

# Global Measurement of DNS Manipulation

Paul Pearce

Ben Jones, Frank Li, Roya Ensafi,  
Nick Feamster, Nick Weaver, and Vern Paxson

**Berkeley**  
UNIVERSITY OF CALIFORNIA

**PRINCETON**  
UNIVERSITY

**INTERNATIONAL**  
COMPUTER SCIENCE  
INSTITUTE

# Censorship


China's scary lesson to the world: Censoring the Internet works

Asia & Pacific

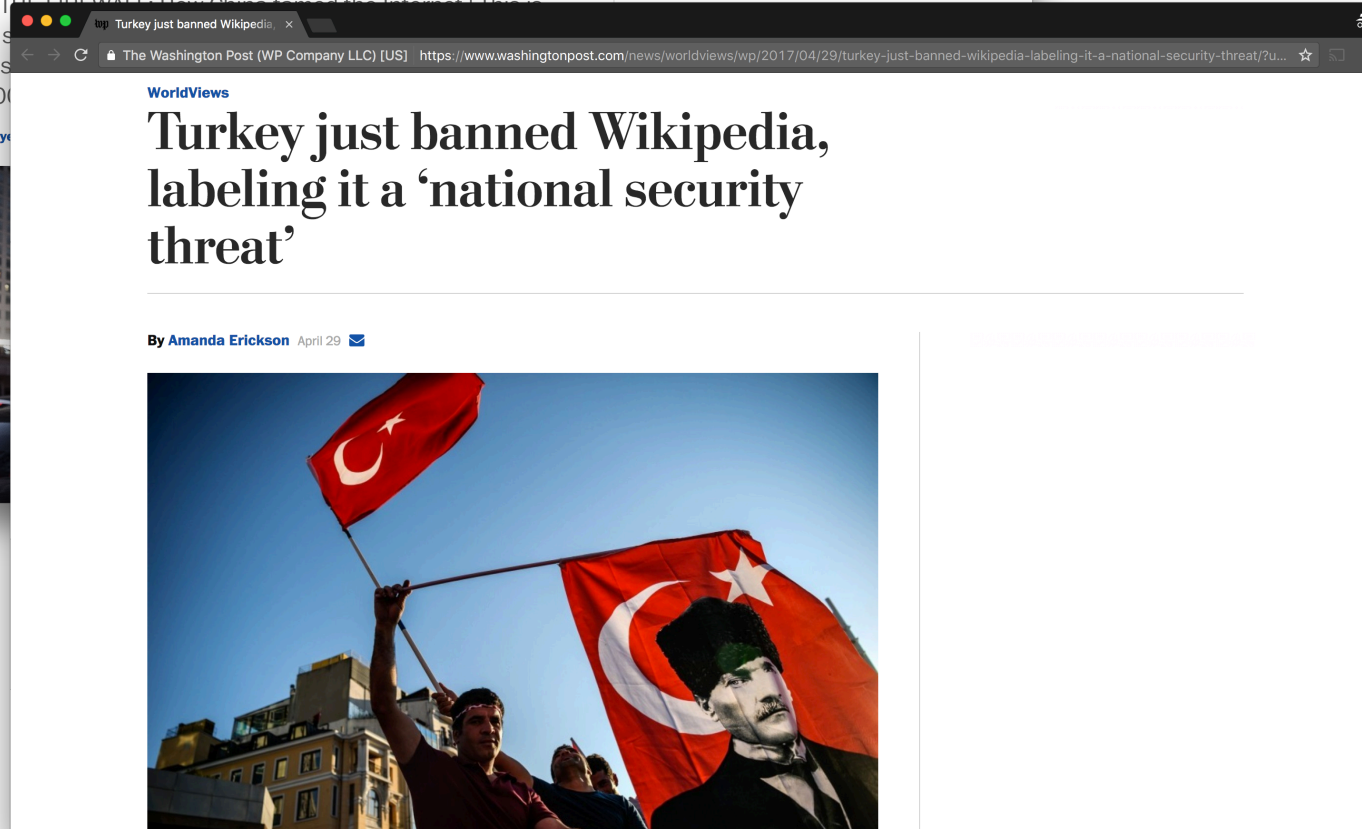
## China's scary lesson to the world: Censoring the Internet works

BEHIND THE FIREWALL: How China tamed the Internet | This is part of a series examining the impact of China's Great Firewall, a mechanism of Internet censorship and surveillance that affects nearly 700 million users.

By [Simon Denyer](#) May 23, 2016



# Censorship




# Censorship

China's scary lesson to the world:  
**Censoring the Internet works**

Asia & Pacific

BEHIND THE FIREWALL: How China tamed the Internet. This is part of a series of reports that look at the mechanisms behind nearly 70 countries' internet restrictions.

By Simon Denyer



Turkey just banned Wikipedia, labeling it a 'national threat'

WorldViews

By Amanda Erickson April 29



Freedom House

https://freedomhouse.org/report/freedom-net/freedom-net-2016

REPORT NAVIGATION

- Silencing the Messenger: Communication Apps Under Pressure
- 2016 Country Scores
- Key Internet Controls by Country
- Methodology
- About Freedom on the Net

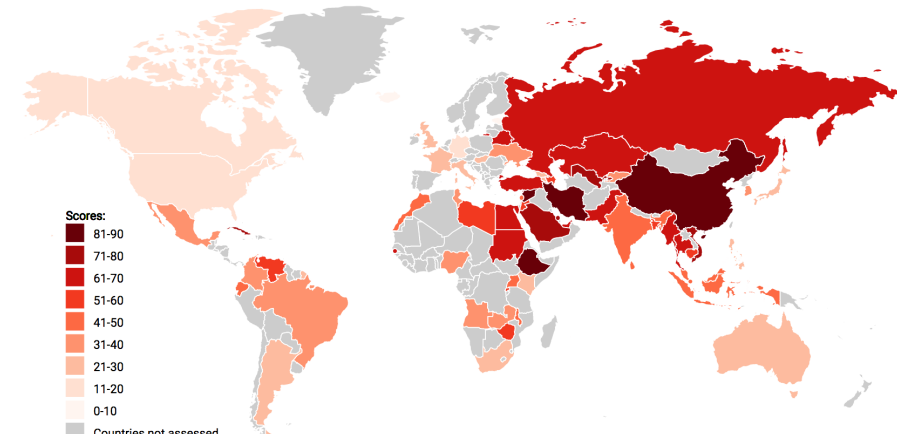
COUNTRY REPORTS

- Select a Country Report -

DOWNLOADS

- Report PDF
- Report + 65 Country Reports
- Report Graphics

Freedom on the Net 2016 Overall Scores



Freedom on the Net Score: 0=Most Free, 100=Less Free

Internet freedom has declined for the sixth consecutive year, with more governments than ever before targeting social media and communication apps as a means of halting the rapid dissemination of information, particularly during anti-government protests.

The increased controls show the importance of social media and online communication for advancing political freedom and social justice. It is no coincidence that the tools at the center of the current crackdown have been widely used to hold governments accountable and facilitate uncensored conversations. Authorities in several countries have even resorted to shutting down all internet access at politically contentious times, solely to prevent users from disseminating information through social

# Understanding Censorship

- Despite prevalence, existing empirical measurement is sparse across:
  - Time
  - Space
  - Content
- Why? Deployed state of the art: Volunteers
- We argue: Continuous, diverse measurement needed to understand the scope, scale, and evolution of Internet censorship

# Our Work

- Censorship techniques vary
  - This work → DNS manipulation
- Measurement Goals:
  - Diverse
  - Longitudinal
  - Does not require participation
  - Ethical
- Design, implement, and deploy **Iris**, a system to identify DNS manipulation globally
- Global measurement study
  - Identifies pervasiveness of manipulation worldwide
  - Heterogeneity across content, countries, and resolvers
  - Heterogeneity within countries

# Approach

- Conceptually simple:
  - Issue DNS queries for sensitive across globally diverse vantage points
  - Look for “wrong” responses
- Challenge 1: Vantage points
  - → Open DNS resolvers
- Challenge 2: Ethics
  - → Identify “Infrastructure” DNS resolvers
- Challenge 3: Repeatable
  - → Design of Iris
- Challenge 4: Identifying “wrong” responses?
  - → Consistency and independent verifiability of structural elements

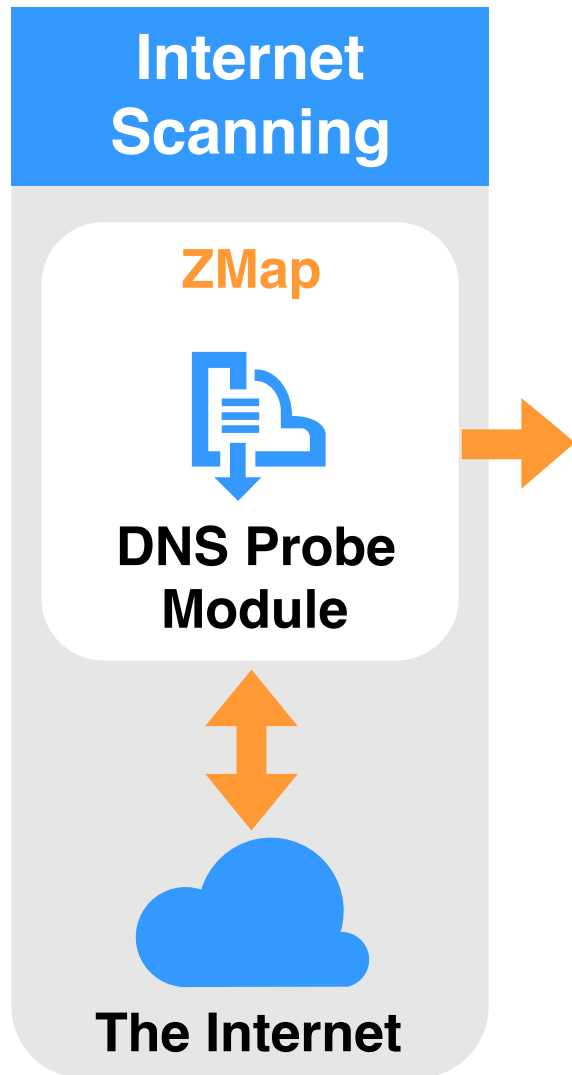
# Ethics

- Guided by ideals laid out by the Menlo Report:
  - Respect for persons
  - Beneficence
  - Respect for law and public
- Only use resolvers *reasonably* attributed to Internet naming infrastructure
- Heavily rate limit queries to resolvers and domains

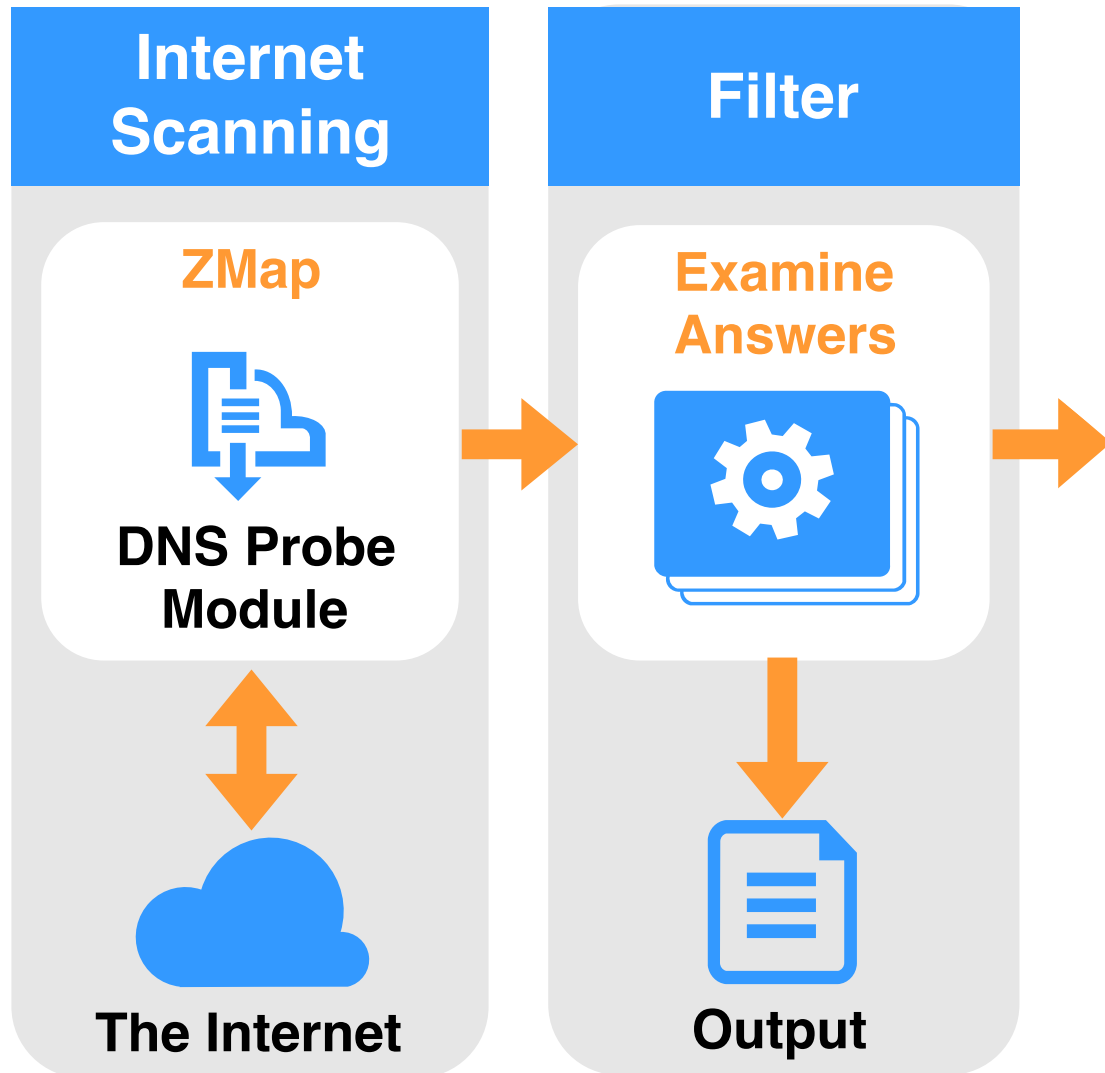


# Finding Infrastructure Resolvers

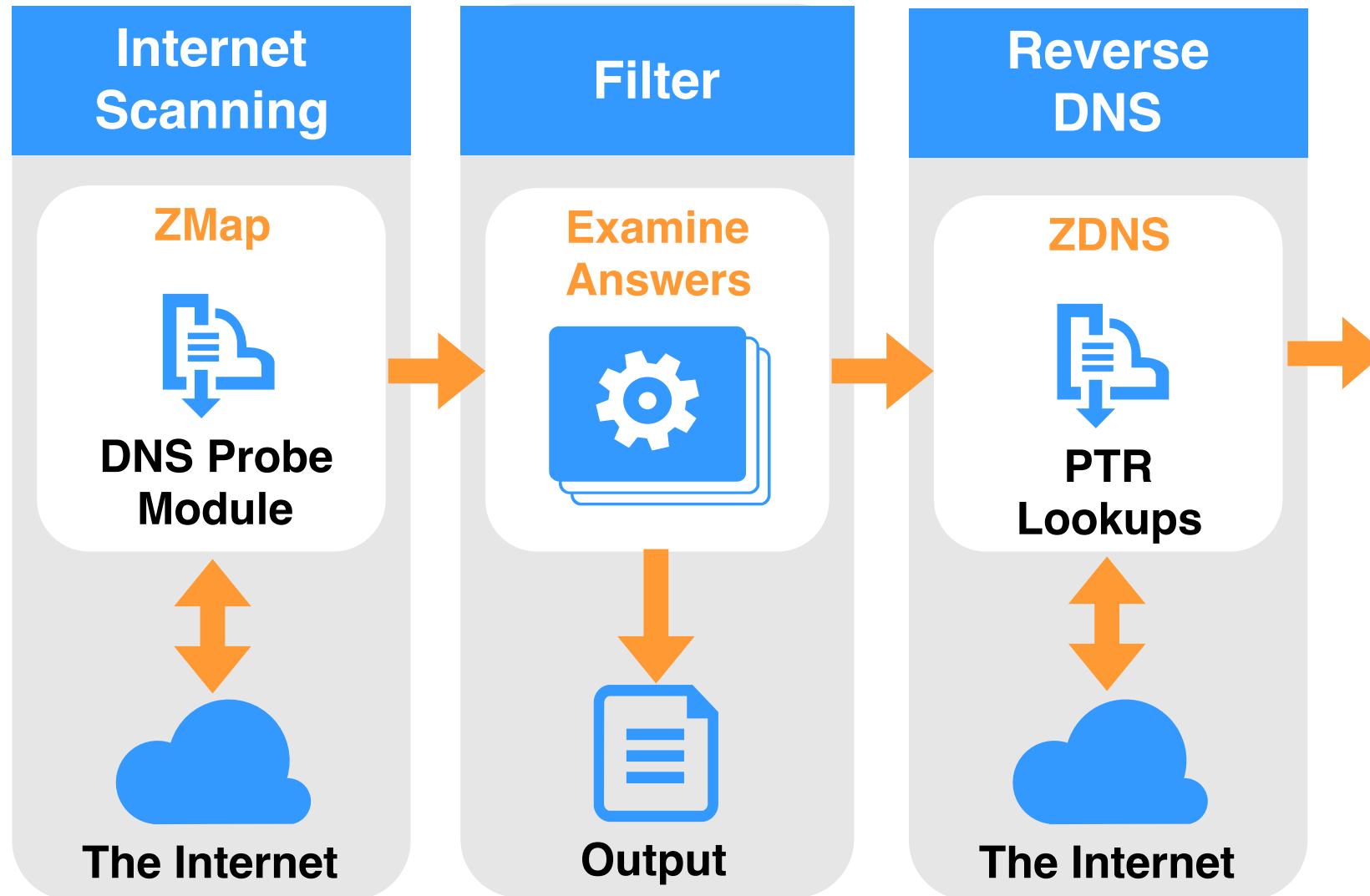
# Finding Infrastructure Resolvers



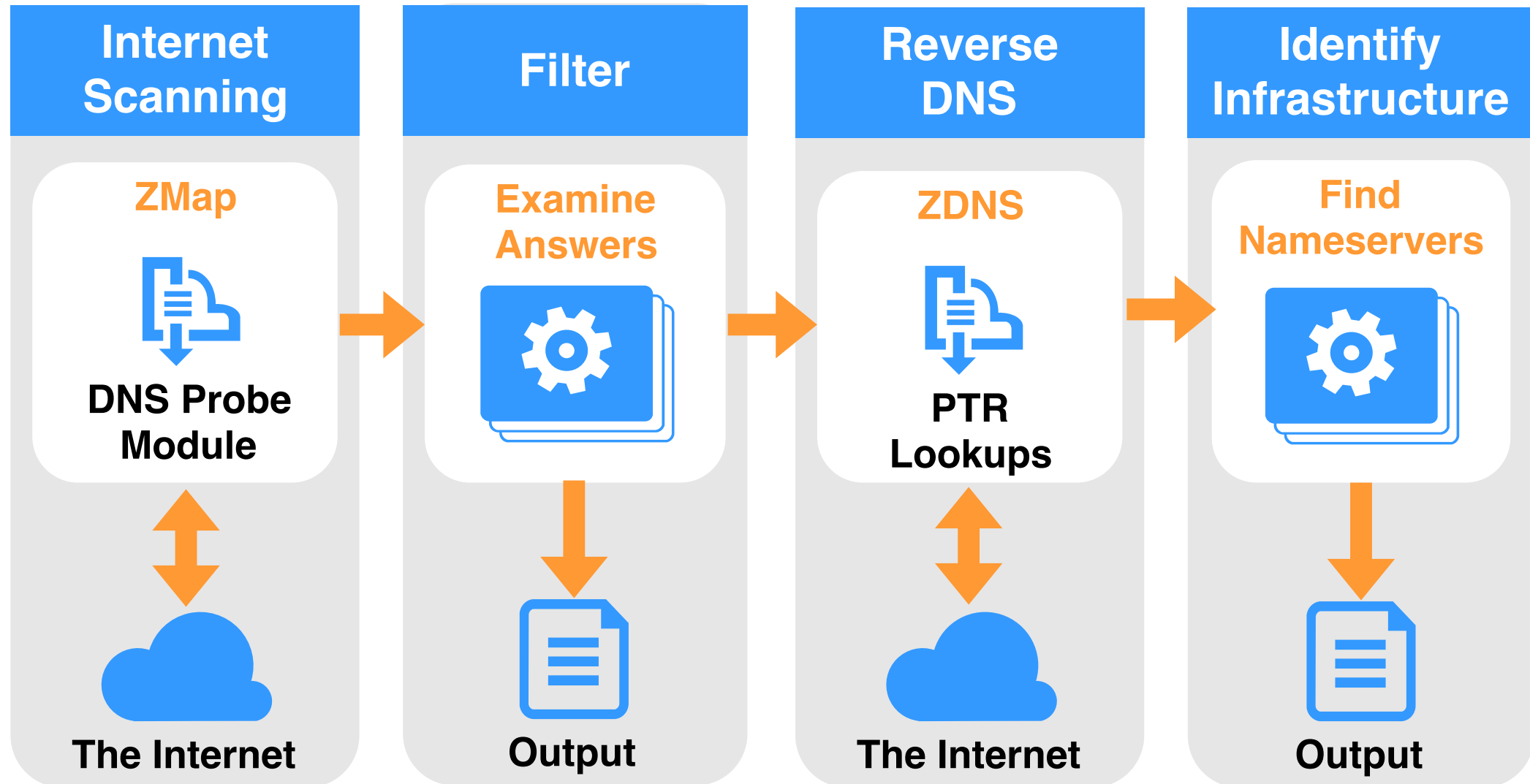
# Finding Infrastructure Resolvers



# Finding Infrastructure Resolvers



# Finding Infrastructure Resolvers



# What to Measure

- Sensitive
  - All domains from the Citizen Lab sensitive test list
- Popular
  - Random subset of Alexa top 10,000
- Feed these artifacts into the Iris pipeline
  - Output → DNS manipulation

# Iris Pipeline

# Iris Pipeline

## Measurement Artifacts

Input



Infrastructure Resolvers

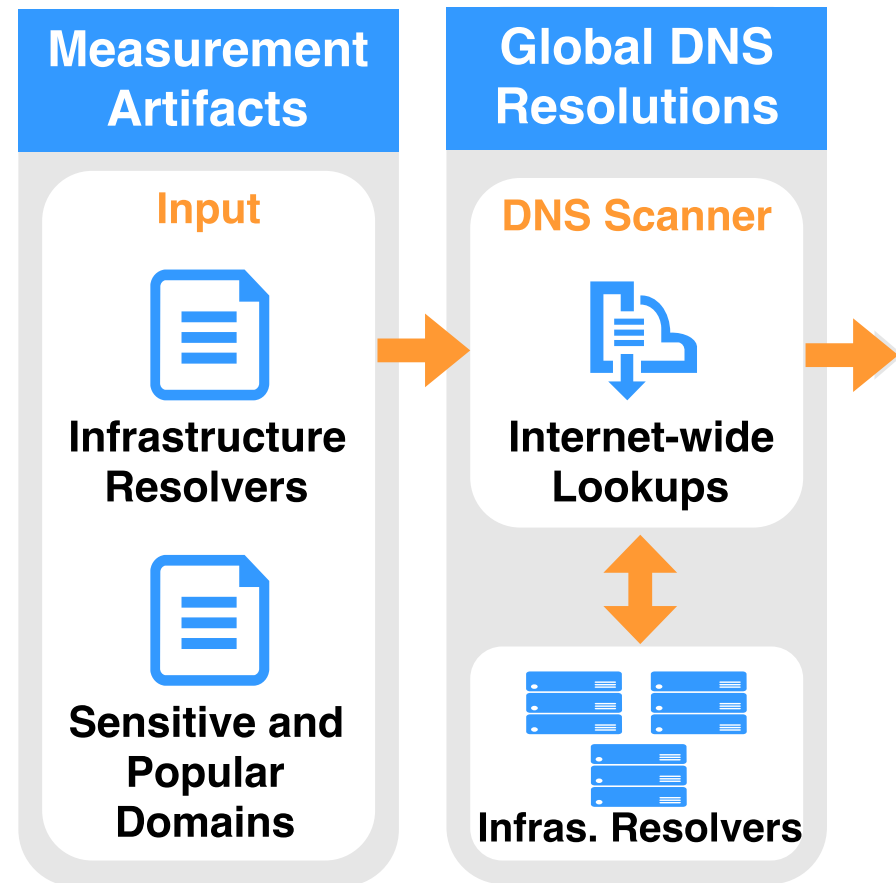


Sensitive and Popular Domains

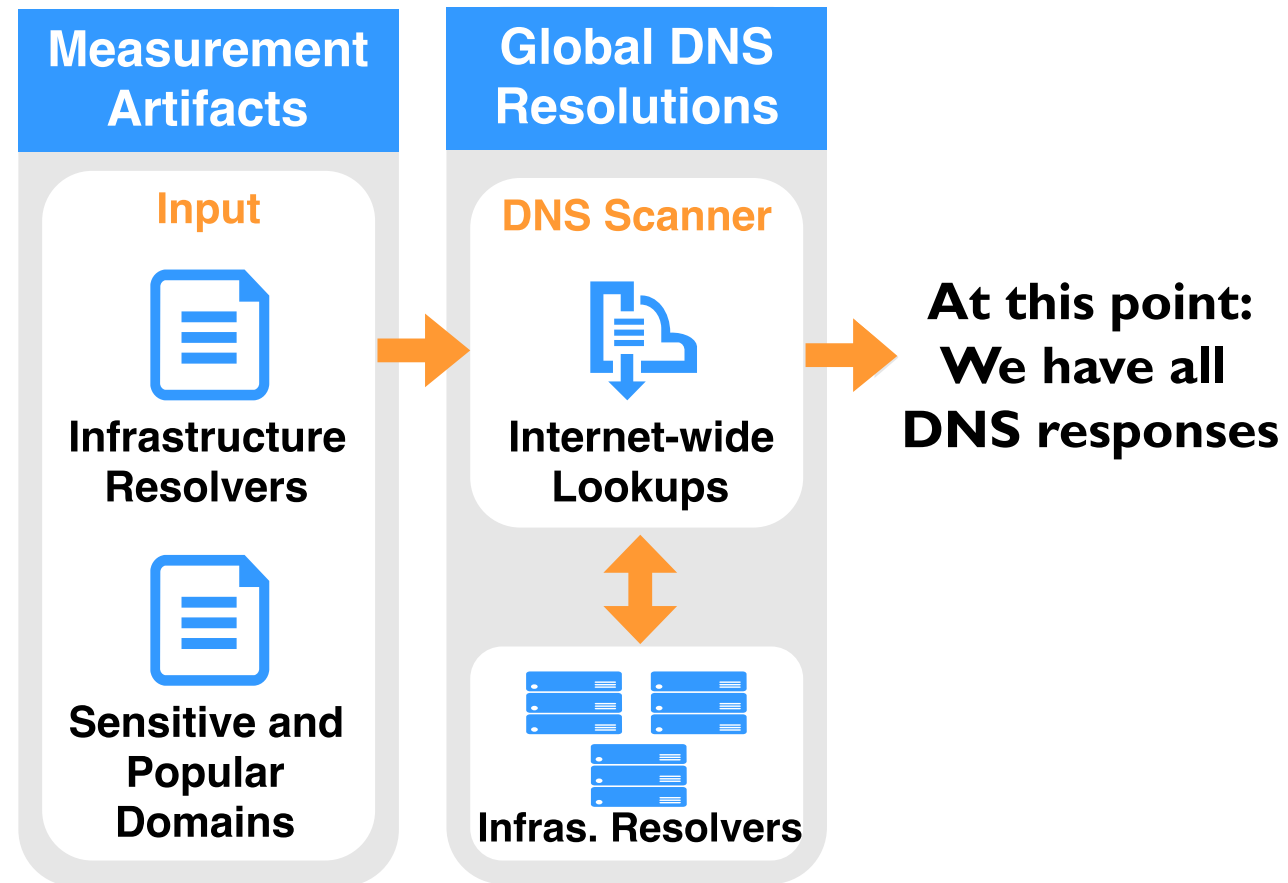




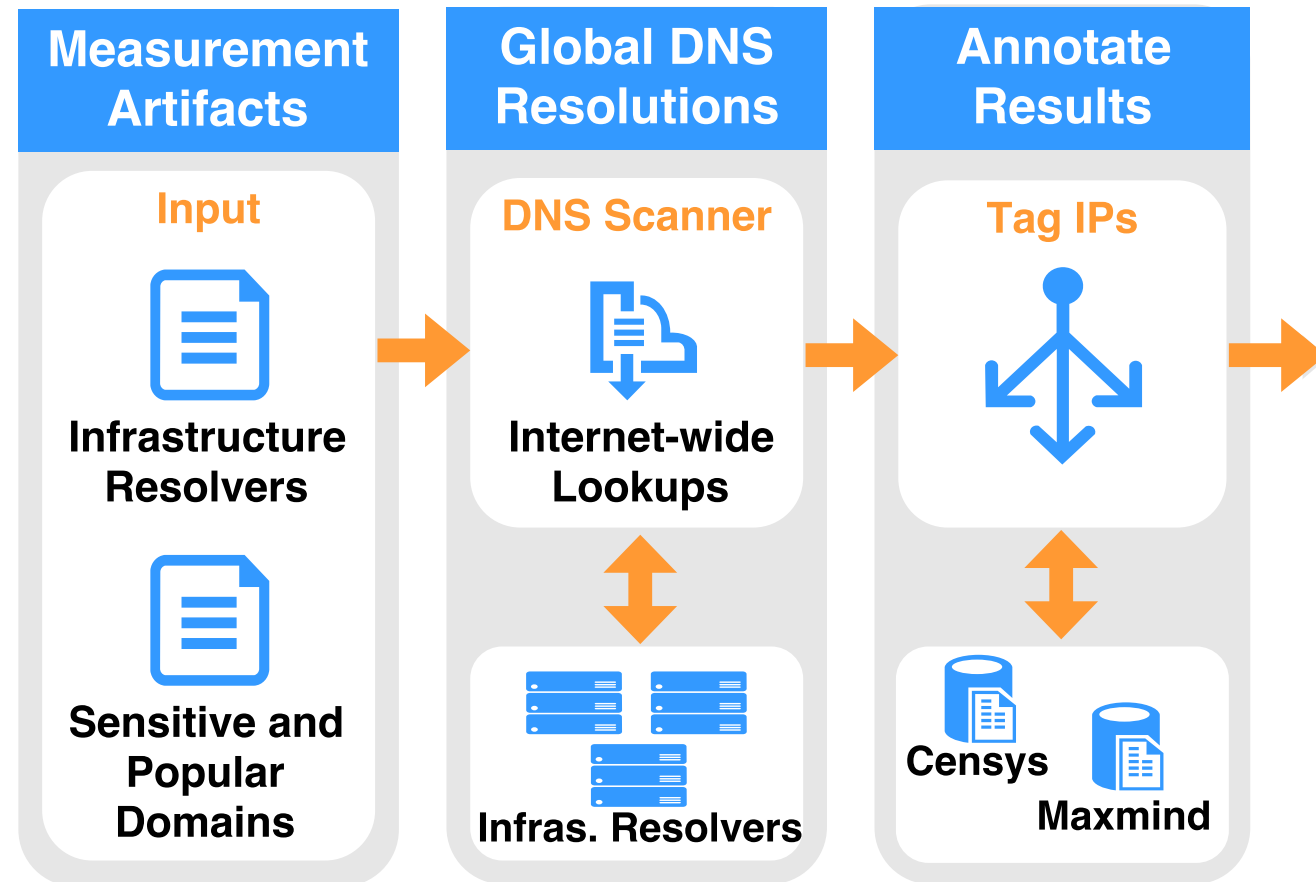
# Iris Pipeline



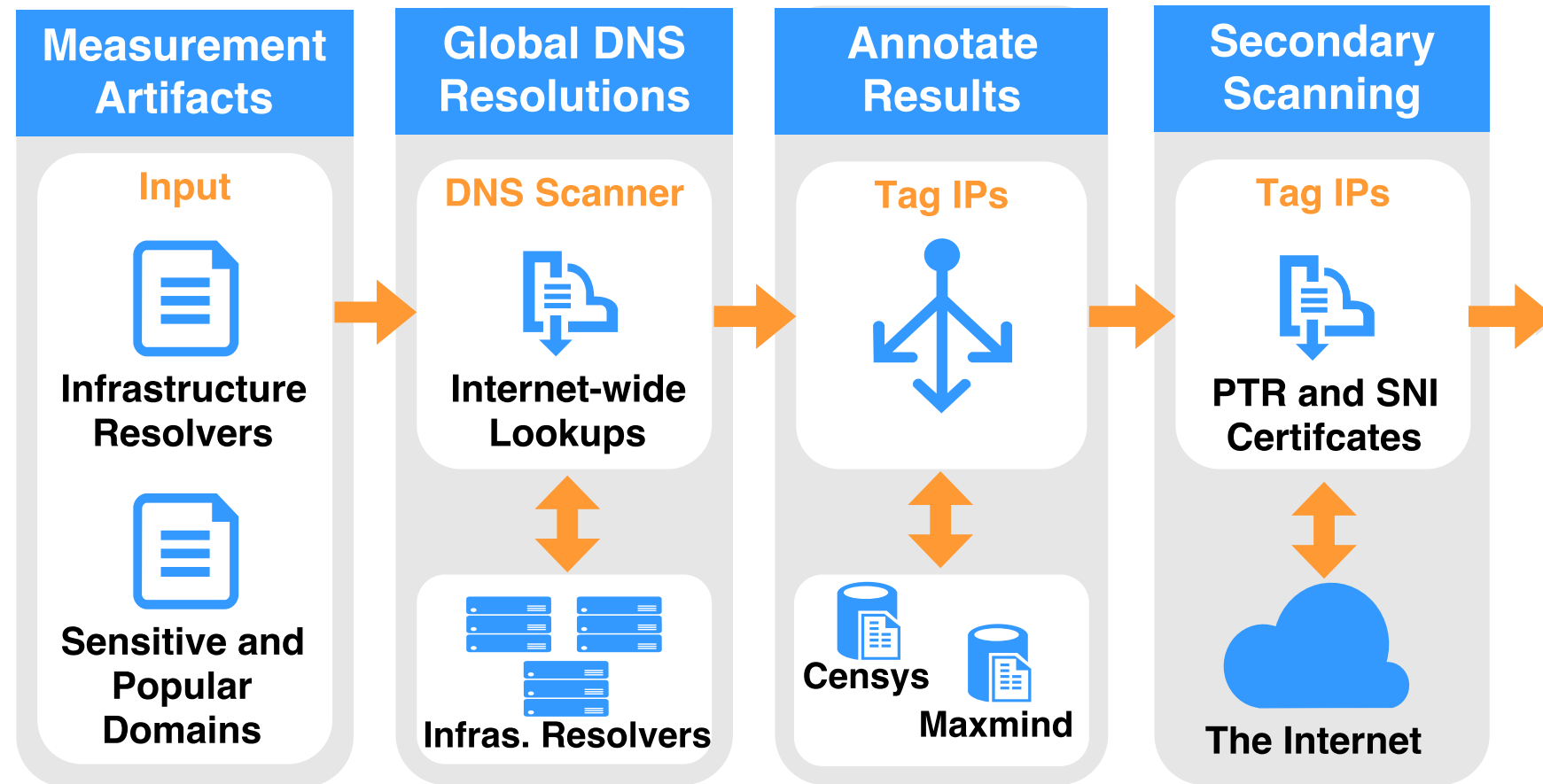
# Iris Pipeline



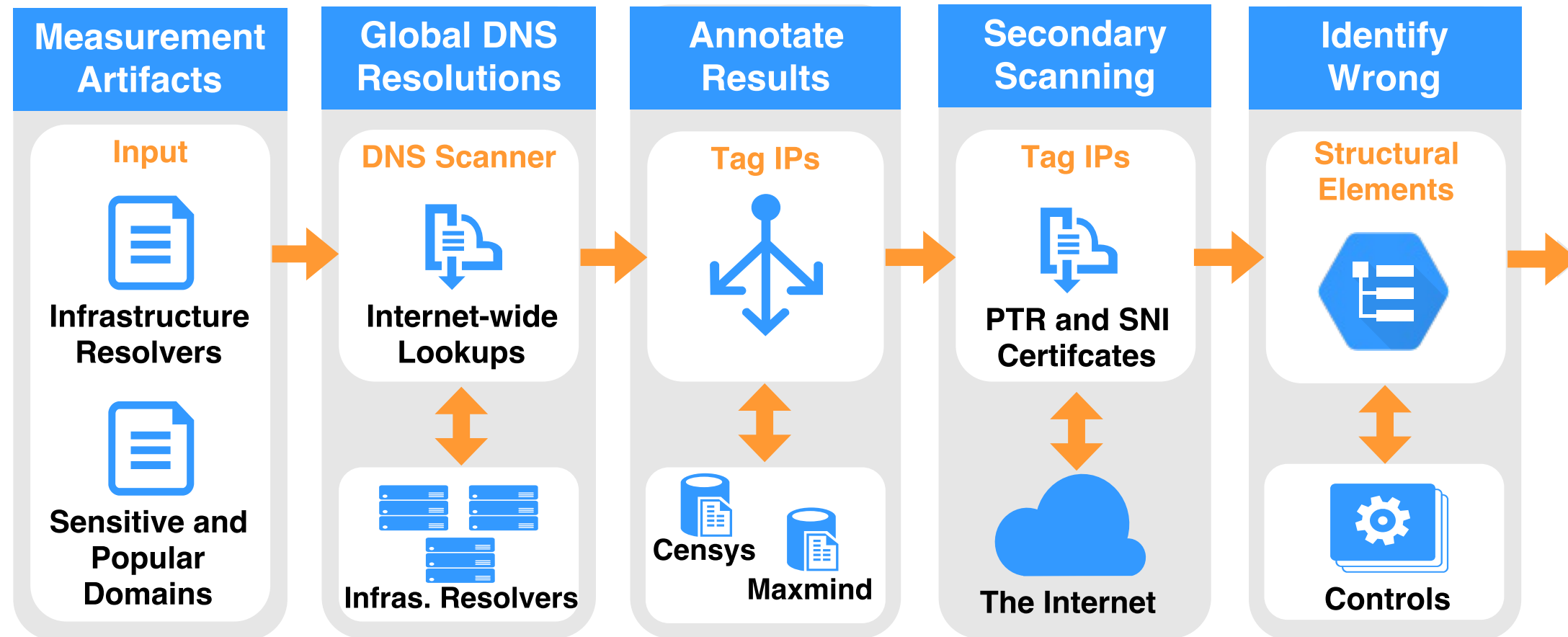
# Iris Pipeline



# Iris Pipeline

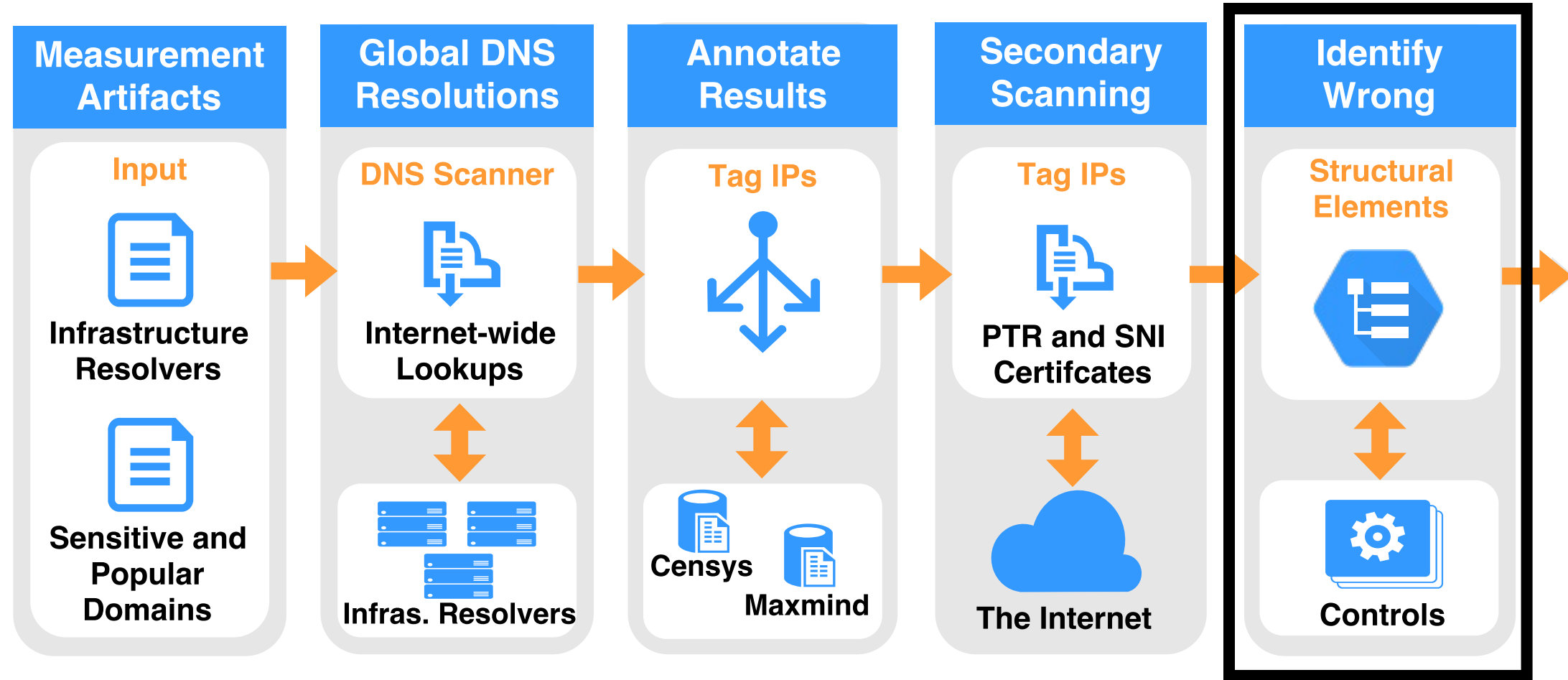


# Iris Pipeline



# Iris Pipeline

## Challenge 4: Identifying wrong responses

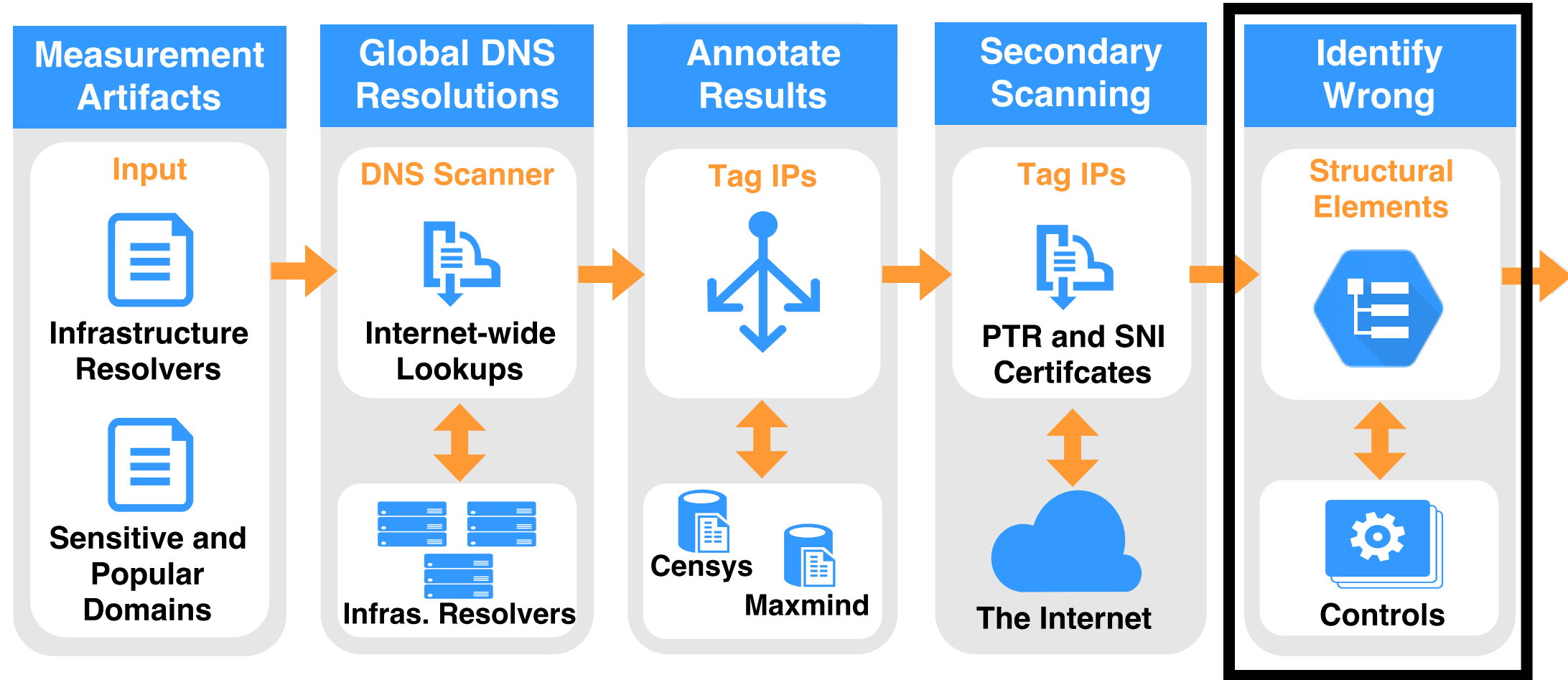


# Identifying Wrong Responses

- Intuition: Shared structural elements, even in global deployments
- Approach: structure consistency and independent verifiability
- Consistency
  - 4, uncensored, geographically diverse **controls**
  - Compare each answer with the control set
- Independent Verifiability
  - *Valid* HTTPS certificate, with and without SNI
- **If any metric consistent/verifiable, response is correct**

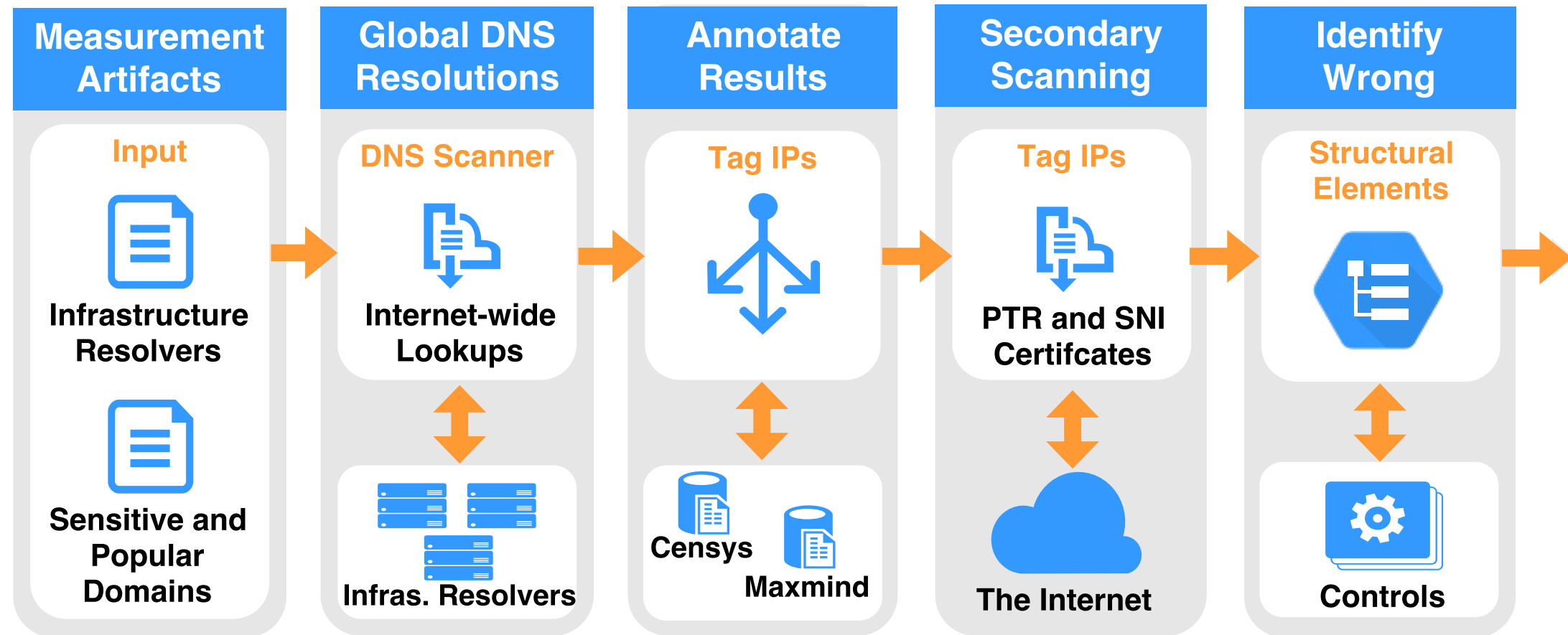
# Iris Pipeline

## Challenge 4: Identifying wrong responses

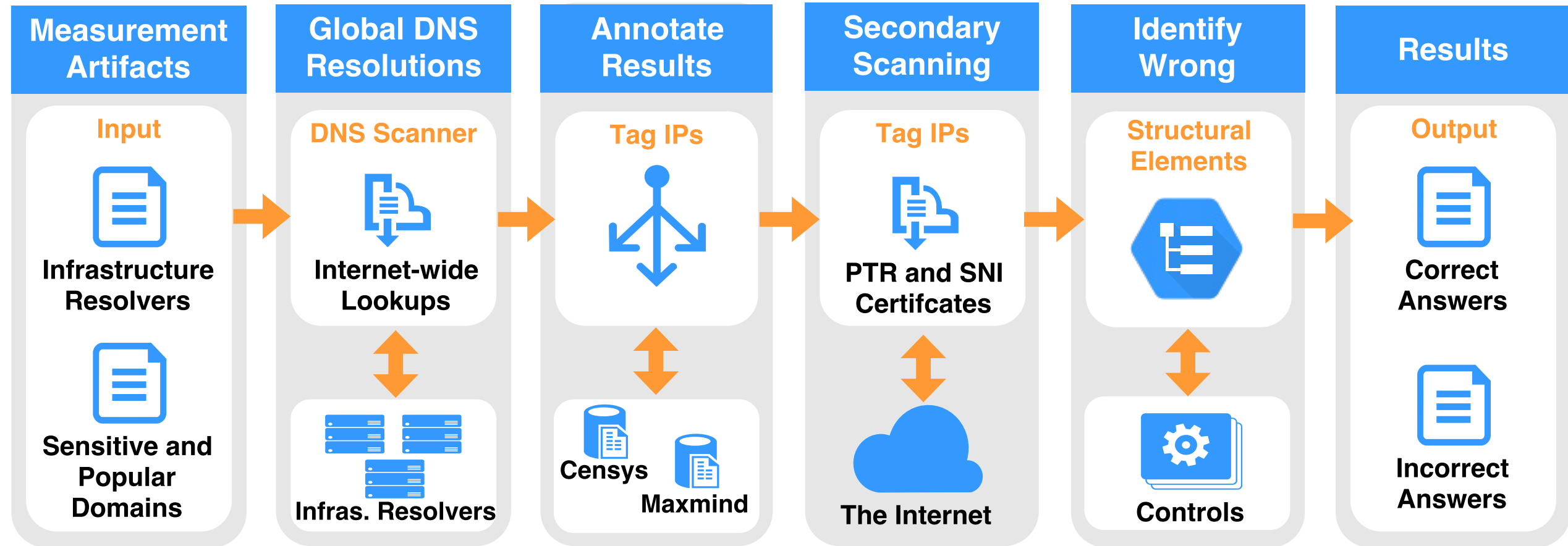




# Iris Pipeline



# Iris Pipeline



# Measurement Study and Dataset

- What is the open resolver population?

Resolver Dataset	Number Resolvers	Number Countries	Median / Country
All Open	4.2M	232	660

- How much does our ethical framework reduce coverage?

- What is the total set of DNS responses we examine?

Number Resolvers	Number Domains	Total Responses
6,564	2,330	14.5M

- **What does our dataset reveal?**

# Manipulation By Country

- What **countries** experience the most manipulation?

# Manipulation By Country

- What **countries** experience the most manipulation?
  - Qualitatively consistent with prior work

Country	Median Manipulated	Number Resolvers
Iran	6.02%	122
China	5.22%	62
Indonesia	0.63%	80
Greece	0.28%	62
Mongolia	0.17%	6
Iraq	0.09%	7
Bermuda	0.04%	2
Kazakhstan	0.04%	14
Belarus	0.04%	18

# Manipulation By Country

- What **countries** experience the most manipulation?
  - Qualitatively consistent with prior work
- Are there outliers within countries?

Country	Median Manipulated	Number Resolvers
Iran	6.02%	122
China	5.22%	62
Indonesia	0.63%	80
Greece	0.28%	62
Mongolia	0.17%	6
Iraq	0.09%	7
Bermuda	0.04%	2
Kazakhstan	0.04%	14
Belarus	0.04%	18

# Manipulation By Country

- What **countries** experience the most manipulation?
  - Qualitatively consistent with prior work
- Are there outliers within countries?
  - High outliers
    - → localized manipulation

Country	Median Manipulated	Number Resolvers	Max Manipulated
Iran	6.02%	122	22.41%
China	5.22%	62	8.40%
Indonesia	0.63%	80	9.95%
Greece	0.28%	62	0.83%
Mongolia	0.17%	6	0.36%
Iraq	0.09%	7	5.79%
Bermuda	0.04%	2	0.09%
Kazakhstan	0.04%	14	3.90%
Belarus	0.04%	18	0.30%

# Manipulation By Country

- What **countries** experience the most manipulation?
  - Qualitatively consistent with prior work
- Are there outliers within countries?
  - High outliers
    - → localized manipulation
  - Low outliers
    - → geolocation error

Country	Median Manipulated	Number Resolvers	Max Manipulated	Min Manipulated
Iran	6.02%	122	22.41%	0%
China	5.22%	62	8.40%	0%
Indonesia	0.63%	80	9.95%	0%
Greece	0.28%	62	0.83%	0%
Mongolia	0.17%	6	0.36%	0%
Iraq	0.09%	7	5.79%	0%
Bermuda	0.04%	2	0.09%	0%
Kazakhstan	0.04%	14	3.90%	0%
Belarus	0.04%	18	0.30%	0%



# Manipulation By Domain

- What **domains** are most frequently manipulated?

# Manipulation By Domain

- What **domains** are most frequently manipulated?
  - Gambling and Pornography
    - → 8 of top 10

Rank	Domain	Category	Countries
1	www.pokerstars.com	Gambling	19
2	betway.com	Gambling	19
3	pornhub.com	Pornography	19
4	youporn.com	Pornography	19
5	xvideos.com	Pornography	19
6	thepiratebay.org	P2P File Sharing	18
7	thepiratebay.se	P2P File Sharing	18
8	xhamster.com	Pornography	18
9	www.partypoker.com	Gambling	17
10	beeg.com	Pornography	17

# Manipulation By Domain

- What **domains** are most frequently manipulated?
  - Gambling and Pornography
    - → 8 of top 10
- Are commonly measured sites the most frequent targets?  
(Anonymity tools, Twitter, Google)
  - No. They experience significantly less manipulation globally
  - → diversity in measured domains

Rank	Domain	Category	Countries
1	www.pokerstars.com	Gambling	19
2	betway.com	Gambling	19
3	pornhub.com	Pornography	19
4	youporn.com	Pornography	19
5	xvideos.com	Pornography	19
6	thepiratebay.org	P2P File Sharing	18
7	thepiratebay.se	P2P File Sharing	18
8	xhamster.com	Pornography	18
9	www.partypoker.com	Gambling	17
10	beeg.com	Pornography	17
80	torproject.org	Anonymity & Censorship	12
181	twitter.com	Twitter	9
250	www.youtube.com	Google Video	8
495	www.citizenlab.org	Freedom of Expression	4
606	www.google.com	Google	3

# Manipulation By Category

- What **categories** are most frequently manipulated?

# Manipulation By Category

- What **categories** are most frequently manipulated?
  - Sites from the Alexa sampling experience widespread manipulation
  - While Gambling and Pornography individual domains were most common, they are not the most common categories

Rank	Category	Countries
1	Alexa Top 10,000	36
2	Freedom of Expression	35
3	P2P File Sharing	34
4	Human Rights	31
5	Gambling	29
6	Pornography	29
7	Alcohol and Drugs	28
8	Anonymity & Censorship	24
9	Hate Speech	22
10	Multimedia Sharing	21

# Manipulation By Category

- What **categories** are most frequently manipulated?
  - Sites from the Alexa sampling experience widespread manipulation
  - While Gambling and Pornography individual domains were most common, they are not the most common categories
- Are commonly measured sites the most frequent targets?
  - No

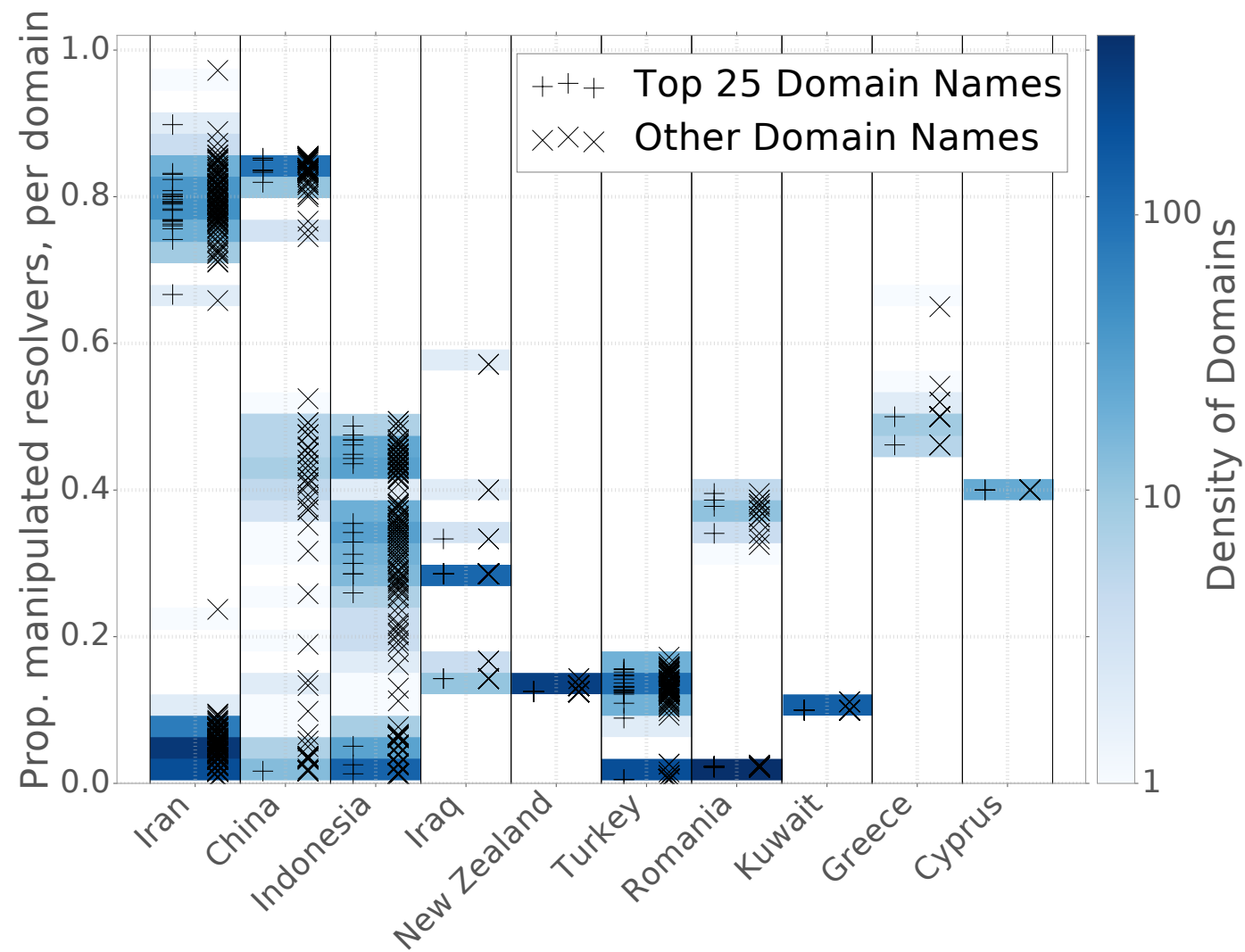
Rank	Category	Countries
1	Alexa Top 10,000	36
2	Freedom of Expression	35
3	P2P File Sharing	34
4	Human Rights	31
5	Gambling	29
6	Pornography	29
7	Alcohol and Drugs	28
8	Anonymity & Censorship	24
9	Hate Speech	22
10	Multimedia Sharing	21
20	Google (All)	16
34	Facebook (All)	10
38	Twitter (All)	9

# Consistency Within Countries

- Is there heterogeneity within countries?

# Consistency Within Countries

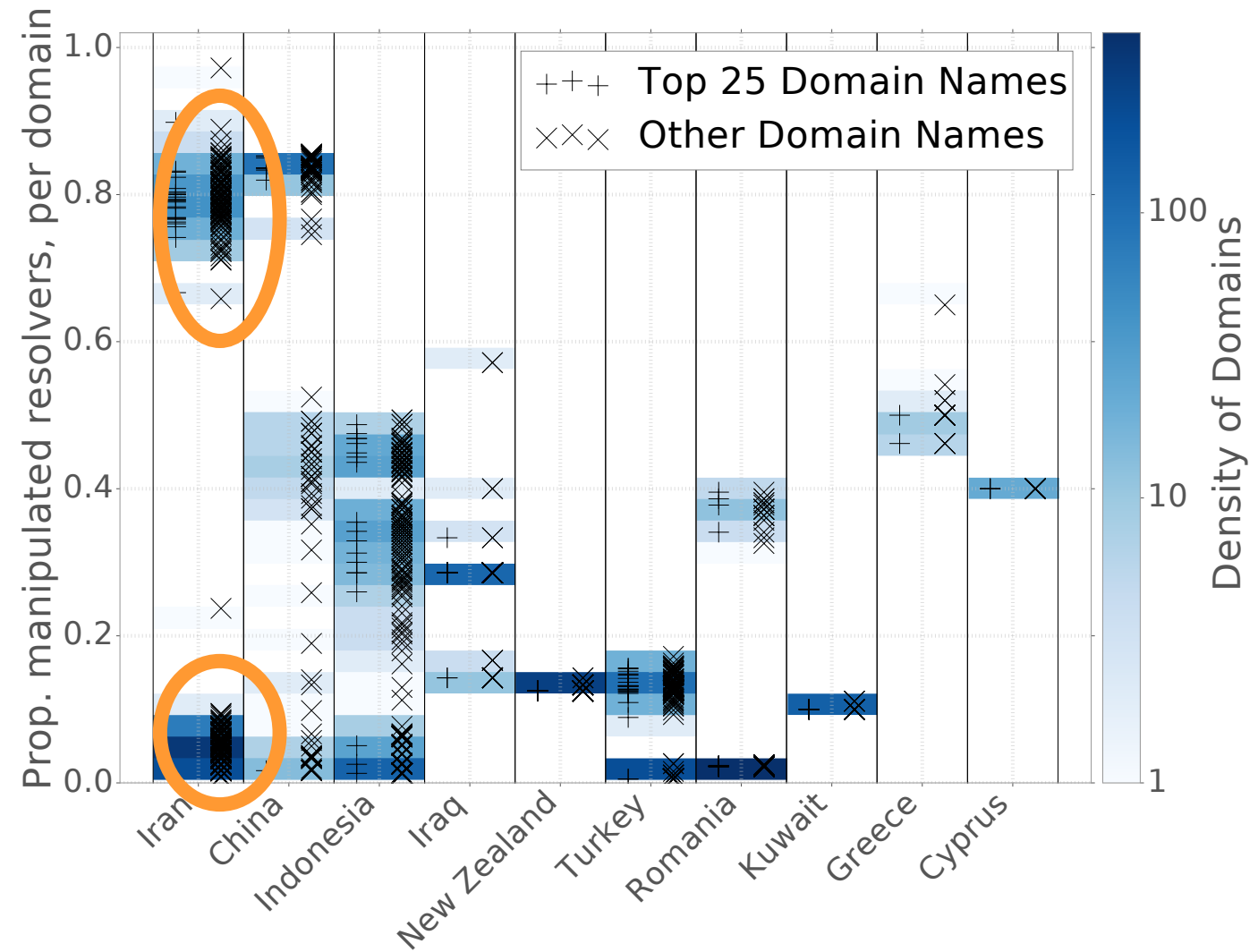
- Is there heterogeneity within countries?





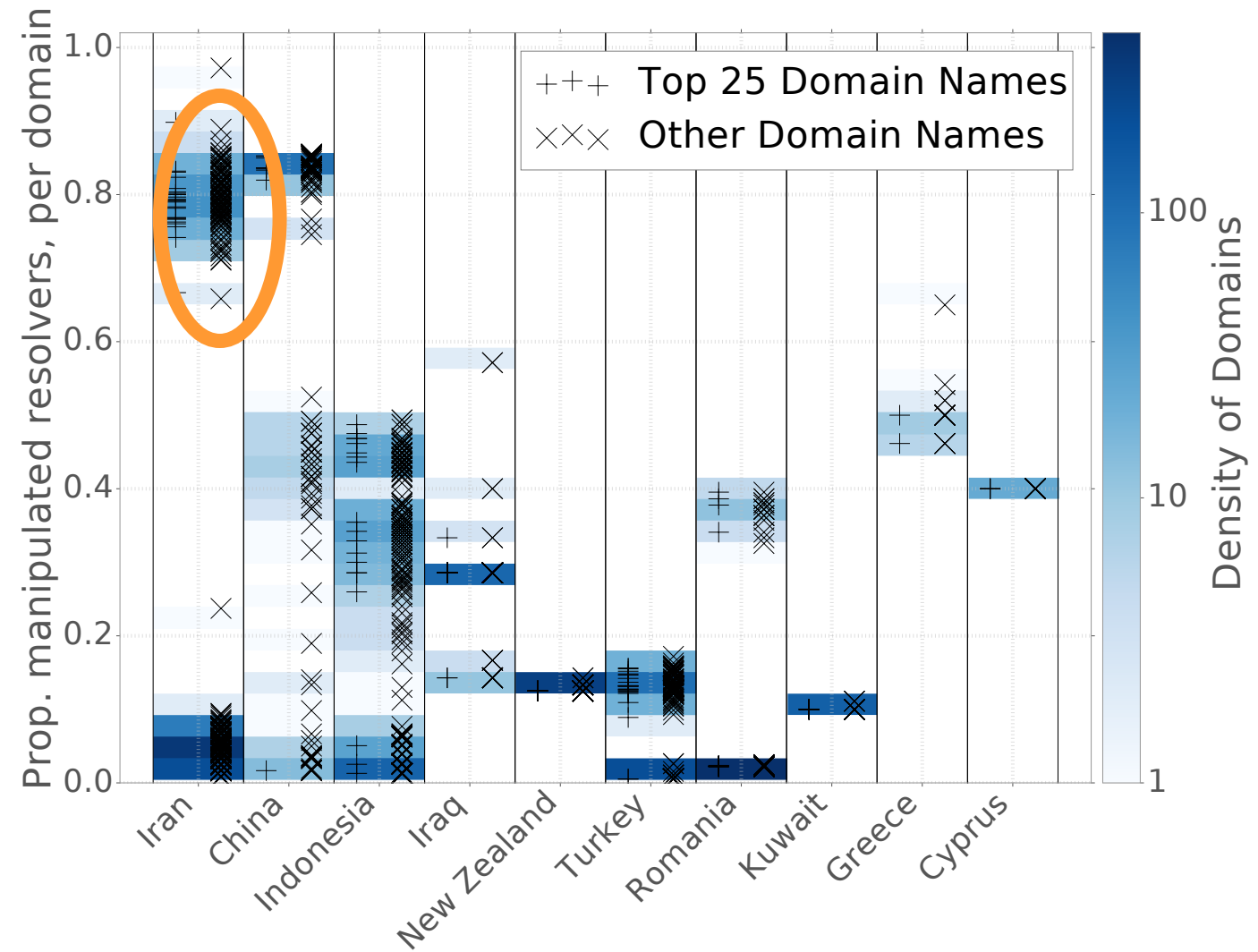
# Consistency Within Countries

- Is there heterogeneity within countries?
  - Yes: Modal effects → multiple systems, localized manipulation



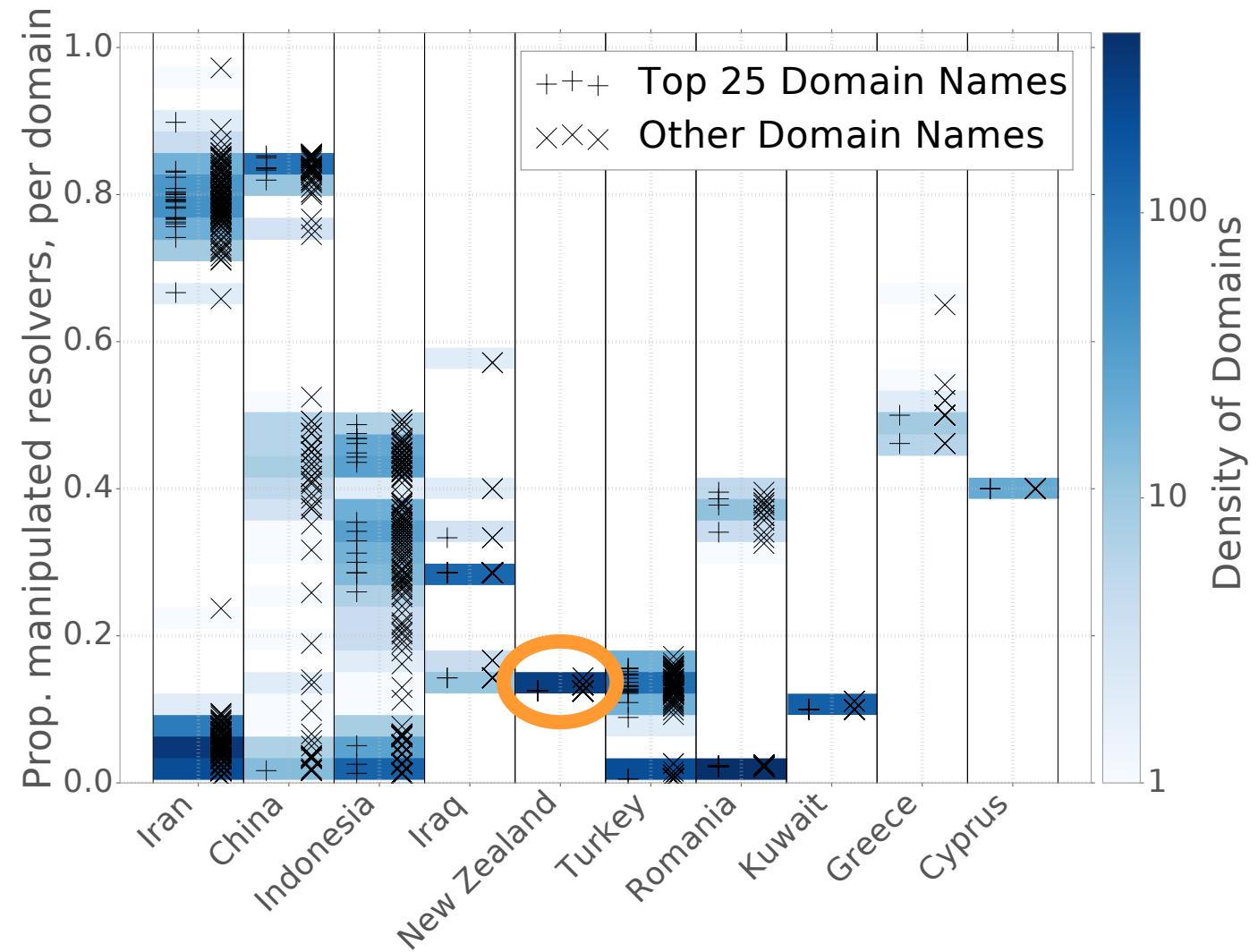
# Consistency Within Countries

- Is there heterogeneity within countries?
  - Yes: Modal effects → multiple systems, localized manipulation
- Is there non-determinism?
  - Yes: Smearing effects



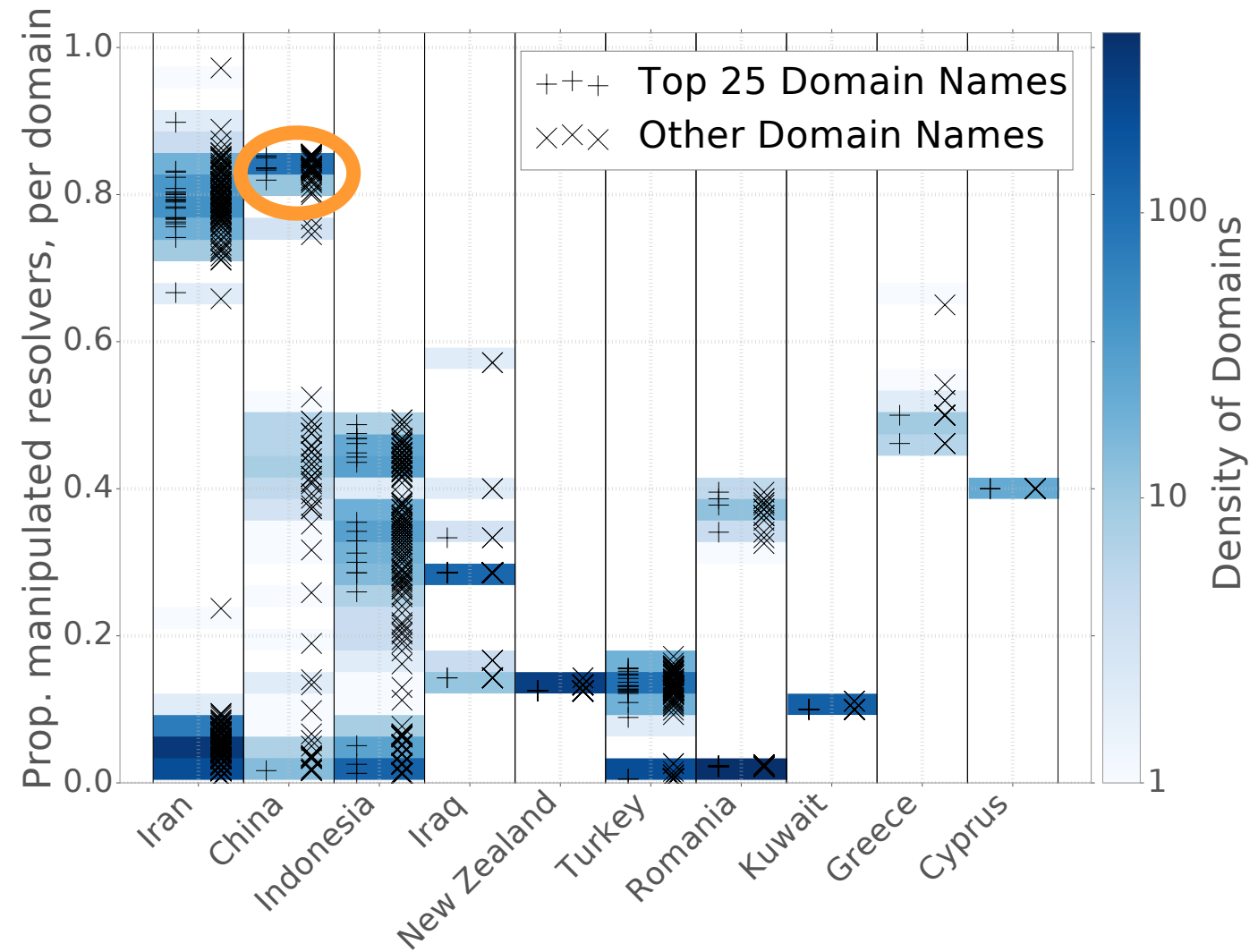
# Consistency Within Countries

- Is there heterogeneity within countries?
  - Yes: Modal effects → multiple systems, localized manipulation
- Is there non-determinism?
  - Yes: Smearing effects
- Is there ISP-level filtering?
  - Yes: Low-but-incomplete countries



# Consistency Within Countries

- Is there heterogeneity within countries?
  - Yes: Modal effects → multiple systems, localized manipulation
- Is there non-determinism?
  - Yes: Smearing effects
- Is there ISP-level filtering?
  - Yes: Low-but-incomplete countries
- Is there geolocation error?
  - Yes: High-but-incomplete countries



# Methodological Takeaways

- Domain selection is critical for comparative studies
  - List biases will influence ranking, comparisons
- Measurement of non-sensitive content is important
  - Lists are inherently limited
- How data is grouped influences results
  - Domain vs Category
- In-country diversity is necessary to accurately depict manipulation
  - Outliers both high and low
  - Heterogeneous manipulation

# Conclusion and Next Steps

- Internet Censorship is prevalent and heterogeneous
- Iris is an ethical system to identify DNS manipulation Internet-wide
- We identified heterogeneity of censorship across multiple dimensions, including variance within countries, highlighting the need for tools such as Iris
- Next Steps: Use of Iris and other Internet-wide techniques for continuous longitudinal measurement

# Thank You

**@paul\_pearce**  
**pearce@berkeley.edu**  
**<https://cs.berkeley.edu/~pearce/>**