



Internet Society
Pulse

Measuring Content Locality

James I. Madeley

Pulse Research Review

11th Dec 2024



About Me



About Me

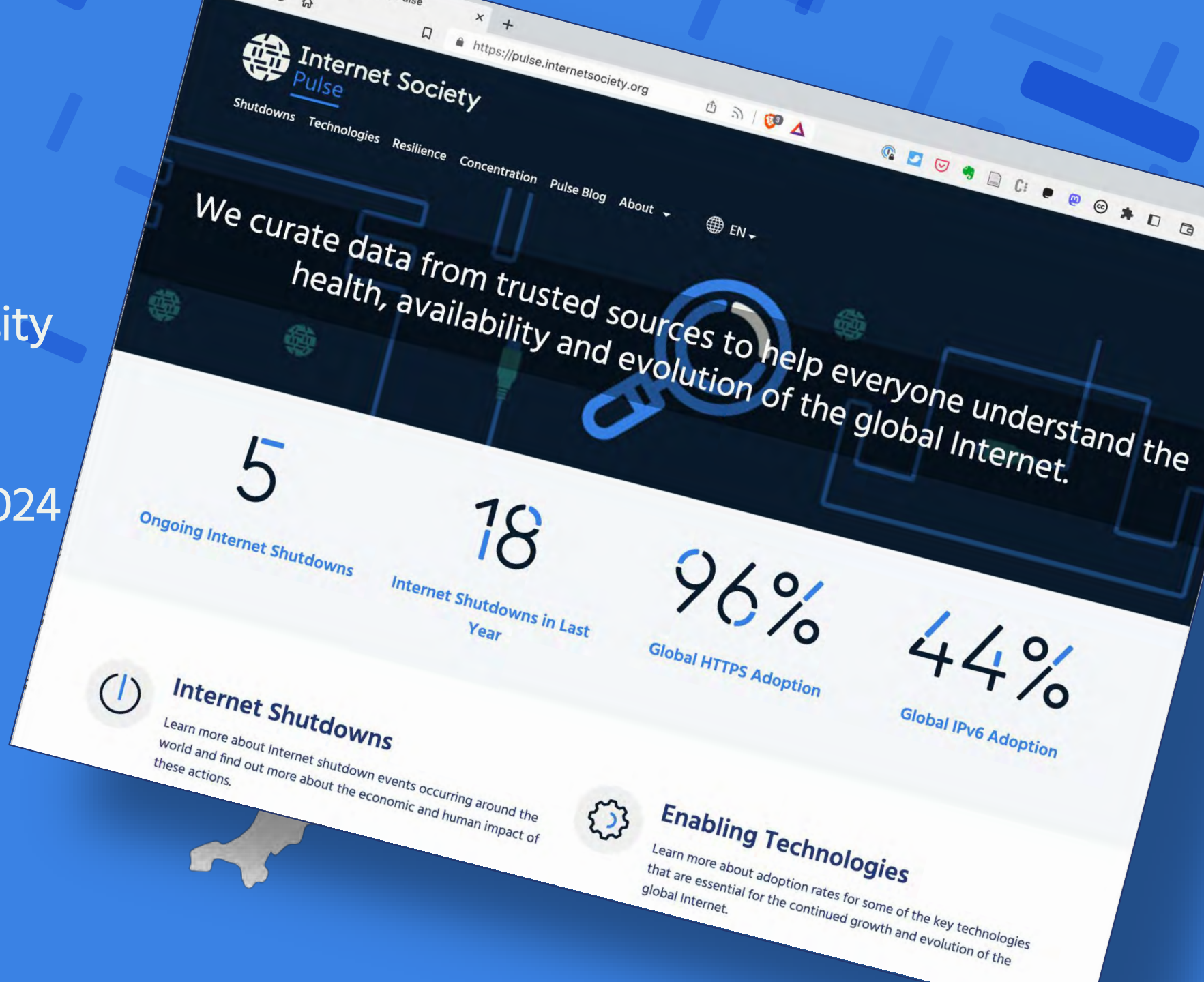
PhD student at
Loughborough University



About Me

PhD student at
Loughborough University

Internet Society Pulse
Research Fellowship 2024



Contents

1. Goals
2. Locality Definition
3. Motivation
4. Methodology
5. Results



Goals



Goals



Goals

50% local traffic in selected economies by 2025



Goals

50% local traffic in selected economies by 2025

50/50 Vision



Goals

50% local traffic in selected economies by 2025

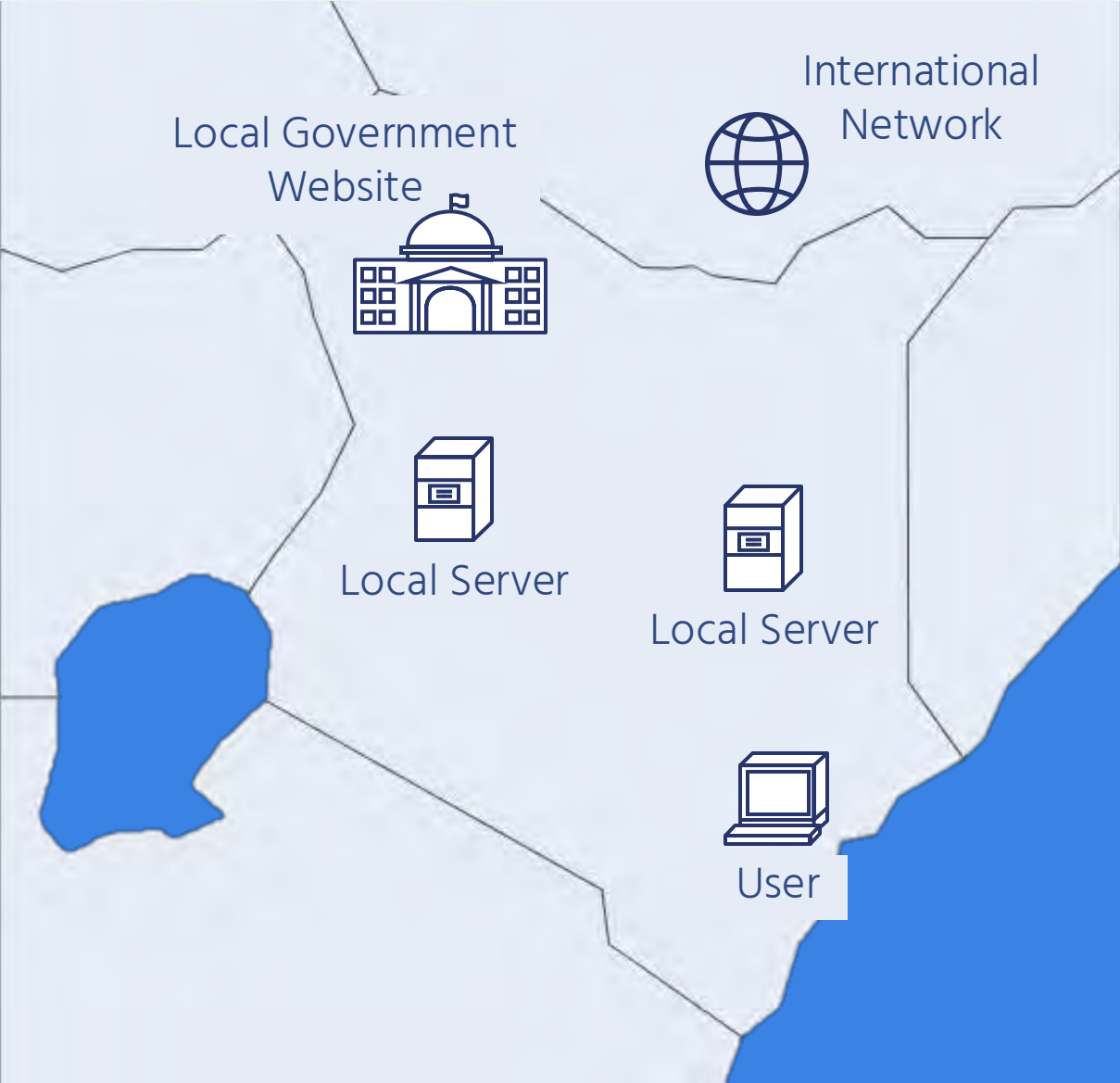
50/50 Vision

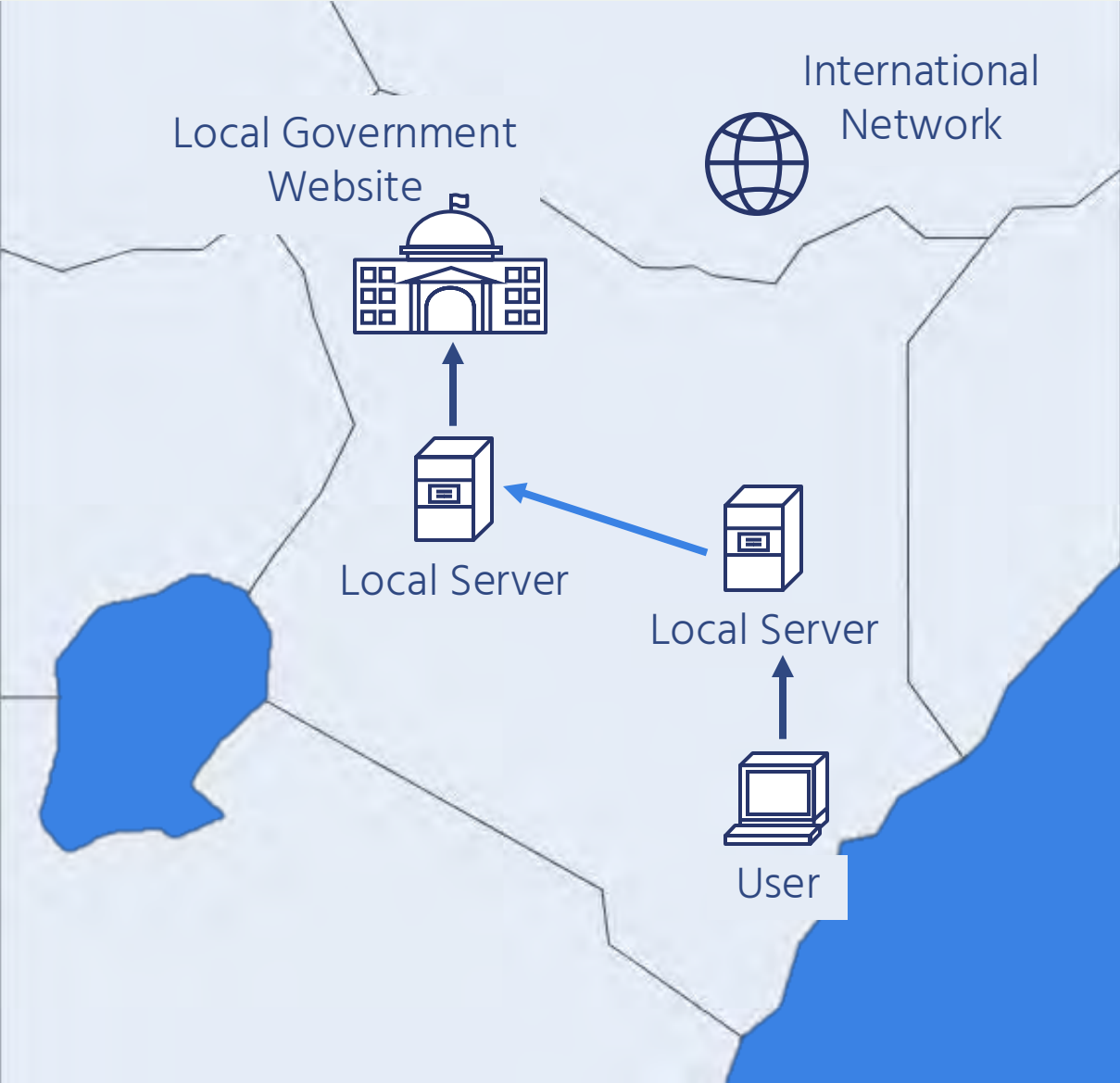
Measurement is key



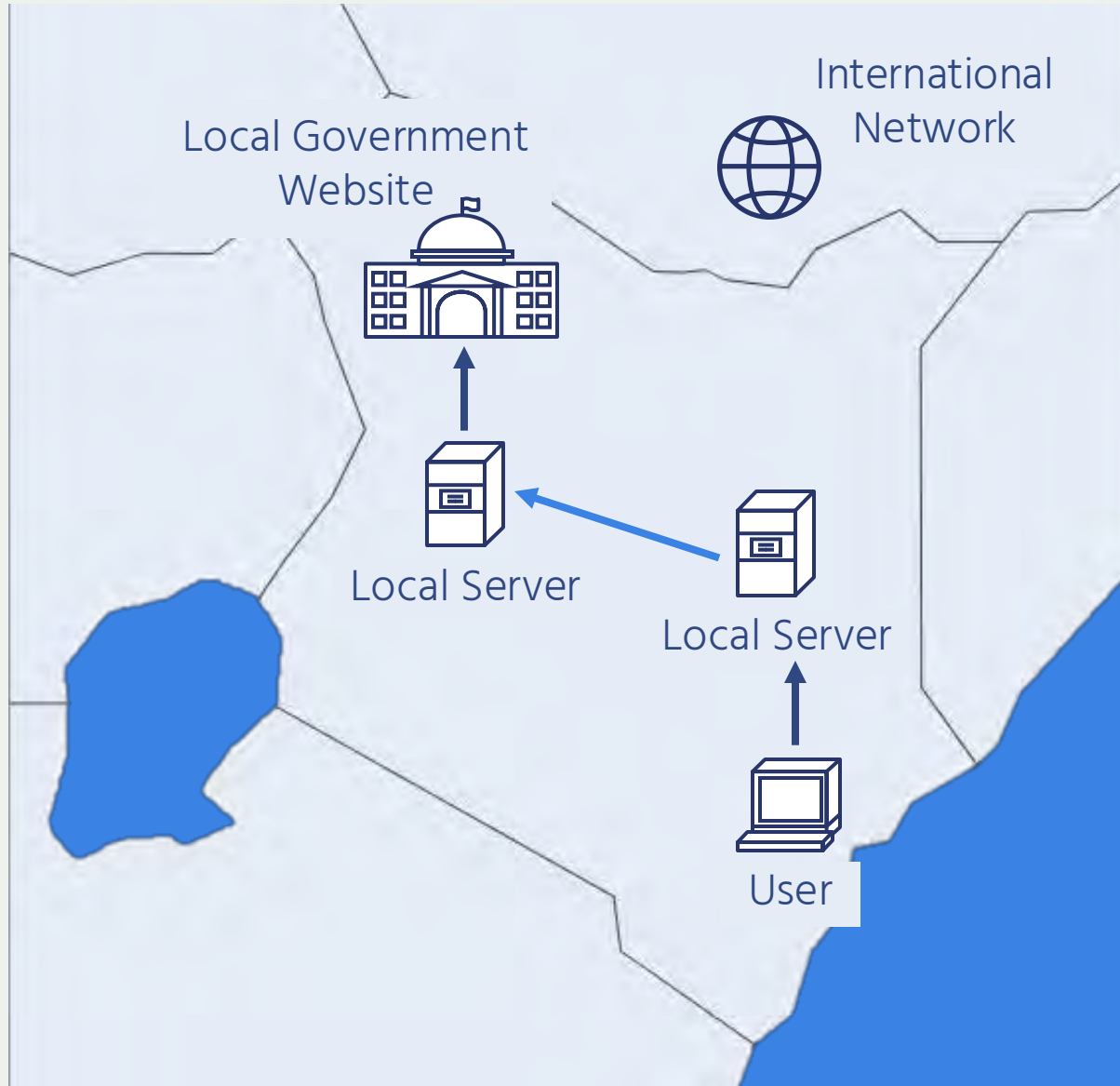
Locality Definitions



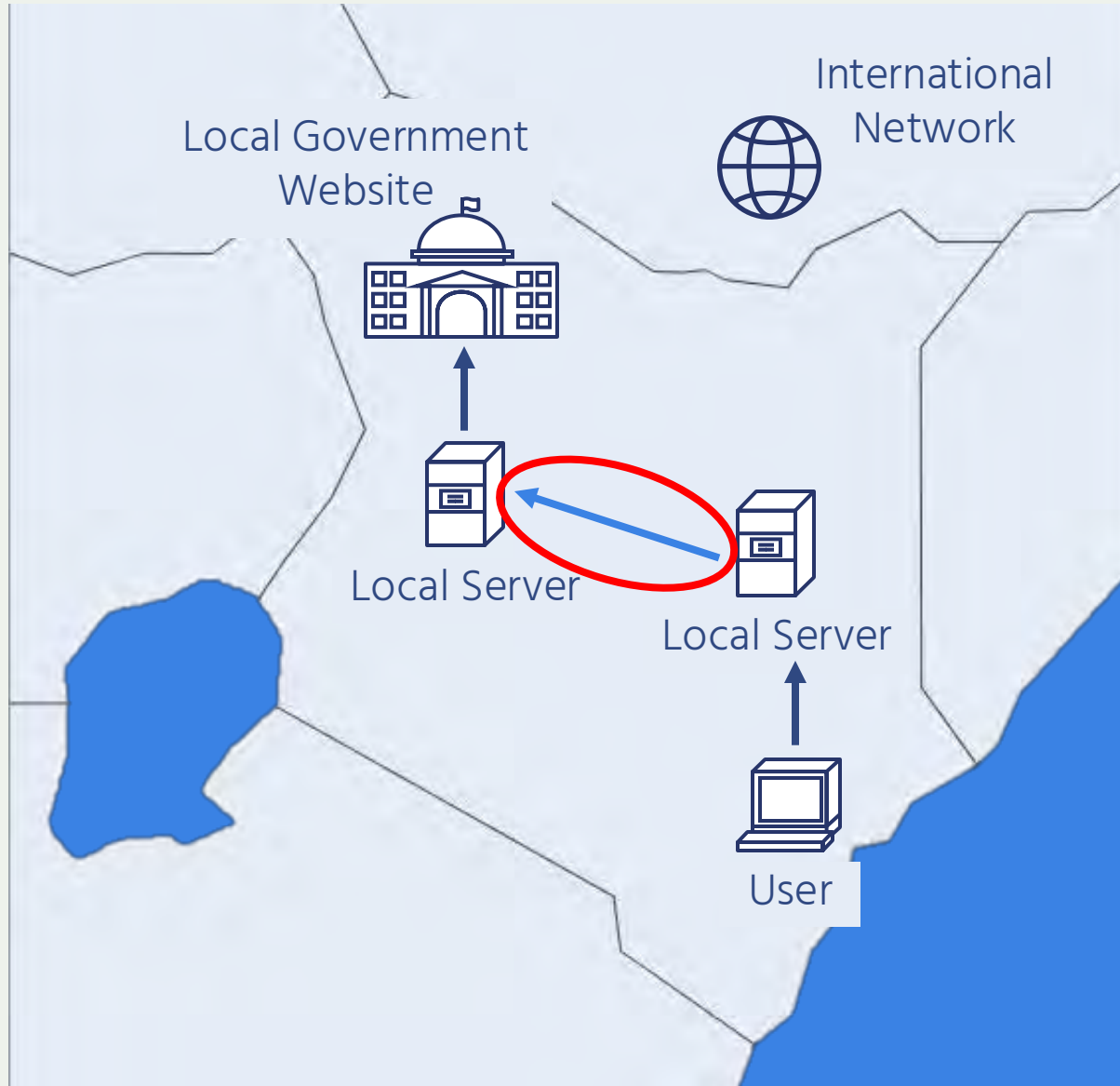


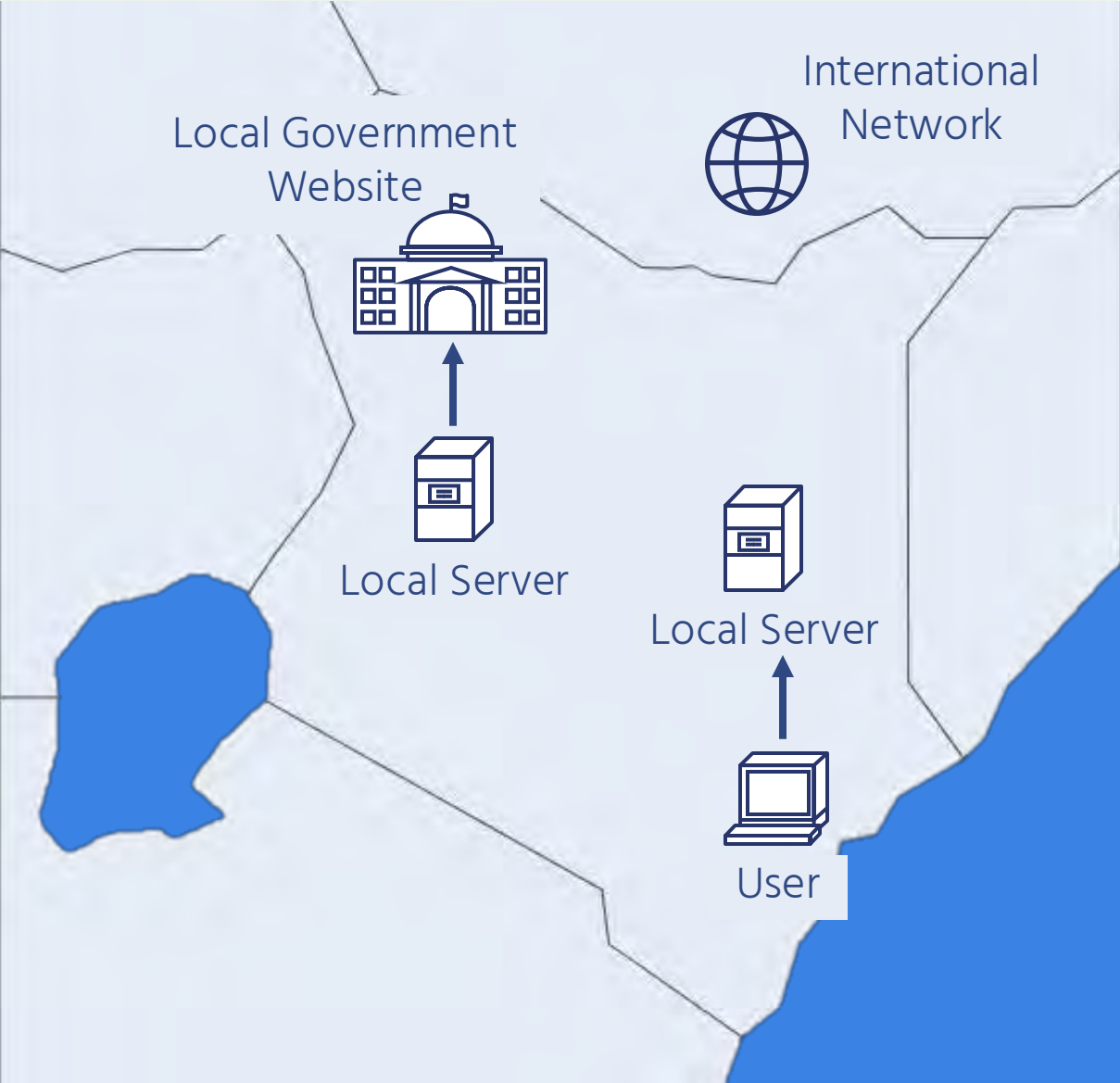


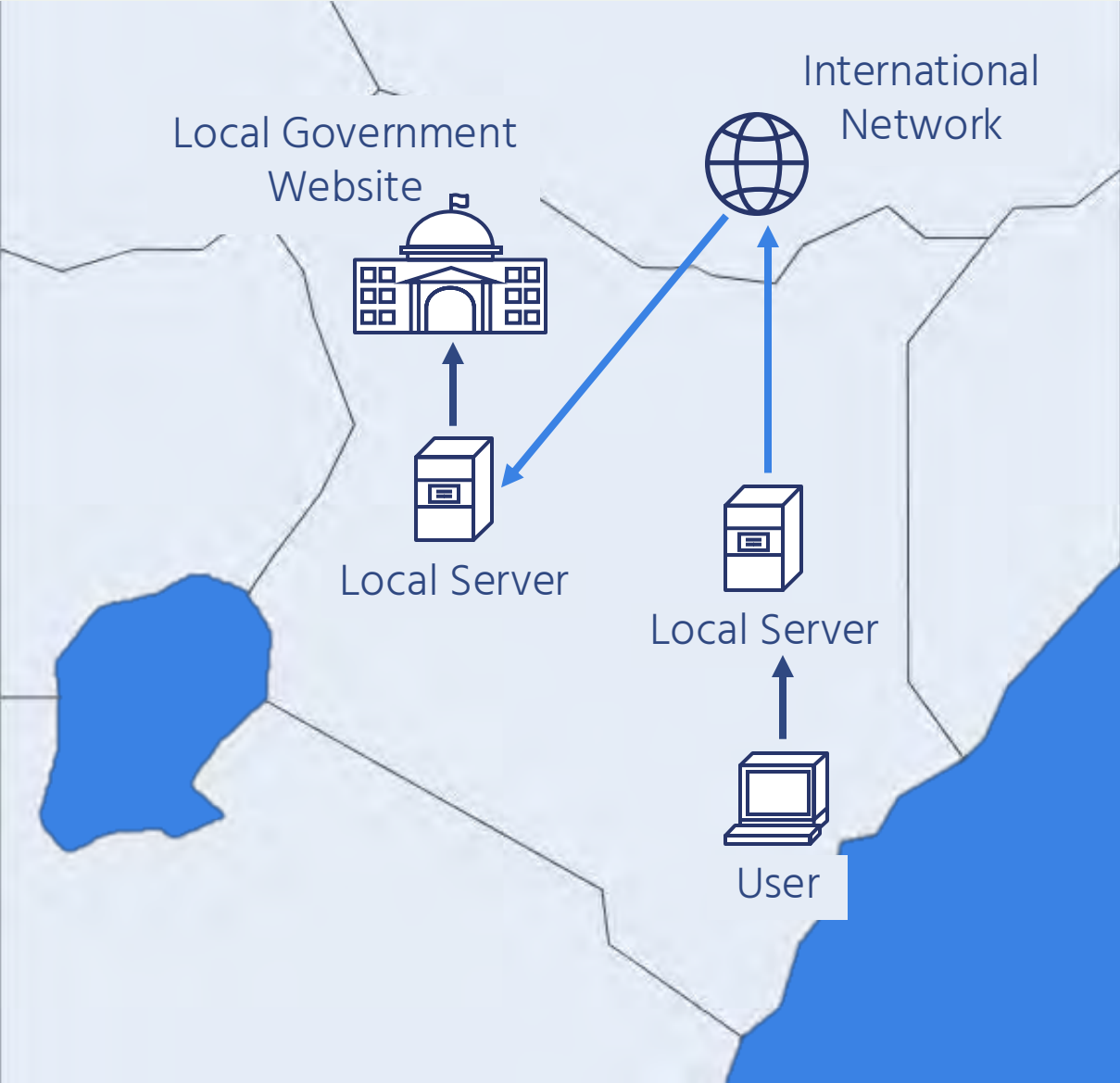
Local Traffic



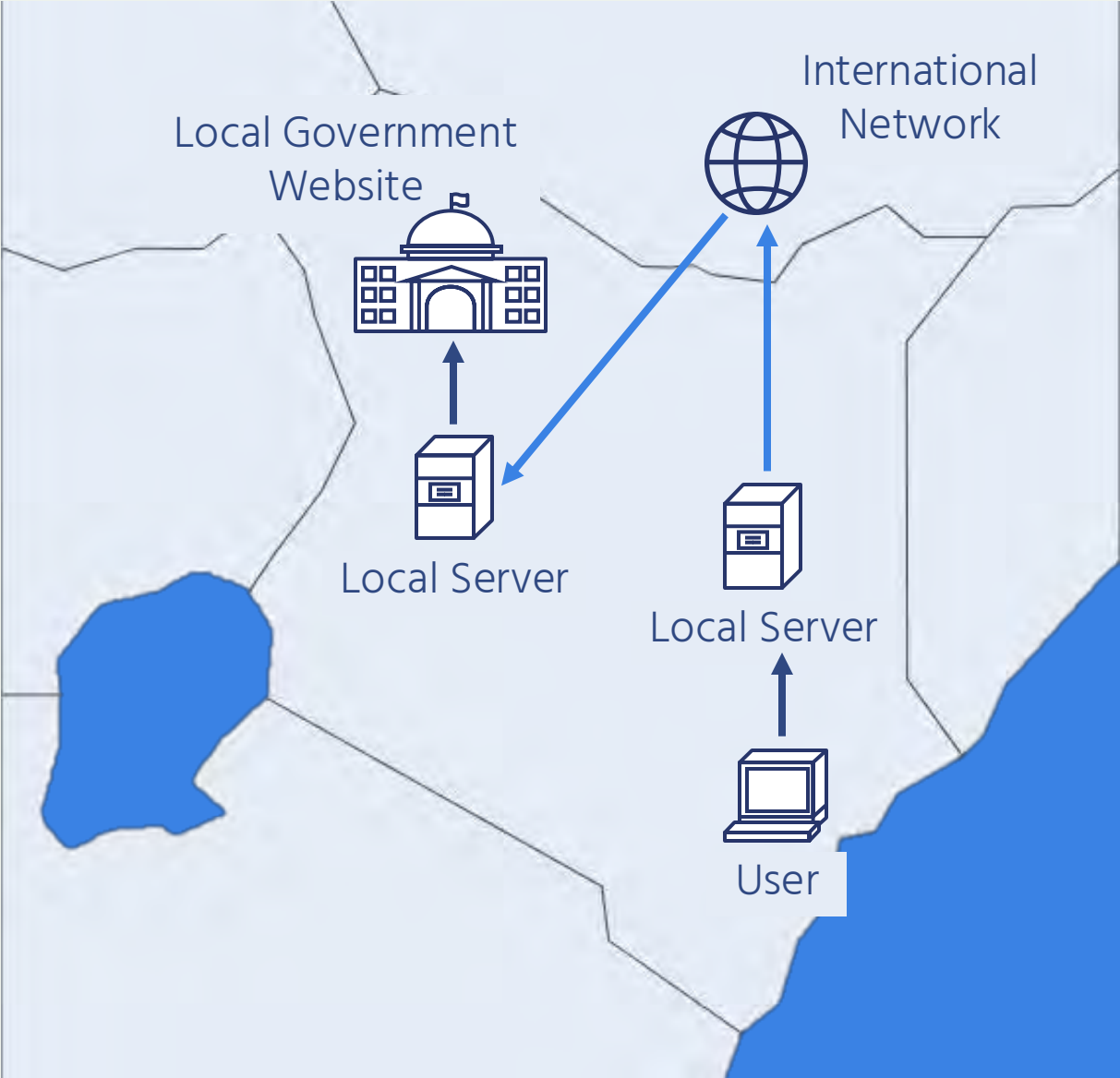
Local Traffic







External Traffic



Motivation



Motivation



Motivation

Policy

Performance

Persistence
(availability)



Motivation

Policy

South Africa's POPI
Act / EU's GDPR

Internet Society's
50/50 Vision

Performance

Persistence (availability)



Motivation

Policy

South Africa's POPI Act / EU's GDPR

Internet Society's 50/50 Vision

Performance

Traveling shorter distances is faster

Improves UX through cheaper connectivity

Persistence (availability)



Motivation

Policy

South Africa's POPI Act / EU's GDPR

Internet Society's 50/50 Vision

Performance

Traveling shorter distances is faster

Improves UX through cheaper connectivity

Persistence (availability)

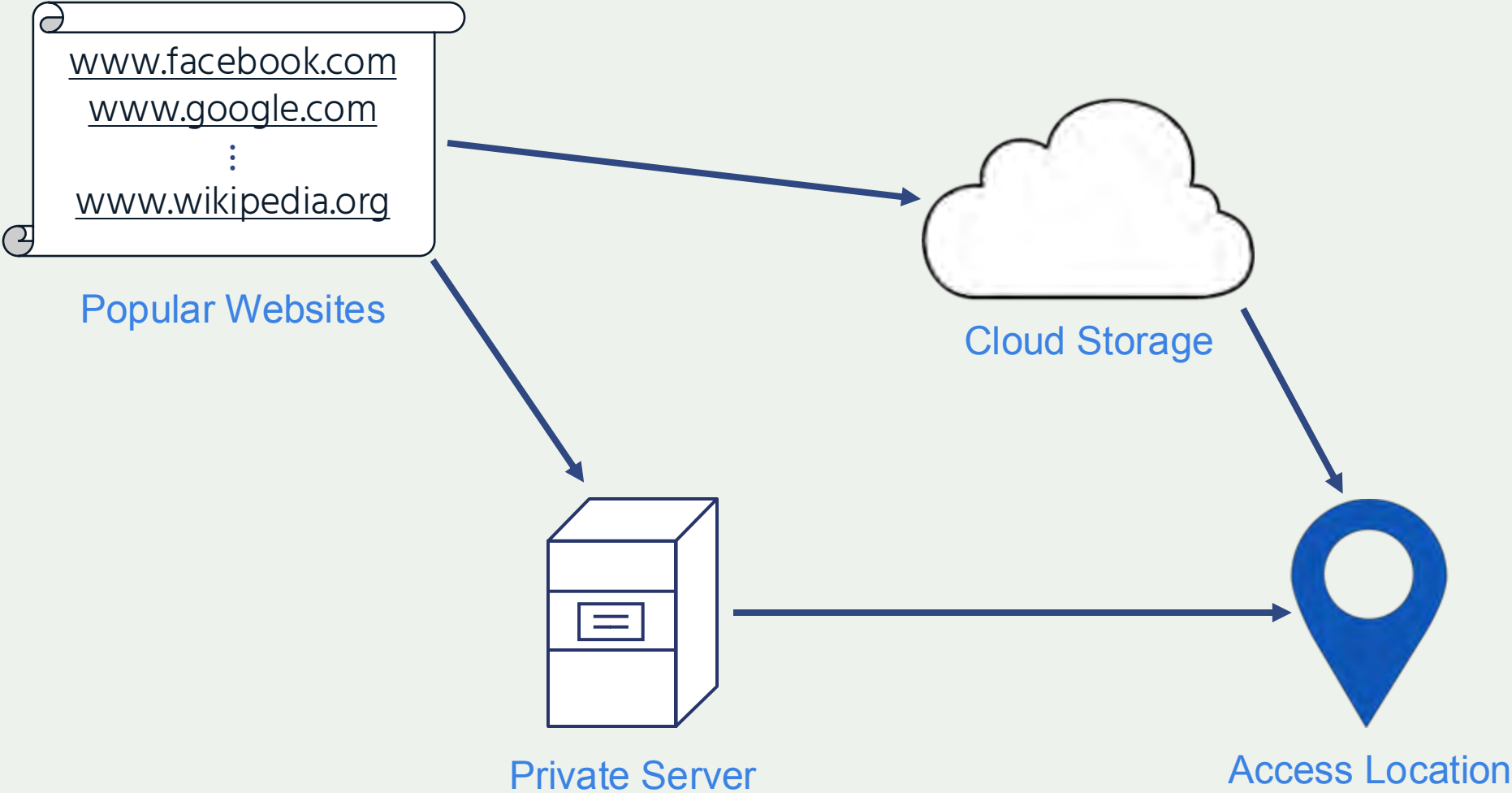
Resilience against faults e.g. cable cuts



Methodology



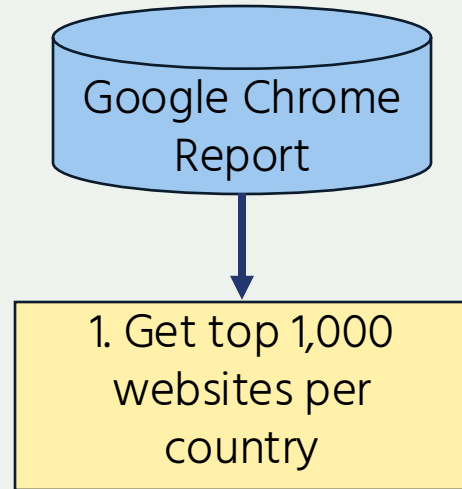
Methodology



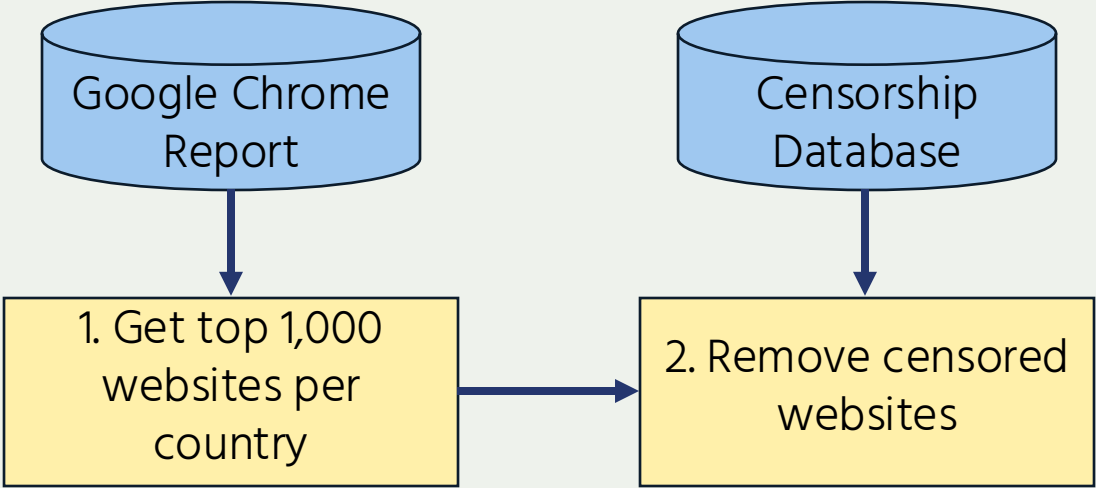
Methodology Steps



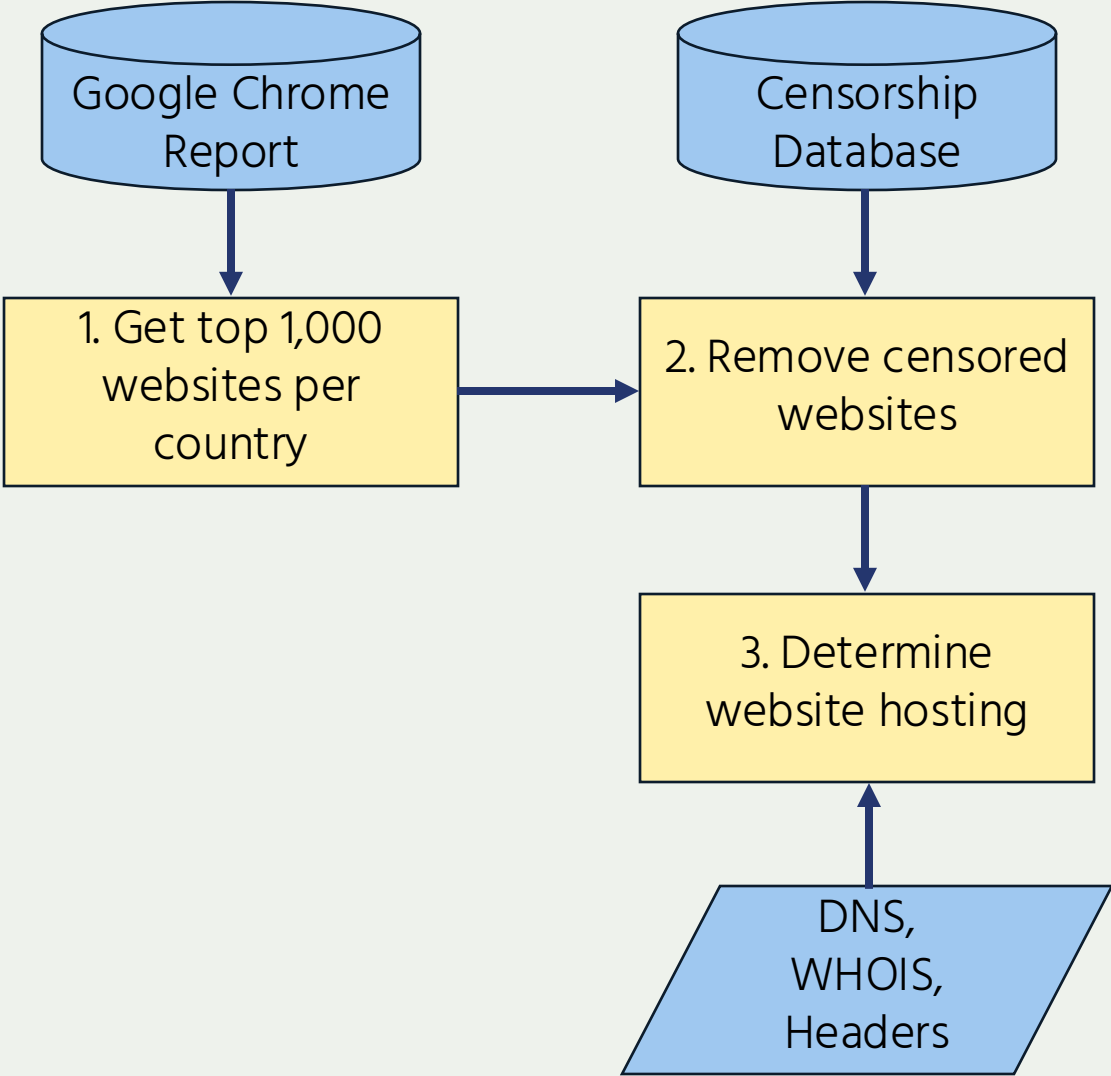
Methodology Steps



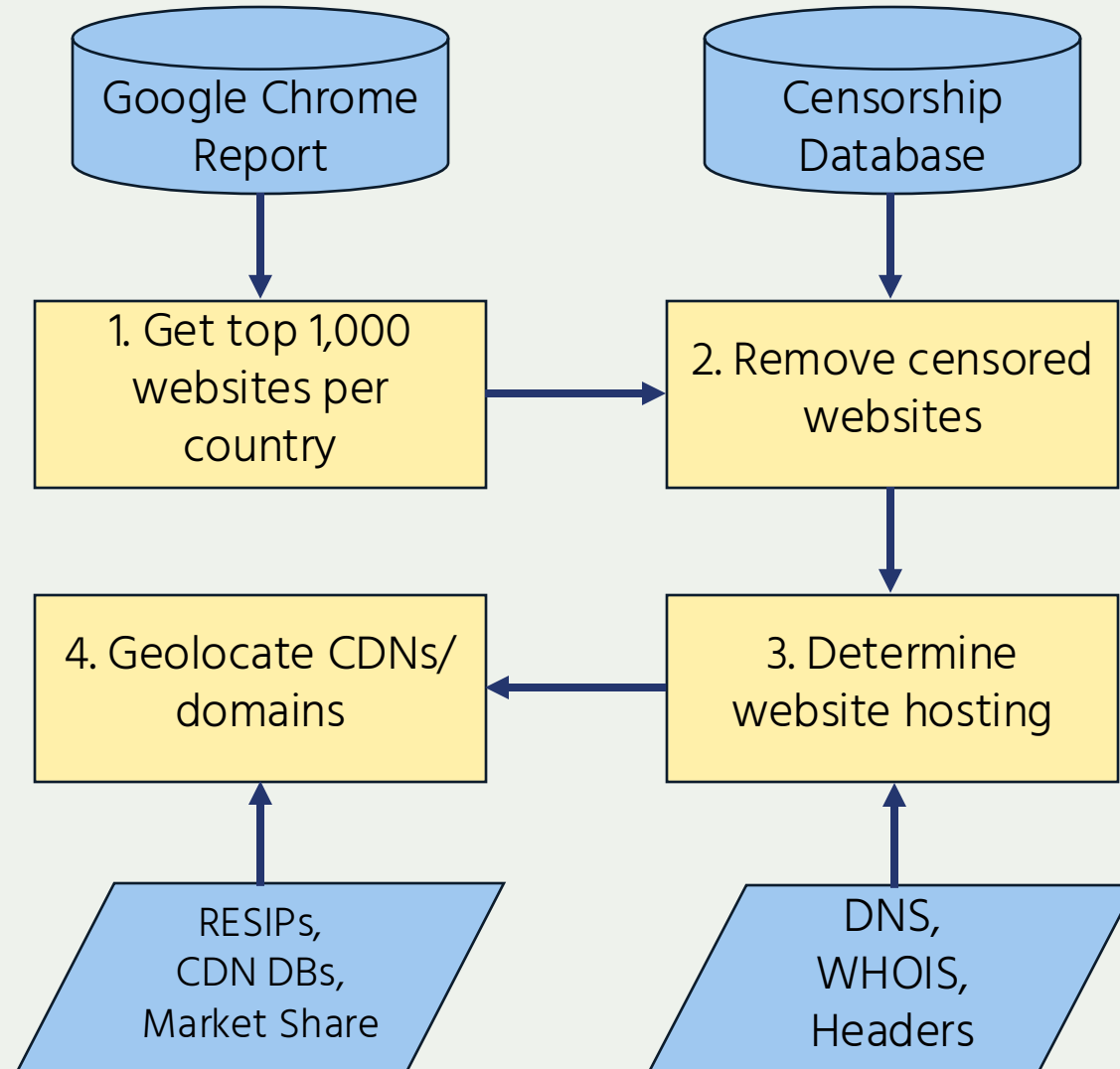
Methodology Steps



Methodology Steps



Methodology Steps

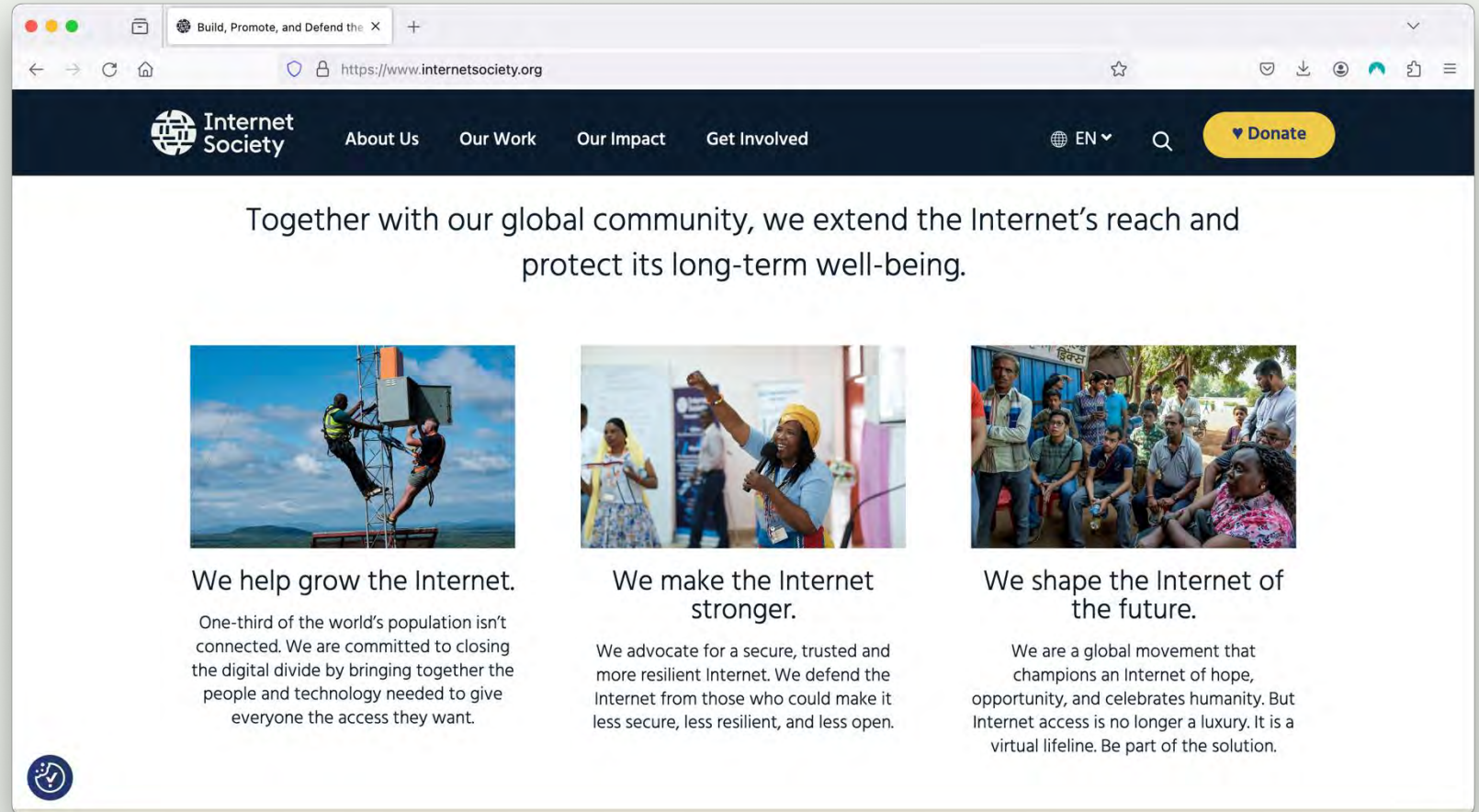


Limitations



Limitations

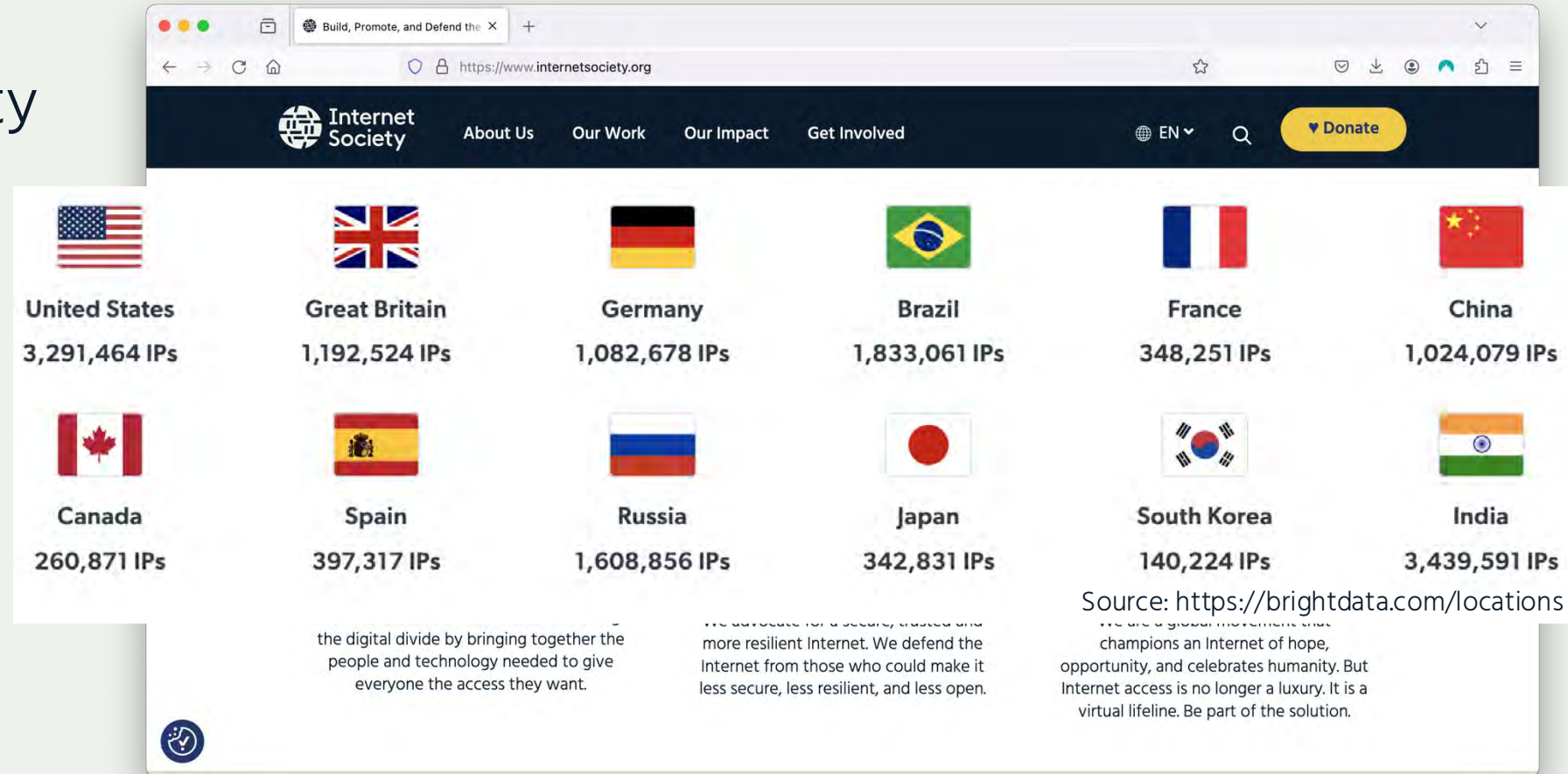
Website Complexity



Limitations

Website Complexity

Proxy Probes



Limitations

Website Complexity

Proxy Probes

Fog of Cloud

The image shows a screenshot of the Internet Society website (https://www.internetsociety.org) with a large white cloud graphic overlaid in the center. Inside the cloud is a large red question mark. The website content is partially obscured by the cloud. Visible text includes the site's name, a 'Donate' button, and a list of countries with their IP counts:

Country	IPs
United States	3,291,464
Great Britain	1,190,000
Canada	260,871
Spain	397,000
China	1,024,079
India	3,439,591

Additional text visible on the page includes: "the digital divide", "people and technology", "everyone the access", "an Internet of hope", "and celebrates humanity. But", "access is no longer a luxury. It is a", "virtual lifeline. Be part of the solution.", and a URL: "https://brightdata.com/locations".



Methodology Improvements



Hosting Platform Identification



Hosting Platform Identification

Pythia – 35 hours



Hosting Platform Identification

Pythia – 35 hours

FindCDN – 3.5 days



Hosting Platform Identification

Pythia – 35 hours

FindCDN – 3.5 days

Our tool – 12 minutes



Hosting Platform Identification

Pythia – 35 hours

FindCDN – 3.5 days

Our tool – 12 minutes

>95% accuracy

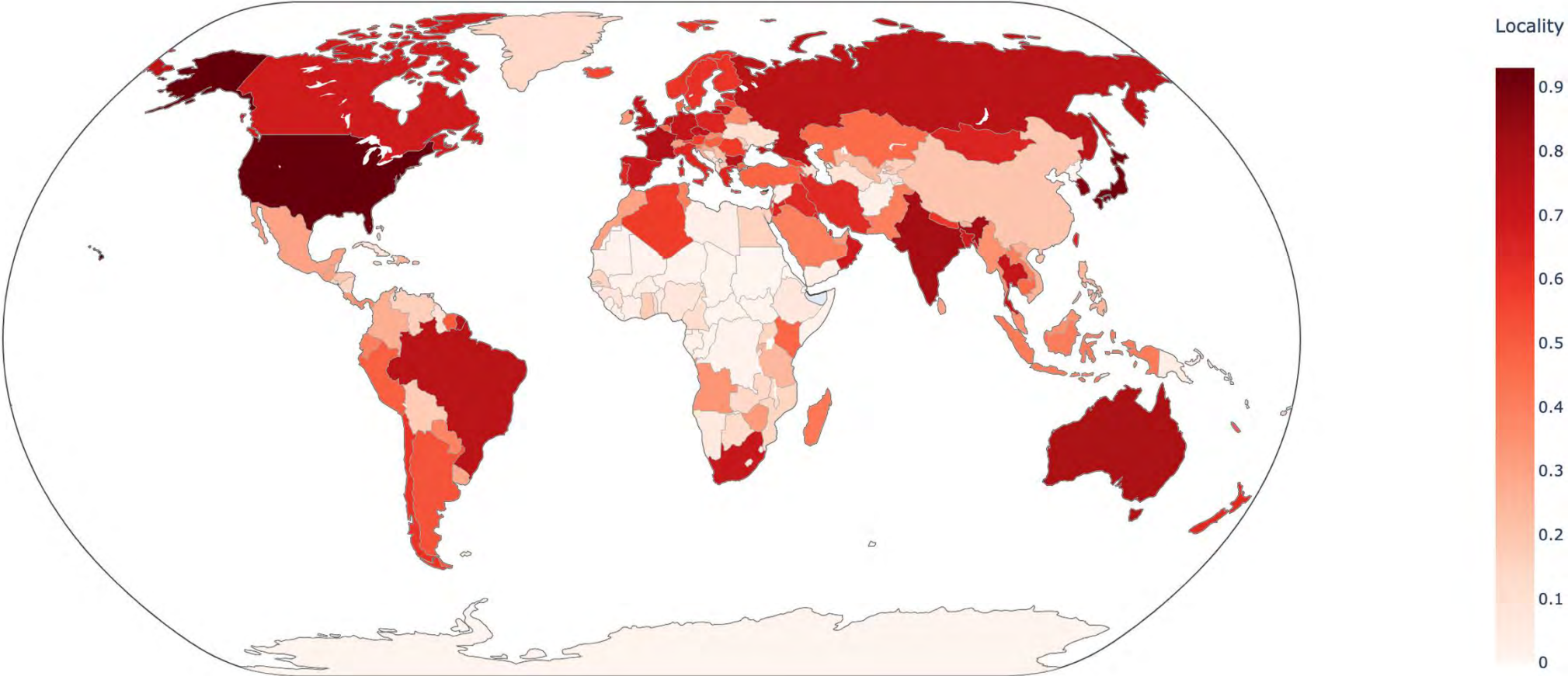


Preliminary Locality Results

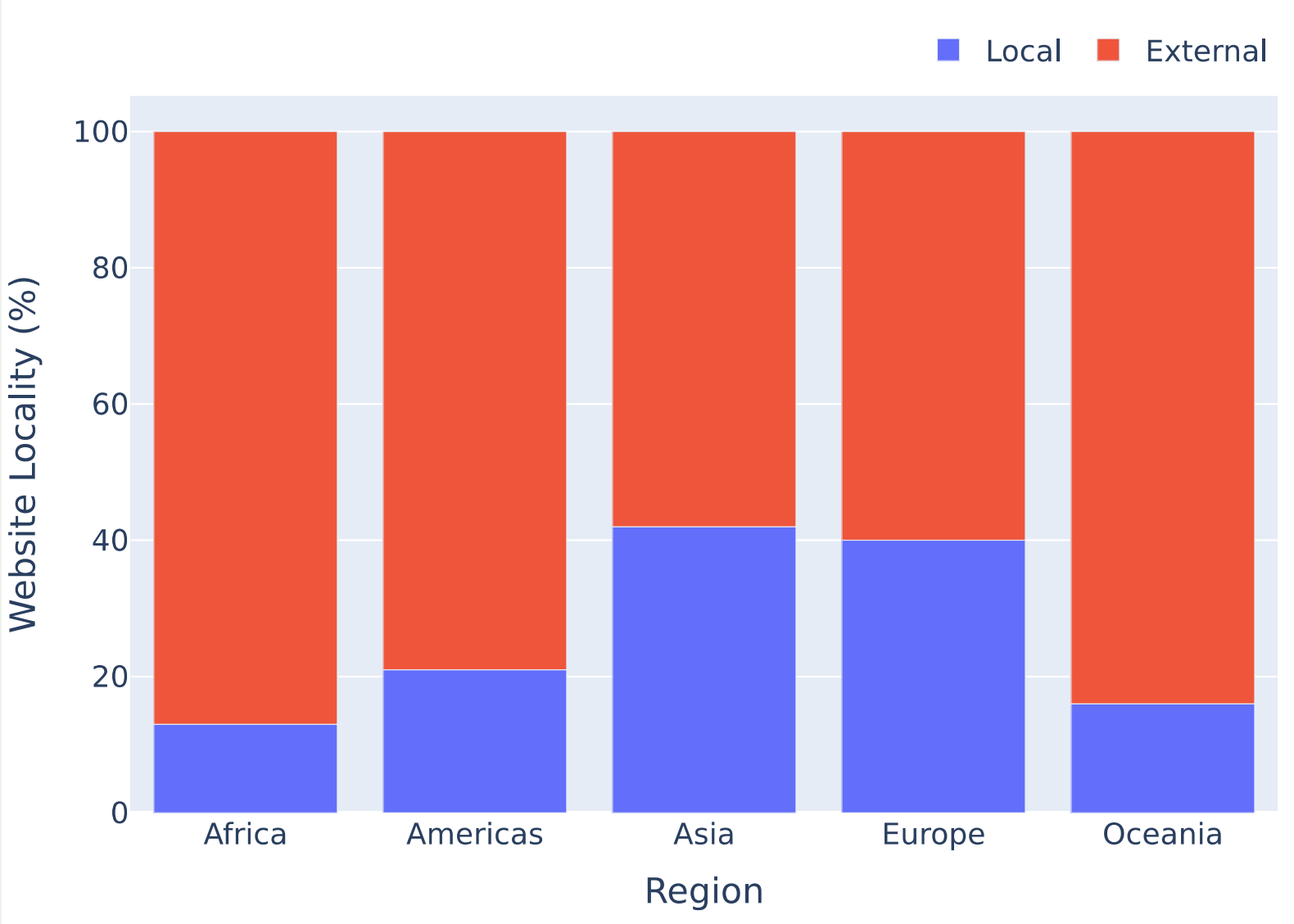


Global Locality

Locality Results: 2024-06-07



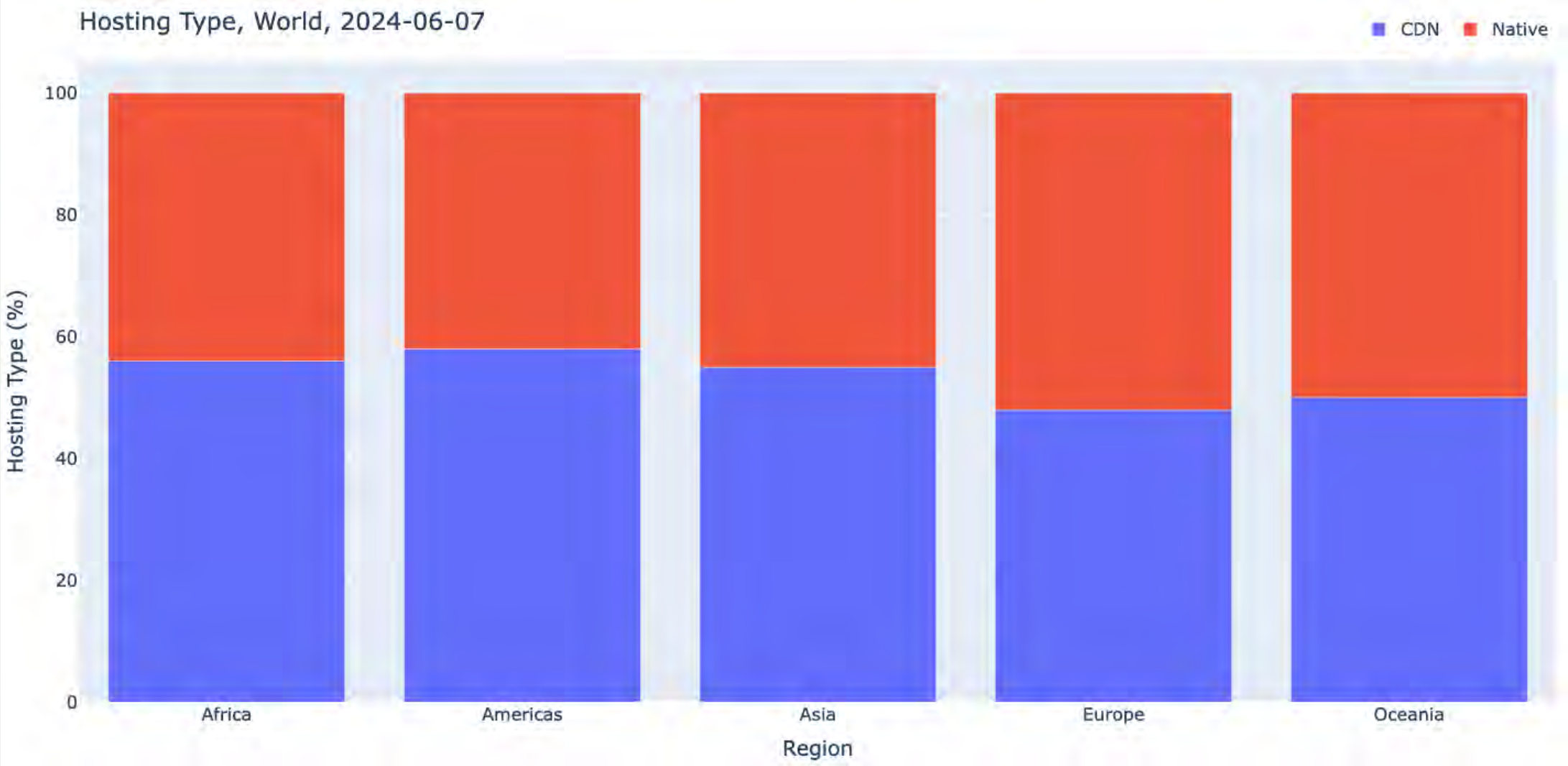
Regional Locality



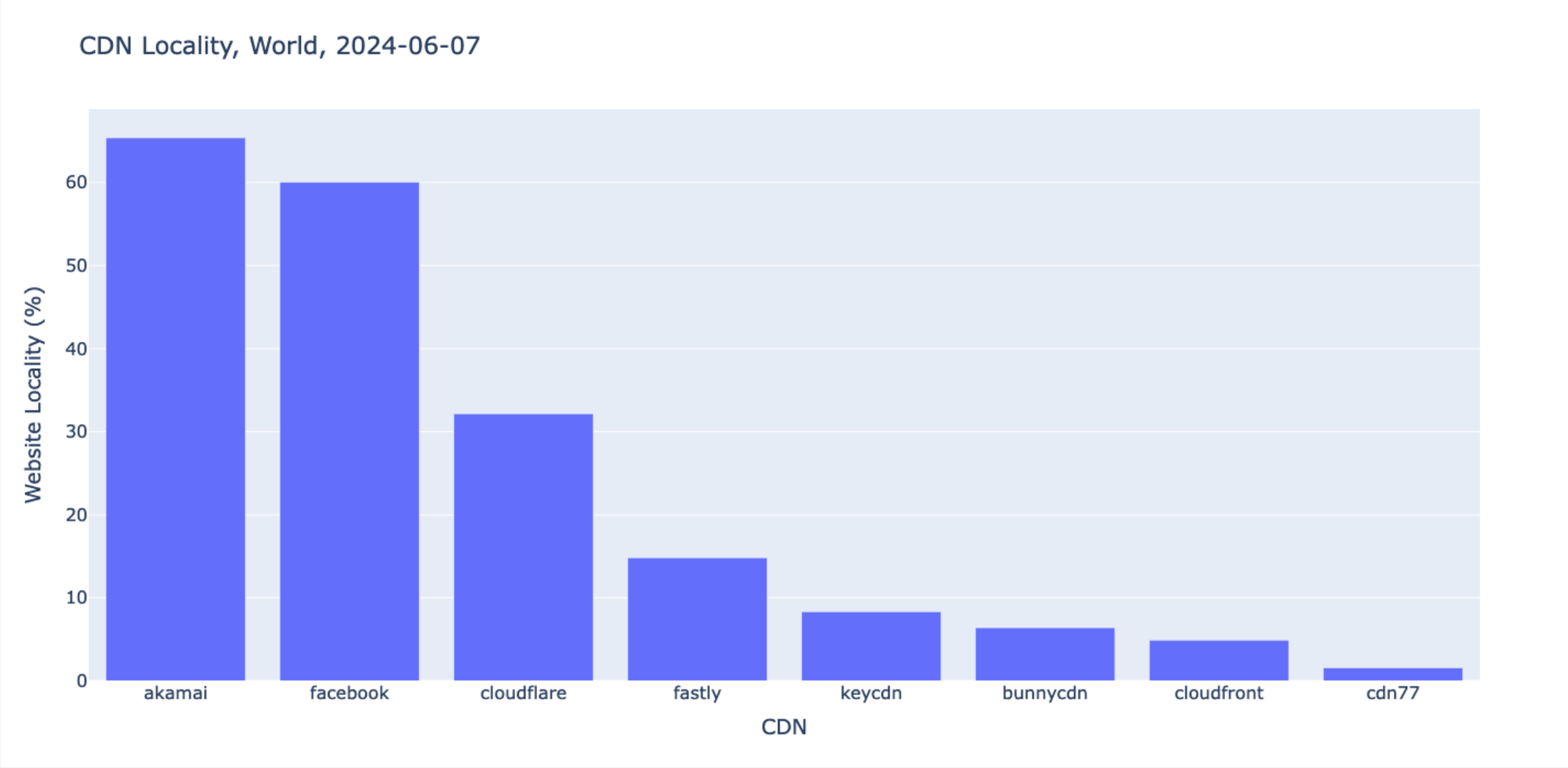
Other Preliminary Results



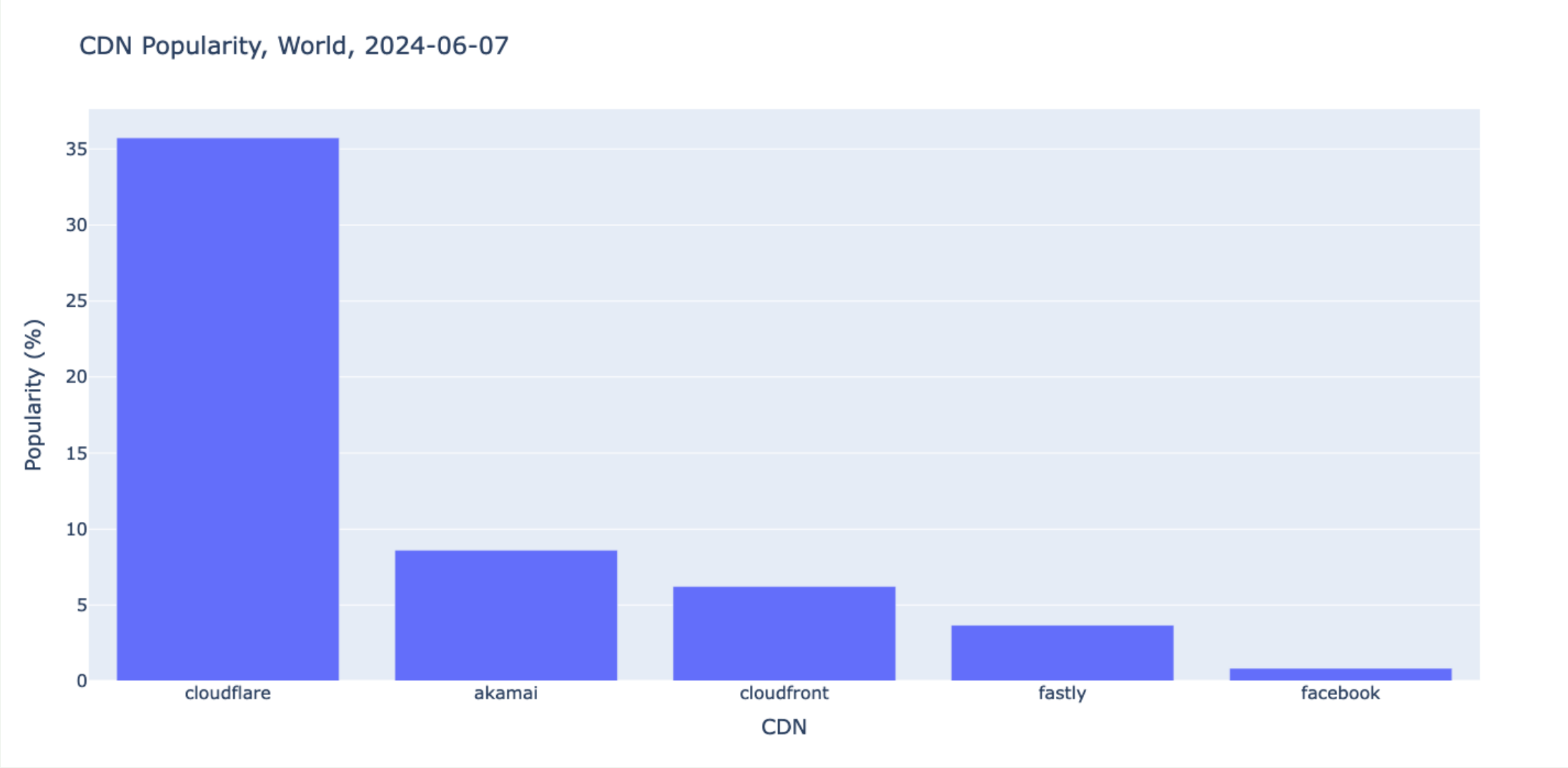
Hosting Type (World)



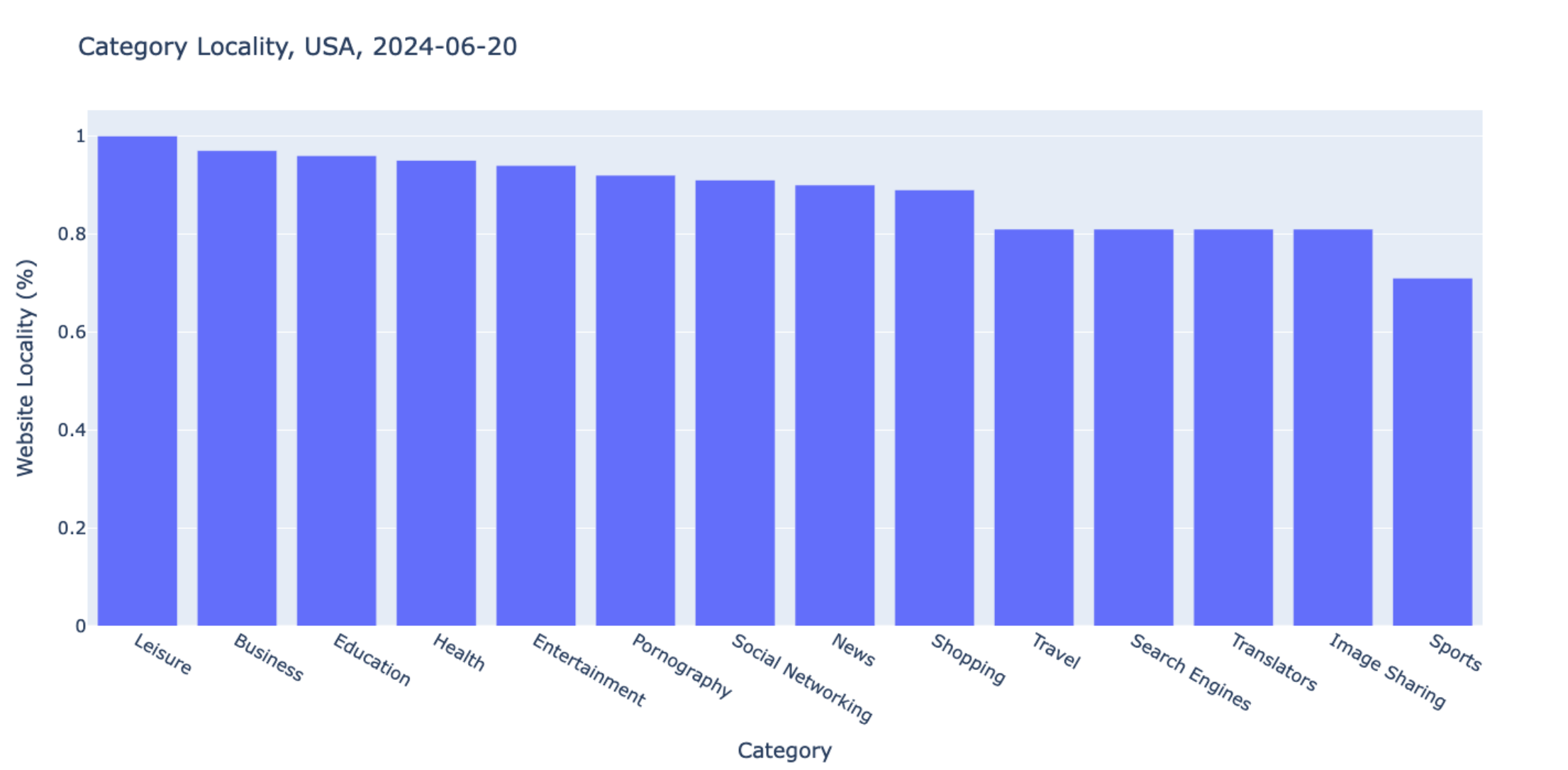
CDN Locality (World)



5 Most Popular CDNs (World)



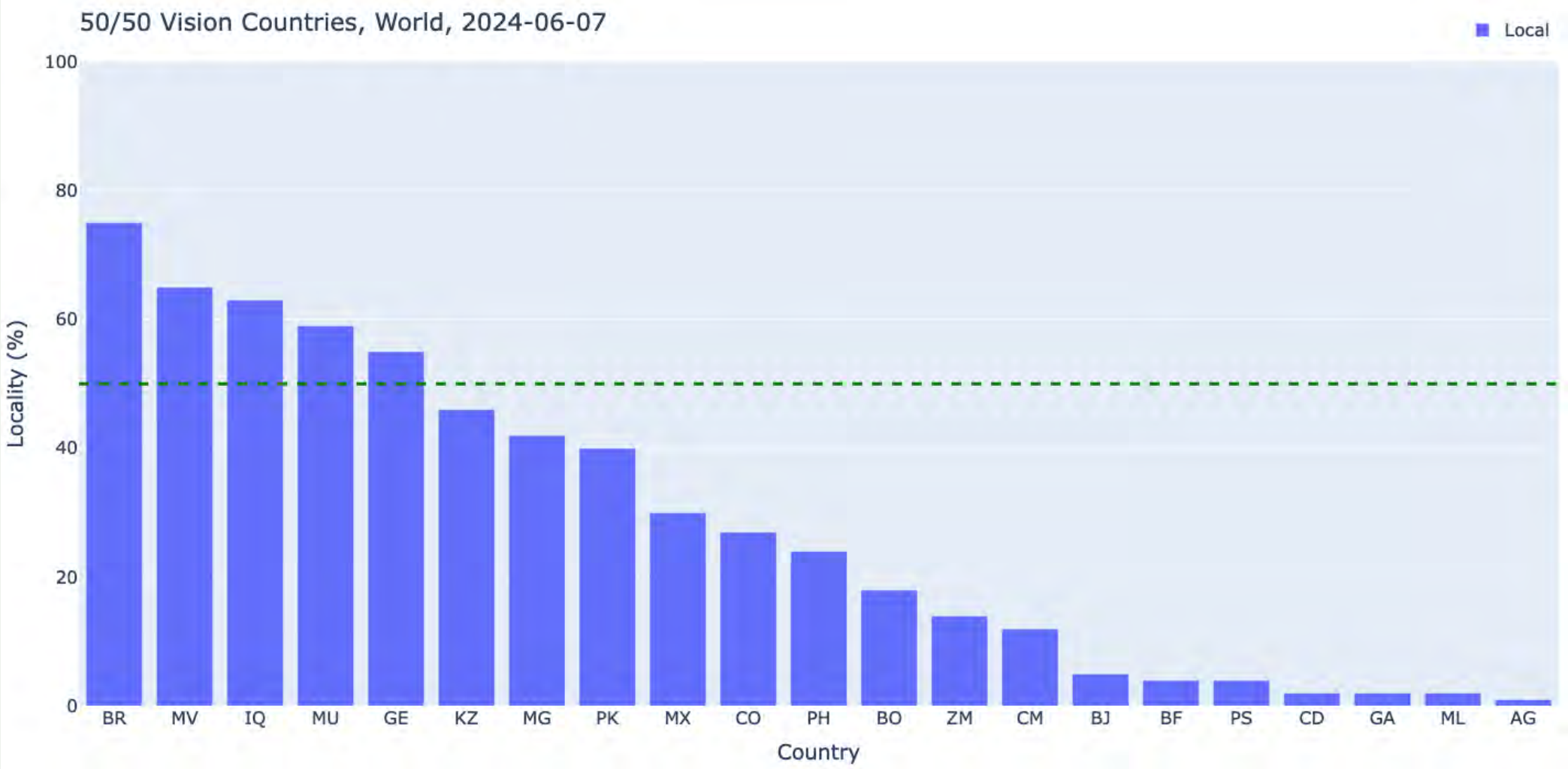
Categories Locality (USA, 100 Sites)



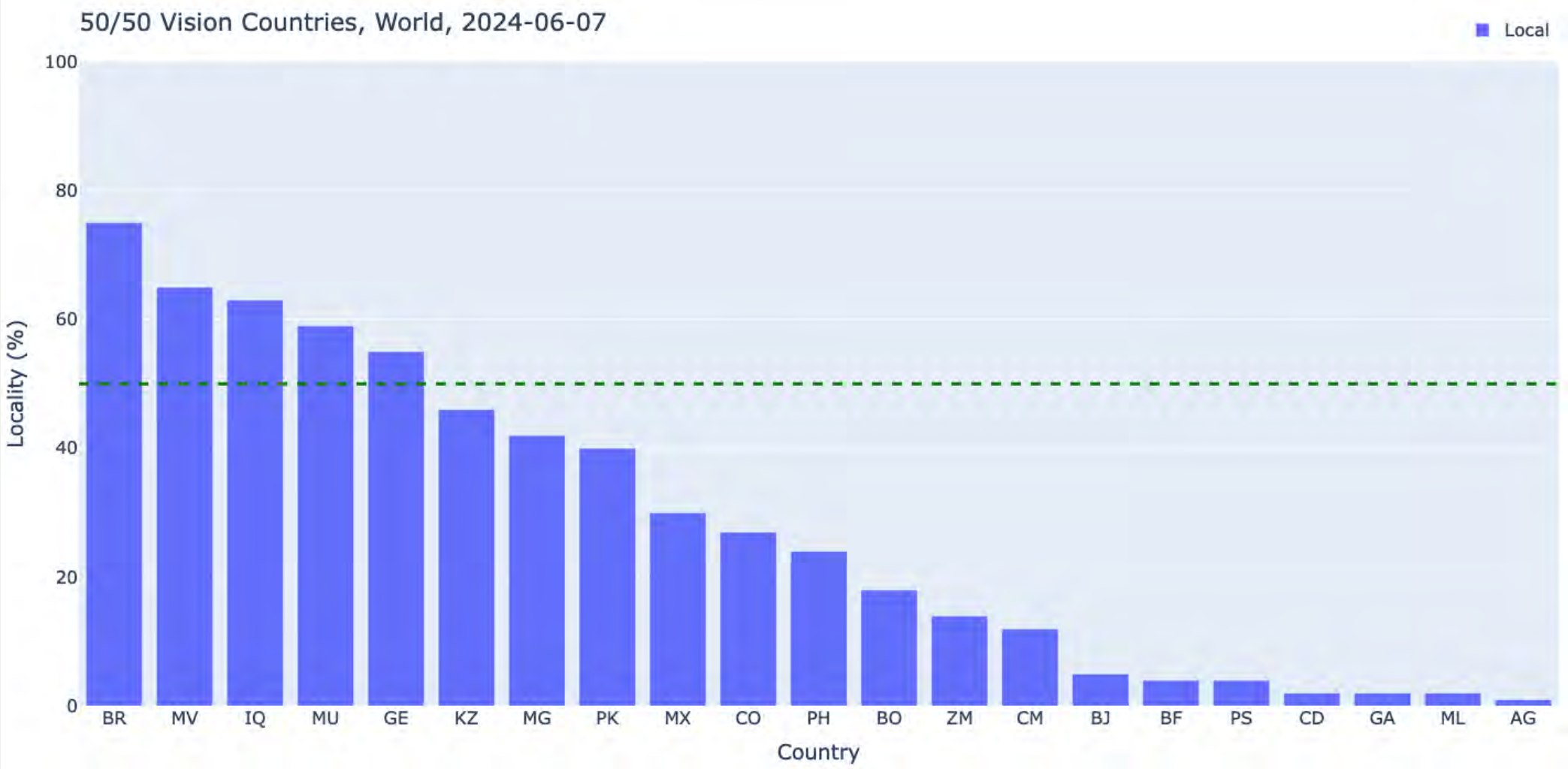
How close to the 50/50 Vision are we?



Locality: 50/50 Vision



Locality: 50/50 Vision



Selected Economies: 24%



Global Traffic Locality

Region	Countries > 50% Local	Countries >50% Local (%)
Africa	3	5%
Americas	7	12%
Asia	21	41%
Europe	21	42%
Oceania	4	18%



Global Traffic Locality

Region	Countries > 50% Local	Countries >50% Local (%)
Africa	3	5%
Americas	7	12%
Asia	21	41%
Europe	21	42%
Oceania	4	18%

Worldwide 50/50 Vision: 24%



Next Steps

Expand the testing methodology

Continue testing to find changes/patterns

Publish our results in papers and via Pulse platform

Towards Measuring Content Locality

James I. Madeley
Loughborough University
j.i.madeley@lboro.ac.uk

Aftab Siddiqui
Internet Society
siddiqui@isoc.org

Amreesh Phokeer
Internet Society
phokeer@isoc.org

Theophilus A. Benson
Carnegie Mellon University
theophilus@cmu.edu

Country Report

 **Mauritius**
Africa, Eastern Africa

Search for countries: ← →

Open Internet Environment

The open Internet allows people and organizations to mix and match technologies without permission and with minimal barriers. Sustaining and growing an open Internet helps to spur innovation and keep it fit for future applications. An open Internet is an accessible Internet – it is easy to connect to the open Internet and use its services.

Internet Use

Individuals using the Internet as a percentage of the total population

68%

Regional Rank: 12

39% Africa avg.



Internet Resilience Score

Top 5

A resilient Internet connection is one that maintains an acceptable level of service in the face of faults and challenges to normal operation

61%

Regional Rank: 1

35% Africa avg.



Transit Provider Diversity

More diversity in routes to the global Internet improves connection resilience

Good



Retail ISP Diversity

Diversity of retail Internet providers improves resilience and user choice

Very Poor



IXP Operator Market

A measure of the diversity and concentration of the local market for Internet Exchange Point operations



Mauritius Inte...
100%

Popular Content Locality

A measure of how much locally popular web content is hosted in-country or in-region

63%

Regional Rank: 2

28% Africa avg.



See details



Summary

Local traffic is beneficial

Measuring Internet traffic locality over time

Focused on 50/50 Vision

Paper at ANRW'24

Any questions?

Contact: j.i.madeley@lboro.ac.uk



Check out the paper



Learn about the
50/50 Vision



Additional Slides



Related Work

Of Choices and Control – A Comparative Analysis of Government Hosting

Rashna Kumar

rashnakumar2024@u.northwestern.edu

Northwestern University

Evanston, IL, USA

Esteban Carisimo

esteban.carisimo@northwestern.edu

Northwestern University

Evanston, IL, USA

Lukas De Angelis Riva

ldeangelis@fi.uba.ar

Universidad de Buenos Aires

Buenos Aires, Argentina

Mauricio Buzzzone

mbuzzzone@fi.uba.ar

Universidad de Buenos Aires

Buenos Aires, Argentina

Fabián E. Bustamante

fabianb@northwestern.edu

Northwestern University

Evanston, IL, USA

Ihsan Ayyub Qazi

ihsan.qazi@lums.edu.pk

LUMS

Lahore, Pakistan

Mariano G. Beiró*

mbeiro@udesa.edu.ar

Universidad de San Andrés

Buenos Aires, Argentina



Future Work



Future Work

Recursive Search

Traceroute/Latency



Future Work

Recursive Search

Traceroute/Latency

Reasons for Locality

Visualisation



Future Work

Recursive Search

Traceroute/Latency

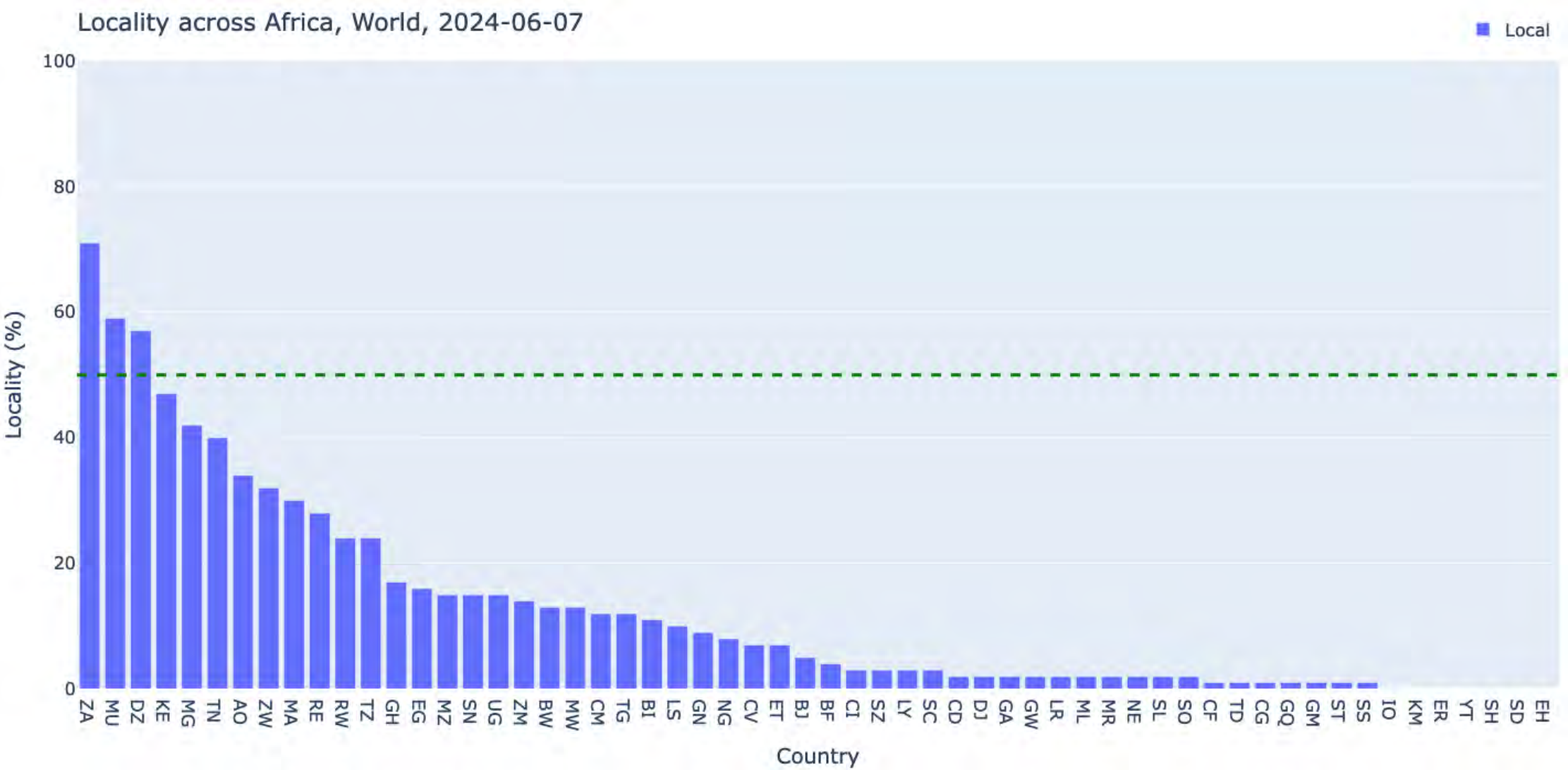
Reasons for Locality

Visualisation

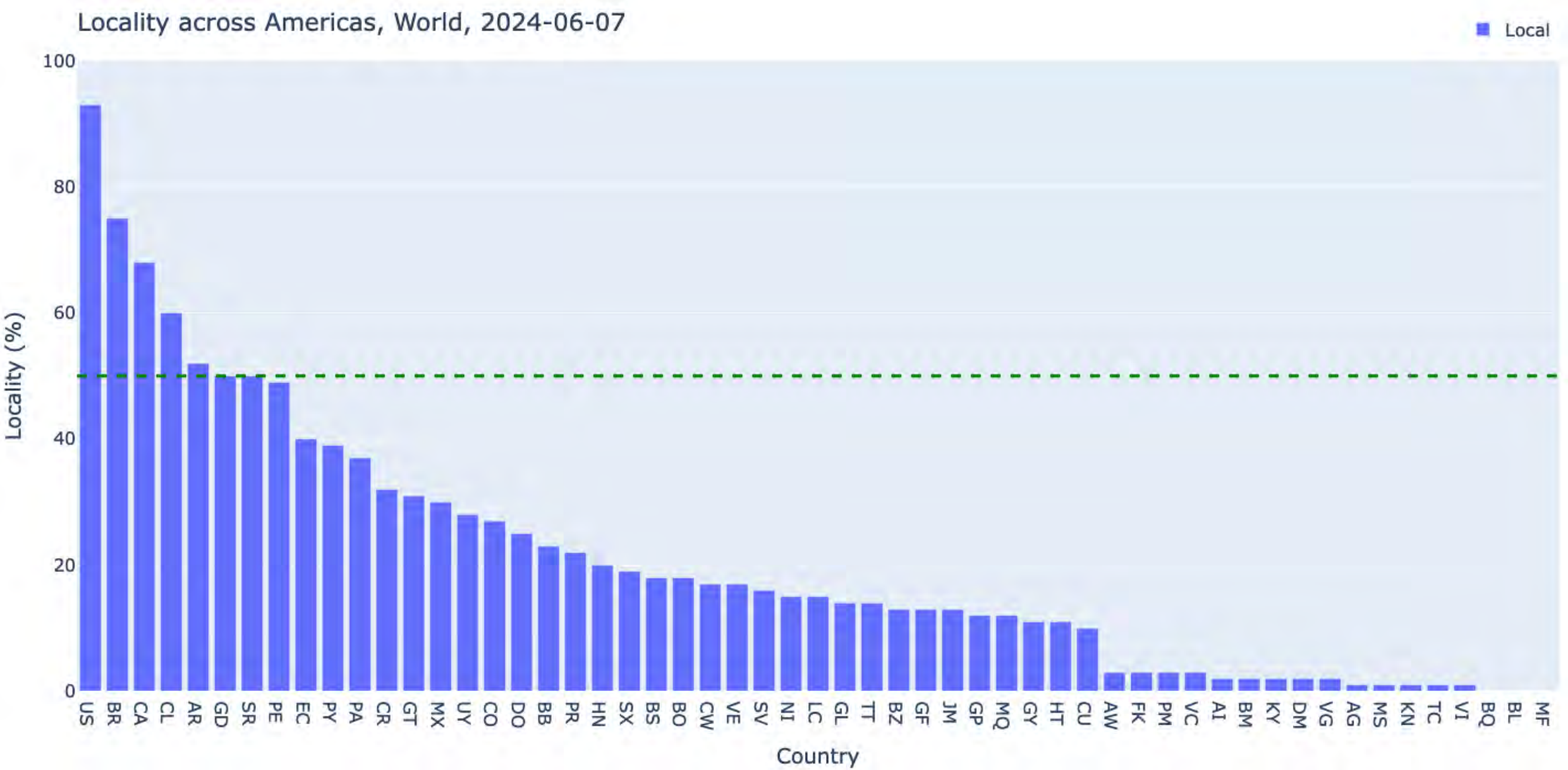
Categorization



Locality: Africa



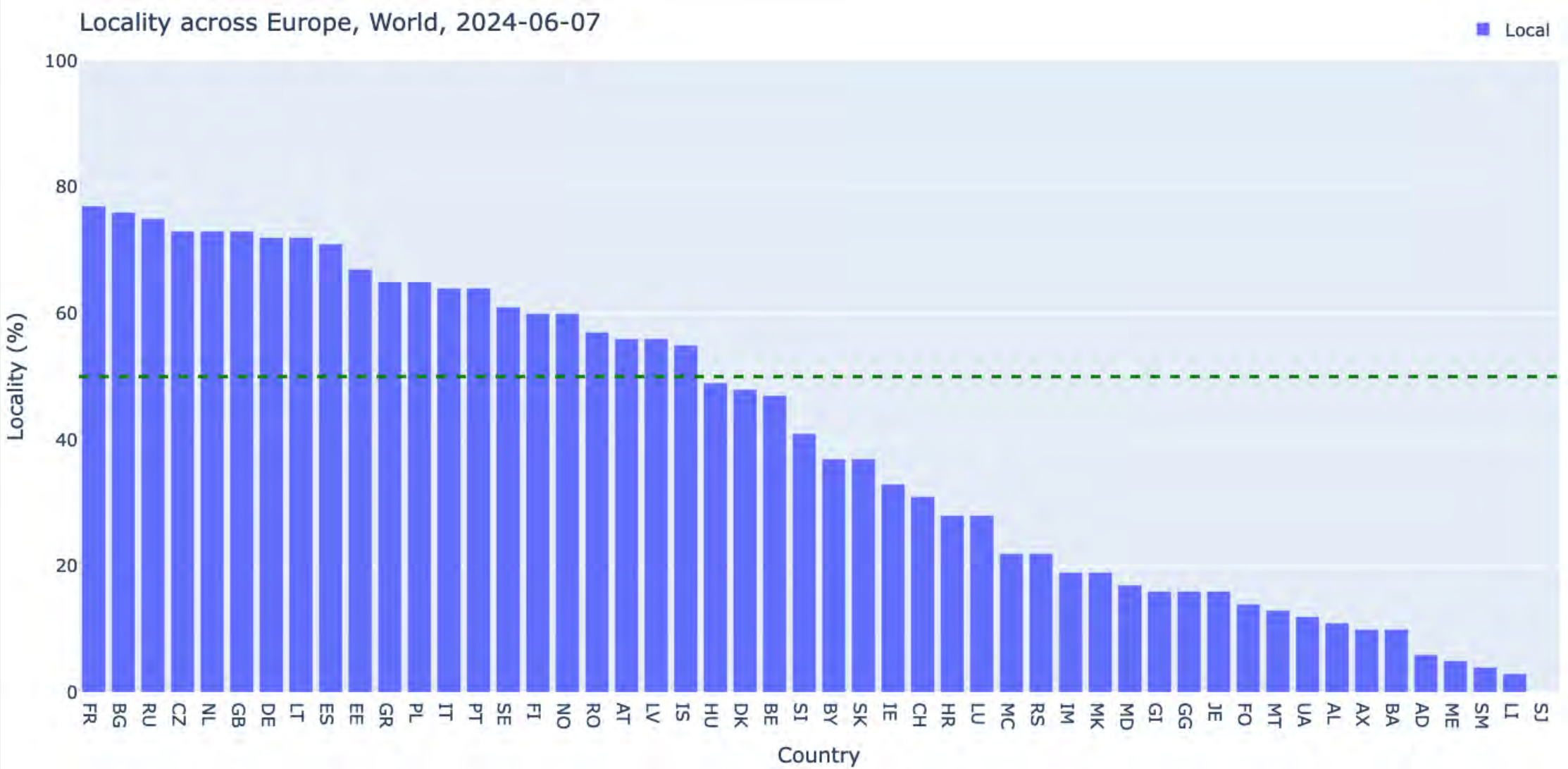
Locality: Americas



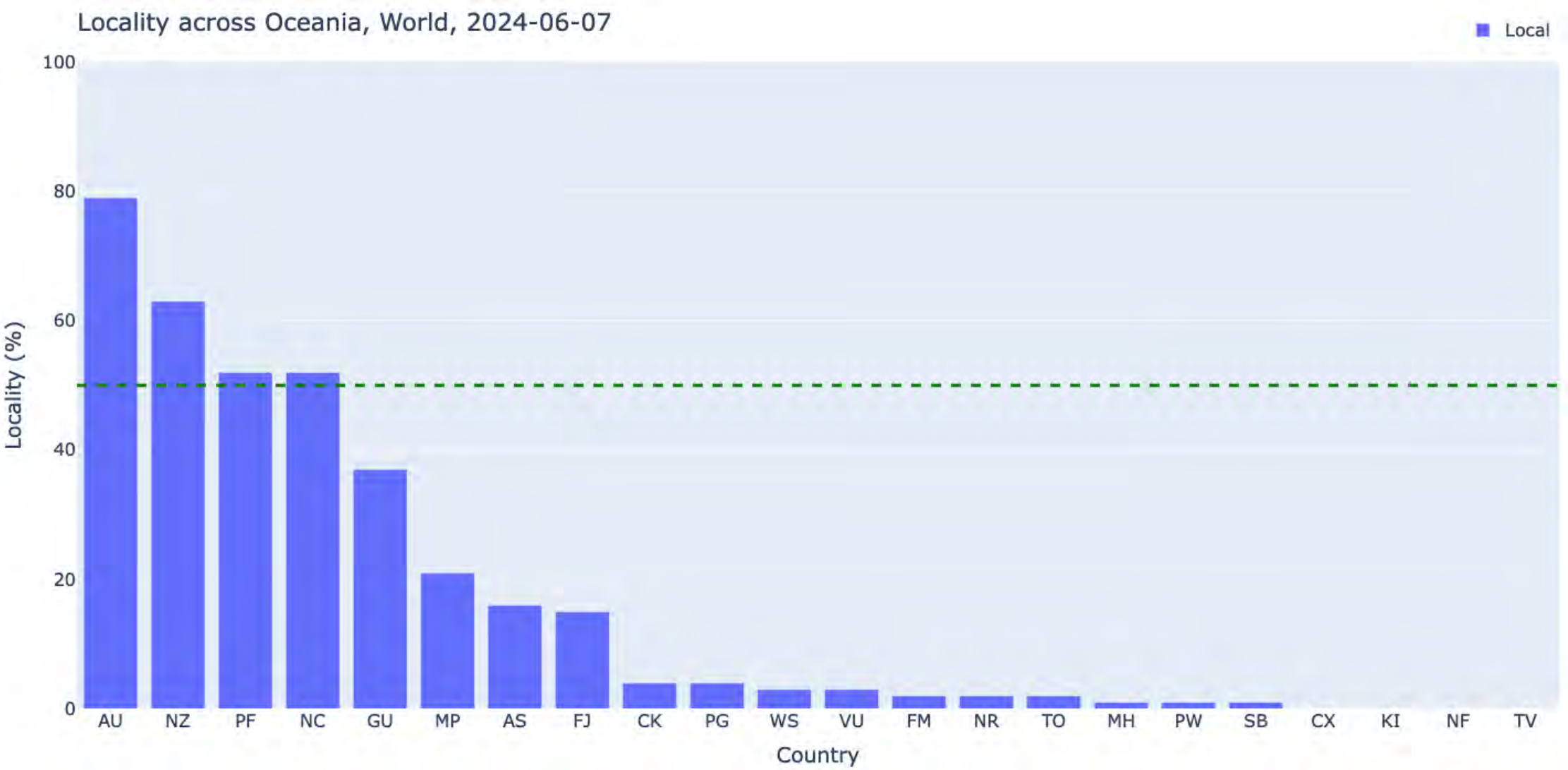
Locality: Asia



Locality: Europe



Locality: Oceania



Previous Works on CDN Deployment

Web Content Cartography

Bernhard Ager
T-Labs/TU Berlin
bernhard@net.t-labs.tu-berlin.de

Georgios Smaragdakis
T-Labs/TU Berlin
georgios@net.t-labs.tu-berlin.de

Wolfgang Mühlbauer
ETU Zurich

CacheLoc: Leveraging CDN Edge Servers for User Geolocation

Mingkui Wei¹[0000-0003-3606-3428], Khaled Rabieh²[0000-0003-2828-6971], and Faisal Kaleem²[0000-0001-6780-1759]

¹ Cyber Forensics Intelligent Center, Computer Science, Sam Houston State University, Huntsville, TX
² Computer Science and Cybersecurity, Metropolitan State University, Saint Paul, MN

The Central Problem with Distributed Content

Common CDN Deployments Centralize Traffic In A Risky Way

Kevin Vermeulen
TASC/ONDC

Loqman Salamatian
Columbia University

Sang Hoon Kim
Columbia University

Ethan Katz-Bassett
Columbia University

Internet development in hosting and distribu

Enrico Calandro¹, Josiah Chavula², and Amreesh Phokeer³

¹ Research ICT Africa, Cape Town, South Africa
ecalandro@researchictafrica.net

² University of Cape Town, Cape Town, South Africa
jchavula@cs.uct.ac.za

³ AFRINIC, Ebene, Mauritius
amreesh@afrrinic.net

Peering vs. Transit: Performance Comparison of Peering and Transit Interconnections

Adnan Ahmed and Zubair Shafiq
The University of Iowa

Harkeerat Bedi and Amir Khakpour
Verizon Digital Media Services



Previous Works on Geolocation



Geobytes

- because everybody's
somewhere

IP2LOCATION



Locating CDN Edge Servers with HTTP Responses

Run Huang, Mengying Zhou, Tiar
Shanghai Key Lab of Intelligent Information Processing, Schc
{runhuang19,myzhou19,tcguo20,ch

Geolocation of IP Hosts in Large Computer Networks with Congestion

Kishan R. Patel

Nadine Moukdad

S. Anand

Towards Geolocation of Millions of IP Addresses*

Computer Science
University,
NY 10458
ukdad@gmail.com

Department of ECE
New York Institute of Technology
New York, NY 10023
Email: asanthan@nyit.edu

Zi Hu John Heidemann Yuri Pradkin
USC/Information Sciences Institute {zihu, johnh, yuri}@isi.edu