AIMS Workshop
April 1, 2015

ICLAB:
A PLATFORM FOR MEASURING ONLINE INFORMATION CONTROLS

PHILLIPA GILL

STONY BROOK UNIVERSITY

ACKS:

SAM BURNETT, MASASHI CRETE-NISHIHATA, JAKUB DALEK, RONALD J. DEIBERT, NICK FEAMSTER, SATHYA GUNASEKARAN, BEN JONES, TZU-WEN LEE, ABBAS RAZAGHPANAH, ADAM SENFT, GREG WISEMAN, JUSTIN WONG

WHY ICLAB?

- **The Citizen Lab has developed a unique network of individuals around the world to measure censorship**
  - …but software support is lacking
- **Running tests requires human coordination**
- **Interpreting results is mostly manual**
- **Existing approach has been in place for nearly 10 years**

Idea: *Let's revisit the problem of designing a measurement platform for online information controls from the network measurement perspective*

CENSORSHIP MEASUREMENTS 101

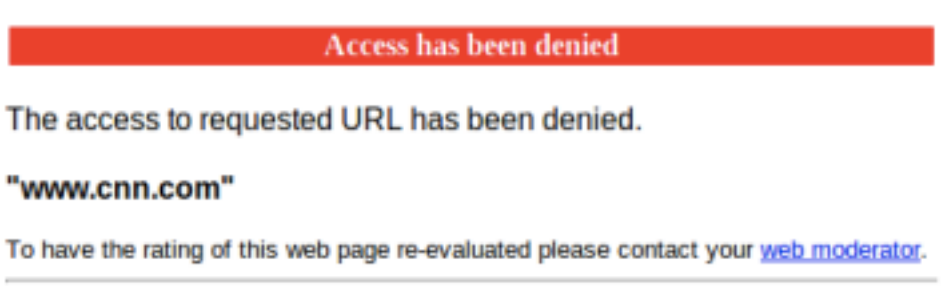
- **Basic approach**
  - Fetch a Web page from a location with suspected censorship – the field
  - Fetch the same Web page simultaneously from a location without censorship – the lab
  - Compare the results

CENSORSHIP MEASUREMENT 101

Example:

Measured in the lab

Measured in the field

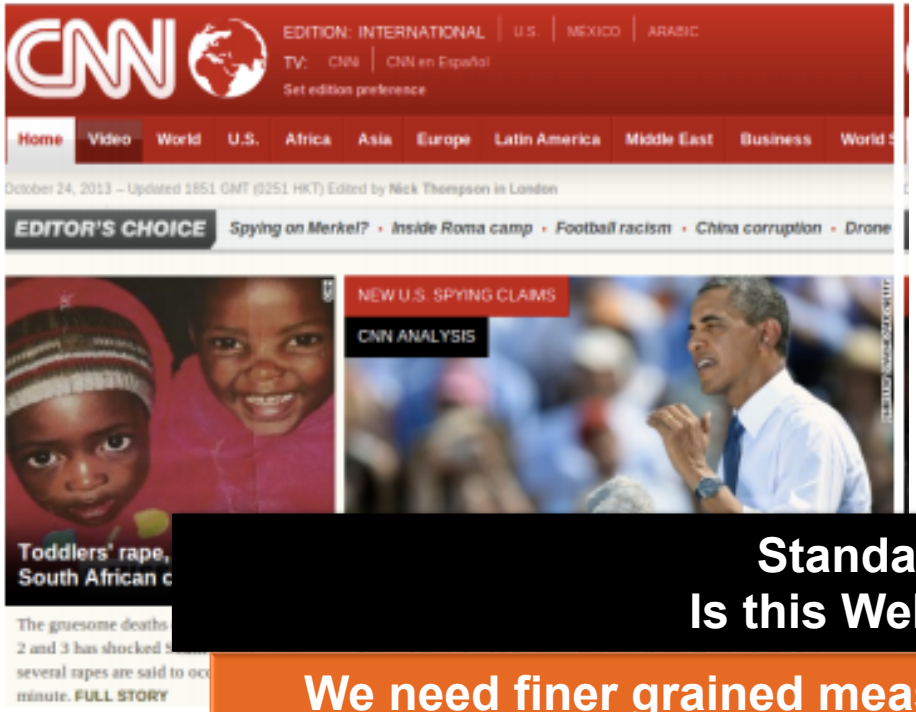


Standard question:
Is this Web site blocked?

CENSORSHIP MEASUREMENT 101

Example:

Measured in the lab



Measured in the field

(no html page returned)

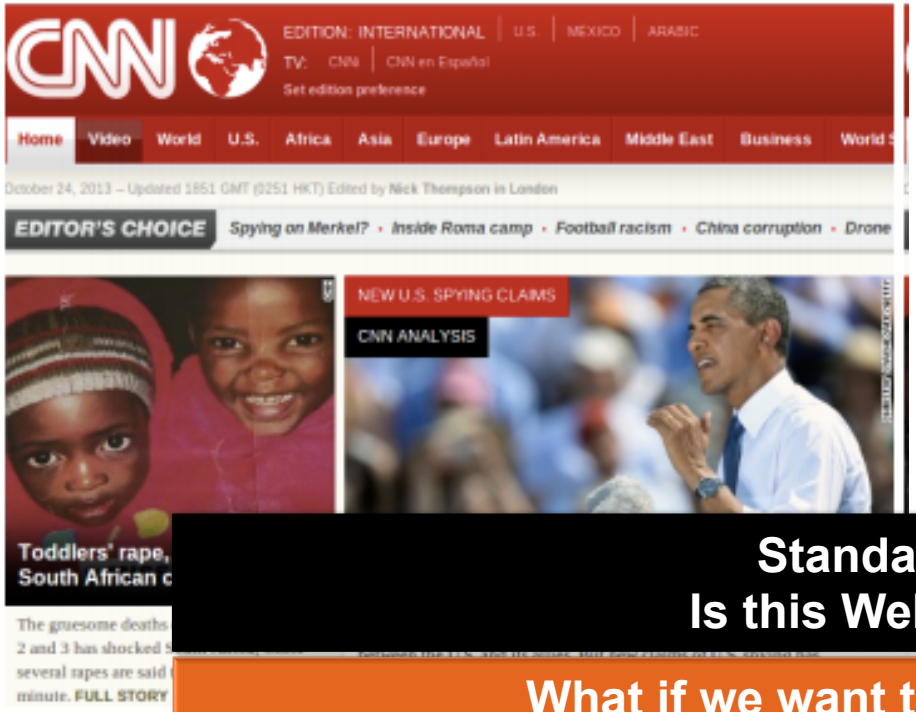
Standard question:
Is this Web site blocked?

We need finer grained measurements to answer this question!

CENSORSHIP MEASUREMENT 101

Example:

Measured in the lab



Measured in the field

(no html page returned)

Standard question:
Is this Web site blocked?

What if we want to ask more questions:
How was this site blocked?
What product was used to block it?
Who is blocking it?

WHAT DOES THIS MEAN FOR ICLAB?

- **Platform should support a wide range of network measurement operations**
  - **Basics**: HTTP request, Traceroute, DNS queries
  - **Not-so-basics:** HTTP header fingerprinting (Netalyzr test)

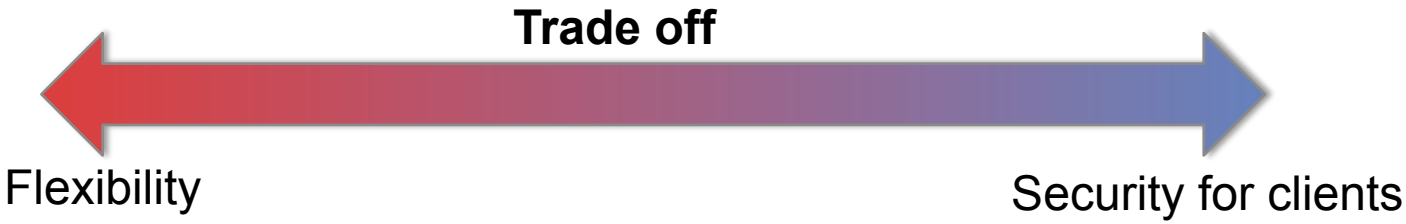
CoNtEnT tYpE: text/html
(sent by client)

CONTENT TYPE: text/html
(received by server)

  - **Even-less-basics:** Customized IP TTL header to localize the censor in the network
  - **Detecting other information controls**: traffic differentiation, surveillance etc.

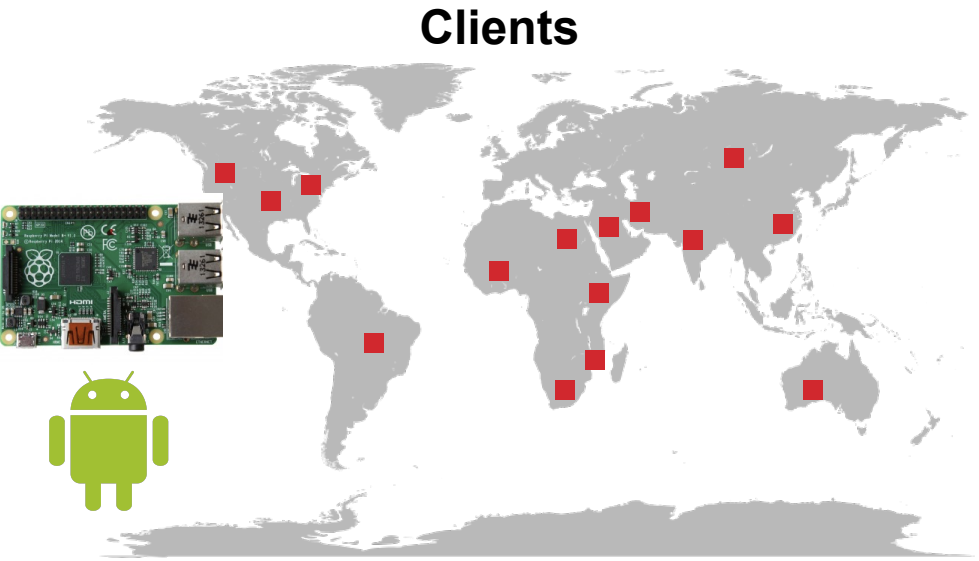
WHAT DOES THIS MEAN FOR ICLAB?

- **Impossible to know the complete set of measurements that need to be supported a priori**
  - New censorship technologies emerge, we need to be able to keep up
- **Need to be able to implement and launch new experiments on demand**
- **Need to be flexible about when, where, and what is run**
- **How to do this well?**



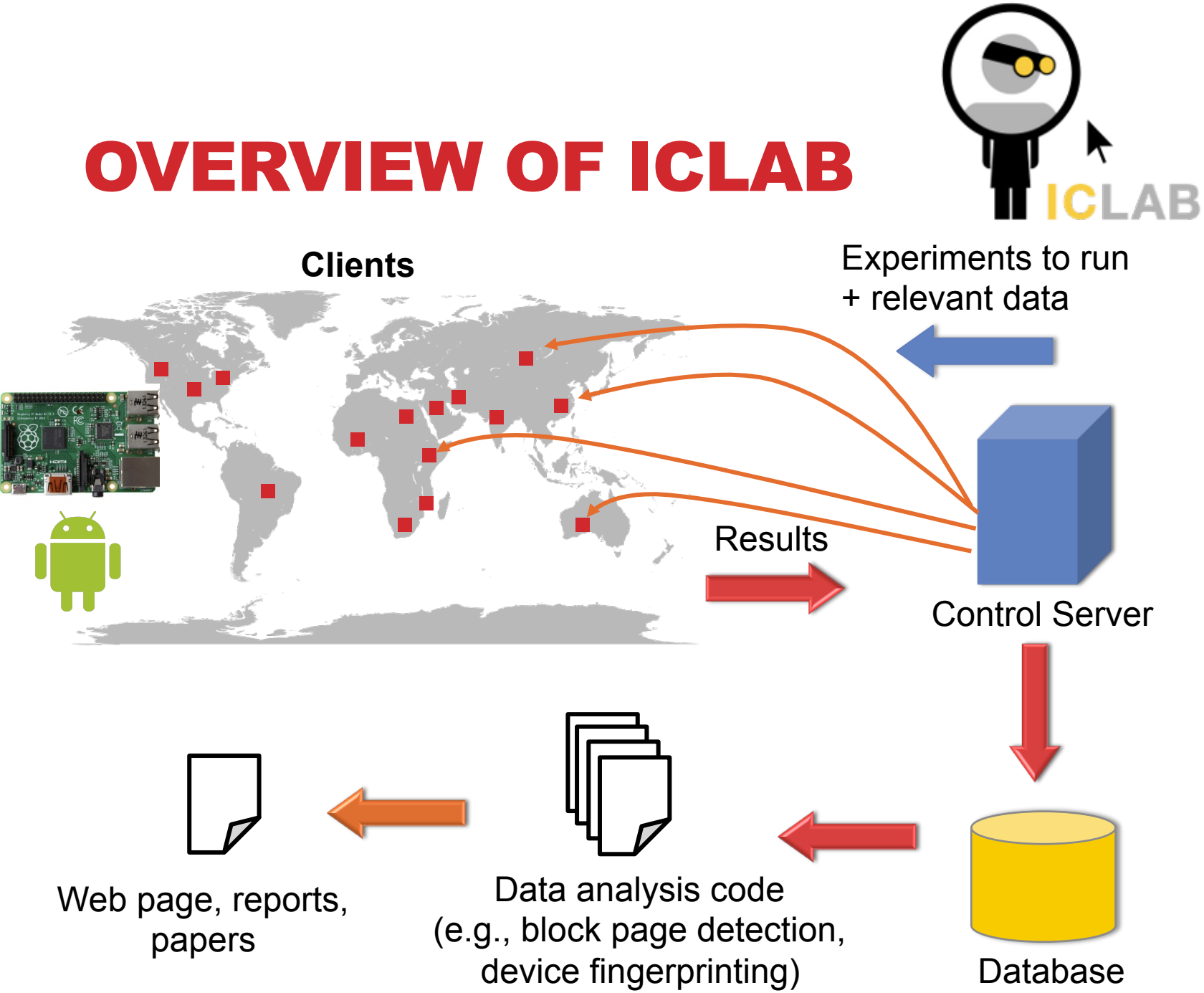
- **Our solution: Python experiment specification + Web UI**

OVERVIEW OF ICLAB



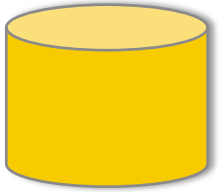
OVERVIEW OF ICLAB

OVERVIEW OF ICLAB



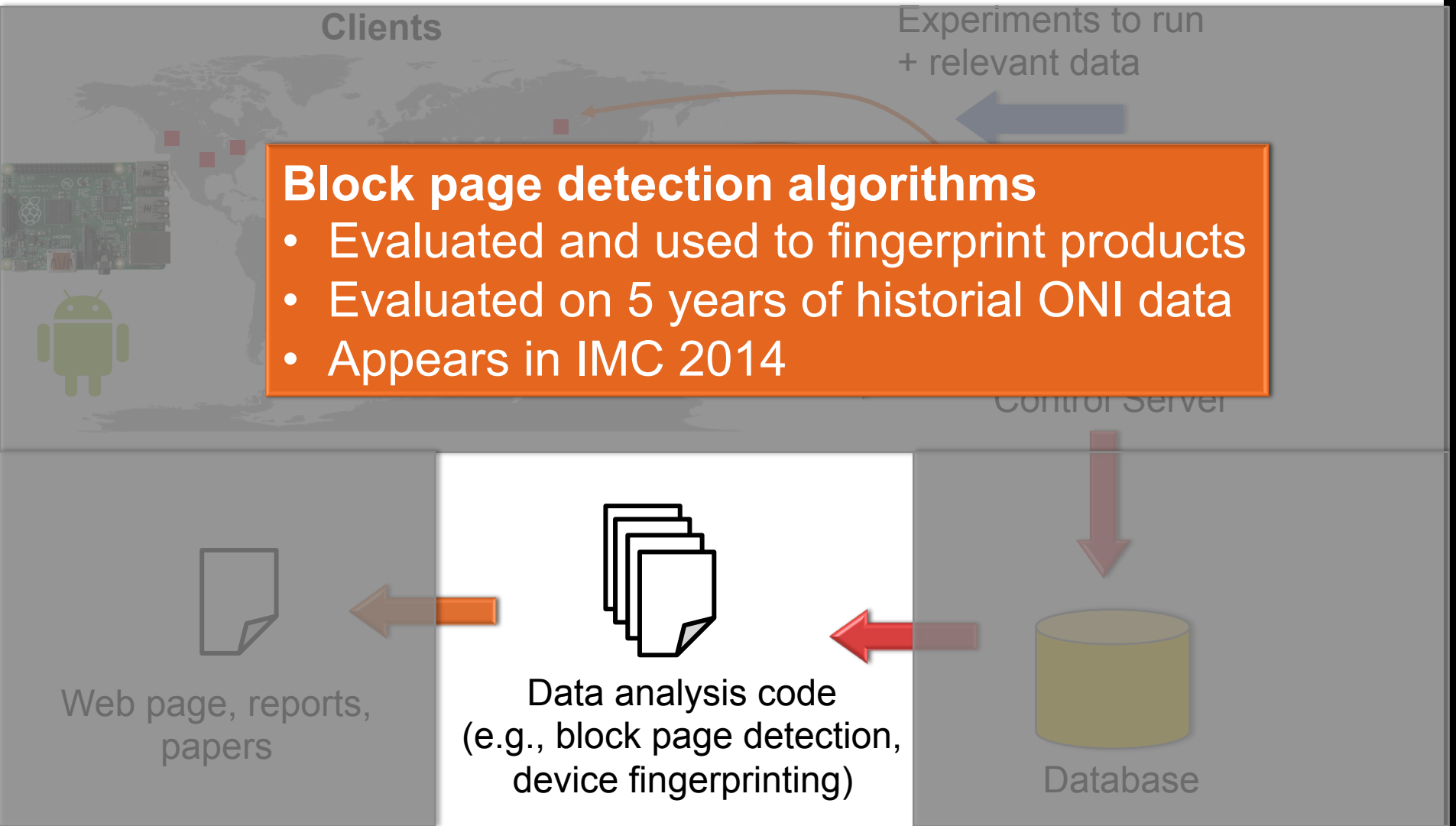
OVERVIEW OF ICLAB

Client + Server in limited beta
Volunteers beginning to deploy nodes
O(100s) of VPN endpoints online



Database

OVERVIEW OF ICLAB



Block page detection algorithms

- Evaluated and used to fingerprint products
- Evaluated on 5 years of historial ONI data
- Appears in IMC 2014

LEVERAGING ARK + ATLAS FOR ICLAB

- **Traceroute, and Ping are common to all platforms**
  - (RIPE Atlas also has SSL in common with ICLab)
- **Challenge to leveraging Ark + Atlas:**
  - Censorship-related destinations can put people at risk!
- **ICLab has censorship-specific consent process**
  - + per country risk information
  - + ongoing monitoring of global politics/situations

Thanks Citizen Lab!

- **This research may not be considered human subjects!**
  - Stony Brook IRB found it not to be human subjects
  - UToronto IRB is working with us

BABY STEPS… ?

- **Not all censored destinations are high risk**
  - E.g., Alexa top 100
- **Monitor their reachability from Ark nodes**
  - Investigate specific outages with more indepth measurements on ICLab nodes

Big Picture Questions

- **What to include in measurement platform consent forms?**
  - Especially when platform can support wide range of tests!
- **How to manage risk to vantage points and measurement targets?**

www.iclab.org

phillipa@cs.stonybrook.edu

Thanks! Work presented is funded by: NSF Grant CNS 1350720 and a Google Faculty Research Award