

DQSEGDB: A time-interval database for storing gravitational wave observatory metadata

R. P. Fisher¹, G. Hemming², M. A. Bizouard³, D. A. Brown⁴,
P. F. Couvares⁵, F. Robinet⁶, D. Verkindt⁷

¹*CNU, 1 Avenue of the Arts, Newport News, VA 23606, USA*, ²*European Gravitational Observatory (EGO), I-56021 Cascina, Pisa, Italy*, ³*Artemis, Université Côte d'Azur, Observatoire Côte d'Azur, CNRS, CS 34229, F-06304 Nice Cedex 4, France*, ⁴*Syracuse University, Syracuse, NY 13244, USA*, ⁵*LIGO, California Institute of Technology, Pasadena, CA 91125, USA*, ⁶*LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, F-91898 Orsay, France*, ⁷*Laboratoire d'Annecy de Physique des Particules (LAPP), Univ. Grenoble Alpes, Université Savoie Mont Blanc, CNRS/IN2P3, F-74941 Annecy, France*

Abstract

The Data Quality Segment Database (DQSEGDB) software is a database service, backend API, frontend graphical web interface, and client package used by the Laser Interferometer Gravitational-Wave Observatory (LIGO), Virgo, GEO600 and the Kamioka Gravitational wave detector for storing and accessing metadata describing the status of their detectors. The DQSEGDB has been used in the analysis of all published detections of gravitational waves in the advanced detector era. The DQSEGDB currently stores roughly 600 million metadata entries and responds to roughly 600,000 queries per day with an average response time of 0.223 ms.

Keywords: database 1, metadata 2, time segments 3, gravitational waves 4, LIGO-Virgo 5

1. Motivation and significance

Gravitational Waves are disturbances in the metric of space-time that propagate through the universe and carry information about the astrophysics of sources that generate them. For current detectors, these sources must be massive objects moving with high accelerations [1]. Although GWs may have very large amplitudes at their origin, they typically travel extra-galactic distances to reach the Earth. Because gravity couples weakly to matter, GWs are difficult to generate and detect. When these waves reach the Earth, their strength is such that their resulting spacetime perturbation changes

measurements of length by 1 part in 10^{20} , equivalent to changing a distance of 1 AU by the width of 1 atom. This results in an extremely small signal even if detected with kilometer-scale detectors [2]. The mission of the LIGO Scientific Collaboration (LSC) and the Virgo Collaboration (LVC) is to detect these weak signals in order to advance our understanding of the universe. Since 2015, GW detections have shown that Einstein’s theory of general relativity holds for colliding black holes and neutron stars [3, 4, 5]. These discoveries allow us to estimate the number of binary black hole (BBH) and binary neutron star (BNS) systems in our local universe [6, 7, 8], and have demonstrated that some gamma-ray burst (GRB) events are powered by the coalescence of neutron stars [4, 9].

The detection of these GWs has been made possible through the colossal effort of thousands of scientists to develop a tremendously precise set of interferometers (IFOs) and an ecosystem of computational infrastructure that enables the capture and analysis of the data generated by these instruments. The Data Quality Segment Database (DQSEGDB) occupies one critical space in this infrastructure. The Data Analysis (DA) algorithms require information about the state of the IFOs to analyze the observatory data. This requires the definition and distribution of metadata about the observatory data, which we call Data Quality (DQ) flags.

A DQ flag is the name given to a set of metadata that describes some portion of the global status of the detector, operation of the instrument, or quality of the data that may impact its analysis. These flags are critical to the data analyses because a category of them mark the times when the IFOs are operating in an optimal state, thereby indicating which observatory data should be analyzed. Additional flags indicate data that should explicitly not be analyzed, such as when hardware injections are ongoing or when electronics faults cause noise in the GW detection channel. These DQ flags are also called DQ vetoes because they can be used to exclude data from being analyzed [10].

The set of data associated with each flag name is the list of times when the state of that flag was known and the list of times when that state was active or inactive, which are compliments within the set of known times. The time periods are contained in a data product known as “segments”, where a segment is a continuous range of time expressed as a half-open GPS time interval [tstart,tend). Within the GW community, the terms DQ segments and DQ flags are often used interchangeably because of this tight relationship. Each flag has a unique name. The flag names are associated with their IFO identifiers, and are combined in the format [IFO]:[FLAG-NAME].

1.1. Initial Detector Databases

The DQSEGDB service and client software were built to replace the aging predecessor services that served the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo collaborations separately during their initial observing runs. These previous services were each able to store hundreds of flags and approximately 2 million individual DQ segment and metadata entries by the end of the final science runs of initial LIGO and Virgo. The LIGO service had become very slow, and would often take 10 to 30 minutes to respond to queries made by GW data analyses. This severely restricted the usability of the service and indicated a strong need for a replacement for advanced GW detectors, where the number of flags and number of segments the databases would need to store would grow by factors of hundreds. An increasing number of new software systems were also unable to use the DQ metadata effectively due to the slow response times of the server. Finally, new user requirements pushed for a redesign of the API and database schema. These issues led to the LVC making the decision to pool their resources and to design a new segment database infrastructure. This led to the development of the DQSEGDB software.

2. LVC Data Landscape and Terminology

Each IFO produces one primary data channel, which contains the measurement of the GW strain, and approximately 200,000 channels of auxiliary data that are used to monitor the status of all the hardware and software components used to produce the primary data. This data set constitutes approximately 2 TB per day per IFO. Customized scripts are used to reduce this huge amount of auxiliary data into approximately 1000 DQ flags per IFO. At the location of the IFOs, a set of real-time processes automatically generate segments for a portion of the total DQ flags. These processes encode the metadata in XML files that each contain information about the status of these flags for 16 seconds of data. Each of these XML files is about 78 kB in size, which translates to roughly 420 MB of metadata generated per day per IFO. This XML is then transferred via rsync from each IFO to the DQSEGDB server, which is hosted at the LIGO Laboratory at California Institute of Technology. The DQSEGDB server then executes all of the code needed to extract the metadata from the XML files, publishes it to the database, and archives the raw XML files.

3. Software Development and Description

3.1. Software Design and Architecture

In addition to the requirements that the DQSEGDB service be able to respond rapidly while storing a large amount of metadata, several other design requirements and elements of design philosophy were also met when the new software was written. The database was required to contain both the DQ segments and enough additional metadata to allow the tracking of their provenance. The service was required to allow remote clients to connect via command line or web GUI, and was to provide the metadata within 15 minutes of its generation. The API was chosen to provide a RESTful set of URIs with a Resource-Oriented Architecture, compatible with multiple programming languages, and restrictive such that data could not be removed from the database. A JSON format was chosen for the returned data, which included an option for all provenance metadata. Additional functionalities were deferred to the client layer to ensure speed at the server.

3.2. Software Functionalities

These design requirements led to the current DQSEGDB software design. The service as a whole is split into three major components. The first is the primary database server, which is generally labelled the DQSEGDB. The second is the client software package, which contains both command line tools and a Python package that can be used to query the database. This set of tools provides many functions as requested by LVC scientists, while also satisfying the design requirements listed. The final component is a graphical web interface to the database, which provides a GUI interface that allows collaboration scientists to rapidly access the metadata without needing to write any code.

The DQSEGDB server consists of an Apache layer that calls a custom Python application via the Apache WSGI module. The Python application uses ODBC to communicate with a MariaDB database instance that hosts a dedicated, normalised database schema that has been designed to optimally host the DQ flags, their associated segments, the associated metadata about those segments, and some overall metadata about the data. The API provides access to all data associated with a given DQ flag through RESTful URIs, formatted as `/dq/IFO/FLAG/VERSION`. Within this URI, the data can be downselected based on the information and time interval of interest, using options such as `/dq/IFO/FLAG/VERSION/active?s=t1&e=t2`. This URI will return all active segments for the given FLAG in the GPS interval [t1, t2).

The DQSEGDB database service with this design is much faster and more stable than its predecessors. The database currently holds over 600 million DQ segment entries stably. The server responds to queries such as a request for the segments that define when the LIGO Livingston IFO is in optimal observing mode over a month of operations, within 3 s. The MariaDB database is well optimized, such that most of this time is spent in the conversion to the JSON output format. Table 1 lists some metrics of the performance of the server in 2016 and 2020. The database has grown by a factor of $O(100)$ over this time, and the results demonstrate that the performance of the server has not changed despite this increase in size. The server also responds to $O(5)$ times more requests per day now than in 2016, averaging 7.178 requests per second.

Date	Requests				Avg. Response Time (s)		
	Get	Patch	Total	Req./s	Get	Patch	All
09/14/16	37,673	80,857	118,530	1.401	3.584	0.018	1.151
02/10/20	39,690	567,588	607,278	7.178	3.238	0.012	0.223

Table 1: Demonstrating performance stability of the DQSEGDB service. In 2020, the database contains $O(100)$ times more data, and responds to $O(5)$ times more requests per second with nearly identical performance compared to 2016 values.

4. Impact

The new DQSEGDB system of servers has been very successful in meeting the needs of the GW community for storing and distributing IFO metadata since 2014. Thanks to the high performance of the DQSEGDB service, nearly all LVC GW searches are using this centralized source of data quality information. It is, thus, also providing a system for careful control and synchronization of the detector status information used by the LVC searches. The DQSEGDB is also used by many automated IFO monitoring processes and many LVC scientists investigating the performance of the IFOs. In particular, the data analyses that concluded in the detection of all GWs thus far have relied on the DQSEGDB infrastructure [11, 4, 12, 13, 14].

The impact of the DQ information hosted in the DQSEGDB on GW searches is significant, as illustrated in Figure 1 [10]. The information is used to mitigate systematic noise issues, and thanks to the speed of the DQSEGDB service, the speed of testing of the different choices of DQ flags has been drastically improved. One example of the types of DQ flags used

to remove a significant amount of noise was the "RF45 flag". This flag indicated times when issues with the electronics that controlled the radio frequency (RF) sidebands used to sense and control LIGOs optical cavities would contaminate the main detection channel with noise that resulted in a significant number of false triggers in DA pipelines [10]. The latency from the time data is collected at the IFO sites to the moment the metadata may be queried by rapid analyses has also been reduced to less than 5 minutes. This functionality is being used by several "medium latency" analyses that are automatically started in response to external events such as observations of gamma-ray bursts.

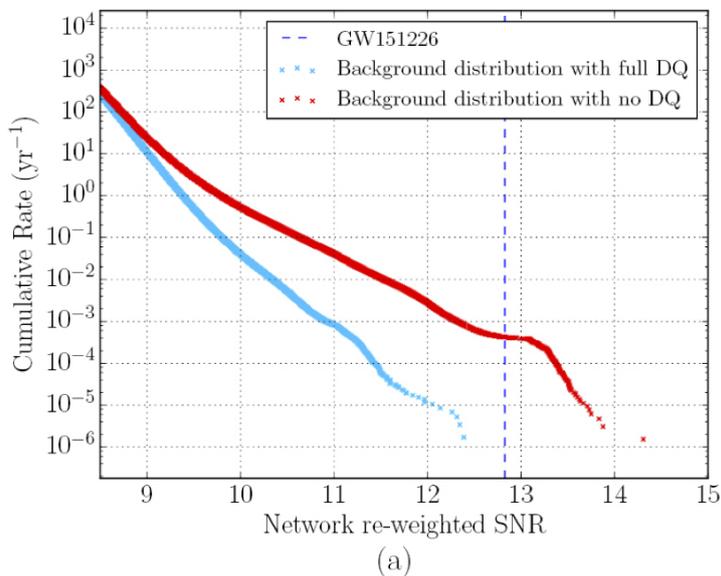


Figure 1: This image is reproduced with permission from [10], original creator T. J. Massinger. This image illustrates the impact of applying DQ flags contained in the DQSEGDB to a LVC GW search in the data from Advanced LIGO's first observing run. The initial background events in the search have SNR values reaching above 14. With the DQ applied, the background is reduced to below an SNR of 12.5. The upper limit of this background is closely tied to the limit at which GW events may confidently be detected. The SNR of the GW event GW151226 is indicated on the figure to demonstrate that this detection would have been missed without the use of the DQ data. Image License <https://creativecommons.org/licenses/by/3.0/>, image was cropped from original.

The new DQSEGDB service has allowed new collaboration tools that automatically, and very frequently, query the DQ flag metadata to be developed. These services provide many different benefits to the large, 1000+ person LVC. In particular, the Summary Page web infrastructure [15] makes heavy use of this ability. One open-access example of these pages is available at https://www.gw-openscience.org/detector_status/day/20200207/.

Many of the plots indicate the state of the interferometers, and all of this metadata is being retrieved from the DQSEGDB. These plots are updated on a rolling basis, requiring very frequent queries to the DQSEGDB service that the old service would not have been able to handle. These pages are used by IFO commissioners, data quality investigators, data analysts and the wider astronomical community to easily assess the state of the interferometers and rapidly investigate systematic issues. They have proven invaluable to data and event validation efforts in addition to daily IFO and collaboration operations.

Finally, due to the speed and reliability of the service, additional GW detectors have also begun using this single instance of the DQSEGDB. The GEO600 (GEO) collaboration now uses this DQSEGDB instance to store its primary detector state data. The Kamioka Gravitational wave detector (KAGRA) collaboration has recently begun storing metadata with this service as well, with the recent start of full observations. Thus, the DQSEGDB infrastructure and service is now used by all IFO-based GW detection efforts in the world.

5. Conclusions

The DQSEGDB software has been tremendously successful in serving the GW astronomy community. This set of database, backend, frontend and client software has provided rapid access to the DQ segments needed by the LVC for all GW detections made thus far. The speed and reliability of the database combined with its clean, RESTful API has resulted in the design of new tools that enable scientists to more rapidly and easily understand the IFOs in the GW detection network.

6. Conflict of Interest

No conflict of interest exists: We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

7. Role of Funding Source

Funding for this project was provided by the National Science Foundation, the European Gravitational Observatory, the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Foundation for Fundamental Research on Matter

supported by the Netherlands Organisation for Scientific Research. The funding sources had no involvement in the work described in this article or the writing or submission of the article.

Acknowledgements

The authors gratefully acknowledge the support of the United States National Science Foundation (NSF) for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. The authors gratefully acknowledge the National Science Foundation Grants PHY-1700765 and PHY-1104371 for the support of this project. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors gratefully acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Foundation for Fundamental Research on Matter supported by the Netherlands Organisation for Scientific Research, for the construction and operation of the Virgo detector, their support in all software developments, and the creation and support of the EGO consortium. The authors also gratefully acknowledge research support from these agencies as well as by the Council of Scientific and Industrial Research of India, the Department of Science and Technology, India, the Science & Engineering Research Board (SERB), India, the Ministry of Human Resource Development, India, the Spanish Agencia Estatal de Investigación, the Vicepresidència i Conselleria d'Innovació, Recerca i Turisme and the Conselleria d'Educació i Universitat del Govern de les Illes Balears, the Conselleria d'Educació, Investigació, Cultura i Esport de la Generalitat Valenciana, the National Science Centre of Poland, the Swiss National Science Foundation (SNSF), the Russian Foundation for Basic Research, the Russian Science Foundation, the European Commission, the European Regional Development Funds (ERDF), the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, the Hungarian Scientific Research Fund (OTKA), the Lyon Institute of Origins (LIO), the Paris Île-de-France Region, the National Research, Development and Innovation Office Hungary (NKFIH), the National Research Foundation of Korea, Industry Canada and the Province of Ontario through the Ministry of Economic Development and Innovation, the Natural Science and Engineering Research Council Canada, the Canadian Institute for Advanced Research, the Brazilian Ministry of Science, Technology, Innovations, and Communications, the International Cen-

ter for Theoretical Physics South American Institute for Fundamental Research (ICTP-SAIFR), the Research Grants Council of Hong Kong, the National Natural Science Foundation of China (NSFC), the Leverhulme Trust, the Research Corporation, the Ministry of Science and Technology (MOST), Taiwan and the Kavli Foundation. The authors gratefully for computational resources provided by the NSF, STFC, INFN and CNRS, and supported by National Science Foundation Grants PHY-1626190, PHY-1700765, PHY-0757058 and PHY-0823459. The authors gratefully acknowledge the contributions of Larne Pekowsky and Ping Wei for the development, support and improvement of the original LIGO segment database and Leone Bosi for the development of the original Virgo database.

References

- [1] K. Thorne, Gravitational radiation, in: S. Hawking, W. Israel (Eds.), *Three Hundred Years of Gravitation*, Cambridge University Press, Cambridge, 1987, pp. 330–458.
- [2] J. D. E. Creighton, W. G. Anderson, *Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis*, 2011.
URL <http://www.wiley-vch.de/publish/dt/books/ISBN3-527-40886-X>
- [3] B. P. Abbott, et al., Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* 116 (6) (2016) 061102. [arXiv:1602.03837](#), [doi:10.1103/PhysRevLett.116.061102](#).
- [4] B. P. Abbott, et al., GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, *Phys. Rev. Lett.* 119 (16) (2017) 161101. [arXiv:1710.05832](#), [doi:10.1103/PhysRevLett.119.161101](#).
- [5] B. P. Abbott, et al., GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, *Phys. Rev. X* 9 (3) (2019) 031040. [arXiv:1811.12907](#), [doi:10.1103/PhysRevX.9.031040](#).
- [6] B. P. Abbott, et al., Upper limits on the rates of binary neutron star and neutron-star–black-hole mergers from Advanced LIGO’s first observing run [arXiv:1607.07456](#).
- [7] B. P. Abbott, et al., Supplement: The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914 [arXiv:1606.03939](#).

- [8] B. P. Abbott, et al., The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914 [arXiv:1602.03842](#).
- [9] B. P. Abbott, et al., Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A, *Astrophys. J.* 848 (2) (2017) L13. [arXiv:1710.05834](#), [doi:10.3847/2041-8213/aa920c](#).
- [10] B. P. Abbott, et al., Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGO's first observing run, *Class. Quant. Grav.* 35 (6) (2018) 065010. [arXiv:1710.02185](#), [doi:10.1088/1361-6382/aaaafa](#).
- [11] B. P. Abbott, et al., Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914, *Class. Quant. Grav.* 33 (13) (2016) 134001. [arXiv:1602.03844](#), [doi:10.1088/0264-9381/33/13/134001](#).
- [12] B. P. Abbott, et al., GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4M_{\odot}$, [arXiv e-prints \(2020\)](#) [arXiv:2001.01761](#)[arXiv:2001.01761](#).
- [13] B. P. Abbott, et al., Multi-messenger Observations of a Binary Neutron Star Merger, *Astrophys. J.* 848 (2) (2017) L12. [arXiv:1710.05833](#), [doi:10.3847/2041-8213/aa91c9](#).
- [14] B. P. Abbott, et al., GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence, *Phys. Rev. Lett.* 119 (14) (2017) 141101. [arXiv:1709.09660](#), [doi:10.1103/PhysRevLett.119.141101](#).
- [15] D. Macleod, A. L. Urban, M. Isi, T. Massinger, paulalain, M. Pitkin, A. Nitz, [gwpy/gwsumm: 1.0.2 \(Dec. 2019\)](#). [doi:10.5281/zenodo.3590375](#).

Required Metadata

Current code version

1.6.1

Nr.	Code metadata description	Please fill in this column
C1	Current code version	1.6.1
C2	Permanent link to code/repository used for this code version	https://github.com/ligovirgo/dqsegdb/tree/dqsegdb-release-1.6.1
C3	Code Ocean compute capsule	N/A
C4	Legal Code License	GNU GENERAL PUBLIC LICENSE v3.0
C5	Code versioning system used	git
C6	Software code languages, tools, and services used	Python, MariaDB, PHP
C7	Compilation requirements, operating environments & dependencies	Scientific Linux 7.5, Python 2.7.5
C8	If available Link to developer documentation/manual	http://ligovirgo.github.io/dqsegdb/
C9	Support email for questions	question@ligo.org

Table 2: Code metadata (mandatory)