National Aeronautics and Space Administration

The LIGO Open Science Center

Michele Vallisneri

Jet Propulsion Laboratory
California Institute of Technology

Jonah Kanner, Roy Williams, Alan Weinstein

Caltech LIGO Lab

Branson Stephens

University of Wisconsin Milwaukee

for the LIGO Scientific Collaboration

LIGO DCC G1400532

PHILOSOPHICAL TRANSACTIONS:

GIVING SOME

ACCOMPT

OF THE PRESENT

Undertakings, Studies, and Labours

OF THE

INGENIOUS

IN MANY

CONSIDERABLE PARTS

OF THE

WORLD.

Vol I.

For *Anno* 1665, and 1666.

In the *SAVOY*,

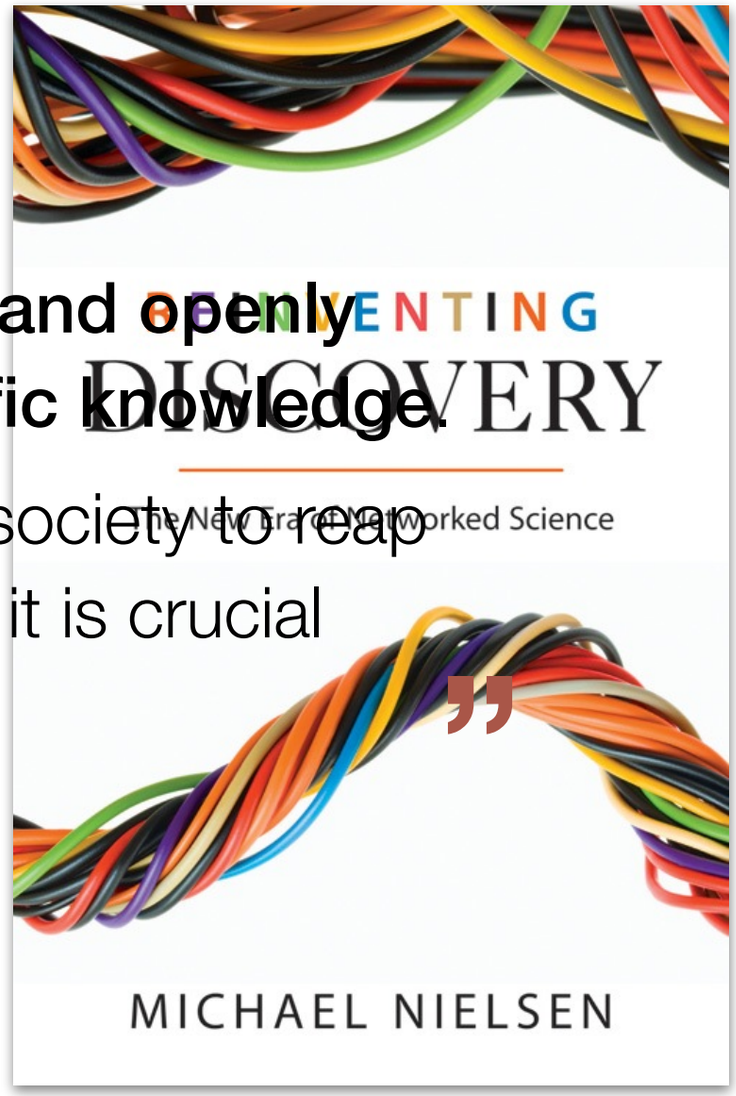
Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,
Printers to the *Royal Society*.

Presented by the Author May. 30th 1667.

Panton principles (2010)

Science is based on **building on, reusing and openly criticising the published body of scientific knowledge.**

For science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that **science data be made open**.



NASA space act (1953)

[*Functions of the Administration*. NASA shall…]

- arrange for **participation by the scientific community** in planning scientific measurements and observations to be made through use of aeronautical and space vehicles, and conduct or arrange for the conduct of such measurements and observations;
- provide for the **widest practicable and appropriate dissemination of information** concerning its activities and the results thereof;
- seek and encourage, to the maximum extent possible, the fullest commercial use of space

“ OMB open data memorandum (2013)

Making information resources accessible, discoverable, and usable by the public can help fuel entrepreneurship, innovation, and scientific discovery—all of which improve Americans' lives and contribute significantly to job creation.

Open data will be: **public**, **accessible**, **described**, **reusable**, **complete**, **timely**, **managed** in post-release. ”

So why should the LIGO data be free and open?

Because the observatory is a public asset.

Because it will **maximize discovery**.

Because it provides excellent opportunities for outreach and teaching.

Because it will help preserve the data.

LIGO Data Management Plan DCC M1000066

Discovery phase

During aLIGO's transition to design sensitivity, the understanding of data is evolving and **detections are rare**. Data is restricted to the collaboration; alerts by MOU.

Release validated discoveries and important non detections, in a manner analogous to post-processed astronomical images.

Already: GRB051103, Big Dog.

Soon: **full S5**, S6.

Observational phase

Mature sensitivity and understanding of the data allow the exploration of the astrophysical content of GWs (populations, distributions, multimessenger).

Release entire GW data, corrected for instrumental and environmental issues (in blocks of 6 months, every 6 months, with **24-month latency**). Release public transient alerts.

Transition when:

- 3×10^7 Mpc^3 y probed **or**
- plentiful detections **or**
- 3.5 y after acceptance

LIGO Data Management Plan DCC M1000066

The LIGO Data System

Has archived 1 PB of LIGO data (1% in the GW channel)
and made it available in real time to hundreds of LIGO scientists.

Implements the elements of an **Open Archival Information System**:

Data ingestion
DAQ, frames

Data storage
LDG, redundant

Metadata annotation
detector, calibration, injections

Data preservation
preserving bits and the **meaning**

Data operations
LDG computing, cloud

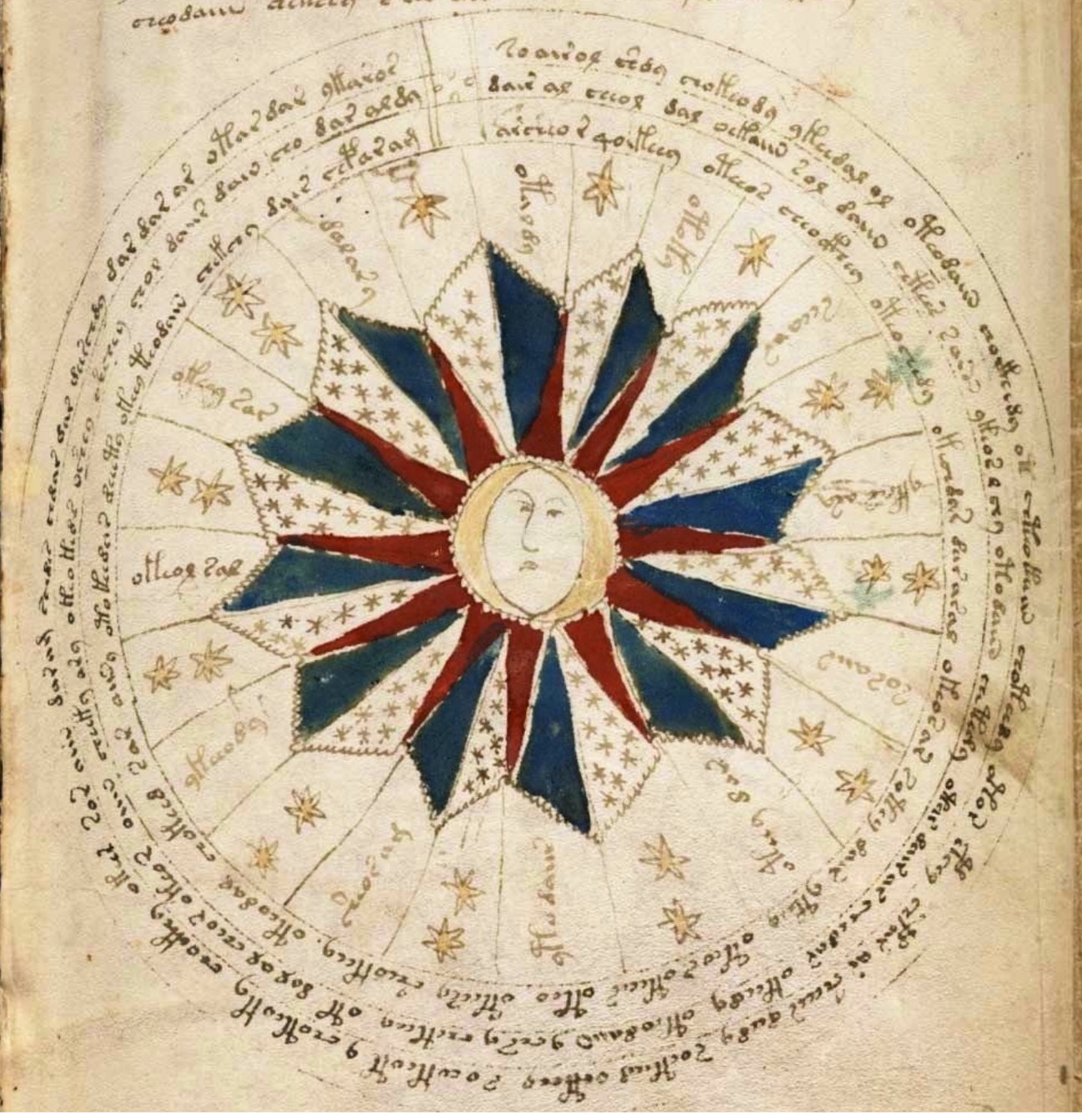
Data access
bulk data, data quality, alerts

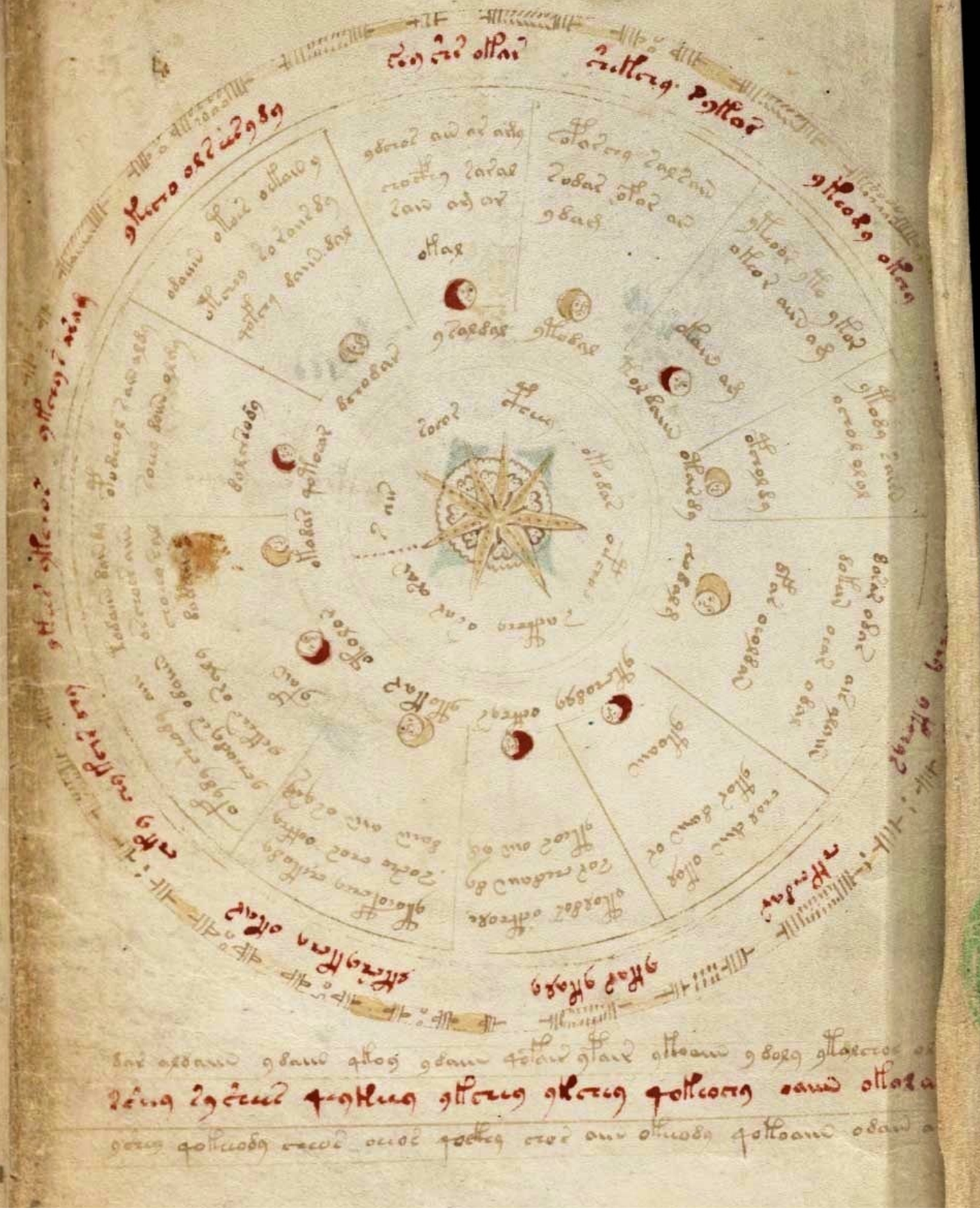
The LIGO Open Science Center (est. 2012)

"A fabric to enable robust scientific discourse about the LIGO data for somebody who has never met any LSC member."

OAIS archival data should be independently understandable by **designated communities**:

- **LSC scientists**
- **non-LSC scientists**
- **the public**





Voynich manuscript (1404–1438?)

LOSC data releases losc.ligo.org

calibrated h(t) strain

(**all of S5** run in 2014, S6 next, aLIGO internally)

interoperable: common formats (HDF5, frames)
I/O software tools

self-describing: in-file **data-quality** information
injection flags/catalogs

published online: queries and selection
by “click and code”

contextualized: getting started, tutorials, FAQ,
examples, references…

supported: helpdesk, workshops/webinars,
student use, outreach

more data types

(TBD…)

baricentered h(t), h(ω)

event repository

real-time transient alerts

LOSC data releases losc.ligo.org

CBC low-mass
CBC high-mass
burst all-sky
CW SFT
stochastic

→

CAT1–4 veto definers

H1:AS_TRIGGER
H1:ASC_Overflow
H1:ASI_CORR_OVERFLOW
H1:CALIB_BAD_COEFFS_60
H1:CALIB_DROPOUT_1SAMPLE
H1:CALIB_DROPOUT_1SEC
H1:CALIB_DROPOUT_AWG_STUCK
H1:CALIB_GLITCH_ZG
H1:CHECKSUM_MISMATCH
H1:CORRUPTED_RDS_C03_LX
...

→

CAT1–4 segments

828657954,828657954
828669551,828669551
830632685,830632685
830632685,830632685
831512303,831512303
839301122,839301122
841119818,841119818
841128304,841128304
841150538,841150538
…

LOSC data quality

a **1-Hz bitmask** (rather than segments!), summarizing the data choices made in the main LSC searches

for S5/H1, CBCHIGH

passes CAT1: 77.3%

passes CAT2: 74.8%

passes CAT3: 71.8%

no science data, or data with severe problems

data with understood instrumental perturbations

data with obvious but less understood problems

good data, use it!

LOSC demo losc.ligo.org

LIGO Open Science Center

LIGO is operated by California Institute of Technology and Massachusetts Institute of Technology and supported by the National Science Foundation of the United States.

Getting Started

Welcome! The LIGO Open Science Center (LOSC) provides access to a variety of LIGO data products, as well as documentation, tutorials, and online tools for finding and viewing data. This page will walk you through a few of the available features. Access to all of the data and documentation is available through the links in the menu bar to the left.

Where's the data?

Time series strain data

Periodically, LIGO time-series data will be released publicly, and made available through these web pages. This "Bulk Data" takes the form of a time series sampled at 4096 Hz, calibrated so that a gravitational wave signal would appear in units of dimensionless strain (ΔL / L). If you are in a hurry to download some strain data right now, just click the "Bulk Data" link in the menu bar at left. However, if you are unfamiliar with LIGO data, we strongly recommend you spend a few minutes working through the Introductory Tutorial. There, you can learn not only how to download LIGO data, but also some basic steps for loading, plotting, and processing data.

What sources has LIGO found?

Event Lists

In addition to providing bulk data, LOSC provides links to access identified events found in LIGO data. This includes low-latency triggers, simulated events, and data releases associated with astronomical events (GRBs, Supernovae, etc.). To see the currently available event lists, or to learn about gaining access to low-latency LIGO triggers, click the Event Lists link in the menu bar at left.

When was LIGO running?

Segment Information

The LIGO detectors do not continuously observe, but rather, collect data in defined observing periods called "Science Runs". Even within a science run, detectors sometimes go off-line for commissioning, due to enviornmental disturbances such as seismic activity, or for other reasons. In addition, all collected data is classified by data quality level.

To provide access to information about when data was collected and at what data quality level, the LOSC provides two web interfaces to segment information. Both the "Timelines" and the "My Sources" applications provide web access to find out when LIGO was collecting science mode data.

Timelines: The Timelines applications provides a graphical view of any segment information. Click on the Timeline link in the

LOSC demo upcoming LSC webinar by J. Kanner and R. Williams

Getting Started

Data

Bulk Data

Event Lists

Timelines

Toolbox

Tutorials

Software

My sources

GPS ↔ UTC

Learn More

About LIGO

Contact

Timeline

Timeline technology provides instant access to time-based metadata on scales from seconds to centuries. Contact the LOSC development team to find out how Timeline can host your metadata. **Timeline** is developed by the LIGO Open Science Center, which is supported by the National Science Foundation of the USA.

Timelines are available to display times when a number of GW instruments were collecting data, as well as data quality and injection information. This information can be retrieved as a plot, an ASCII list of segments, or as JSON data. A simple URL pattern is all you need to find the segment information you want. To see how these URLs are constructed, fill in the information below, and see the URL that appears at the bottom of the screen.

Timeline Query Form for the S5 dataset

S5 start: GPS 815155213 (2005-11-04T16:00:00)
S5 end: GPS 875232014 (2007-10-01T00:00:00)

Enter the **starting GPS**: 843895454 (GPS ↔ UTC converter)

Enter the **end GPS or duration** in seconds: 5400

Select display type:
- (•) **Plot**: Plot fraction of time that passes the chosen DQ level in each time bin
- ( ) **Segment List**: ASCII segment list
- ( ) **JSON**: Data for timeline plot in JSON format

OPTIONAL Bin level:

The "bin level" tells what size time bins to use. The bin size will be 2^L seconds, where L is the "bin level". So, for example, if you want each bin in your plot to be 16 s, the bin level should be 4.

Select flag names. For plots and JSON, multiple selections can be made, but for segments, only one can be selected. A short description of each flag is available by hovering the mouse over its name, or on the inventory page.

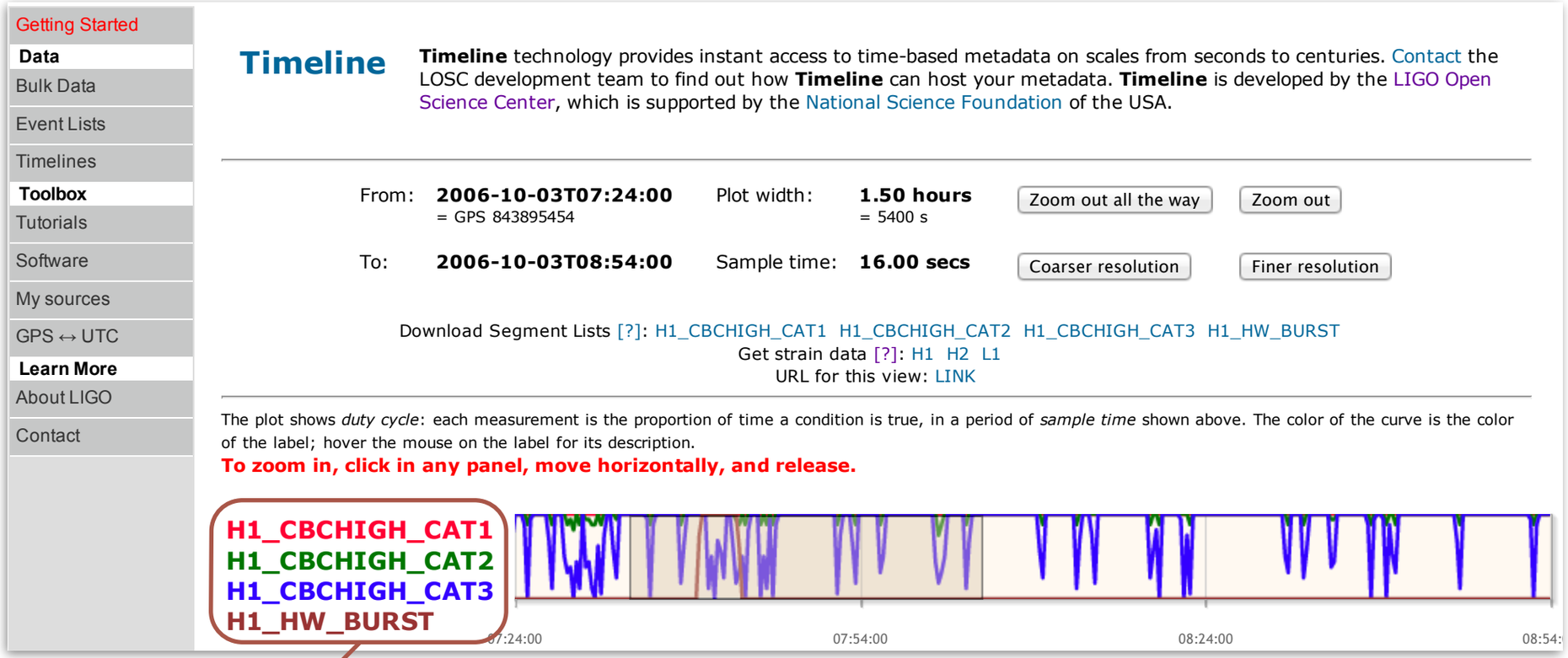
☐ H1_SCI	☐ H2_SCI	☐ L1_SCI
☐ H1_PEM	☐ H2_PEM	☐ L1_PEM
☑ H1_CBCHIGH_CAT1	☐ H2_CBCHIGH_CAT1	☐ L1_CBCHIGH_CAT1
☐ H1_HW	☐ H2_HW	☐ L1_HW
☑ H1_CBCHIGH_CAT2	☐ H2_CBCHIGH_CAT2	☐ L1_CBCHIGH_CAT2
☐ H1_HW_CBC	☐ H2_HW_CBC	☐ L1_HW_CBC
☑ H1_CBCHIGH_CAT3	☐ H2_CBCHIGH_CAT3	☐ L1_CBCHIGH_CAT3
☑ H1_HW_BURST	☐ H2_HW_BURST	☐ L1_HW_BURST
☐ H1_CBCHIGH_CAT4	☐ H2_CBCHIGH_CAT4	☐ L1_CBCHIGH_CAT4
☐ H1_HW_STOCH	☐ H2_HW_STOCH	☐ L1_HW_STOCH
☐ H1_CBCLOW_CAT1	☐ H2_CBCLOW_CAT1	☐ L1_CBCLOW_CAT1
☐ H1_CW	☐ H2_CW	☐ L1_CW
☐ H1_CBCLOW_CAT2	☐ H2_CBCLOW_CAT2	☐ L1_CBCLOW_CAT2
☐ H1_CBCLOW_CAT3	☐ H2_CBCLOW_CAT3	☐ L1_CBCLOW_CAT3
☐ H1_CBCLOW_CAT4	☐ H2_CBCLOW_CAT4	☐ L1_CBCLOW_CAT4
☐ H1_BURST_CAT1	☐ H2_BURST_CAT1	☐ L1_BURST_CAT1
☐ H1_BURST_CAT2	☐ H2_BURST_CAT2	☐ L1_BURST_CAT2
☐ H1_BURST_CAT3	☐ H2_BURST_CAT3	☐ L1_BURST_CAT3
☐ H1_BURST_CAT2E	☐ H2_BURST_CAT2E	☐ L1_BURST_CAT2E
☐ H1_BURST_CAT3E	☐ H2_BURST_CAT3E	☐ L1_BURST_CAT3E
☐ H1_CW_CAT1	☐ H2_CW_CAT1	☐ L1_CW_CAT1
☐ H1_STOCH_CAT1	☐ H2_STOCH_CAT1	☐ L1_STOCH_CAT1
☐ H1_STOCH_CAT2_H1L1	☐ H2_STOCH_CAT2_H1L1	☐ L1_STOCH_CAT2_H1L1
☐ H1_STOCH_CAT2_H2L1	☐ H2_STOCH_CAT2_H2L1	☐ L1_STOCH_CAT2_H2L1

Get your Timeline at:

https://losc-dev.ligo.org//timeline/show/S5/H1_CBCHIGH_CAT1-H1_CBCHIGH_CAT2-H1_CBCHIGH_CAT3-H1_HW_BURST/843895454/5400/

Query the LOSC **Timeline**
to inspect data of desired quality.

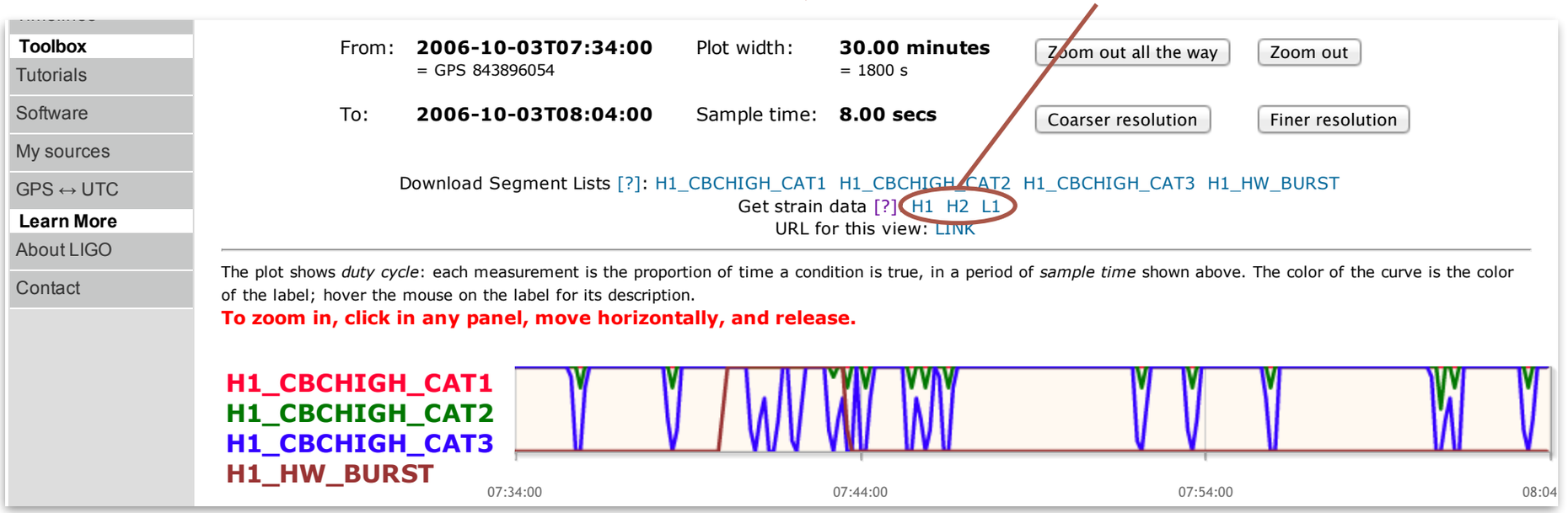
LOSC demo upcoming LSC webinar by J. Kanner and R. Williams



Visualize fraction of time in data-quality category

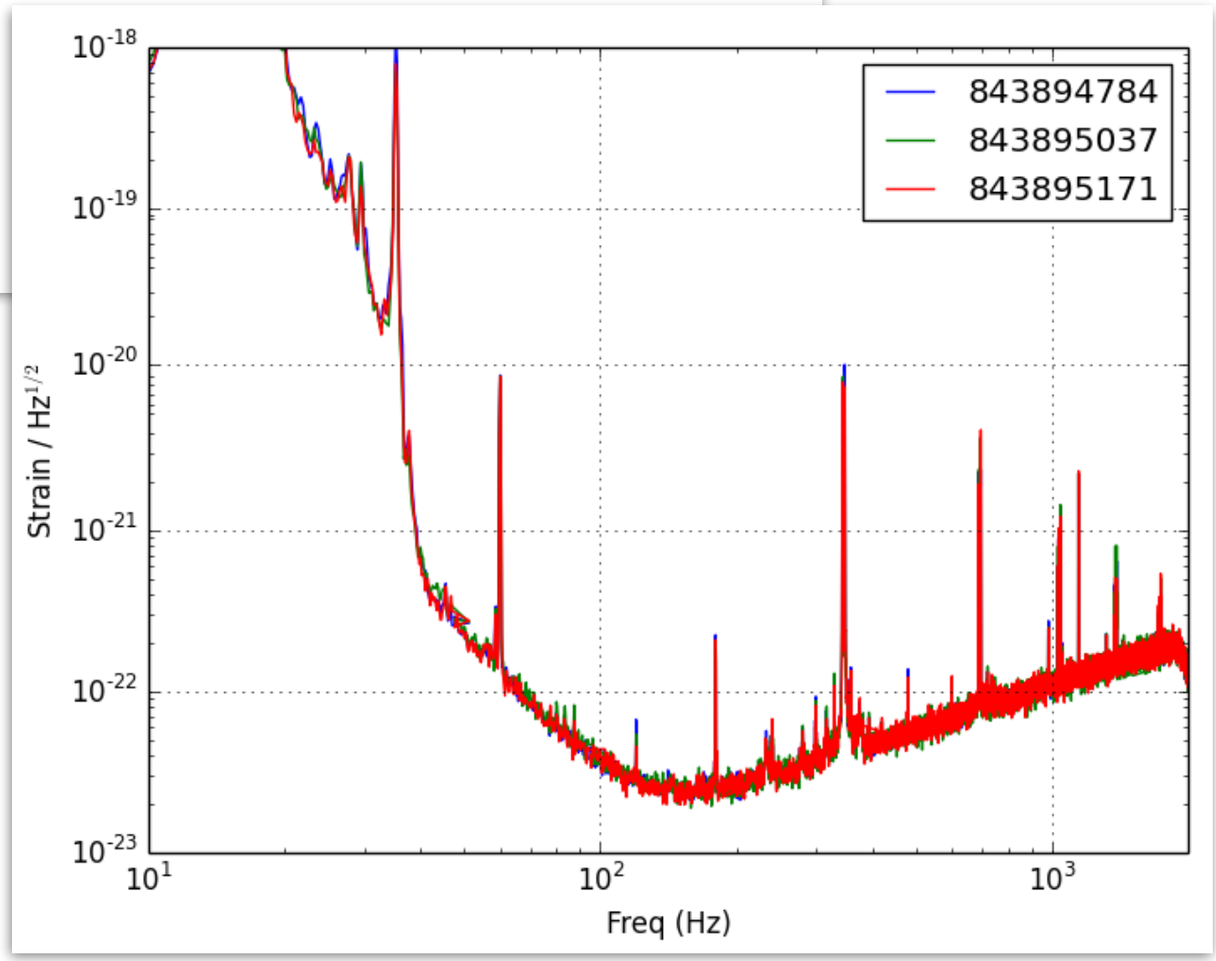
Zoom

Download HDF data file:
losc.ligo.org/archive/data/S5/843055104/H-H1_LOSC_4_V1-843894784-4096.hdf5



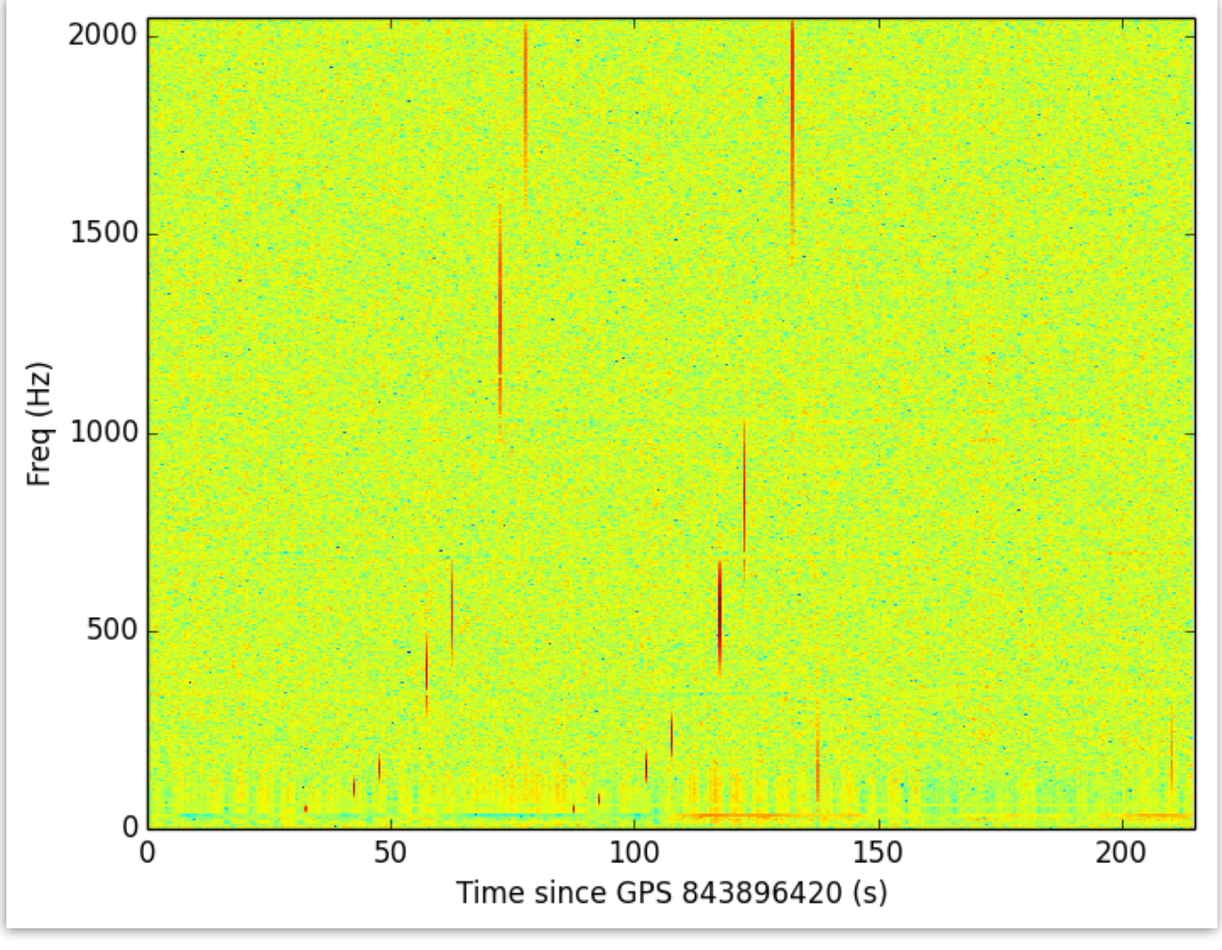
LOSC demo upcoming LSC webinar by J. Kanner and R. Williams

```
import readligo as rl
import numpy as np, matplotlib.pyplot as plt, matplotlib.mlab as mlab

# create a list of CBCHIGH post-CAT2 segments
seglist = rl.getsegs(843894784, 843894784+4096, 'H1', flag='CBCHIGH_CAT2')

# loop over the segments
for (start, stop) in seglist[0:3]:
    # load the data and data quality as numpy arrays
    strain, meta, dq = rl.getstrain(start, stop, 'H1')

    # plot the ASD
    fs = int(1.0/meta['dt'])
    Pxx, freqs = mlab.psd(strain, Fs=fs
                          NFFT=4*fs)
    plt.loglog(freqs, np.sqrt(Pxx),
               label=str(start))
```



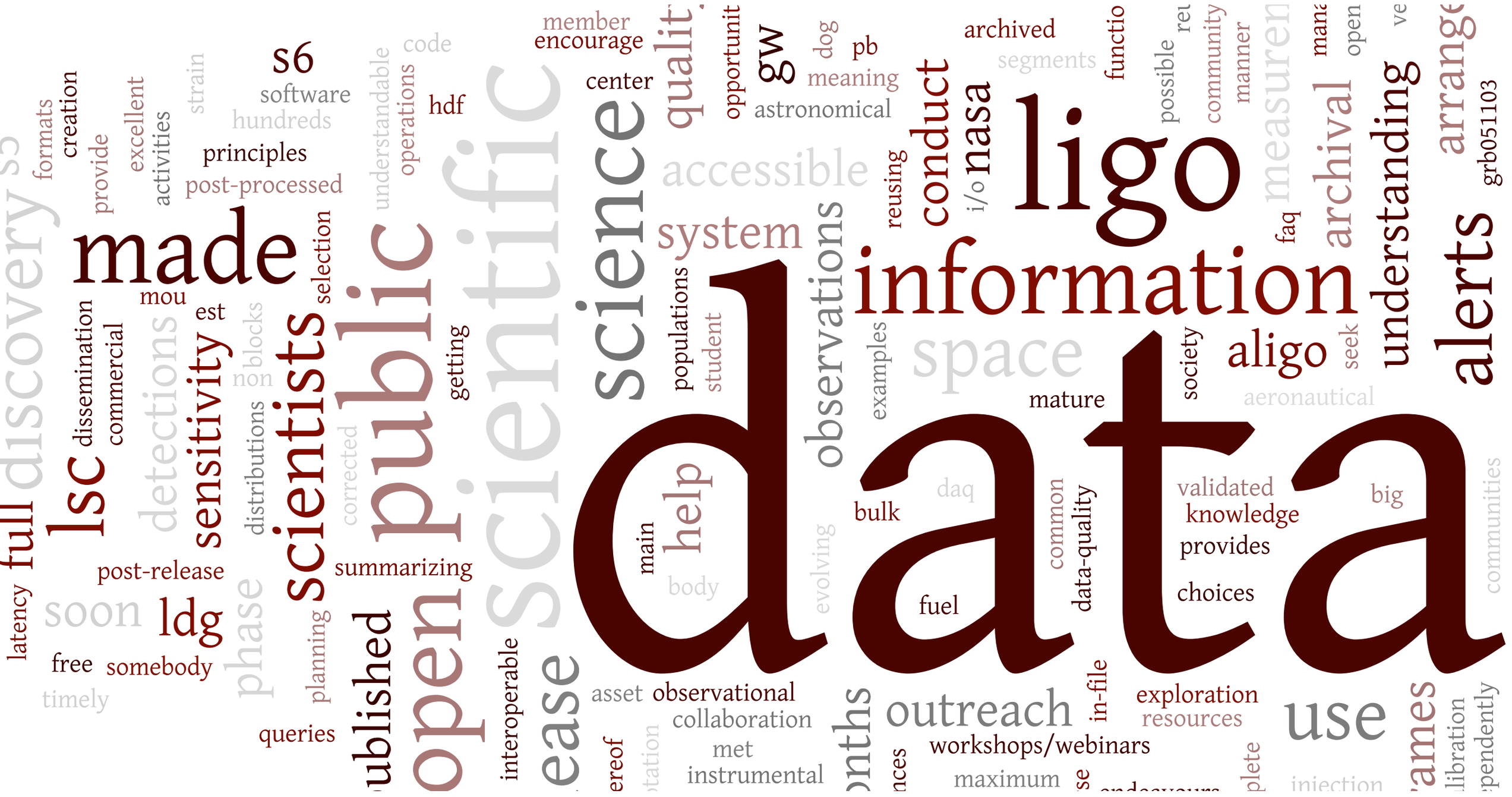
LOSC demo upcoming LSC webinar by J. Kanner and R. Williams

```
for (start, stop) in seglist[0:3]:
    strain, meta, dq = rl.getstrain(start, stop, 'H1')

    # make a spectrogram
    fs = int(1.0/meta['dt'])
    window = np.blackman(NFFT)
    spec_power, freqs, bins = mlab.specgram(strain, NFFT= fs/2, Fs=fs, window=window)

    # normalize each frequency bin to its median power
    spec_power /= np.median(spec_power,axis=1)[:,np.newaxis]
    plt.pcolormesh(bins, freqs, np.log10(spec_power))
```



Compare with **list of burst injections** on losc.ligo.org; it's a match!

GPS	IFO	Waveform	Scale	Message	SNR
843896462.5	H1	wfsg100Q9	6.4	Successful	88.7
843896467.5	H1	wfsg153Q9	3.2	Successful	84.5
843896472.5	H1	wfsg235Q9	3.2	Successful	72.8
843896477.5	H1	wfsg393Q9	6.4	Successful	90.5
843896482.5	H1	wfsg554Q9	8	Successful	88.7
843896487.5	H1	wfsg850Q9	11.2	Successful	82.5
843896492.5	H1	wfsg1304Q9	12.8	Successful	61.4
843896497.5	H1	wfsg2000Q9	19.2	Successful	20.2
843896502.5	H1	wfsg3068Q9	7.2	Successful	0.0
843896507.5	H1	wfsg50Q9	96	Successful	110.3
843896512.5	H1	wfsg70Q9	32	Successful	140.3
843896517.5	H1	wfsg100Q9	12.8	Successful	177.5
843896522.5	H1	wfsg153Q9	6.4	Successful	168.9
843896527.5	H1	wfsg235Q9	6.4	Successful	145.6
843896532.5	H1	wfsg393Q9	12.8	Successful	181.0
843896537.5	H1	wfsg554Q9	16	Successful	177.4
843896542.5	H1	wfsg850Q9	22.4	Successful	165.1
843896547.5	H1	wfsg1304Q9	25.6	Successful	122.8



The LIGO Open Science Center

Jonah Kanner, Roy Williams,
Alan Weinstein, Michele Vallisneri,
Branson Stephens, and more…