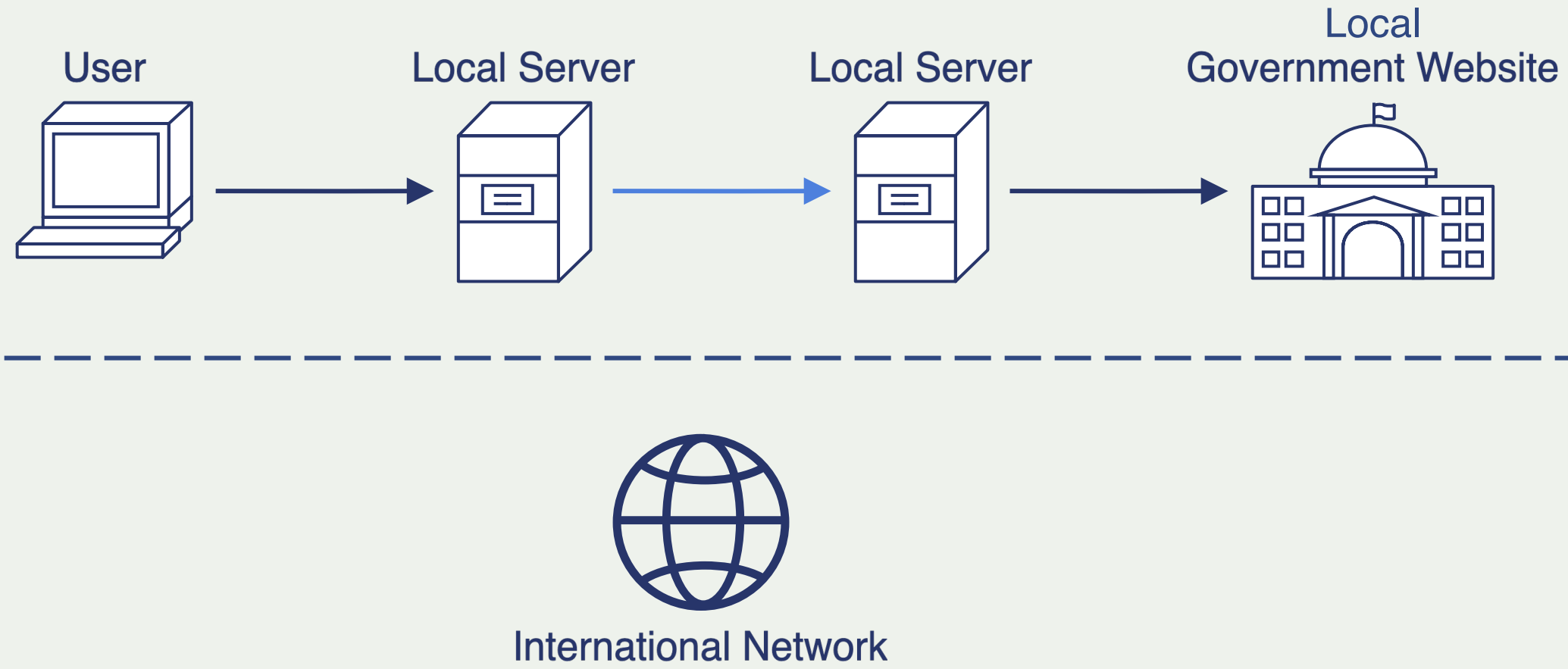
Internet Society
Pulse

# Towards Measuring Content Locality

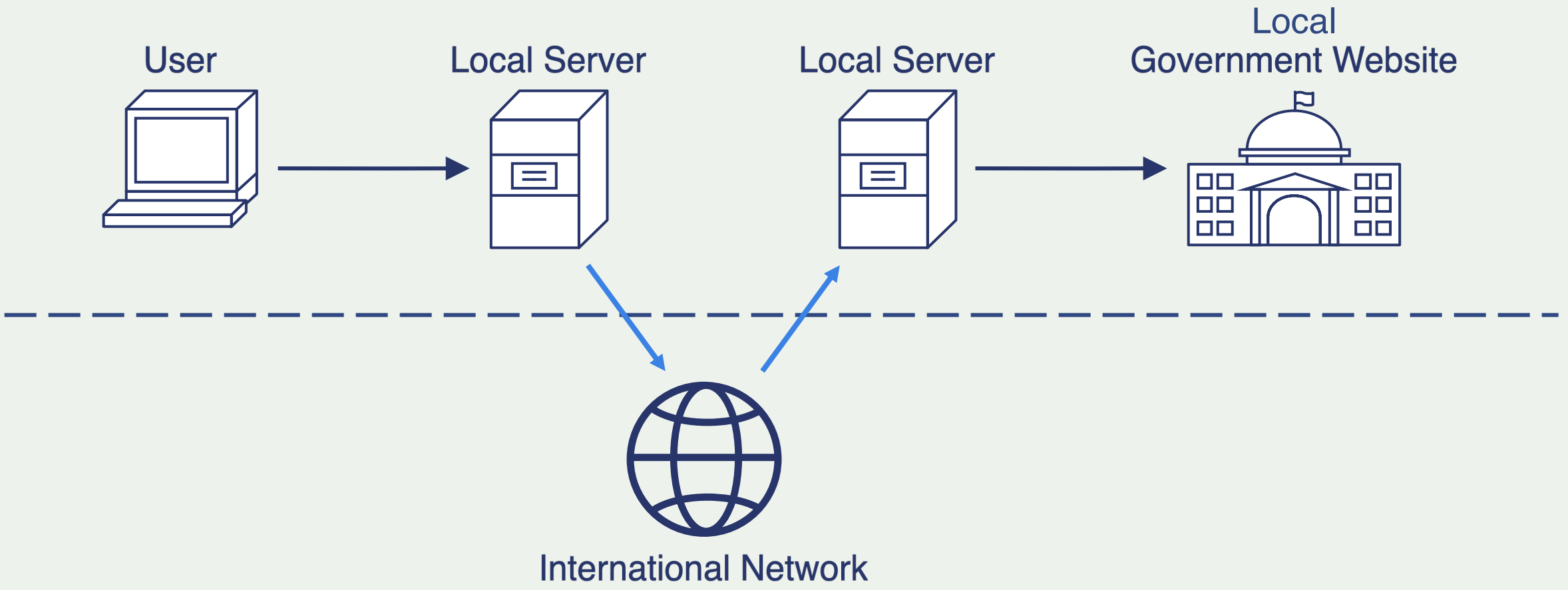
James I. Madeley, Amreesh Phokeer,
Theophilus A. Benson, Aftab Siddiqui

ANRW'24
23rd July 2024

# Local Traffic



# External Traffic



# 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]

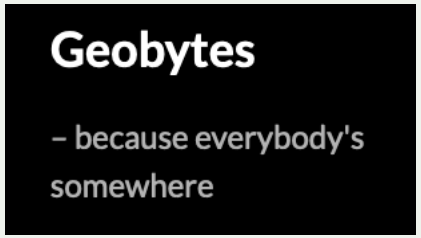
[1] Research ICT Africa, Cape Town, South Africa
ecalandro@researchictafrica.net
[2] University of Cape Town, Cape Town, South Africa
jchavula@cs.uct.ac.za
[3] AFRINIC, Ebene, Mauritius
amreesh@afrinic.net

## 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
Computer Science
University,
NY 10458
ukdad@gmail.com

Nadine Moukdad

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

# 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 | Live streaming favours local content | |
| Internet Society's 50/50 Vision | | |

# Motivation

## Policy

South Africa's POPI Act / EU's GDPR

Internet Society's 50/50 Vision

## Performance

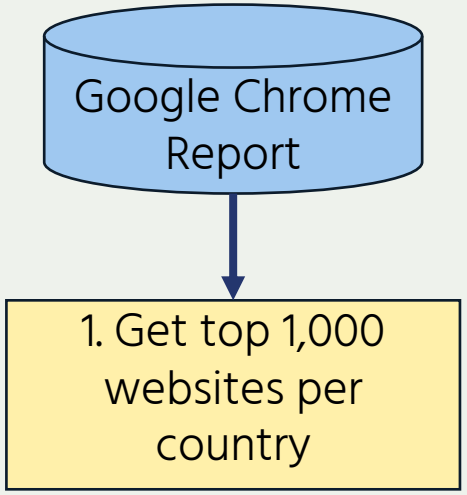
Live streaming favours local content

## Persistence
(availability)

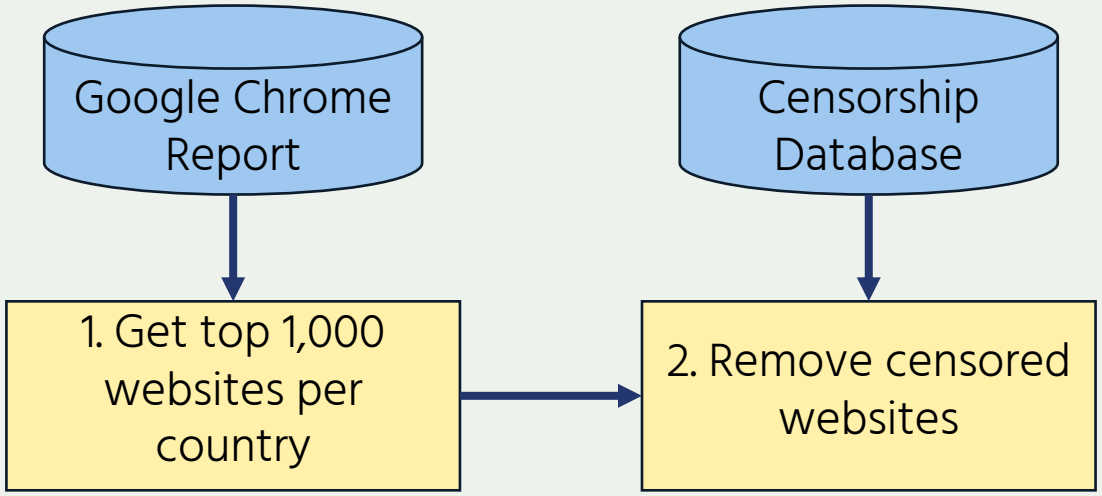
Resilience against faults e.g. cable cuts

# Methodology

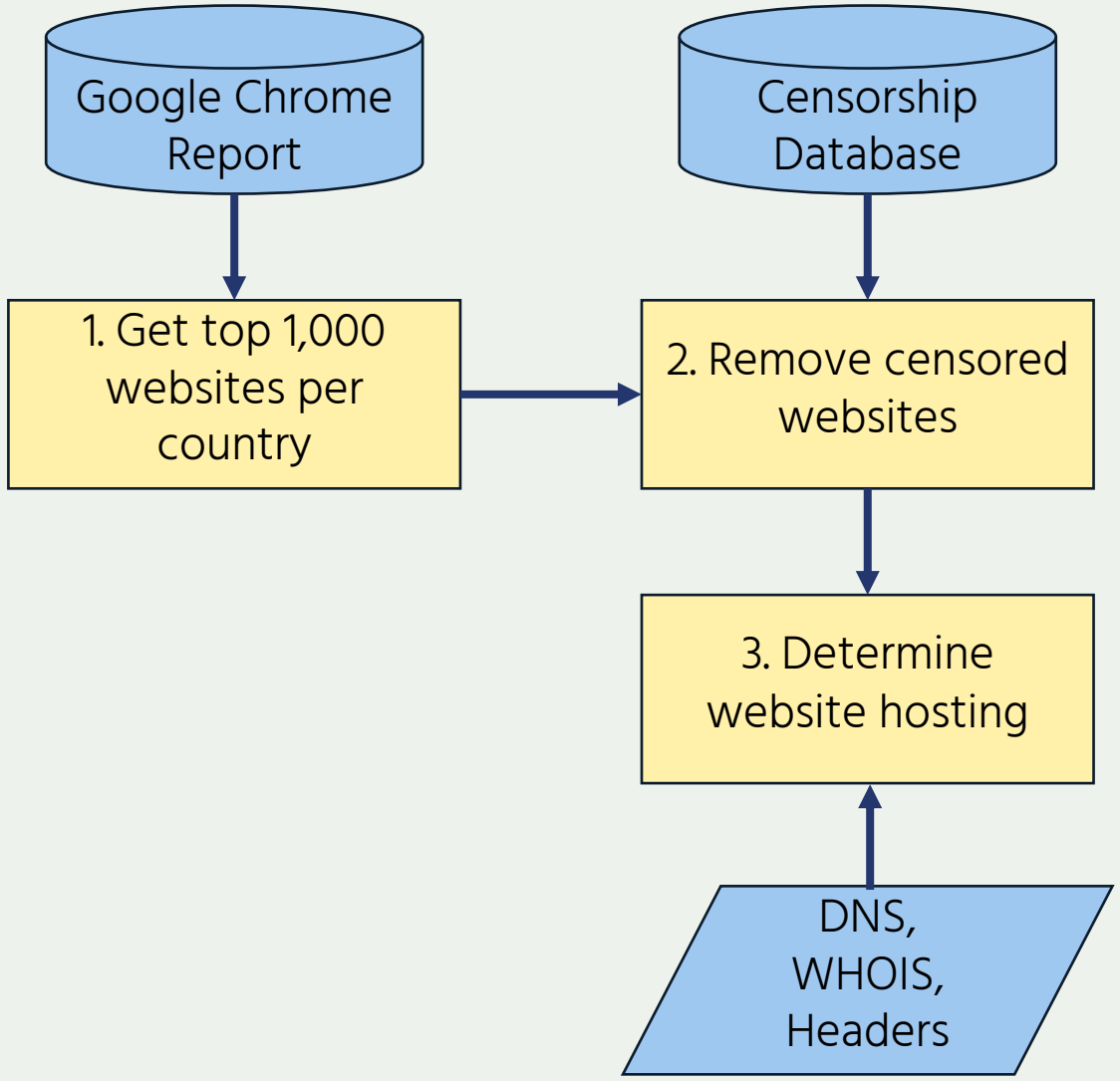
# Methodology



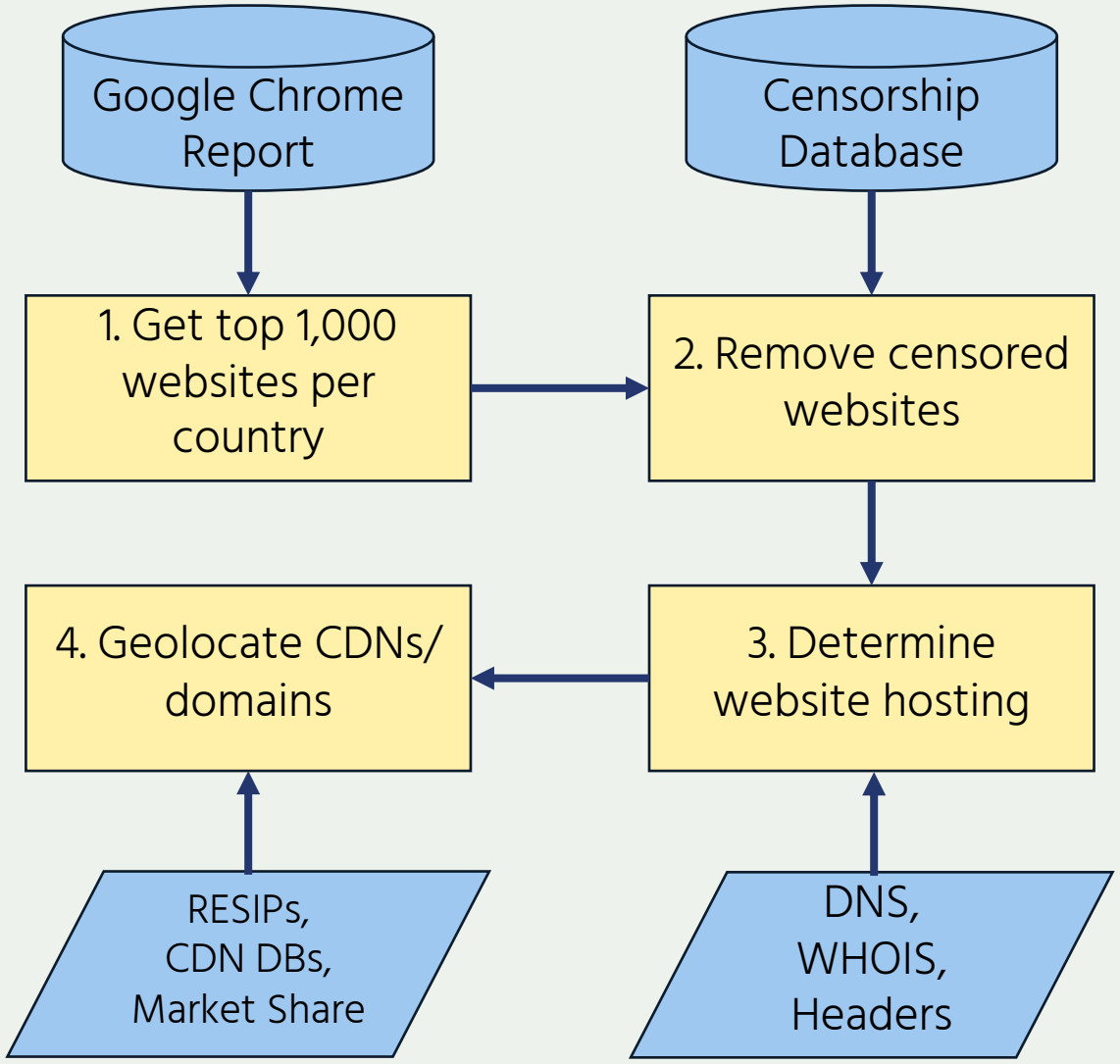
# Methodology



# Methodology



# Methodology



# Methodology – Determine Website Hosting

www.studysmarter.co.uk, NL (Netherlands)

| Service | Returned | Parsed |
|---|---|---|
| WHOIS | CDN77 _, GB | CDN77 |
| DNS CNAME | prod-web-uk.b-cdn.net. | BunnyCDN |
| | Result | Tie |

# Methodology – Determine Website Hosting

www.studysmarter.co.uk, NL (Netherlands)

| Service | Returned | Parsed |
|---|---|---|
| WHOIS | CDN77 _, GB | CDN77 |
| DNS CNAME | prod-web-uk.b-cdn.net. | BunnyCDN |
| | **Result** | **Tie** |
| Headers | BunnyCDN-DE1-1080 | BunnyCDN |
| | **Result** | **BunnyCDN** |

# Methodology – Geolocate Domains without CDNs

wikipedia.org, NL (Netherlands)

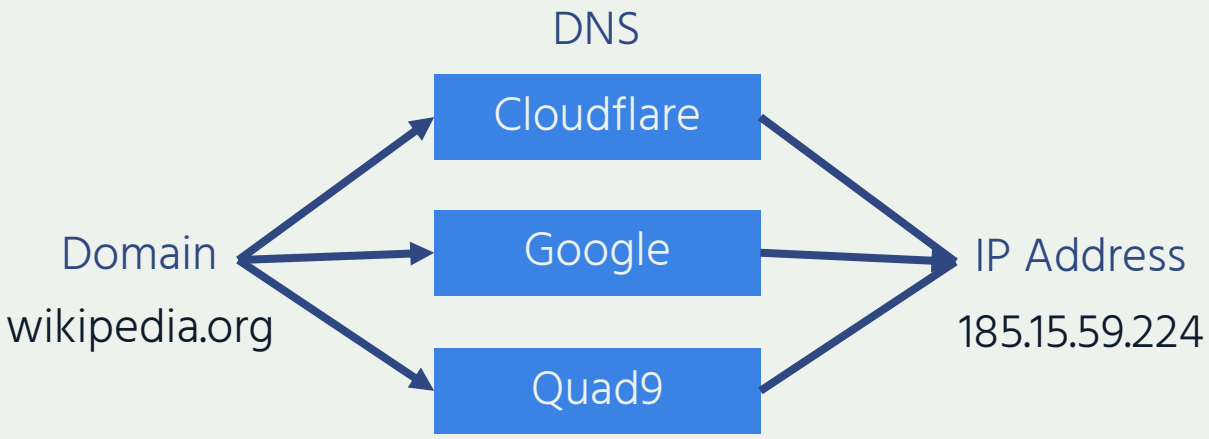
# Methodology – Geolocate Domains without CDNs

wikipedia.org, NL (Netherlands)

| Domain |
| --- |
| wikipedia.org |

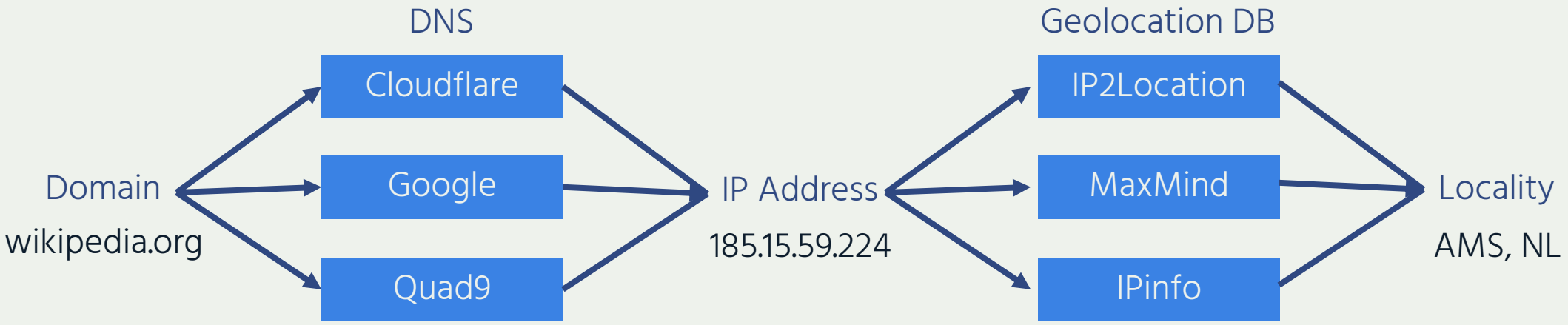
# Methodology – Geolocate Domains without CDNs

wikipedia.org, NL (Netherlands)



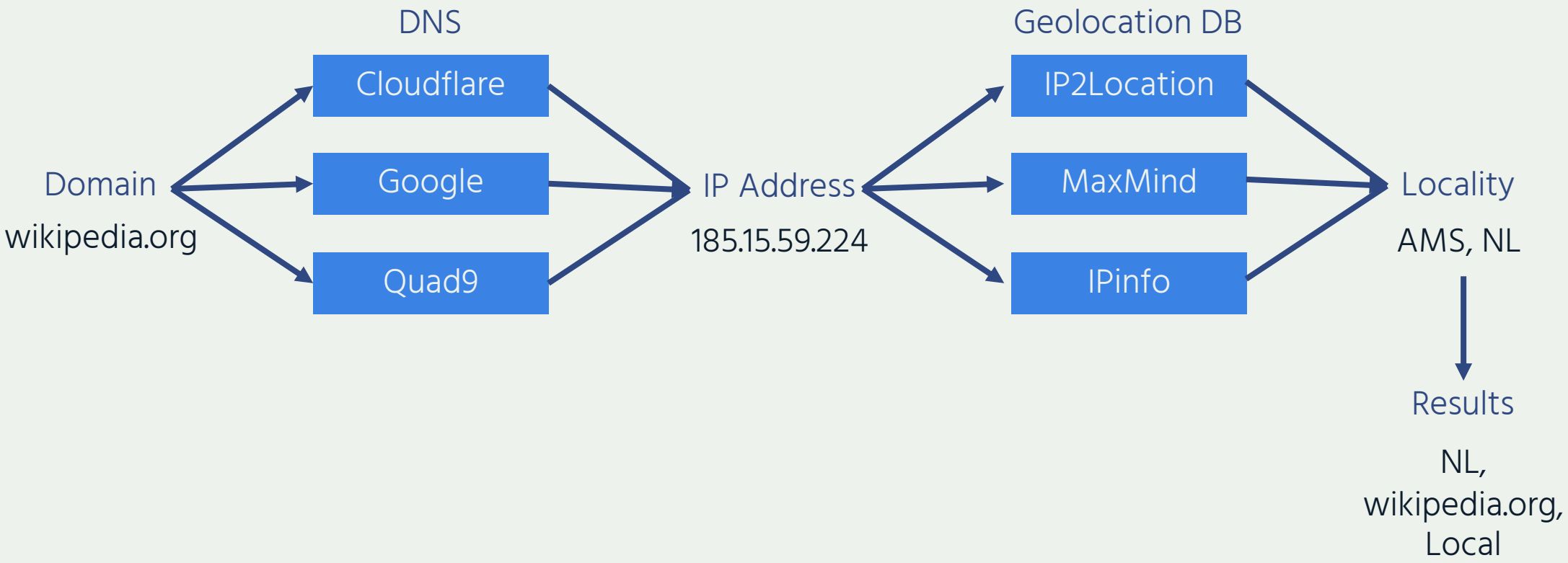
# Methodology – Geolocate Domains without CDNs

wikipedia.org, NL (Netherlands)



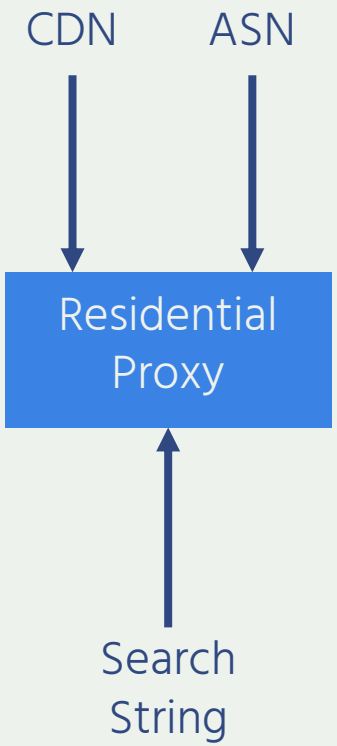
# Methodology – Geolocate Domains without CDNs

wikipedia.org, NL (Netherlands)

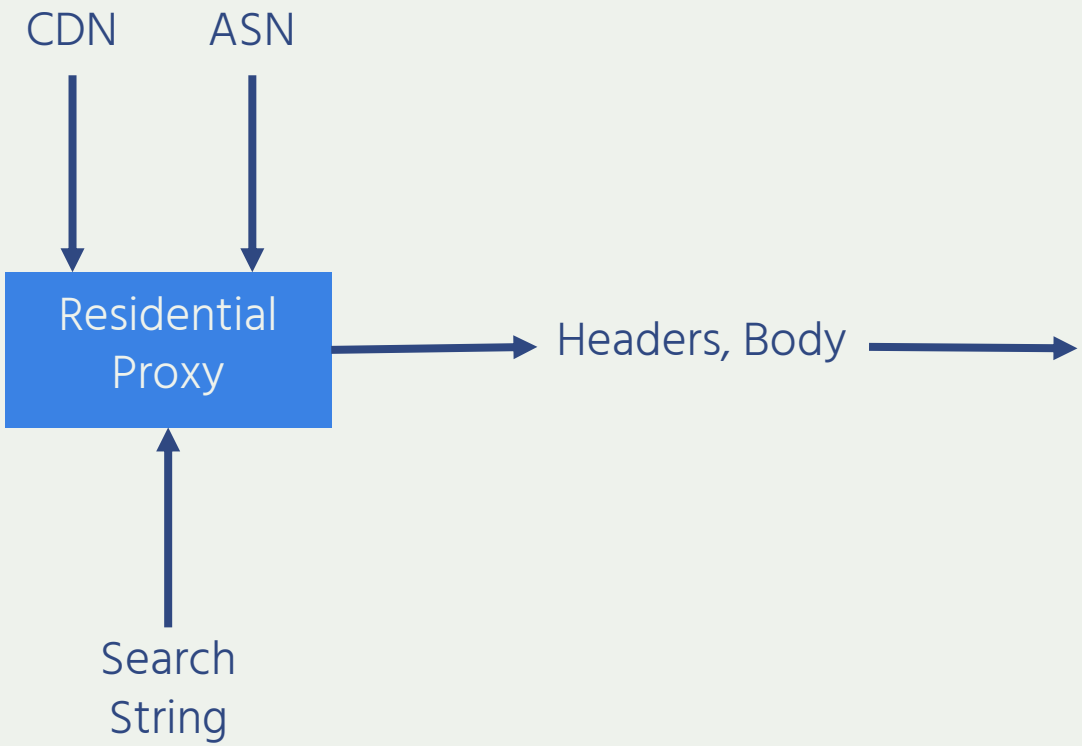


# Methodology – Geolocate CDNs

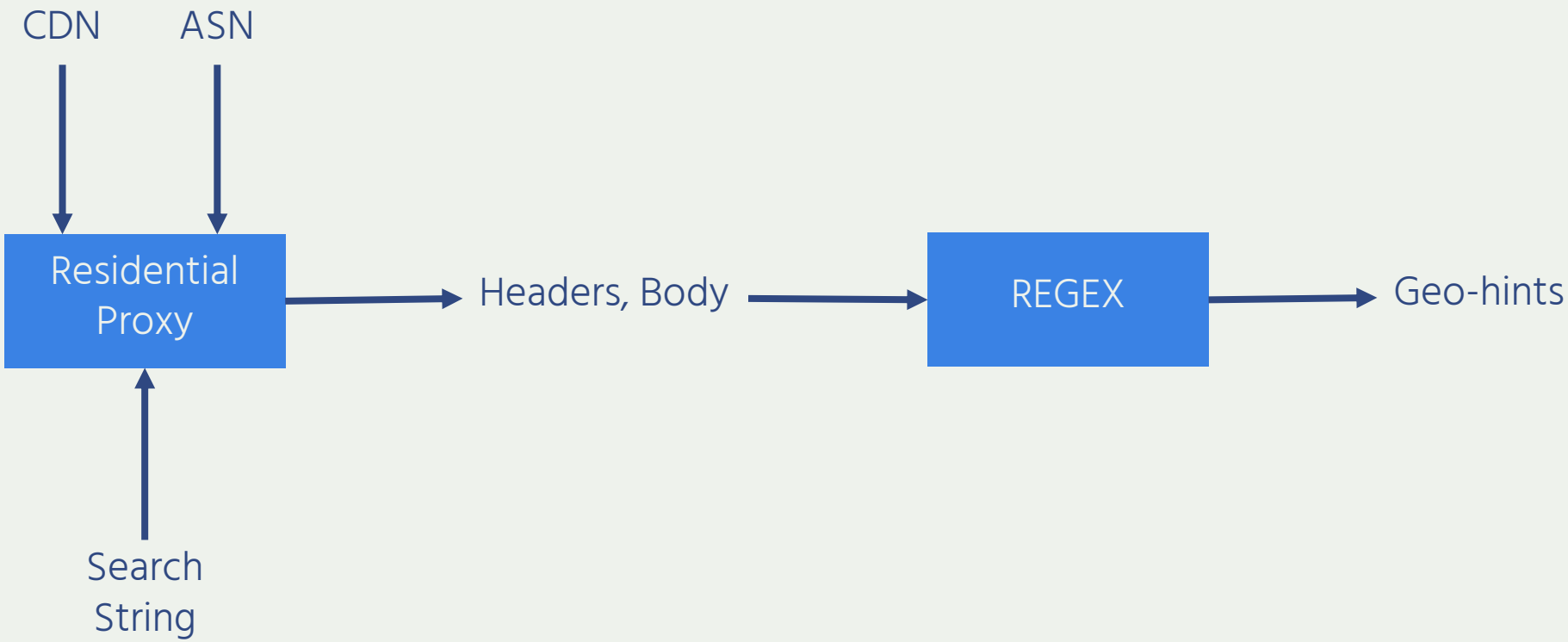
# Methodology – Geolocate CDNs



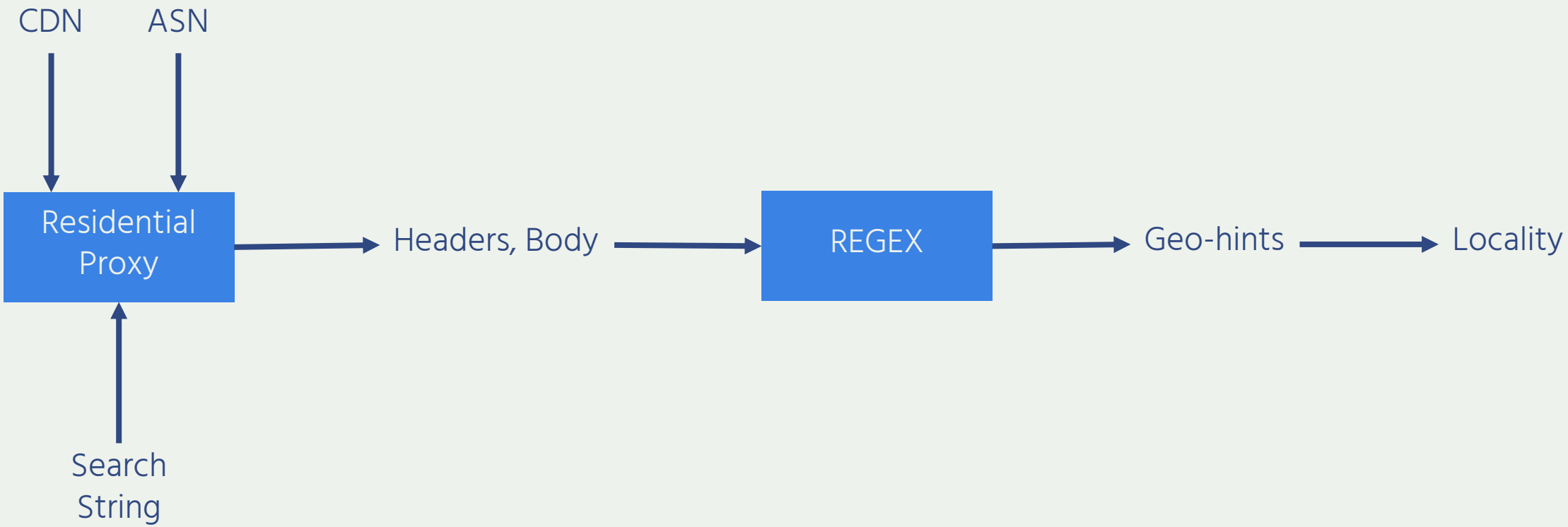
# Methodology – Geolocate CDNs



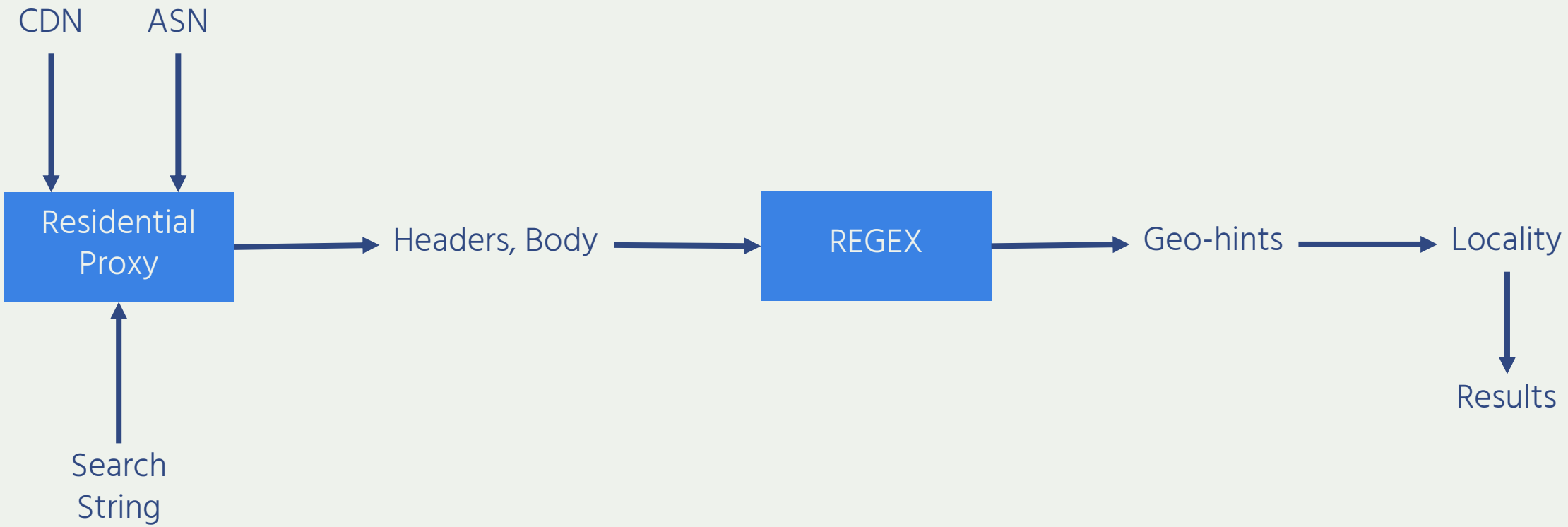
# Methodology – Geolocate CDNs



# Methodology – Geolocate CDNs

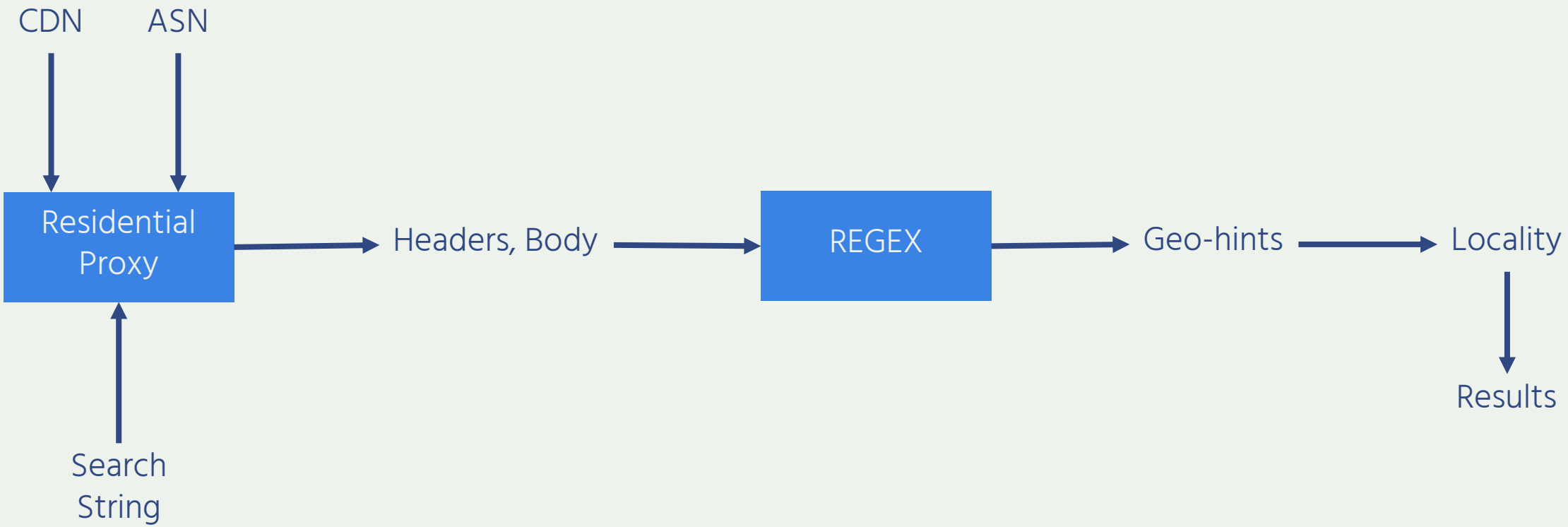


# Methodology – Geolocate CDNs



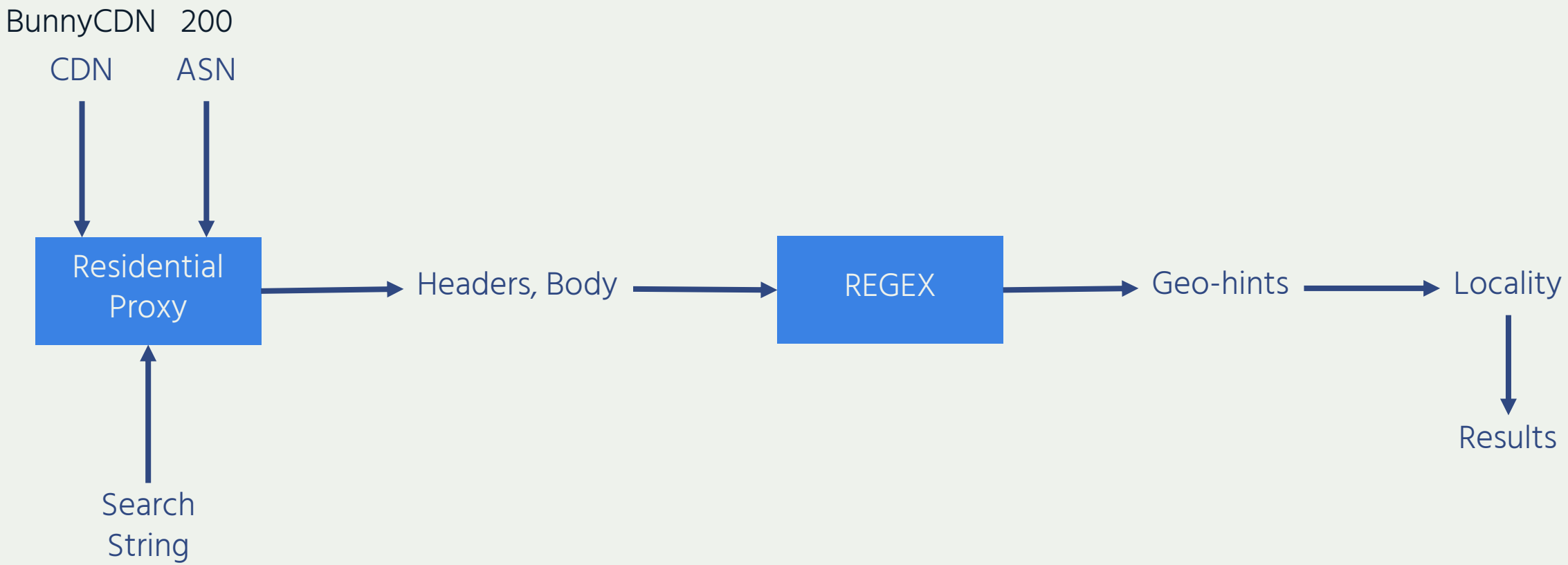
# Methodology – Geolocate CDNs

BunnyCDN, AS 200, NL (Netherlands)



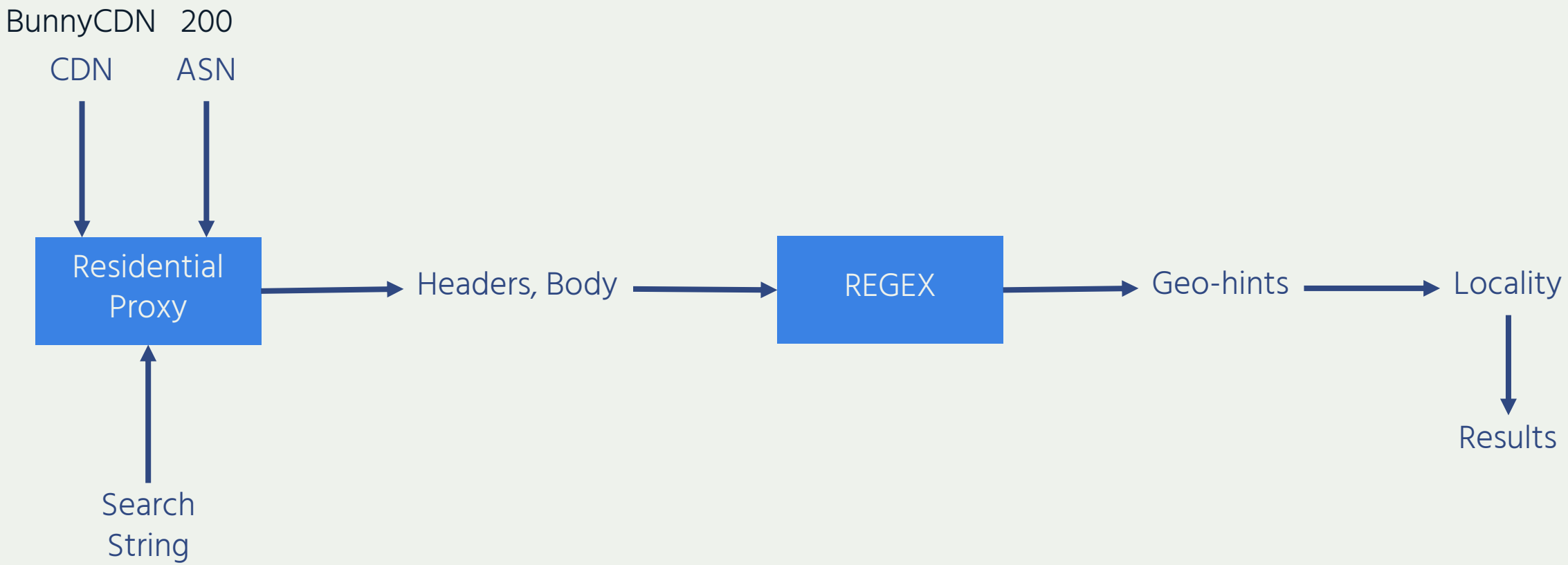
# Methodology – Geolocate CDNs

BunnyCDN, AS 200, NL (Netherlands)



# Methodology – Geolocate CDNs

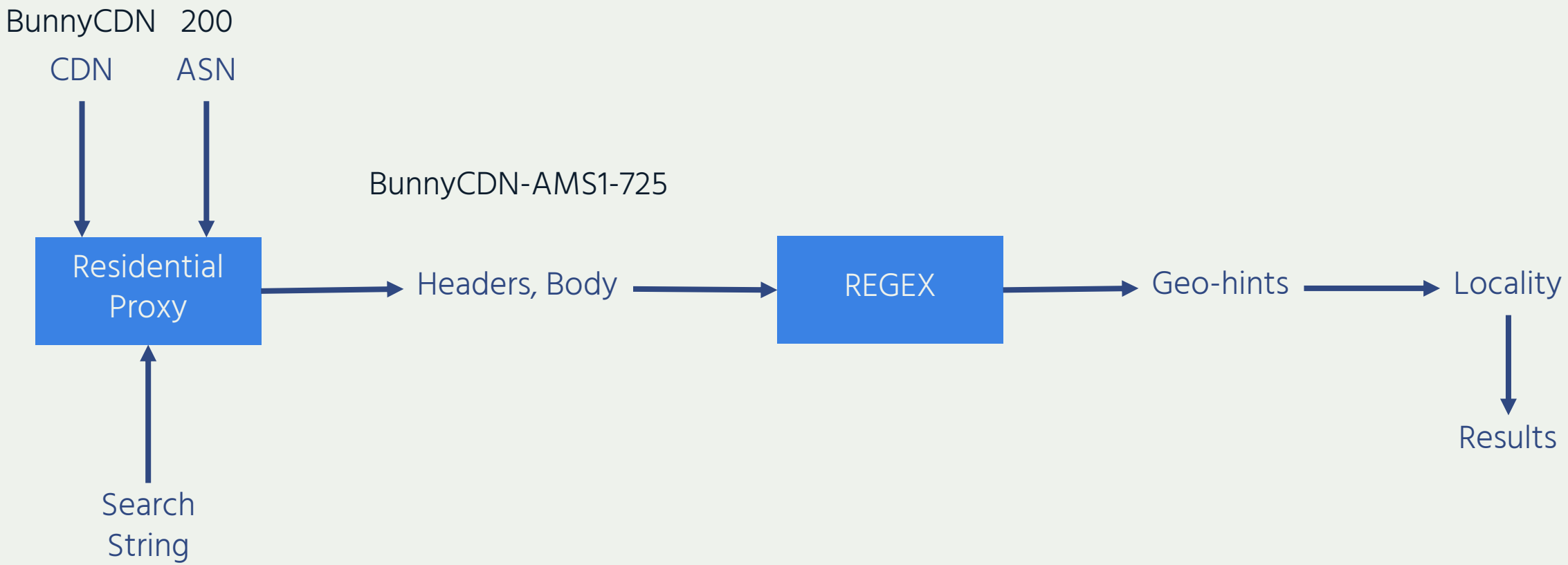
BunnyCDN, AS 200, NL (Netherlands)




https://bunny.net/images/bunnynet-the-best-content-delivery-network-cdn.jpg

# Methodology – Geolocate CDNs

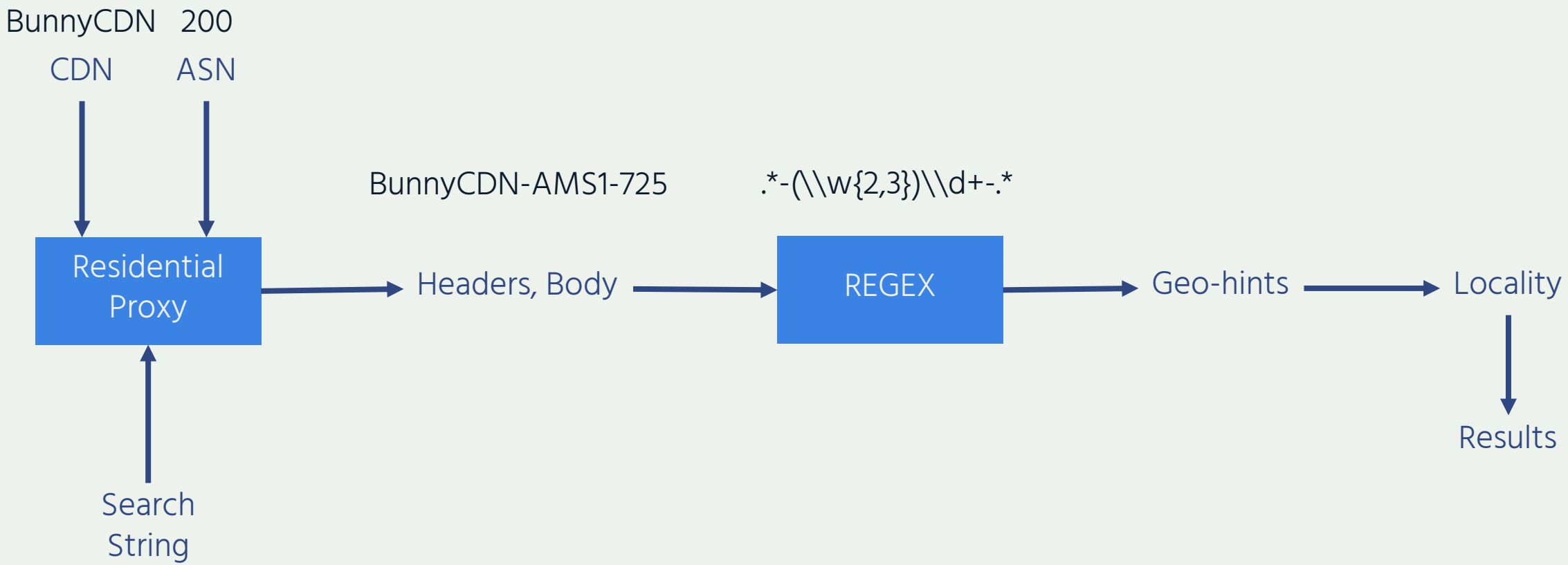
BunnyCDN, AS 200, NL (Netherlands)




https://bunny.net/images/bunnynet-the-best-content-delivery-network-cdn.jpg

# Methodology – Geolocate CDNs

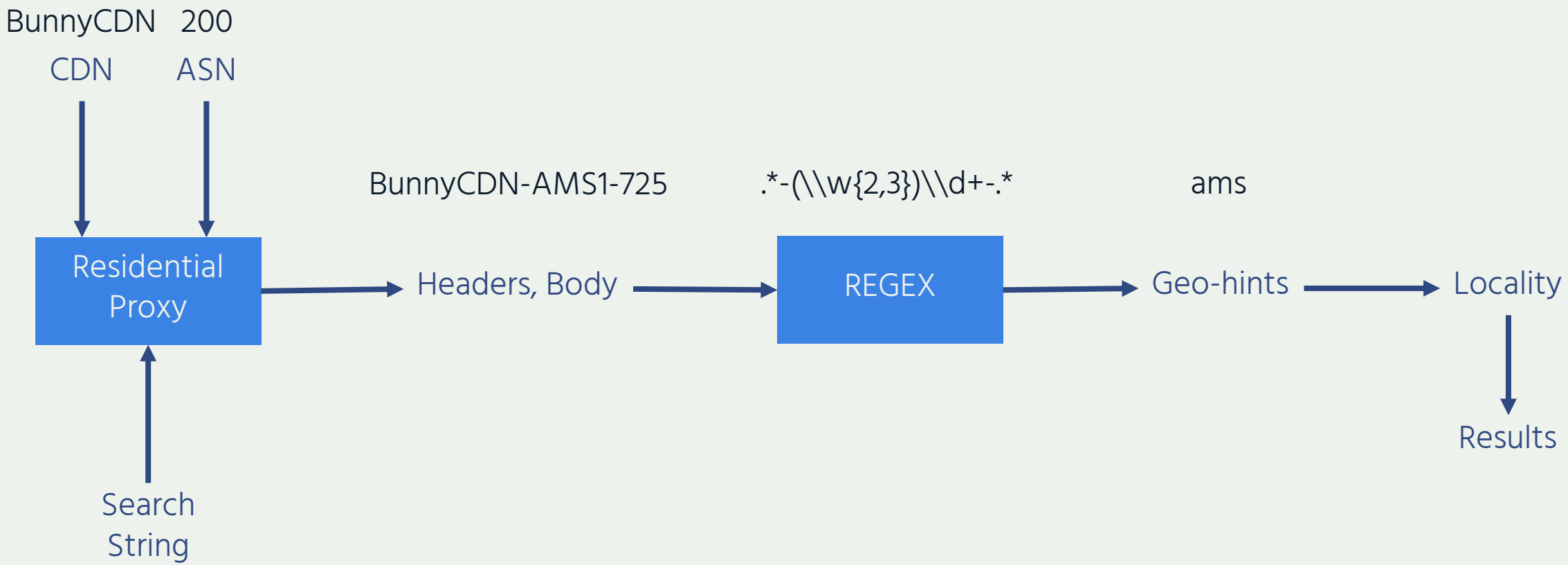
**BunnyCDN, AS 200, NL (Netherlands)**




https://bunny.net/images/bunnynet-the-best-content-delivery-network-cdn.jpg

# Methodology – Geolocate CDNs

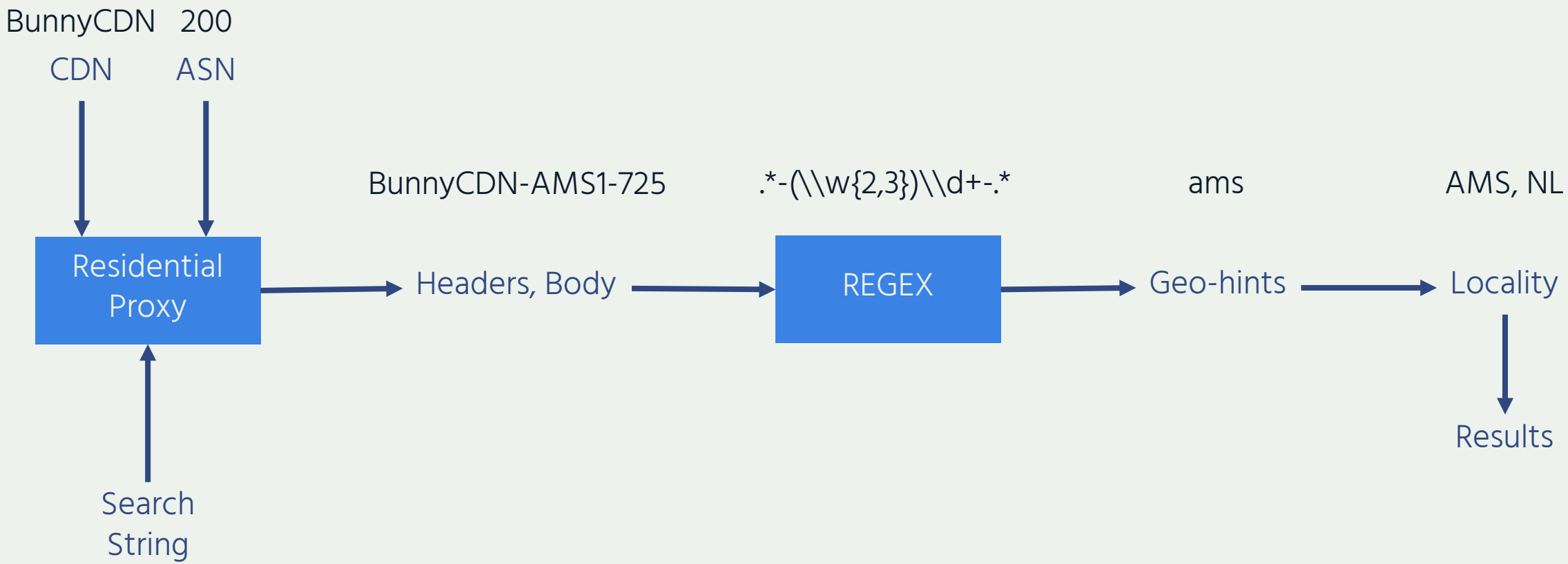
**BunnyCDN, AS 200, NL (Netherlands)**




https://bunny.net/images/bunnynet-the-best-content-delivery-network-cdn.jpg

# Methodology – Geolocate CDNs

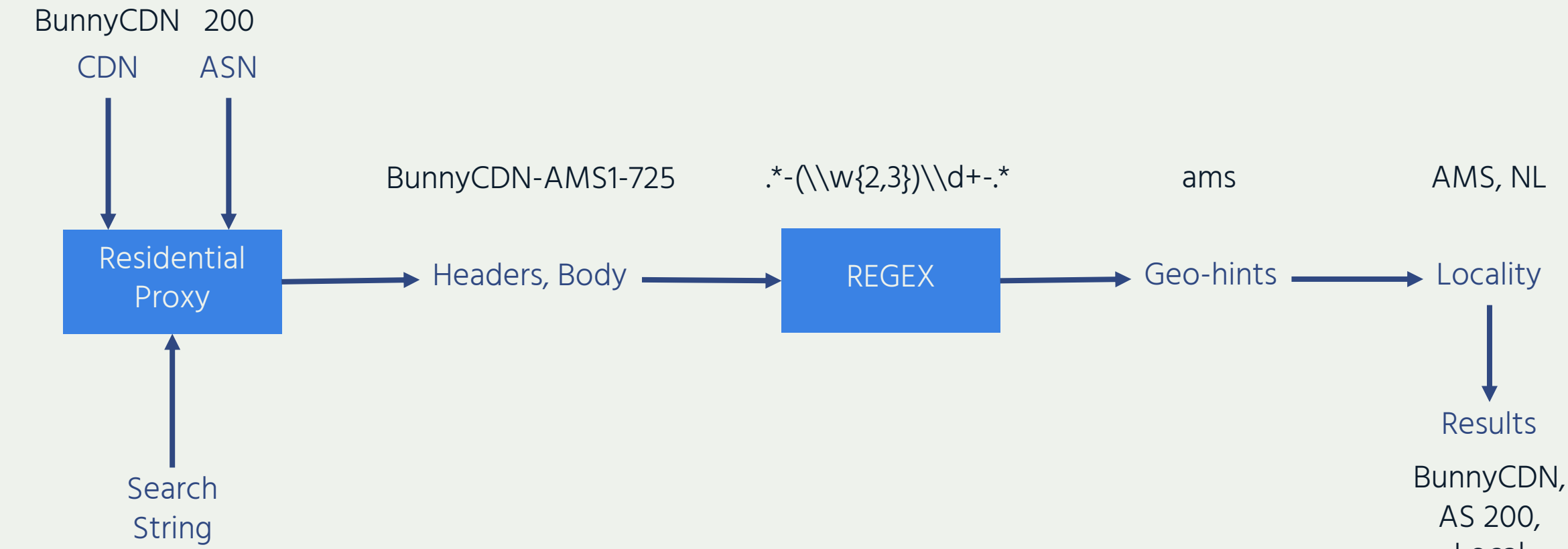
**BunnyCDN, AS 200, NL (Netherlands)**




https://bunny.net/images/bunnynet-the-best-content-delivery-network-cdn.jpg

# Methodology – Geolocate CDNs

**BunnyCDN, AS 200, NL (Netherlands)**




https://bunny.net/images/bunnynet-the-best-content-delivery-network-cdn.jpg

# Methodology – Combining Results

# Methodology – Combining Results

| CDN | ASN | Market Share | Local (Y/N) |
|---|---|---|---|
| BunnyCDN | AS 200 | 50% | Y |
| | AS 250 | 50% | N |

# Methodology – Combining Results

| CDN | ASN | Market Share | Local (Y/N) |
|---|---|---|---|
| BunnyCDN | AS 200 | 50% | Y |
| | AS 250 | 50% | N |

| Hosting Method | Domain Count | Locality | Local Domains |
|---|---|---|---|
| BunnyCDN | 10 | 50% | 5 |
| Native | 5 | 20% | 1 |

# Methodology – Combining Results

| CDN | ASN | Market Share | Local (Y/N) |
|---|---|---|---|
| BunnyCDN | AS 200 | 50% | Y |
| | AS 250 | 50% | N |

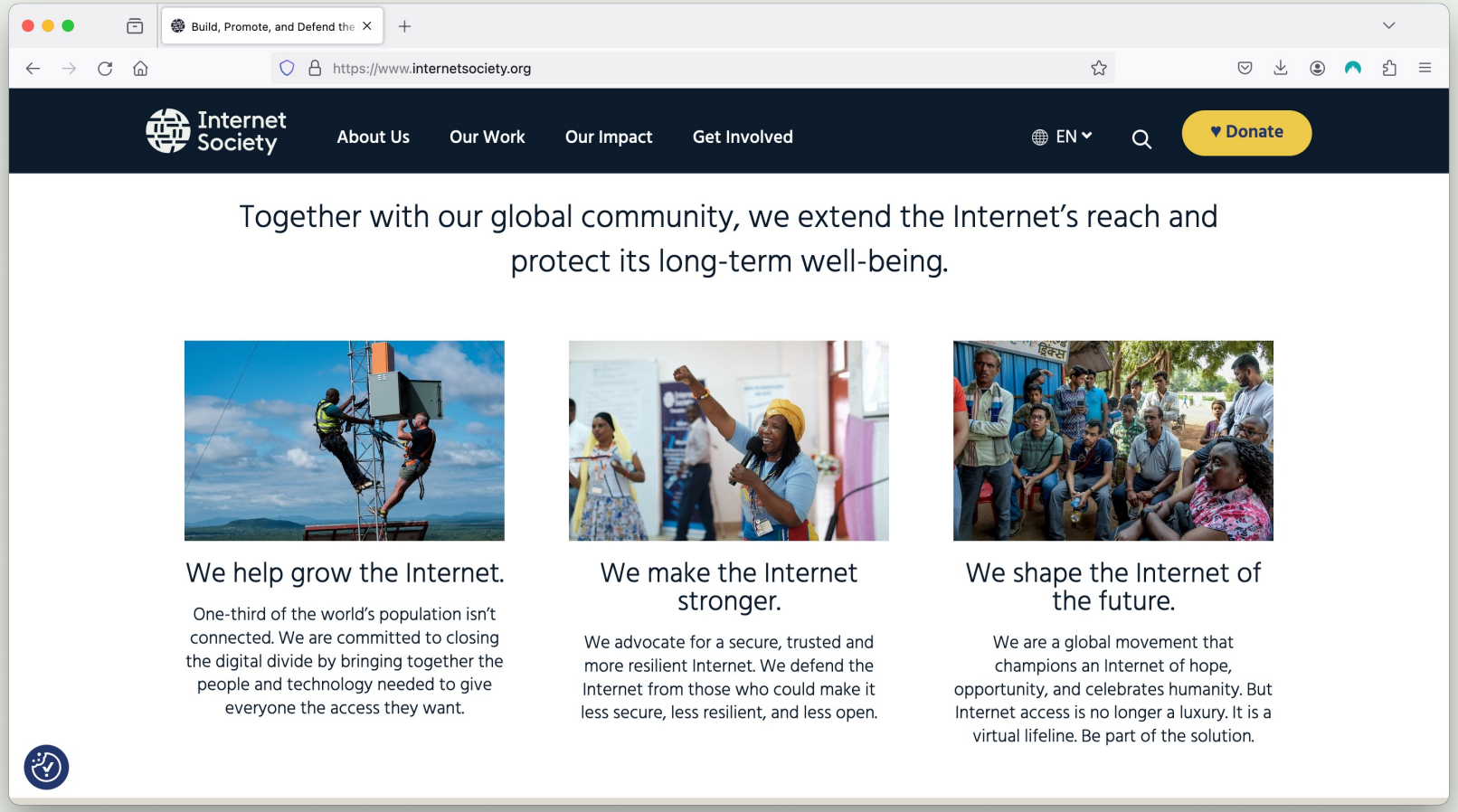
| Hosting Method | Domain Count | Locality | Local Domains |
|---|---|---|---|
| BunnyCDN | 10 | 50% | 5 |
| Native | 5 | 20% | 1 |

| Country Locality | 40% |
|---|---|

# Limitations

# Limitations

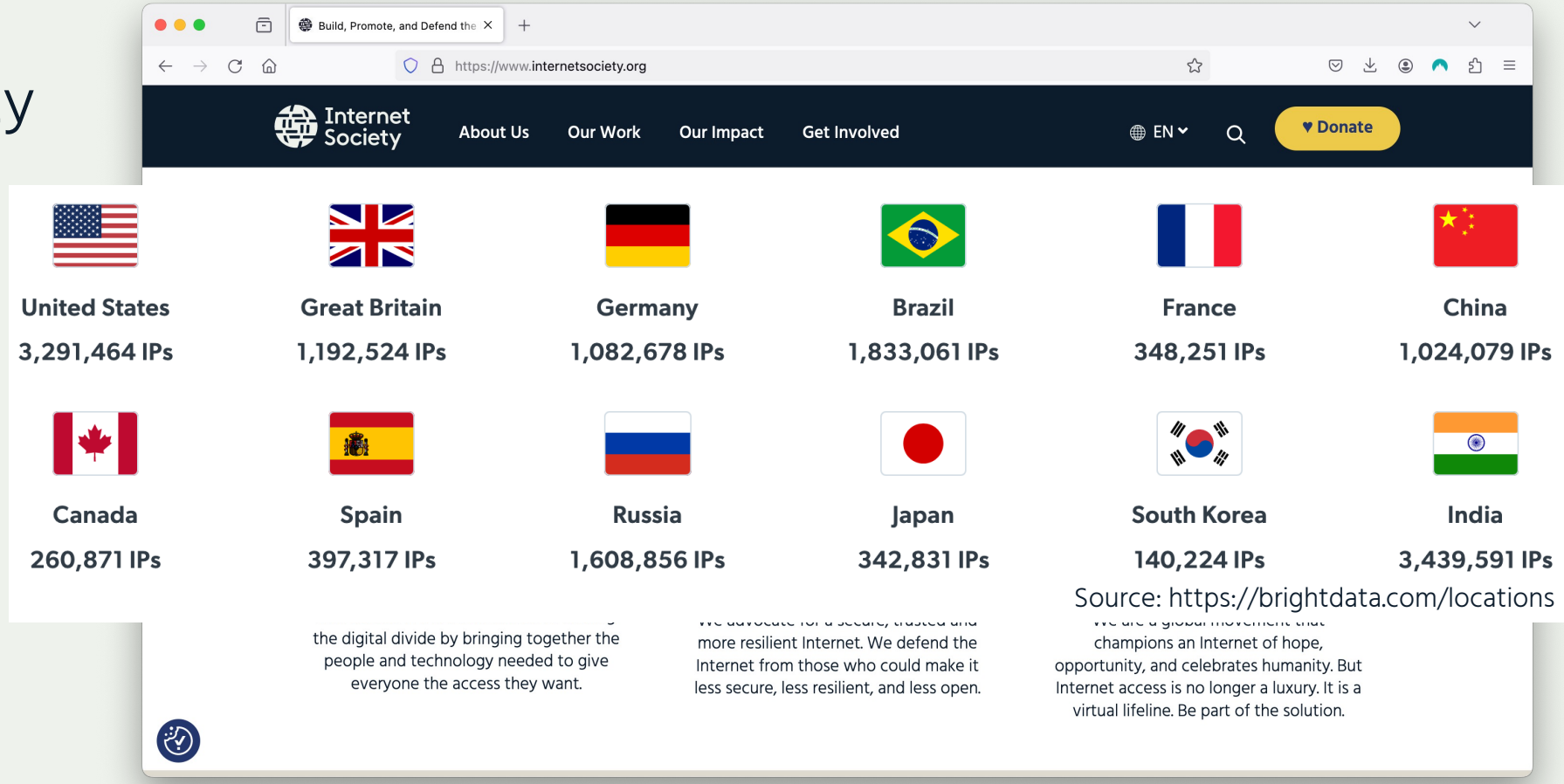
## Website Complexity



# Limitations

Website Complexity

Proxy Probes

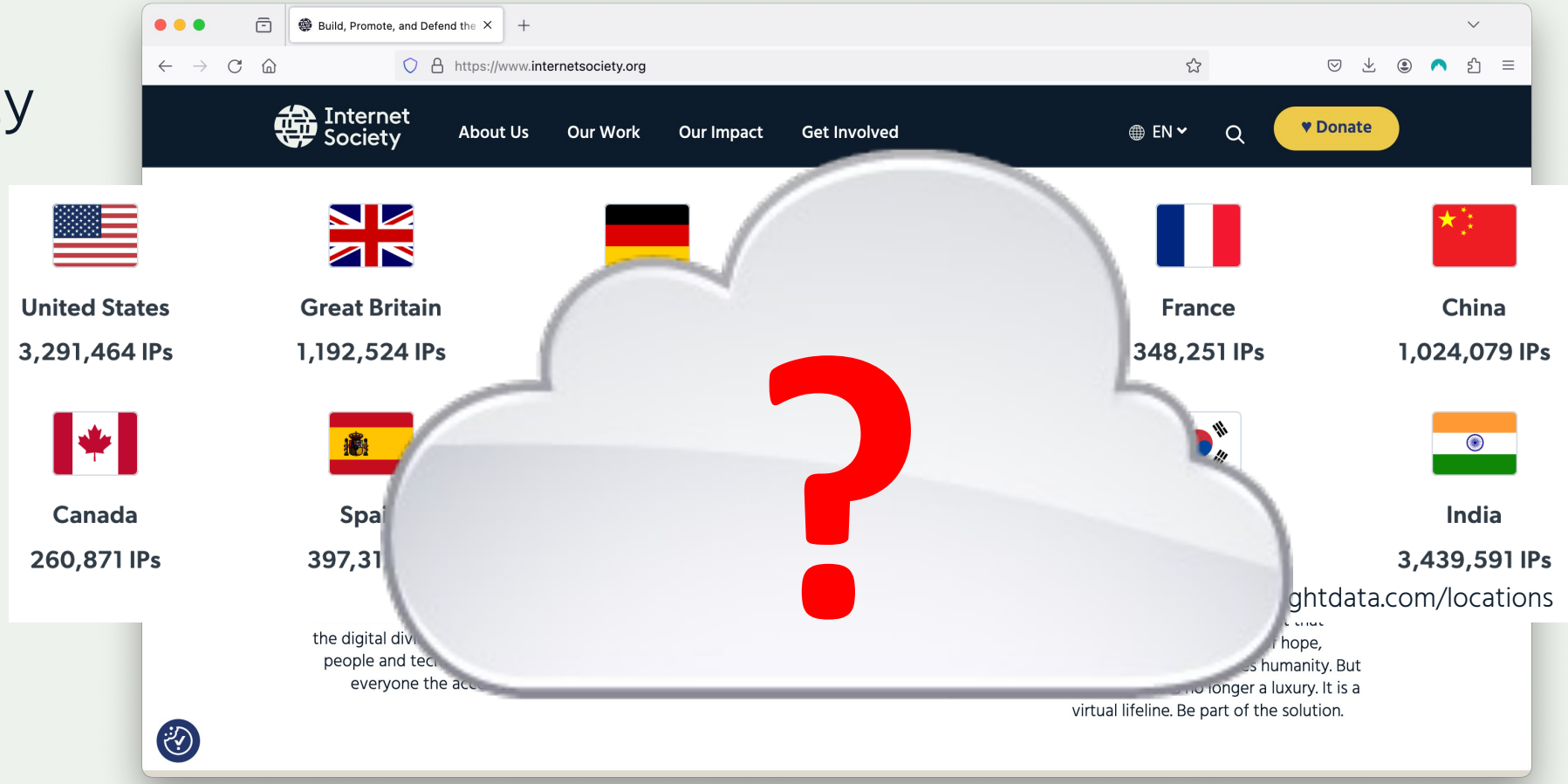


# Limitations

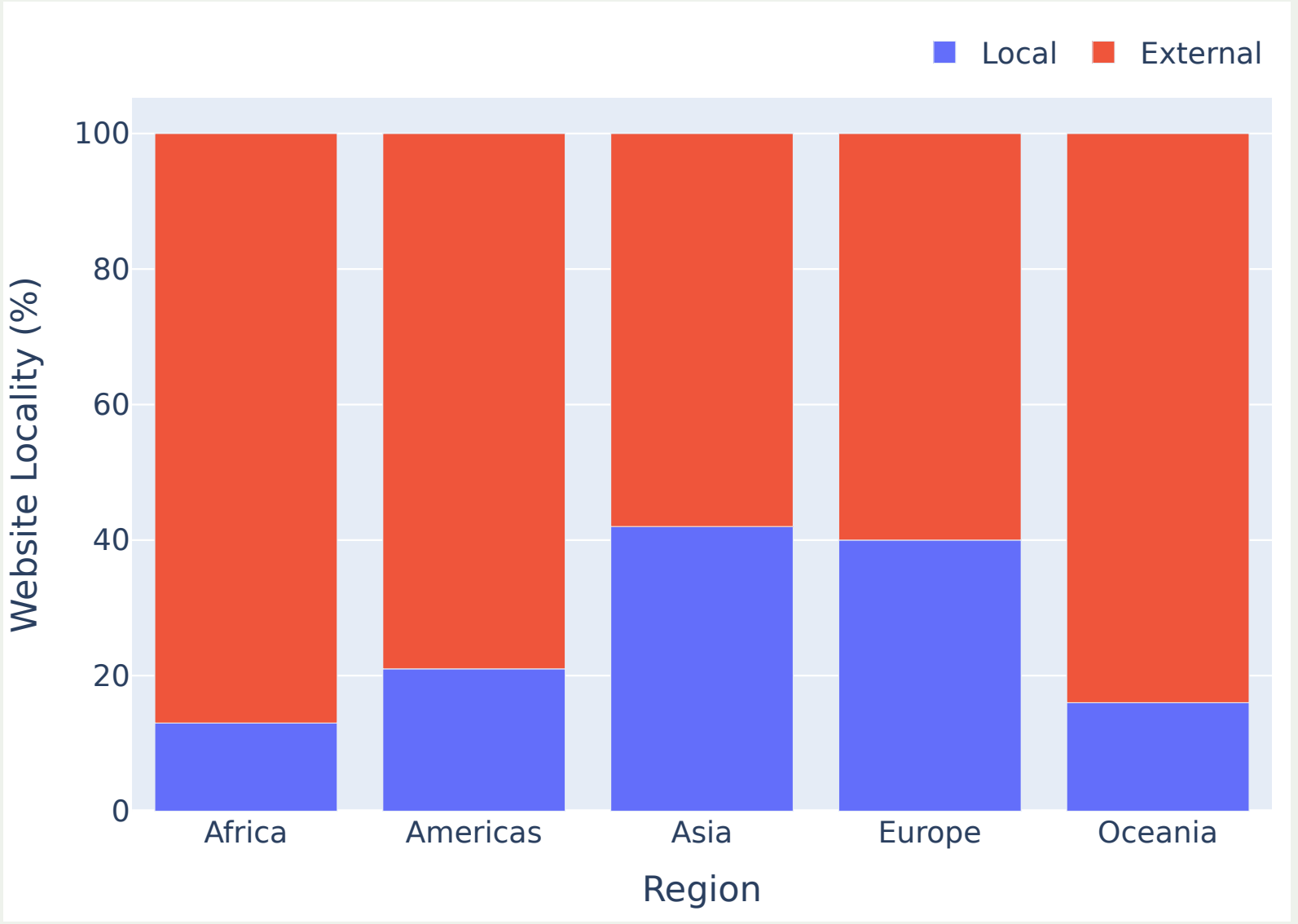
Website Complexity

Proxy Probes

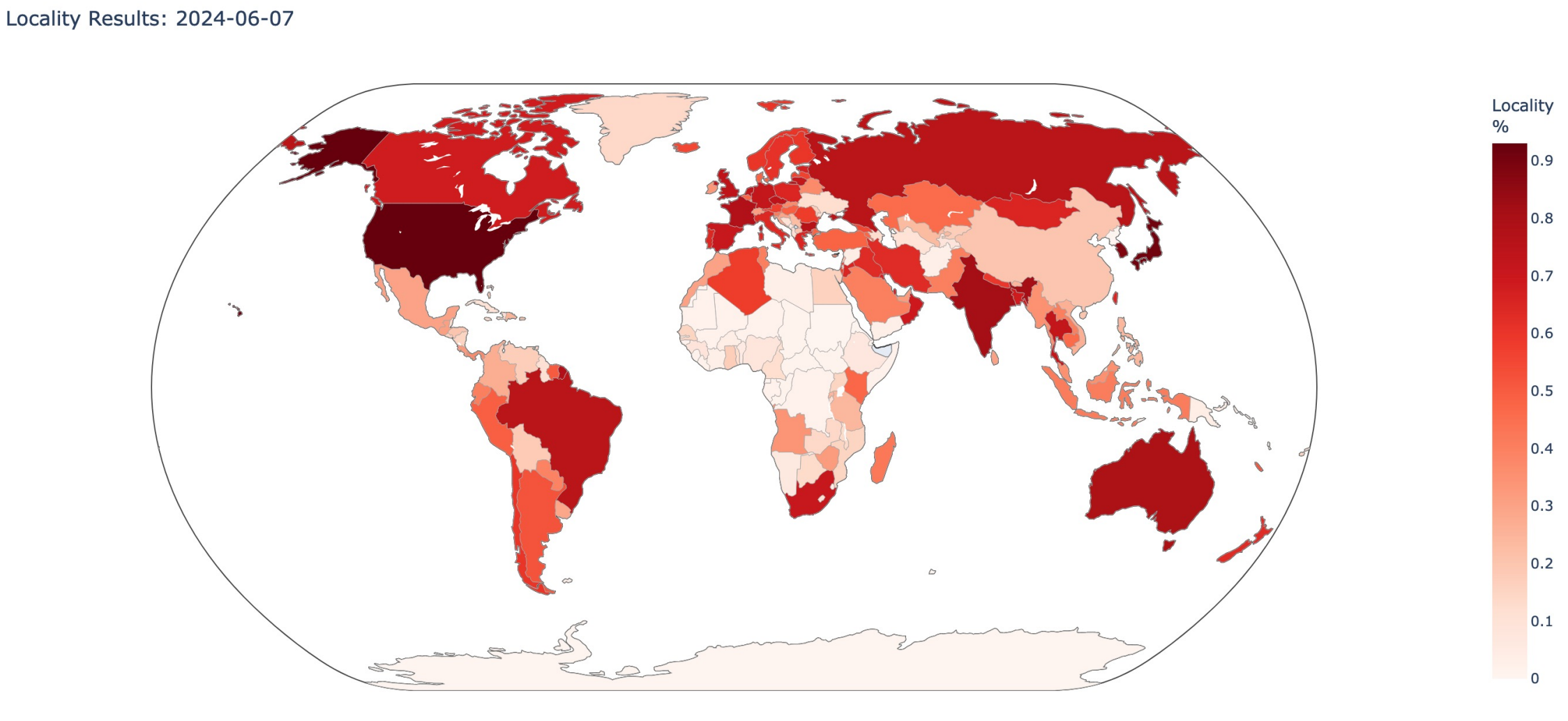
“Fog of Cloud”



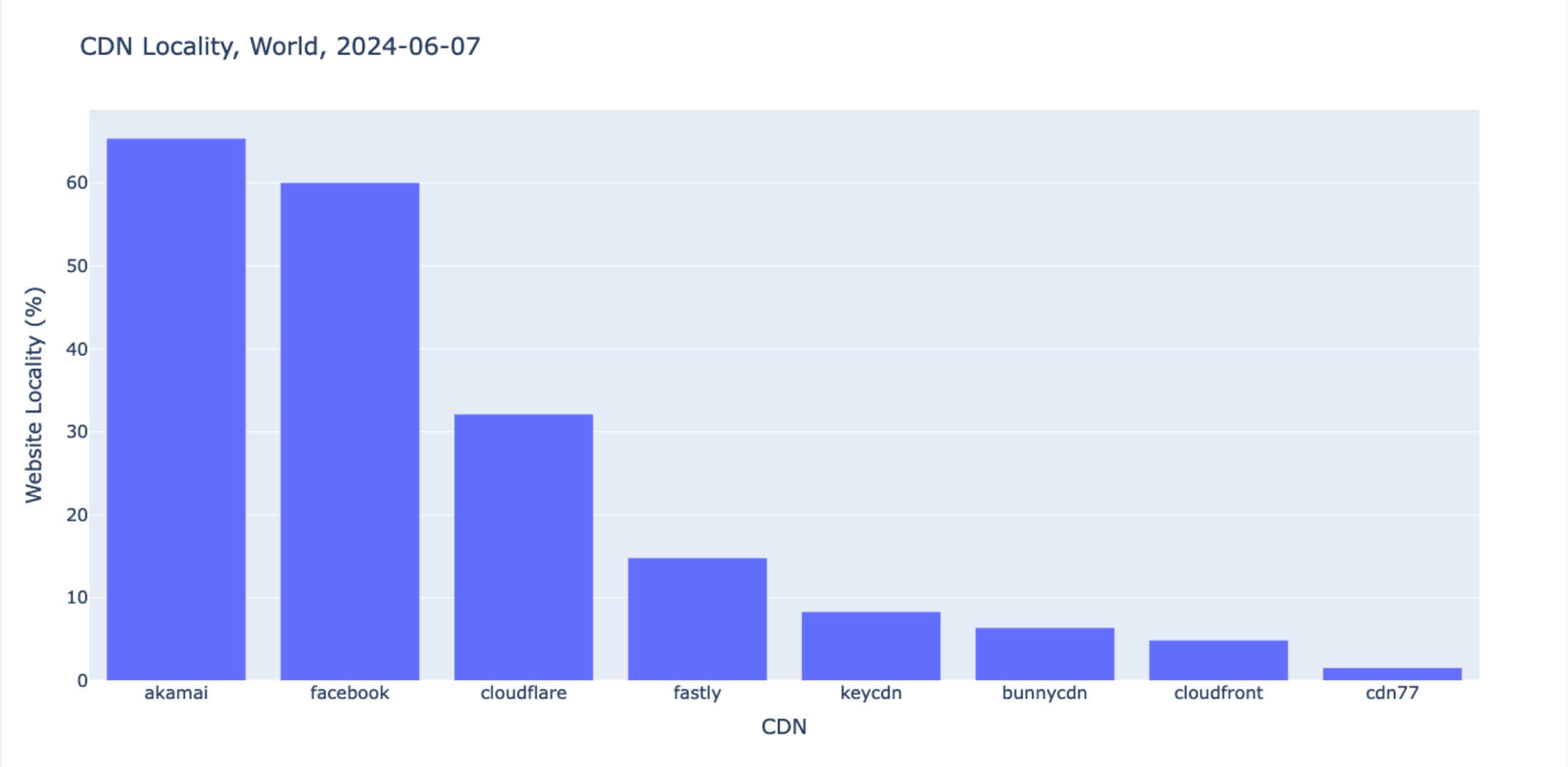
# Preliminary Results – Regional Locality



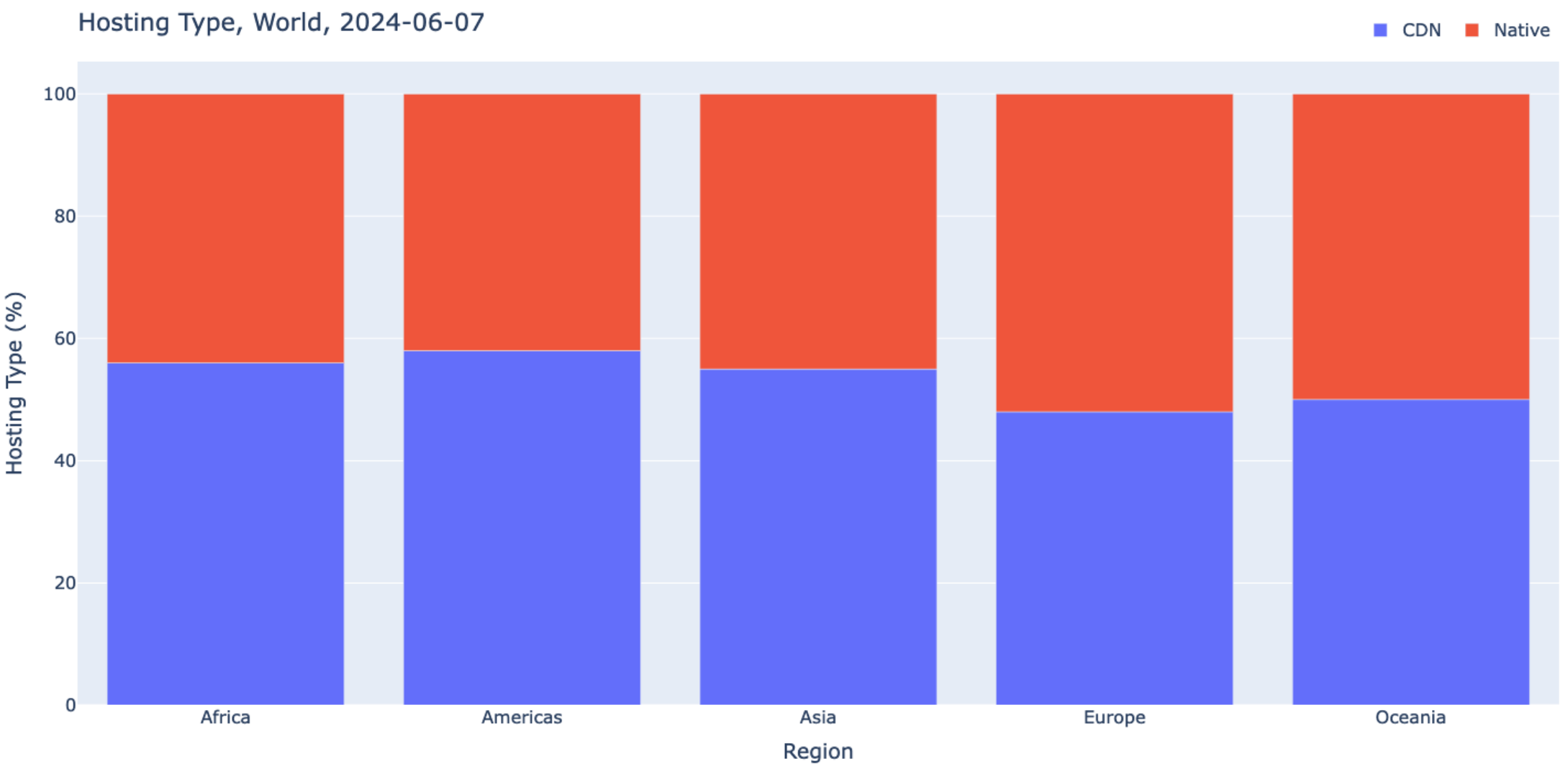
# Preliminary Results – Global Locality



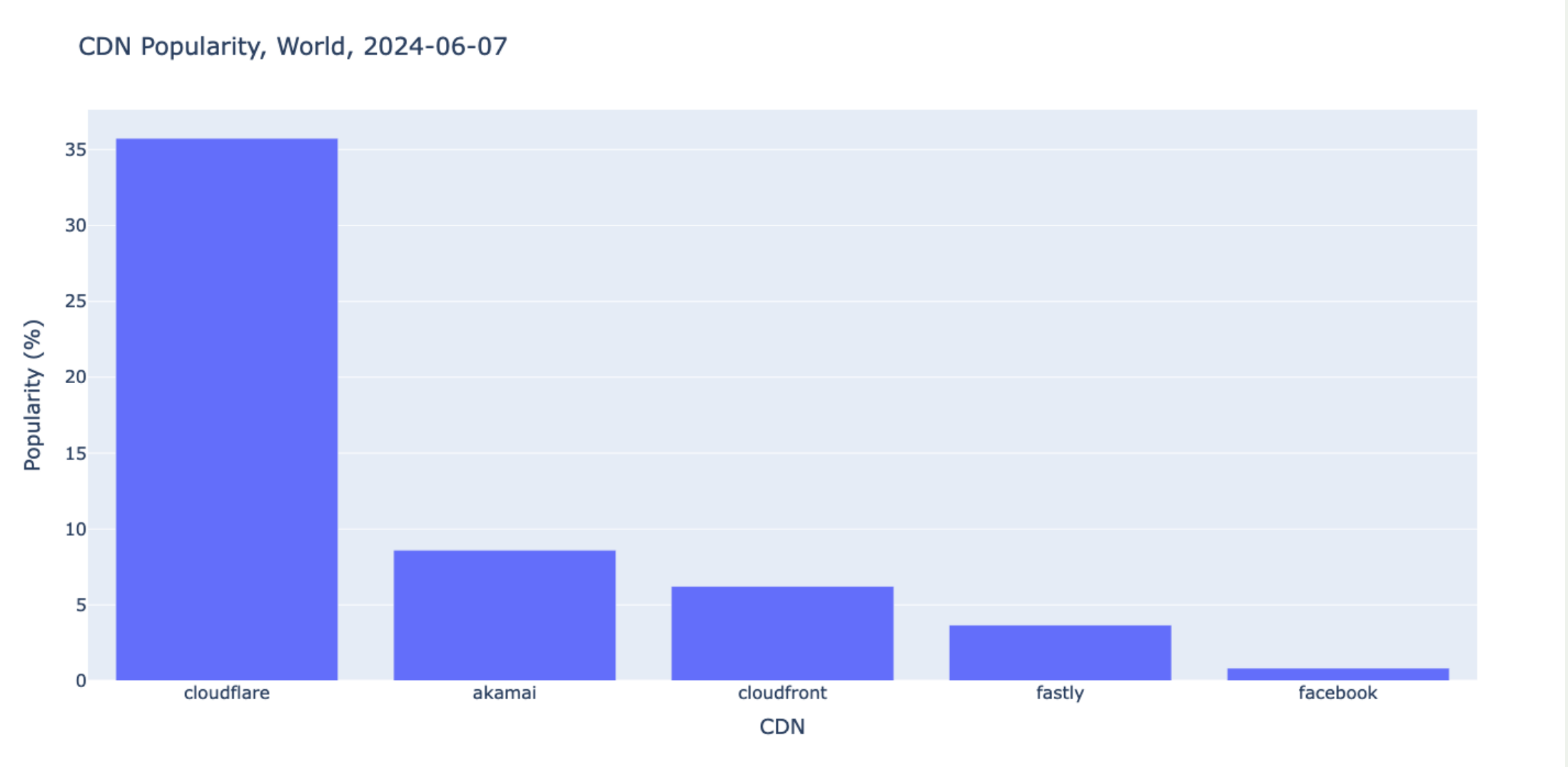
# Preliminary Results – CDN Locality (World)



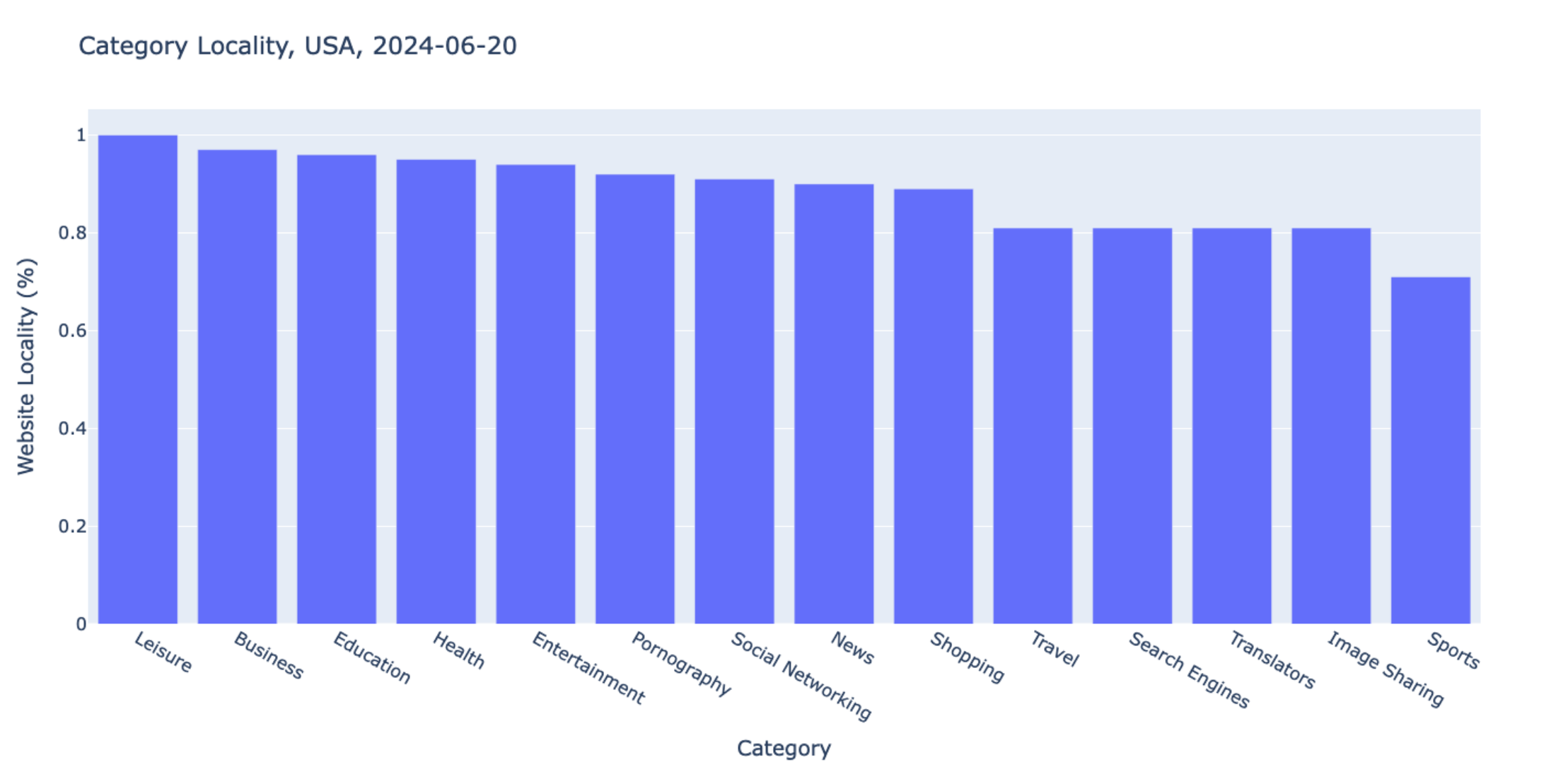
# Preliminary Results – Hosting Type (World)



# Preliminary Results – 5 Most Popular CDNs (World)



# Preliminary Results – Categories Locality (USA, 100 Sites)



# 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

# Summary

Local traffic is (often) beneficial

Conducting global Internet traffic locality measurements

Expanding methodology

Any questions?

Check out the **paper**

Learn about the
**50/50 Vision**