Open Internet Performance Data, and How to Get it

Meredith Whittaker
Tiziana Refice

OECD, WPIIS, December 14, 2012

Why do we care about measurement?

To maintain and grow the open, multi-stakeholder Internet for the future, all stakeholders must have access to data about the Internet today.

To get this data, we must measure.

M-Lab's founding principle: Openness

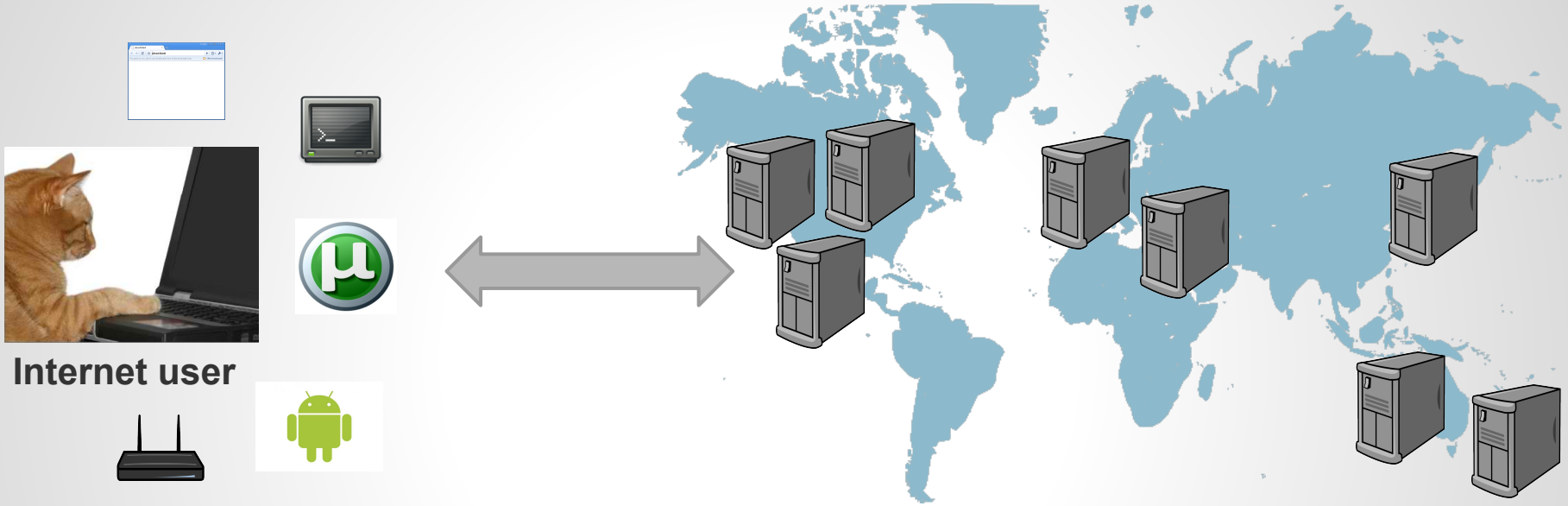
Openness means making room for real science

Real science means replicable facts, not conjecture

How does M-Lab do this?

- **Open source**, publicly documented **server platform**
- **Open source** **experiments** built by researchers
- **Openly available**, freely accessible **data**

Open-source measurement tools



- Client-server, web-based, CPE & mobile
- Active
- Open-source

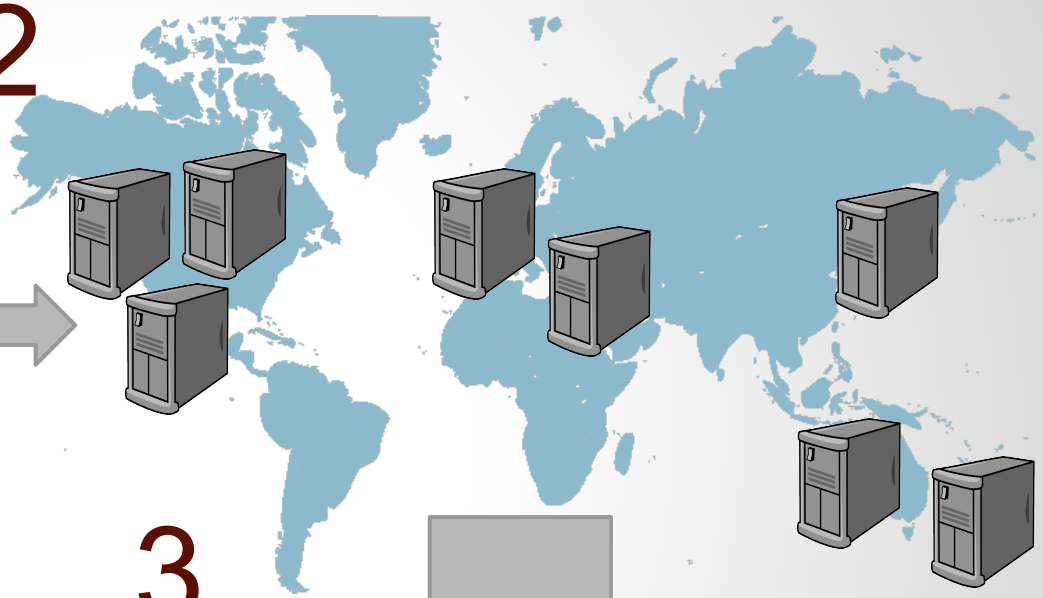
M-Lab's open data ecosystem

1

Internet user

Sam Knows

2



3

+600TB!

Cloud Storage

BigQuery

raw data

700B rows

4

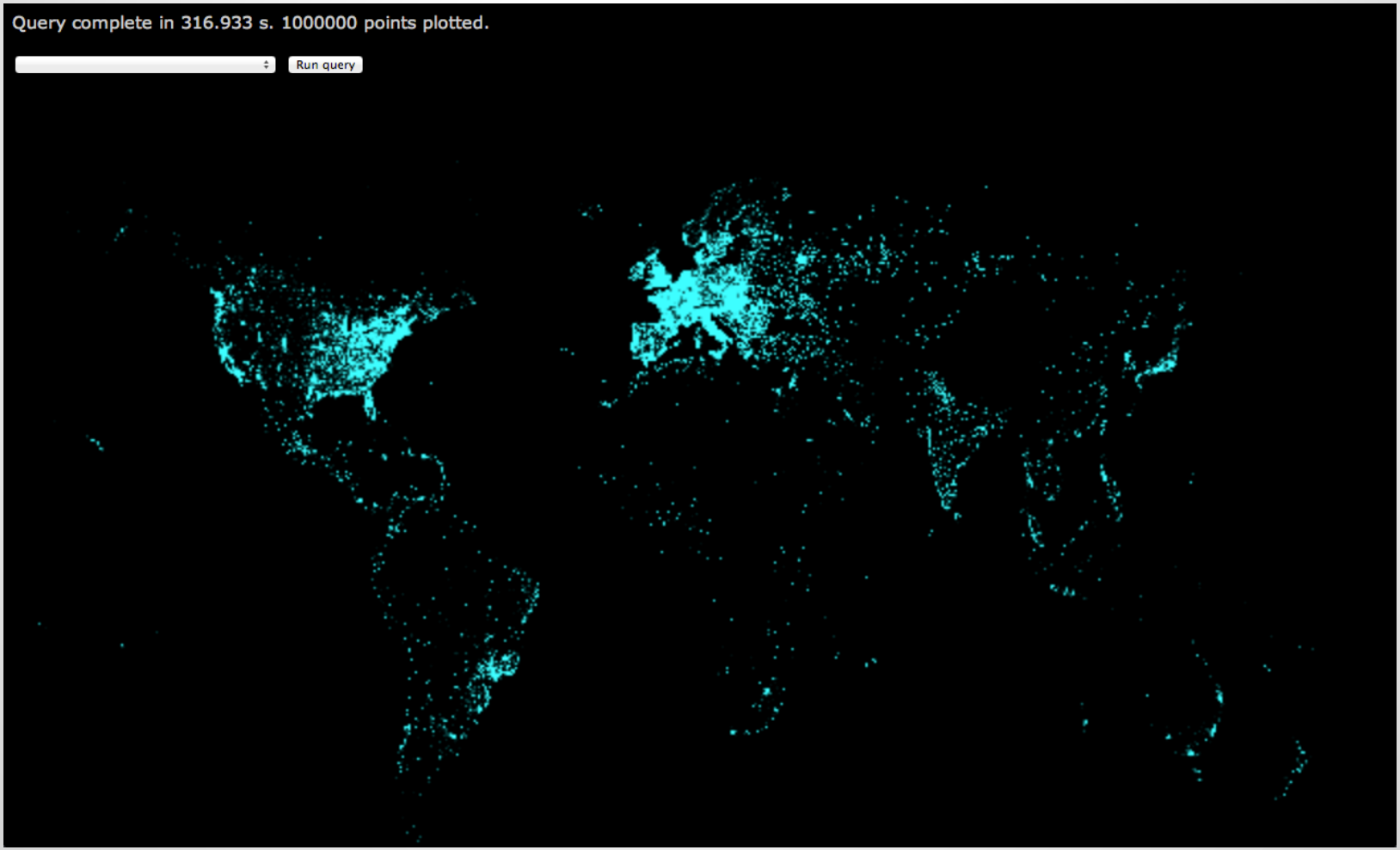
Internet Stakeholders

Researcher

Regulator

Open data, world-wide

Global tests over 3 months, more here.



Measuring from the device

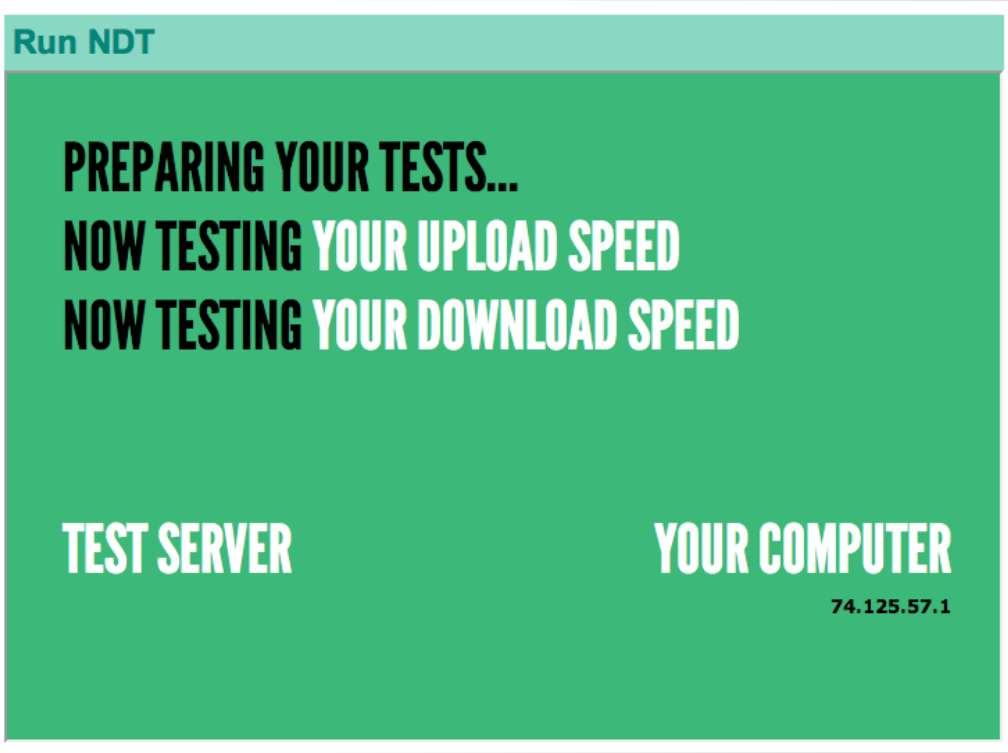
M-Lab examples:

NDT, NPAD,

Glasnost

Pros:

Broad distribution;
Data on real experience;
Data on device impact



Cons:

Difficult to isolate access link,

Data potentially impacted by resource use

Measuring from the router (CPE)

M-Lab examples:

Bismark; SamKnows

Pros:

Able to isolate access link;
Controlled panels/sampling;
Controlled environment

Cons:

Only measures access link
(not user experience);
Expensive; Difficult to scale;

Mobile measurement

M-Lab examples:

Mobiperf; NDT mobile app

Evolving and new.

Challenges:

Measurement space contains new dimensions;

Useful testing models;

Rapidly expanding devices and use-cases

Measurement: it isn't just about the client

Server end-point impacts measurement

Accurate measurement requires:

- Control of the servers
- Deterministic server selection
- Consistently instrumented and deployed infrastructure
- Broad geographical coverage

Measurement: it isn't just about the client

Good data takes good stewardship

Meaningful, useful data requires:

- Clear, accurate versioning
- Useful, machine-readable formats
- Robust collection pipeline
- Consistent measurement over time
- Creative solutions to protect privacy while retaining openness

Conclusions

To be useful and accurate, ***measurement and analysis should be open.*** Openness makes room for the scientific process.

- There are many ways to measure, each with benefits. ***Make room for diversity and change***

- The ***measurement platform (servers) is as important*** as the measurement tool