GWOSC: Gravitational Wave Open Science Center

Agata Trovato[*] **for LIGO The LIGO Scientific Collaboration and Virgo Collaboration**
APC, Université Paris Diderot, CNRS/IN2P3, CEA/IRFU, Observatoire de Paris, Sorbonne
Paris Cité, Paris 75205, France
E-mail: trovato@apc.in2p3.fr

The LIGO Scientific Collaboration and Virgo Collaboration actively pursue an open science programme. Gravitational wave strain data are publicly released through the website http://www.gw-openscience.org along with the information and tools necessary to understand and use the data. This contribution will give an overview of past and current activities related to this topic, including the recent release of the Catalog of Compact Binary Mergers (GWTC-1) and will give a sense of its impact on science, teaching and other applications.

The New Era of Multi-Messenger Astrophysics - Asterics2019
25 - 29 March, 2019
Groningen, The Netherlands

---

[*]Speaker.

1. Introduction

The Advanced LIGO detectors [1] began scientific operations in 2015 with the first observing run (O1) that took place from the 12th September 2015 to the 19th January 2016. After an upgrade of the detectors, the second observing run (O2) took place from the 30th November 2016 to the 25th August 2017. On the 1st of August 2017 Advanced Virgo [2] joined the observing run. The O1 and O2 bulk data were released respectively on January 2018 and on February 2019 (2 years and 1.5 year after the end of the runs). Short segments of data surrounding published event discoveries were released just after the publication of the corresponding paper. At the moment there are 11 confident detections (GW150914, GW151012, GW151226, GW170104, GW170608, GW170729, GW170809, GW170814, GW170817, GW170818, GW170823) and 14 marginal triggers, all detailed in the gravitational wave transient catalog (GWTC-1) [3].

Access to these data is provided by the Gravitational Wave Open Science Center (GWOSC) [4] in the website *gw-openscience.org* along with documentation, tutorials, and online tools for finding and viewing data. GWOSC reflects the growing awareness for the importance of open data in science. Sharing the data allows reproducibility of scientific results and discoveries made from the data. It maximises the experiment's discovery potential by allowing anyone to perform their own analysis. In the present case, it also enables multi-messenger astronomy, i.e., joint analysis with other astronomical observation channels. Finally, it is possible to involve a wider community, not only professional scientists but also students, amateur astronomers and artists. Many projects[1] have already been possible through the LIGO and Virgo public data release, including about 60 scientific papers so far and also student projects, online courses, an iPhone App and art installations.

The third observing run (O3) is ongoing since April 2019 and it will last roughly one year. The current plans for the bulk O3 data release foresee a release of the first block of six months of data in April 2021 (with a latency of 18 months from the end of acquisition of that observing block), followed by the release of the second observing block after another 6 months.

This paper provides some insights on the way the data are collected and are repackaged for the public release in the section 2. The files' format and their content is described in section 3, while section 4 provides some recommendations and suggestions for the use of the data.

2. Methods

The Laser Interferometer Gravitational-Wave Observatory (LIGO) includes two identical 4-km arm interferometers situated in Hanford in Washington State and Livingston in Louisiana. The Advanced Virgo detector is an interferometer with 3 km arms located in Italy, near Pisa.

Those detectors measure the strain that gravitational waves (GW) exert on space-time. The GW strain is the main science output of the detectors and it is measured through the difference in phase of the light travelling in the arms of the interferometers. The data are repackaged and resampled by GWOSC to make it more accessible to casual users both within the LIGO Scientific Collaboration and Virgo Collaboration and outside. More specifically the strain channel is sampled in LIGO and Virgo data at 16384 Hz (2^{14} Hz), while GWOSC data are made available both at 4096 and 16384 Hz. Events releases consist of data segments centred at the time of each event provided

---

[1]https://www.gw-openscience.org/projects

in two durations: 32 s and 4096 s. Strain data taken during a whole observation run are divided into many files of length 4096 s.

Additional data are essential to qualify the reliability of the strain measurement, including the status of the detectors, the sub-systems, and the environment, all of which are continuously monitored with hundreds of thousands of auxiliary channels. From these auxiliary channels a handful of "Data Quality" channels are extracted and released in addition to the strain by the GWOSC.

Further technical details are given in the GWOSC webpages [2]. For each run GWOSC also provides a representative power spectral density (PSD) of the background noise from the entire frequency range, along with a list of spectral lines obtained from a highly-resolved PSD estimate.

3. Data Records

Gravitational wave open data are hosted in the GWOSC webpages [3]. The files can be directly downloaded one by one from this webpage. However, to download large amounts of data (as in the case of a whole observing run) it is suggested to use the distributed filesystem CernVM-FS that, once installed, allows the users to access GWOSC data as files in a directory tree mounted on their local computer.

GWOSC data are available in 3 files format: "GWF", which is a specialized format of GW detectors; "HDF5" that is readable with many coding languages (python, MATLAB, C/C++, IDL,...); simple ASCII file format containing two columns, one with the time and the other with the corresponding strain values (only for the events release).

The HDF5 files contain:

- Metadata: Description of the data, URL of the GWOSC website, detector and observatory concerned, duration of the segment of data, starting time both in GPS and UTC.

- Strain: main "data" of the interferometer, sampled at 4 or 16 kHz depending on the file and accompanied by some attributes such as the starting GPS time, the spacing in time between the points in the time series and the number of points.

- Quality: 1 Hz time series that summarize the data cuts and vetoes made by LIGO and Virgo scientists who identify and flag detector misbehavior that can affect the searches for GW signals in those data. There are also 1 Hz time series that flags the simulated GW injections that were inserted in the data to test search pipelines and characterize their performance.

4. Usage Notes

Gravitational wave detectors are complex instruments, and their data reflect this complexity. Precaution needs to be taken when searching for GW signals in the LIGO and Virgo data. The background noise is not stationary and evolves over multiple time scales. The detectors are not always in science mode and, even when they are, it is possible that data quality does not meet the

---

[2]gw-openscience.org/O1 and gw-openscience.org/O2 respectively for O1 and O2. The current and past detector status is given in gw-openscience.org/detector_status

[3]gw-openscience.org/data

requirements of a given analyses. For these reasons it is necessary to work with segments of data. On the GWOSC website a tool called *Timeline*[4] provides a visual representation of the duty cycle of each data-quality (or injection) bit across a prescribed time interval. From the timeline page it is also possible to select specific segments and download the corresponding data.

The data are not calibrated in the extreme parts of the observed frequency range (below 10 Hz and above 5 kHz). In most searches for astrophysical sources, data below 20 Hz are not used because the noise is too high. Moreover, the data quality and calibration are less well studied above 2 kHz.

Simulated signals, called hardware injections, are also present in the Advanced LIGO data. Their times can be retrieved in the GWOSC files by the injection data quality information.

The GWOSC website contains many tutorials[5] that can be used to understand how to read and analyze the data. To ease these analyses, the LIGO Scientific Collaboration and Virgo Collaboration have developed various software packages[6], mainly python based. The tutorials help also in using these packages, especially the on-line courses obtained from recordings of the two Open Data Workshops[5] that the GWOSC team have organized in 2018 and 2019.

5. Conclusion

The basic features of the Virgo and LIGO open science program have been presented. This model is quite successful and, for example, it has been cited in a recent issue of the CERN courier about open science [5].

6. Acknowledgement

References

[1] B. P. Abbott et al. (LIGO and Virgo Collaborations) Phys. Rev. Lett. 116 (2016) 131103.

[2] F. Arcenese et al. (Virgo Collaboration) Class. Quant. Grav., 32 (2015) 024001.

[3] B. P. Abbott et al. (LIGO and Virgo Collaborations) Submitted to PRX [arXiv:1811.12907].

[4] M. Vallisneri et al. , proceedings of the 10th LISA Symposium, University of Florida, Gainesville, May 18-23, 2014 [arXiv:1410.4839].

[5] CERN COURIER, Volume 59, Issue 2, Match/April 2019.

---

[4]gw-openscience.org/timeline

[5]gw-openscience.org/tutorials

[6]gw-openscience.org/software