

Gravitational Wave Data: The Last Mile

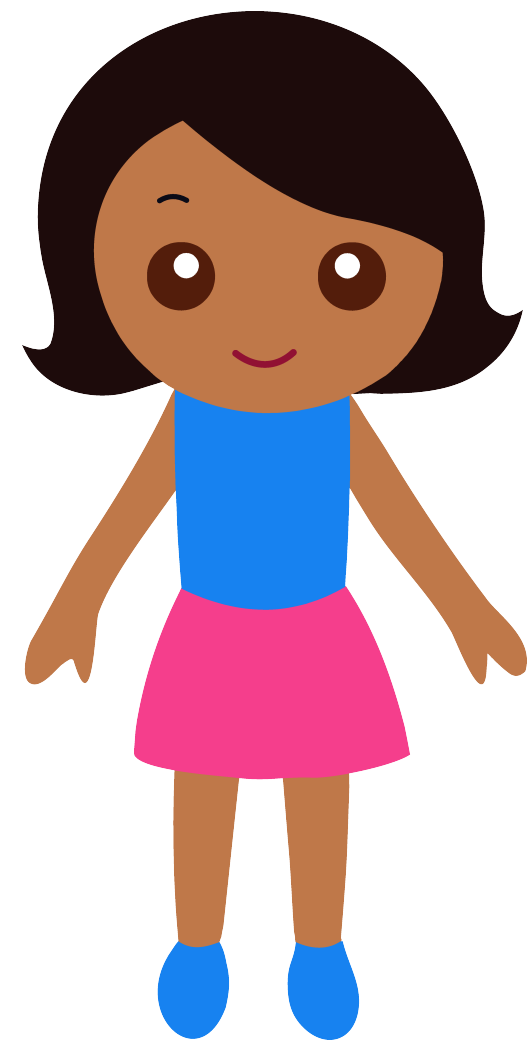
Jonah Kanner

LIGO Lab, Caltech

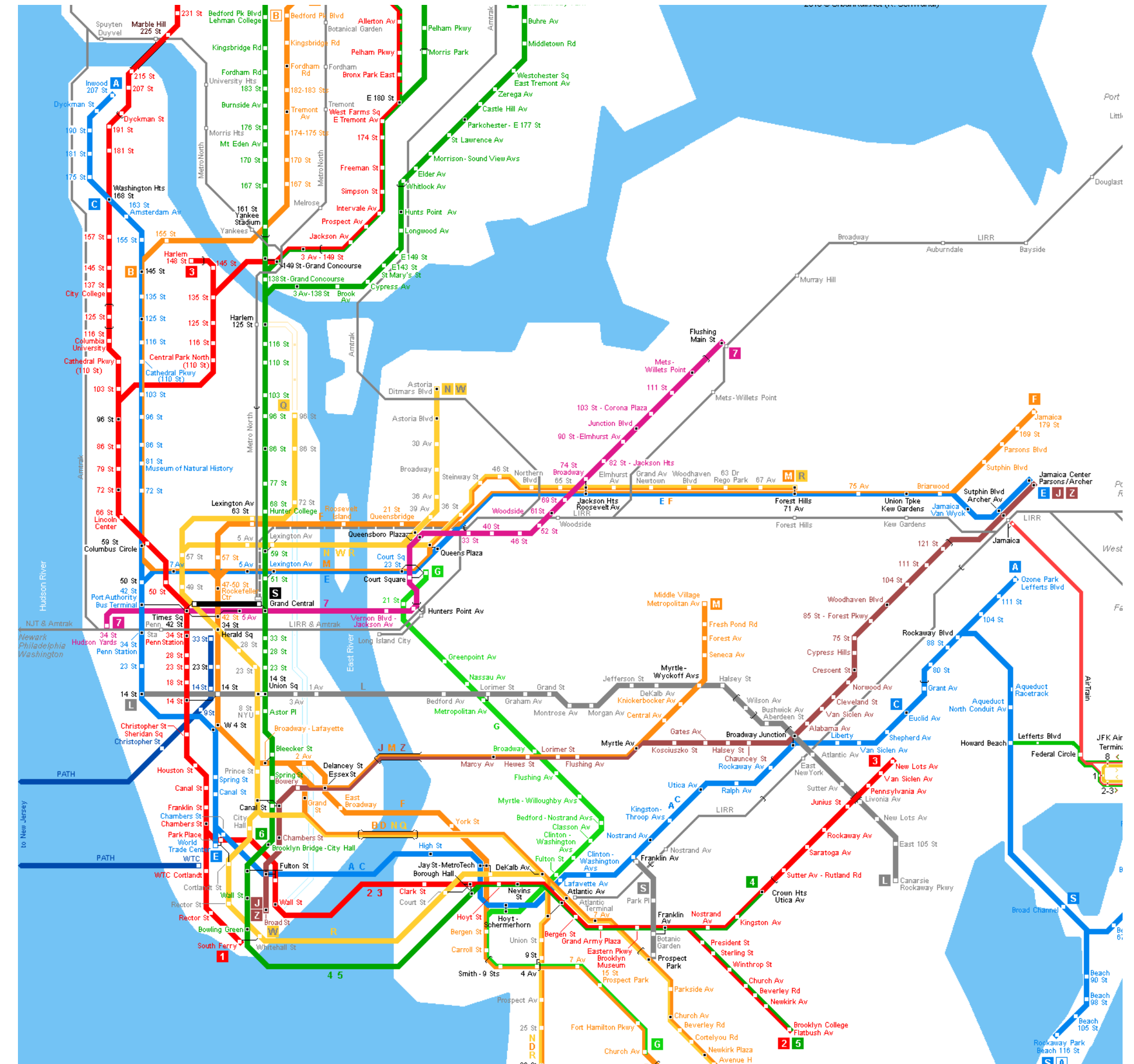
June 15, 2022 - G2200946-v2

The Last Mile Problem

Transportation

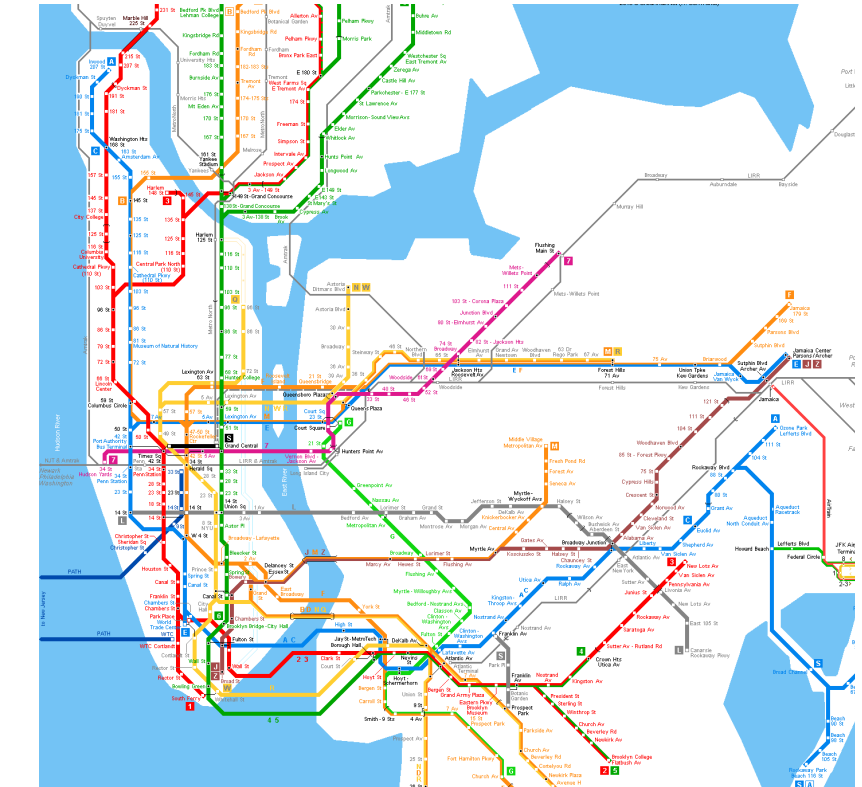
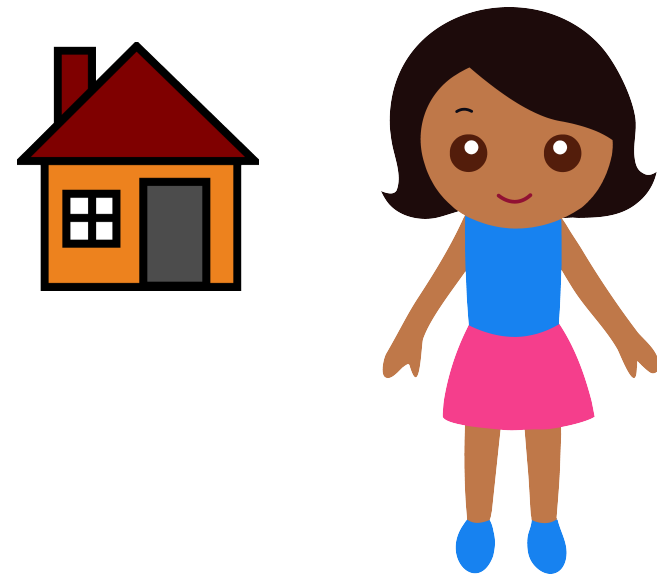


Last Mile



The Last Mile Problem

Public Data



Expert Networks

“Last Mile”



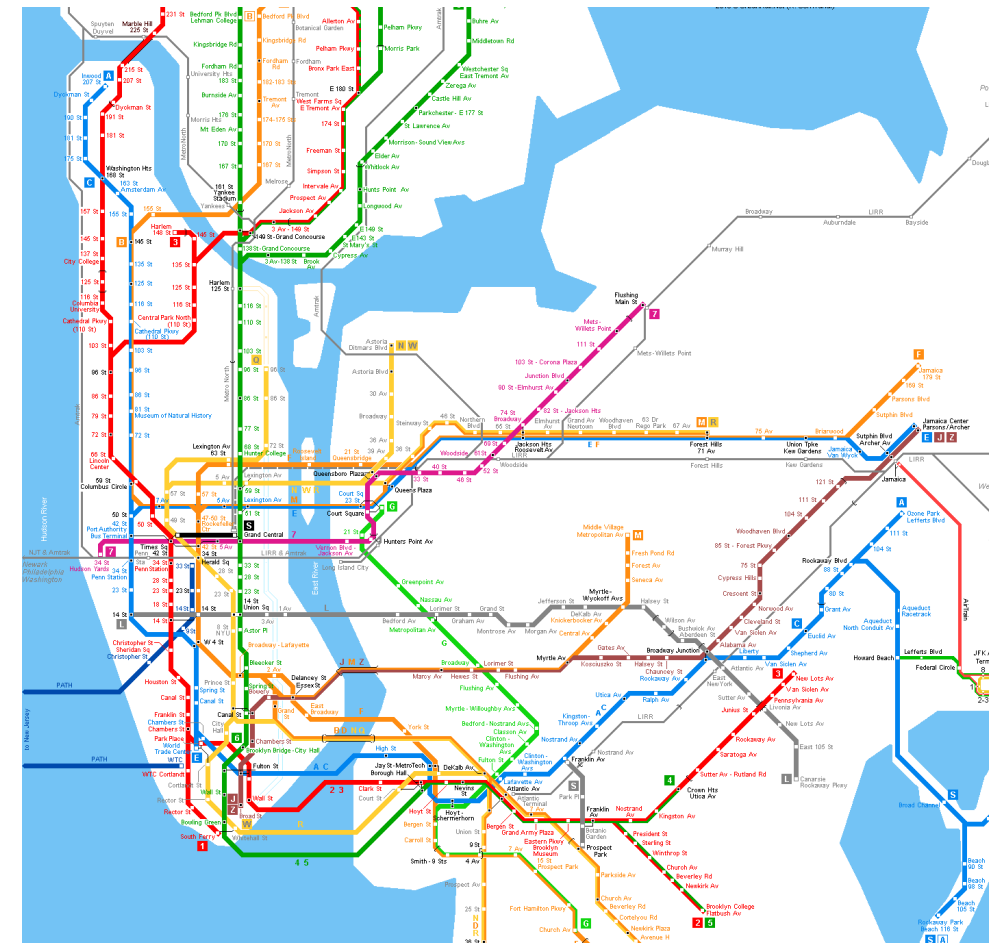
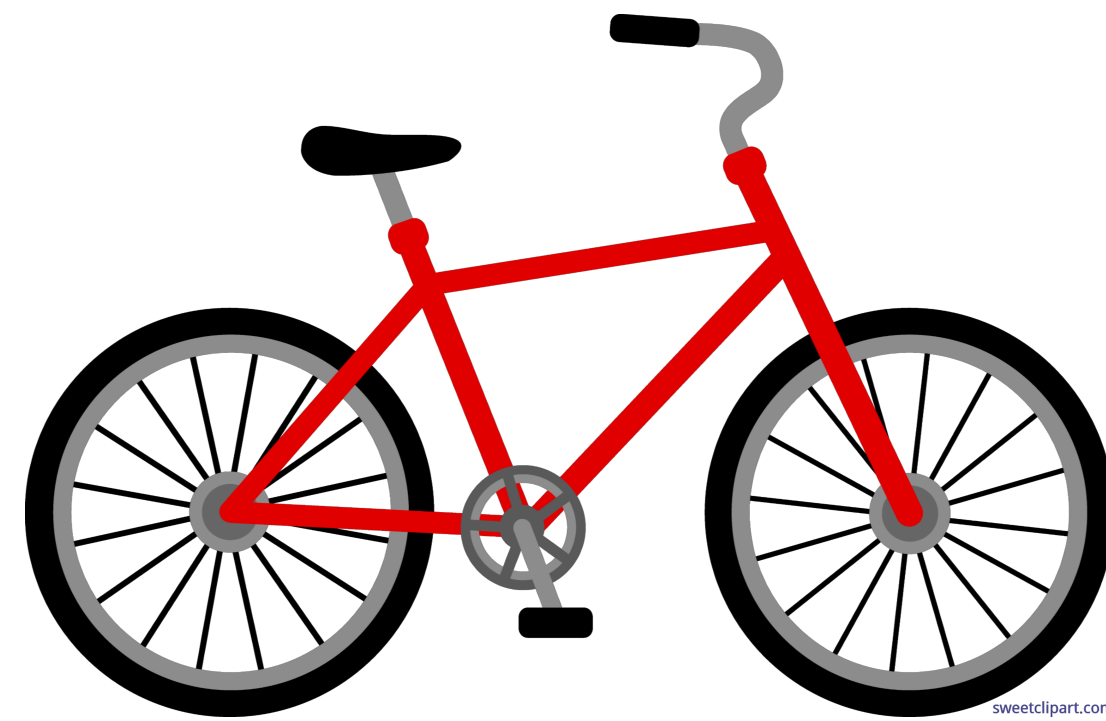
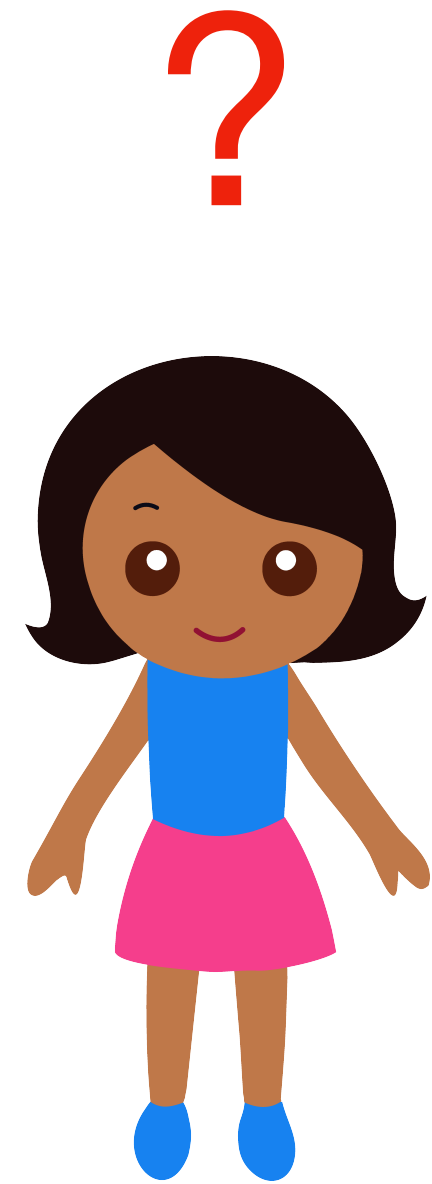
K-12
Undergraduates
Grad Students
Pros in other fields
Amateurs
Artists

Gaps in:
Access
Knowledge
Resources

Data
Software
Journal Articles
Conferences
Colleagues

The Last Mile Problem

High Stakes



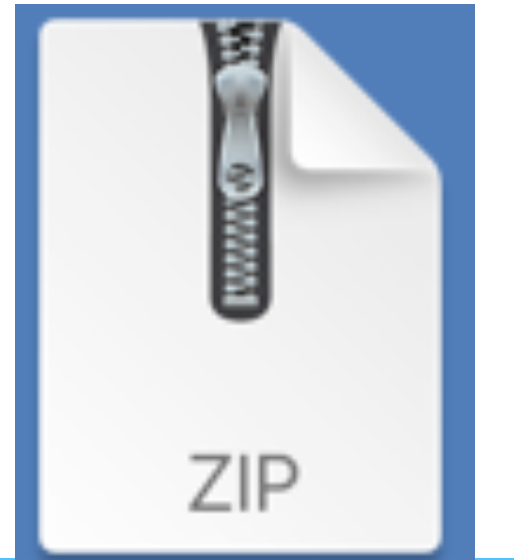
Diversity, Equity,
and Inclusion

Efficiency and
Productivity

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?



data.wtf.gz

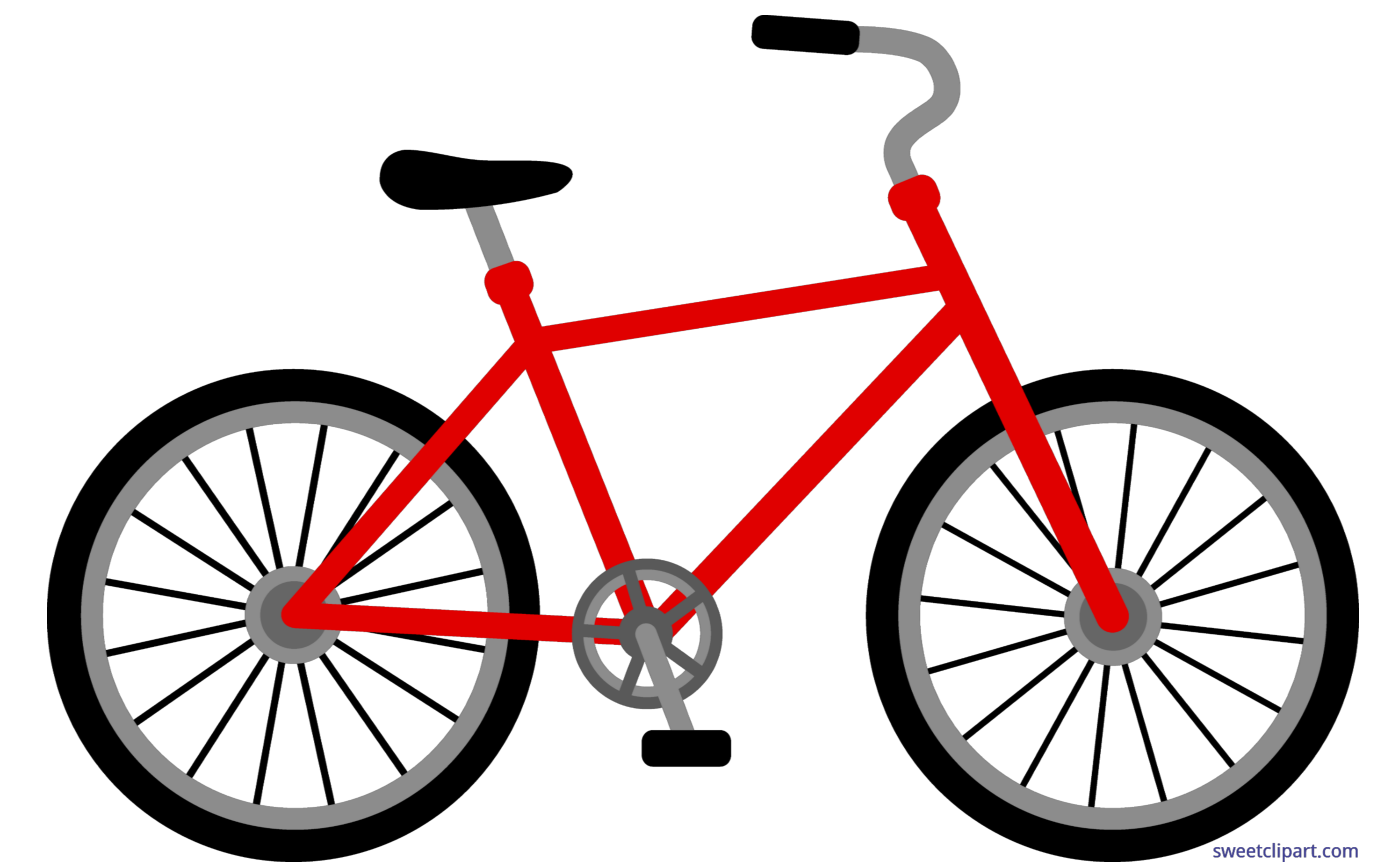


Access Gaps for LIGO Data (circa 2014)

- 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>

Open Data Workshops

2022 Open Data Workshop
1000+ Participants
15 Locations + Virtual

- 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

Shreejit Jadhav

PhD Student

Inter-University Centre for Astronomy and Astrophysics (IUCAA), Pune, India



Leïla Haegel

Researcher

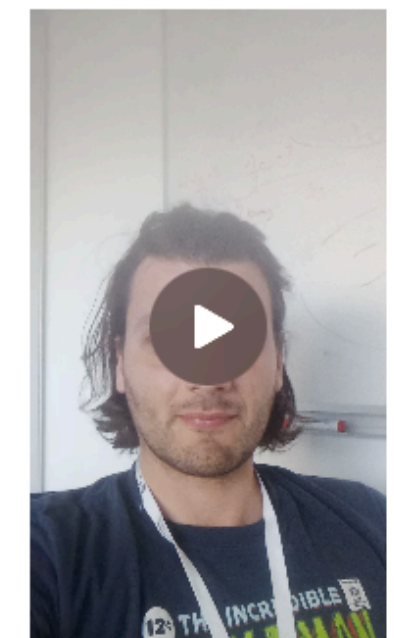
Astroparticles and Cosmology Laboratory, France



Simone Mastrogiovanni

Postdoc

ARTEMIS, Nice Observatory, France





Pasadena, CA



Trieste, Italy



San Luis Potosí, Mexico



European Virtual Hub



Guadalajara, Mexico



LIGO Livingston

Software Examples In Your Browser

Jupyter Notebooks

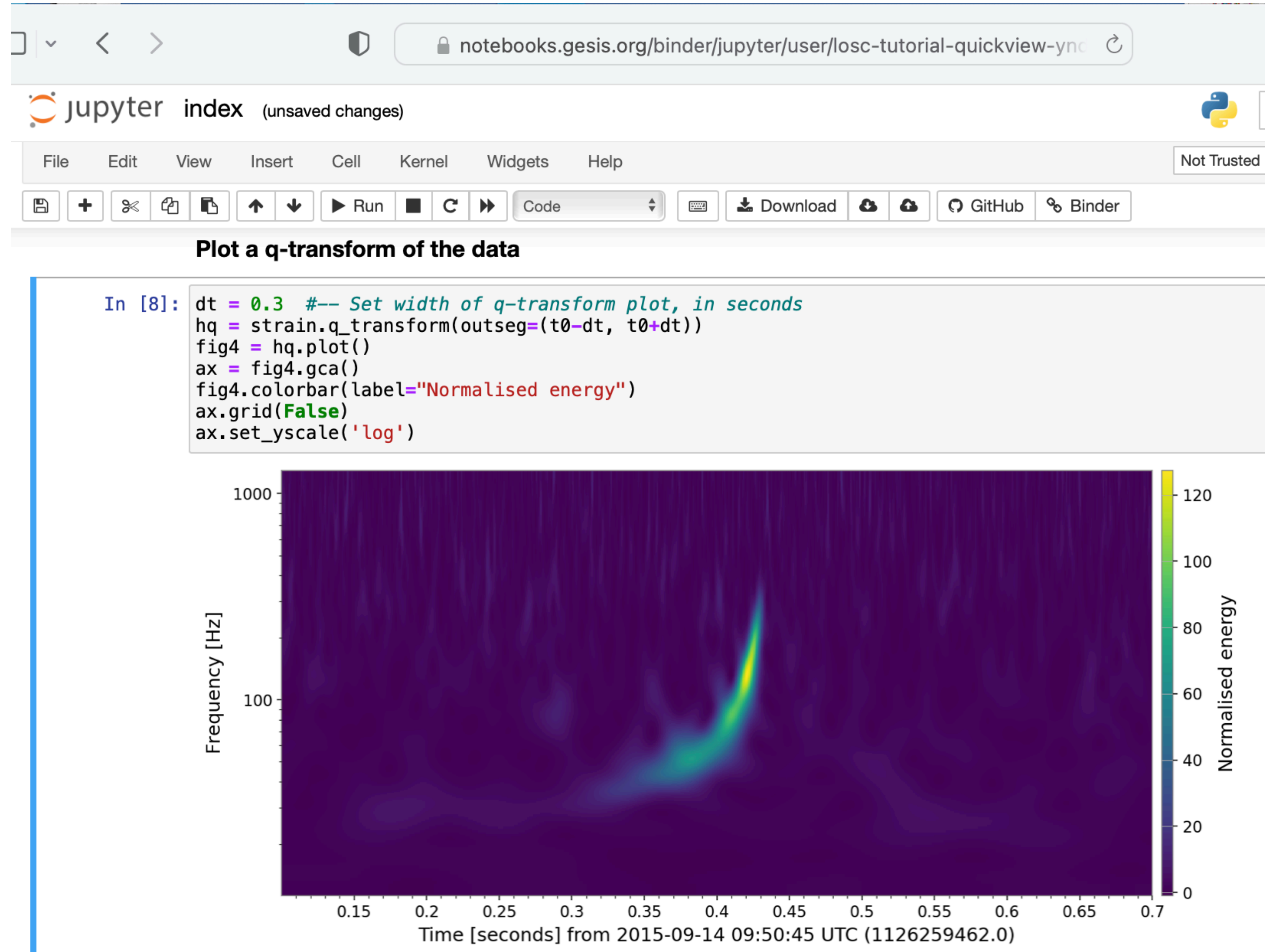
google co-lab
mybinder

Specialized libraries:

GWpy, pyCBC, bilby

No installation

gwosc.org/tutorials



Web Apps or GUIs

Remove the need to program!

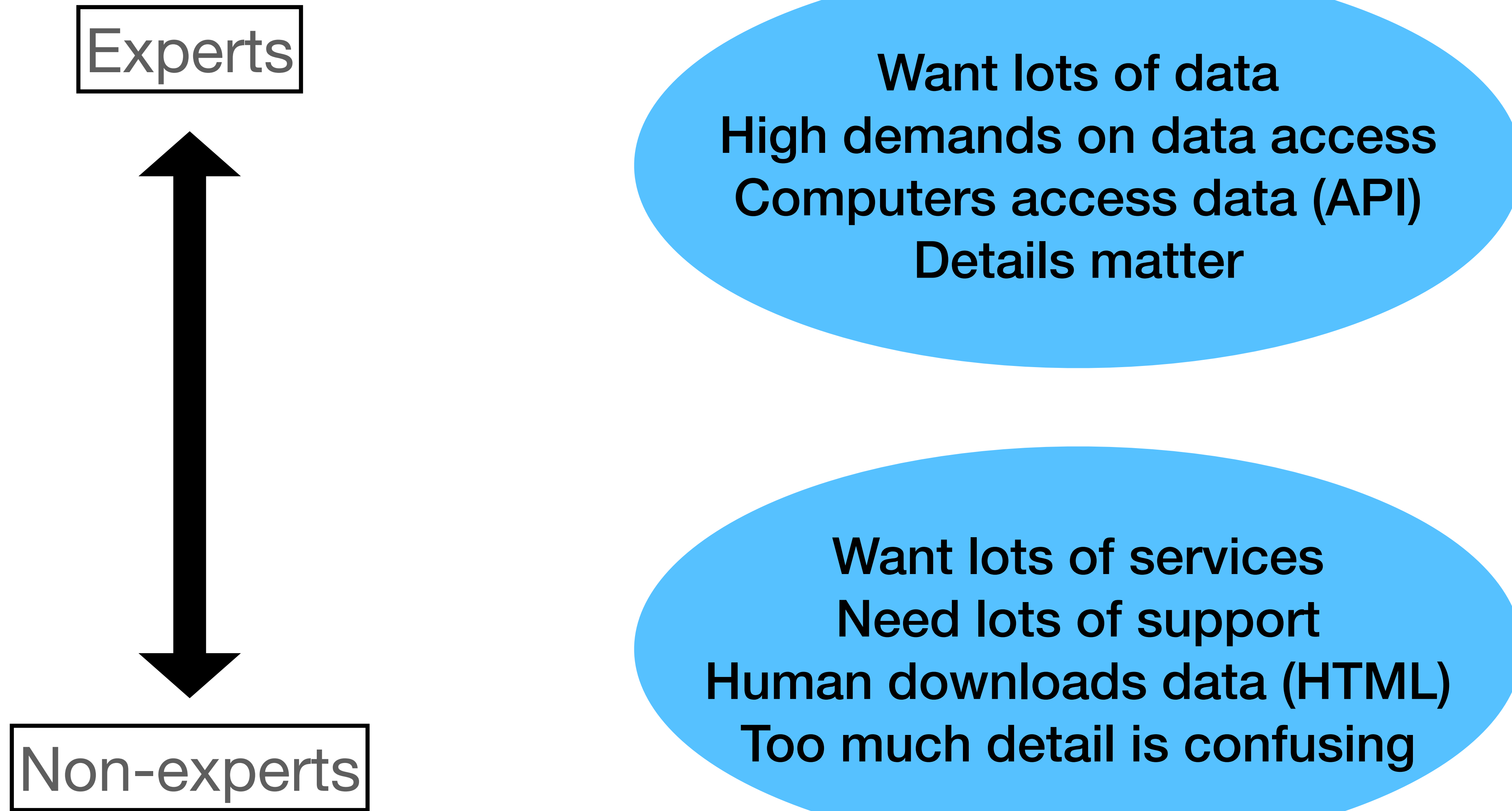
- Plot data with no programming
- “Pre-process” data (whiten, filter, etc.)
- Export common file types (e.g. CSV)
- Introduction to signal processing

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

<https://gwosc.org/path>

The screenshot shows a web browser window displaying a Streamlit application titled "Gravitational Wave Quickview". The browser address bar shows the URL "share.streamlit.io/jkanner/streamlit-dataview/app.py". The application interface is split into two main sections. On the left is a control panel titled "Select Data Time and Detector" with a close button (X). It contains several dropdown menus: "How do you want to find data?" (set to "By event name"), "Select Event" (set to "GW151012"), and "Detector" (set to "H1"). There is also a checkbox for "Full sample rate data" which is currently unchecked. Below these is a "Set Plot Parameters" section with a "Time Range (seconds)" input field showing "0.44". On the right is the main content area with the title "Gravitational Wave Quickview". It includes a list of instructions: "Use the menu at left to select data and set plot parameters" and "Your plots will appear below". Below this, it displays event details for "GW151012": "GPS: 1128678900.4", "Mass 1: 23.2 M_⊙", "Mass 2: 13.6 M_⊙", and "Network SNR: 10". A link for the "Event page" is provided: "https://gw-osc.org/eventapi/html/event/GW151012". At the bottom of the right section, it says "Loading data...done!". In the bottom right corner of the application, there is a "Manage app" button.

Diversity of Experience



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>

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 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

The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. Below the header, the main content area displays the title 'LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration Data Releases'. Underneath, there is a section for 'Recent uploads' with a search bar. Two upload entries are visible:

- May 14, 2022 (v9)** Dataset Open Access: GWOSC Event Portal Snapshots. Description: LIGO Scientific Collaboration; Virgo Collaboration; KAGRA Collaboration; This repo contains "snapshots" of the information available through the GWOSC Event Portal API, as seen at: https://gw-openscience.org/eventapi Snapshots are made about a day after any updates to the Event Portal database, and the date of each snapshot can be seen in the file name. Each s. Uploaded on May 17, 2022. 8 more version(s) exist for this record.
- April 22, 2022 (v1)** Dataset Open Access: GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run - Data Quality Products for GW Searches. Description: LIGO Scientific Collaboration and Virgo Collaboration;

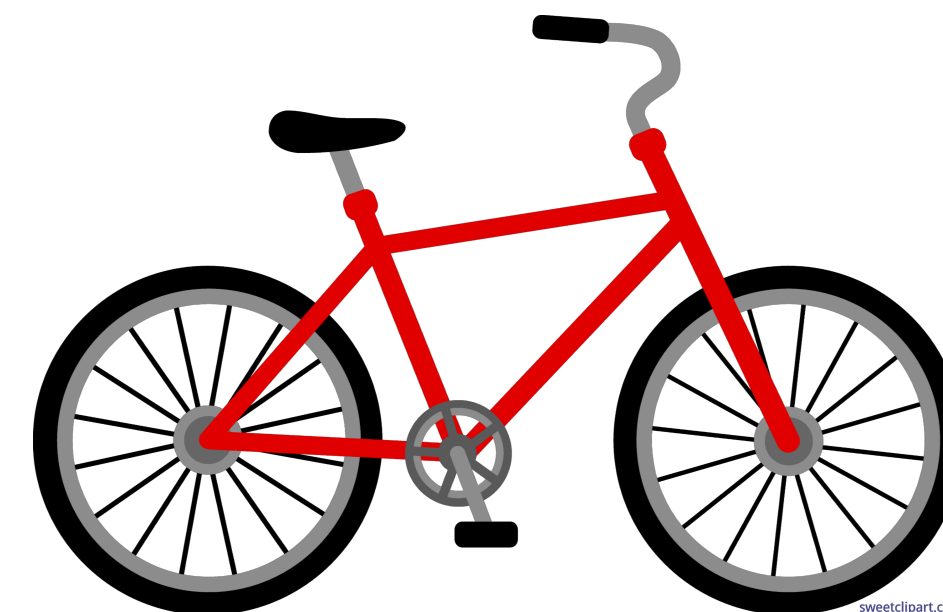
On the right side, there is a sidebar with a green 'New' button and a community description for 'LIGO Scientific Co Collaboration and Data Releases'. The sidebar also includes metadata: 'Curated by: LVK-publications', 'Curation policy: Not specified', 'Created: July 20, 2021', and 'Harvesting API: OAI-PMH Interface'.

<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 for everyone
 - 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!

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_⊙

Mass 2: 13.6 M_⊙

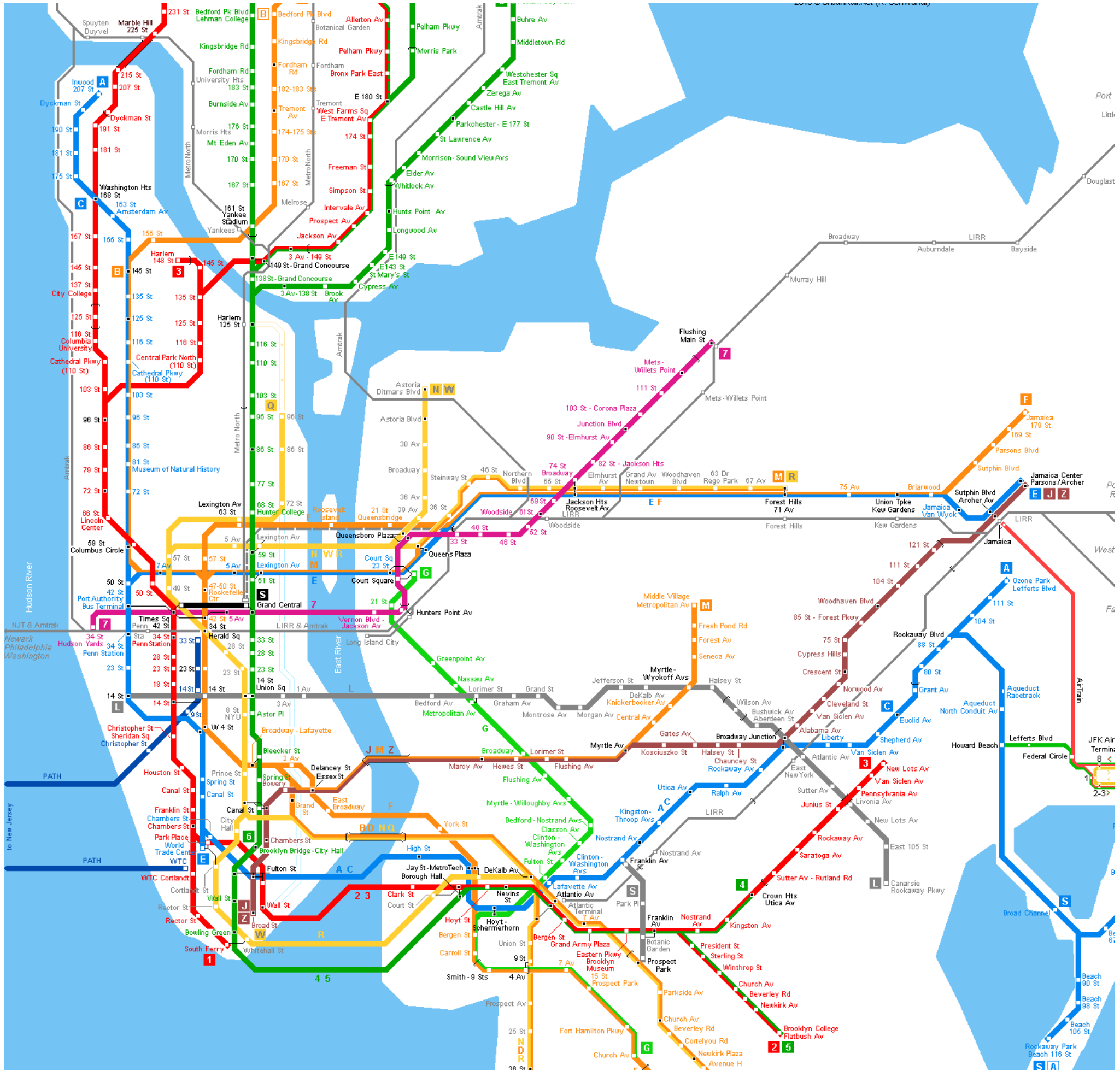
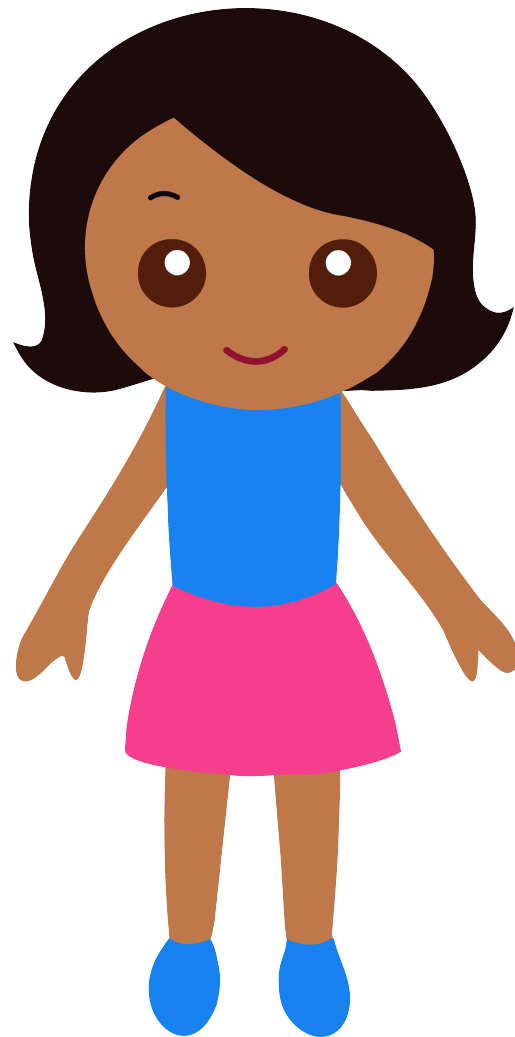
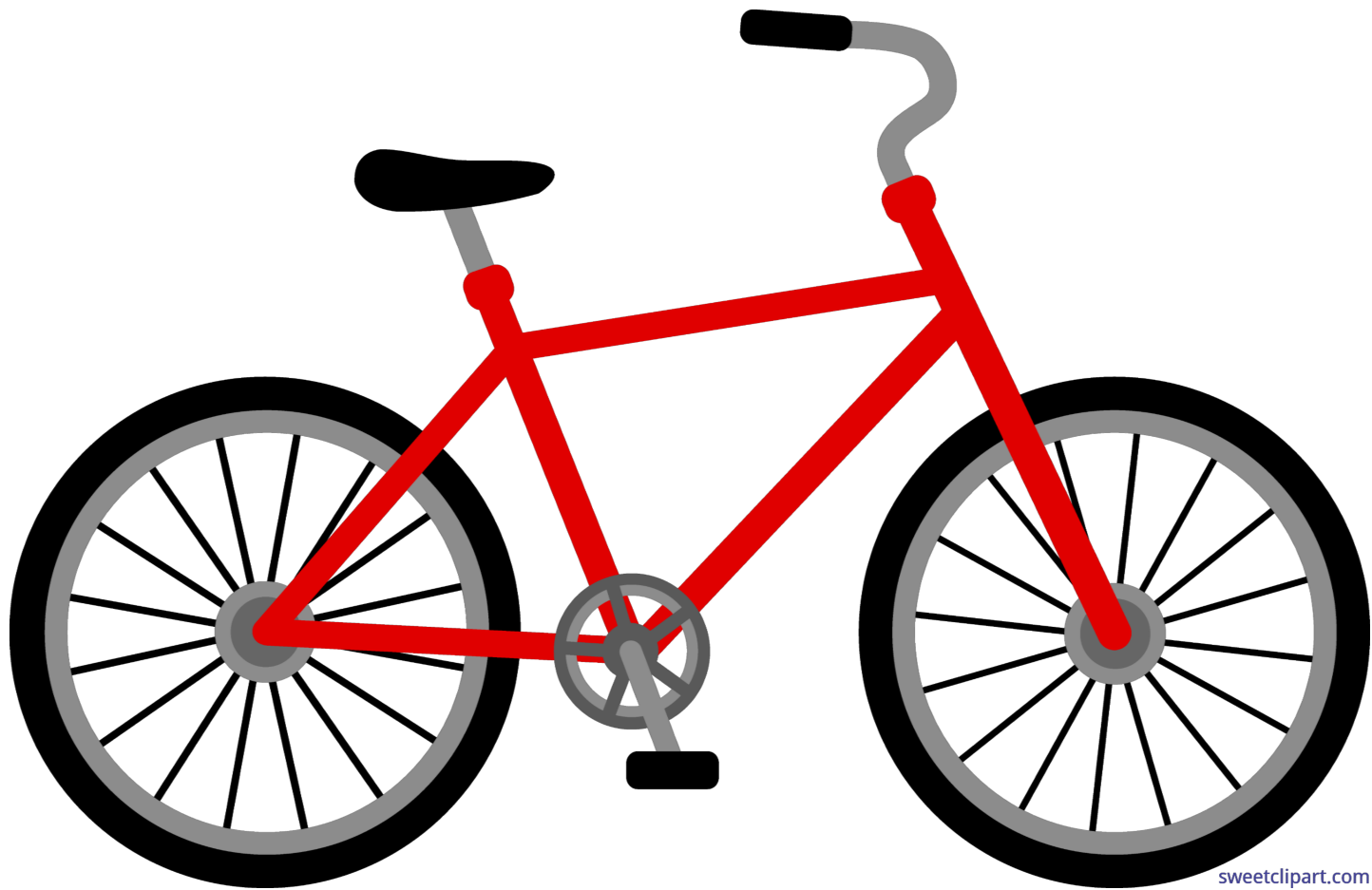
Network SNR: 10

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

Loading data...done!

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

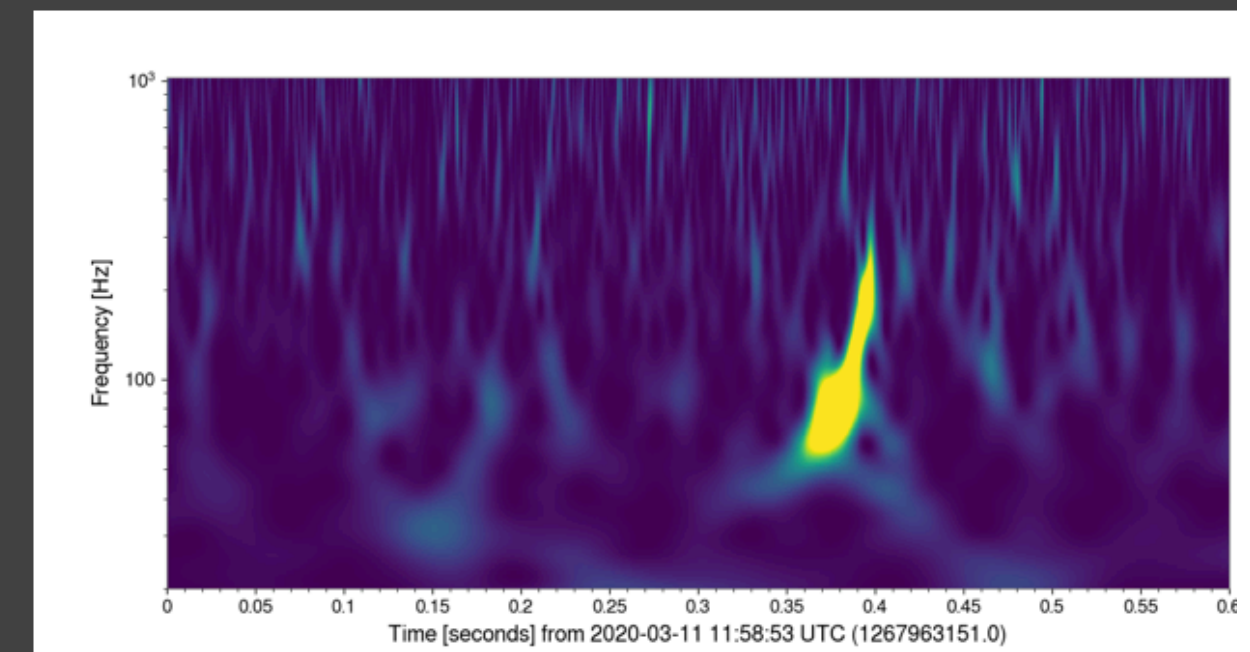
GPS: 1267963151.3

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 _⊙) | Mass 2 (M _⊙) |
|---------------------------------|---------|----------------------------------|--------------|---------------------------------------|---------------------------------------|
| 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>

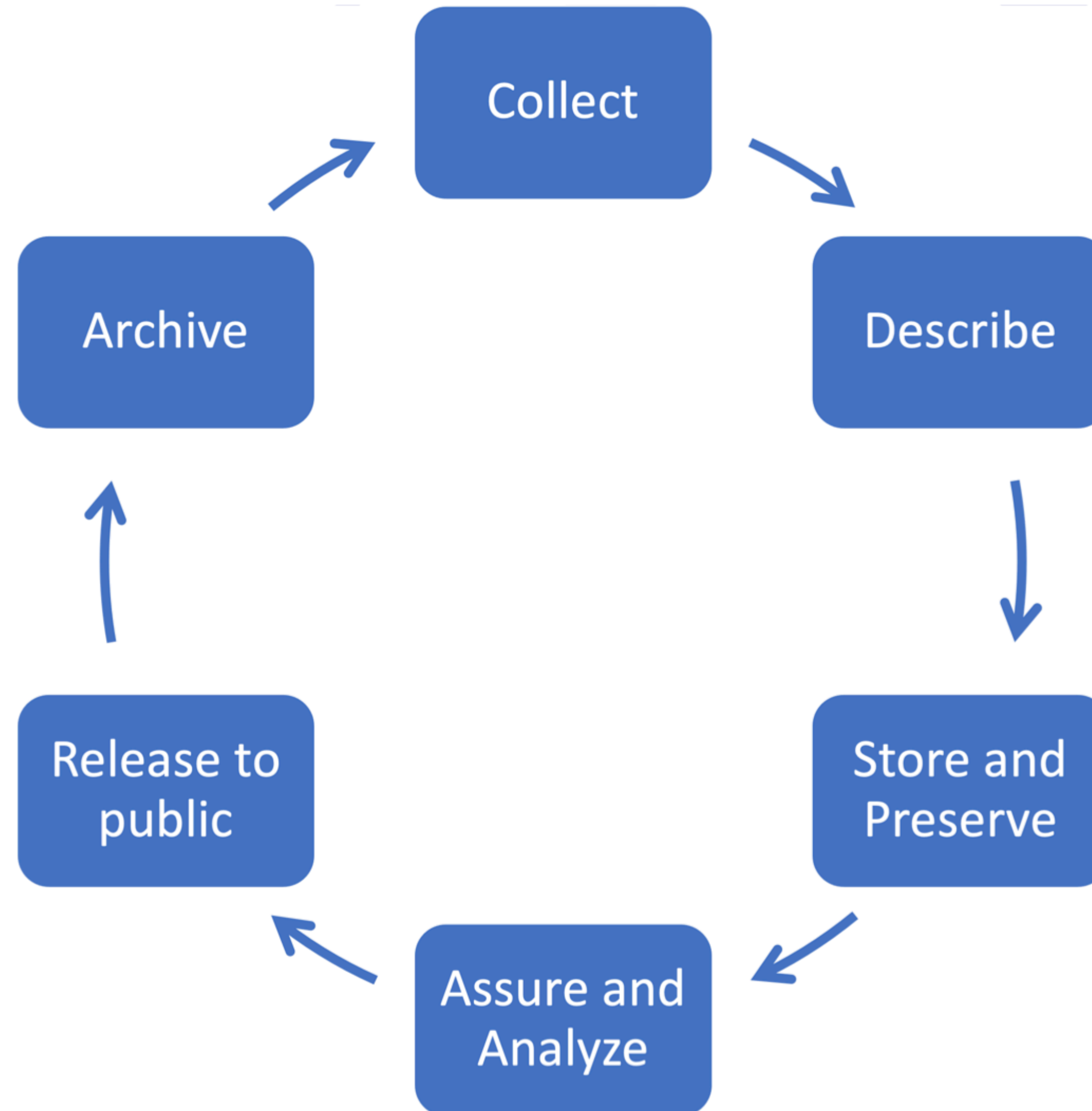
Getting Help

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

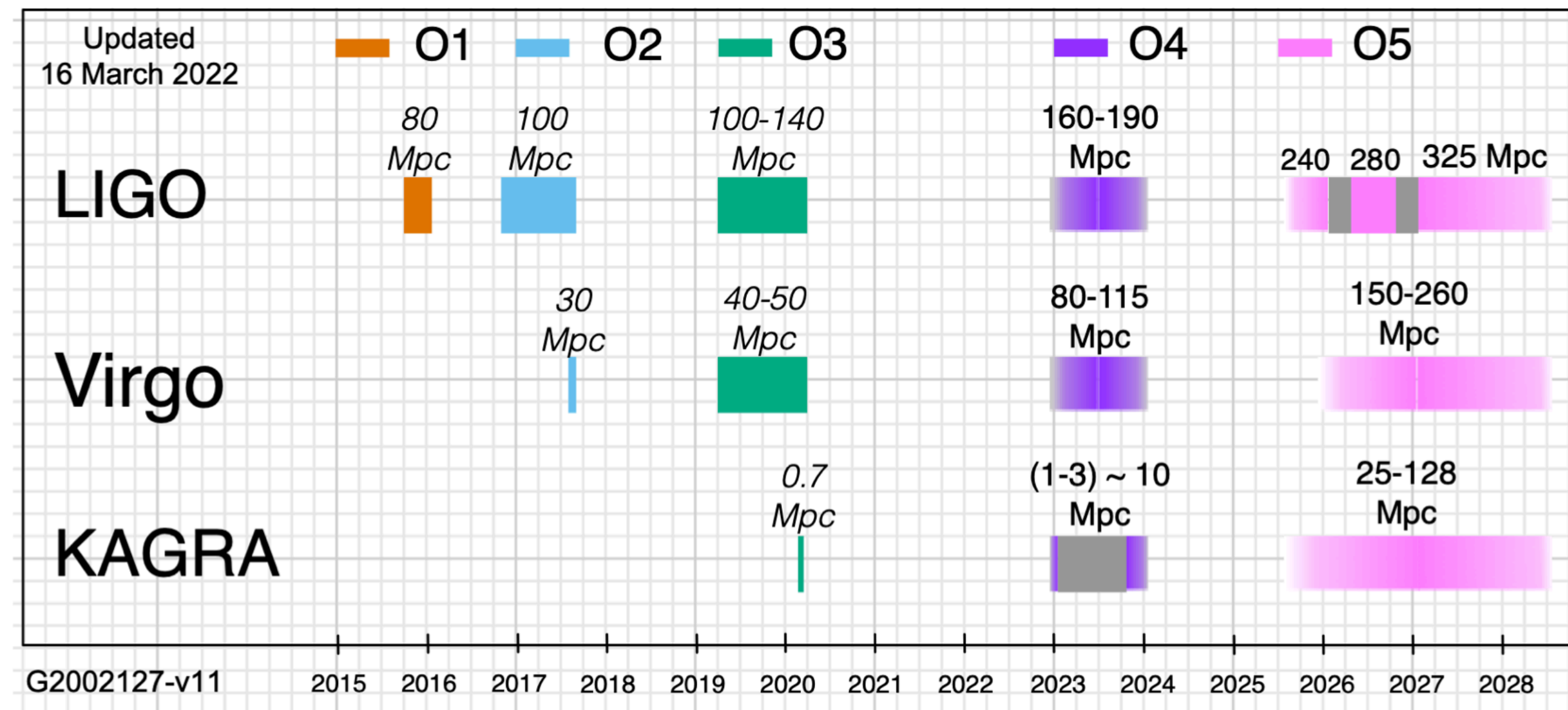
Provide direct support AND Learn about gaps

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



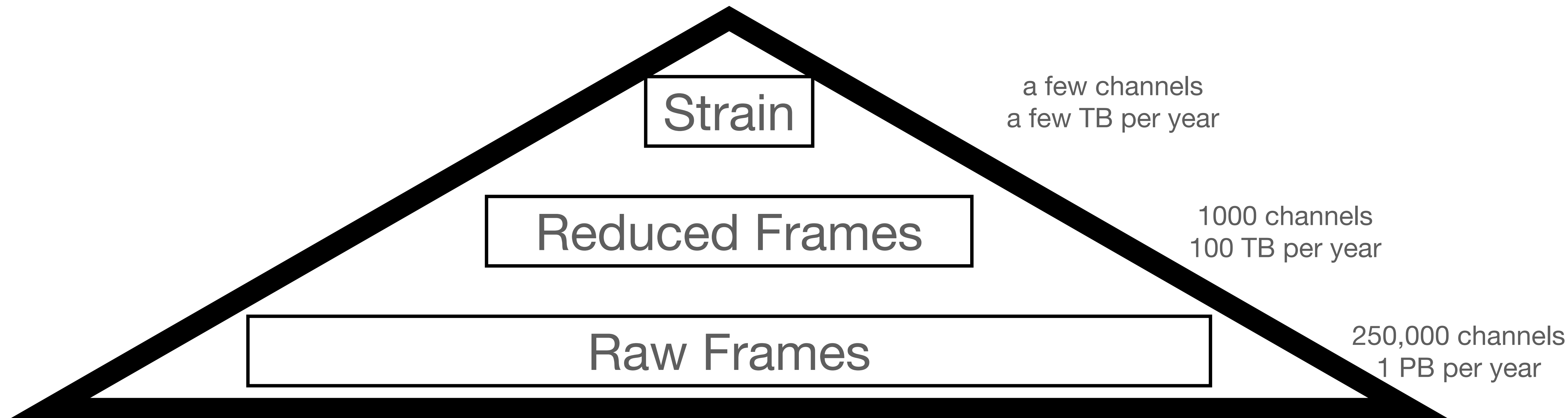
Describe

- 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

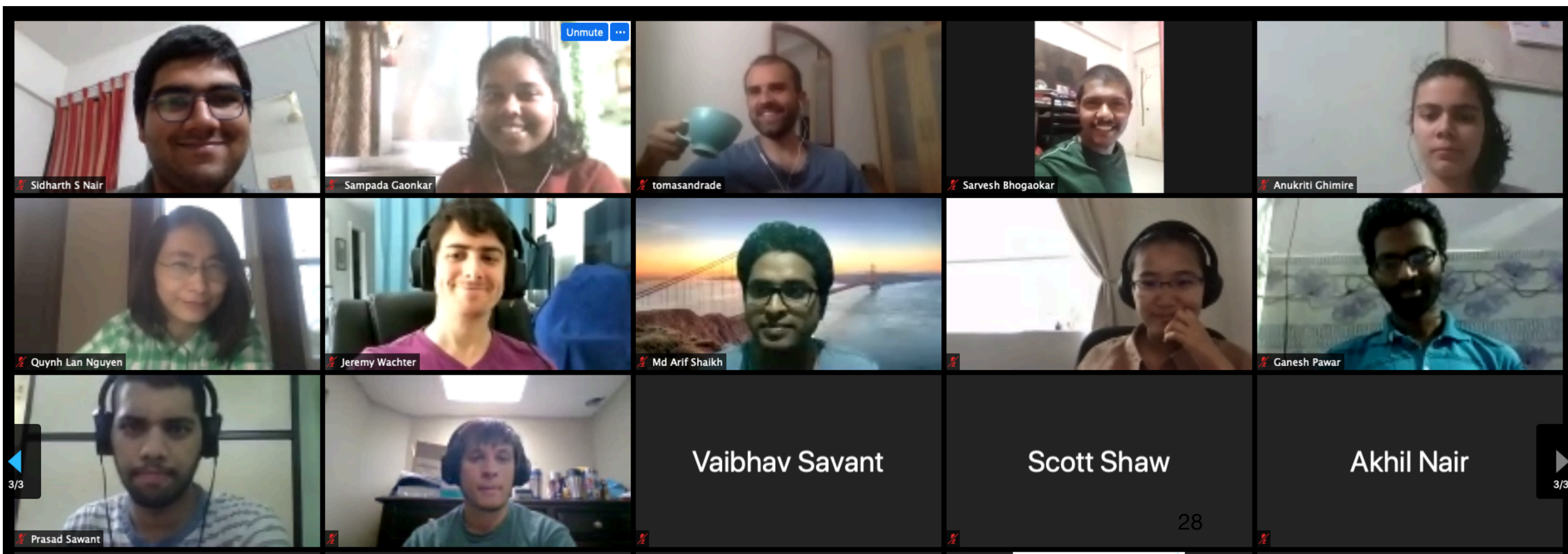
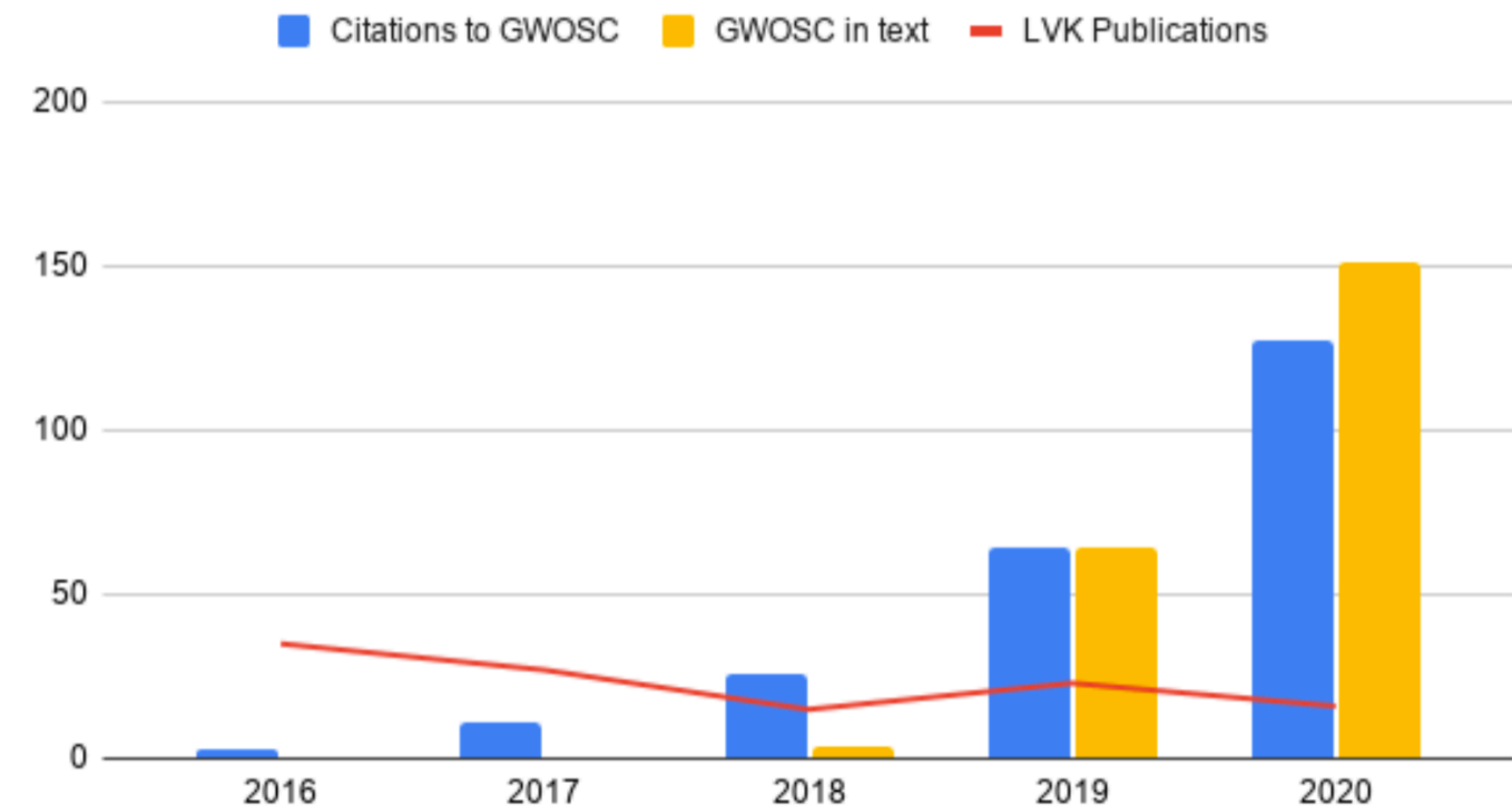
Around 6,000 visitors (12,000 sessions) to GWOSC each month

Over a million strain file downloads over 6 months

250 Papers in 2 years (2020 + 2021)

Open Data Workshops with hundreds of participants

Number of papers using LIGO/Virgo data



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>

GW200129_065458

Documentation

Release: [GWTC-3-confident](#)

Event UID: GW200129_065458-v1

Names: GW200129_065458

GPS: 1264316116.4

UTC Time: 2020-01-29 06:54

GraceDB: [S200129m](#)

GCN: [Notices](#) • [Circulars](#)

Timeline: [Query for segments](#)

DOI: <https://doi.org/10.7935/b024-1886>

Data sourced from frame channels.

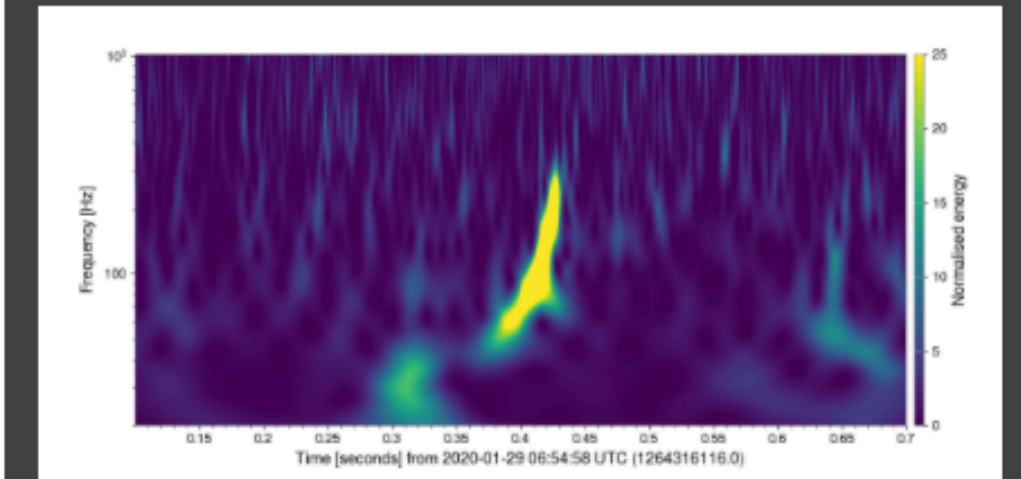
FrameChannels: [H1:DCS-CALIB_STRAIN_CLEAN_SUB60HZ_C01, L1:DCS-CALIB_STRAIN_CLEAN_SUB60HZ_C01, V1:Hrec_hoft_16384Hz]

Data sourced from frame types:

FrameTypes: [H1_HOFT_CLEAN_SUB60HZ_C01, L1_HOFT_CLEAN_SUB60HZ_C01, V1Online]

To open GWF files, use channels names as shown for GWTC-1:
<https://doi.org/10.7935/82H3-HH23>

H1 strain



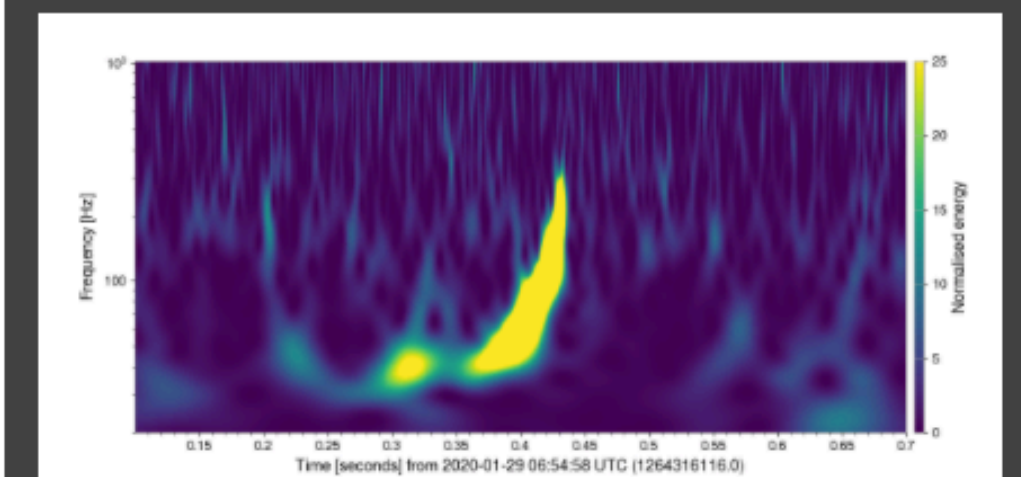
32sec • 16KHz: [GWF](#) [HDF](#) [TXT](#)

32sec • 4KHz: [GWF](#) [HDF](#) [TXT](#)

4096sec • 16KHz: [GWF](#) [HDF](#) [TXT](#)

4096sec • 4KHz: [GWF](#) [HDF](#) [TXT](#)

L1 strain



Supporting the Community

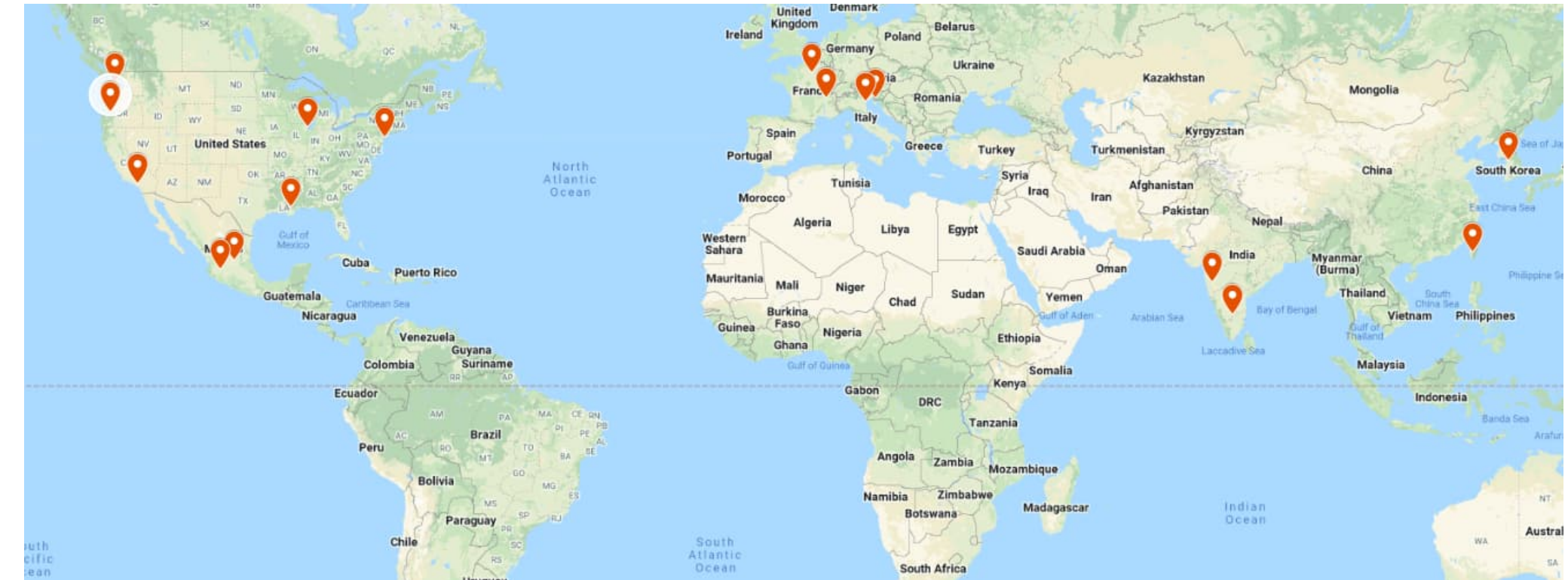
Discussion forum: <https://ask.igwn.org>

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>



A screenshot of the GWOSC community gallery website. The top navigation bar includes 'Cloud', 'Gallery', 'Components', 'Community', 'Docs', and 'Blog', along with 'Sign in' and 'Sign up' buttons. A left sidebar lists categories: 'Streamlit templates', 'Science & technology' (highlighted), 'NLP & language', 'Computer vision & images', 'Finance & business', 'Data visualization', 'Geography & society', 'Education', and 'Other'. The main content area displays three featured apps: 1. 'Bayesian Deep Learning for Galaxy Zoo DECaLS' by Mike Walmsley, featuring a grid of galaxy images. 2. 'CloneRetriever' by Eitan Halper-Stromberg and team, featuring a data visualization interface. 3. 'Gravitational Wave Quickview' by Jonah Kanner, featuring a spectrogram plot. Each app card includes a 'View source code' and 'Go to app' link.

Supporting the Community

2022 Open Data Workshop
1000+ Participants
15 Locations + Virtual

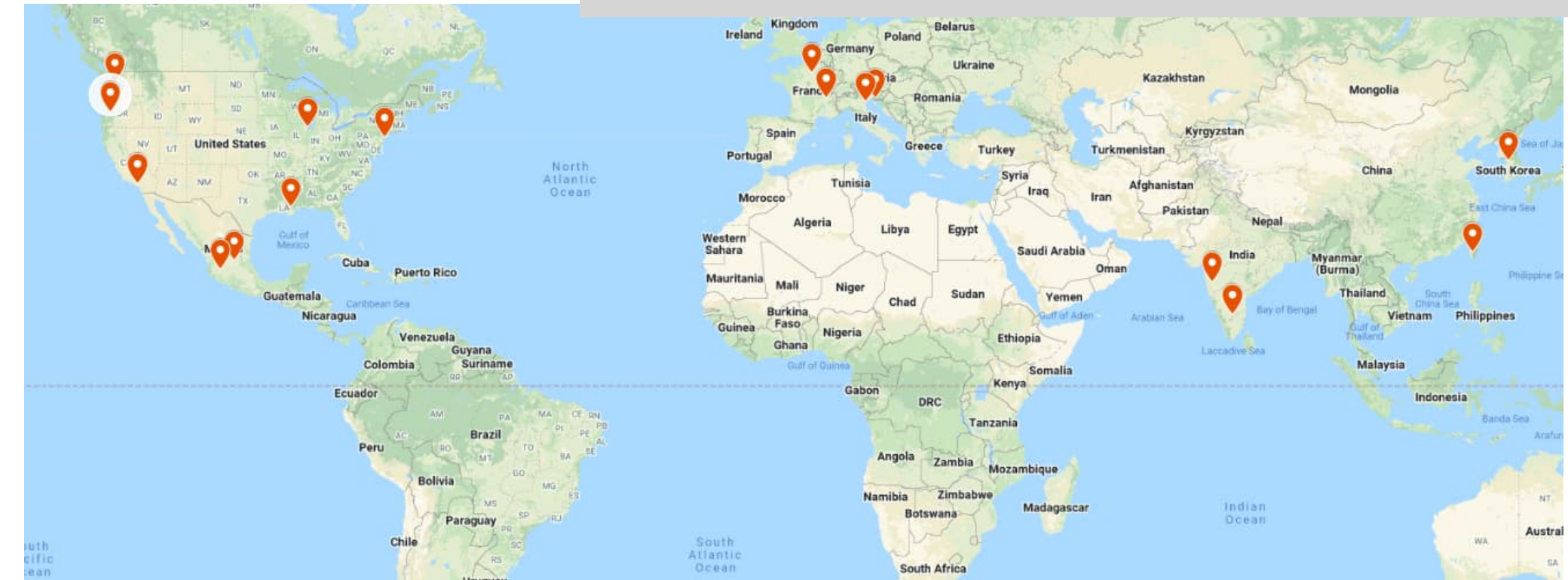
Discussion forum: <https://ask.igwn.org>

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>



Cloud Gallery Components Community Docs Blog

Sign in Sign up

CATEGORIES

- Streamlit templates
- Science & technology
- NLP & language
- Computer vision & images
- Finance & business
- Data visualization
- Geography & society
- Education
- Other

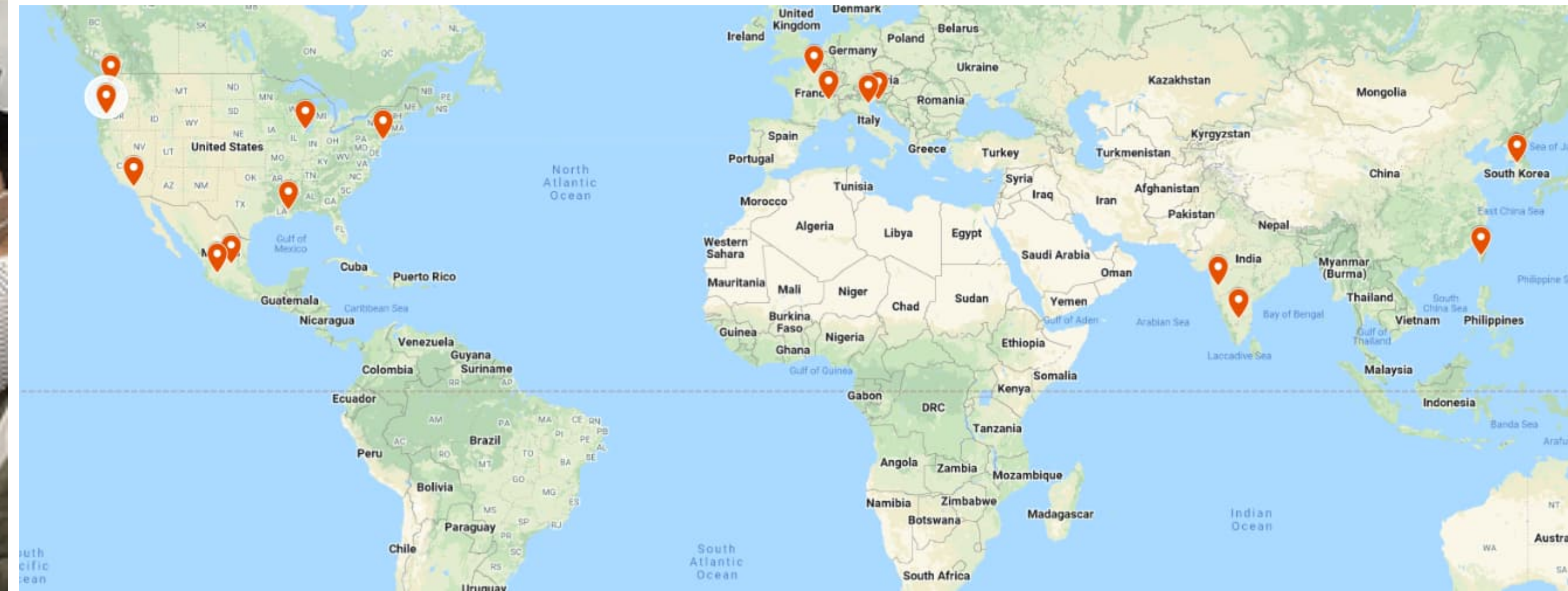
Bayesian Galaxy Zoo
This app in...
classification
model learns from volunteers and
by Mike Walmsley
View source code →
Go to app →

GW Quickview App
Featured on Streamlit Home Page
Attracted 9,000 views per month
by Eitan Halper-Stromberg and team
Go to app →

Gravitational Wave Quickview
This app downloads and displays a few seconds of data from the Gravitational Wave Open Science
by Jonah Kanner
View source code →
Go to app →



New York



Paris



Evanston, IL