MLAB @ CS&S Code for Science & Society

Measurement Lab

Internet Measurement - Open Source, Open Data

Chris Ritzo critzo@measurementlab.net

Slides: http://bit.ly/mlab-GÉANT-2020-03-04

Outline for Today

MLAB @ CS&S Code for Science & Society

- Introduction to M-Lab
- Government & Other broadband data sources
- M-Lab Tools for Community-Driven Data Collection
  - Survey, Speed Test, and Mapping - *Piecewise*
  - Automated, on-premise Measurement - *Murakami*
  - Self-hosted speed test server
- Wrap up / Questions

MLAB

@ CS&S Code for Science & Society

Introduction to M-Lab

M-Lab's Mission

MLAB
@ CS&S Code for Science & Society

Measure the internet.

Save the data.

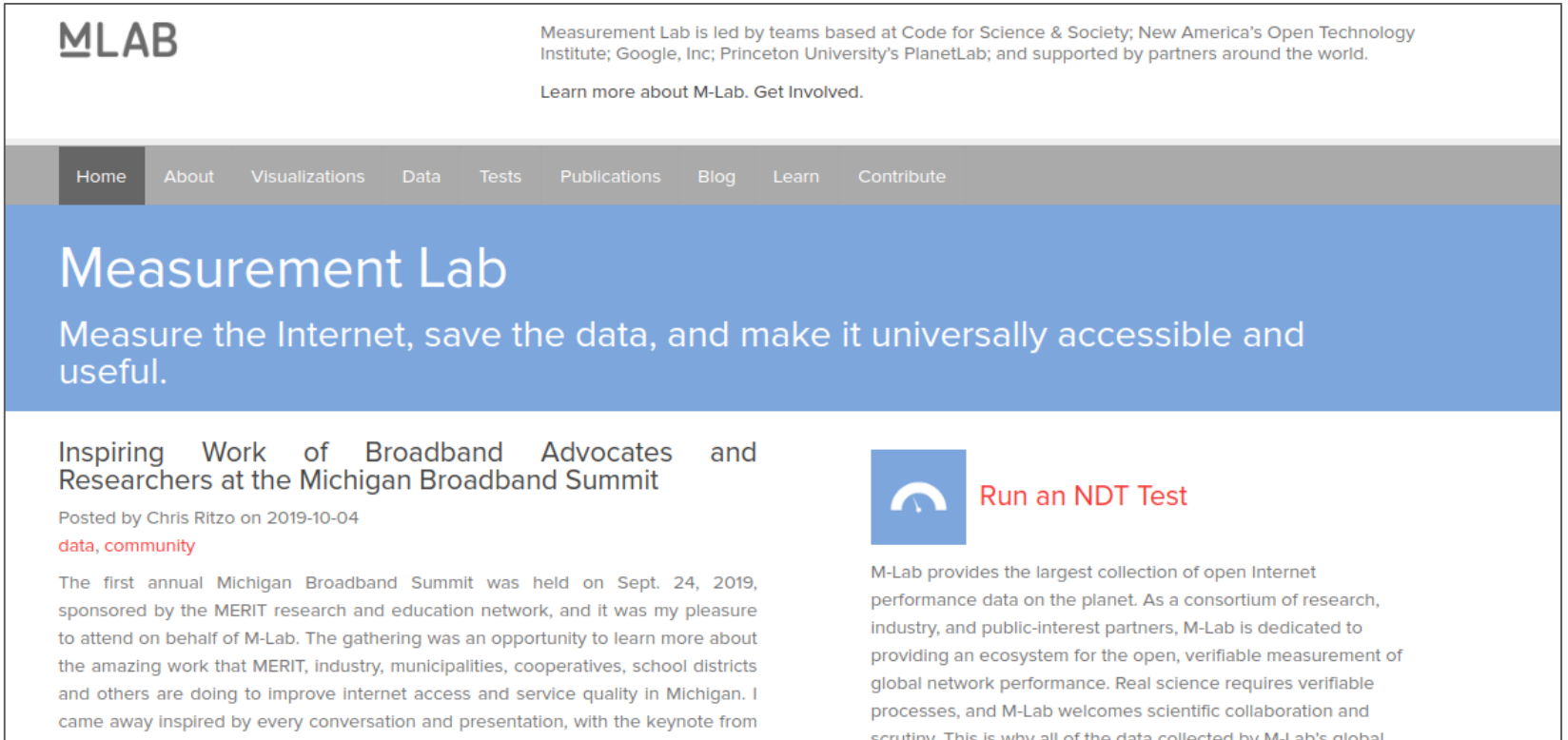
Make it universally accessible and useful.

Note: we don't measure the internet by ourselves -- people measure the Internet, using their own computers/phones and our servers, and we collect the data, and support them in their measurements

https://www.measurementlab.net/

MLAB

CS&S Code for Science & Society



What is M-Lab?

MLAB @ CS&S Code for Science & Society

- People/Organization: A joint initiative between staff at Code for Science & Society, Google, and many other contributing partners.
- Data: An open repository of user-contributed, longitudinal, open-source derived Internet infrastructure data
- Infrastructure: A global infrastructure deployed, built, and run to aid in the creation of that data repository

https://www.measurementlab.net/status

MLAB

CS&S Code for Science & Society

Infrastructure Map

The M-Lab infrastructure map displays information about our server pods around the world. Clusters of M-Lab pods are displayed as dark-blue and light-blue circles with a count of pods in the center. Clicking on a cluster will display a pop up with info about each pod in the cluster. Single small dots show locations with only one server. Clicking on a single pod will display a pop up with info about that pod. Double-click on the map to zoom in, or scroll in and out to zoom.

Our Organizing Principles

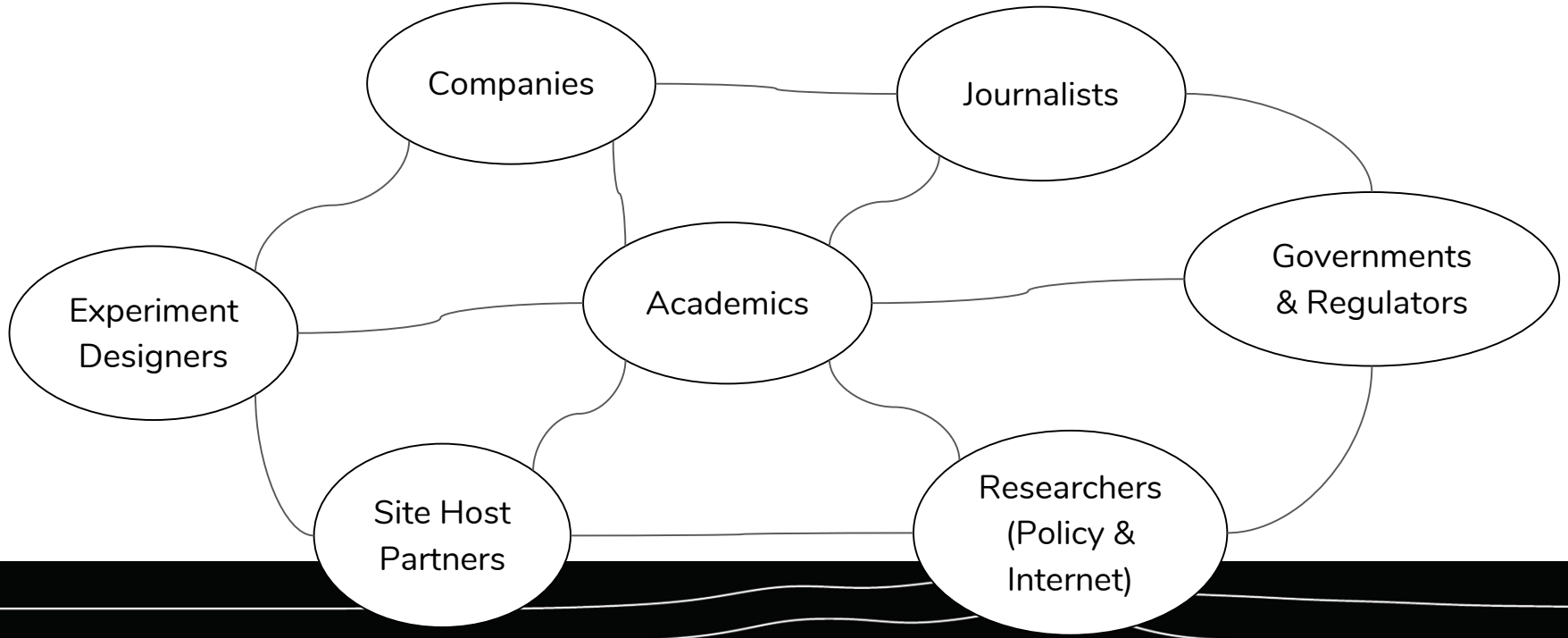
MLAB @ CS&S Code for Science & Society

- Active Measurements
  - All synthetic data, we take user privacy seriously.
  - Client initiated tests only, Servers do not start tests on their own
- Experiments are curated and approved by a panel of reviewers, primarily developed by academics
  - Longitudinal research and data
- Clients come from the community
  - Anyone can develop them
  - Anyone can run tests against the platform
- Openness
  - All of the data is openly licensed (CC0 1.0 Universal (CC0 1.0) Public Domain)
  - All of the code is open source.

Learn more about M-Lab: 10th Anniversary Convening blog post, videos: https://www.measurementlab.net/blog/mlab-10year-wrapup/

Our Community

MLAB @ CS&S Code for Science & Society



MLAB
@ CS&S Code for Science & Society

M-Lab Broadband Data

How M-Lab Collects Tests

MLAB

https://www.measurementlab.net/faq/

1. The consumer starts an M-Lab test

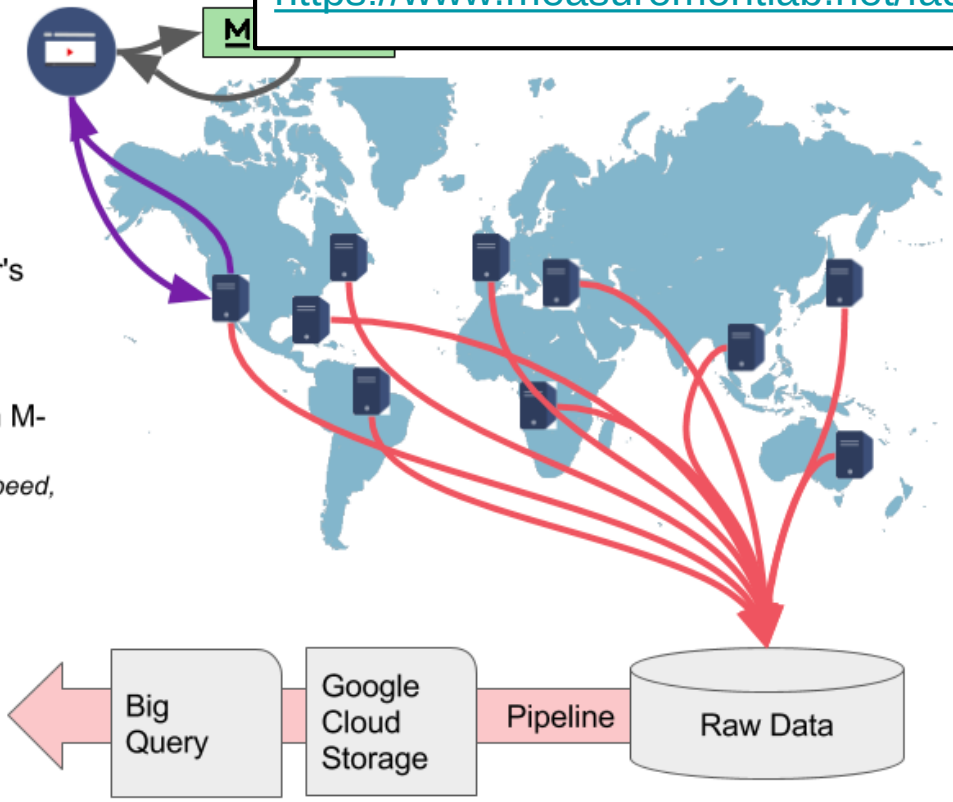
2. The M-Lab service provides closest server to use for the test and for collecting data

3. Test is then run between consumer's browser and M-Lab's server

4. The measurements are shown to consumer, and the data are stored on M-Lab's server
(Measurements include upload/download speed, round trip time, etc.)

5. Completed test data are then sent to Google's Cloud Storage and to Google's BigQuery

6. Through applications that use the collected data, the consumer's questions are answered



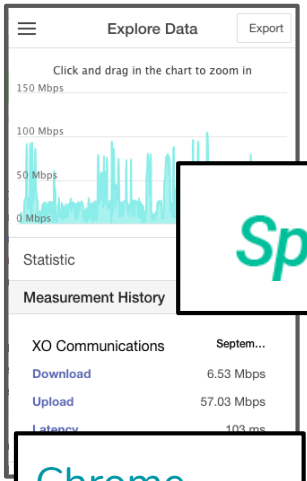
NDT tests come from lots of places

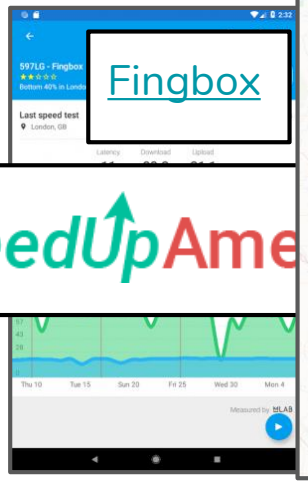
MLAB

Network Diagnostic Tool

MLAB INTERNET 2

CS&S Code for Science & Society

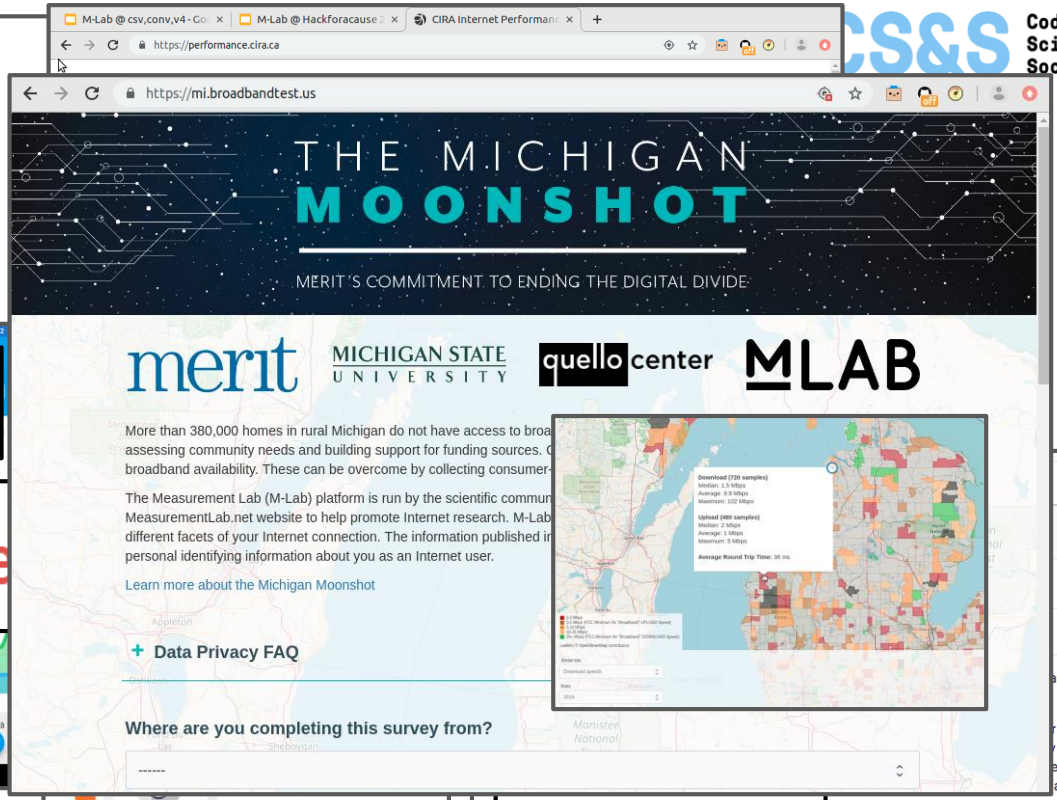


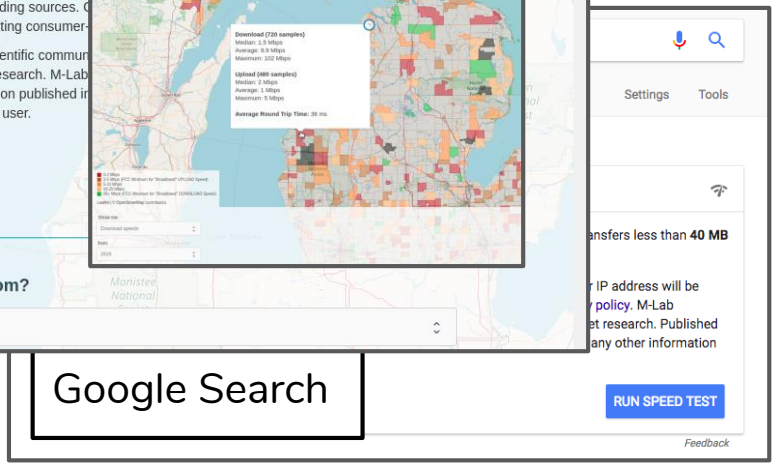


Fingbox

SpeedUpAme

Chrome Extension



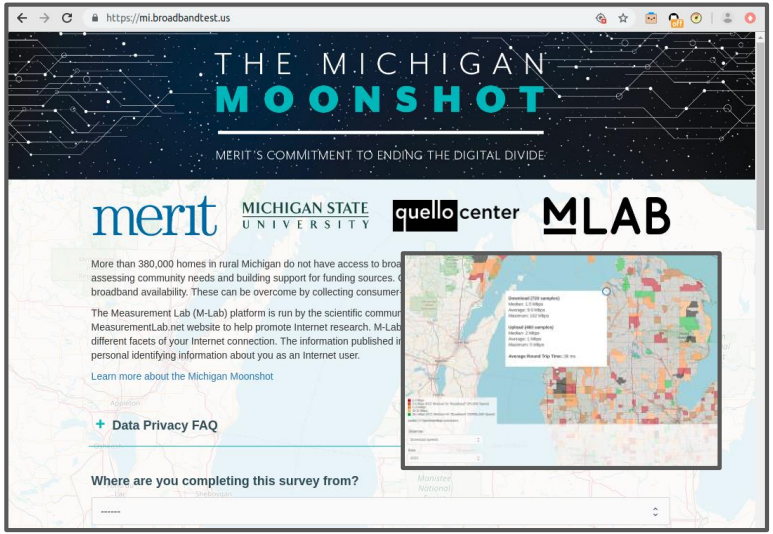


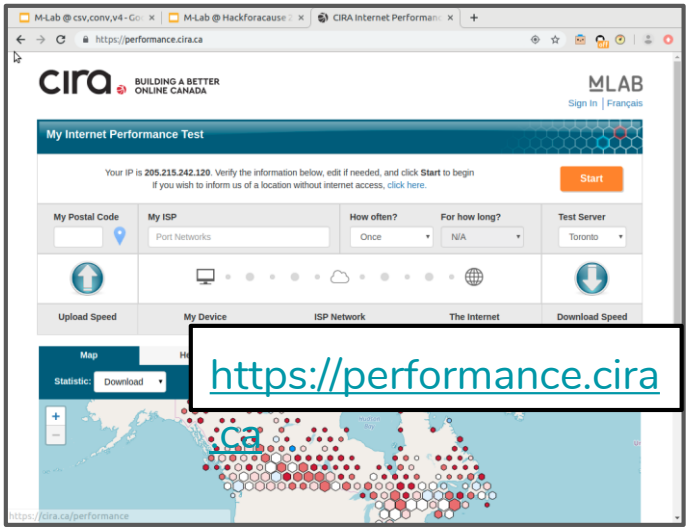
Google Search

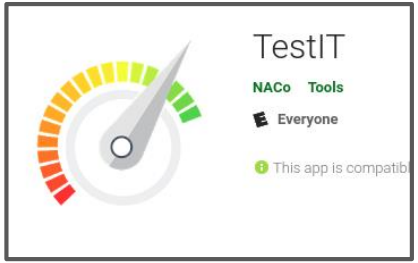
Custom NDT Integrations Gather Enhanced Geolocation, but Privately @ CS&S

MLAB

Code for Science & Society







SpeedUpAmerica

M-Lab Data Sources

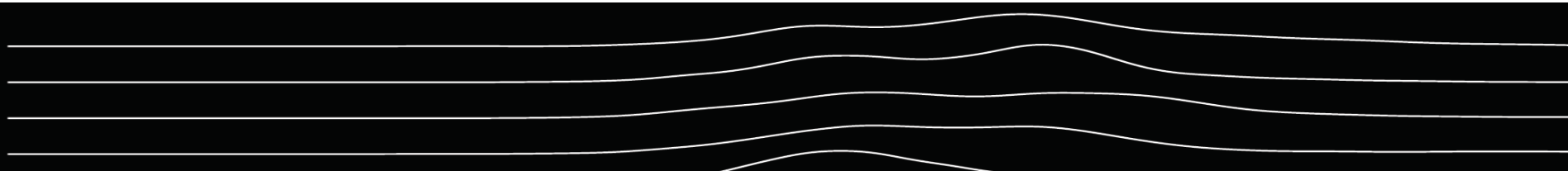
Network Diagnostic Tool - NDT

MLAB @ CS&S Code for Science & Society

- NDT measures “single stream performance” or “bulk transport capacity” as defined in IETF’s RFC 3148 to the nearest available “off-net” M-Lab server

- M-Lab measurements are not expected to be equal to other speed test sources https://www.measurementlab.net/faq/#why-are-my-m-lab-results-different-from-other-speed-tests

- https://tools.ietf.org/html/rfc3148
- https://github.com/m-lab/ndt-server/tree/master/ndt7
- https://www.measurementlab.net/blog/tcp-and-bbr/
- https://www.measurementlab.net/blog/speed-tests-accuracy



MLAB

@ CS&S Code for Science & Society

Government & Other Broadband Data Sources

FCC Data Sources

MLAB @ CS&S Code for Science & Society

Measuring Broadband America Program

Fixed Broadband Annual Study & Report

- Conducted by SamKnows
- Methodology & reports published by FCC

Mobile Broadband App

- Android & iOS app
- De-identified data published by FCC

MBA-Assisted Research Studies (MARS)

- FCC partnership with leading Internet measurement academic researchers
- Projects, data, and methods published by the FCC

FCC Data Sources

MLAB @ CS&S Code for Science & Society

Form 477 Provider Coverage Data

- All facilities-based broadband providers are required to file data with the FCC twice a year (Form 477) on where they offer Internet access service at speeds exceeding 200 kbps in at least one direction.
- Fixed providers file lists of census blocks in which they can or do offer service to at least one location, with additional information about the service.
- Mobile providers file maps of their coverage areas for each broadband technology (e.g., EV-DO, HSPA, LTE).

Differences in Data Sources

	M-Lab NDT	NDT Integration with enhanced Geo	Form 477 Fixed	Form 477 Mobile
Geography	Aggregate by any geography. Test geolocation based on IP address	Aggregate by any geography. Test geo based on GPS, HTML5, address	Census Block	Shapefile defining coverage areas
Speed metrics	*Measured* speeds to closest *off-net* server	*Measured* speeds to closest *off-net* server	Maximum advertised speeds. Likely reported *on-net* speeds.	Minimum *advertised* speeds are collected, but not made public. Can be requested
Mobile vs. Fixed	All results in M-Lab BigQuery dataset. Can delineate mobile/non-mobile using 3rd party services.	All results in M-Lab BigQuery, *and* in 3rd party integrator maintained database	Fixed only	Mobile only
Provider Information	ASN - Autonomous System Number/Name	ASN - Autonomous System Number/Name, and/or via user survey	- Provider Name - DBA Name - "Doing business as" - Holding Company Name	- Provider Name - DBA Name - "Doing business as" - Holding Company Name

Recommended Use of M-Lab NDT Data and 3rd Party “Enhanced” NDT Data

MLAB @ CS&S Code for Science & Society

M-Lab’s tools and data are useful because of their openness, transparency, and accessibility.

Our data are most useful in demonstrating the overall network connection speed and service quality. It is not a replacement for FCC data, or an apples-to-apples comparison to Form 477.

M-Lab data is a well respected, public dataset of *measurements* not simply *reported or advertised* service levels.

Gathering “Enhanced” NDT data provides more precise location accuracy, and can help support your dialogue and advocacy with legislators and agency staff.

MLAB

@ CS&S Code for Science & Society

M-Lab Tools for Community-Driven Data Collection

MLAB

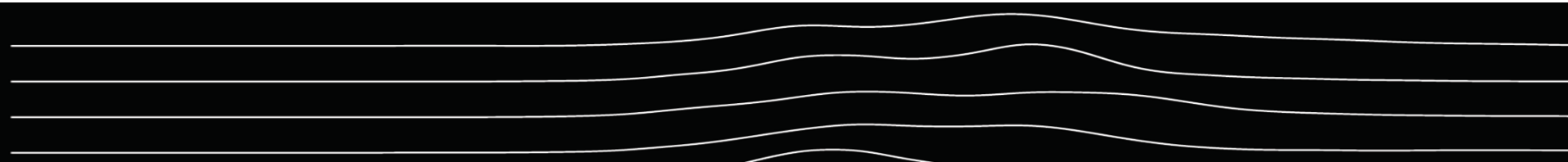
@ CS&S Code for Science & Society

Survey, Speed Test, and Mapping: *Piecewise*

Piecewise

MLAB @ CS&S Code for Science & Society

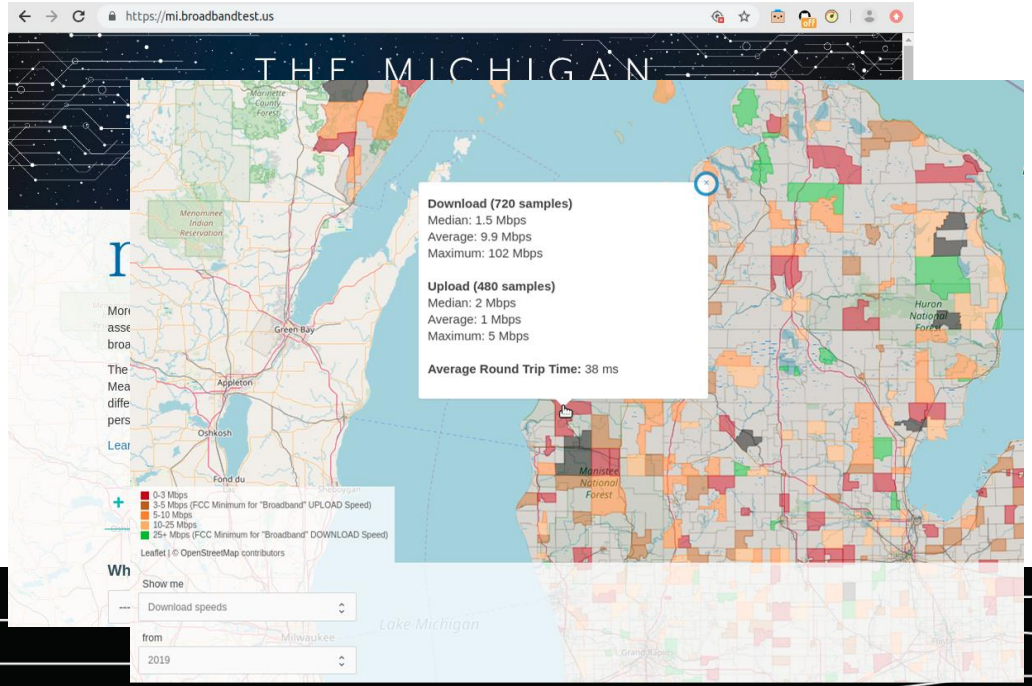
Piecewise is an open-source public engagement portal that collects both user-volunteered survey responses and speed test data using the Measurement Lab platform. Data collected by Piecewise is visually aggregated on the web and mapped on top of M-Lab's public dataset.



Piecewise

MLAB @ CS&S Code for Science & Society

A Survey & Mapping tool



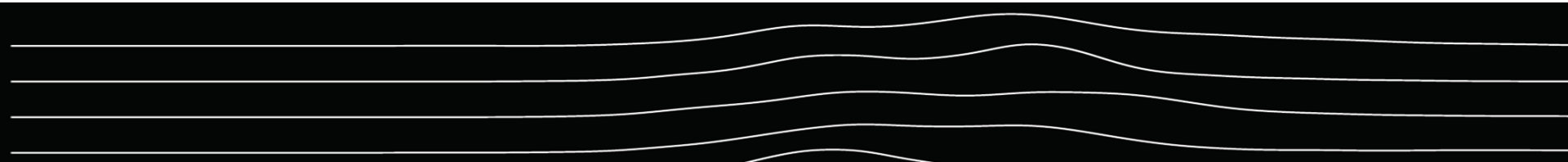
- Michigan / MERIT - broadband testing & mapping with a homework gap hook
  - R&E network, K12 statewide systems
  - Homework assignment to run a test
- https://mi.broadbandtest.us/

Piecewise

MLAB @ CS&S Code for Science & Society

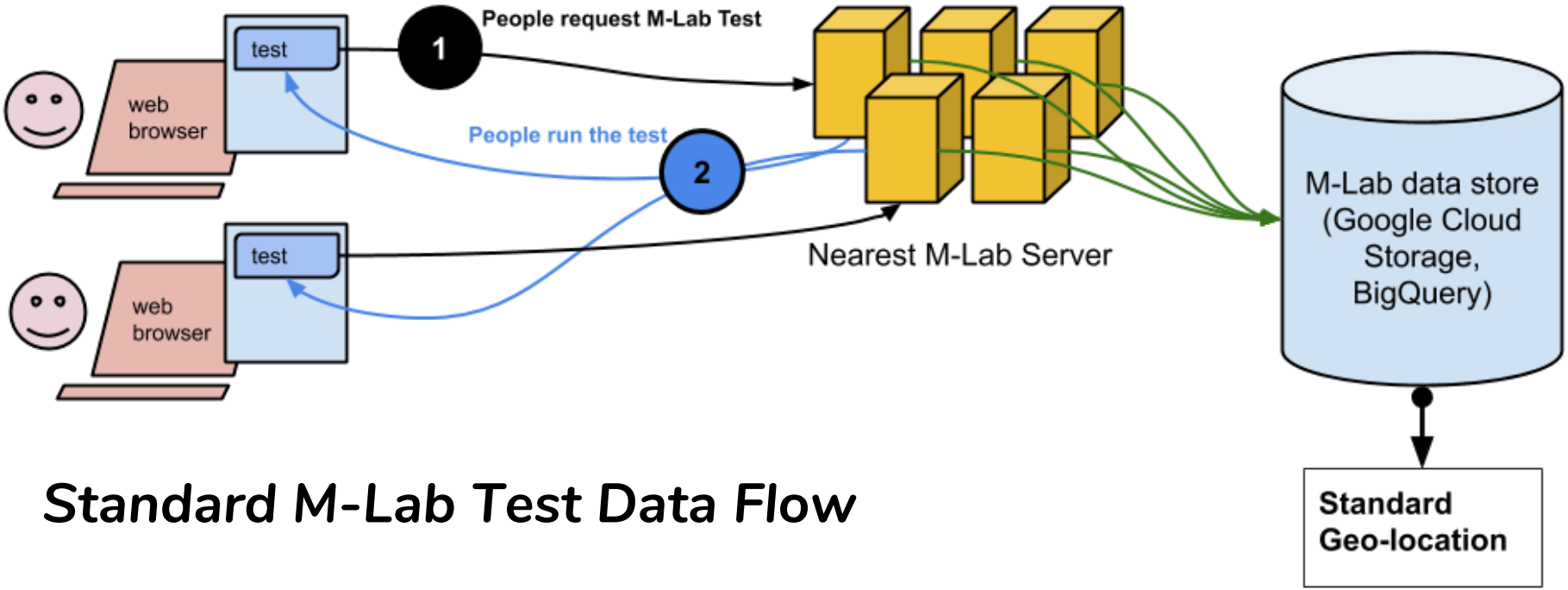
A Survey & Mapping tool

- Status?
  - ~Q2/Q3 2020, Piecewise will become available as a SaaS app
- What do you use it for?
  - Collecting enhanced geolocation from data in specific area(s)
    - For example: state collecting data about each of its counties, or an advocacy organization collecting information for a report
  - Mapping the collected data on the web
- Goal:
  - Communities collecting crowd-sourced data to advocate for their broadband needs



How Does Piecewise Work?

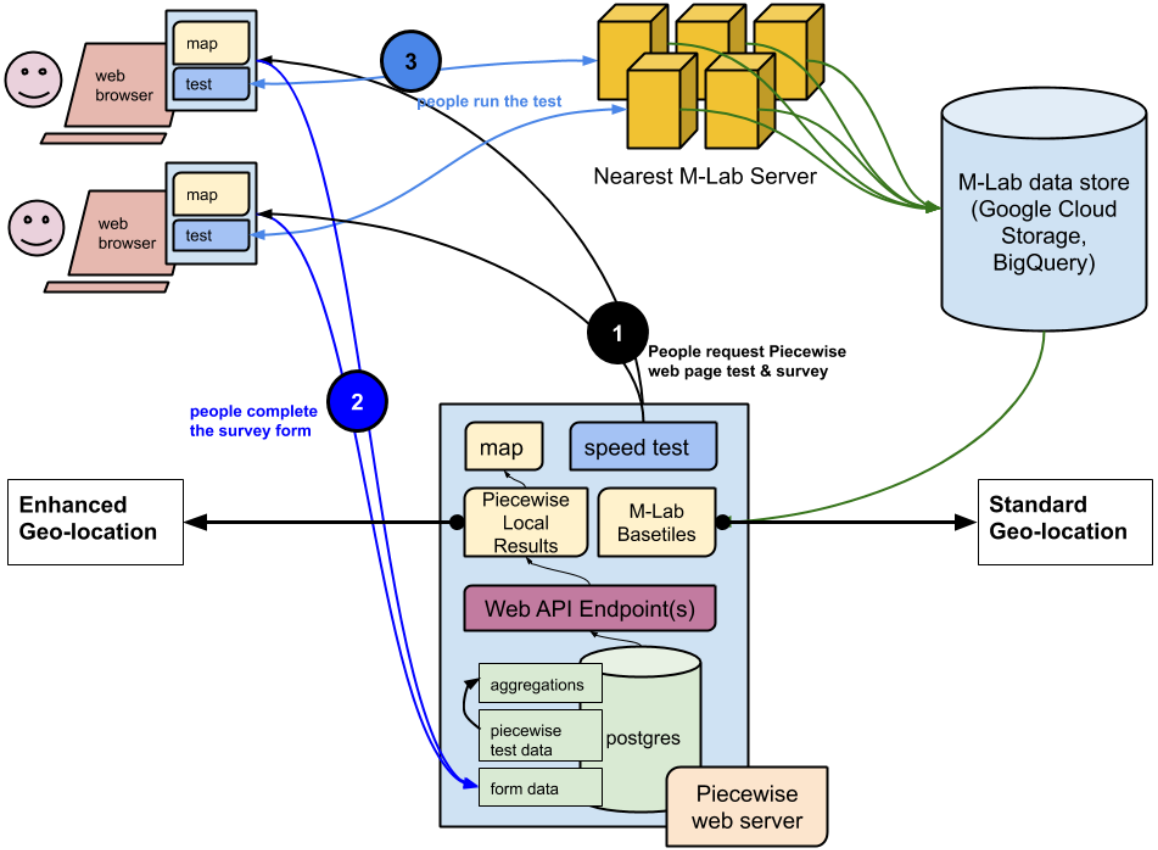
MLAB @ CS&S Code for Science & Society



Standard M-Lab Test Data Flow

How Does Piecewise Work?

MLAB @ CS&S Code for Science & Society

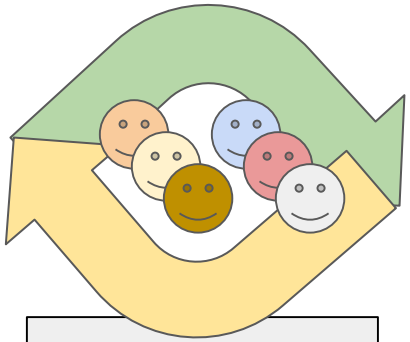


Piecewise Instance Test Data Flow

How Does Piecewise Work?

MLAB @ CS&S Code for Science & Society

- Design & publish survey questions
- Customize branding
- Publicize



People share their location, complete your survey, run speed test

- Standard test data sent to M-Lab
- Test results, survey answers, and enhanced location saved in your Piecewise instance

You have M-Lab Measurements with geographic accuracy, along with survey results for your analyses, planning, etc.

MLAB

@ CS&S Code for Science & Society

Automated, On-premise Measurement:

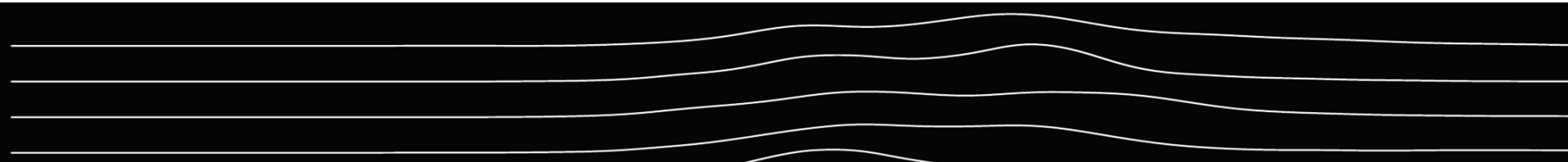
Murakami

Measurement Devices (Murakami)

MLAB @ CS&S Code for Science & Society

An automatic data collection tool

Murakami is a tool for creating an automated internet measurement service, running in a Docker container. A Murakami measurement container will automatically run supported tests four times a day using a randomized schedule, and can be configured to export each test result to a local storage device, to one or more remote servers via SCP, or to a Google Cloud Storage bucket. Results are saved as individual files in JSON new line format (.jsonl).

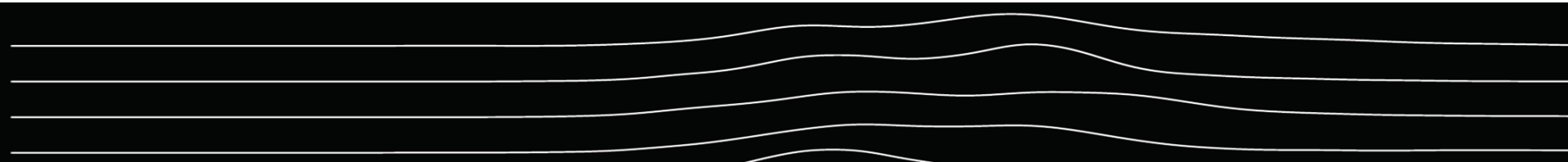


Measurement Devices (Murakami)

MLAB @ CS&S Code for Science & Society

An automatic data collection tool

- Status?
  - Currently in development as part of an IMLS grant in partnership with Simmons University and Internet2
    - More info about Measuring Libraries Broadband Networks here
  - M-Lab is starting to do research about users outside of libraries
- What do you use it for?
  - Running regular tests where broadband is regularly provided
    - For example: schools, libraries, community centers
- Goal:
  - Communities collecting data to advocate for their broadband needs



MLAB

@ CS&S Code for Science & Society

Measurement Devices (Murakami)

How does it work?

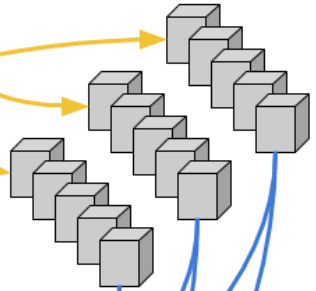
Code managed in Github
https://github.com/m-lab/murakami

- Modular “test runners”
  - ndt5
  - ndt7
  - DASH
  - speedtest-cli (single & multi stream)
- Data exporters
  - Google Cloud Storage
  - Secure Copy Protocol
  - Local storage

balenaCloud

- IoT device provisioning & management
- Remote access
- Remote updates

- Build and push software releases to provisioned devices



Tests run and push results to defined exporter locations

SCP Server

Google Cloud Storage

Visualization Web Service
(in development now)

- Provide charts of collected data
- Select / interact with data
- Export / download data
- Annotation feature

- Will use Docker so it can be hosted anywhere
- GCP VM, other cloud service VM, local server

Measurement Devices (Murakami)

MLAB @ CS&S Code for Science & Society

Provisioning a single device

1. Flash SD card with stock Balena OS

2. Run configure script, make any required manual configurations

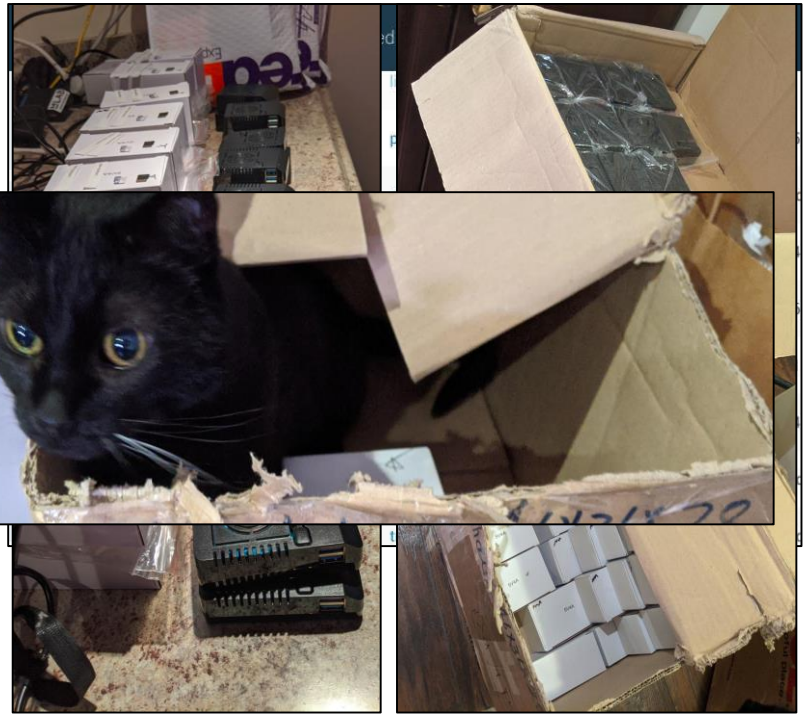
3. Obtain MAC address

```
[connection]
id=resin-ethernet
type=ethernet
interface-name=eth0

[ethernet]
mac-address-blacklist=

[ipv4]
method=auto
dns={ custom DNS servers };
dns-search=
```

4. Connect them!



MLAB

@ CS&S Code for Science & Society

Self-hosted speed test server:

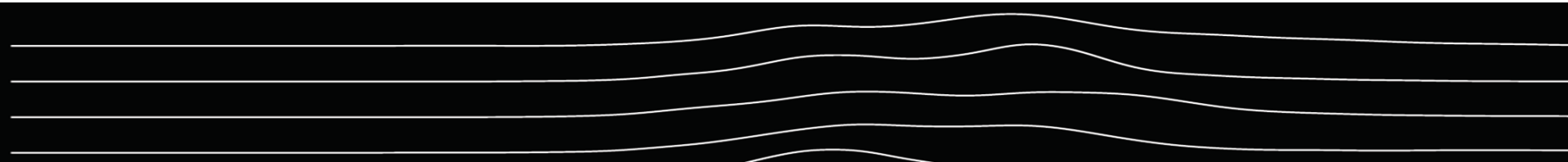
Run M-Lab NDT server in a non- M-Lab setting

Host your own NDT server

A standalone speed test

MLAB @ CS&S Code for Science & Society

- Status?
  - Able to “test drive” now using M-Lab’s image on Docker Hub
  - Includes all M-Lab “sidecar services”
  - Build your own Docker images for your production deployment (your certs, storage, etc)
- What do you use it for?
  - Test the performance of networks that are not on the public Internet
  - Test between campus science DMZ & cloud storage
  - Test performance of layer 2 networks
- Goal:
  - Allow anyone to run M-Lab tools on their own infrastructure



Test drive your own ndt-server

MLAB @ CS&S Code for Science & Society

On a Linux machine with docker & updated kernel, run:

```
docker run --net=host measurementlab/ndt
```

Then point your browser to:

ndt5 (original proto, http)	http://localhost:3001/static/widget.html
ndt5 (original proto, TLS)	https://localhost:3010/static/widget.html
ndt7	https://localhost/static/ndt7.html

More information: https://www.measurementlab.net/blog/run-your-own-ndt-server/#running-your-own-ndt-server

Running your own production ndt-server

MLAB @ CS&S Code for Science & Society

- Obtain the server code https://github.com/m-lab/ndt-server
- Build a “full stack” image using details from your host:
  - https://github.com/m-lab/ndt-server/tree/master/fullstack
- Run your full stack image instead of our demo image

MLAB

@ CS&S Code for Science & Society

Wrap up / Questions

MLAB @ CS&S Code for Science & Society

Measurement Lab

Internet Measurement - Open Source, Open Data

Chris Ritzo critzo@measurementlab.net

Slides: http://bit.ly/mlab-GÉANT-2020-03-04