, the broader community comments on the extent to which LIGO should strive to achieve event alerts at the sub-minute latency.

Question from the Survey: *Significant effort will be required to realize our goal of transient alerts within ~60 sec of data acquisition. Are there any reasons to reduce this latency further?*

“We are working hard to build a 1-meter robotic telescope for astroparticle physics, with a time of reaction ~30 seconds, so 1 minute is OK for me. The main problem is the size of the error boxes. Our telescope will have a FOV which is a significant fraction of 1 square degree, and can observe about 60 sq. deg. per hour, which is OK for GW candidate follow-up if the error boxes are not too large...”

“Yes. Coherent emission (radio) may last less time than 60 sec. And we don't know how bright at optical wavebands an EM candidate will be at early times, but we do know widefield imagers can respond in seconds.”

“Not at this time. However, early results of finding transient events in LIGO together with EM detection may alter these needs. An improvement from 60 seconds to 50 seconds isn't meaningful. An improvement from 60 seconds to 10 seconds could be meaningful. I think we just don't know yet.”

“Short GRBs are sometimes accompanied by extended high energy emission lasting ~ 100 s (its origin unknown). If short GRBs are in fact associated with NS mergers, it might be of use to be able to trigger and slew Swift BAT on a short timescale. Though highly uncertain, there is also a possibility that NS mergers may be associated by prompt radio emission detectable by wide-field instruments such as LOFAR. To the best of my understanding, LOFAR can only hold the full data from its observations for seconds (currently, but possibly longer in the next several years), prompt alerts would be useful in this case as well.”

“For some gamma-ray bursts, prompt optical flashes have been seen as soon as 20-30 seconds after the burst itself. If robotic telescopes such as ROTSE were notified, they could

slew within seconds. Thus there is reason to reduce the latency further, which must of course be balanced against the difficulty of doing so.”

“No, not from the PTF-2 perspective.”

“All sky-monitors can always be correlated after the fact, so the latency limits just pointing of other telescopes (optical, radio, space). The slew times for these are typically 15-60 sec or more. So I would say there is not much point in pushing the latency shorter than the slew times. The one exception would be a search for a burst of coherent radio emission using electronically steered phased arrays like LOFAR. Currently LOFAR stores the data for the whole sky in a buffer for 1 second, and dumps it if no trigger is sent. Their goal is to follow Moore’s law for the next few years buying ever cheaper buffer memory to be able to store for 30 sec in several years. Ideally LIGO’s latency would intersect LOFARs buffer duration when advanced LIGO turns on.”

“I think for this kind of events it is important to be observed optically as fast as possible. They may be very short and decaying quickly. Quickest mounts can reach a given position within ~10-20 seconds thus in order to catch event in the most interesting phase it is extremely important to have alerts as fast as possible. Therefore it is always important to release as fast as possible and every further improvement would still be good.”

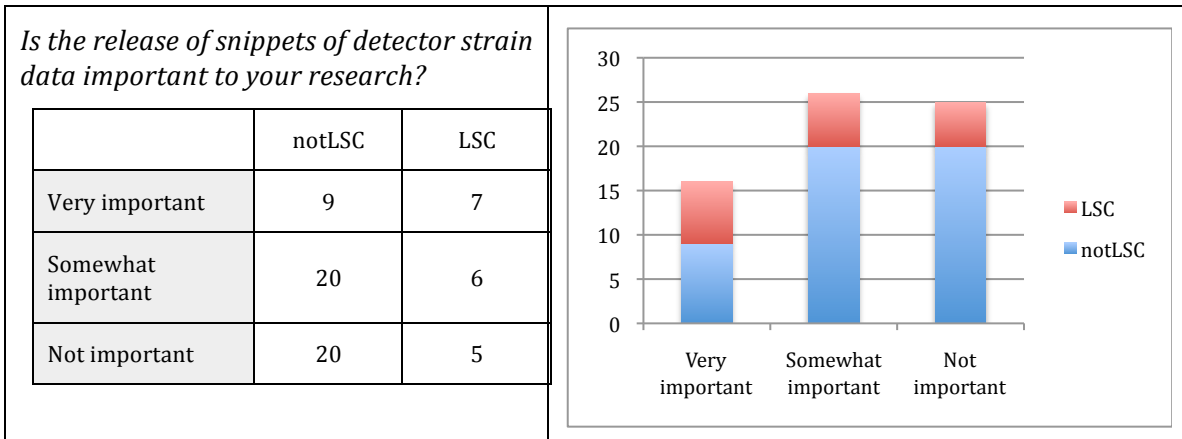
“There is some hints that the EM signal may arrive as soon as 20sec after the GW. Since it is likely a short burst, with dim afterglow, then reducing the delay is of importance. Of course, a 60sec. delay will be already an important achievement.”

“It depends on which radio telescopes have a buffer and how long it may last. For LOFAR, 60sec would be too long.”

3. Source Catalog

This section considers how to build a GW source catalog, once detections are plentiful. What data should be in the catalog entry, who wants the data, and how can it interoperate with other astronomical datasets?

Question from the Survey:



“A gravitational-wave catalog ... is an object catalog of what we as a collaboration believe are gravitational-wave signals. A list of what we thought the various contents would be ... is the event time, the position on the sky, the various parameters ... the strain, either time or frequency, just some cleaning of the noise out ... some noise-power spectrum from around the

time of the event, maybe some audio files. Perhaps also the underlying strain signal, de-noised strain in time or frequency.”

Question from the Survey: *Best estimates of physical parameters, sky location probability maps, summary spectra and snippets of strain data will be released as part of the gravitational-wave source detections. What other information would you like to receive in these catalogs?*

“Statistics of events just below the significance cutoff, for independent assessment of the probability the event is real...”

“Some data quality estimate, perhaps, e.g., an estimate of the detector noise spectrum during that period?”

“Would be good to have secondary possible metrics on performance: e.g., instrument uptime prior to trigger, anything regarding status, etc.”

“Visual representation of the data in time-frequency plane, eg., as produced by Omega pipeline etc.”

“Amplitudes and orientations of different modes, including dipole, and corresponding electromagnetic observations.”

“Different types of significance (using different automated programs)?”

“Possible nearest location in optical catalog or Radio source nearest to above parameter. With standard reference catalog.”

“Estimated distance from the source in the case of coalescing binaries, making some prior assumption about the masses involved, would be useful in reducing the redshift phase required for searches.”

“Timing information: when detected, duration of signal, periodicities, chirp, etc. If the periodicity information is part of the "summary spectra", then perhaps some of this is not needed. SNR/likelihood information: from all detectors. Correlation information: when detected where.”

4. Full data

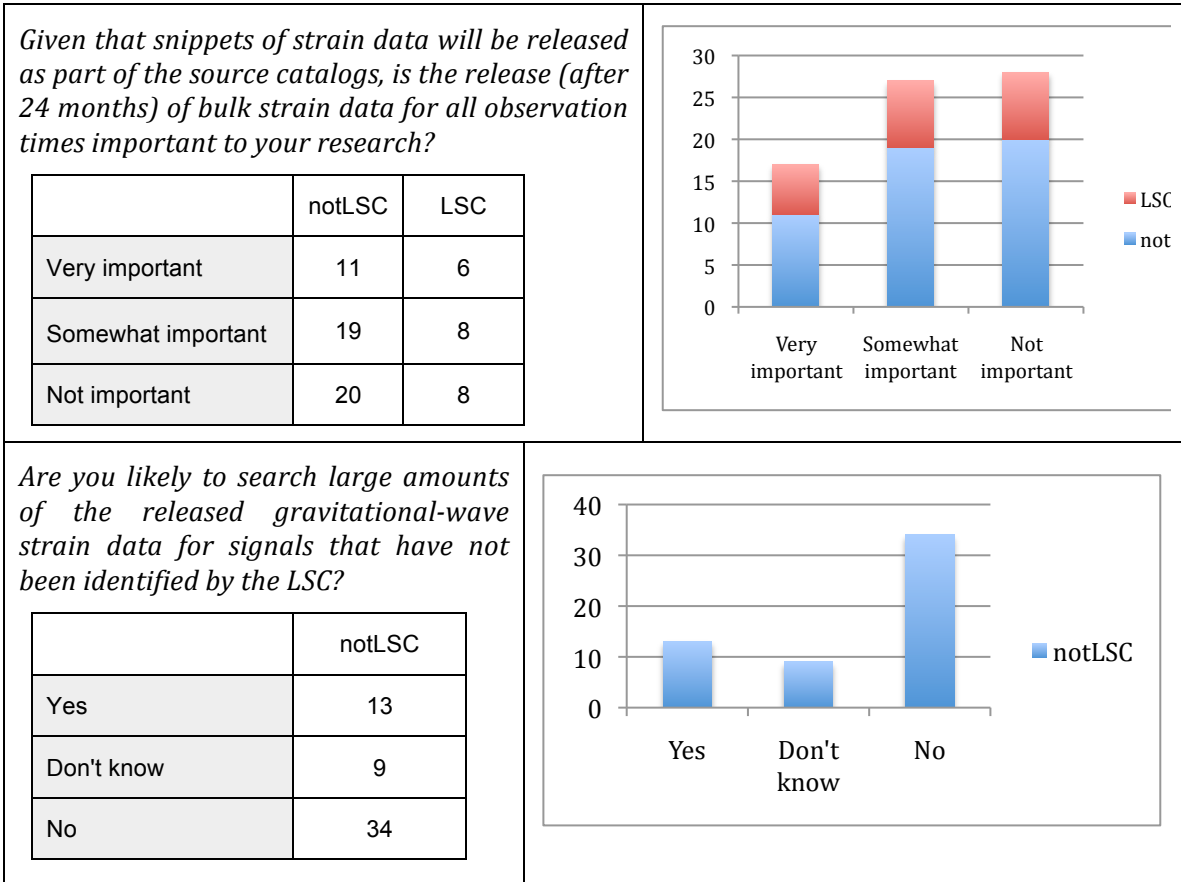
Once LIGO declares the start of open data phase 2, it will be another two years, according to the Data Management Plan [1], before the full $h(t)$ strain channel is released. Here we consider how to make that most useful, how to include data quality information, who might be most interested in that data release.

“If we put out the strain data, then how to represent the data quality? We discussed that: the basic idea is to have data-quality bit streams. There might be more than one, they might be classified by search type and how to talk about data quality for the different search types, and these would be summaries. These would be bit-stream summaries of the data quality that gets used by the collaboration to identify glitches, vetoes, all these things.”

“We did have a brief discussion about the other channels that’re recorded. Half a petabyte per year is a lot of data, a lot of auxiliary channels. And the suggestion that’s floated ... is to make sure to document the instrumentation at each site so that it’s understood what is available ... how many magnetometers, what types they are, etc, and to then make sure that it’s possible for people to say, ‘Hey, I’ve got this cool thing I wanna do with your magnetometer data.’ “

“It is very difficult for me to predict how interested I will be in searching the raw data. That depends on how easy it is, how many sources are detected, whether it is possible to form small collaborations with LSC members (who know the data formats and data peculiarities) to quickly develop new ideas and then search the data for them. Nevertheless, scientific data is best open, and as such I would encourage LIGO to make the entire strain-data available, even if currently there is only small interest from the community.”

Question from the Survey:



5. Building the LIGO Open Science Center

To provide open data access to the broader community, LIGO expects to build an ‘open science center’. This section has comments and advice about that.

“The most important questions IRSA users want answered are: ‘how do I get (and use) my data?’ and ‘whom do I ask if I have a problem?’”

“Openly releasing your data maximizes the scientific potential of the whole experiment. Do not underestimate the personnel and person-hours that it will require.”

“It’s really hard to create a good database. It’s even harder when you are dealing with people who like to continuously monkey with their data and (to be fair) whose data is continuously updated. It’s doubly harder when you are dealing with people who don’t really want to be bothered with the database because it distracts them from their ‘real science.’”

“Trends: Archives as analysis environments. (1) Browse-and-download: identify data of interest, take it home for further study (2) Complex queries, finding “interesting” data within large datasets (3) In-database analysis, more complex queries over larger data volumes, where data size is growing faster than communications, so it is best to bring the software to the data.”

“The Fermi mission suffered (still suffers?) from improper division of responsibility, where the instrument team is responsible for tool development, but the science center team is responsible for dealing with user issues.”

5.1. Documentation and Curation

“You’ll see this theme quite a bit a few places here, documentation, documentation, documentation. You must tell people what this stuff really means. This will be new stuff for the gamma-ray burst and the general transient community; they [need to understand] how far you can take them, how soft, how competent, confidence level, things like that. So we need to put some effort into the big picture, the context, how these things are derived, so that people making these follow-ups can effectively control their robotic telescopes.”

“Maybe that’s already planned but it would be good to have a designated website both for the open data as well as for the plans leading to open data so one can follow the development.”

“I just want to add a word of caution. It’s not so clear that all these scientists who spent years developing the analysis tools will be so greatly enthusiastic about now documenting the tools so that everybody else can use them before they’ve had a chance to exploit them.”

“Should also be easy for HS and college undergrads to pursue simple science fair or research projects with the data. Should be possible to upload results into forums and electronic lab notebooks and journals. Should have a lot of documentation and background on GW physics, astrophysics, detectors, etc, with lots of links to papers, resources, and cool graphics.”

“On curation, let me just jump down to the bottom here. In the NASA astrophysics community we have HEASARC and IPAC, who do that for us, independent, standalone, self-funded. I don’t think I heard anything that you guys have such a thing, so you might want to lobby NSF for creating one for you, or you could do it yourself, but think in two decades plus time scale.”

5.2. Mock Data Challenges

“Give them the cookbooks with the example data sets, that’s the best thing. They can go there and they can practice ahead of time. What might be practice data? Well, probably the big dog event is a good one. “

“[Include] some noise events in there; one of each kind or what constitutes the bulk majority of them or something like that. Then they can practice on those and ... recognize what they are.”

“Providing complete simulated data sets from expected events would very much help in prepping the EM community for follow-up efforts.”

5.3. Software

“Software doesn’t write itself. Make sure you have enough dollars and people to do it, dedicated people that’re not wrapped up at the daily operations and getting distracted. It has to be more robust than what you guys are used to, because people will apply it to a given data set and it will blow up in their face, and you guys don’t encounter that problem, because you know

that that tool is not applicable to that kind of data set; you just know it from years of experience. It has to be robust to handle all those kind of gotchas.”

“A suggestion here: a large community of these follow-up people are going to be coming from the astrophysics community, hardcore X-ray, gamma ray, optical [community]. They’re used to HEASoft and FTOOLS, and if you produce something that’s entirely alone internal to LIGO I’m sure it’s wonderful, but there’s a whole architecture, philosophy, design approach ... you might take advantage of that.”

“Cloud computing environments are going to be key in the future - so many users may not even download their data, just run their analysis remotely on some virtual machine which has all the LIGO software prepackaged.”

“There wasn’t any need expressed for software *per se*, just information, and maybe you need software to decode the information and maybe not.”

“End-to-end tutorials that lead you through what you would do, how you would see the data, what the transformations of the data do. A step-by-step walkthrough with some very simple example codes, something that reads in the time series, that whitens it, cleans it, calculates spectra, filters it and tells you what the outputs should be at each step and what’s been done.”

“It doesn’t matter what the data format is, but provide a library that works, that is documented that has some examples about how to use it to read and write, and that’s the thing that people care about. It goes back to not wanting to know what’s inside the sausage; just [make] a way to get at it.”

“The one thing that I’ve heard from numerous people ... is that we should settle on the format, and release that format with tools built around it, so that we don’t establish an expectation and then later say, ‘No. By the way, the release format’s really different ... you’ve got the simulated one”

“Many of the LSC codes are actually already released under GPL. They’re released at regular intervals, so they’re already out there, not great documentation. You really do need to get an expert to run them, but they’re there.”

Question from the Survey: *Given that data will be delivered primarily via web services, what software would you like us to provide?*

Here are the numbers of people who checked each category:

- Simple examples of command-line clients to access the web services: 45
- I/O libraries that can read the data returned by the services: 38
- Sample applications to read, filter, and plot the data: 47
- More advanced software, for conditioning, searching, correlating, etc: 17
- Other: 16

6. Appendix

6.1. Gravitational Wave Open Data Workshop

The workshop program, with all the slides in pdf, is available at: <http://www.ligo.caltech.edu/gwodw2011/program.shtml>

Thursday Oct 27

- 9.20 - 10.00 Christian Ott and Roy Williams: *Introduction*
- 10.00 - 10.20 Gaby Gonzalez: *What is LIGO*
- 10.20 - 10.40 Patrick Brady: *What is LIGO data*
- 10.40 - 11.00 Cole Miller: *Connect LIGO to astrophysics*
- 11.20 - 11.40 Neil Gehrels: *Connect LIGO to observational astronomy*
- 11.40 - 11.55 Roy Williams: *LIGO Open Data plans*
- 11.55 - 12.10 Benoit Mours: *Virgo Open Data plans*
- 12.10 - 12.20 Andrea Lommen: *Open data and pulsar timing*
- 12.20 - 12.40 Roy Williams: *Charge to the breakout groups*
- 1.20 - 1.40 Alicia Soderberg: *Science opportunities*
- 1.40 - 2.00 Brian Metzger: *Electromagnetic Counterparts of Neutron Star Binary Mergers and their Detection in the Era of Advanced LIGO*
- 2.00 - 3.00 Breakout groups convene
- 3.20 - 3.40 Tom Stephens: *Experiences from the Fermi Data Archive*
- 3.40 - 4.00 Steve Groom: *IPAC architecture for open data*
- 4.00 - 5.00 Breakout groups convene
- 5.00 - 5.20 Wen-fai Fong: *Short Gamma-Ray Bursts and their Progenitors*
- 5.20 - 5.40 Daniel Holz: *Science opportunities*

Friday Oct 28

- 9.20 Reports from breakout leaders:
 - A: **Information Technology** (presenter: Scott Barthelmy)
 - B: **Smoking Gun** (presenter: Peter Shawhan)
 - C: **Full Data** (presenter: Patrick Brady)
 - D: **New Searches** (presenter: Cole Miller)
- 11.40 - 12.00 Geoffrey Bower: *Radio Surveys*
- 12.40 - 1.00 Dave Reitze: *Towards Advanced LIGO*
- 1.00 - 2.20 Background on aLIGO and tour of work
- 2.20 End of workshop

6.2. Open Data Survey text

The survey is a reference [3]. The following is a list of the questions, with tabular answers where appropriate. Written responses have been used to build the rest of this report.

The LIGO Data Management Plan [1] defines plans for data release by the LIGO Scientific Collaboration (LSC) in the advanced detector era. In the discovery phase (phase 1) these plans foresee limited data releases around detections and around significant non-detections, for example in conjunction with interesting electromagnetic/neutrino observations. In the observational phase (phase 2) a much broader data release is planned, including near-real-time alerts. Furthermore, 24 months after data taking, all LIGO data will become publicly accessible.

The LIGO Data Analysis Council would appreciate your feedback on the following questions, to enable planning for phase 2. Anonymous submission of this form is welcome, however, we will weigh responses with contact information more heavily in our considerations. You need not answer all the questions. You may make just one response to one question. Your input and thought is appreciated.

Your name / Your Email

1. What is your background in gravitational-wave related astronomy?

Member of LIGO Scientific Collaboration	22 28%
Member of other gravitational-wave detection effort	16 20%
Numerical Relativist	16 20%
Observational Astronomer	27 34%
Theorist	24 30%
Other	12 15%

2. The phase 2 plan includes low-latency (seconds, minutes) release of transient triggers including time, sky location probability maps, and false alarm rate estimates. Is this important to the science you want to pursue?

	nonLSC	LSC
Very important	39	18
Somewhat	8	3
Not important	11	1

3. If you are interested in transient alerts in phase 2, do you want all of them (with significance estimate), or do you want only those of highest significance?

An alert each 24 hours but probably it is noise	27 34%
Once a month with good chance that its a real GW detection	19 24%
Inform me only when it is very likely to be a genuine detection	18 23%
Other	16 20%

4. Significant effort will be required to realize our goal of transient alerts within ~60 sec of data acquisition. Are there any reasons to reduce this latency further?

5. If you are interested in transient alerts in phase 2, what level of review would you want?

As fast as possible, machine generated, with no human review, may be detector noise	29 36%
One hour later when humans have reviewed the information and confirmed that everything is operating normally	40 50%

6. Is the release of snippets of detector strain data important to your research?

	notLSC	LSC
Very important	9	7

Somewhat important	20	6
Not important	20	5

7. Best estimates of physical parameters, sky location probability maps, summary spectra and snippets of strain data will be released as part of the gravitational-wave source detections. What other information would you like to receive in these catalogs?

8. Given that snippets of strain data will be released as part of the source catalogs, is the release (after 24 months) of bulk strain data for all observation times important to your research?

	notLSC	LSC
Very important	11	6
Somewhat important	19	8
Not important	20	8

9. Are you likely to search large amounts of the released gravitational-wave strain data for signals that have not been identified by the LSC?

	notLSC
Yes	13
Don't know	9
No	34

10. What data formats would be most convenient for you for strain data, event notification, query responses, etc??

The simplest ASCII text	47	73%
Frame files	10	16%
VOTable	11	17%
VOEvent	16	25%
HDF5	12	19%
SQLite file	5	8%
FITS images	16	25%
Other	7	11%

People may select more than one checkbox, so percentages may add up to more than 100%.

11. Given that data will be delivered primarily via web services, what software would you like us to provide?

Simple examples of command-line clients to access the web services	45	66%
I/O libraries that can read the data returned by the services	38	56%
Sample applications to read, filter, and plot the data	47	69%
More advanced software, for conditioning, searching, correlating, etc	17	25%

Other **16** 24%

People may select more than one checkbox, so percentages may add up to more than 100%.

12. Would your research benefit from integration of LIGO open data with the Virtual Observatory?

Yes	27	34%
No	9	11%
Don't know	33	41%
Other	11	14%

13. For Phase 1, the LSC seeks partners to carry out broad joint observations. Would you be interested in taking part in these campaigns and what contribution would you be willing to commit?

14. Please provide further comments and suggestions related to the release of LIGO data.

7. References

[1] The LIGO Data Management Plan

<https://dcc.ligo.org/cgi-bin/DocDB/RetrieveFile?docid=9967>

[2] Gravitational Wave Open Data Workshop

<http://www.ligo.caltech.edu/gwodw2011/program.shtml>

[3] LIGO Open Data Questionnaire

<https://docs.google.com/spreadsheets/viewform?formkey=dFJuS1d4UINiLW9Wdm5OQTIKWFE4ekE6MQ>

[4] Example of the data products that the LSC and Virgo might release for their first gravitational wave transient detections. This particular event was not a real detection; it was a "blind injection". <http://www.ligo.org/science/GW100916/>

[5] "Providing Open Access of LIGO Data to the Broader Research Community"

<https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=100>