Gravitational Wave Data: The Last Mile

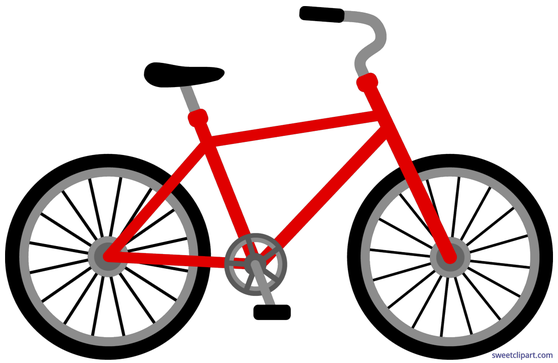
Jonah Kanner
LIGO Lab, Caltech

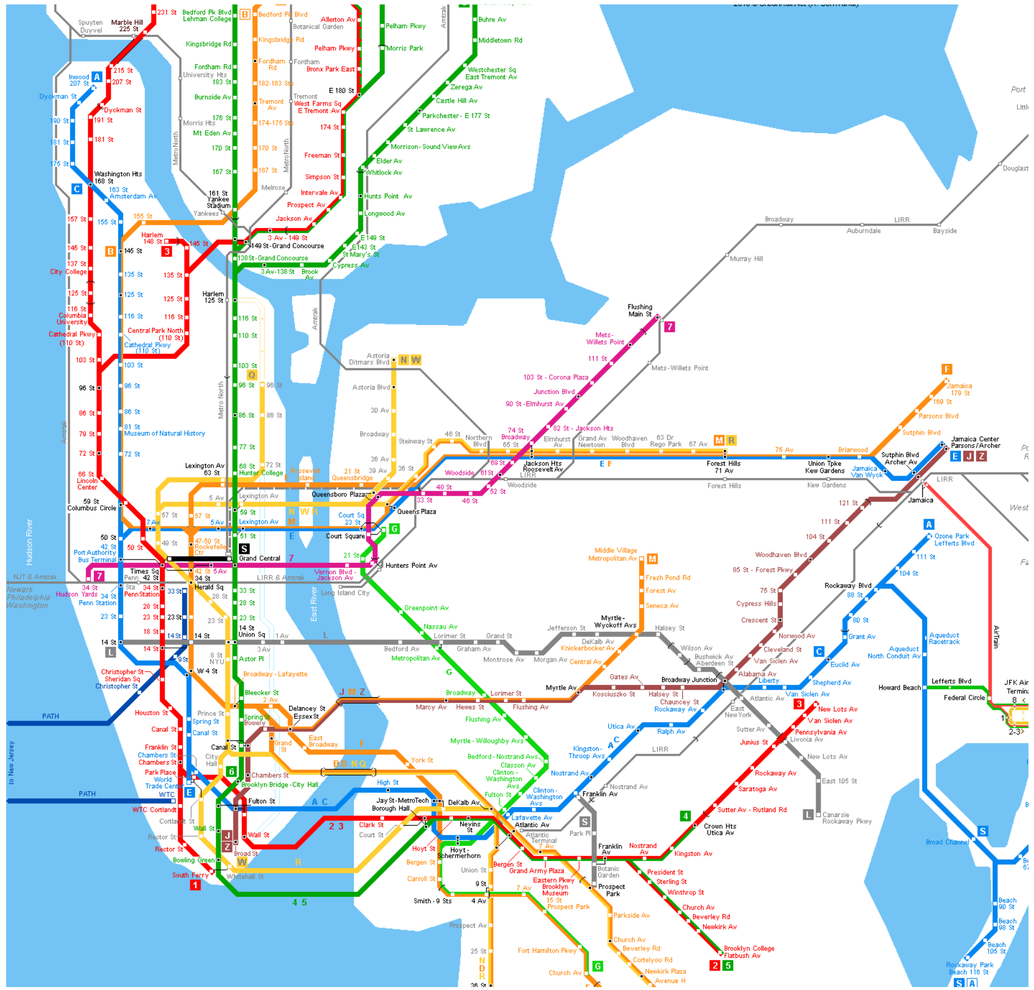
June 15, 2022 - G2200946-v3

The Last Mile Problem

Transportation

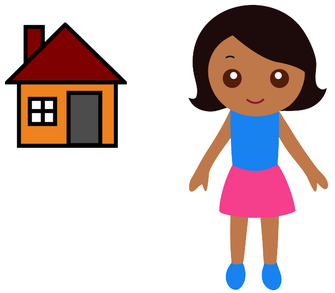
Last Mile

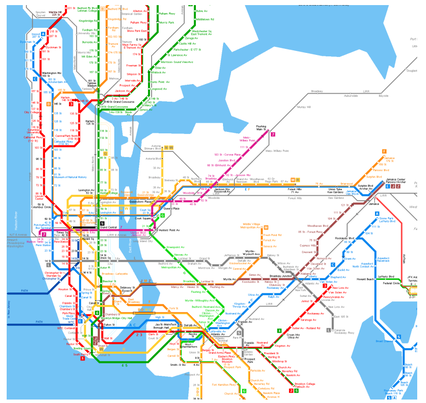




The Last Mile Problem

Public Data





Expert Networks

"Last Mile"

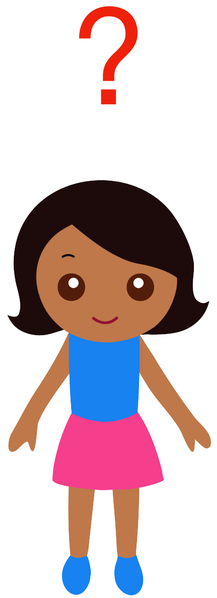
High School Student
Undergraduates
Grad Students
Experts in other field
Amateurs
Artists

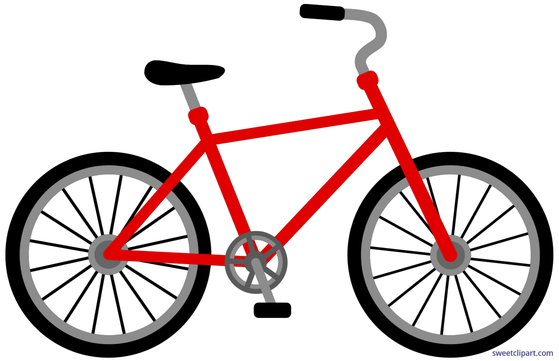
Gaps in:
Access
Knowledge
Resources

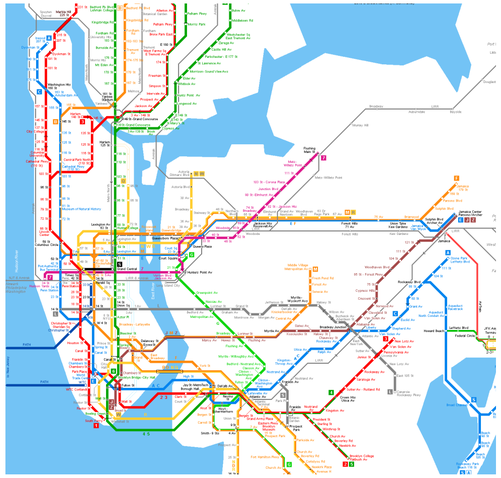
Data
Software
Journal Articles
Conferences
Colleagues

The Last Mile Problem

High Stakes







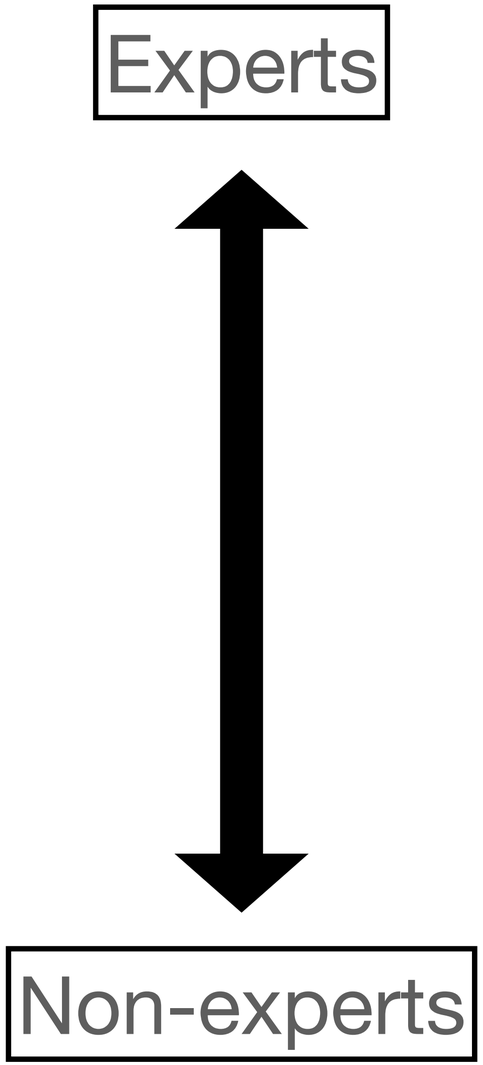
Diversity, Equity, and Inclusion

Efficiency and Productivity

Synergy

Climate Change

Diversity of Experience

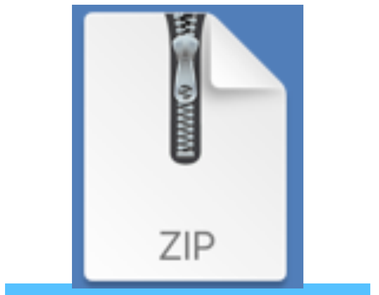


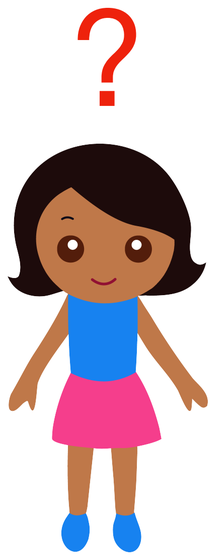
Want lots of data
High demands on data access
Computers access data (API)
Details matter

Want lots of services
Need lots of support
Human downloads data (HTML)
Too much detail is confusing

Sure, our data are public … but:

- Are the data easy to find and download?
- Do I recognize the file format? Can I figure out how to open it?
- Can I load the data in a spreadsheet or text file?
- Are there “secret steps” to processing the data?
- Can I find the software? Can I get it installed on my computer?
- Once the software is installed, can I figure out how to use it?
- Do I know where to ask for help when I get stuck?



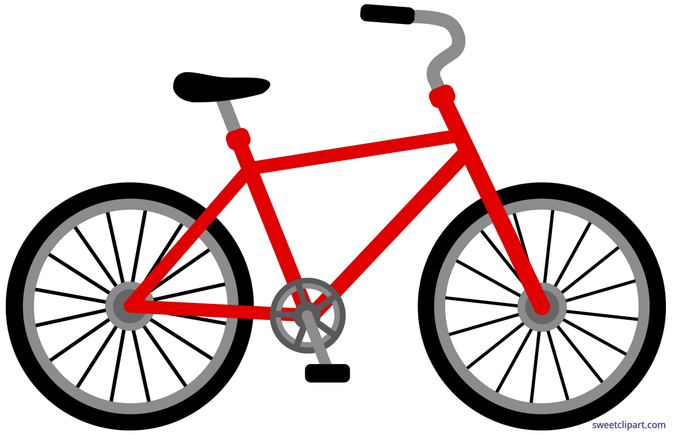


Access Gaps for LIGO Data (circa 2014)

- All data stored in “special” file format (GWF)
  - Won’t work with outside tools
  - Won’t work on Windows (90% of computers !!)
- All data access requires programming (e.g. in python)
- Specialized libraries lacked examples / documentation
- Some signal processing required
- Data contain detector artifacts

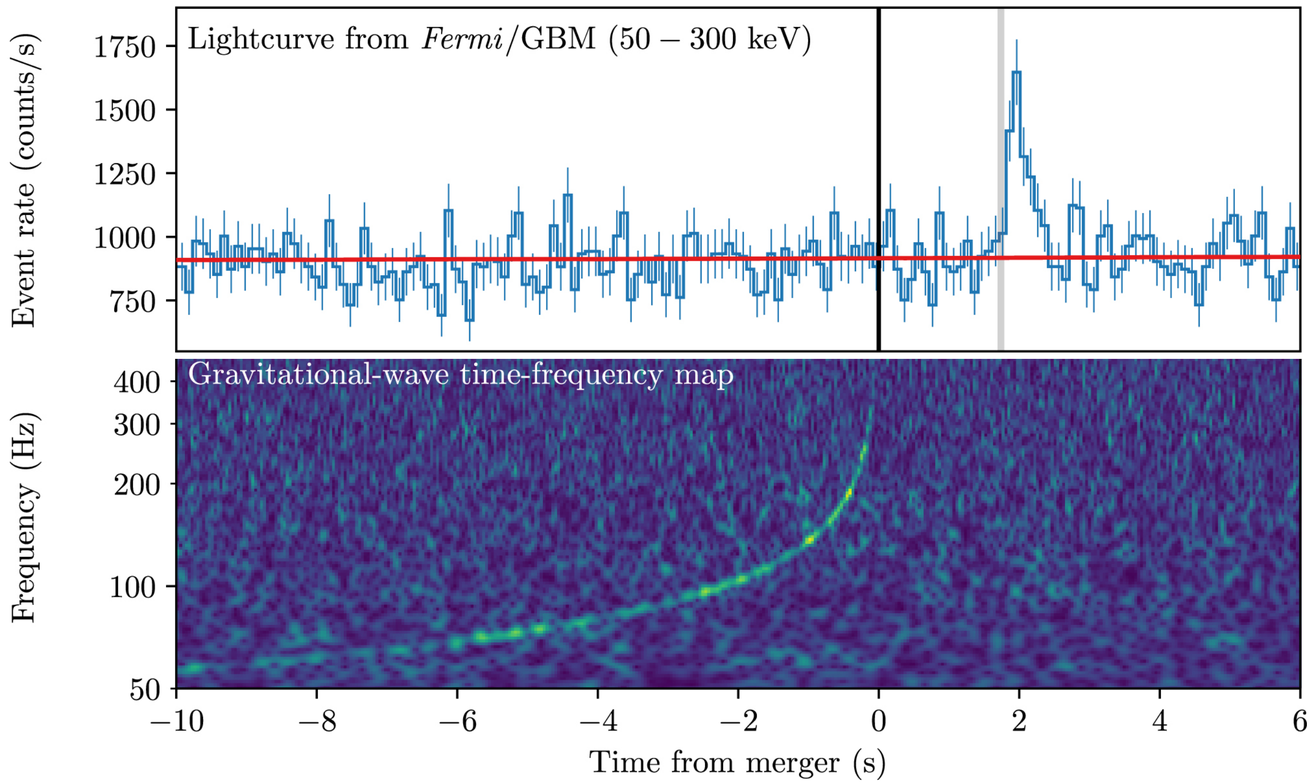
Solutions for LIGO data

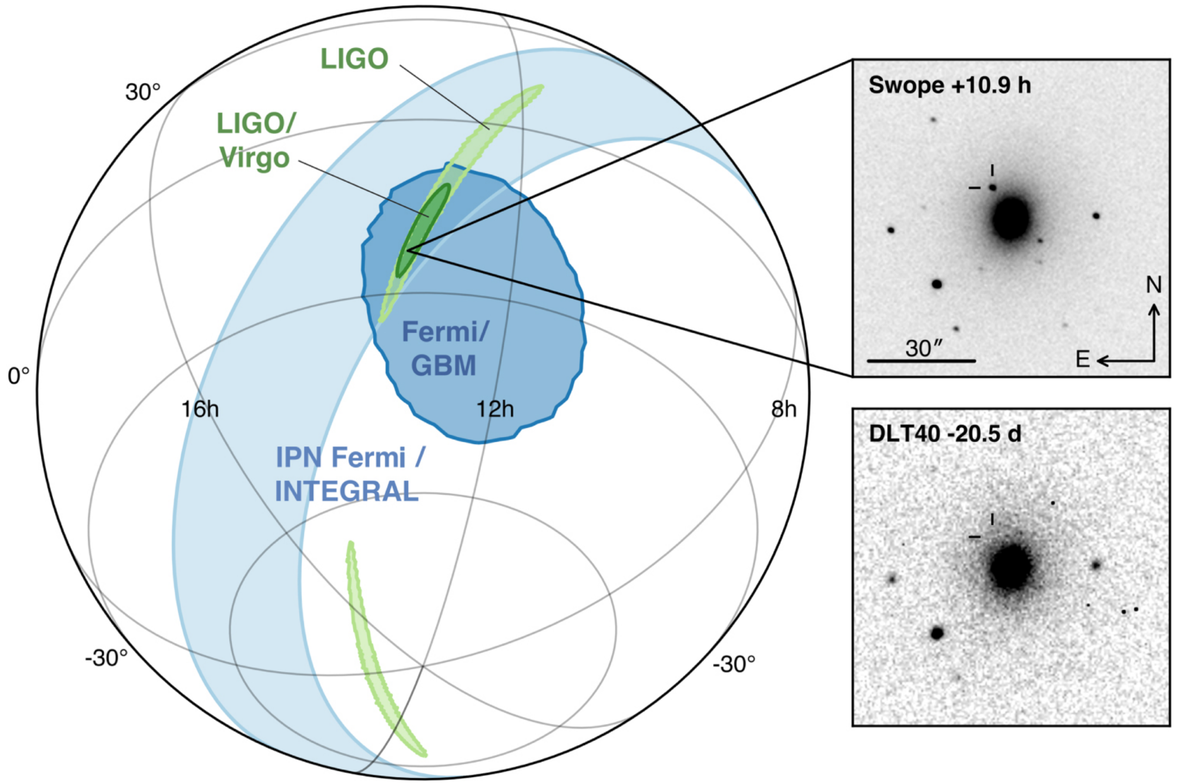
- Data in multiple formats (GWF and HDF5 and “streaming”)
- Software examples to show people ***exactly*** how to get started
  - Focus on basic tasks: loading, pre-processing, and plotting
- Use online tools, so no software installation is needed
  - (Google co-lab , mybinder , streamlit)
- Link to resources: software libraries, related data, papers, tools, web services
- Workshops and online courses
- Help Desk and Discussion Forum
- Integrated platform: gwosc.org



Everything at https://gwosc.org

Synergy and Multi-messenger Astrophysics





LIGO / Virgo / KAGRA share data and perform low-latency analysis

Results public within minutes

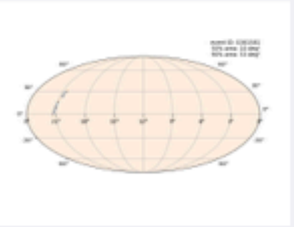
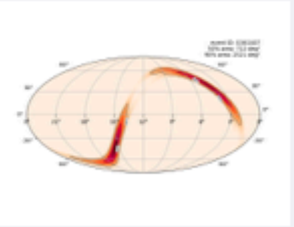
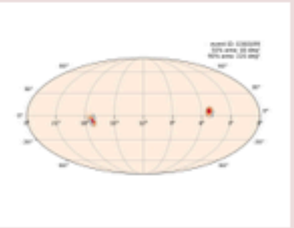
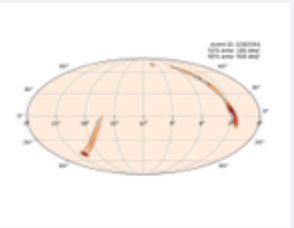
Telescopes perform follow-up observations

GW170817
"Most Observed Transient"
1st Observation of a BNS Merger

B. P. Abbott *et al* 2017 *ApJL* **848** L12

GraceDB **Public Alerts** Latest Search Documentation Login

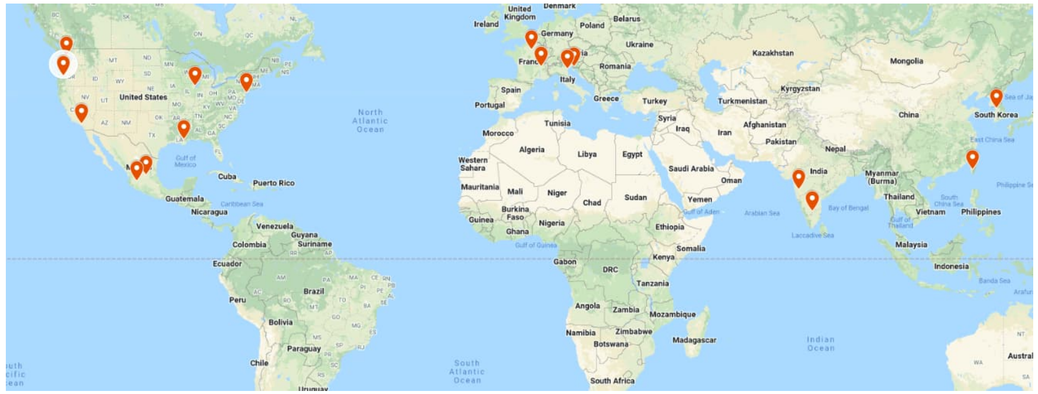
Please log in to view full database contents.

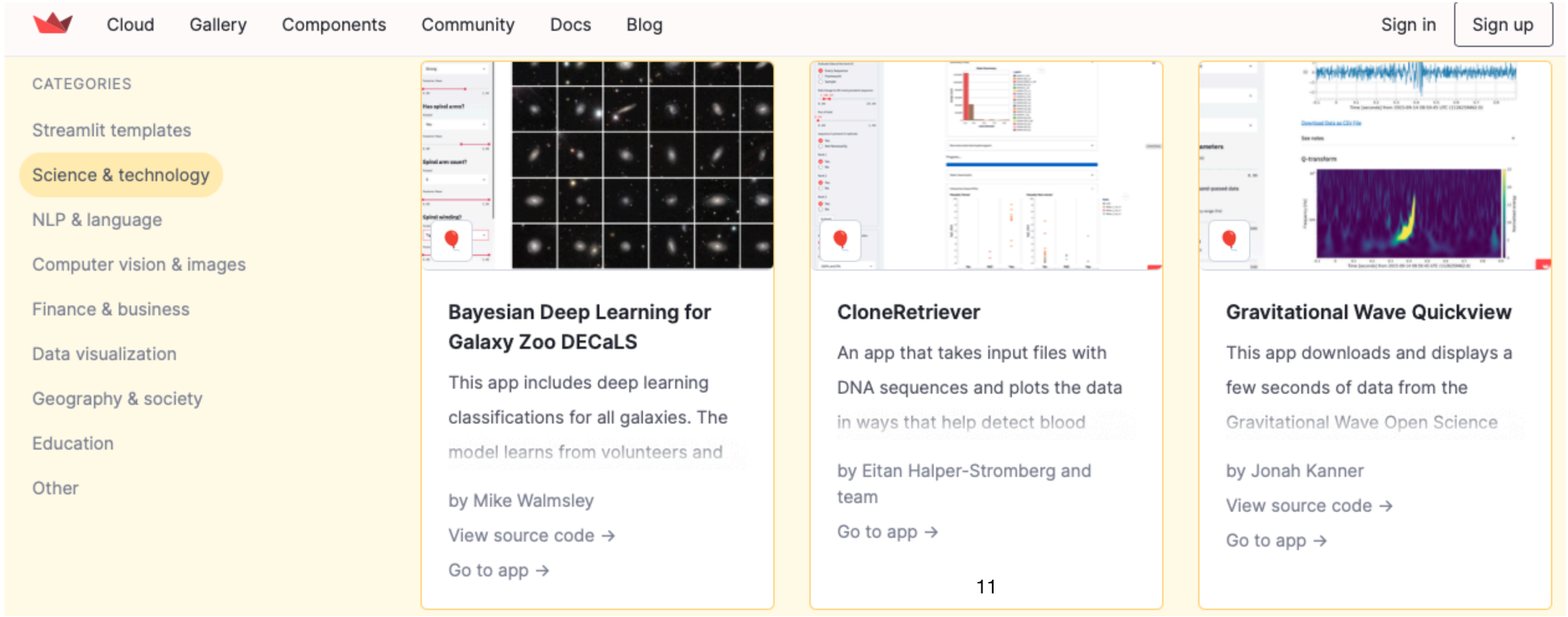
S200129m	BBH (>99%)	Jan. 29, 2020 06:54:58 UTC	GCN Circulars Notices | VOE	
S200128d	BBH (97%), Terrestrial (3%)	Jan. 28, 2020 02:20:11 UTC	GCN Circulars Notices | VOE	
S200116ah	NSBH (>99%)	Jan. 16, 2020 11:56:42 UTC	GCN Circulars Notices | VOE	
S200115j	MassGap (>99%)	Jan. 15, 2020 04:23:09 UTC	GCN Circulars Notices | VOE	

https://gracedb.ligo.org

Supporting the Community

Discussion forum: https://ask.igwn.org
E-mail help desk: gwosc@igwn.org
Online Course: https://gw-odw.thinkific.com
Web apps: https://gwosc.org/path
Tutorials & Workshops: https://gwosc.org/tutorials





Open Data Workshops

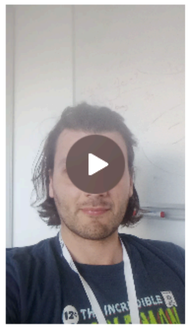
- Annual Event
- Junior scientists prepare material, lecture, and mentor
  - Visibility and experience
- Includes “hands on”
  software examples + challenge problems
- This year: Hybrid and Scalable
- Live Event —> Online course

2022 Open Data Workshop
1000+ Participants
15 Locations + Virtual

Shreejit Jadhav
PhD Student
Inter-University Centre for Astronomy and Astrophysics (IUCAA), Pune, India

Leïla Haegel
Researcher
Astroparticles and Cosmology Laboratory, France

Simone Mastrogiovanni
Postdoc
ARTEMIS, Nice Observatory, France




Pasadena, CA


GW Open Data Workshop
Study Hub
LIGO VIRGO KAGRA
San Luis Potosí, Mexico


Trieste, Italy


Maria Paola Vaccaro
Kumar
Ankur C
Arnab Lahiry
Simon Prunet
Simone Mastrogiovanni [Mentor]
European Virtual Hub


Applications Places System
Zoom Meeting
Recording
Manuel David
Claudia Moreno
César Eduardo Tiznado Alonso
Laura Villegas
Javier M. Antelis
Aldo Alvarez
Ricardo Escobedo
Rafael Hernandez-Jimenez
Erick Madrigal
Jorge Toral
Kevin Sanbilán
José Carlos Martínez Cuevas
MARIO GOMEZ INDA
Johan Villa
Anuar Sánchez
Mariana Lira
Jaime Alejandro Montaño Carrillo
Enrique Herrera Arevalo
Luis Josue Diaz Alvarez
Yadhyra Itzel Estrada
Francisco Ruelas
Lucy Bravo
Josue Prieto
Yuridia Rojas
Guadalajara, Mexico


LIGO Livingston

Software Examples In Your Browser

Jupyter Notebooks

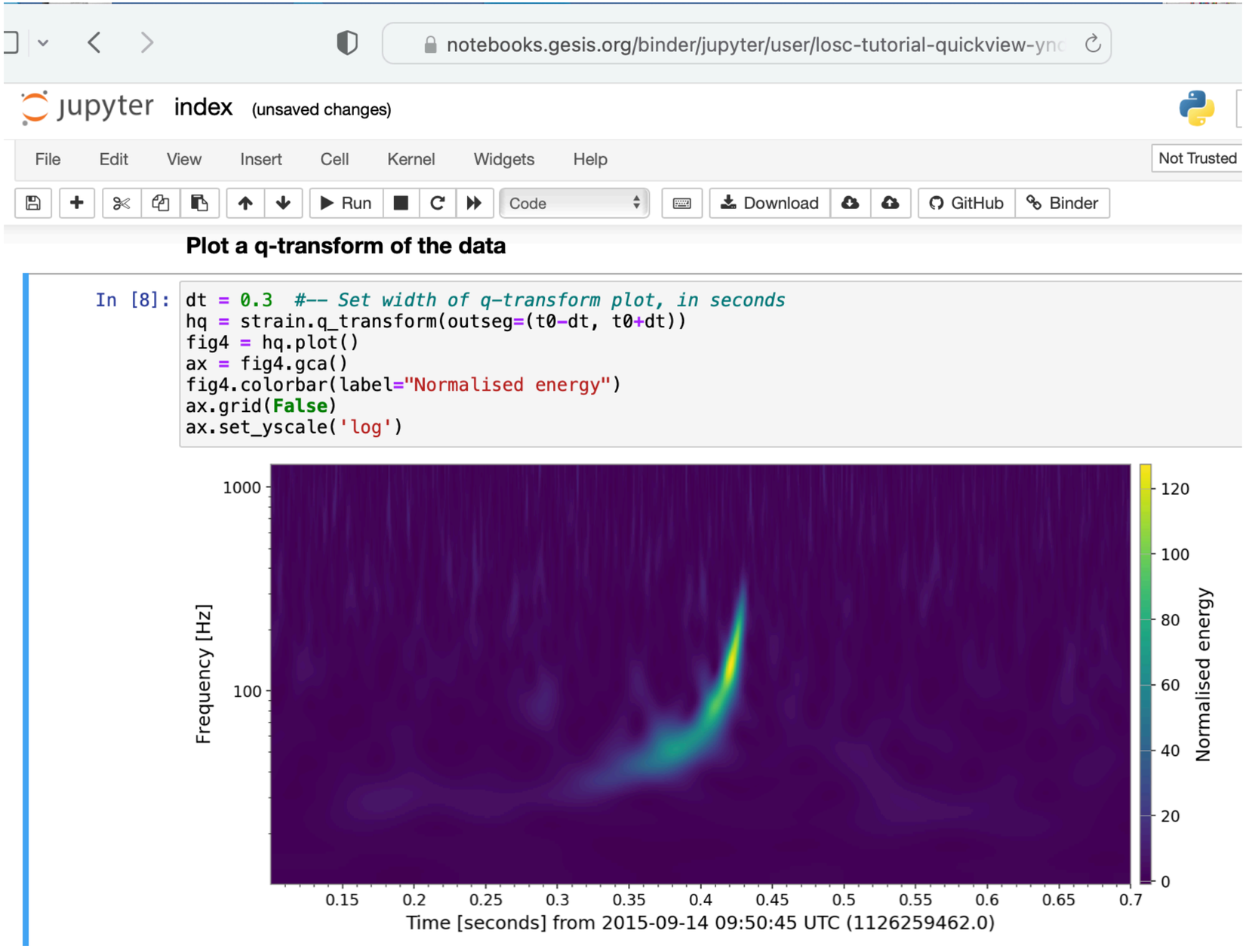
google co-lab
mybinder

Specialized libraries:

`GWpy, pyCBC, bilby`

No installation

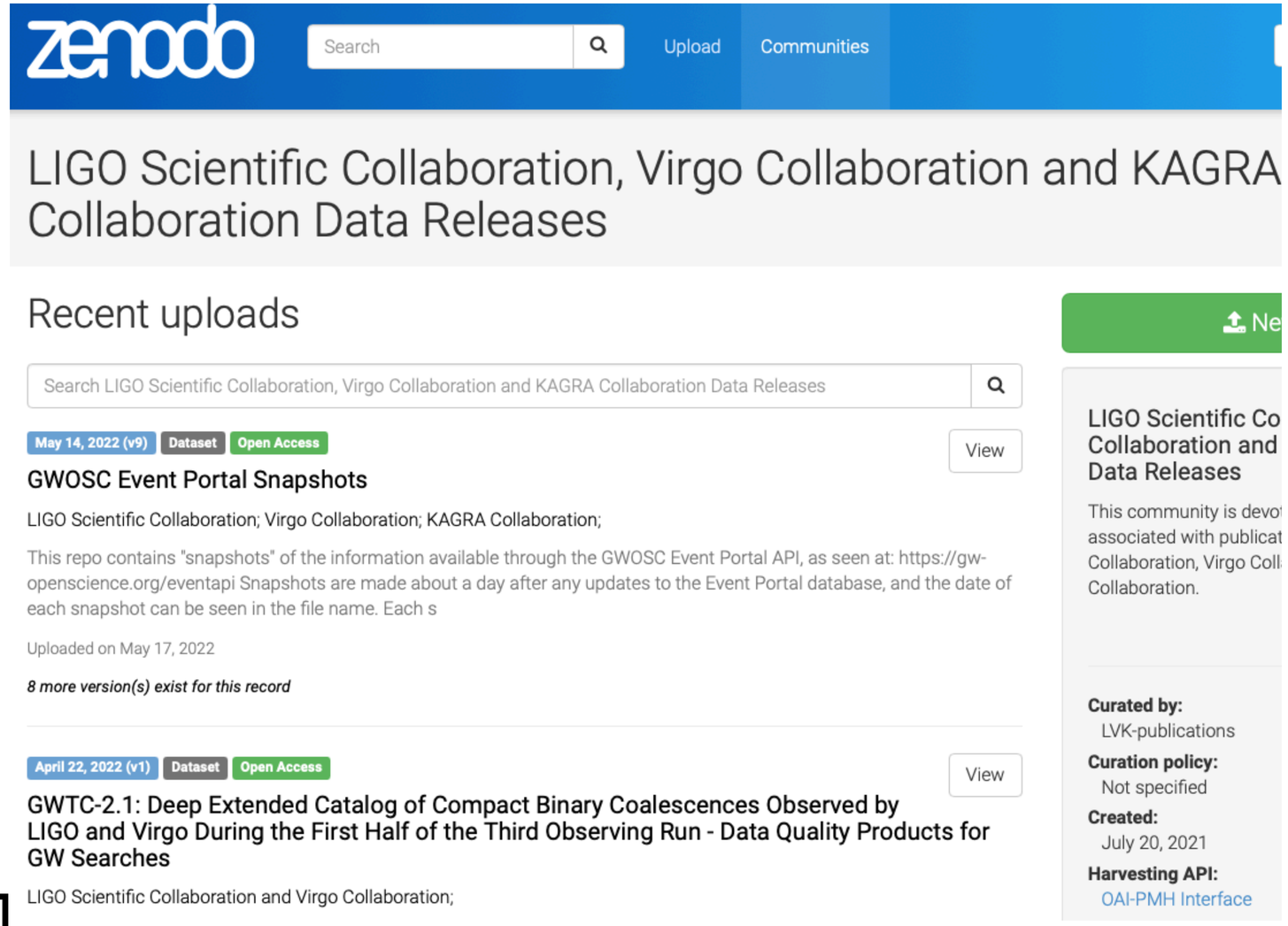
gwosc.org/tutorials



Data Access for Analysis Results

The long tail of public data

- LIGO/Virgo/KAGRA now releases public analysis results in zenodo
  - CERN funded data archive
  - Trigger lists, PE samples, skymaps, etc.
  - LVK community makes these easy to find
  - Authors manage own data



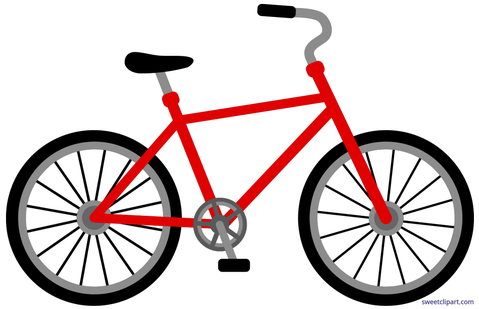
https://zenodo.org/communities/ligo-virgo-kagra/

Summary

- **Solving the “last mile” problem for public data is high stakes!**
  - Important for diversity, equity, and inclusion
  - Improves efficiency, creates synergy across fields
  - Find gaps in resources and create solutions

- **Essential to consider needs of both experts and non-experts**
  - Need lots of data for experts, lots of services & support for non-experts

https://gwosc.org



Thank you!

share.streamlit.io/jkanner/streamlit-dataview/app.py

Select Data Time and Detector

How do you want to find data?

By event name

Select Event

GW151012

Detector

H1

Full sample rate data

Set Plot Parameters

Time Range (seconds)

0.44

Gravitational Wave Quickview

- Use the menu at left to select data and set plot parameters
- Your plots will appear below

GW151012

GPS: `1128678900.4`

Mass 1: `23.2` $M_\odot$

Mass 2: `13.6` $M_\odot$

Network SNR: `10`

Event page: https://gw-osc.org/eventapi/html/event/GW151012

```
Loading data...done!
```

Manage app

Event Catalogs and Queries

GWOSC Event Portal

- Provide easy access to lists of Gravitational Wave Transients
- Web interface: No programming required
- Query by name or physical parameters
- Browse catalogs
- Includes physical parameters, instrument data, analysis results, and documentation
- Scriptable against a REST API

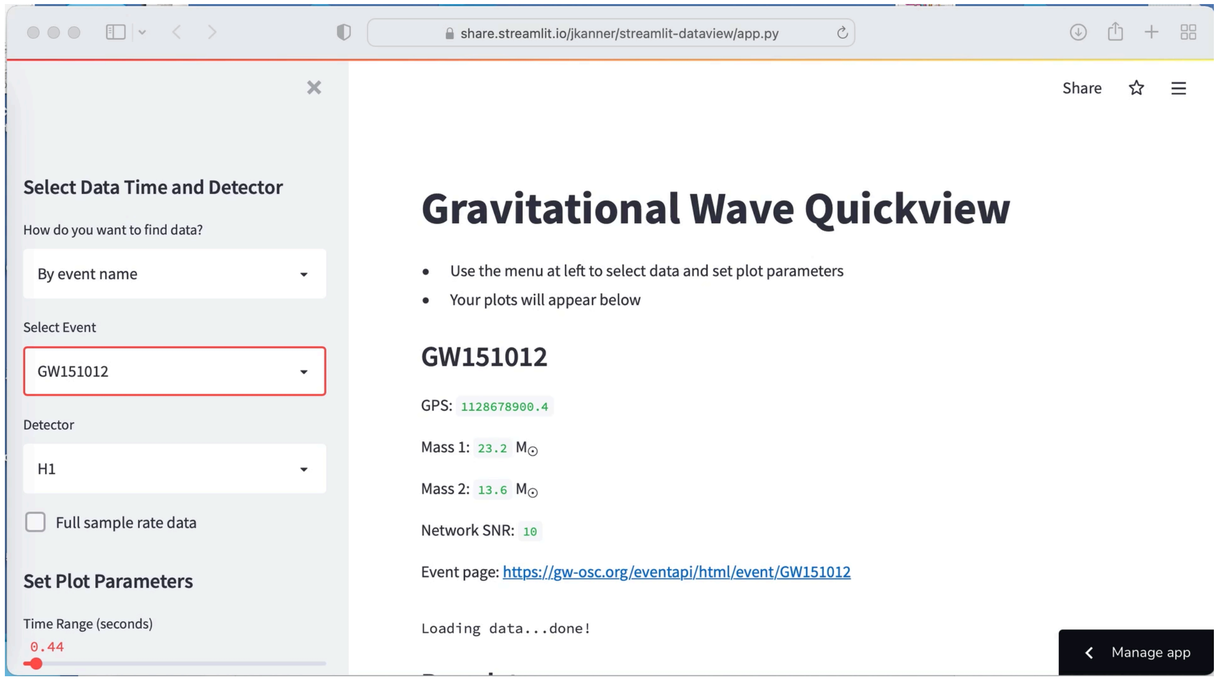
https://gwosc.org/eventapi

Web Apps or GUIs

Remove the need to program!

https://gwosc.org/path

- Plot data with no programming
- “Pre-process” data (whiten, filter, etc.)
- Export common file types (e.g. CSV)
- Introduction to signal processing



- **Common Request:**
  “I’d like to download processed data to in a CSV or text file”

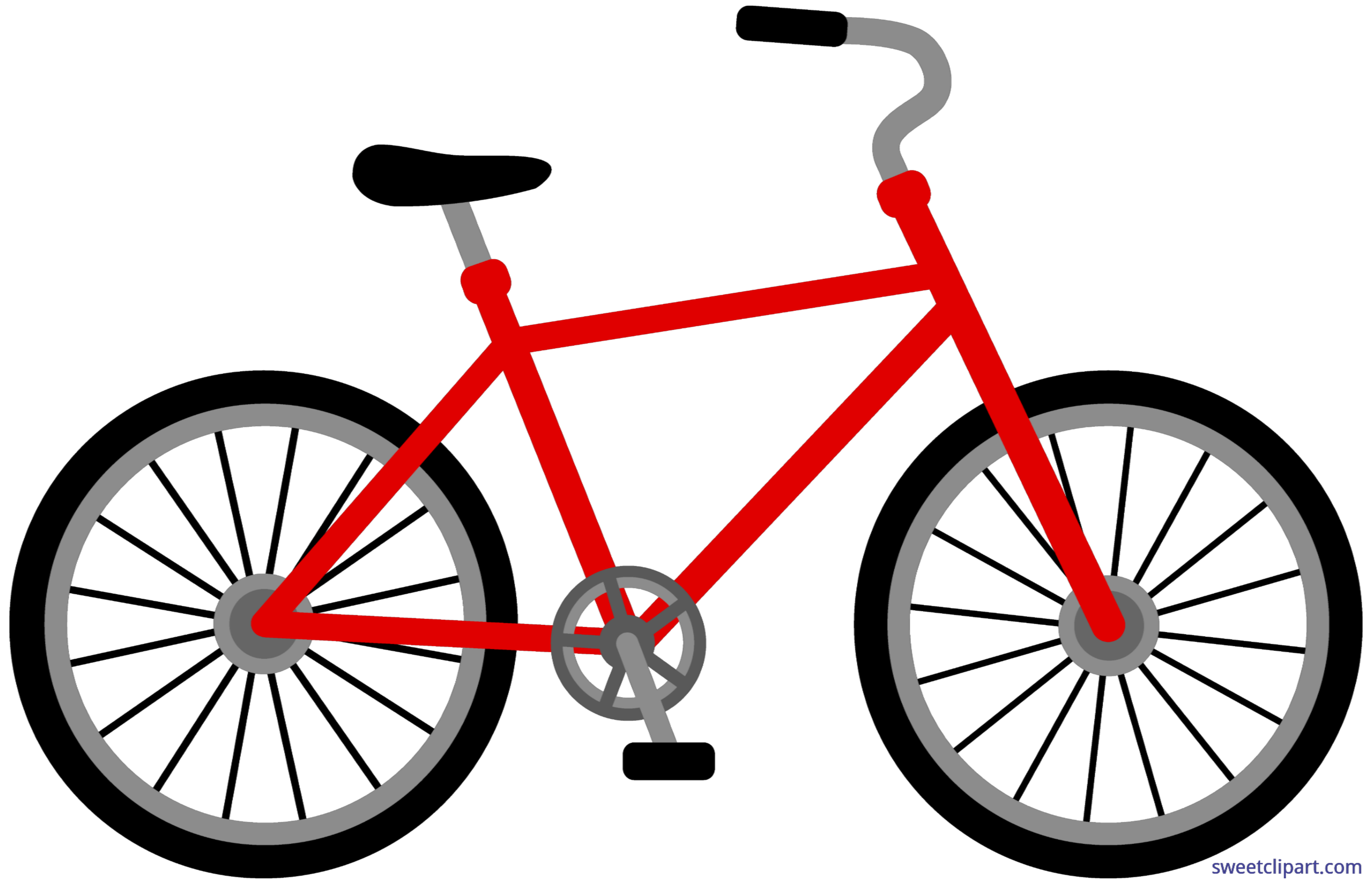
Data Access on GWOSC (Instrument Data)

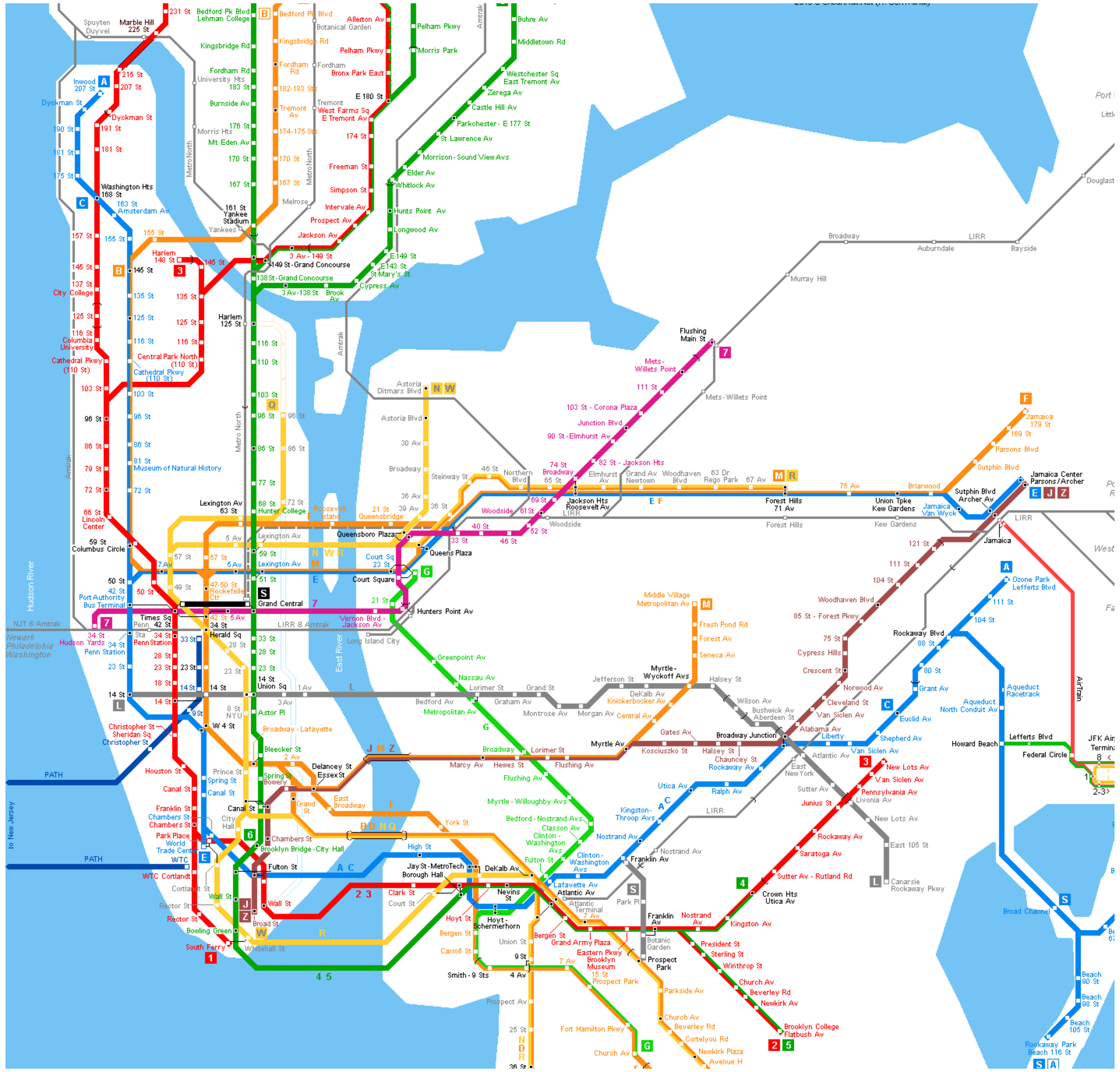
- **Web Access:** Query for data by time or event (HTML or REST API)
  - Easy access for everyone, one file at a time
- **CernVM File System:** Needed for high performance
  - Works well for access by computing clusters
- **Network Data Server (NDS2)**
  - Provides access to data “snippets” - don’t need to download whole file
  - Fast and convenient data access

https://gwosc.org/data

The Last Mile SOLUTION

Transportation





Lyon, France


Seoul, Korea


Padova, Italy


ICTS, Bangalore, India

GW200311_115853

Event Portal

Documentation

Release: GWTC-3-confident

Event UID: GW200311_115853-v1

Names: GW200311_115853

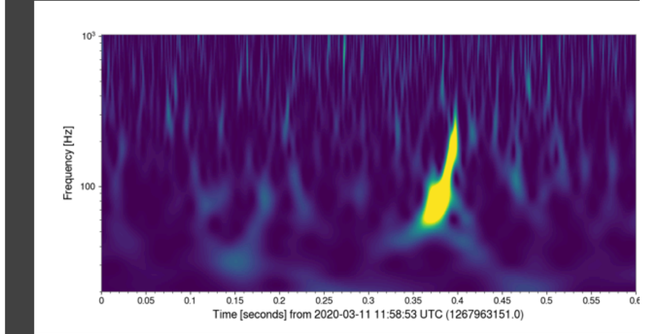
GPS: 1267963151.3

UTC Time: 2020-03-11 11:58

GraceDB: S200311bg

GCN: Notices • Circulars

H1 strain



32sec • 16KHz:	GWF	HDF	TXT
32sec • 4KHz:	GWF	HDF	TXT
4096sec • 16KHz:	GWF	HDF	TXT
4096sec • 4KHz:	GWF	HDF	TXT

SORT: GPS ↓

Name	Version	Release	GPS ↓	Mass 1 ($M_\odot$)	Mass 2 ($M_\odot$)
GW200322_091133	v1	GWTC-3-confident	1268903511.3	34^{+48}_{-18}	$14.0^{+16.8}_{-8.7}$
GW200316_215756	v1	GWTC-3-confident	1268431094.1	$13.1^{+10.2}_{-2.9}$	$7.8^{+1.9}_{-2.9}$
GW200311_115853	v1	GWTC-3-confident	1267963151.3	$34.2^{+6.4}_{-3.8}$	$27.7^{+4.1}_{-5.9}$

https://gwosc.org/eventapi

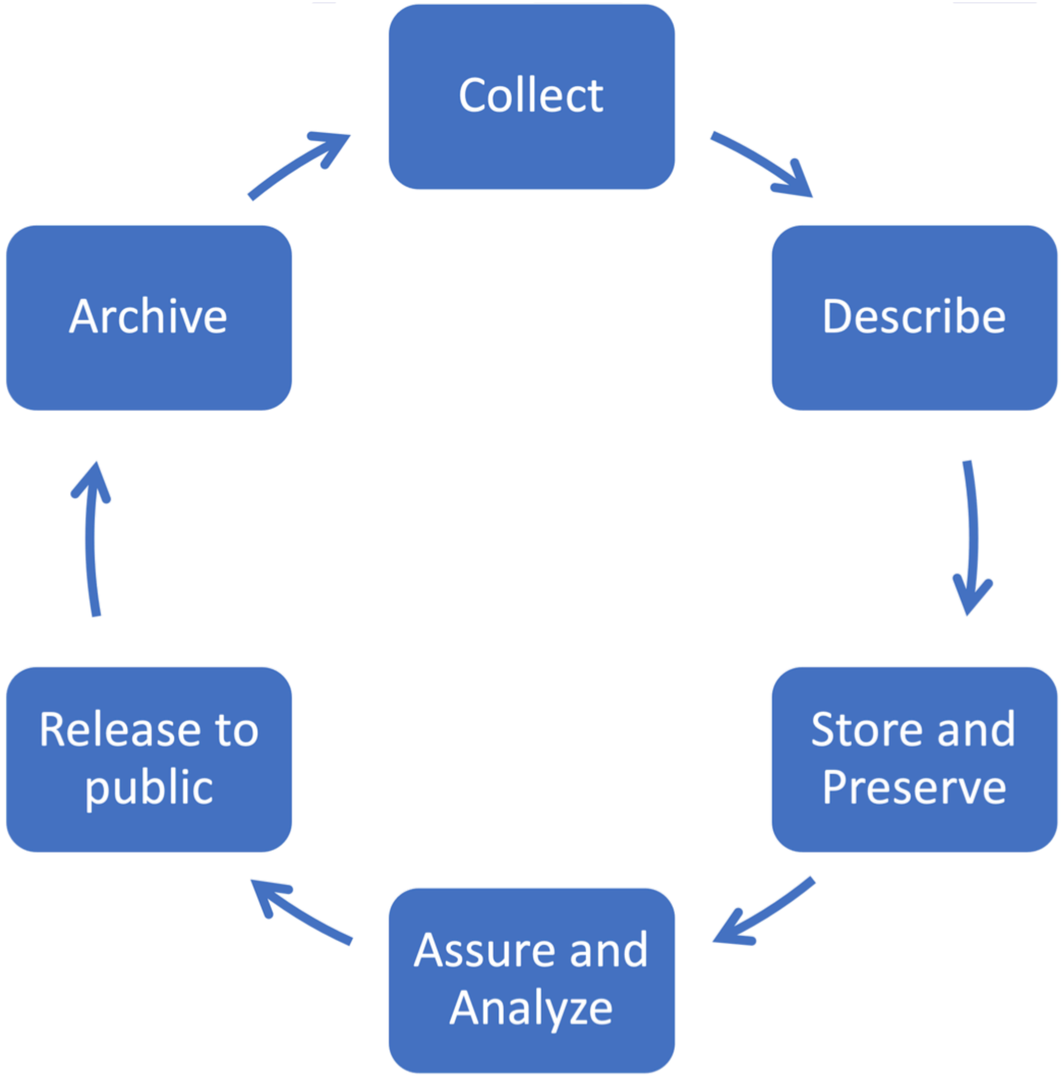
Getting Help

Need to hear from people using data

- GWOSC Help Desk, via e-mail: gwosc@igwn.org
- New: LIGO/Virgo/KAGRA discussion forum: https://ask.igwn.org
  - Vera Rubin Telescope has an active discussion forum, with thousands of posts
- Discussion Board / Help Desk monitored both by GWOSC staff and volunteers in LIGO/Virgo/KAGRA collaboration

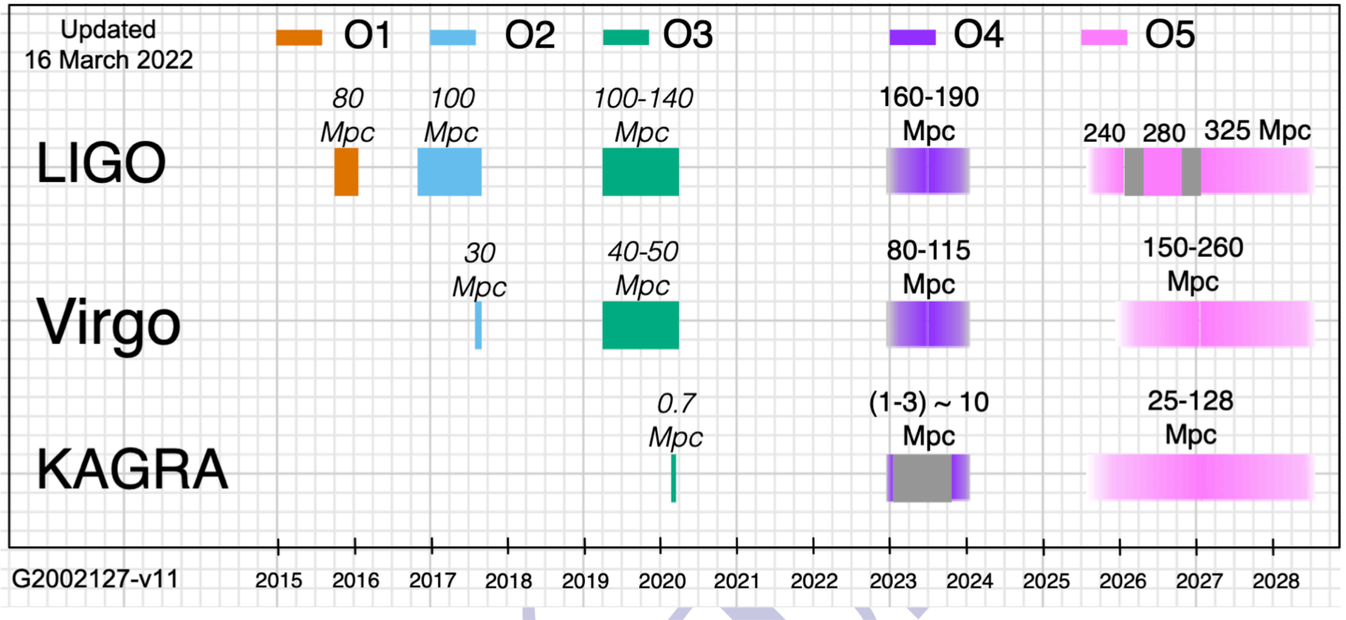
Provide direct support AND Learn about gaps

LIGO Data Life Cycle



Data Collection

- Data collected in a series of observing runs
- “Raw” frames contain 250,000 channels per IFO,
  - ~petabyte per year
- Calibrated STRAIN in own frames
  - ~terabytes per year
  - 99% of astrophysics in 1% of data



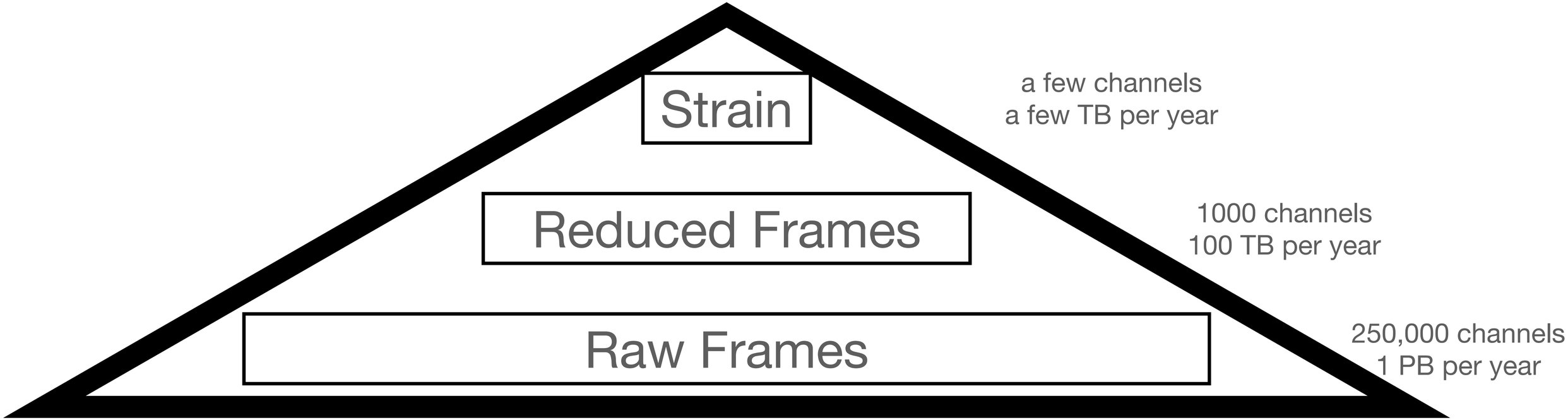
Describe

- All data stored in GWF files, with self-describing meta-data for each channel
  - Defined in *International Gravitational Wave Detectors (IGWD)* data format, established 1997 (https://dcc.ligo.org/LIGO-T970130/public)
- Acronyms for decoding: https://dcc.ligo.org/LIGO-M080375-v1/public

```
>>> from gwpy.timeseries import TimeSeries
>>> data = TimeSeries.fetch('H1:DCS-CALIB_STRAIN_CLEAN_C01_AR', start=1240559616, end=1240559626, host='losc-nds.ligo.org')
>>> print(data)
TimeSeries([3.45188295e-20, 5.52788219e-20, 6.79233525e-20, ...,
            6.73696363e-20, 3.88823380e-20, 4.08627208e-20]
           unit: strain,
           t0: 1240559616.0 s,
           dt: 6.103515625e-05 s,
           name: H1:DCS-CALIB_STRAIN_CLEAN_C01_AR,
           channel: H1:DCS-CALIB_STRAIN_CLEAN_C01_AR)
```

Store and preserve

- Raw frames during observing runs preserved for life of lab
- Raw frames between observing runs “reduced” after set time period
- All data stored at multiple locations



FAIR Public Data Release

https://gwosc.org

Gravitational Wave Open Science Center

- FINDABLE: Data are easily discoverable through the GWOSC web server, with human readable and machine readable options
- ACCESSIBLE: Strain data can be accessed via http, CVM-FS, or NDS2
- INTEROPERABLE: Available in both GWF and HDF5 formats. Identical formats for LIGO, Virgo, & KAGRA
- REUSABLE: Open source software, documentation, tutorials, and workshops

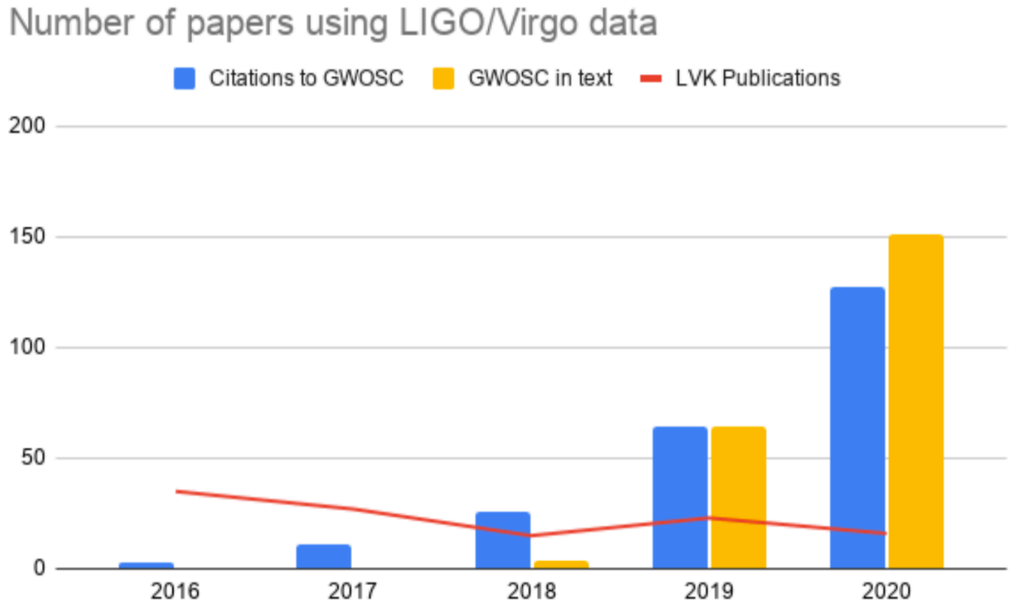
Impacts of Open Data

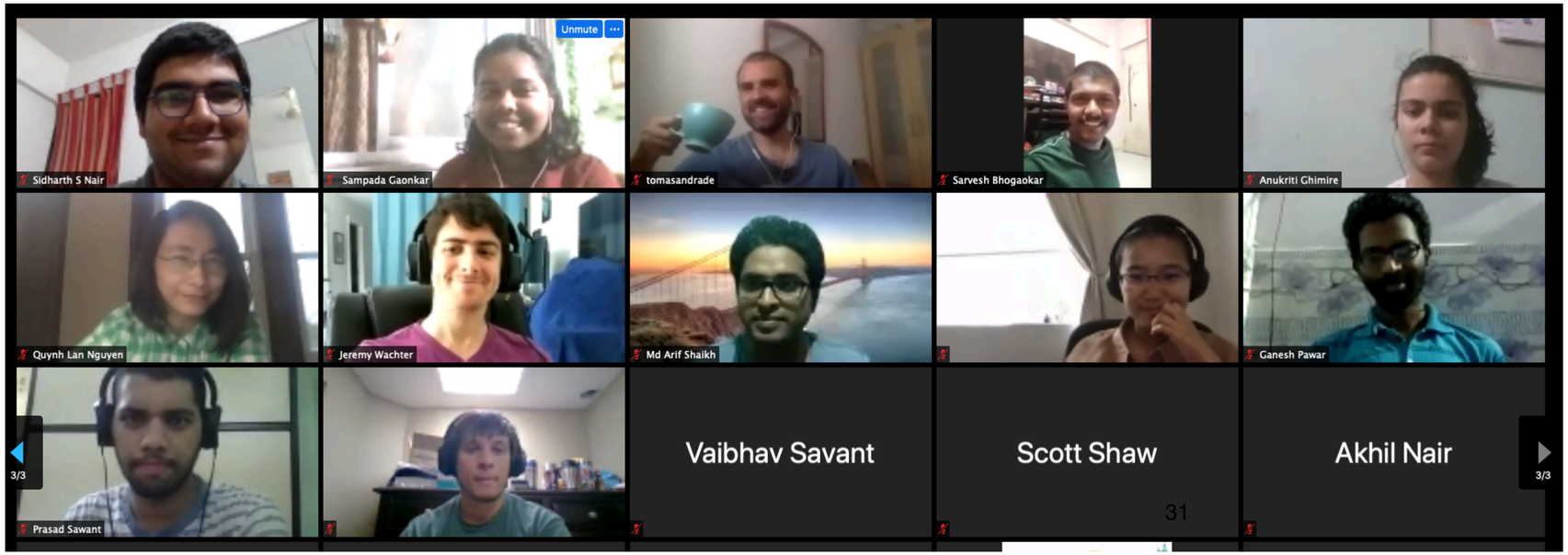
Around 6,000 visitors (12,000 sessions)
to GWOSC each month

Over a million strain file downloads over 6 months

250 Papers in 2 years (2020 + 2021)

Open Data Workshops with hundreds of participants

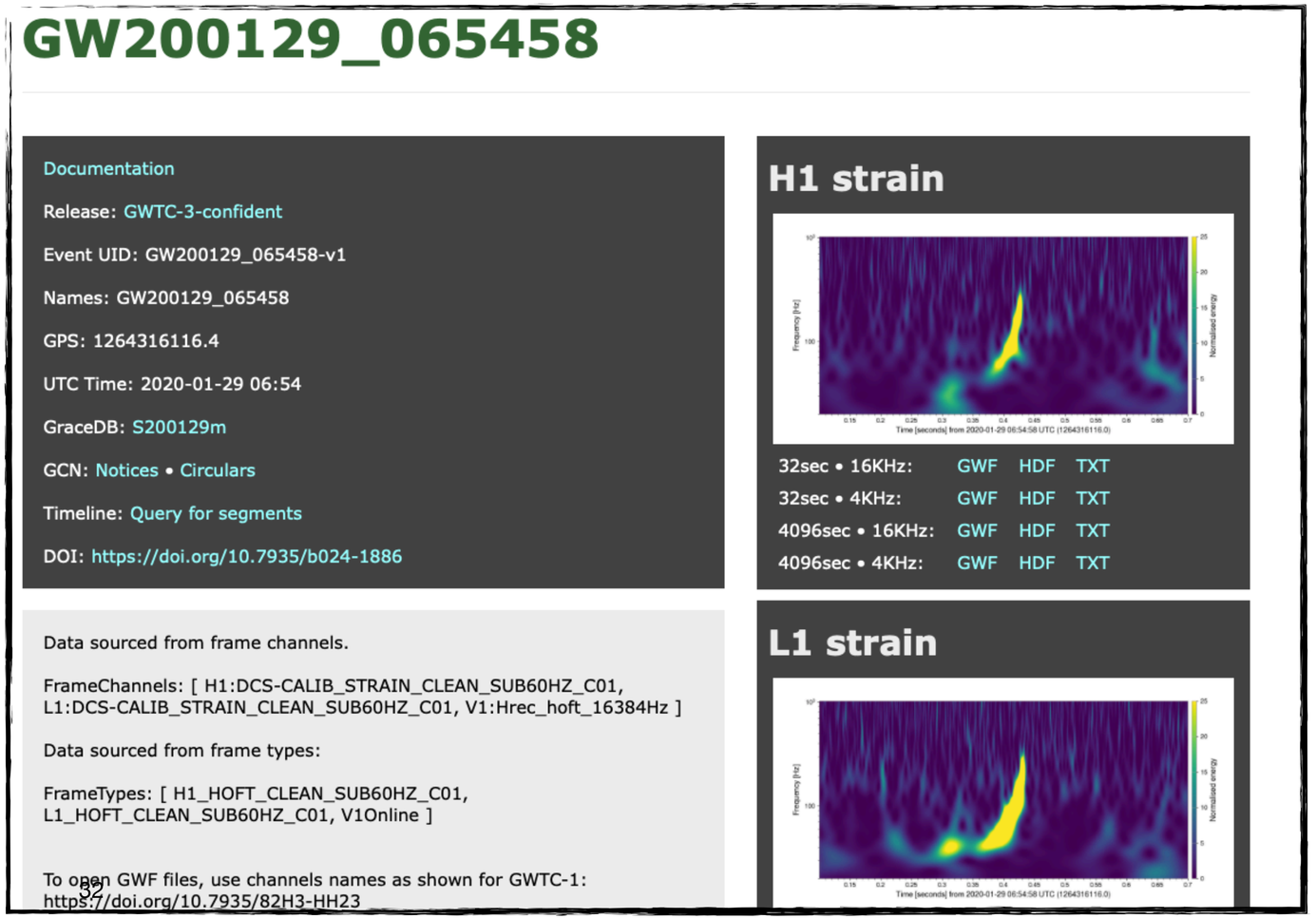




GWOSC Event Portal

- Includes catalogs of LVK discoveries, with PE results and strain data
- Reflects only published results
- Includes “GWTC” - a cumulative catalog of all LVK detections
- Snapshots archived in zenodo to preserve history

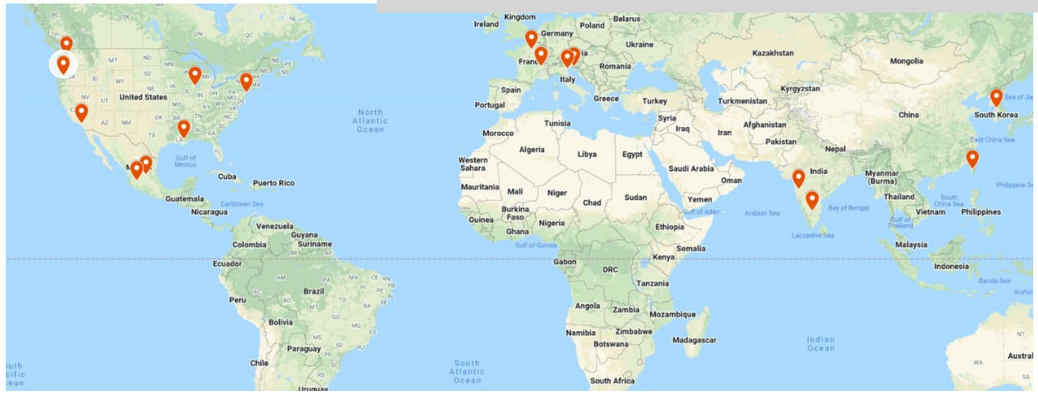
https://gwosc.org/eventapi

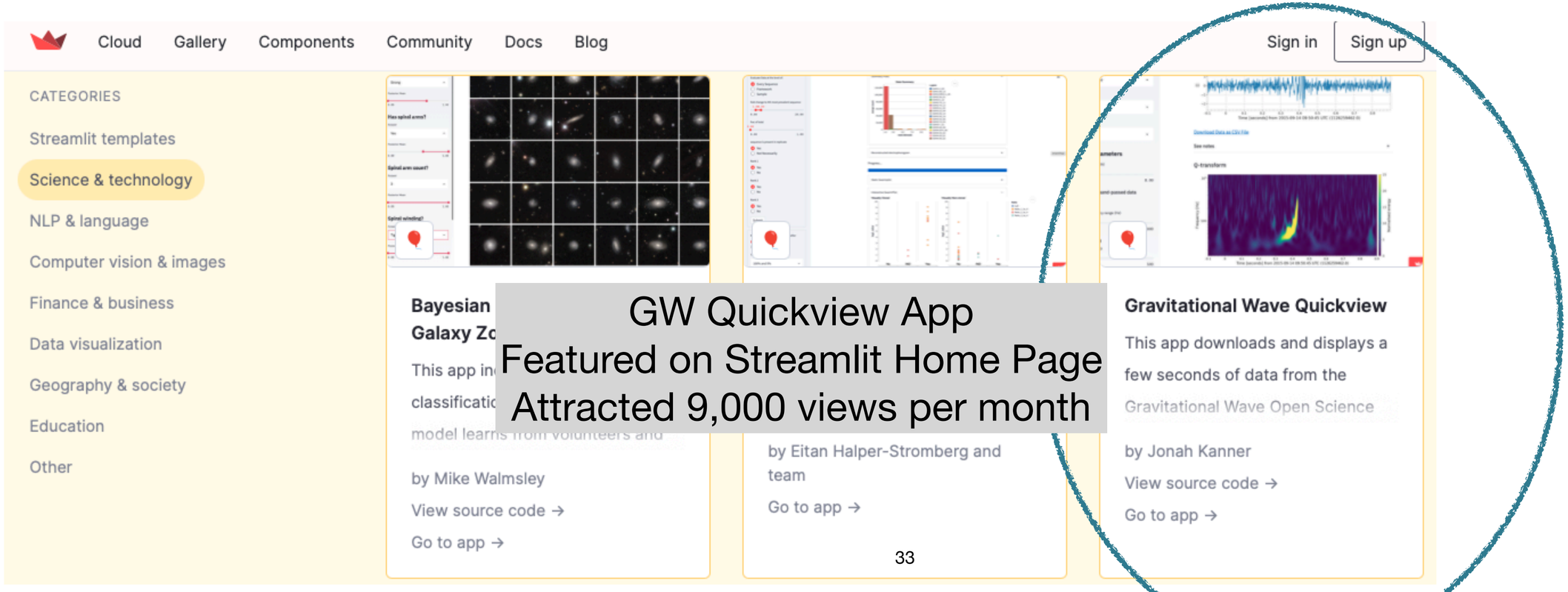


Supporting the Community

2022 Open Data Workshop
1000+ Participants
15 Locations + Virtual

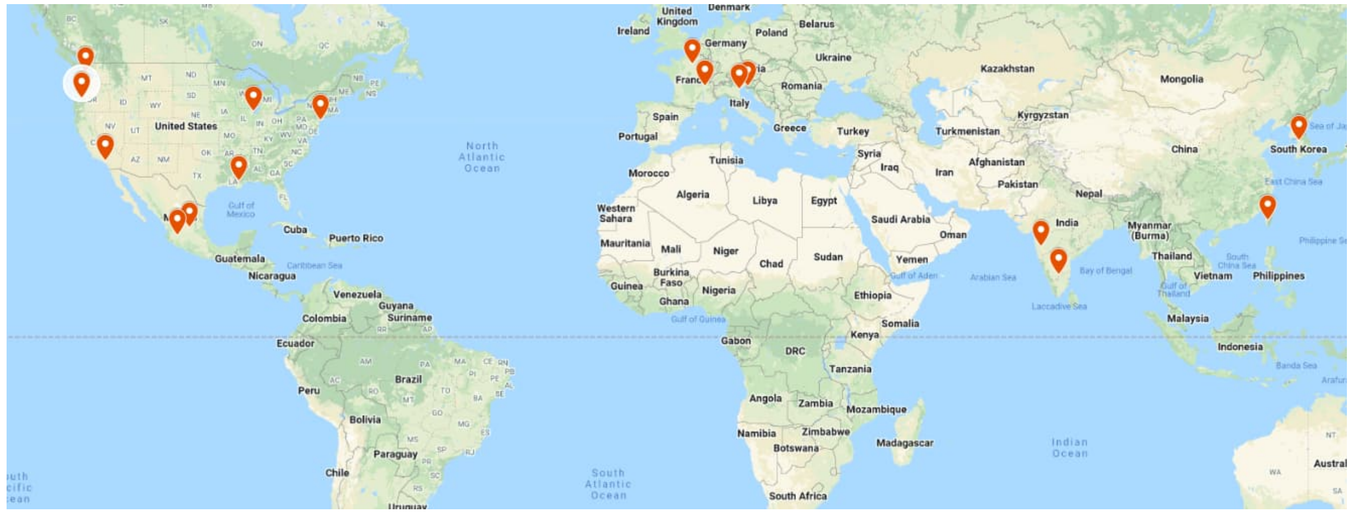
Discussion forum: https://ask.igwn.org
E-mail help desk: gwosc@igwn.org
Online Course: https://gw-odw.thinkific.com
Web apps: https://gwosc.org/path
Tutorials & Workshops: https://gwosc.org/tutorials





GW Quickview App
Featured on Streamlit Home Page
Attracted 9,000 views per month


New York


United States
Cuba
Puerto Rico
Guatemala
Nicaragua
Venezuela
Guyana
Suriname
Colombia
Ecuador
Brazil
Peru
Bolivia
Paraguay
Chile
North Atlantic Ocean
South Atlantic Ocean
United Kingdom
Ireland
Denmark
Belarus
Poland
Germany
Ukraine
Kazakhstan
Mongolia
Romania
Italy
Spain
Portugal
Greece
Turkey
Turkmenistan
Kyrgyzstan
China
South Korea
Tunisia
Syria
Iraq
Iran
Afghanistan
Pakistan
Nepal
Morocco
Algeria
Libya
Egypt
Western Sahara
Saudi Arabia
India
Myanmar (Burma)
Mauritania
Mali
Niger
Chad
Sudan
Oman
Yemen
Thailand
Vietnam
Philippines
Burkina Faso
Guinea
Nigeria
Ghana
Ethiopia
Somalia
Kenya
Malaysia
Indonesia
Gabon
DRC
Tanzania
Angola
Zambia
Mozambique
Namibia
Zimbabwe
Botswana
Madagascar
South Africa
Arabian Sea
Bay of Bengal
Indian Ocean


Paris


Evanston, IL