Gravitational Wave Data: The Last Mile

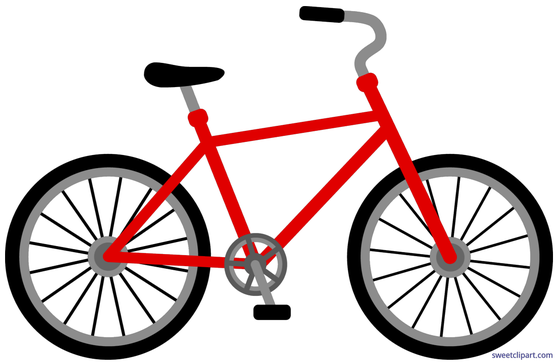
Jonah Kanner
LIGO Lab, Caltech

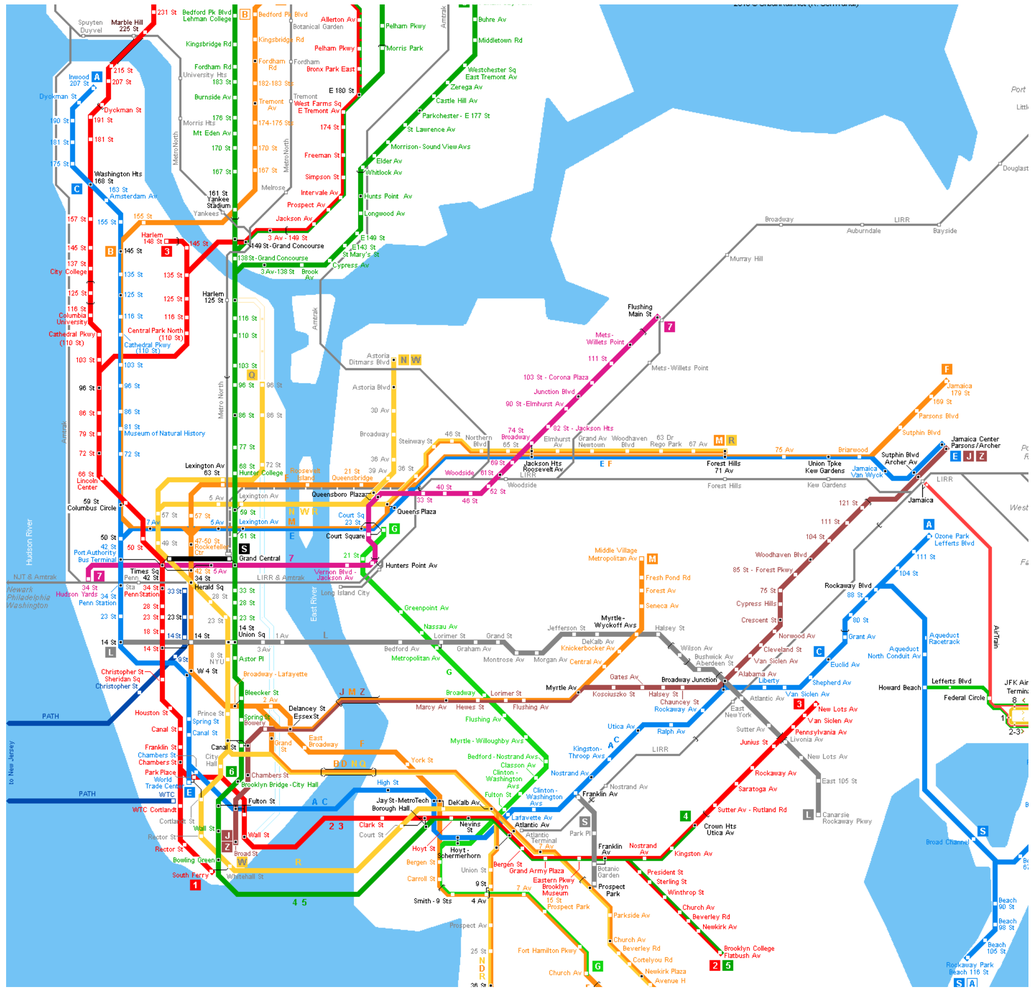
June 15, 2022 - G2200946-v5

The Last Mile Problem

Transportation

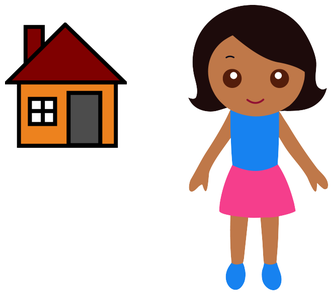
Last Mile

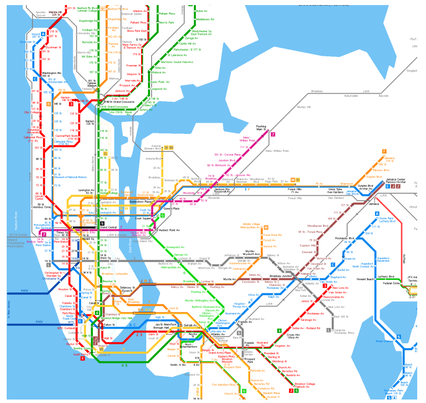




The Last Mile Problem

Public Data





Expert Networks

"Last Mile"

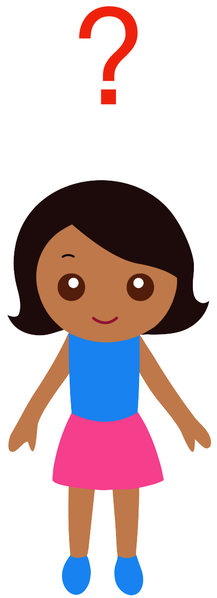
High School Student
Undergraduates
Grad Students
Experts in other field
Amateurs
Artists

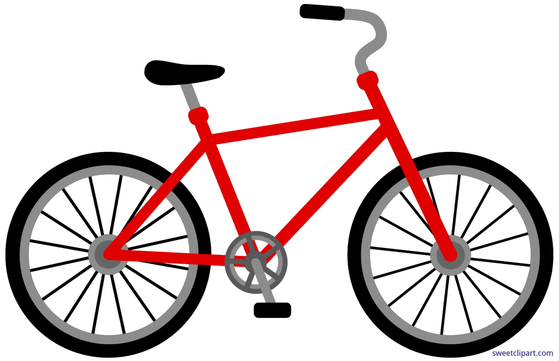
Gaps in:
Access
Knowledge
Resources

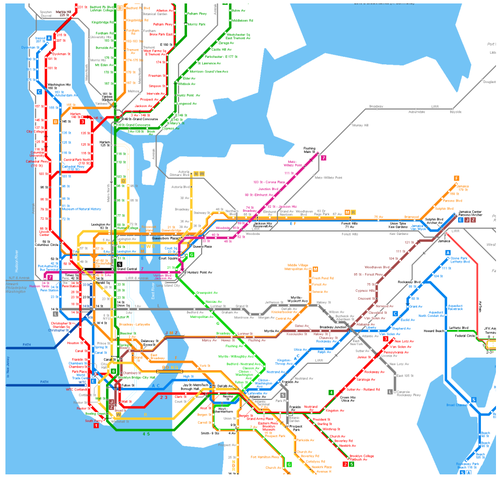
Data
Software
Journal Articles
Conferences
Colleagues

The Last Mile Problem

High Stakes







Diversity, Equity, and Inclusion

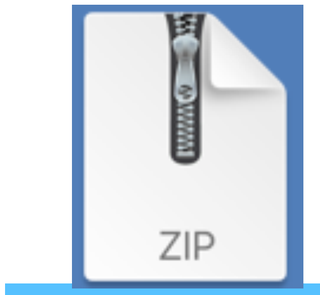
Efficiency and Productivity

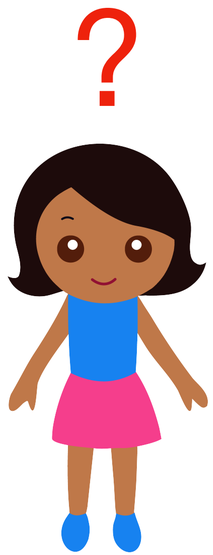
Synergy

Climate Change

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?



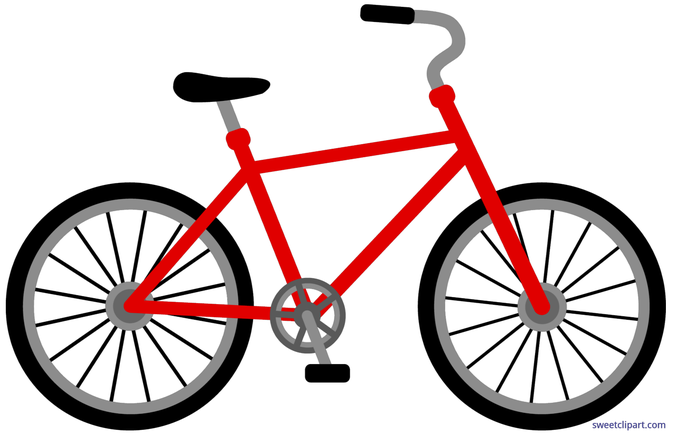


Access Gaps for LIGO Data (circa 2014)

- 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

Mountain of data

and the garden on top

Used broadly
Synergy with other projects

Mostly used internally

Results

a few TB per year

Strain

a few channels
a few TB per year

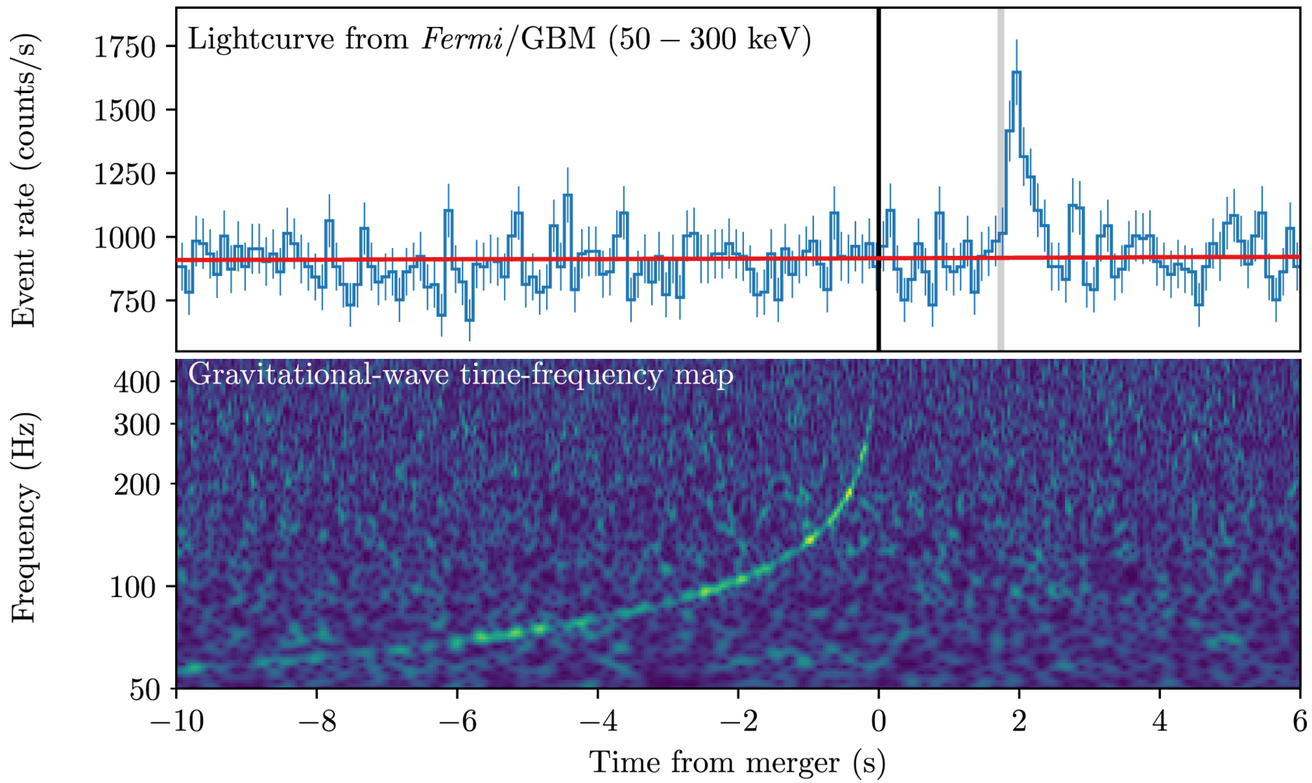
Reduced Frames

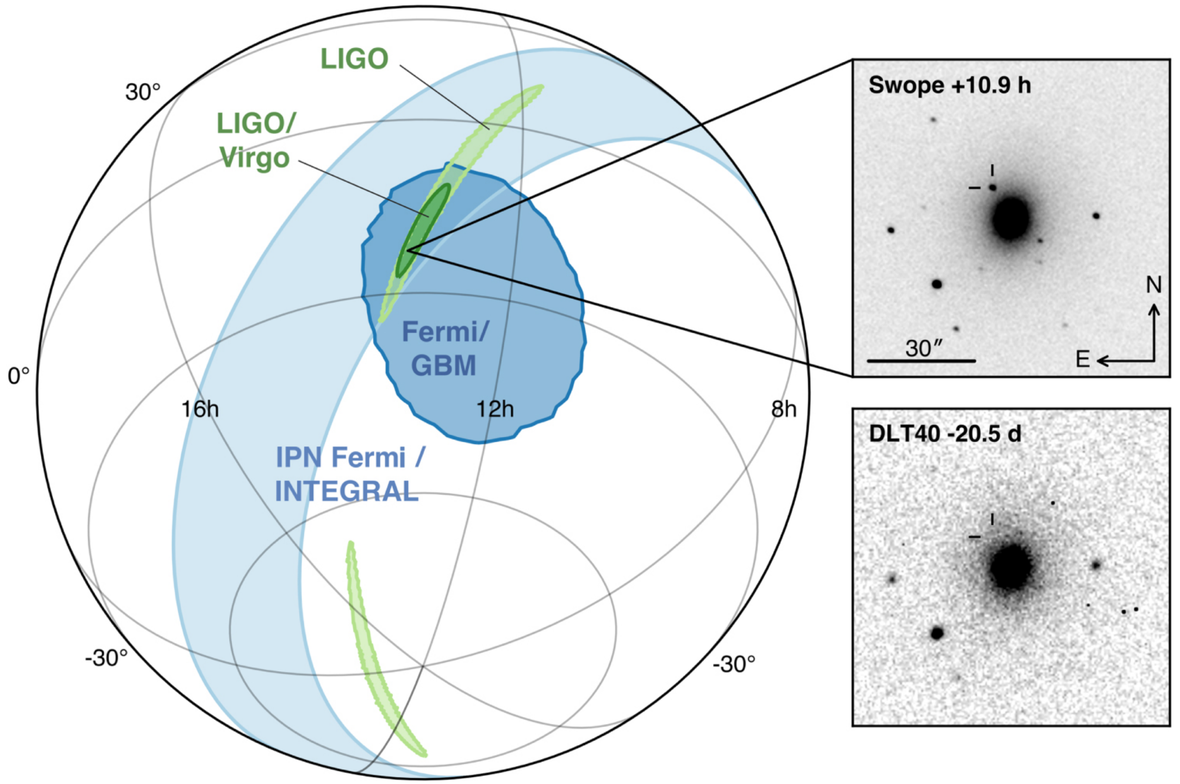
1000 channels
100 TB per year

Raw Frames

250,000 channels
1 PB per year

Synergy and Multi-messenger Astrophysics





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

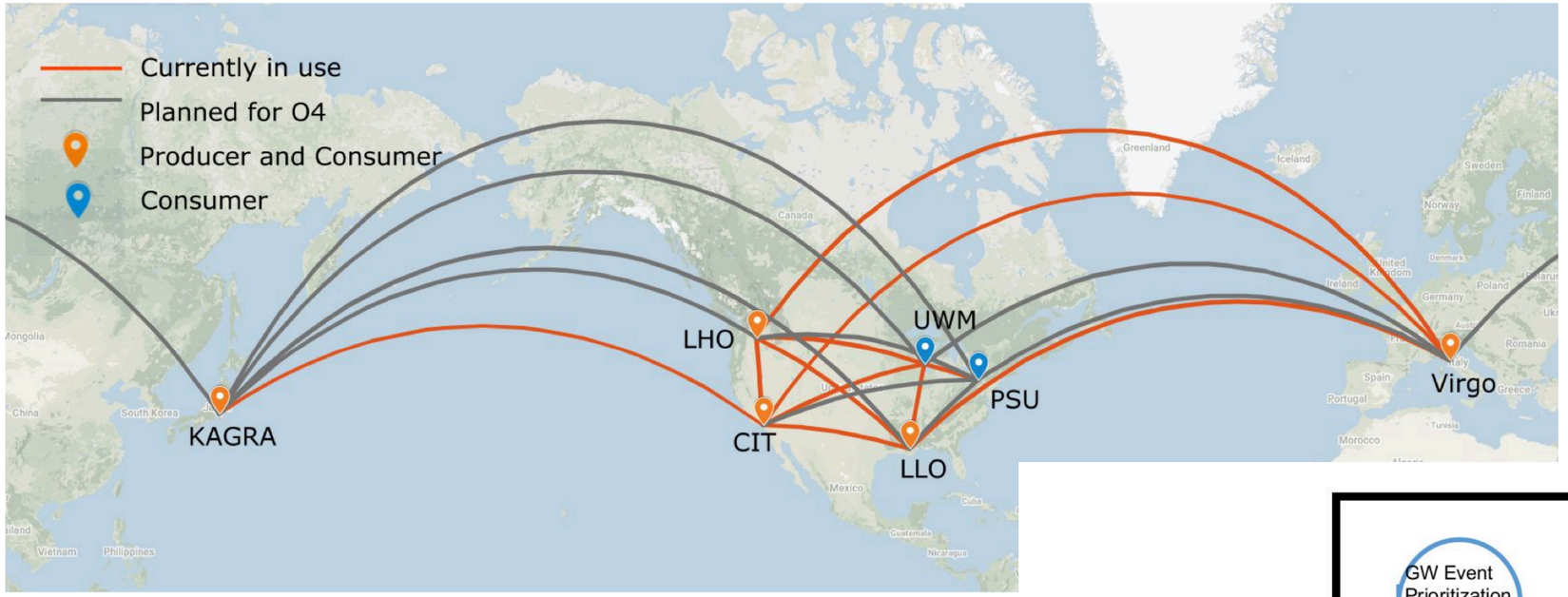
Results public within minutes

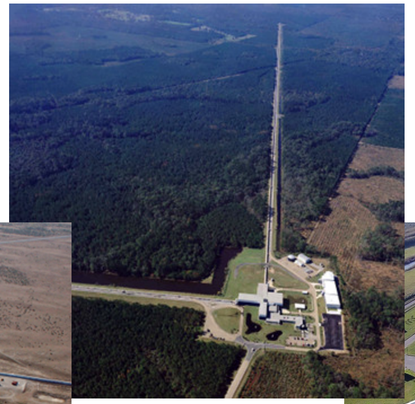
Telescopes perform follow-up observations

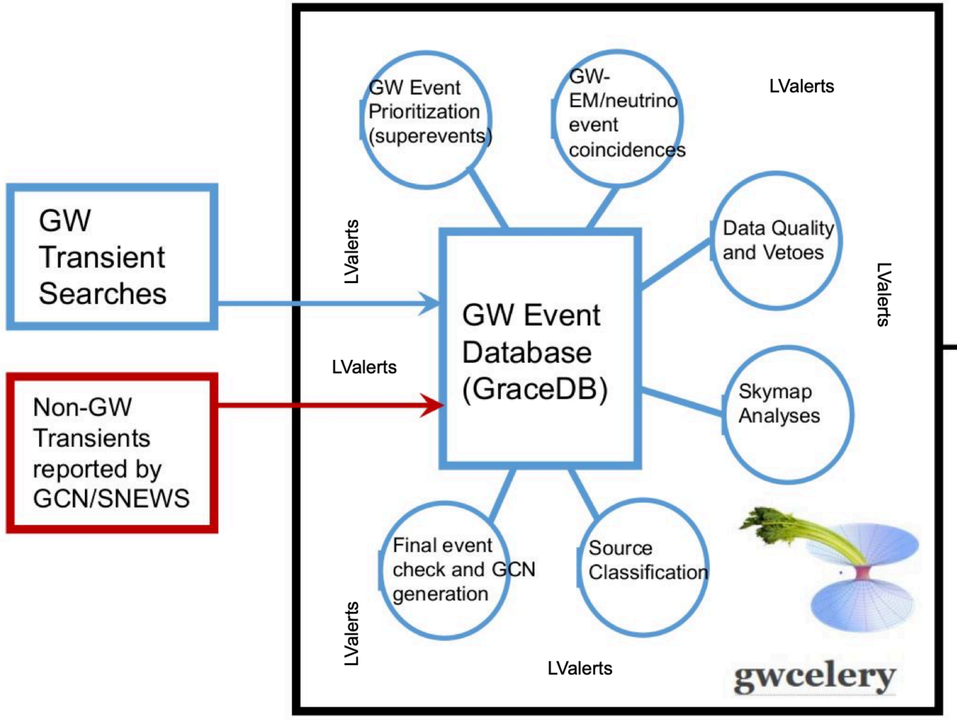
GW170817
"Most Observed Transient"
1st Observation of a BNS Merger

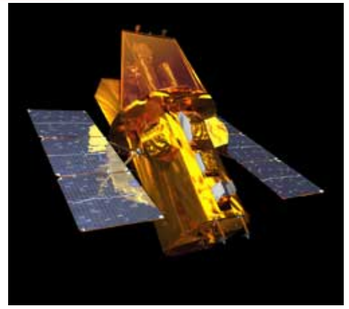
B. P. Abbott *et al* 2017 *ApJL* **848** L12

Low Latency Data Pipeline



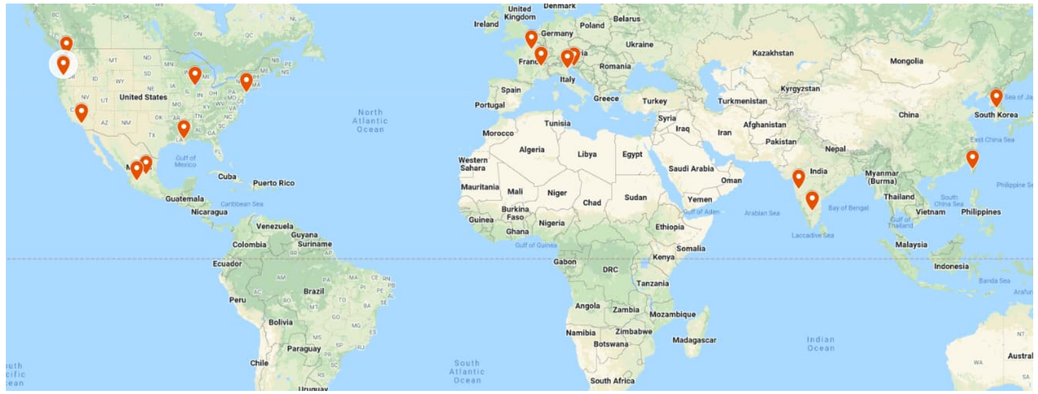


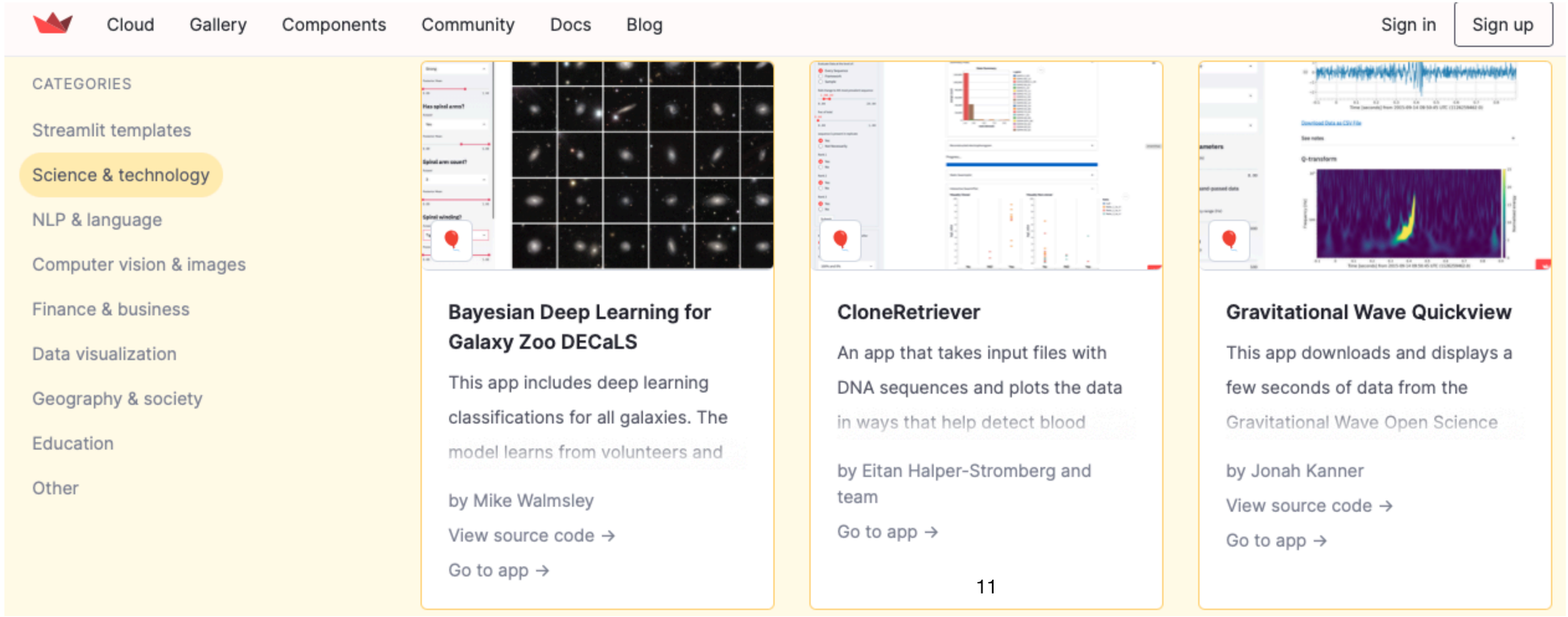




Supporting the Community

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

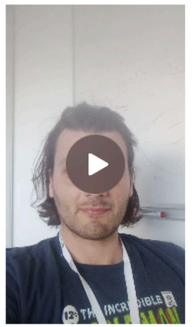
- Annual Event
- Junior scientists prepare material, lecture, and mentor
  - Visibility and experience
- Includes “hands on”
  software examples + challenge problems
- This year: Hybrid and Scalable
- Live Event —> Online course

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 Sanbrán
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

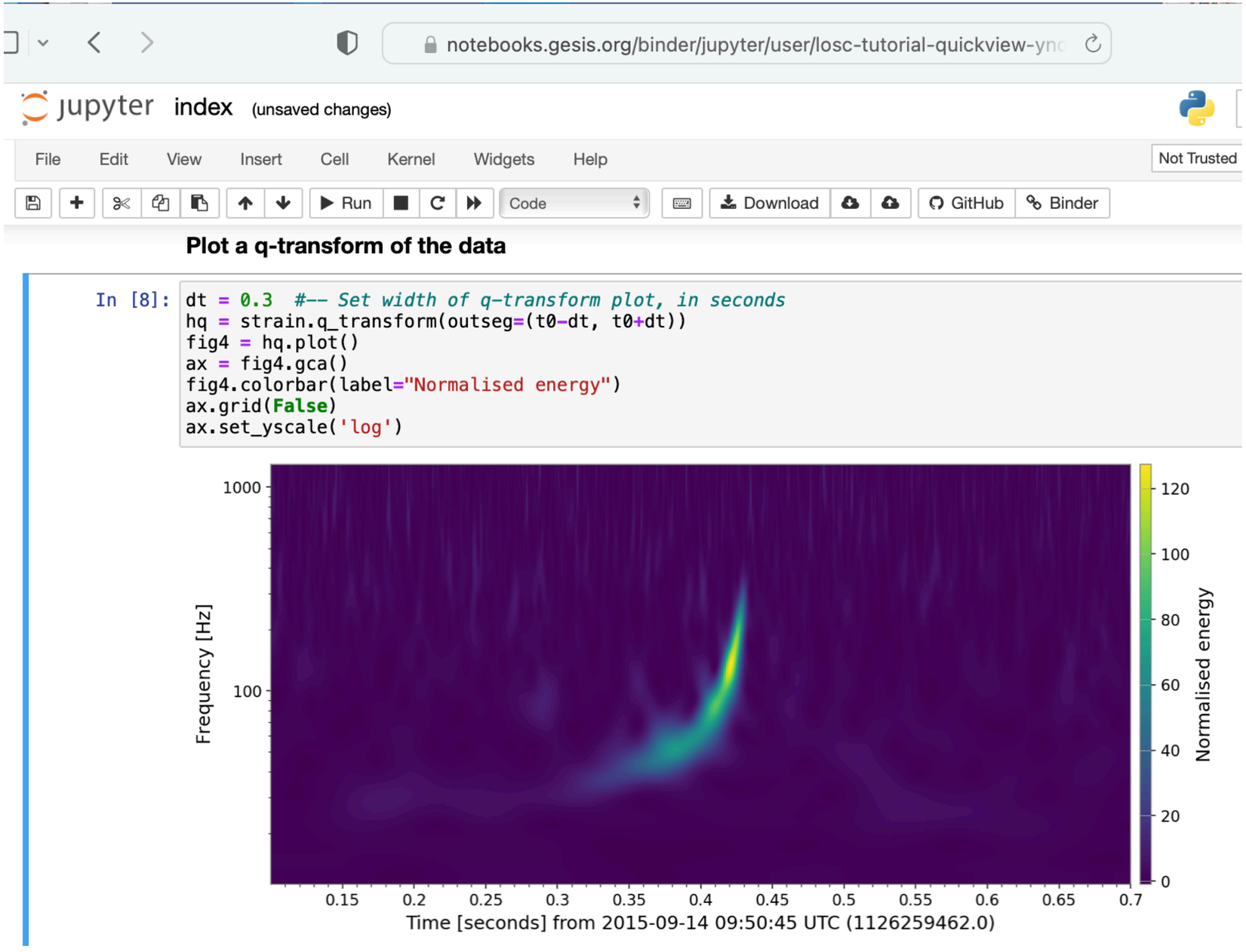
google co-lab
mybinder

Specialized libraries:

`GWpy, pyCBC, bilby`

No installation

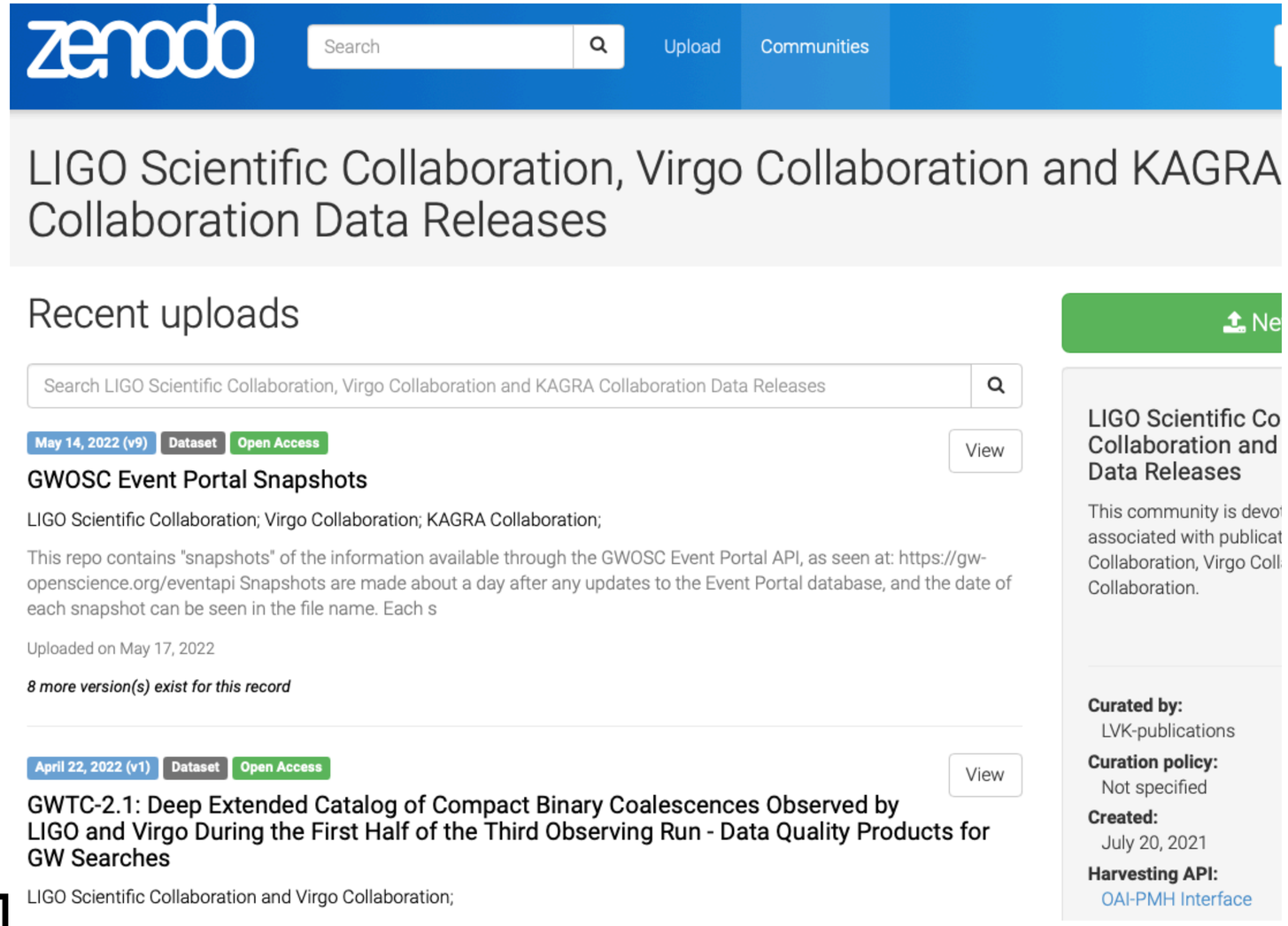
gwosc.org/tutorials



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



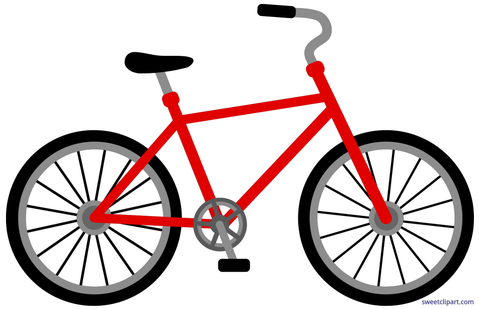
https://zenodo.org/communities/ligo-virgo-kagra/

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

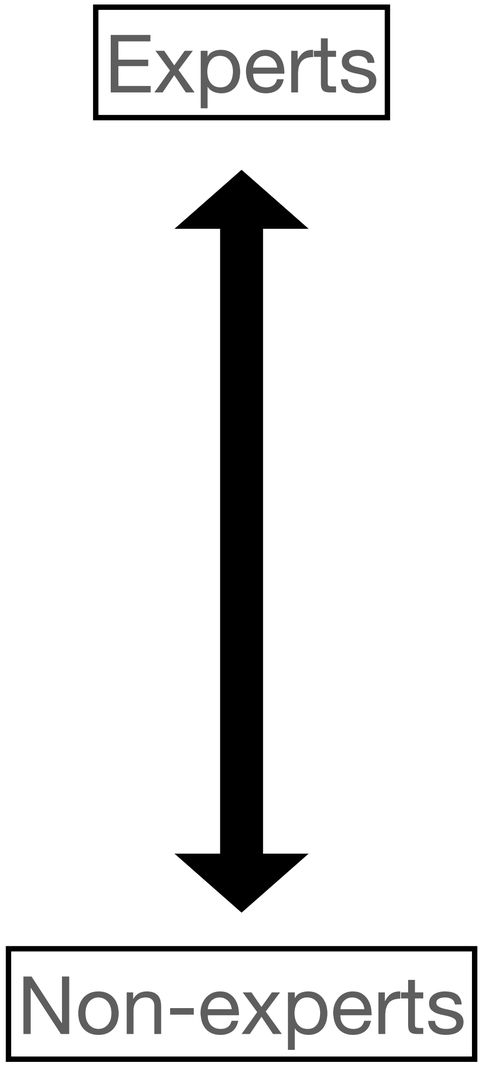
- **Essential to consider needs of both experts and non-experts**
  - Need lots of data for experts, lots of services & support for non-experts

https://gwosc.org



Thank you!

Diversity of Experience

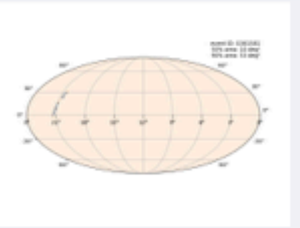
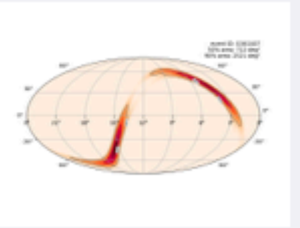
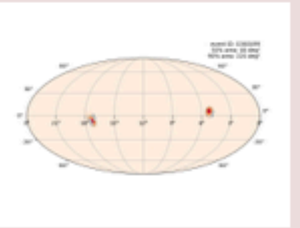
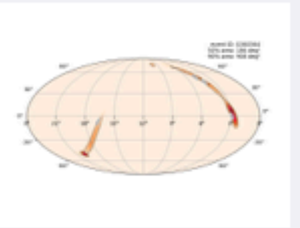


Want lots of data
High demands on data access
Computers access data (API)
Details matter

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

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

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!
```

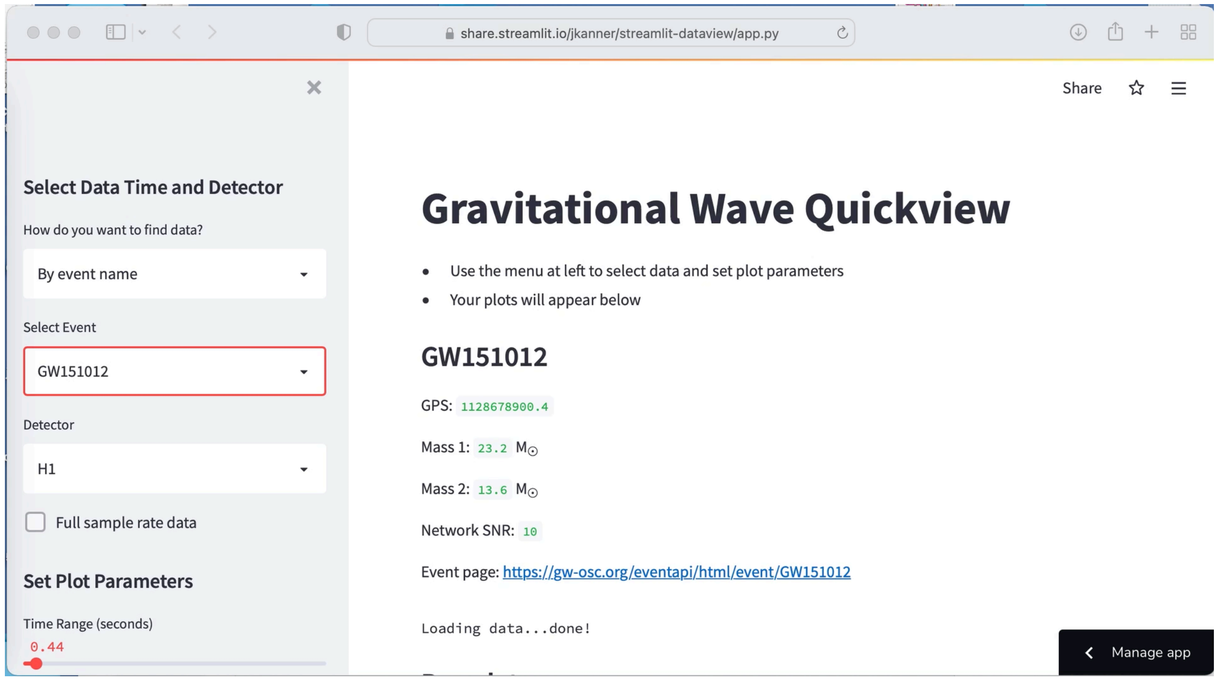
Manage app

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”

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

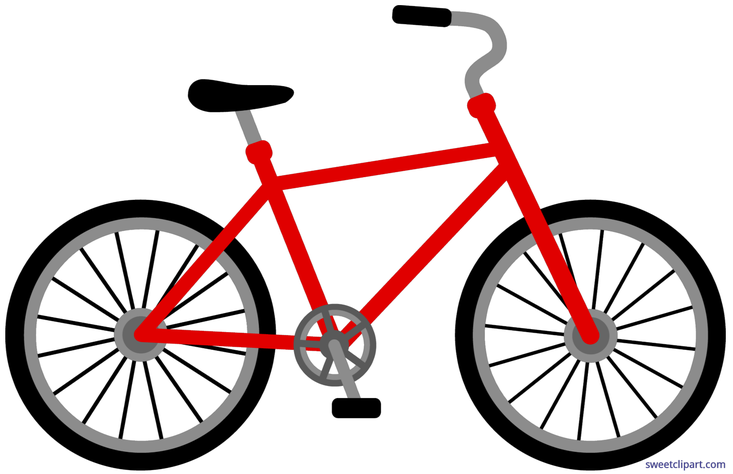
Data Access on GWOSC (Instrument Data)

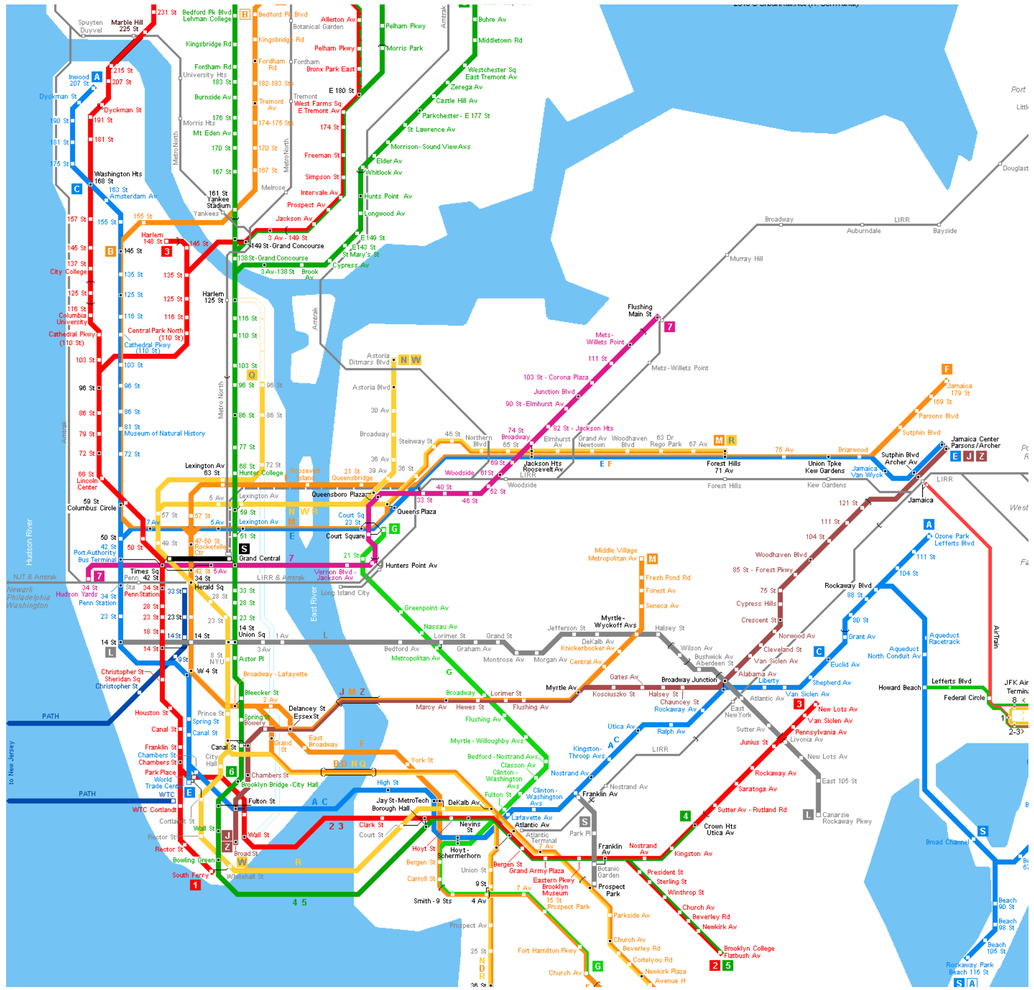
- **Web Access:** Query for data by time or event (HTML or REST API)
  - Easy access for everyone, one file at a time
- **CernVM File System:** Needed for high performance
  - Works well for access by computing clusters
- **Network Data Server (NDS2)**
  - Provides access to data “snippets” - don’t need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

https://gwosc.org/data

The Last Mile SOLUTION

Transportation




Marble Hill 225 St
231 St
Bedford Pk Blvd Lehman College
Kingsbridge Rd
Fordham Rd
Burnside Av
Inwood 207 St
Dyckman St
191 St
181 St
Washington Hts 168 St
163 St Amsterdam Av
155 St
Harlem 148 St
145 St
137 St City College
125 St
116 St Columbia University
Cathedral Pkwy (110 St)
103 St
96 St
86 St
81 St Museum of Natural History
72 St
66 St Lincoln Center
59 St Columbus Circle
50 St
42 St Port Authority Bus Terminal
Times Sq 42 St
34 St Penn Station
Hudson Yards
28 St
23 St
18 St
14 St
8 St NYU
Christopher St Sheridan Sq
Houston St
Canal St
Franklin St
Chambers St
Park Place
World Trade Center
WTC Cortlandt
Cortlandt St
Rector St
Bowling Green
South Ferry
Whitehall St
Wall St
Fulton St
Chambers St Brooklyn Bridge - City Hall
Clark St
High St
Jay St - MetroTech
DeKalb Av
Lafayette Av
Atlantic Av Barclays Ctr
Nevins St
Hoyt - Schermerhorn
Bergen St
Carroll St
Union St
9 St
4 Av
Smith - 9 Sts
7 Av
Grand Army Plaza
Eastern Pkwy Brooklyn Museum
15 St Prospect Park
Prospect Park
Parkside Av
Church Av
Beverley Rd
Cortelyou Rd
Newkirk Plaza
Avenue H
Fort Hamilton Pkwy
Church Av
Newkirk Av
Winthrop St
Sterling St
President St
Kingston Av
Crown Hts Utica Av
Sutter Av - Rutland Rd
Saratoga Av
Rockaway Av
Junius St
Pennsylvania Av
Van Siclen Av
New Lots Av
East 105 St
Canarsie Rockaway Pkwy
Broad Channel
Beach 90 St
Beach 98 St
Beach 105 St
Rockaway Park Beach 116 St
Broadway Junction
Atlantic Av
Liberty
Van Siclen Av
Shepherd Av
Cleveland St
Norwood Av
Crescent St
Cypress Hills
75 St
85 St - Forest Pkwy
Woodhaven Blvd
104 St
111 St
121 St
Jamaica
Jamaica Center Parsons/Archer
Sutphin Blvd Archer Av
Sutphin Blvd
Jamaica 179 St
Parsons Blvd
Union Tpke Kew Gardens
Forest Hills 71 Av
Jackson Hts Roosevelt Av
Queensboro Plaza
Queens Plaza
Court Sq 23 St
Hunters Point Av
Vernon Blvd Jackson Av
Long Island City
Greenpoint Av
Nassau Av
Metropolitan Av
Lorimer St
Broadway
Flushing Av
Myrtle Av
Myrtle - Willoughby Av
Bedford - Nostrand Avs
Classon Av
Clinton - Washington Avs
Franklin Av
Nostrand Av
Kingston - Throop Avs
DeKalb Av
Knickerbocker Av
Central Av
Gates Av
Kosciuszko St
Halsey St
Chauncey St
Rockaway Av
Ralph Av
Utica Av
Halsey St
Wilson Av
Bushwick Av Aberdeen St
Jefferson St
Myrtle - Wyckoff Avs
Seneca Av
Forest Av
Fresh Pond Rd
Middle Village Metropolitan Av
Grand Av Newtown
Woodhaven Blvd
Queens Blvd
63 Dr Rego Park
67 Av
75 Av
Forest Hills
Kew Gardens
Union Tpke
Briarwood
Jamaica Van Wyck
Mets - Willets Point
111 St
103 St - Corona Plaza
Junction Blvd
90 St - Elmhurst Av
82 St - Jackson Hts
74 St Broadway
65 St
46 St Northern Blvd
40 St
33 St
Woodside
Steinway St
Astoria Ditmars Blvd
Astoria Blvd
30 Av
36 Av
39 Av
21 St
Broadway
Ditmars Blvd
Flushing Main St
Murray Hill
Broadway
Auburndale
LIRR
Bayside
Douglaston
Little Neck
Port Washington
Hudson River
Harlem 125 St
Lexington Av 63 St
Roosevelt Island
Tramway
Lexington Av 59 St
5 Av
57 St
47-50 Sts Rockefeller Ctr
Grand Central 42 St
34 St Hudson Yards
34 St Penn Station
28 St
23 St
18 St
14 St Union Sq
8 St NYU
W 4 St
Astor Pl
Broadway - Lafayette
Bleecker St
Spring St
Prince

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

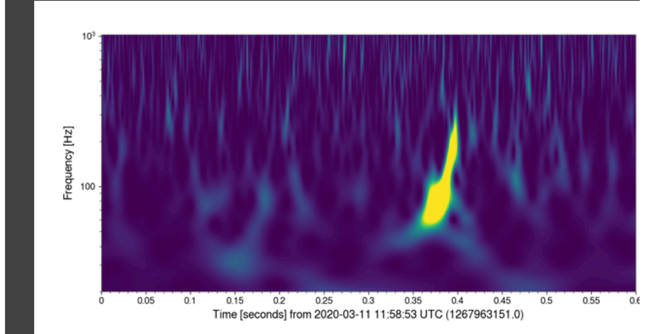
GPS: 1267963151.3

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

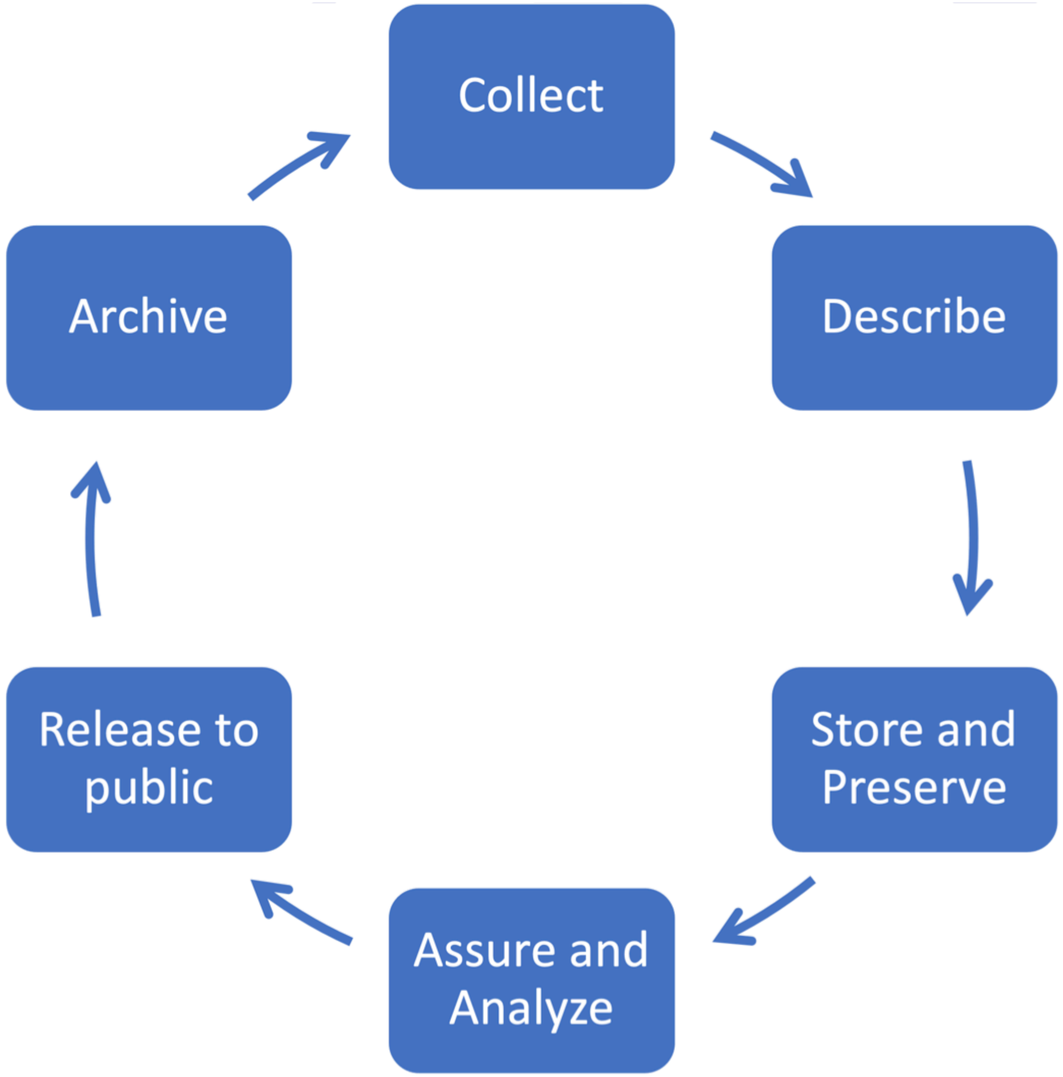
Getting Help

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

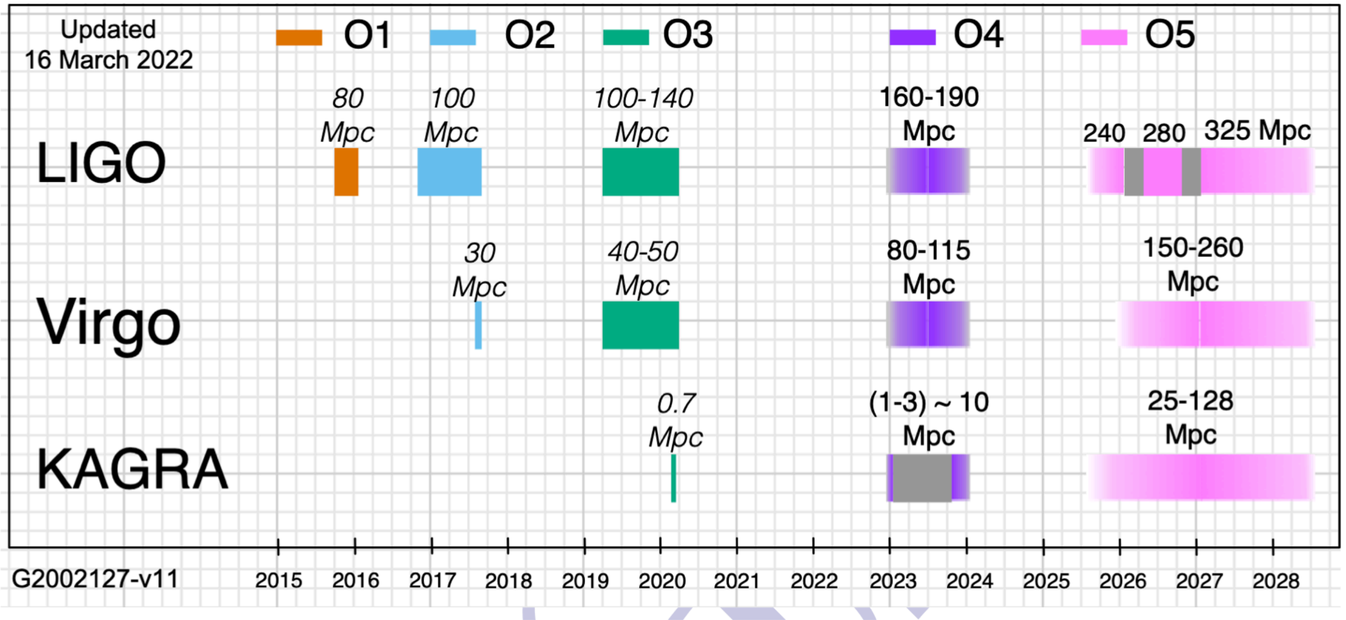
Provide direct support AND Learn about gaps

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



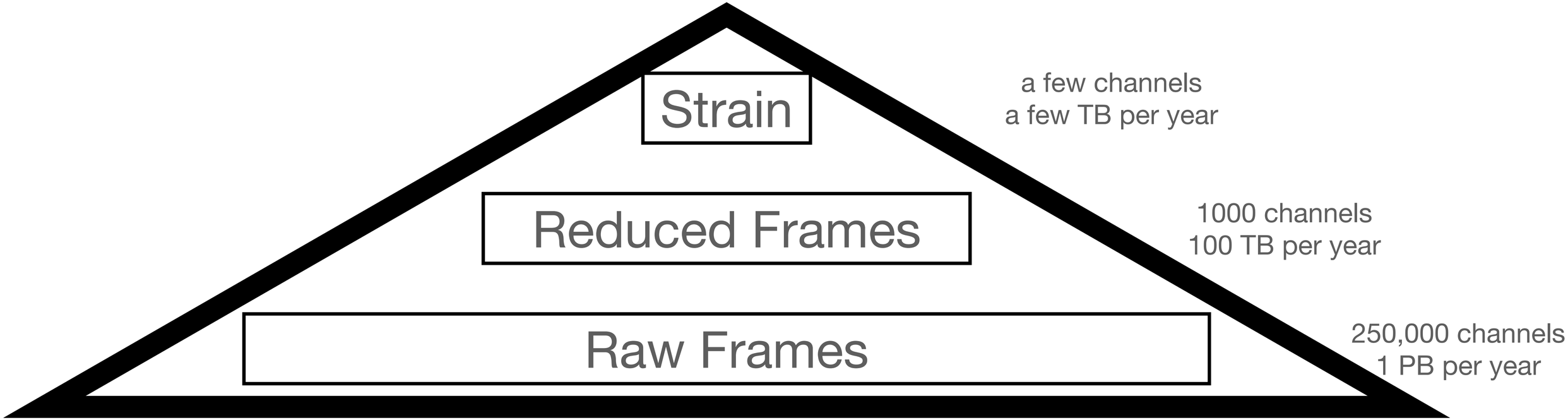
Describe

- 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

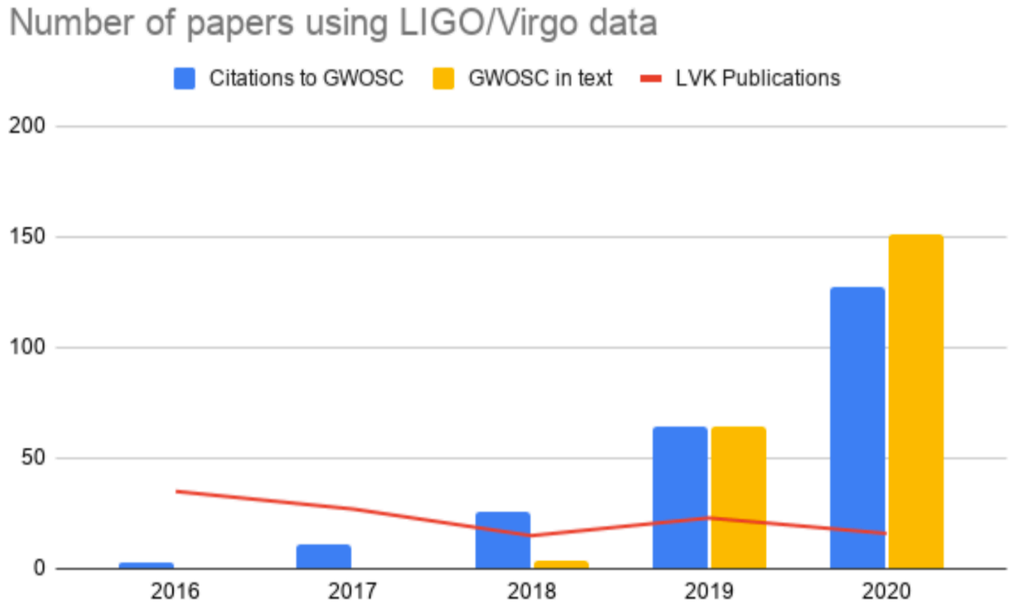
Impacts of Open Data

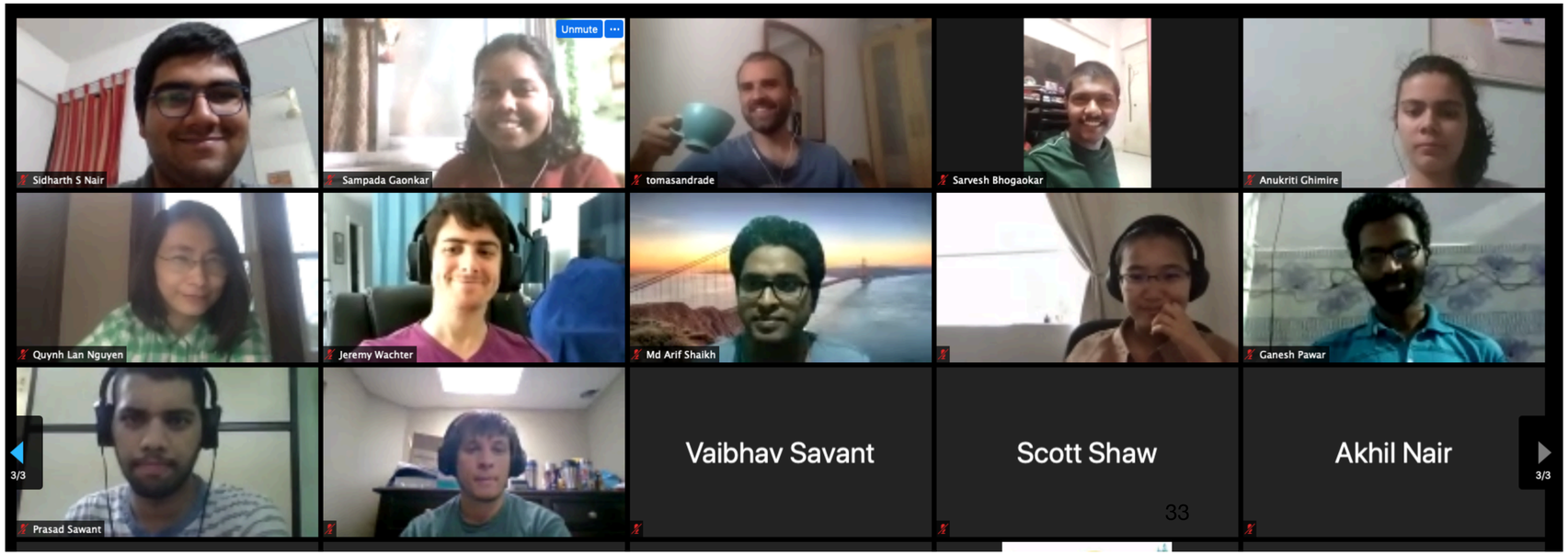
Around 6,000 visitors (12,000 sessions)
to GWOSC each month

Over a million strain file downloads over 6 months

250 Papers in 2 years (2020 + 2021)

Open Data Workshops with hundreds of participants

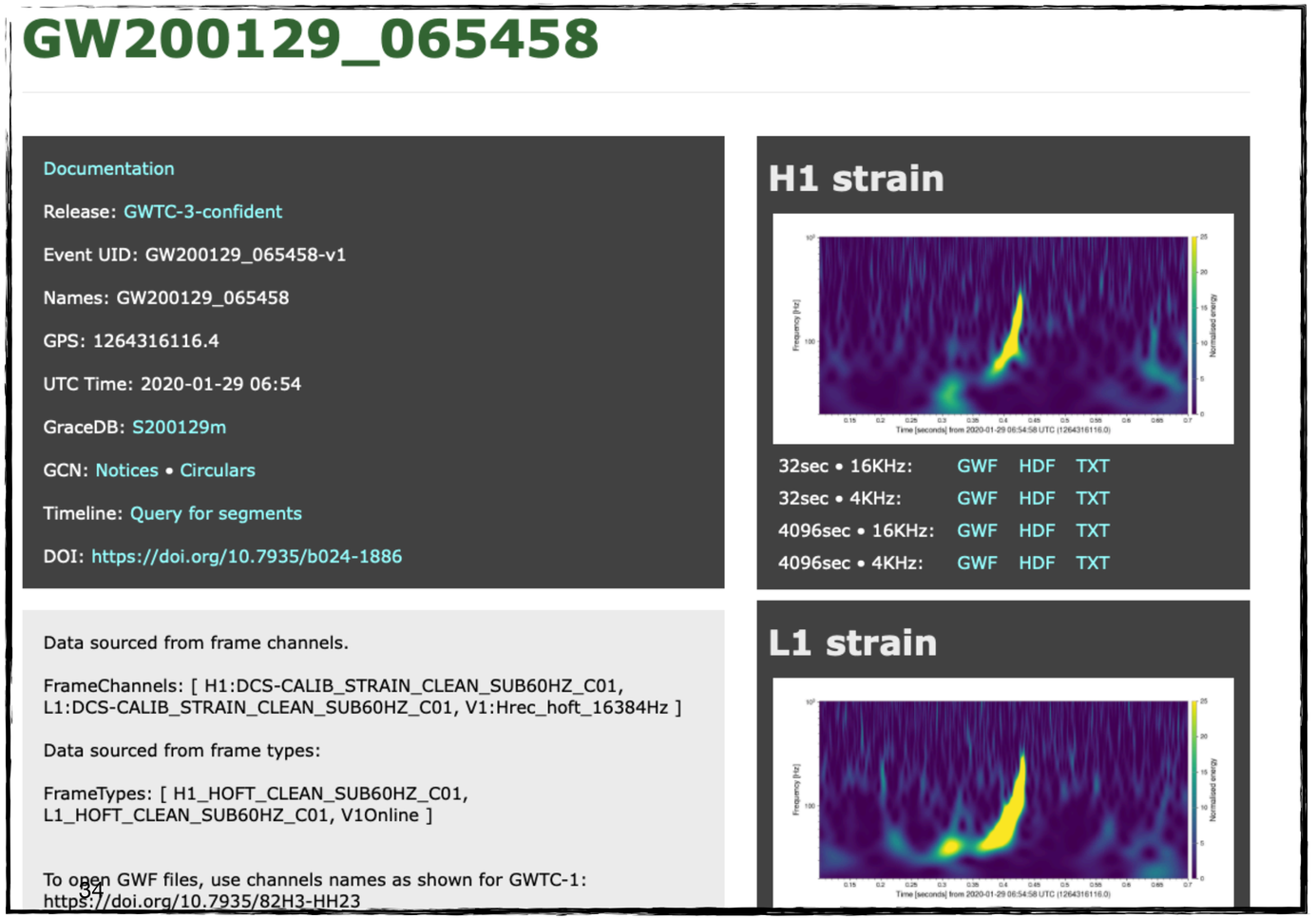




GWOSC Event Portal

- 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

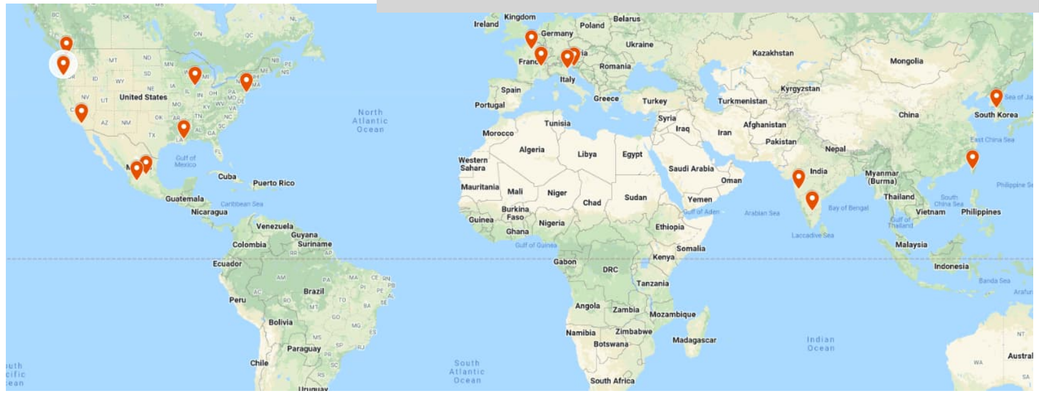
https://gwosc.org/eventapi

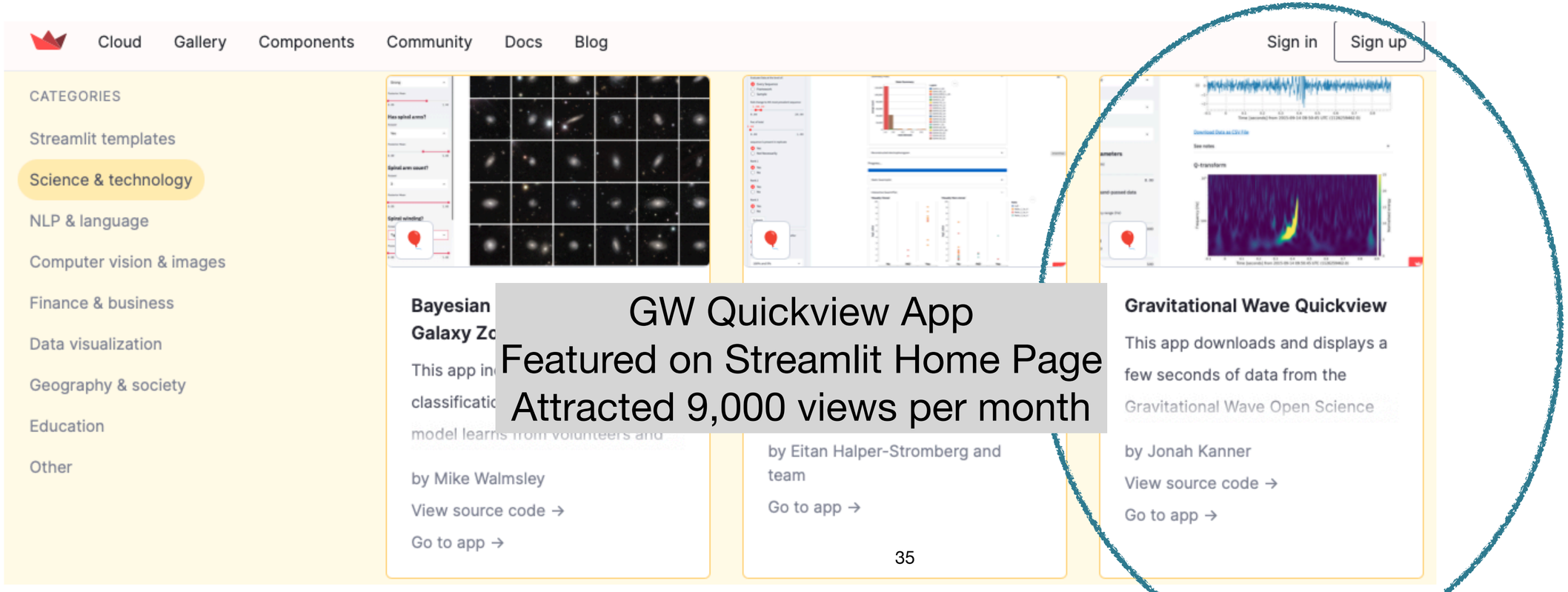


Supporting the Community

2022 Open Data Workshop
1000+ Participants
15 Locations + Virtual

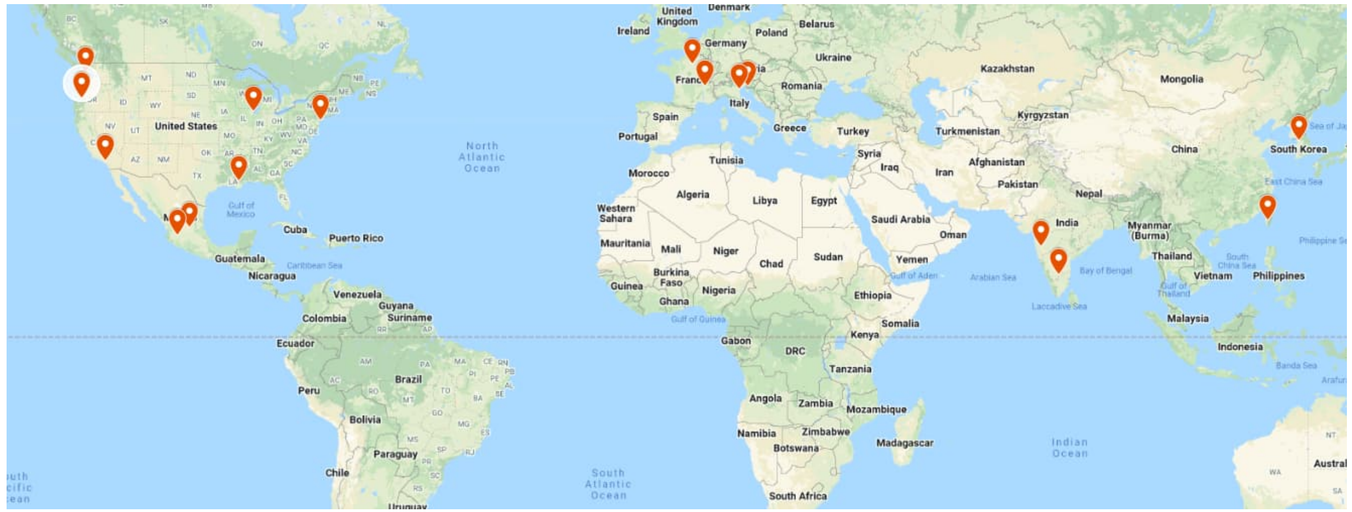
Discussion forum: https://ask.igwn.org
E-mail help desk: gwosc@igwn.org
Online Course: https://gw-odw.thinkific.com
Web apps: https://gwosc.org/path
Tutorials & Workshops: https://gwosc.org/tutorials





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

New York


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

Paris

Evanston, IL