Internet Society
Pulse

Measuring Content Locality

James I. Madeley

Pulse Research Review
11th Dec 2024

About Me

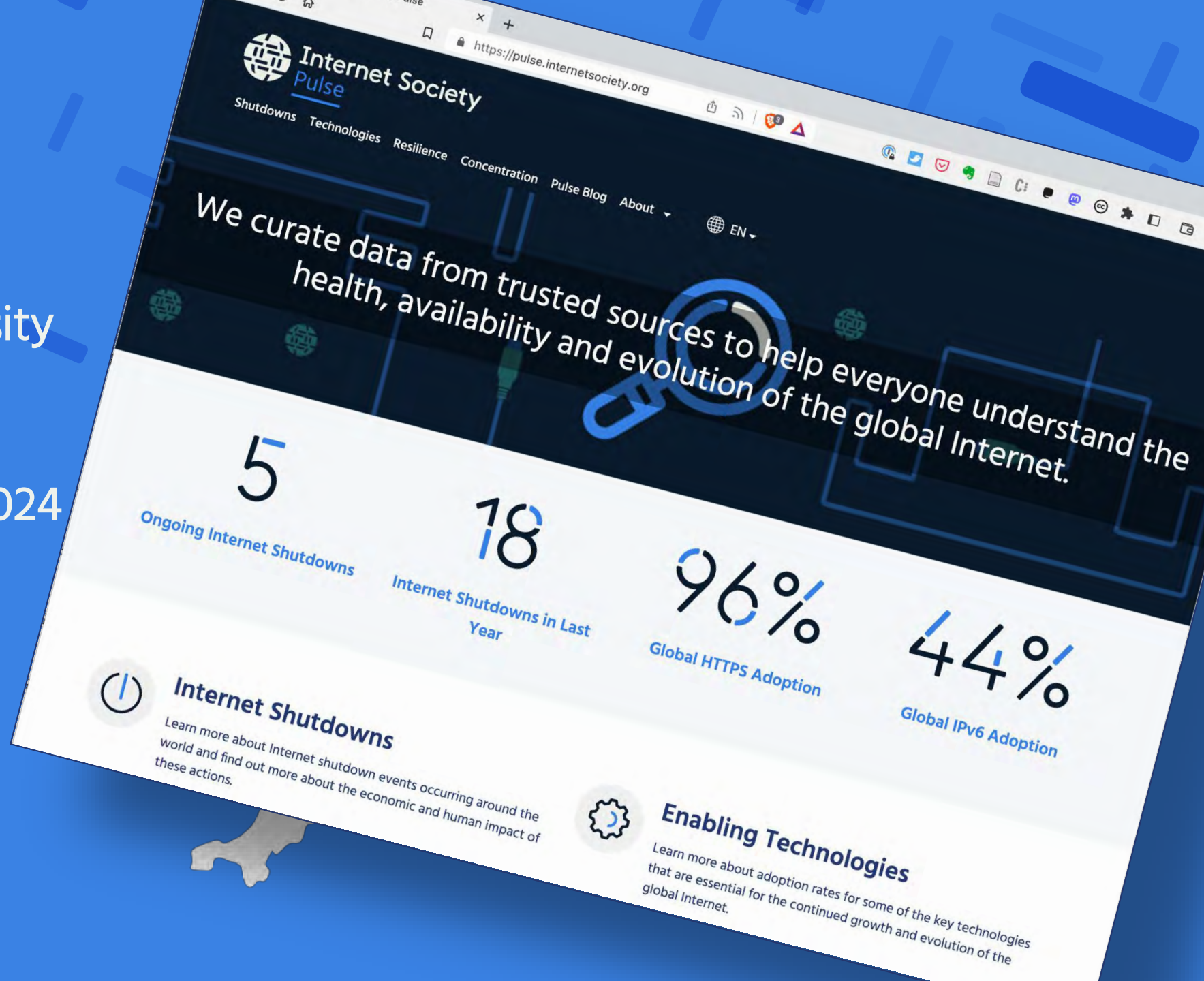
About Me

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

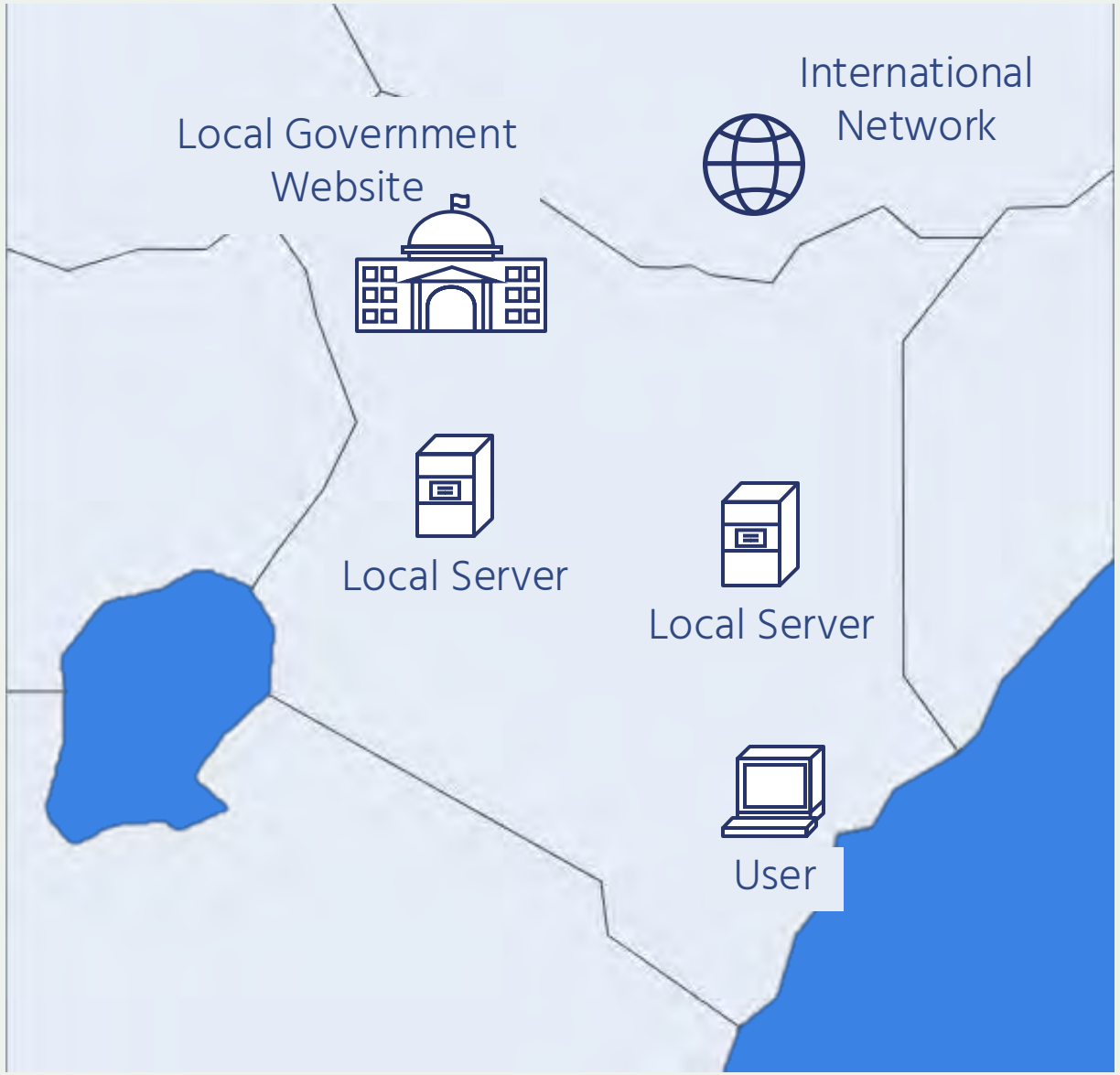
Goals

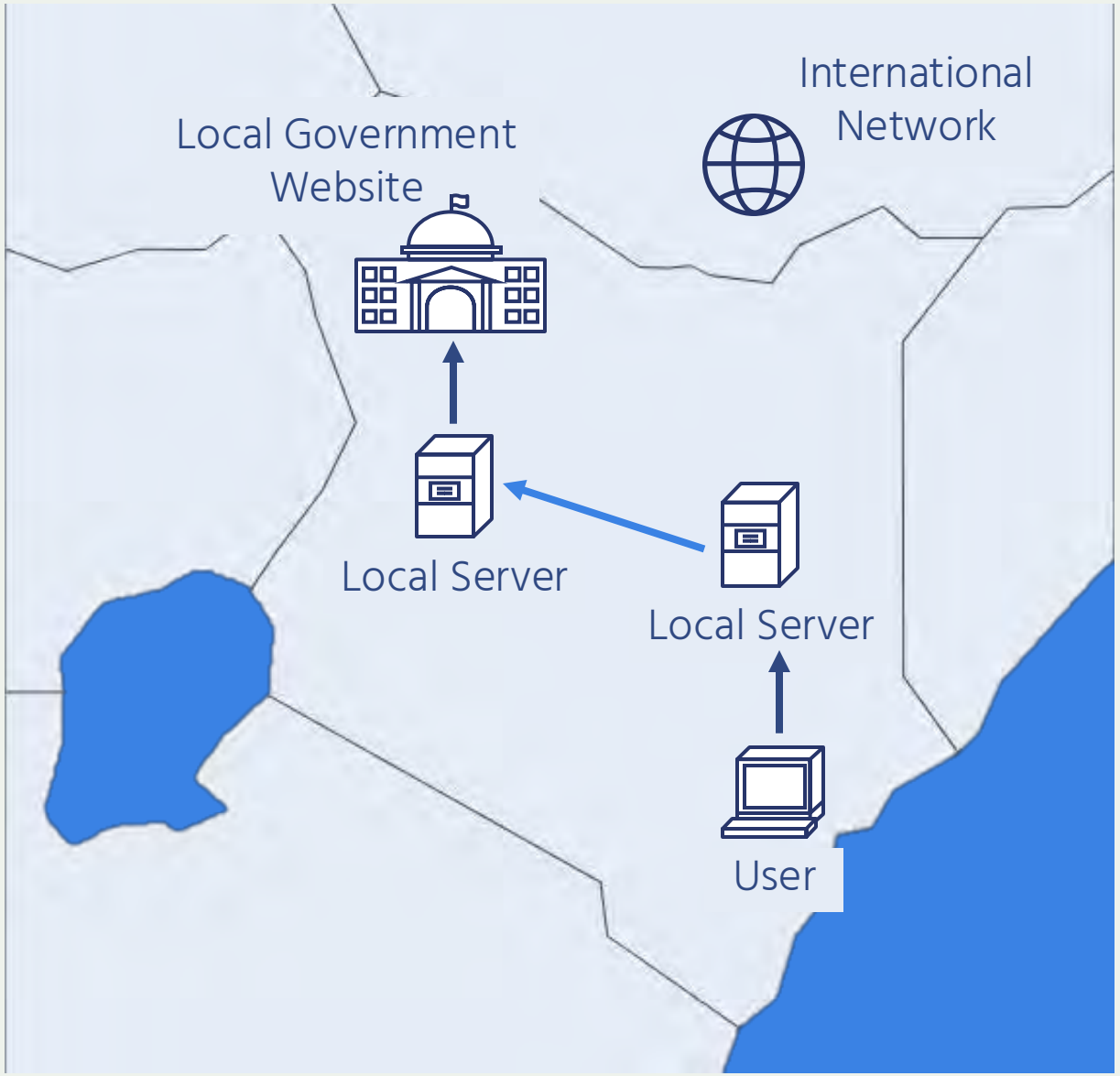
50% local traffic in selected economies by 2025

50/50 Vision

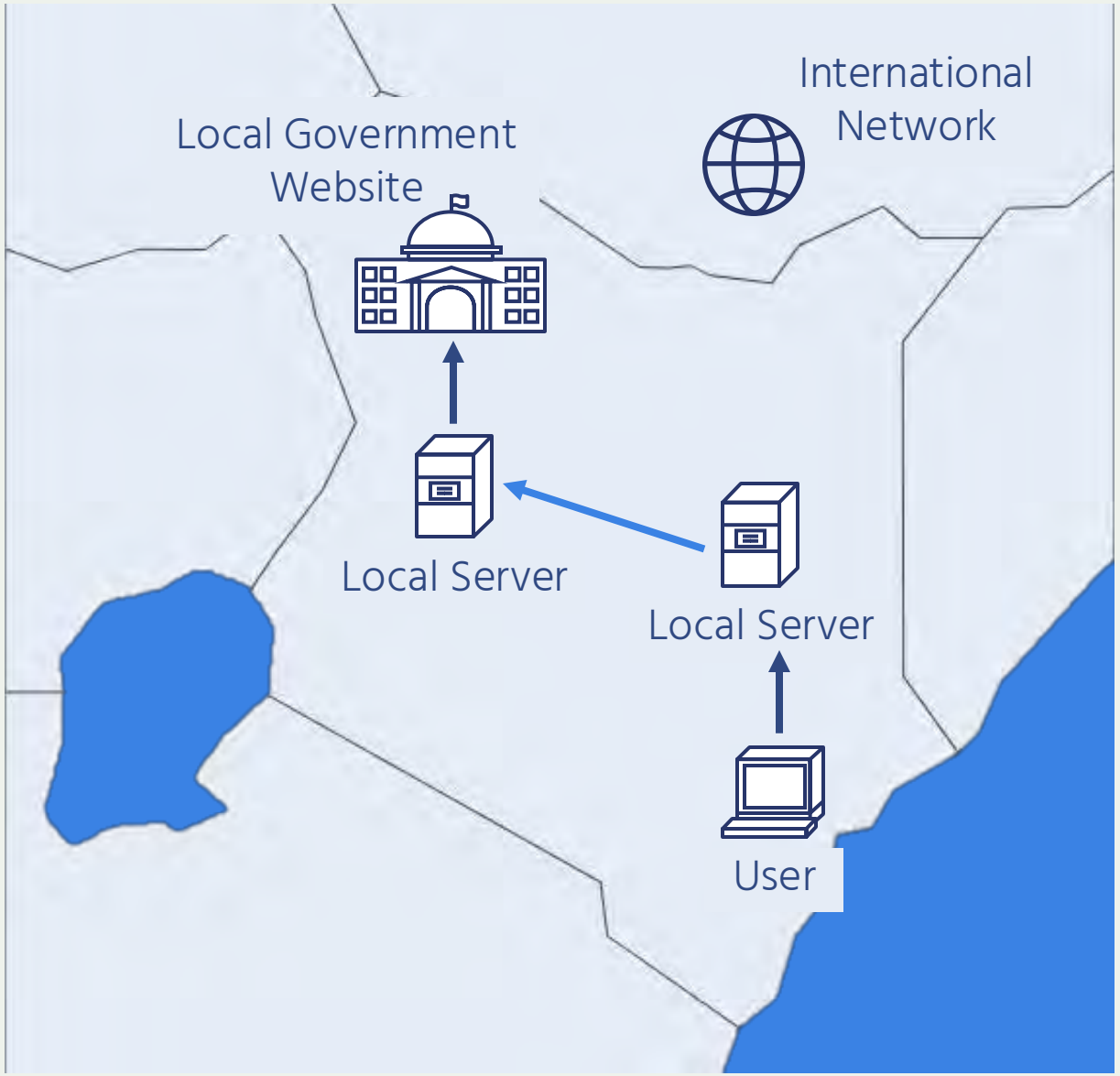
Measurement is key

Locality Definitions

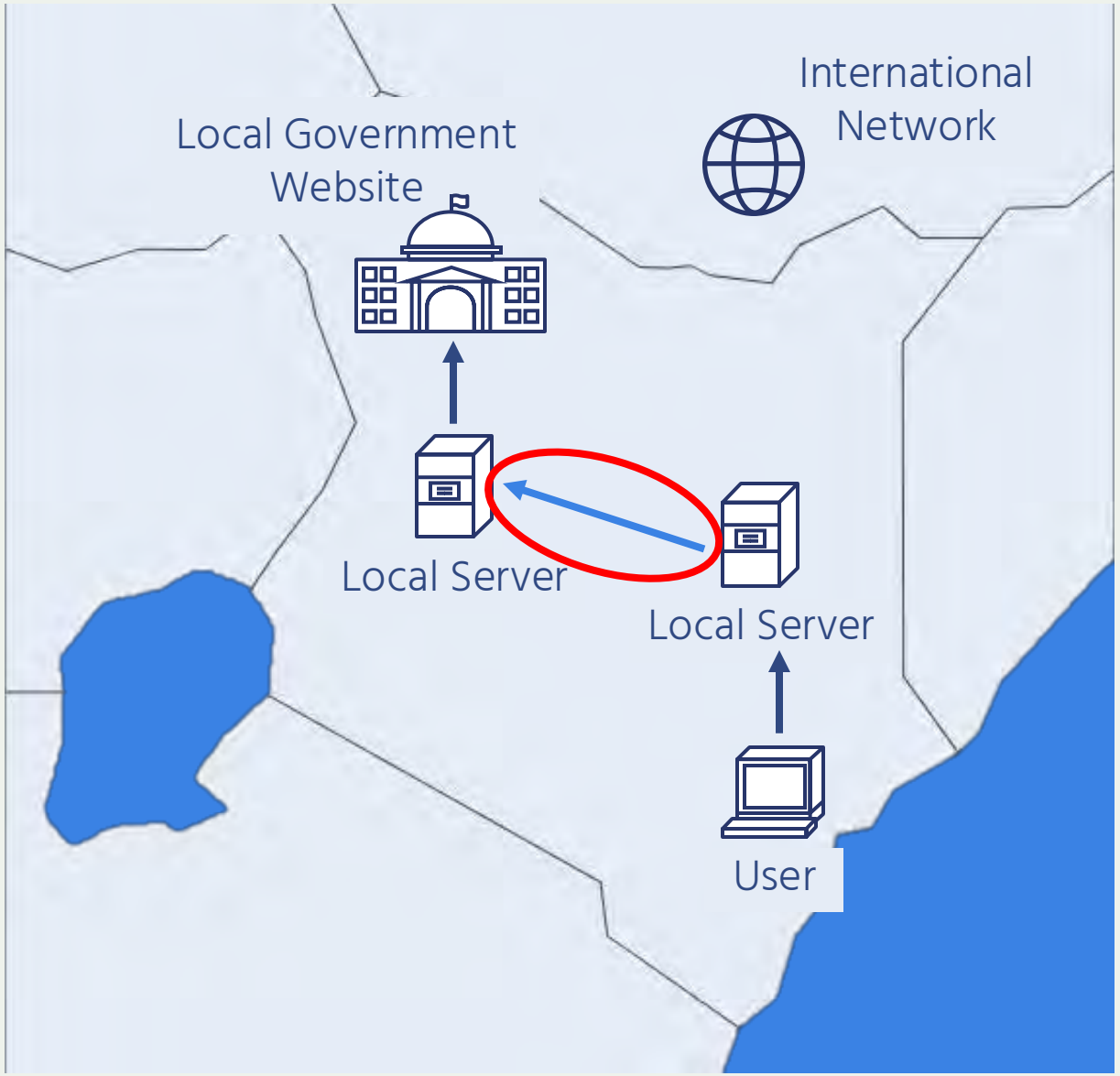

International Network
Local Government Website
Local Server
Local Server
User

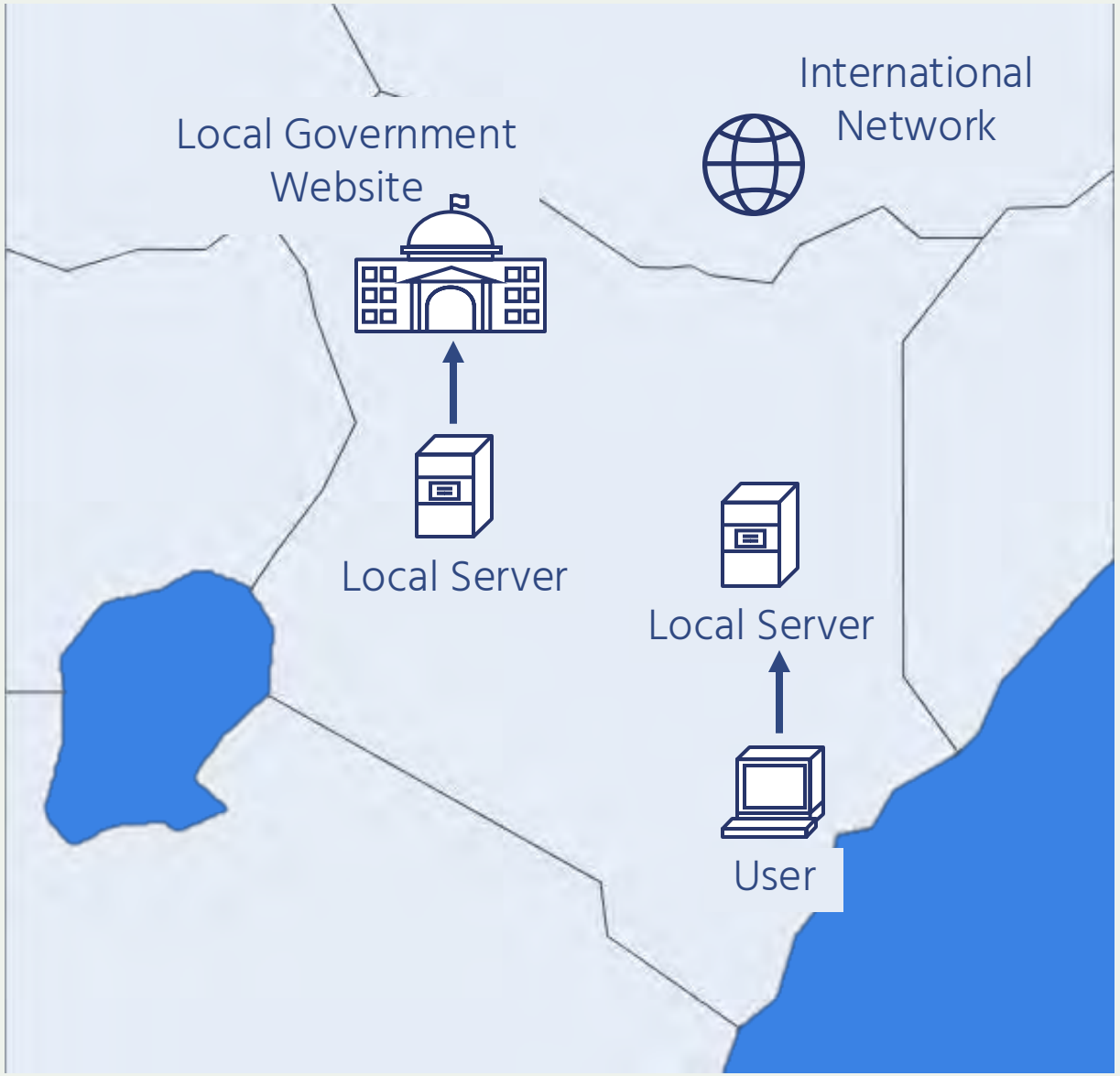

International Network
Local Government Website
Local Server
Local Server
User

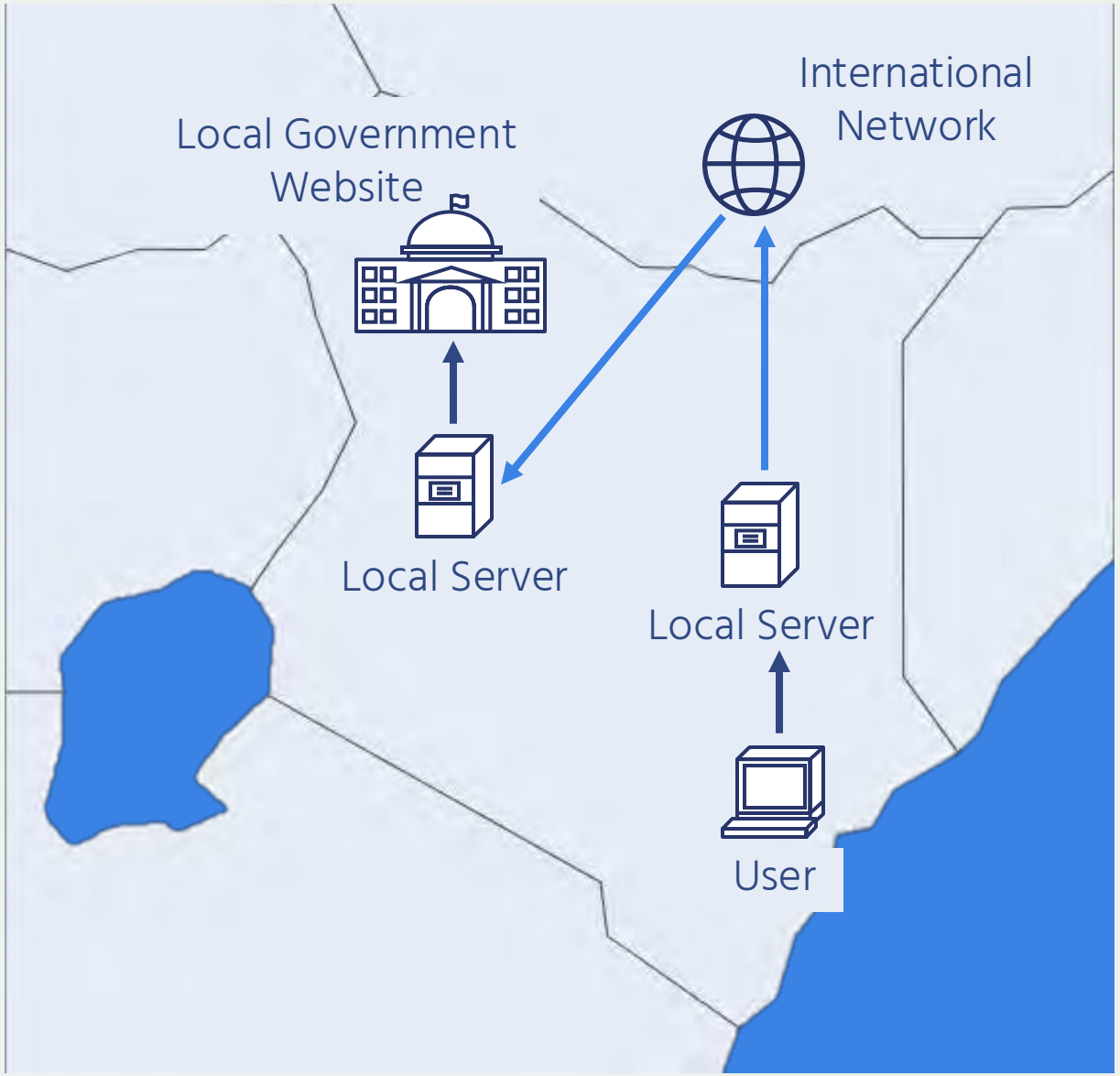
Local Traffic



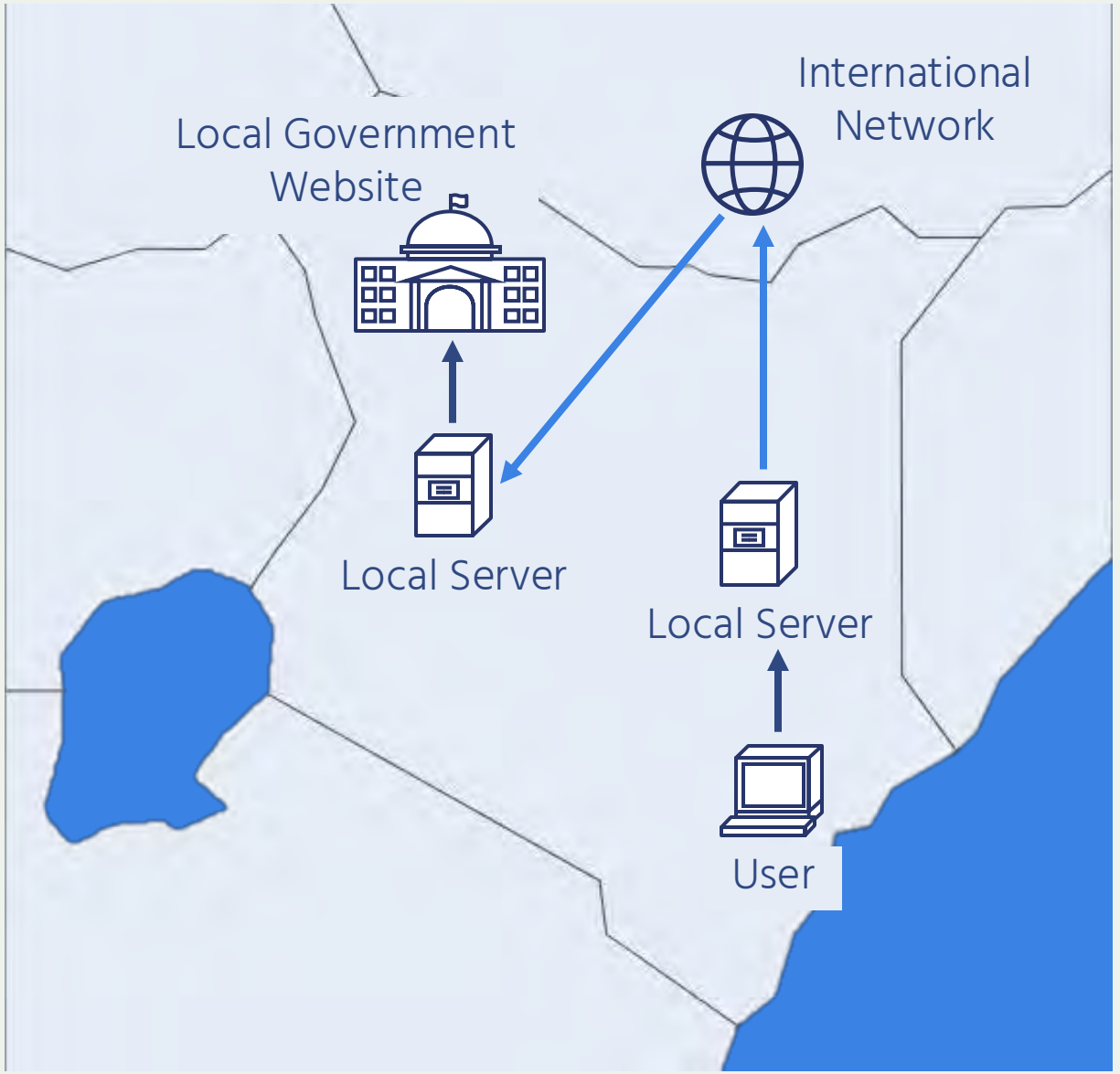
Local Traffic




International Network
Local Government Website
Local Server
Local Server
User


International Network
Local Government Website
Local Server
Local Server
User

External Traffic



Motivation

Motivation

Motivation

Policy

Performance

Persistence
(availability)

Motivation

Policy	Performance	Persistence (availability)
South Africa's POPI Act / EU's GDPR		
Internet Society's 50/50 Vision		

Motivation

Policy	Performance	Persistence (availability)
South Africa's POPI Act / EU's GDPR	Traveling shorter distances is faster	
Internet Society's 50/50 Vision	Improves UX through cheaper connectivity	

Motivation

Policy

South Africa's POPI Act / EU's GDPR

Internet Society's 50/50 Vision

Performance

Traveling shorter distances is faster

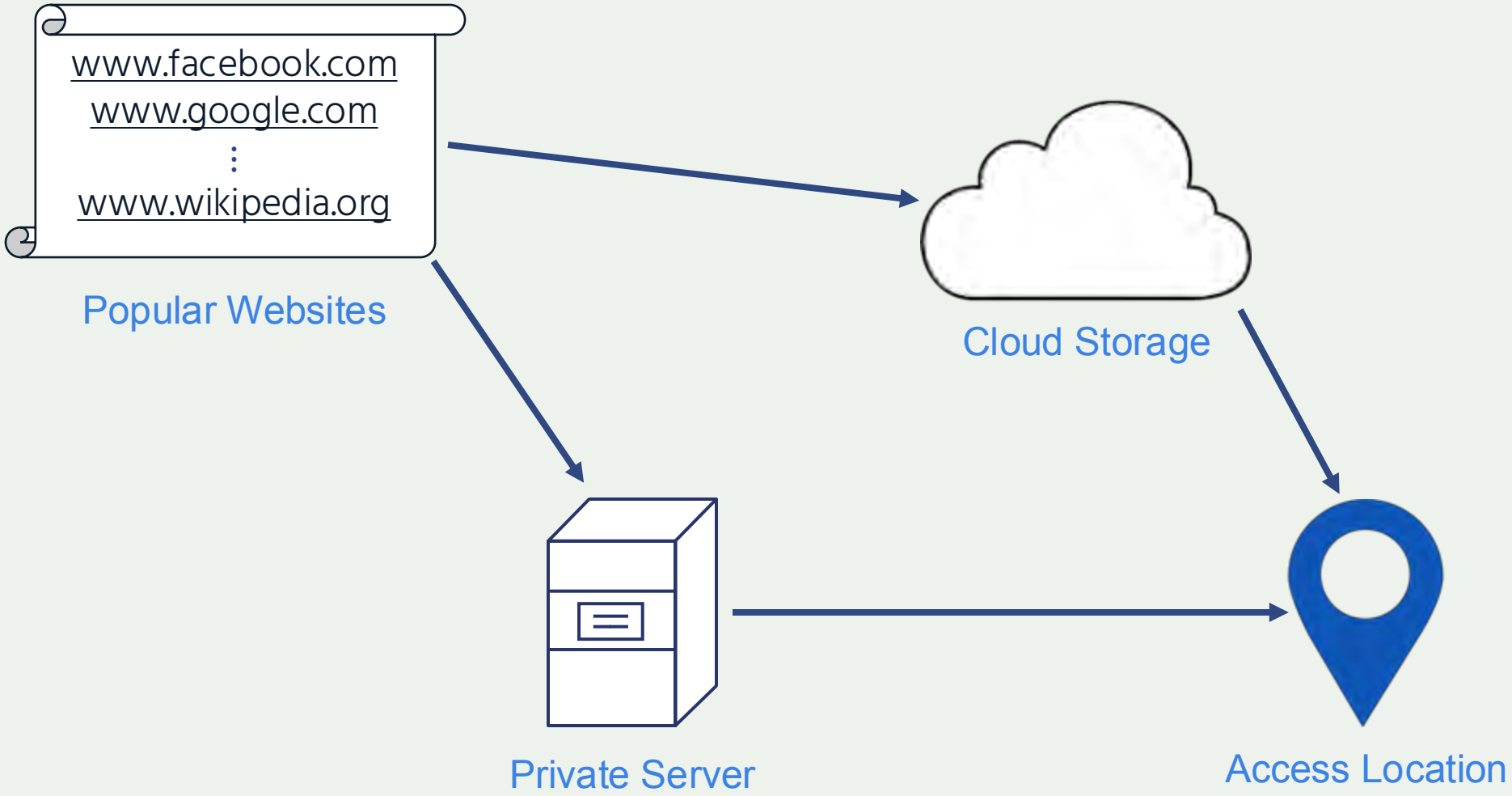
Improves UX through cheaper connectivity

Persistence
(availability)

Resilience against faults e.g. cable cuts

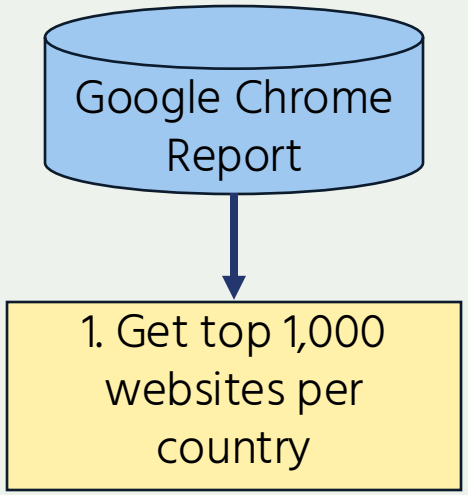
Methodology

Methodology

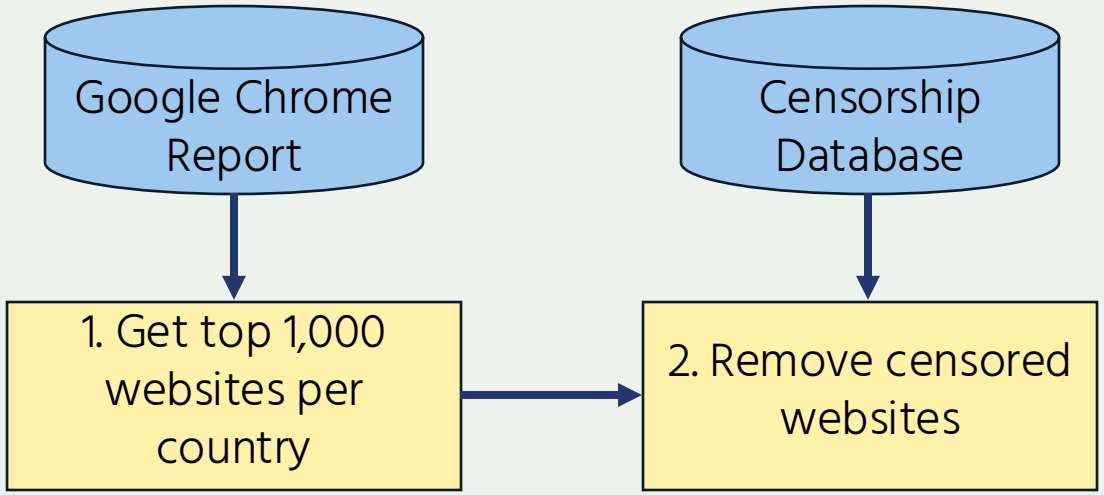


Methodology Steps

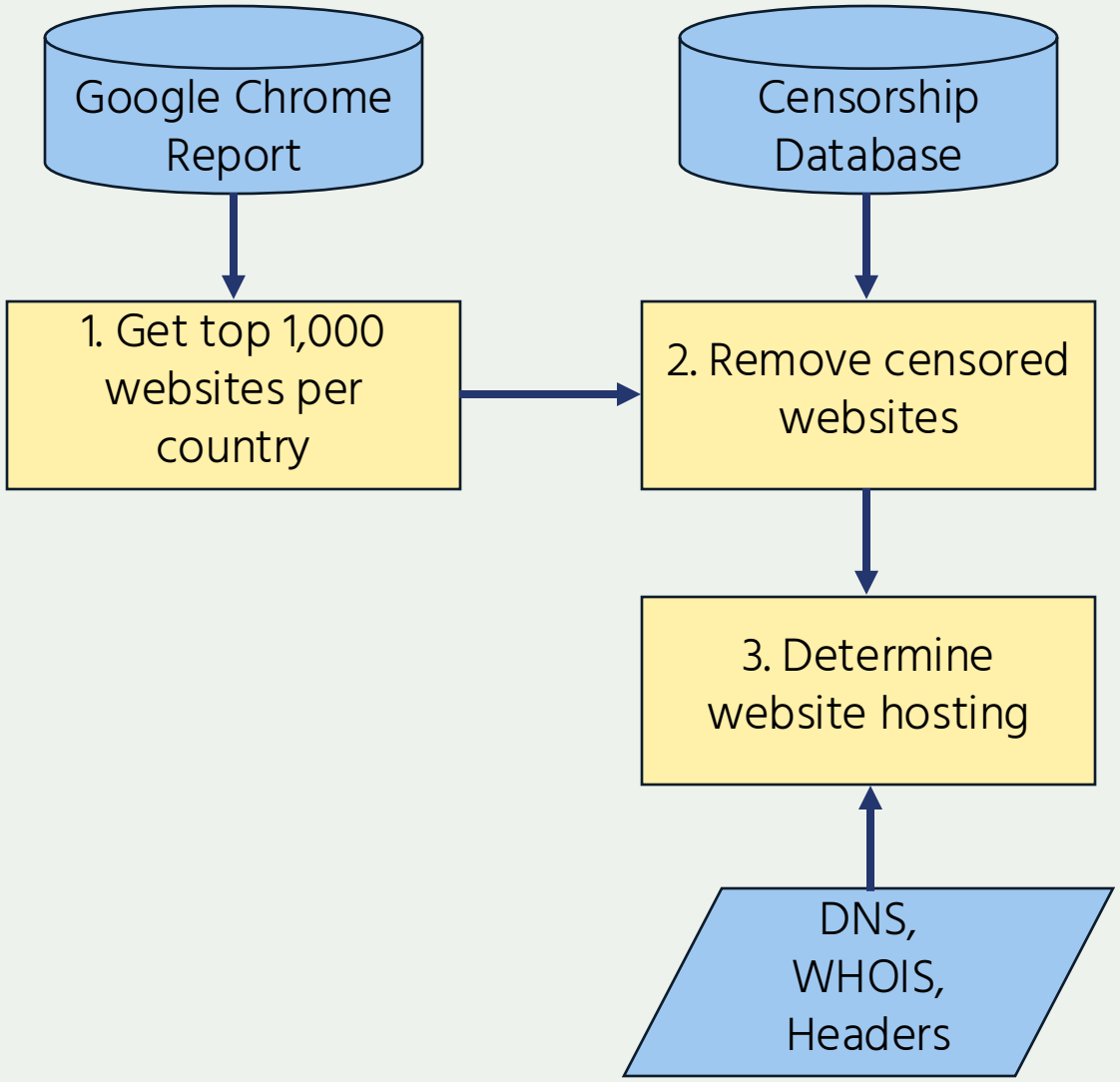
Methodology Steps



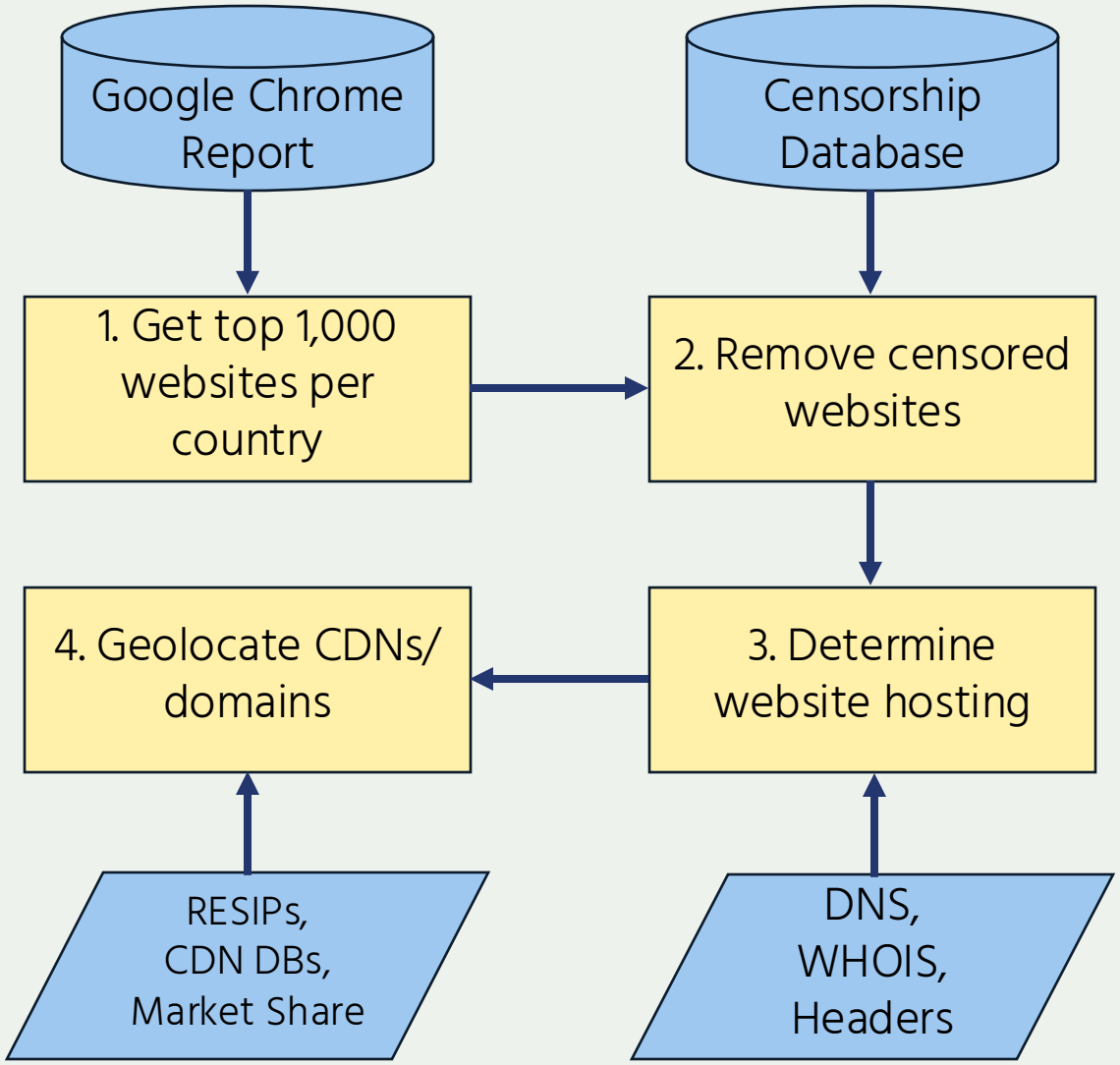
Methodology Steps



Methodology Steps



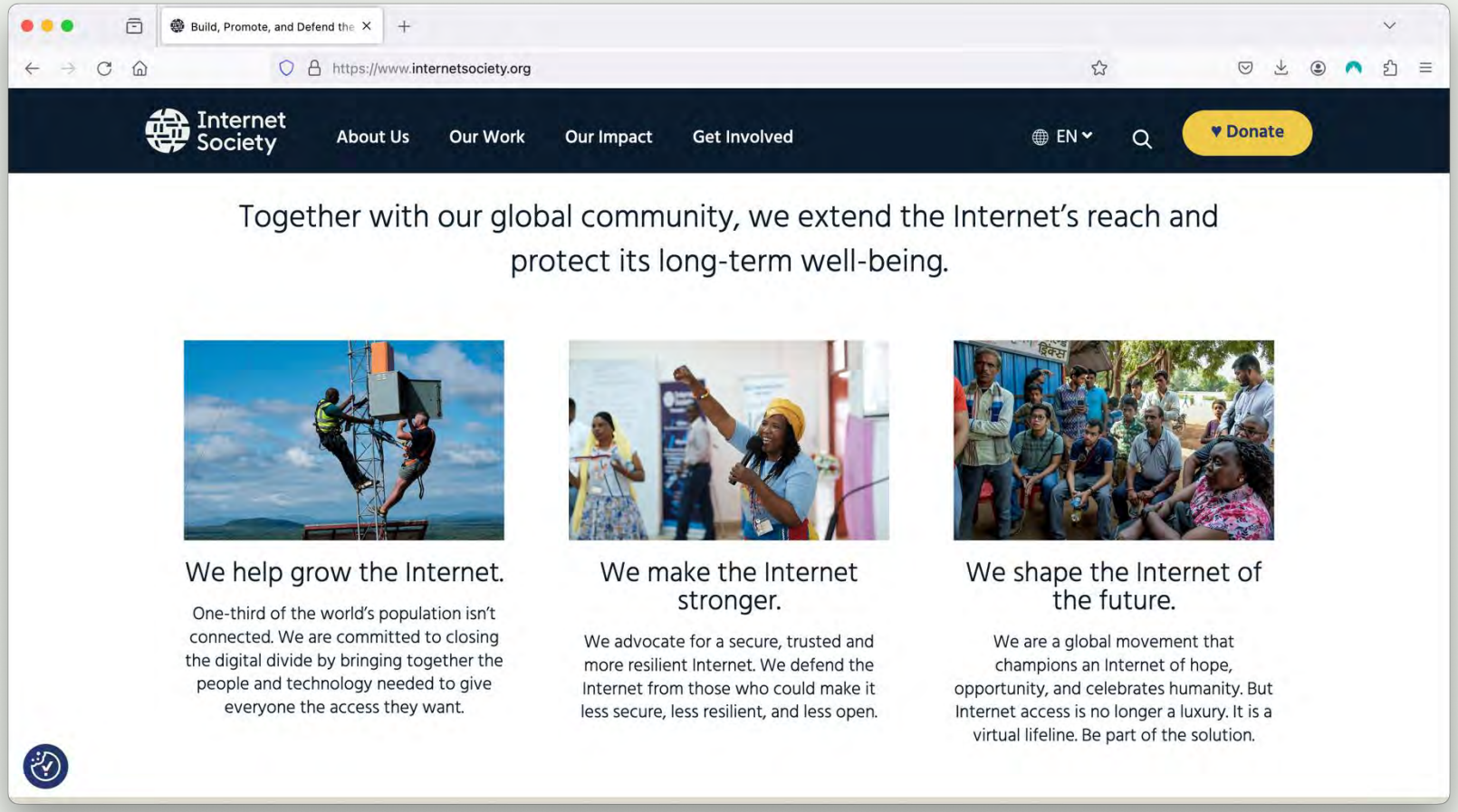
Methodology Steps



Limitations

Limitations

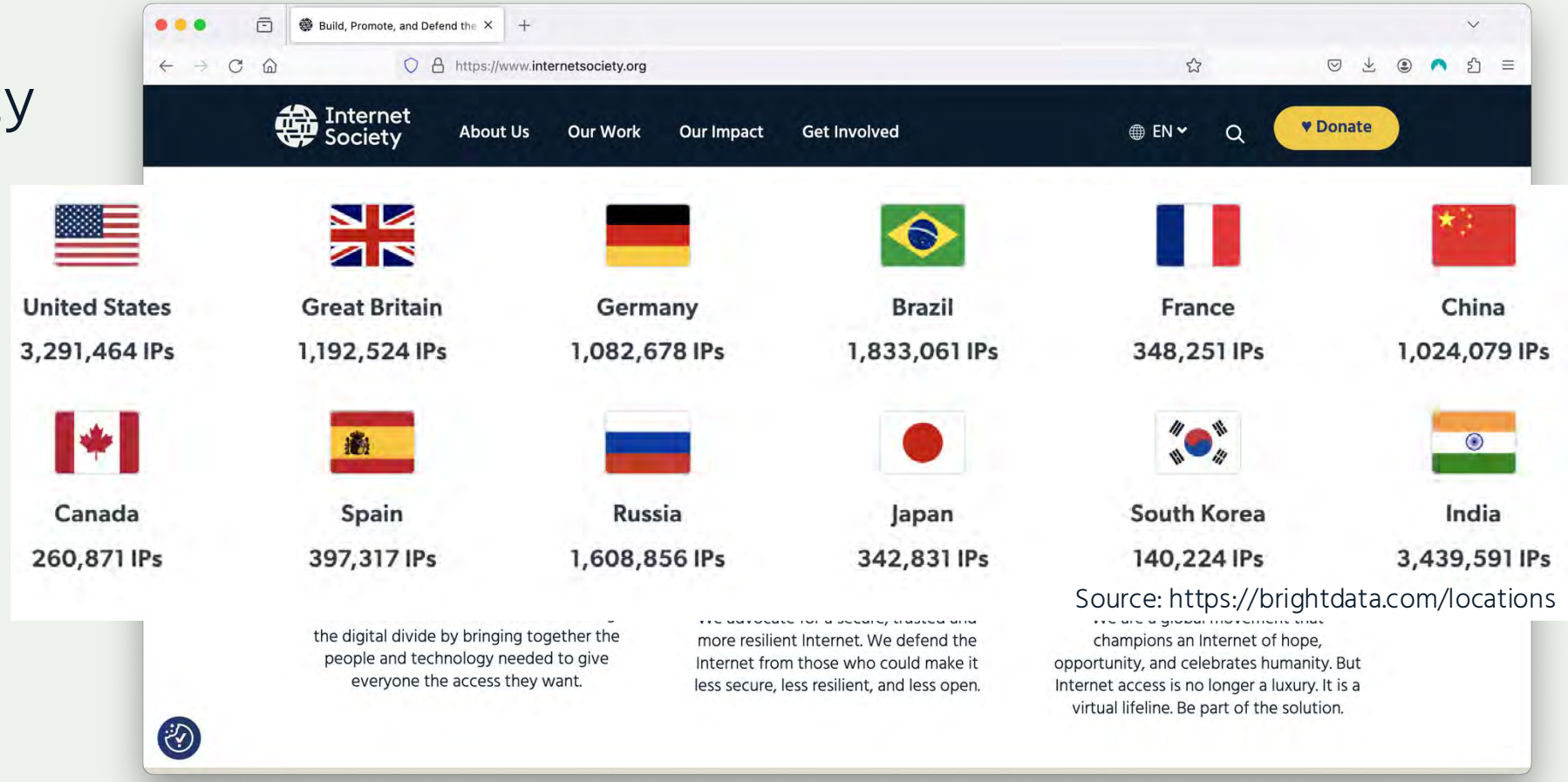
Website Complexity



Limitations

Website Complexity

Proxy Probes



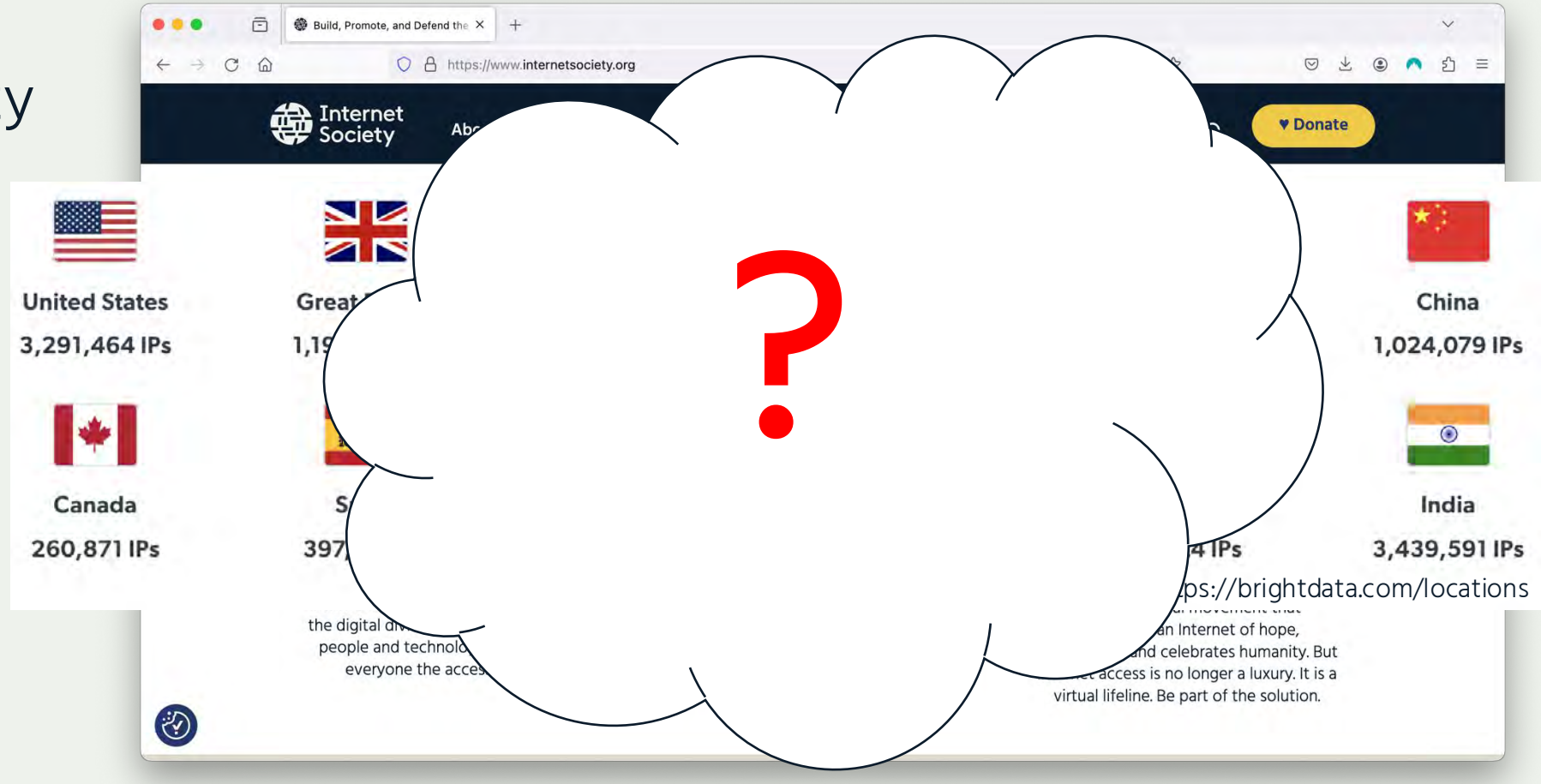
Source: https://brightdata.com/locations

Limitations

Website Complexity

Proxy Probes

Fog of Cloud



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

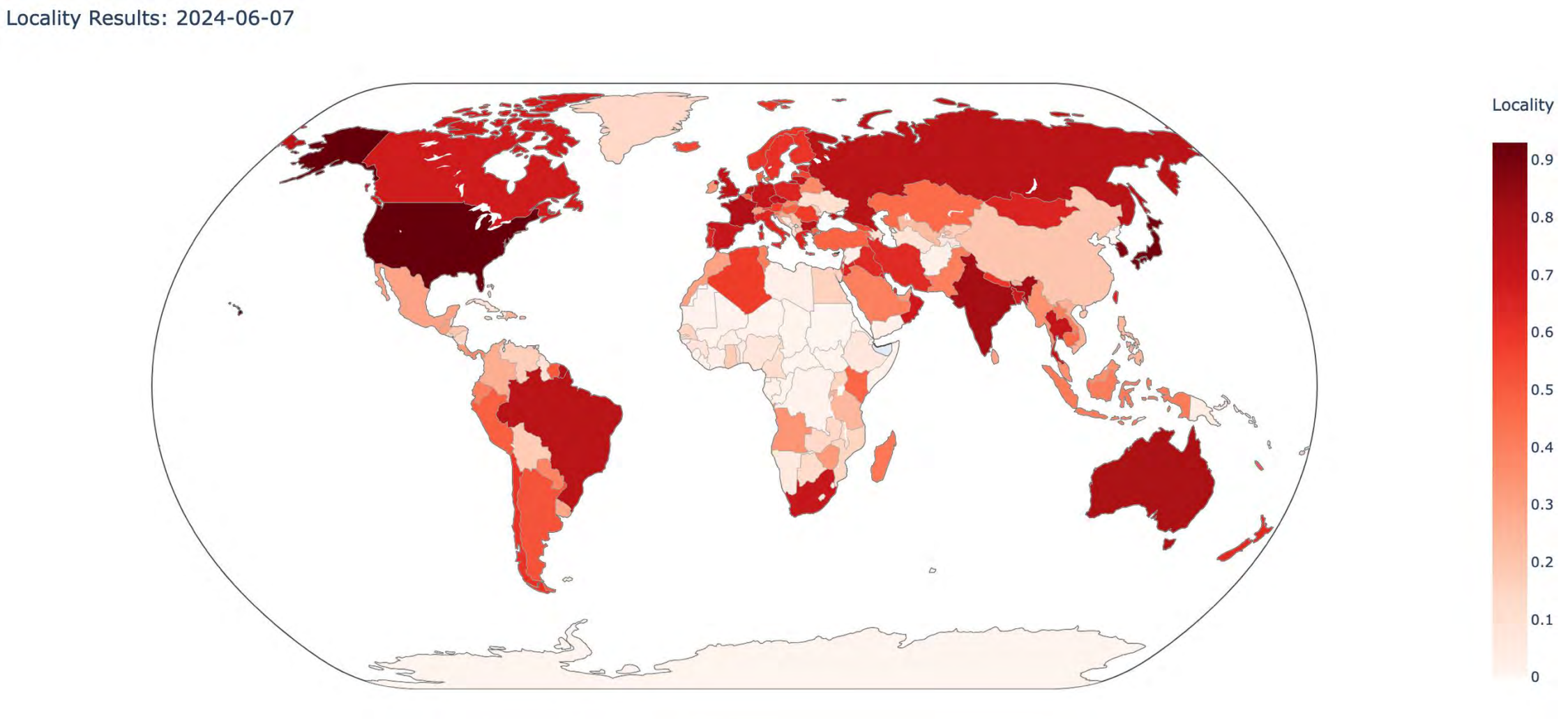
FindCDN – 3.5 days

Our tool – 12 minutes

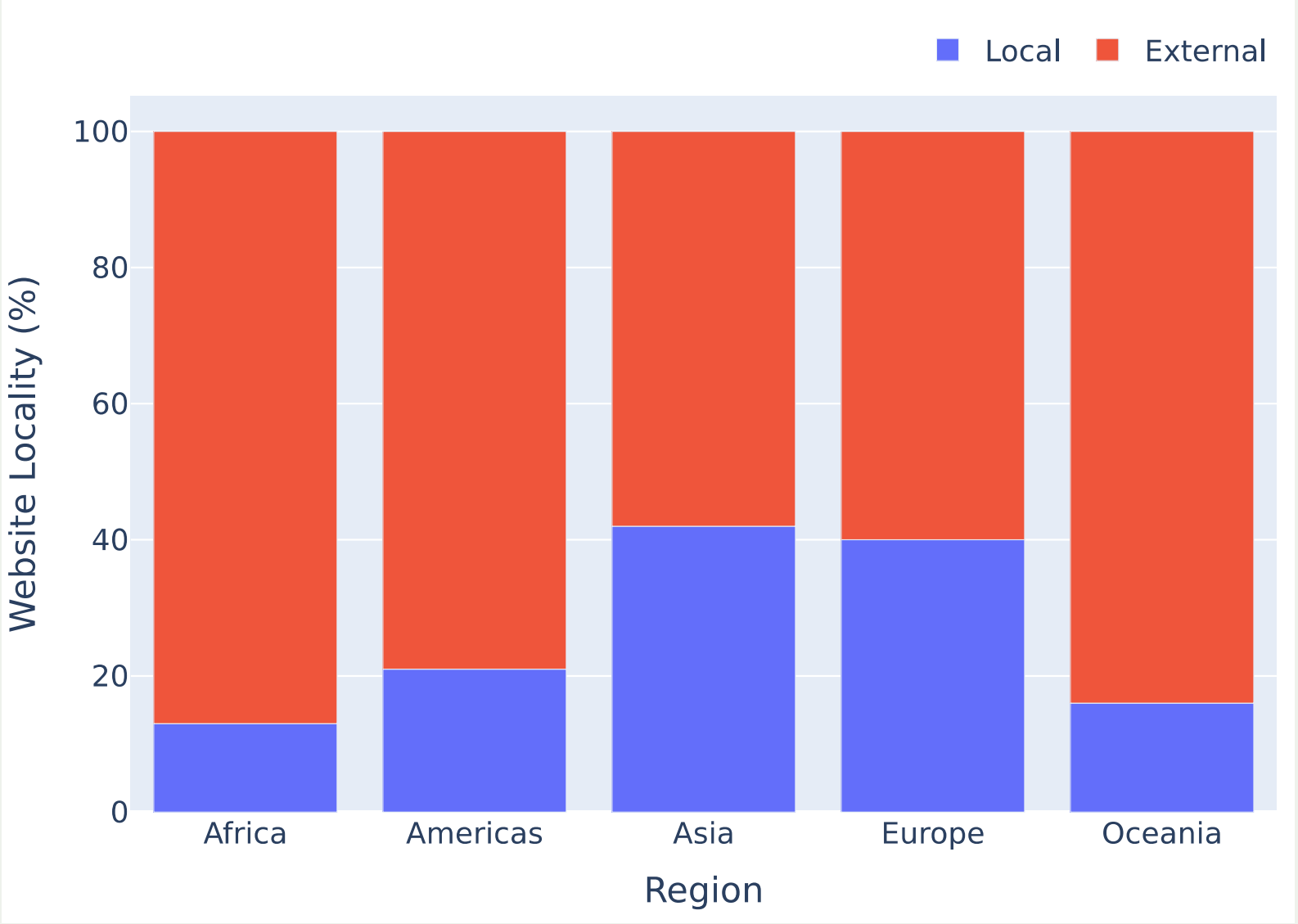
>95% accuracy

Preliminary Locality Results

Global Locality

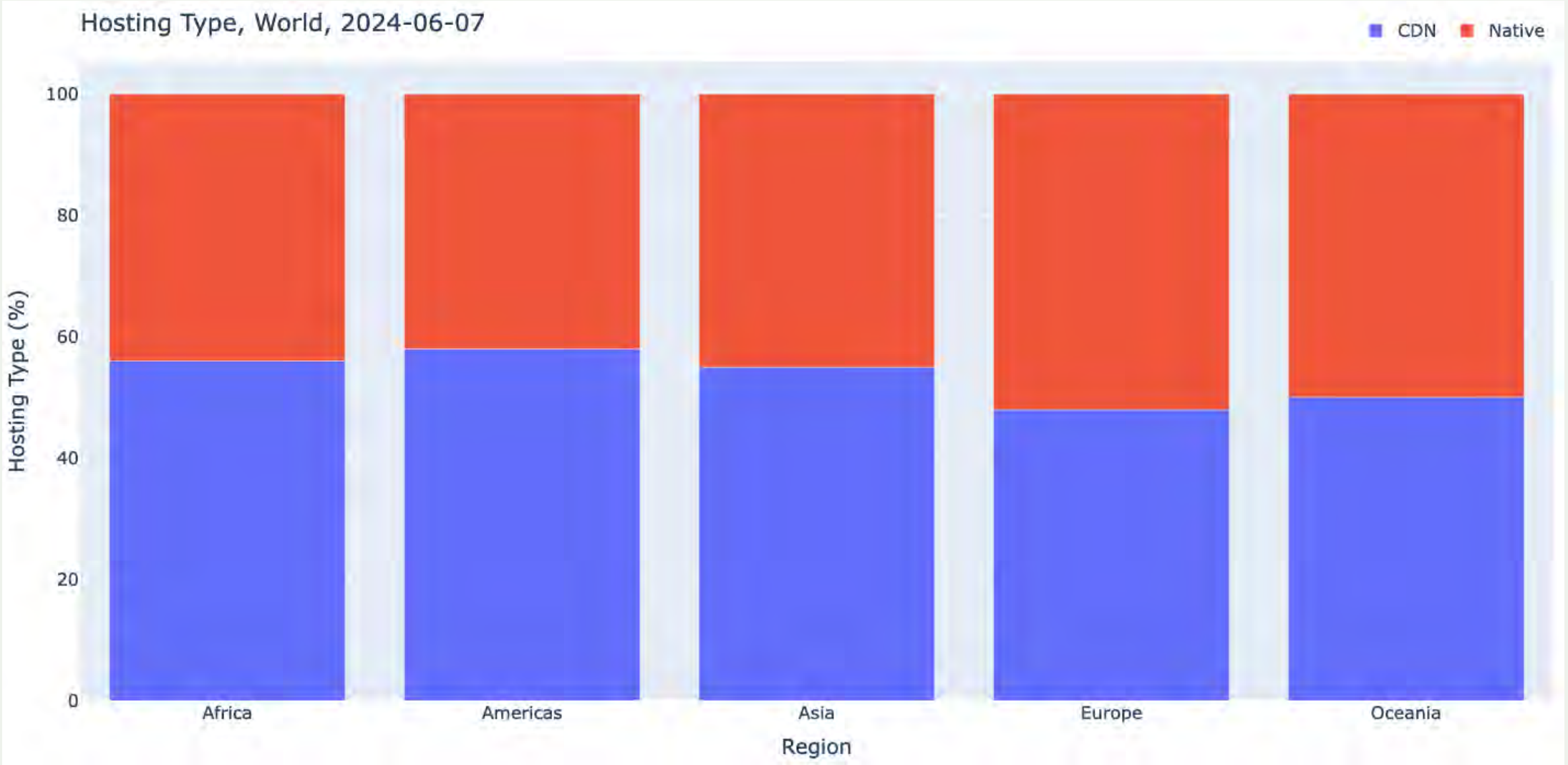


Regional Locality

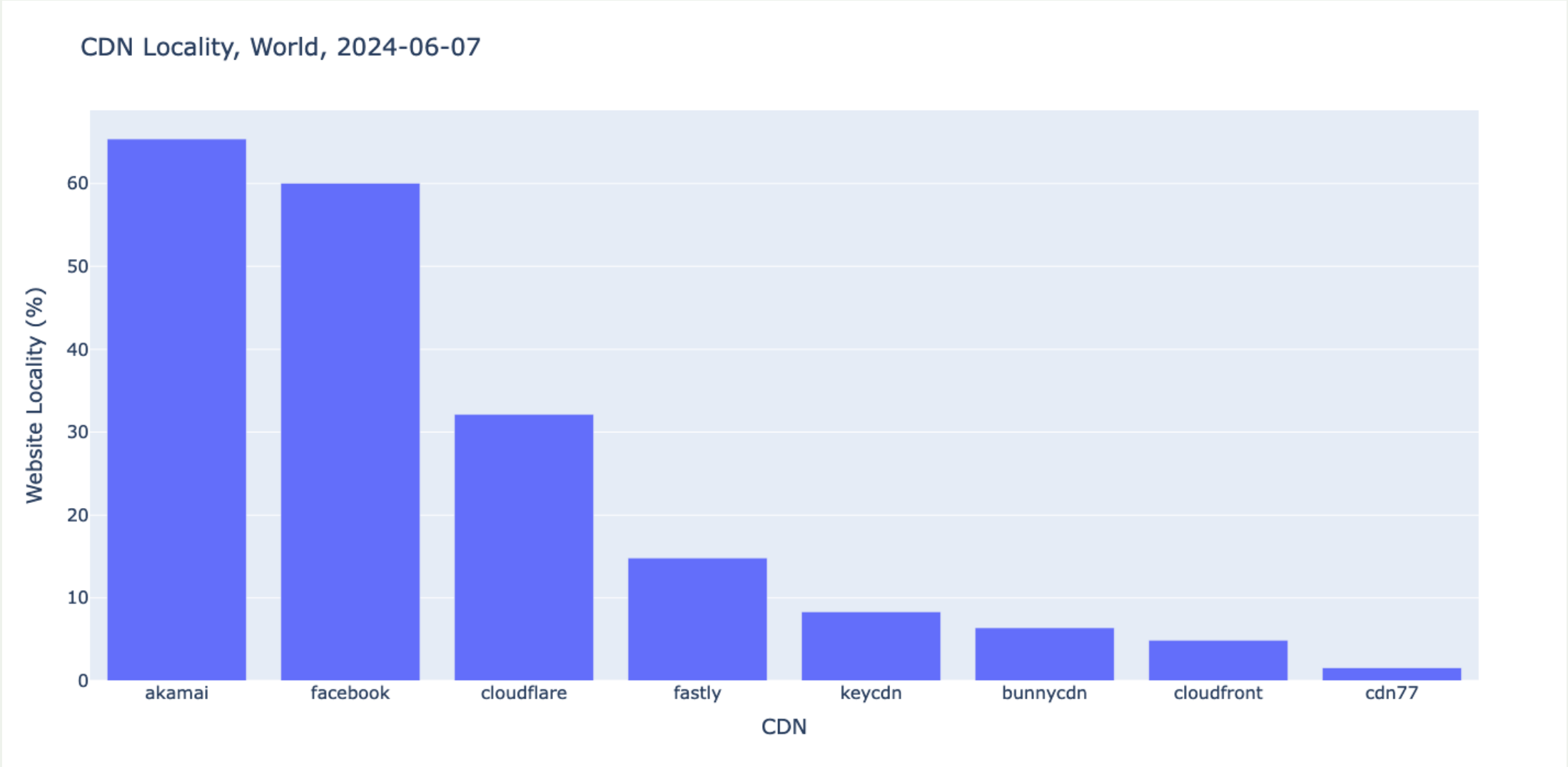


Other Preliminary Results

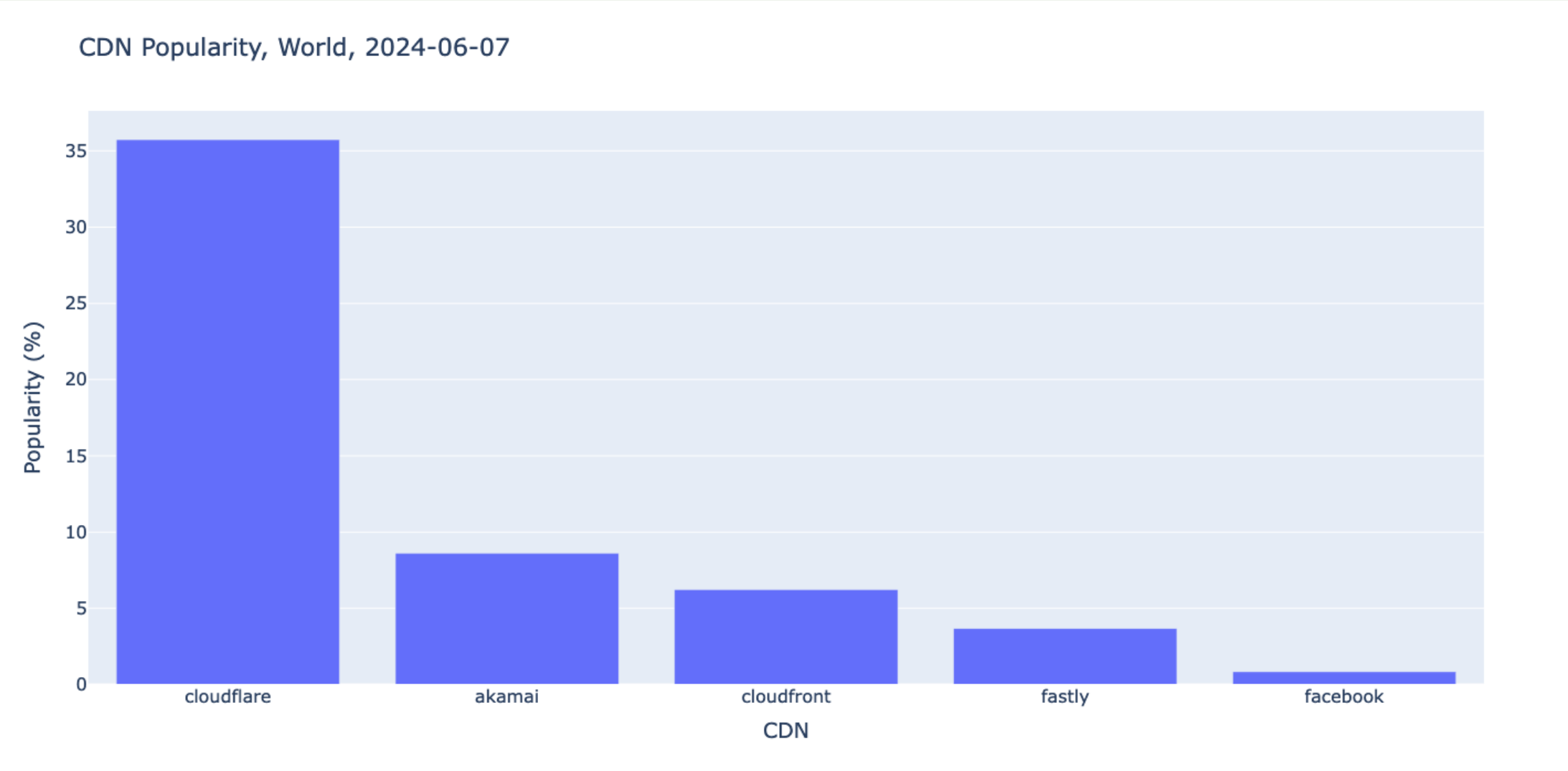
Hosting Type (World)



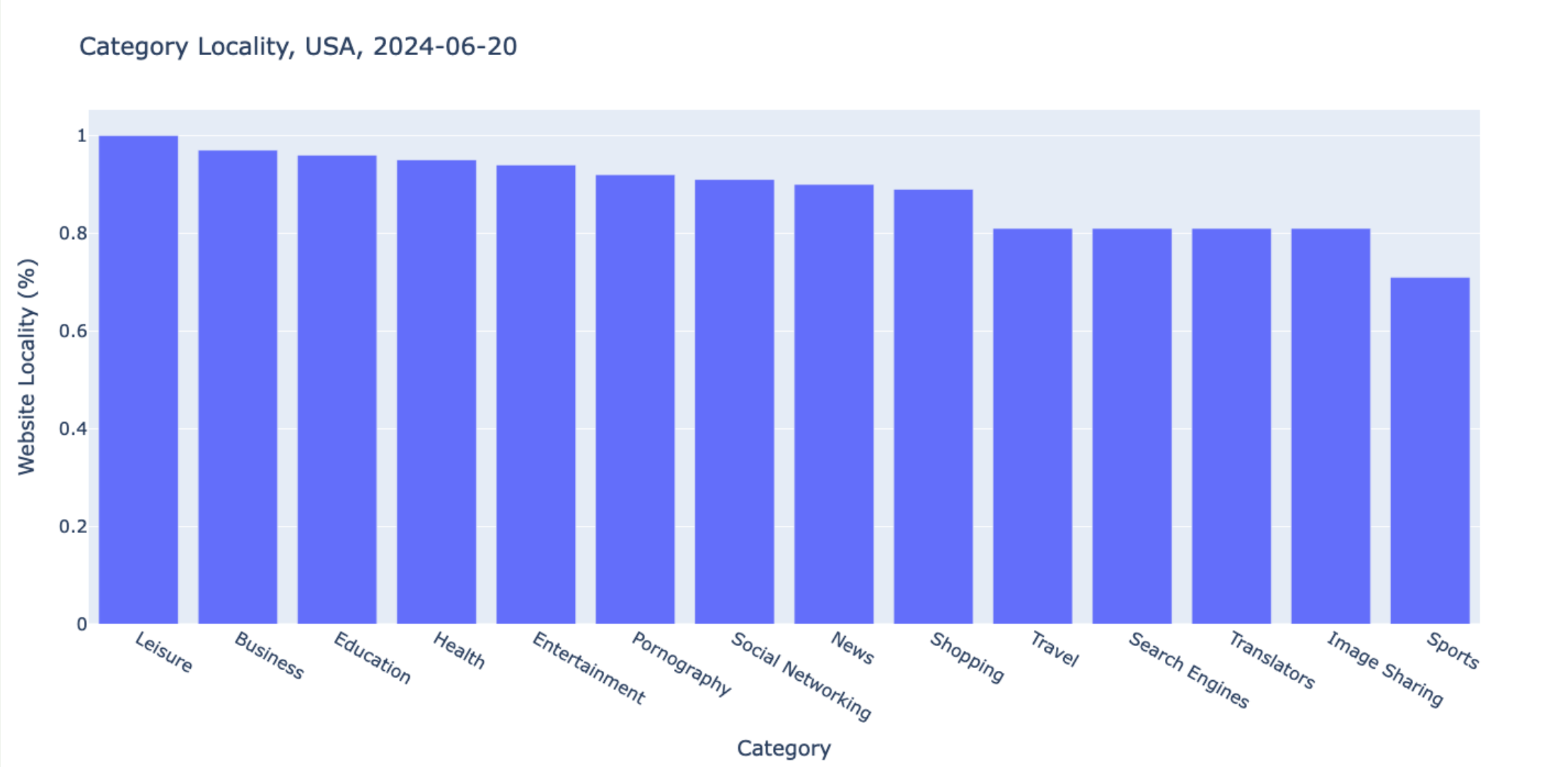
CDN Locality (World)



5 Most Popular CDNs (World)

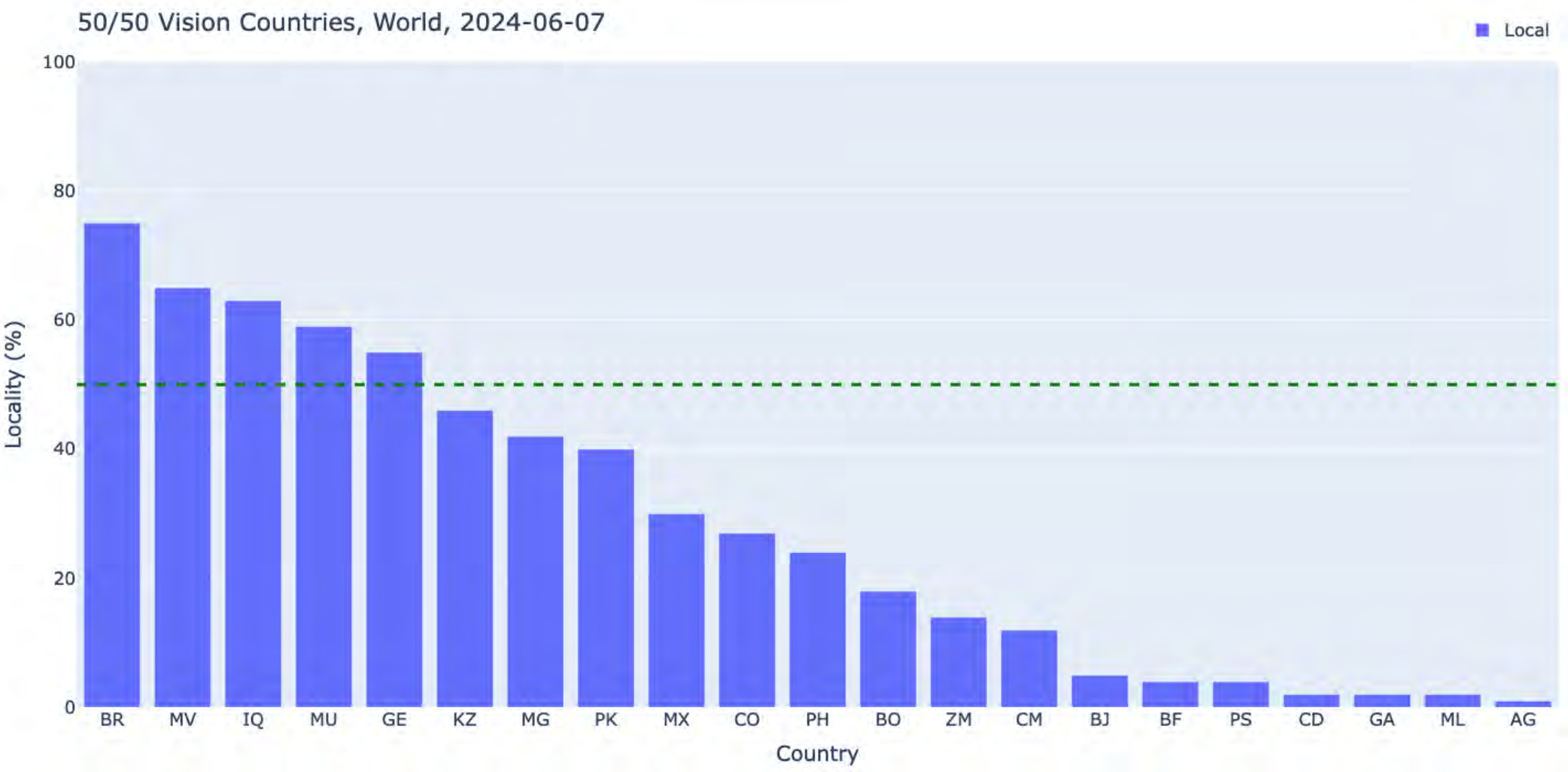


Categories Locality (USA, 100 Sites)

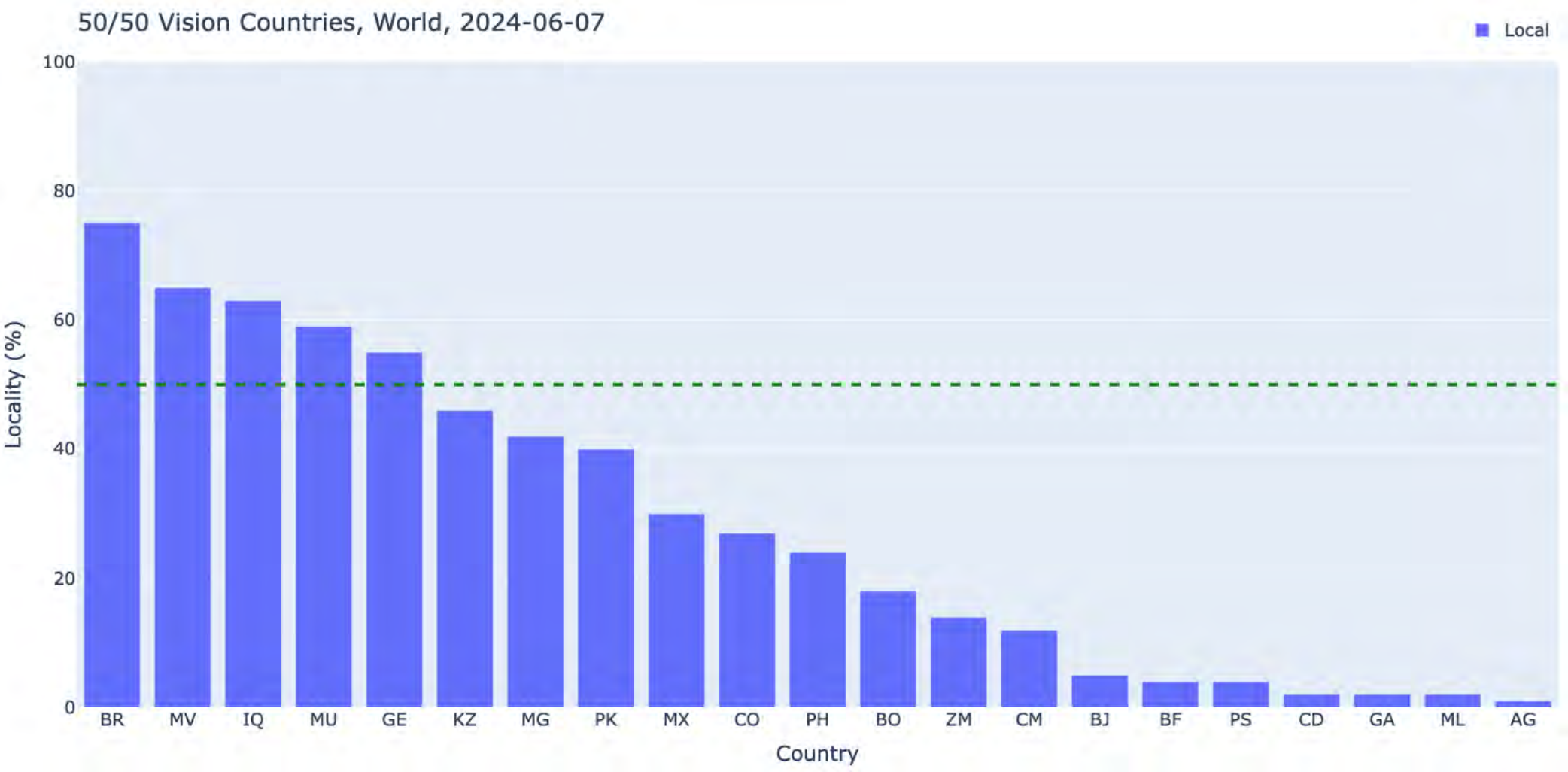


How close to the 50/50 Vision are we?

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%

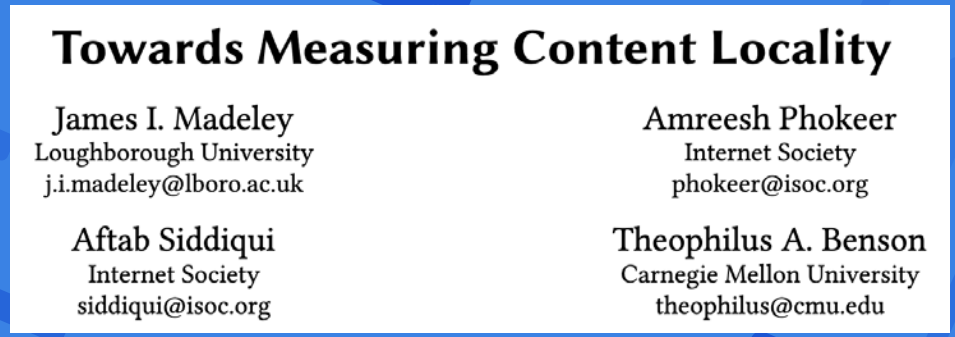
Worldwide 50/50 Vision: 24%

Next Steps

Expand the testing methodology

Continue testing to find changes/patterns

Publish our results in papers and via Pulse platform



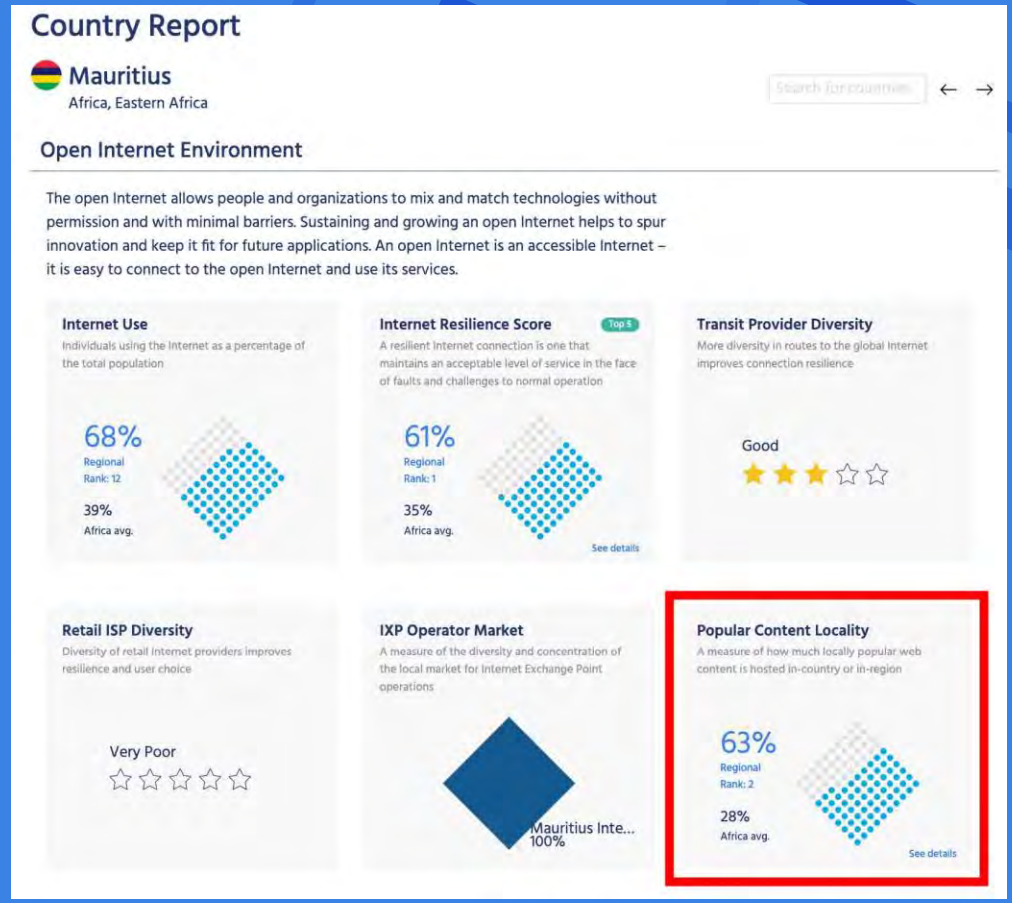
Towards Measuring Content Locality

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

Amreesh Phokeer
Internet Society
phokeer@isoc.org

Aftab Siddiqui
Internet Society
siddiqui@isoc.org

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



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 Buzzone
mbuzzone@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

Reasons for Locality

Traceroute/Latency

Visualisation

Future Work

Recursive Search

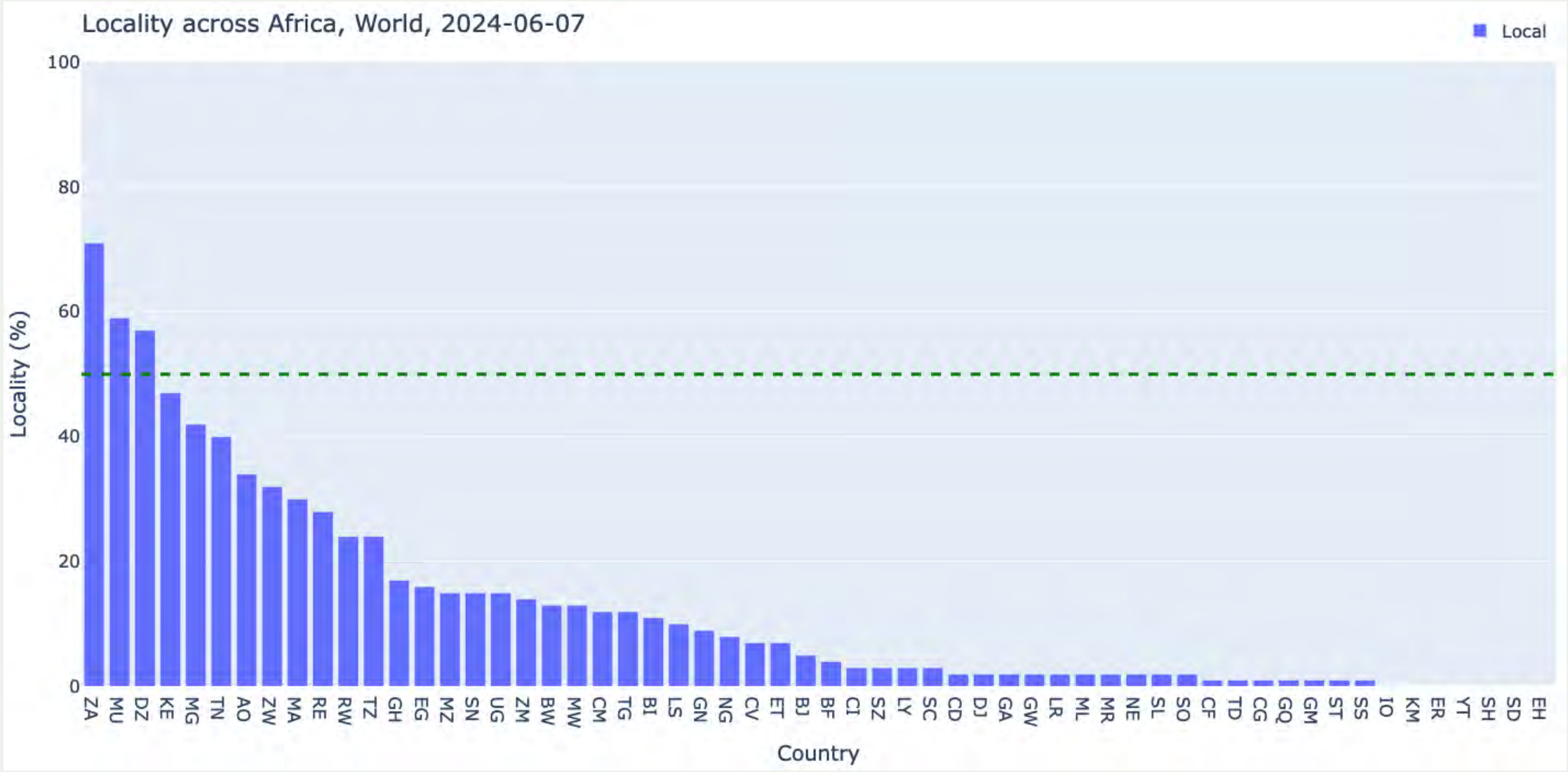
Reasons for Locality

Categorization

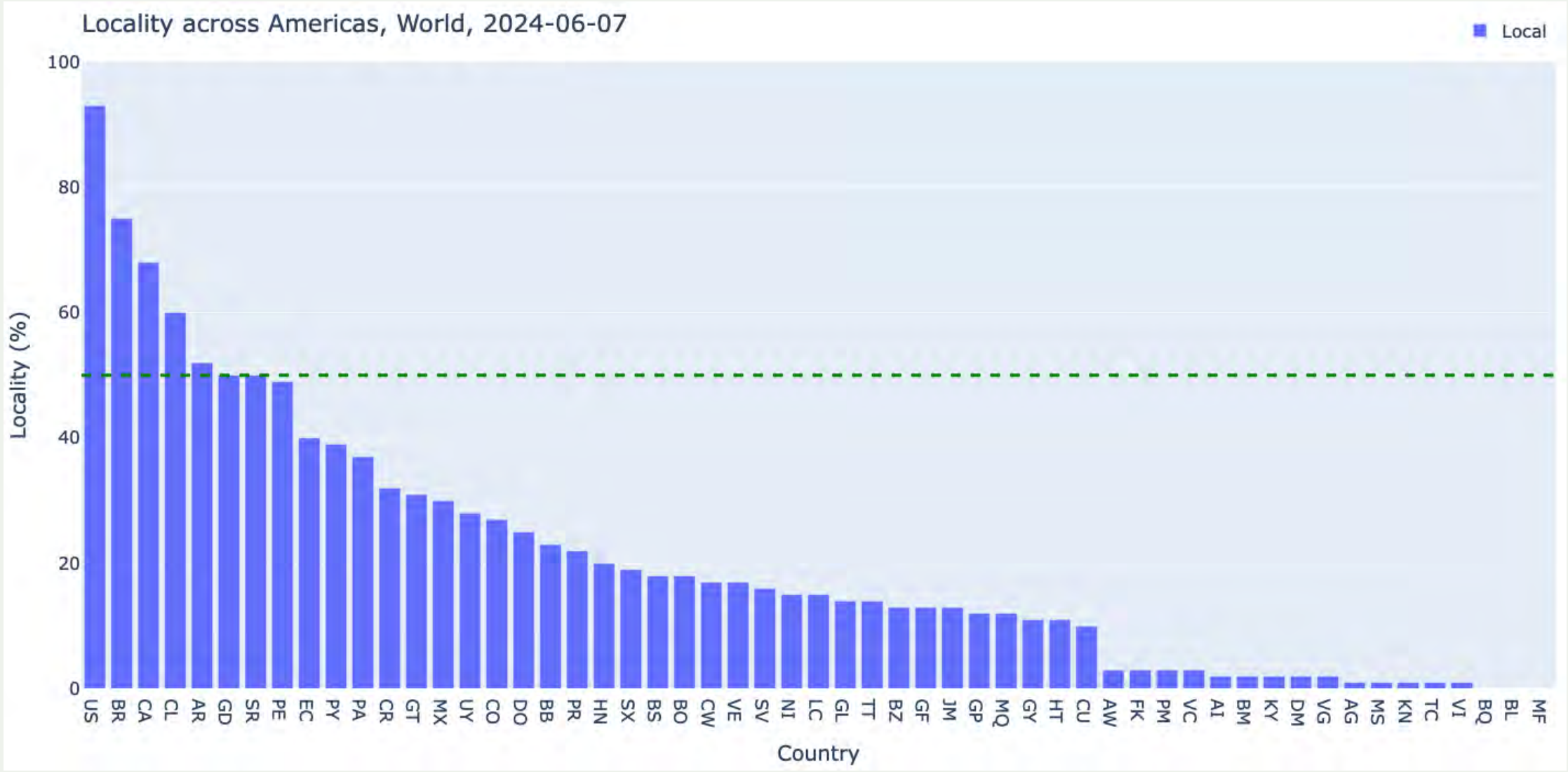
Traceroute/Latency

Visualisation

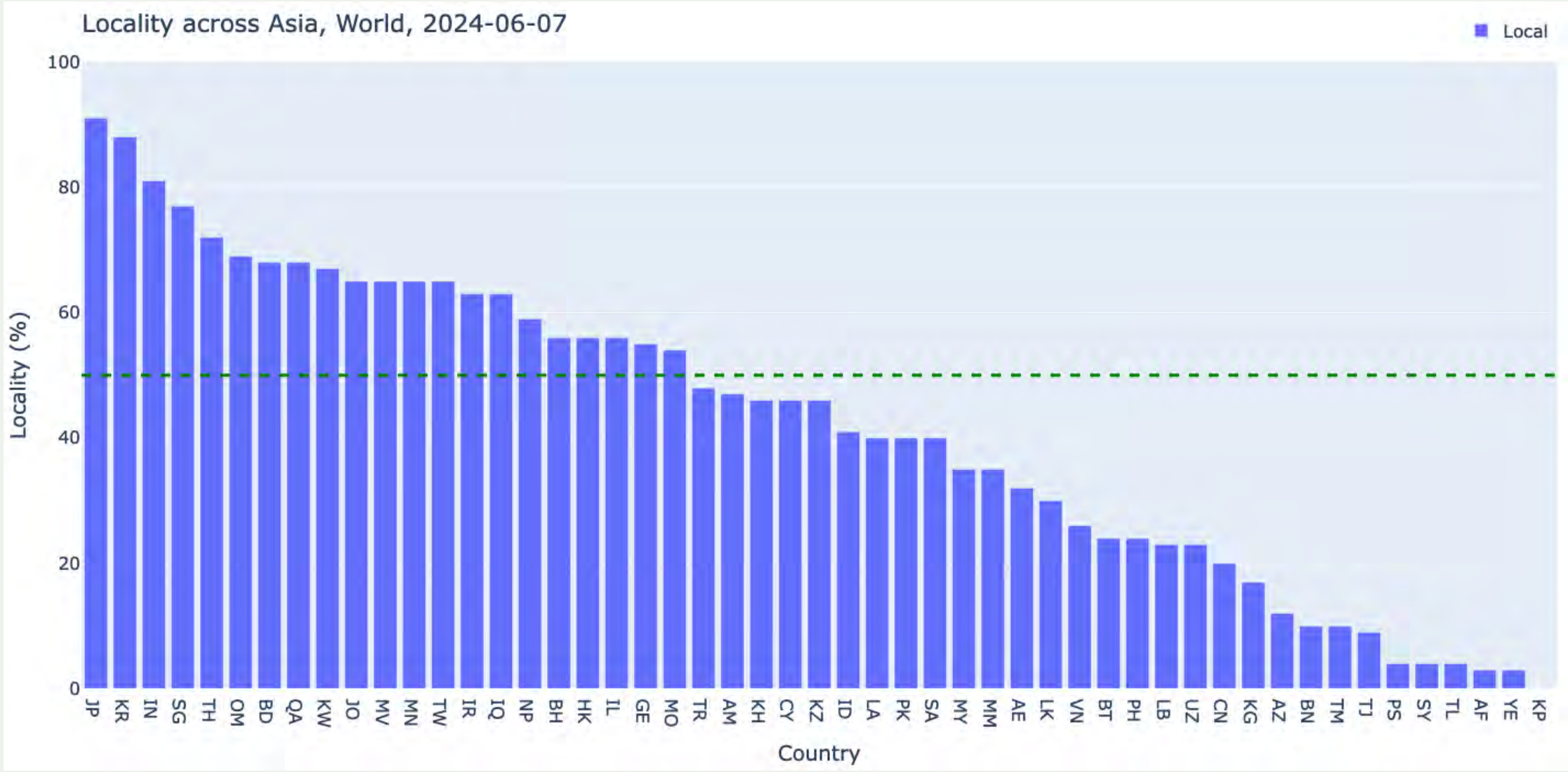
Locality: Africa



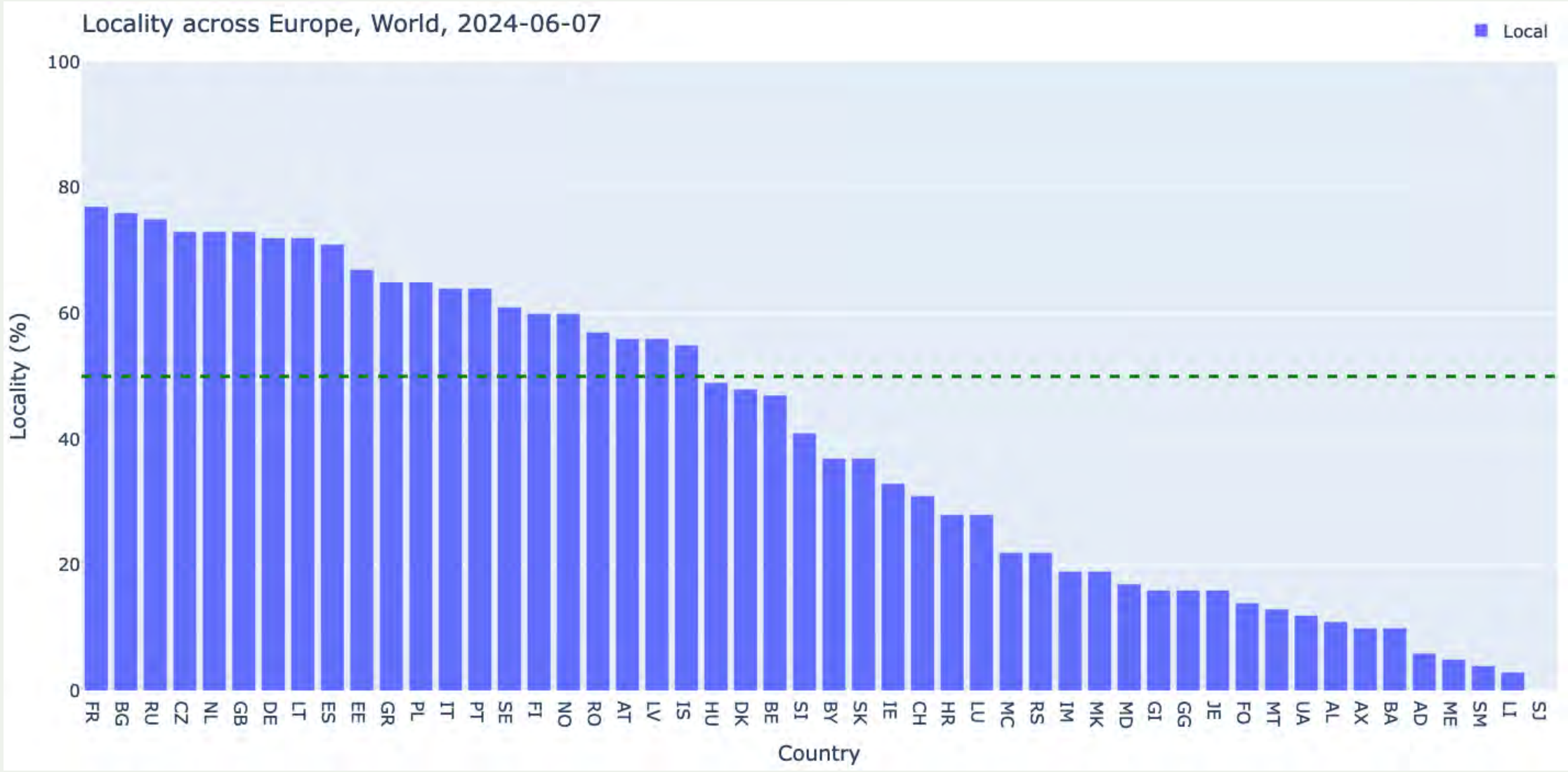
Locality: Americas



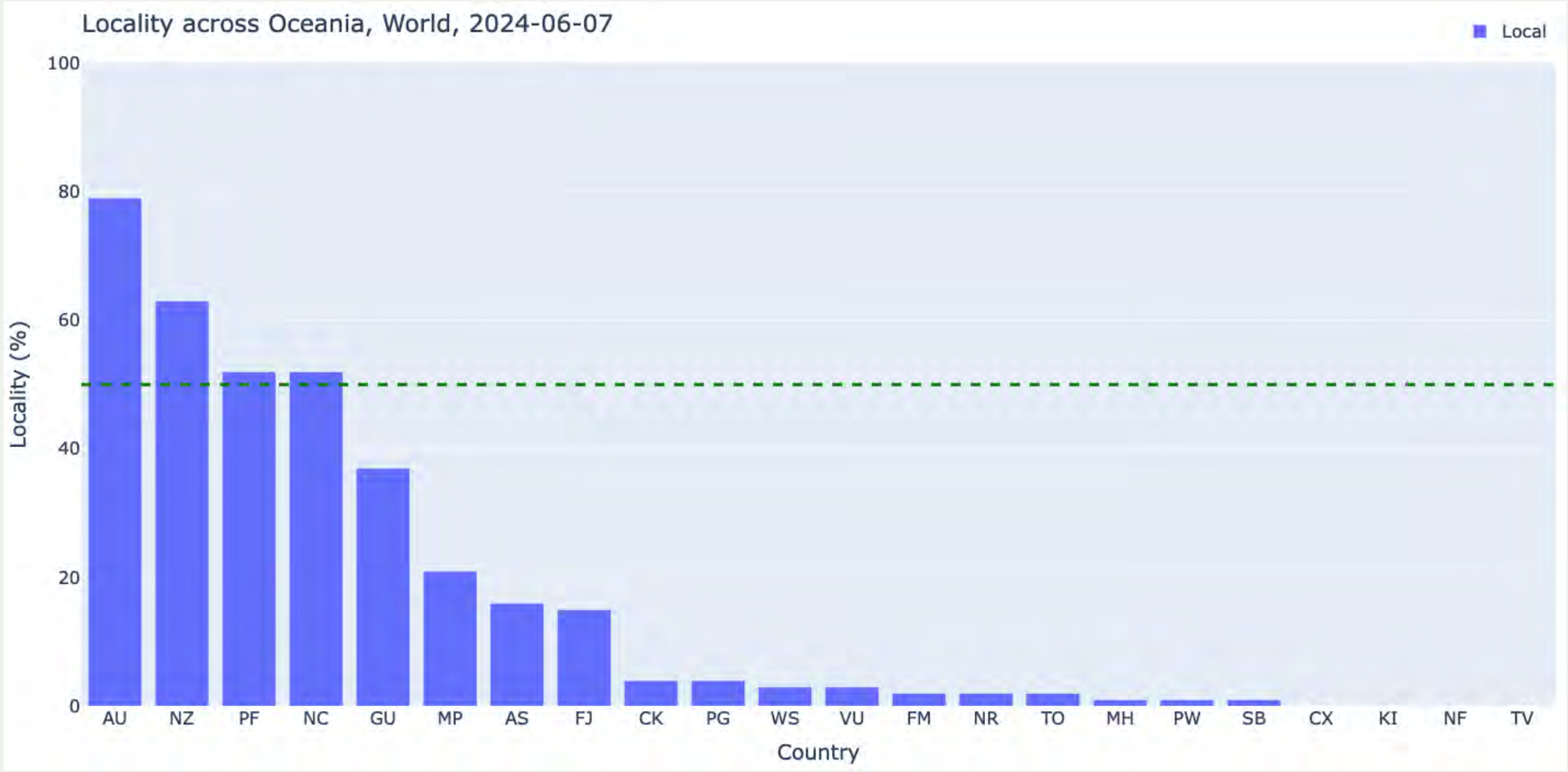
Locality: Asia



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

Wolfgang Mühlbauer
ETH Zurich

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

The Central Problem with Distributed Content

Common CDN Deployments Centralize Traffic In A Risky Way

Kevin Vermeulen
LAAS-CNRS

Loqman Salamatian
Columbia University

Sang Hoon Kim
Columbia University

Ethan Katz-Bassett
Columbia University

CacheLoc: Leveraging CDN Edge Servers for User Geolocation

Mingkui Wei[1][0000−0003−3606−3428], Khaled Rabieh[2][0000−0003−2828−6971], and Faisal Kaleem[2][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

Internet development in … hosting and distribu…

Enrico Calandro[1], Josiah Chavula[2], and Amreesh Phokeer[3]

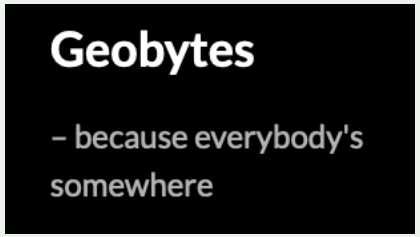
[1] Research ICT Africa, Cape Town, South Africa
ecalandro@researchictafrica.net
[2] University of Cape Town, Cape Town, South Africa
jchavula@cs.uct.ac.za
[3] AFRINIC, Ebene, Mauritius
amreesh@afrinic.net

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



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 B. Patel — Nadine Moukdad
Computer Science
University,
NY 10458
ukdad@gmail.com

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

Towards Geolocation of Millions of IP Addresses*

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