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

Abstract

We present the first large-scale analysis of the adoption of third-party serving infrastructures in government digital services. Drawing from data collected across 61 countries spanning every continent and region, capturing over 82% of the world's Internet population, we examine the preferred hosting models for public-facing government sites and associated resources. Leveraging this dataset, we analyze government hosting strategies, cross-border dependencies, and the level of centralization in government web services. Among other findings, we show that governments predominantly rely on third-party infrastructure for data delivery, although this varies significantly, with even neighboring countries showing contrasting patterns. Despite a preference for third-party hosting solutions, most government URLs in our study are served from domestic servers, although again with significant regional variation. Looking at overseas located servers, while the majority are found in North America and Western Europe, we note some interesting bilateral relationships (e.g., with 79% of Mexico's government URLs being served from the US, and 26% of China's government URLs from Japan). This research contributes to understanding the evolving landscape of serving infrastructures in the government sector, and the choices governments make between leveraging third-party solutions and maintaining control over users' access to their services and information.

CCS Concepts

• **Applied computing** → **E-government**; • **Networks** → **Location based services**; *Network measurement*; *Public Internet*; • **Social and professional topics** → Governmental regulations.

Keywords

Government, Third-party, Centralization, Measurement, Hosting, Cross-border Dependency

ACM Reference Format:
Rashna Kumar, Esteban Carisimo, Lukas De Angelis Riva, Mauricio Buzzone, Fabián E. Bustamante, Ihsan Ayyub Qazi, and Mariano G. Beiró. 2024. Of Choices and Control – A Comparative Analysis of Government Hosting. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24), November 4–6, 2024, Madrid, Spain.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3646547.3688447

1 Introduction

Digital transformation has fundamentally changed government communication, establishing new channels for disseminating policies and information while providing citizens with direct access to essential services [88, 89]. The importance of digital government is evident in cases like federal websites in the US, which attract nearly two billion visits monthly and result in approximately 80 million hours of public interaction [35], and in the Asia-Pacific region where 77% of citizens primarily use a digital platform to access government services [21]. This transformation underscores the need for understanding the infrastructure behind public-facing government websites.

The transition from on-premise servers to third-party infrastructure is a widespread trend across numerous sectors [1, 3, 29, 38, 50, 53, 62, 63, 92], as highlighted by recent studies. For governments, this shift presents a particularly challenging dilemma. While third-party providers offer specialized content delivery solutions with benefits such as flexibility, scalability, and enhanced security, they also introduce the potential risks of multi-tenancy [2, 16, 48], centralization [1, 7, 8, 39, 40, 58, 92], and a lack of control over data

[*] Also with CONICET.

IMC '24, November 4–6, 2024, Madrid, Spain
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0592-2/24/11
https://doi.org/10.1145/3646547.3688447

placement [14], raising critical considerations for governmental authorities.

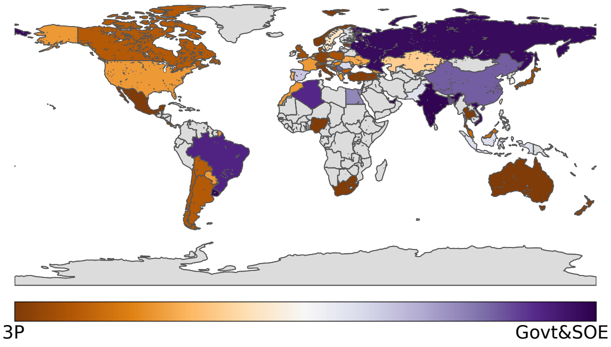


Figure 1: Brown-shaded countries have the majority of their traffic served by third parties (3P), while purple-shaded from government or State-Owned Enterprises (SOE).

In this paper, we present the first comprehensive study of hosting models employed by public-facing government digital services. Our analysis draws on data from 61 countries, covering every continent and representing over 82% of the world's Internet population. We identify government-related sites within these countries, collect resources from the landing pages and recursively crawl internal pages up to seven levels deep [79]. Our dataset comprises over 1 million unique resources, providing a broad and detailed snapshot of government digital service hosting. Building on this dataset, we conduct an extensive measurement study to analyze government hosting strategies, cross-border dependencies, and the level of centralization in government web services.

Among the key findings of our study, we highlight:

- Governments predominantly rely on third-party infrastructure for data delivery, using them to deliver 62% of URLs and 53% of bytes.
- The adoption of third-party providers varies significantly across and within regions, as illustrated in Fig. 1. For instance, in North America, third-party providers account for 68% of government bytes, whereas in South Asia, this reliance drops to just 5%. Similarly, neighboring countries, such as Argentina and Uruguay, show contrasting patterns with Argentina relying 90% and Uruguay 2% on third-party providers.
- Despite a preference for third-party hosting solutions, 87% of government URLs in our study are served from domestic servers, although with significant regional variation. For example, South Asia, East Asia and Pacific, and North America deliver less than 10% of their government URLs from international servers. In contrast, Sub-Saharan Africa relies on international servers for 48% of its government URLs.
- Of the servers located abroad, 57% are in North America and Western Europe. Our analysis reveals interesting bilateral relationships with 40% of New Zealand's government URLs being served from Australia, 79% of Mexico's URLs from the US, and 26% of China's URLs from Japan.
- Unlike other sectors [50, 53], global centralization on third-party providers appears to be more concentrated, with Cloudflare serving 49 different governments, nearly twice as many as the following two providers (Microsoft and Amazon).
- Diversification appears correlated with reliance on third-party providers. For example, 63% of countries that rely mainly on government infrastructure serve the majority of their content from a single network. In contrast, only 32% of countries that depend primarily on Global Providers rely on a single network.

We addressed several challenges throughout the study (§3). For starters, identifying and classifying governments' websites is complicated; while some governments adopt common strings for their sites (e.g., *.gov*), the practice is far from universal (e.g., defensie.nl, parlement.ma, orniss.ro, landkreistag.de). Key to our analysis is correctly identifying the location of servers hosting a country's government resources which requires accessing resources from within the specific country. Anycast services further complicate this task.

In summary, we make the following main contributions:

- We describe a methodology to characterize government approaches to domain hosting, determining their service infrastructure and location (§3).
- We apply this methodology to gather a comprehensive dataset of government URLs and annotated networks across 61 countries (§4).
- We report on the first extensive measurement study to analyze government hosting strategies (§5), cross-border dependencies (§6), and the level of centralization (§7) in government web services.
- We will make our dataset available upon request.

Overall, our research contributes to the understanding of the evolving landscape of serving infrastructures in governments and provides valuable perspectives on the choices between leveraging third-party solutions and managing control over access to government services and information.

2 Governments' Digital Dilemma

Our work is motivated by two concurrent trends: (1) a global shift toward third-party providers [57, 61] and (2) a push towards both digital government, digital sovereignty and strategic autonomy [23, 54].

The transition from on-premises to third-party providers is a global trend that has reshaped the Internet during the last decade. Large providers offer access to computing resources in data centers across various regions, flexibility in resource allocation, high service availability, and comparatively low capital and operational expenses [9, 31, 41]. This shift, however, has also led to increasing concerns about Internet consolidation and centralization. The 2019 Global Internet Report [1] provides an early overview of this trend in every aspect of the Internet economy, from access provision to service infrastructure and applications. It argues that while consolidation is often seen as an expected result of maturing markets and industries, the combination of society's increased dependency on the Internet, business agility, and the almost total lack of regulation is leading to a handful of platforms in control of much of the Internet's functionality and interoperability. Since that report, several efforts have explored this trend [3, 29, 38, 50, 53, 62, 63, 92] and its economic, political and reliability implications [1, 7, 8, 39, 40, 58, 92].

Somewhat concurrently, governments worldwide began considering cyber sovereignty – from data sovereignty to digital privacy and security and Internet governance – out of concern with foreign interference and dependence on foreign governments, foreign tech platforms, and infrastructure [18, 22, 71, 82**?**]. While early discussions followed the 2013 Snowden revelations of surveillance [60], different initiatives worldwide have focused on the issue in the context of geopolitical and economic tensions and the growing recognition of the Internet as critical infrastructure [11, 27, 69, 76, 78].

These concerns have motivated the development of legal frameworks, including the European Union's General Data Protection Regulation (GDPR) [15], in effect since May 2018, California's Consumer Privacy Act (CCPA) [68], which became effective in January 2020, and Brazil's General Data Protection Law (LGPD) [55], operational since September 2020. Together, these frameworks reflect a concerted effort to protect and manage data within the respective jurisdictions, highlighting the increasing importance of data sovereignty.

At the same time some cloud providers have began offering solutions tailored to specific governments. For example, Amazon Web Services [25] and Microsoft Azure [26] have developed solutions tailored to meet the requirements of the U.S. government. Nevertheless, for the majority of countries, third-party services are foreign-based, forcing them to strike a balance between external expertise and maintaining sovereign control over their digital assets. Our work aims to empirically characterize the various ways in which governments navigate and resolve this emerging dilemma.

3 Methodology

To characterize governments' approaches to domain hosting, we (*i*) collect government sites, and (*ii*) identify the resources they rely on, excluding those of external contractors. We then determine (*iii*) the serving infrastructure of those resources and (*iv*) their location. The following paragraphs describe this process in detail.

3.1 Gathering Government Websites

The first step in our methodology is to compile a comprehensive list of government sites. In this study, we focus specifically on federal-level (or equivalent) resources, including various segments of the federal administration (e.g., the presidency, ministries, and secretaries), federal agencies, often referred to as decentralized agencies (e.g., the US National Science Foundation and the US Internal Revenue Service) and state-owned enterprises. To consider State-Owned Enterprises (SOEs), we follow the International Monetary Fund (IMF) guidelines and only include companies where the federal government holds more than 50% of the shares [28].

This step requires searching through a country's government sources that may provide insights into the organizational structure, identifying digital directories and authoritative resources that provide details on these structures and links to corresponding government sites. As this information is typically in the country's official languages, we rely on translation tools for this part of the process.

3.2 Scraping Government Websites

We scrape the collected government websites to identify the resources they rely on. For this, we use Selenium [77], a web automation tool, to capture the URL of each resource that constitutes the queried websites, which are then consolidated into an HTTP Archive (HAR) file. We move beyond the landing pages using the collected HAR files to recursively navigate internal pages up to seven levels deep, a threshold informed by previous work [79].

The geographic location of our vantage points can impact website rendering, replica selection, or determine resource accessibility, with some sites restricting access to non-domestic devices.[1] To avoid these and other potential problems, we rely on different VPN services including NordVPN [66], Surfshark [85], Hotspot Shield [34], to access these sites from within the target country.

3.3 Internal Government URLs

As we scrape seven levels deep into a government domain, we run the risk of leaving the government domain (e.g, into an external contractor's site). After completing data collection we identify internal government URLs and filter out non-government ones following the steps summarized in Table 1.

Approach	Method
Government TLDs	All domains including `.gov`, `.govern`, `.government`, `.govt`, `.mil`, `.fed`, `.admin`, `.gouv`, `.gob`, `.go`, `.gub`, `.guv`
Domain Matching	If the hostname of the internal page aligns with those listed in the government websites section (§3.1).
SAN	If the hostname is included under domains specified as Subject Alternative Names (SANs) in the TLS certificates of landing pages

Table 1: Steps of the methodology to identify government domains.

We first label as government resources those with domains under government top-level domains (TLDs). We adopt the pattern-matching rules defined by Singanamalla et al. [79], which account for the different government TLDs that vary based on each country's definitions and official languages. This includes TLDs such as `.gov`, `.gouv`, `.gob`, and `.go`, among others, as listed in Table 1.

We then identify government resources that do not fall under government top-level domains (TLDs), either because the country does not utilize government TLDs or chooses not to use them for some agencies or state-owned enterprises (§3.1). If the hostname of an internal page matches the hostname of any of the sites comprised

[1] For instance, Mexico's Taxpayers Defense Attorney (in Spanish Procuraduría de la Defensa del Contribuyente, www.prodecon.gob.mx).

in our list of government websites, we classify it as a government hostname.

Finally, we identify government resources included under domains specified as Subject Alternative Names (SANs) in the TLS certificates of landing pages [12]. When the hostname of an internal page appears in the SANs list of landing pages, we manually verify that the hostname corresponds to a government resource. This last step allows us to select additional government-affiliated resources that may not be directly evident through their domain names or top-level domains (e.g., orniss.ro, energia-argentina.com.ar). At this stage, any hostnames that cannot be verified as government hostnames are discarded from our analysis.

3.4 Identifying the Serving Infrastructure

We identify the serving infrastructure utilized by government hostnames. This involves determining the IP address, Autonomous System (AS) number, organization, the registered location and the geolocation of the serving infrastructure. Table 2 shows an example of the information we collect for a government hostname in Uruguay.

Field	Value
URL	`www.gub.uy`
IP address	179.27.169.201
ASN	6057
Organization	Administracion Nac. de Telecom.
Registration	Uruguay
Geolocation	Uruguay

Table 2: An example of the information of serving infrastructure that is collected for each government resource.

To obtain registration and topological data on government website infrastructure, we connect to a VPN within the country and resolve all government hostnames to their IP addresses. Once we have the IP addresses, we determine the corresponding AS number, organization and country of registration using public WHOIS services managed by organizations responsible for IP address registration.

We then determine whether content is served from on-premise infrastructure within government-operated networks. While a recent study has made progress in identifying state-owned Internet providers [13], there is no dataset with annotations of government networks. We thus manually examine the entity behind all identified ASes to determine government ownership. It is important to differentiate between state-owned Internet providers – government-controlled companies participating in the Internet market – and government networks used exclusively by government institutions.

We combine various data sources to identify government ownership of networks. We examine PeeringDB records, searching for any indicator of government ownership, which may be revealed in the network's name, associated organization, or note, as in the entry of AS26810 indicating the organization as "U.S. Dept. of Health and Human Services". We also leverage the website reported on PeeringDB records and investigate whether the associated website reveals any information that could indicate a connection with the government. Given the limitations of PeeringDB's coverage, we use WHOIS records to complement our classification. This involves querying WHOIS databases to check if the organization's name refers to the government (e.g., ministry) or the domain of the contact person's email is linked to a government domain (e.g., ".gov"). Finally, for cases where we are unable to find direct matches, we resort to Google searches. We utilize domain information extracted from WHOIS records to search for these companies' websites. This process also allows us to identify domains of state-owned enterprises that may not always be identifiable as government domains (e.g., AS27655 - Yacimientos Petrolíferos Fiscales).

3.5 Server Geolocation

The last step of our process consists of determining the geographic location of the infrastructure serving government websites. Given the limitations of existing geolocation heuristics and databases, we outline our specific methodology to address these challenges.

Step #1: Geolocation databases. We first query IPInfo [45], a widely-used open geolocation database, with the addresses of all the collected government hostnames. Darwich et al. [17] report that 89% of the geolocation targets in IPInfo have an error of less than 40 km (i.e., within a city).

Step #2: Identifying Anycast addresses. IP Anycast challenges latency-based geolocation. To determine if a server address is anycast, we rely on a recent data snapshot from MAnycast2, generated based on the idea of using anycast IPs as VPs to launch active measurements to candidate anycast destinations [80].

Step #3: Verifying country-level geolocation. To enhance the accuracy of our geolocation data, we deploy active-probing measurements to validate the reported geolocations.

For anycast addresses, we select five RIPE Atlas probes situated in the vantage country to send three pings to anycast addresses and calculate the minimum latency to each address. Our methodology integrates active measurements with the country's road infrastructure data to derive a threshold that determines whether a server address is located within a country. When the latency to a specific server address is less than the threshold, we conclude that the anycast address has servers within the country. Anycast addresses with latencies higher than this threshold are excluded from the analysis.

For unicast addresses, we also use five RIPE Atlas probes in the country assigned by IPInfo to send pings to each reported address in that country. To confirm the server's location reported by IPInfo, we calculate the minimum latency to each address and, following [4], check if the latency to a specific server address is less than this threshold calculated using the country's road infrastructure data. Discrepancies trigger additional verifications for unicast addresses, explained in Step #4.

Given the different shapes and sizes of countries, rather than settling for a single global threshold, we determine a per-country threshold based on the intercity road distance between the two furthest cities in that country and convert this distance into latency values.

Step #4: Geolocating Unicast Addresses. To verify the location of remaining unicast addresses we use CAIDA's HOIHO methodology [59], which leverages geolocation hints found in PTR DNS records, with additional regular expressions (e.g., for NTT). We also consult the cached results from RIPE's IPmap [74] and, if not available, we resort to active probing following a single-radius approach for geolocation.

4 Government Hosting Dataset

To capture a global view of trends in government hosting, we select a sample of 61 countries across all world regions, and apply our methodology for identifying government approaches to domain hosting. We first describe our criteria for including a country in our sample before providing general statistics on the collected dataset.

4.1 A Sample of Countries

We create a representative dataset encompassing countries from all regions worldwide. Regional divisions allow us to identify global and regional trends for governments' digital approaches. We set criteria for sampling countries across these regions, balancing our scope with technical and logistic limitations (such as the absence of verifiable VPN servers[2] or insufficient information on e-governments).

World's Regional Slicing. To explore regional patterns in government digital strategies, we rely on the World Bank's regional division [10]. This division groups countries into seven regions: North America (NA), Latin America and the Caribbean (LAC), Europe and Central Asia (ECA), North Africa and the Middle East (MENA), Sub-Saharan Africa (SSA), South Asia (SA), and East Asia and Pacific (EAP).

Country Selection Criteria. Covering each region, we select countries that, combined, capture a wide range of key development indices, specifically: (1) the E-Government Development Index (EGDI) [64], (2) the Human Development Index (HDI) [72], and (3) the International Telecommunication Union/World Bank Internet Penetration rates [83]. This combination of indices allows us to capture a broad spectrum of countries in various stages of development and digital advancement. We integrate these indices at a regional level and select countries from five different quintiles.

While aiming for uniform coverage across these quintiles, we encounter some limitations. Specifically, the challenge is set by the lack of commercial VPN services in countries from the lower quintile, particularly in regions like Sub-Saharan Africa and Latin America and the Caribbean.

Our final selection of 61 countries from across the globe includes 2 countries from North America, 8 from Latin America and the Caribbean, 29 from Europe and Central Asia, 5 from North Africa and the Middle East, 2 from Sub-Saharan Africa, 3 from South Asia, and 12 from East Asia and Pacific (EAP). These countries combined represent 82.70% of the global Internet population. To access government URLs across these countries, we use 3 VPN services: NordVPN (49), Surfshark VPN (10), and Hotspot Shield VPN (2).

Table 9 in Appendix B provides a list of the countries in our sample, including their regions, index values, the specific VPN services used for each country, and additional statistics.

4.2 Dataset Characteristics

We apply our methodology to the set of countries in our sample. Table 3 offers a high-level overview of the extent and scope of our data collection.

Category	Element	Value
Government Websites	Landing URLs	15,878
	Internal URLs	1,017,865
	Total Unique URLs	1,033,743
	Total Unique Hostnames	13,483
Serving Infrastructure	ASes	950
	Govt ASes	347
	Unique IP addresses	4,286
	Anycast addresses	433
	Countries with servers located	68

Table 3: Landing URLs, unique hostnames and unique URLs in our dataset.

Government Websites. The dataset includes 15,878 unique landing pages from governments of 61 countries, and 1,017,865 internal government URLs obtained through scraping across seven levels. In total, the dataset comprises 13,483[3] unique hostnames and 1,033,743 distinct URLs. The vast majority of URLs, 84%, were collected directly from the landing pages, with 95% obtained from one additional level below the landing page. Detailed statistics on the collected government websites for each country are provided in Table 8 in Appendix C.

Internal Government URLs. We apply a set of heuristics (Table 1) to identify government URLs and filter out non-government ones from the set of URLs obtained. This step identified 285,767 (27.6%) internal government URLs using the government TLDs, 745,358 (72.1%) using the domain-matching approach and 2,618 (0.3%) using SANs.

Serving Infrastructure. We identified 950 ASes connecting to 4,286 server addresses associated with 13,483 hostnames. We discovered 347 (36.5%) of these ASes are operated by government entities.

We localize the serving infrastructure of the 4,286 addresses. MAnycast2 identified 433 (10.10%) of them as anycast addresses. Active-probing confirmed that 361 anycast addresses (83.37% of all anycast addresses identified) are within the country's borders. We excluded the remaining 72 anycast addresses from the analysis due to insufficient confidence in their location.

[2] We gained confidence in the claimed VPN locations of the countries in our set, by validating the VPN vantage points' IPs using the geolocation approach described in (§3.5)

[3] Note that the number of unique hostnames is less than the number of unique landing pages. This is because landing pages can include URLs like https://www.gov.br/secretariageral/pt-br, https://www.gov.br/abin/pt-br, representing different pages with the same hostname.

Type of Address	AP	MG	UR
Unicast Addresses	0.41	0.57	0.02
Anycast Addresses	0.83	0.00	0.17

Table 4: Fraction of unicast and anycast addresses validated by Active Probing (AP) and Multistage Geolocation (MG), or Unresolved (UR).

From the 3,853 unicast addresses, IPInfo identifies 3,349 addresses (86.92%) in the same country as the government they are serving and 504 unicast addresses (13.08%) outside the country borders. To increase our confidence, we tried to confirm IPInfo geolocation. Through active-probing, we confirmed the location of 40.77% (1,571) of these addresses. Through a multistage geolocation approach (§3.5) we confirmed the geolocation of an additional 2,198 addresses. In total, we confirmed 3,769 (97.8%) of all unicast addresses. We exclude 84 instances where the geolocation obtained at this stage conflicts with IPInfo. Table 4 summarizes the output of this validation process for unicast and anycast addresses spanning 68 countries.

5 Trends in Government Hosting

Building on the collected dataset, in this section, we explore global and regional trends in government domain hosting and compare them with trends among popular websites. We close the section examining the similarities in governments serving strategies across the countries in our study.

5.1 Global Trends

We first take a global perspective, exploring governments' preferences in choosing the serving infrastructure powering their websites. *Do governments prefer on-premises or third-party hosting?* For governments opting for third-party providers, we further explore their preferences towards global, regional, or local providers.

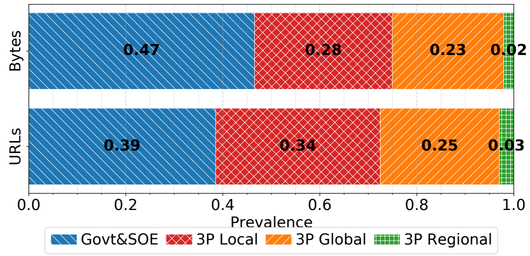


Figure 2: Global fraction of URLs and Bytes served by each provider category.

We examine the adoption of on-premises solutions, labeled *Government and State-Owned Enterprises* (Govt&SOE 🔵), versus third-party providers. We categorize third-party (3P) services into three groups: (1) Local (3P Local 🔴), (2) Regional (3P Regional 🟢), and (3) Global providers (3P Global 🟠), with 3P Global 🟠 defined as networks that serve governments across multiple continents, and 3P Local 🔴 as those registered in the same country as the government they serve. The remaining category, 3P Regional 🟢, includes networks registered outside the country they serve, but that do not span beyond one continent.

Using this classification, Figure 2 illustrates the global prevalence of each server URL category and quantifies the content by aggregating the total bytes of government URLs to account for variations in URL sizes.

Overall, governments show a preference for 3P infrastructure for data delivery, using them to deliver 62% of URLs and 53% of bytes, compared to only 39% of URLs and 47% of bytes hosted by Govt&SOE 🔵. When focusing on the categories of 3P, the figures show a more balanced reliance on Govt&SOE 🔵, 3P Global 🟠, and 3P Local 🔴 although with a preference for Govt&SOE 🔵 for bytes.

Interestingly, the analysis reveals that governments rarely consider 3P Regional 🟢, preferring to depend on their own infrastructure, collaborate with global partners, or engage with local providers. Utilizing their own infrastructure provides the maximum degree of control, but involves capital and operational expenditures. Global partners, on the other hand, offer the benefit of mature, large-scale infrastructure, while local providers may combine the benefits of third-party expertise and specialization under government jurisdiction.

Governments vs. Topsites. To compare the hosting strategies of governments and popular websites, we select a subset of 14 countries (described in Table 6 in Appendix D), including two from each region from different digital development strata and compare the adoption of third-party providers between those countries' governments and regional popular sites. We use Google's Chrome User Experience Report (CrUX) to compile a list of popular websites in these countries. To mirror our methodology, we employ VPNs and limit our scraping to resources one level beyond the landing pages. This depth limit is due to the intensive nature of deeper scraping of commercial sites (i.e., particularly broad trees) and the observation that a significant majority (95%) of government URLs are found just one level down. By leveraging the methodology described in (§3.4) and (§3.5), we then determine the serving infrastructure and geolocation of the organizations responsible for the infrastructure serving these top sites in each selected country. We also identify the fraction of non-government topsites that use either on-premise or third-party solutions to deliver content. This mirrors our government site analysis and redefines categories as (1) self-hosting, (2) global, (3) local, and (4) foreign providers. To identify self-hosted solutions, we use a heuristic from previous research [50, 53]. A detailed explanation of this methodology can be found in Appendix D.

This comparison (Fig. 3) shows that top sites predominantly rely on 3P Global 🟠, using them to deliver 78% of URLs and 74% of bytes, more than twice as commonly as government sites with 32% for URLs and 16% for bytes. In contrast, on-premise infrastructure is much more prevalent across governments, with an average of 46% of URLs and 69% of bytes, compared to only 18% and 17% for top sites. The difference suggests the relative weight of considerations, beyond market forces, behind government hosting decisions.

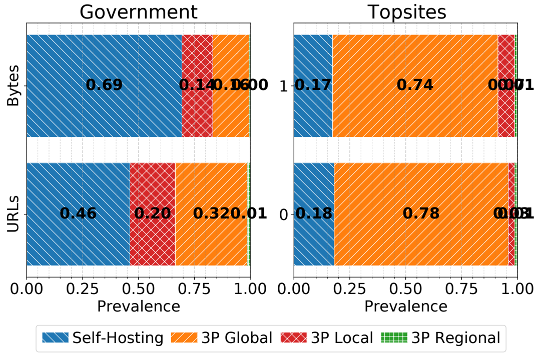
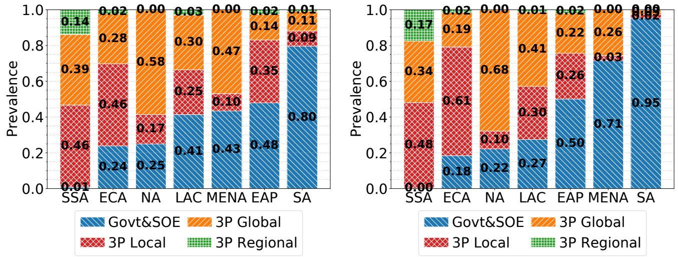


Figure 3: Comparison of self-hosting (on-premises) and third-party hosting between government websites and top sites within our selected subset of countries.

5.2 Regional Trends

In this section, we replicate our previous analysis now using the World Bank's regional division (§4.1) to investigate unique patterns or singularities that might exist in different regions. This regional-focused approach provides valuable insights into how factors like shared geography[4] and common cultural backgrounds may influence government decisions regarding digital infrastructure.



(a) URLs

(b) Bytes

Figure 4: Fraction of URLs and Bytes served by each provider category per region.

We assess both on-premises and third-party providers using the same four categories (Govt&SOE ●, 3P Global ●, 3P Local ● and 3P Regional ●) from a regional perspective. Figure 4 illustrates the regional prevalence of each category, represented separately for URLs (Fig. 4a) and bytes (Fig. 4b).

Both perspectives consistently reveal a significant variation in adopting Govt&SOE ● or 3P infrastructures across different regions. For instance, in regions like South Asia (SA) and North Africa and the Middle East (MENA) most bytes originate from government infrastructures (95% and 71%, respectively). In the case of North America, most bytes and URLs originate from 3P Global ● (68% and 58%, respectively). Sub-Saharan Africa (SSA), on the other hand, delivers most of their URLs and bytes through a combination of 3P Global ● and 3P Local ● infrastructures (85% and 82%), highlighting the complexity and variability in regional hosting strategies.

5.3 Countries' (dis)Similarities

We conclude our evaluation of hosting trends by examining the similarities in governments' serving strategies across the countries in our study.

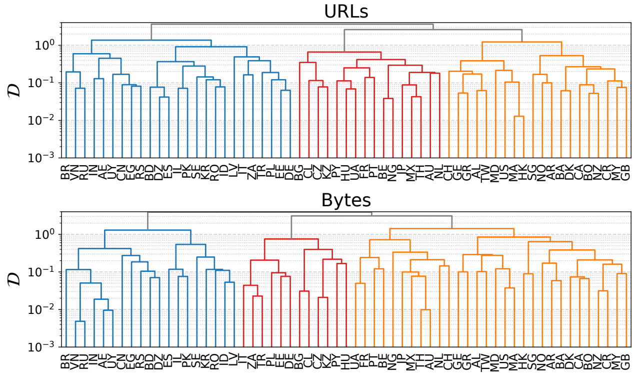


Figure 5: Similarities in governments' serving strategies across countries.

We use the same four categories of government hosting options and look at both URLs and bytes. The distribution of URLs and bytes across these sources creates a unique pattern, which represents the signature of a government's digital serving strategy. Our goal is to identify commonalities in these signatures across different countries.

We apply Hierarchical Agglomerative Clustering (HCA) using the Ward distance on a matrix that includes these four categories across countries, each represented by a row. This process results in the two three-branch dendograms shown in Figure 5. Each branch corresponds to the principal type of hosting sources (e.g., Govt&SOE ●).

The analysis shows the absence of strong regional patterns in government hosting strategies. For example, within the Southern Cone, Argentina, Brazil, and Chile each adopt a different approach, predominantly relying on 3P Global ●, Govt&SOE ●, and 3P Local ●, respectively. A similar diversity is observed in Southeast Asia, where Malaysia primarily depends on 3P Global ● in contrast to Indonesia's reliance on Govt&SOE ●. Even more remarkable is the situation within the European Union, a region bound by a common legislative framework yet displaying varied hosting preferences. For instance, Spain, Italy, and the Netherlands each show a distinct inclination, with major dependencies on Govt&SOE ● (64%), 3P Local ● (93%), and 3P Global ● (41%), respectively.

At the same time, it reveals similarities in the hosting strategies of countries from different regions despite having no apparent connections. For example, Brazil, Vietnam, and Russia share the same sub-tree due to their hosting similarities. We note, however, the challenges of generalizing from the observed trends and similarities. Apparent similar hosting practices may be driven by significantly different policies. In this case, Brazil's hosting choices may be the

[4] Geographical considerations affect choices to host content with providers whose serving infrastructure is distant from a particular country.

result of a comprehensive GDPR-like regulation, known as the LGPD [55], whereas Russia's [70] and Vietnam's [56] hosting models may respond to a focus on data localization and state control. France and Canada, though both predominantly rely on global providers (3P Global ●) for hosting, differ significantly in the extent of their reliance, with 42% and 79% of bytes, respectively, sourced from these providers. Likewise, Uruguay and Indonesia, primarily depending on government and state-owned enterprises (Govt&SOE ●), show considerable variance in their reliance, with 98% and 58% of bytes, respectively, attributed to government sources. These examples highlight the diverse approaches and degrees of dependency on specific hosting types, even among countries with similar strategies.

Key Findings:

- Governments predominantly use 3P infrastructure for data delivery, using them to deliver 62% of URLs and 53% of bytes. This adoption changes across regions with, for example, SA (95%) and MENA (71%) hosting most bytes on Govt&SOE ●, NA (68%) on 3P Global ● and SSA (82%) on a combination of 3P Global ● and 3P Local ●.
- Countries within the same region show diverse hosting preferences despite similar strategies. For instance, within the Southern Cone, Argentina, Brazil, and Chile each adopt a different approach, predominantly relying on 3P Global ●, Govt&SOE ●, and 3P Local ●, respectively.

6 Hosting Registration and Server Locations

The previous section focuses on government preferences between on-premise and third-party hosting. Even when opting for third-party service, a government could have its content hosted within its jurisdiction. In this section, we explore this aspect of hosting, specifically answering: *What are the jurisdictions where the organizations serving government content are registered? What is the location of the servers hosting the content of government sites?*

We explore these starting with a global overview (§6.1), followed by a regional perspective (§6.2), and concluding with an analysis of cross-country dependencies (§6.3).

6.1 Global Trends

We examine the country of registration and the location of the servers hosting the government URLs in our dataset. Figure 2 categorizes this data globally into two distinct groups: (1) Domestic ●, and (2) International ●. While a majority of the URLs, to different extents, are served from servers located within the country (87%) and from addresses allocated to domestic organizations (77%), *23% of URLs are served from internationally registered organizations and 13% are served from servers located outside the country.* Note that foreign-registered organizations of domestically provided services may still need to comply with local legislation.

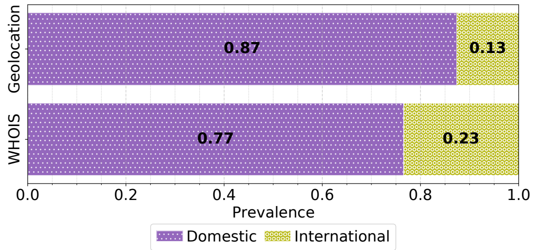


Figure 6: Fraction of Government URLs registered and served by Domestic or International Organizations.

Governments vs. Topsites. As in the previous section, we compare the hosting strategies of governments and popular websites, focusing on their use of domestic and international hosting solutions for the 14 selected countries.

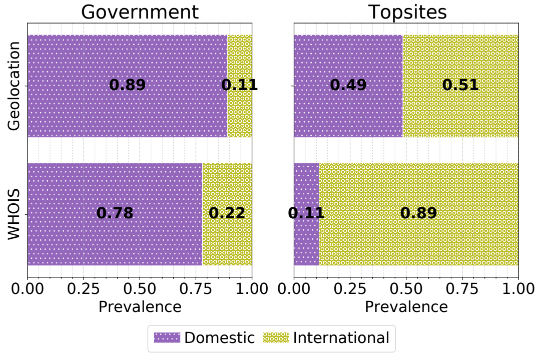


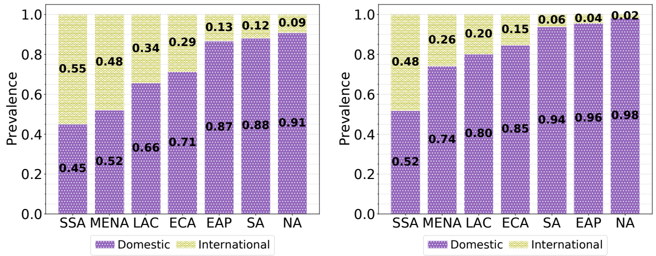
Figure 7: Comparison of domestic and international hosting between government websites and top sites within our selected subset of countries.

Figure 7 shows this comparison, displaying: (1) the country of registration of the organization and (2) the server locations serving the URLs in our dataset for this analysis. This comparison (Fig. 7) shows that governments predominantly opt for domestic hosting, with 78% of their URLs served by in-country registered organizations and 89% hosted within their borders. In contrast, popular websites prefer domestic hosting less; only 11% of their URLs are from domestically registered organizations, and just 49% of URLs are served from servers within their borders. This comparison highlights the different priorities between government entities, which favor control and jurisdictional autonomy, and popular websites that follow a more varied approach to digital service hosting.

6.2 Regional Trends

At a regional level, we analyze the country of registration and the physical location of servers hosting government URLs in our dataset.

Figure 8 presents this analysis, dividing organizations into two main categories: (1) Domestic and (2) International, with separate plots for their countries of registration (Fig. 8a) and server locations (Fig. 8b).



(a) Country of Registration

(b) Server Location

Figure 8: Fraction of Government URLs registered and served by Domestic or International Organizations per region.

While most URLs in all regions are served from servers within their respective countries, the extent of this adoption varies significantly across regions. For example, in North America (NA), 98% of URLs are served domestically, compared to the Middle East and North Africa (MENA), where this drops to 74% and Sub-Saharan Africa (SSA) where the number of URLs hosted in the country drops to 52%. These variations are even more pronounced regarding the nationality of registrations. In North America, 91% of content is hosted by domestic companies, while in East Asia and the Pacific (EAP), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA) and Sub-Saharan Africa (SSA), the percentages of URLs served by companies registered domestically are 87%, 66%, 52% and 45%, respectively. This may be partially explained by the maturity of digital markets in the US and Western Europe, where these third-party providers are registered.

6.3 Cross-Border Dependencies

We now explore the cross-border dependencies of government websites to determine whether there are any preferences across the regions when selecting foreign countries from which this content is served.

Our analysis of cross-border dependencies examines both the country of registration and the location of servers from which governments' URLs are served.

Figure 9 presents this analysis through two circular Sankey diagrams, where countries are grouped using the World Bank's regional division, with one diagram showing the country of registration for these organizations (Fig. 9a) and the other showing the server locations (Fig. 9b). The plots reveal several interesting trends.

Inter-region dependency. This high-level analysis shows a clear trend with most governments largely relying on US-registered organizations in cases of foreign dependence. It also reveals that reliance on servers located abroad is generally confined to the same region; Table 5 shows this through interregional percentages.

There are, however, some notable exceptions, such as the Middle East and North Africa (MENA) region relying on servers in Western European countries and Latin America and the Caribbean (LAC) predominantly depending on servers in the US.

In the case of Mexico and Costa Rica, we observe significant reliance on US-based servers with Mexico hosting 79.22% and Costa Rica 49.70% of their government URLs on servers in the US. In countries like Morocco, Egypt, and Algeria, the percentages of government URLs hosted on foreign servers are 48.38%, 21.1%, and 18.62%, respectively, similarly highlighting a pattern of dependence on international hosting solutions.

In sum, we observe that servers in North America and Europe serve 57% of government URLs crossing their respective country's borders. Brazil stands as the only exception in Latin America and the Caribbean, with only 1.78% of the URLs being served from the US, likely following Brazil's data regulation policy LGPD [55].

Region	%
Europe and Central Asia	94.87
East Asia and Pacific	80.79
North America	59.89
Latin America and Caribbean	3.41
Sub-Saharan Africa	2.95
Middle East and North Africa	0.00
South Asia	0.00

Table 5: Percentage of the cross-border dependencies that remain in the region.

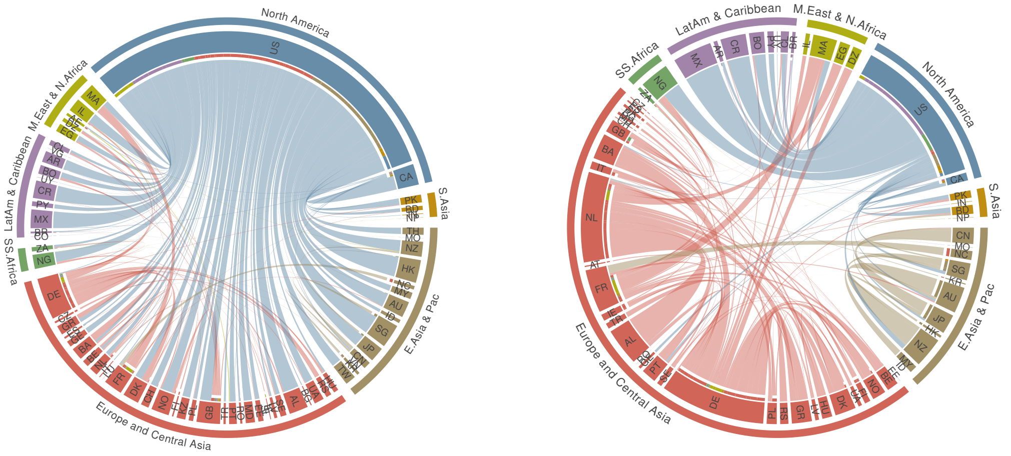
Regional Affinity. When looking at cross-border dependencies within the same region – resources from other countries within the region – we find that South Africa hosts 100% of regional cross-border dependencies in Sub-Saharan Africa, Brazil hosts 85% in Latin America and the Caribbean (LAC), the US 83% in North America (NA), 76%, Japan hosts 60% in the East Asia and Pacific region and Germany accounts for 36% in Europe and Central Asia.

We also find some specific bilateral cases, such as New Zealand and Australia (with 40% of the URLs in New Zealand served from Australia). In general, we observe that 42% of government URLs crossing their respective country's borders are served by servers within the same region.

GDPR Compliance. As part of our regional analysis, we explore compliance of government websites with the General Data Protection Regulation (GDPR) [15]. This EU regulation establishes that digital content within the European Union must be hosted on servers located within the member countries. Focusing on government websites, which might be more sensitive yet more likely to comply with their own regulations, we find a high level of compliance. Our analysis reveals that 98.3% (41,109 / 41,813) of URLs from EU countries are indeed served from servers within the EU's borders, indicating a strong alignment with GDPR requirements in the governmental digital sphere [44].

France and (former) colonies. We find interesting trends involving France with its historical and territorial connections. For instance, Morocco, which was a French protectorate from 1912 to 1956 [91], hosts 29.82% of its government URLs (that belong to 6 unique hostnames e.g., social.gov.ma) on servers located in France. On the other hand, 18.03% of the URLs of the French government are hosted on servers in New Caledonia, a French overseas territory in the southwest Pacific Ocean.

While New Caledonia is technically a part of France, its status is unique: it is not part of the European Union [86], it is an independent



(a) Country of Registration

(b) Server Location

Figure 9: Cross Border Dependencies. Flows represent the fraction of government resources that rely on a foreign country, either because the serving organization is registered there according to WHOIS records (Fig. 9a) or because the server itself is located there (Fig. 9b). Colors represent the region of the foreign country, while the color band connecting the flow to the foreign country represents the region of the source country that relies on it.

member of APNIC [5], listed by the UN as a non-self-governing territory [90], and has been engaged in long-standing discussions with France about independence [87]. Significantly, all URLs of the French government served from this territory are hosted by New Caledonia's state-owned provider, *Office des Postes et des Telecomm de Nouvelle Caledonie* (OPT-AS18200), and belong to the hostname *gouv.nc*. This highlights the complex interplay of historical, political, and technological factors in determining the hosting locations of government digital services.

China and India. China and India, two of the world's largest economies, show contrasting trends. Despite both countries predominantly depending on their domestic and government infrastructures, the extent of their reliance varies. For China, despite historical tensions with Japan [65, 84], we find 26.4% of its URLs hosted by third-party providers in Japan. India, on the other hand, strongly prefers government hosting, with 99.3% of its URLs served domestically. This approach may relate to India's recent efforts to enhance data privacy, as reflected in the Digital Personal Data Protection (DPDP) Act passed in August 2023 [43].

Bilateral relationships and server deployments. The Dutch government adopts a singular approach to domain hosting, deploying servers abroad to support services linked to its bilateral relationships. For instance, *dutchculturekorea.com*, a cultural blog of the Embassy of the Kingdom of the Netherlands in Seoul, is hosted on a server located in Korea. Similarly, *nbso-brazil.com.br*, the website for The Netherlands Business Support Offices in Brazil, is served from a server within this South American country.

Key Findings:

- 23% of government URLs are served from organizations registered internationally; 13% from servers located internationally.
- Significant regional variation in URLs served by international addresses. For example, while SA, EAP and NA deliver less than 10% of government URLs from international servers, SSA delivers 48% of government URLs from international servers.
- Foreign server reliance is usually within the same region, except LAC, MENA, and SA.
- Governments commonly use US-registered organizations for foreign hosting.
- Strong compliance with data regulations such as GDPR, DPDP, LGPD.
- Evidence of colonial ties: Morocco and France show mutual hosting dependencies.

7 Global providers and diversification

In the last section of our analysis, we focus on the networks responsible for serving government websites. The goal is to understand the role of Global Providers in this context (§7.1), and the degree of diversification among government providers (§7.2).

7.1 The Role of Global Providers

We have seen that governments are also engaged, if to a lesser extent, in the trend towards adopting third-party global providers for their digital services. In the following paragraphs, we characterize these providers, examining their global footprints, and analyzing countries' reliance on them.

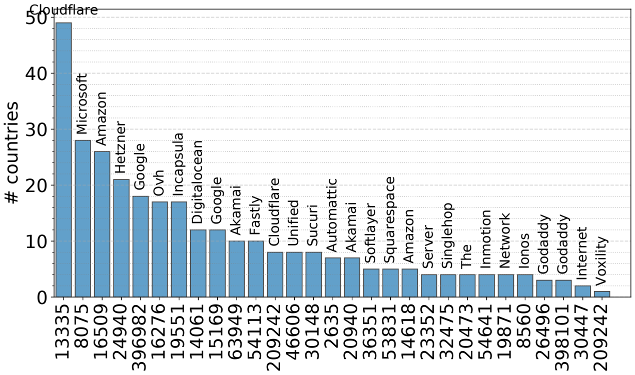


Figure 10: No. of countries that rely on Global Providers and CDF of Frac. of bytes served by Global Providers.

Figure 10 shows a histogram of the number of countries with government sites relying on one of the 28 global providers we identified. Cloudflare (AS13335) appears as the clear leader, serving content for 49 out of the 61 countries in our study. Cloudflare is followed by two other major cloud providers, AWS (AS16509, AS14618) and Azure (AS8075), hosting content for 31 and 28 countries, respectively.

To understand the degree of reliance on any given provider, we analyze the proportion of each country's data bytes served by each provider. At the top of the list, Amazon (AS16509) stands out by serving 97% of the bytes for an East Asian country, while Cloudflare (AS13335) is responsible for 72%, 58%, and 56% of the bytes for a country in Eastern Europe, in South America, and a small Asian country, respectively. Additionally, Hetzner (AS24940) delivers 57% of the bytes for the government of a Scandinavian country.

7.2 Diversification of Hosting Providers

Diversification in hosting strategies can enhance the resilience of government services by reducing the risk of a digital shutdown caused by organizational failure. It also helps in creating isolation of data access across different domains. We explore whether governments tend to adopt more diversified hosting strategies and how this strategy correlates with their preference for using Govt&SOE 🔵, 3P Local 🔴 or 3P Global 🟠 for hosting their services.

To assess diversification in the networks serving government websites, we utilize the Herfindahl-Hirschman Index (HHI) [73], a common measure of market concentration. This index provides a score ranging from 0 to 1, indicating the level of network diversification, where a score closer to 0 indicates high diversification and a score closer to 1 indicates higher concentration. Figure 11 illustrates the HHI distribution for both the fraction of URLs and bytes served per network in each country. These are further categorized into three groups (Govt&SOE 🔵, 3P Local 🔴 or 3P Global 🟠) based on the predominant source of bytes for each country.

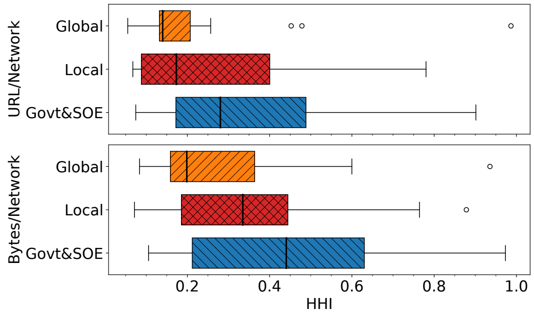


Figure 11: HHI distribution for the fraction of URLs and bytes served per hosting category.

While there is some overlap in the boxplots, governments mostly reliant on 3P Global 🟠 tend to adopt more diversified strategies compared to those using 3P Local 🔴, and even more so than those relying on Govt&SOE 🔵. For example, while 63% (12 / 19) of the countries in the Govt&SOE 🔵 category serve over 50% of their bytes from a single network, just 32% (8 / 25) of the countries in the 3P Global 🟠 category depend on a single network for their bytes. Diversification is simpler with third-party providers, as it typically involves just contractual agreements. With on-premises hosting (Govt&SOE 🔵), on the other hand, diversification is more complex and may require significant capital investment.

Key Findings:

- Cloudflare serves 49 countries, with a high of 72% bytes in one Eastern European country.
- Governments with on-premise infrastructure (63% in Govt&SOE 🔵) are less diversified than those using third-party services (32% in 3P Global 🟠 rely on a single network).

8 Limitations

Our study is subject to a number of limitations. For starters, our compilation of government websites predominantly relies on self-reported information from governments (§3.1). We benefited from a global trend among governments towards developing data repositories to centralize government digital resources. In some countries, this process is part of legislative initiatives, such as in Brazil with the Digital Government Law, while in others, or efforts from the executive branch, such as Argentina's Ministry of Modernization and Spain's Ministry of Digital Transformation. This data is made available in different formats (e.g., HTML items, CSV files) and through different types of resources, from webpages to GitHub repositories, as in the case of the US Cybersecurity and Infrastructure Security Agency (CISA). Despite of this, the criteria for including services on these lists vary, often due to unique governmental administrative structures, legal frameworks, and cultural idiosyncrasies among other factors. *We make our data available upon request to enable replication efforts.*

Our findings also reveal a lack of a standard convention for naming government domain names. While numerous countries adopt the ".gov" subdomain (or variations either in English or its equivalent in other languages) exclusively for government services, there are notable exceptions. For instance, state-owned enterprises rarely fall under this categorization and may use different domain structures. Furthermore, certain countries, including Germany, Poland, and the Netherlands, do not adhere to a specific subdomain convention for their government domains, indicating a varied approach to the digital identification of government entities across the globe which may impact our data collection effort.

In addition, our methodology focuses on public-facing services and excludes resources behind login portals, so it remains unclear if the same infrastructure supports publicly accessible and restricted resources. Despite recent advancements in understanding the potential use of Single-Sign-On (SSO) on top sites [6], these heuristics are not applicable to government sites that rarely accept third-party logins.

While we combine multiple approaches to minimize geolocation inaccuracies, we do not completely solve the problem. For instance, although active probing is the most accurate technique, it depends on factors such as server ICMP responsiveness and proximity of probes. In scenarios where active probing is not feasible, we resort to a multistage geolocation process, which can be costly. We opted for a conservative approach in our analysis, omitting (a small number of) IPs with geolocation from commercial databases that we could not validate.

Finally, we conduct a preliminary analysis to explore the possible mechanisms driving the hosting of government websites outside of their country, detailed in Appendix E. Key findings indicate that countries with higher GDP [52] and advanced network readiness [67] tend to host fewer services externally, while those with larger Internet populations [83] tend to do the opposite. This could be suggestive of higher traffic to government websites in these countries. Exploring the specific drivers in depth is part of future work.

9 Related Work

While prior research has explored the involvement of governments in managing network services and infrastructure, to the best of our knowledge, our work is the first large-scale study that focuses on the serving infrastructure used for the delivery of government services worldwide.

In a related context, Jansen et al. [47] examined the infrastructure used by government services in six countries with tense relationships with their neighbors. While our findings align with theirs on the shared countries examined, our study's broader perspective allows us to explore global and regional patterns, as well as the consolidation of global providers. In a study with a narrower scope in terms of the number of countries and government websites, Hsiao et al. [37] examined the external dependencies of public-facing government websites in seven countries over a two-year period. Jonker et al. [49] explored the changes in the infrastructure used by domains in the Russian Federation before and after the invasion of Ukraine. They found that a significant majority (70%) of Russian sites were fully hosted within Russia long before the conflict, and that thousands of Russian sites lost access to Western service providers post-conflict. Our findings confirm these observations. Sommese et al. [81] and Houser et al. [36] focused on evaluating the DNS infrastructure of government websites, reporting among other findings a growing reliance on single third-party DNS service providers.

Other studies have also analyzed government sites in various contexts. Habib et al. [32] investigated the affordability of public service websites in developing countries, highlighting the issue of large webpage sizes and suggesting potential solutions. Singanamalla et al. [79] analyzed the HTTPS prevalence in secure web communication for government websites, revealing that over 70% of global government sites lack valid HTTPS. Samarasinghe et al. [75] investigated the privacy practices of government websites and mobile apps, focusing on how personal data is handled and the associated privacy risks. Gotze et al. [30] studied popular governmental websites across different countries to examine the use of cookies and found that in some countries, up to 90% of these websites create cookies of third-party trackers without user consent.

Taking an AS-level perspective, Carisimo et al. [13] examined diverse data sources to compile a list of state-owned Internet Operators, showing that state involvement in offering Internet services is a widespread phenomenon. Furthermore, Fontugne et al. [24] studied the routing concentration during the 3-year transition of Crimea's Internet infrastructure following Russia's annexation where they found the creation of a choke point in the region's Internet architecture.

Internet centralization has gained notable attention in recent years. Kashaf et al. [50] studied the third-party dependency of the web ecosystem, finding that 89% of the top-100K websites rely on third-party DNS, CDN, or CA providers, with concentrated dependencies, suggesting potential vulnerabilities with large-scale consequences, such as the major Dyn outage in 2016. Kumar et al. [53] conducted a large-scale, longitudinal analysis of third-party dependency and centralization around the world. Their findings revealed that while dependencies on a third-party DNS, CDN or CA provider vary widely around the world, ranging from 19% to 76%, there is a highly concentrated market of third-party providers. Three providers across all countries serve an average of 92% of the surveyed websites. Even more concerning, these differences persist a year later with increasing dependencies across the countries. Doan et al. [20] focused on web consolidation within Content Delivery Infrastructures (CDI), noting an 83% increase in the fraction of webpages hosted on CDIs between 2015 and 2020, with Google, Akamai, Amazon, and Cloudflare emerging as the dominant providers.

10 Conclusion

We reported on the first comprehensive study exploring the hosting strategies of government digital services worldwide. Drawing from data collected across 61 countries spanning every continent and region in the world, we examined preferred hosting models for public-facing government sites, cross-border dependencies, and the level of centralization in government services. Our work provides the empirical basis for an understanding of hosting approaches in government sectors and can inform national and international policy agendas on digital sovereignty.

11 Acknowledgements

This work was supported by National Science Foundation grant CNS-2246475, with additional partial support from the Comcast

Innovation Fund Award. The opinions, findings, conclusions, and recommendations expressed are those of the authors and do not necessarily reflect the views of the funding sources. The authors would like to thank the reviewers for their valuable feedback and our shepherd for their comments on early versions of this manuscript.

References

[1] 2020. Consolidation in the Internet Economy. https://future.internetsociety.org/2019/

[2] Hussain AlJahdali, Abdulaziz Albatli, Peter Garraghan, Paul Townend, Lydia Lau, and Jie Xu. 2014. Multi-tenancy in cloud computing. In *Proc. of IEEE SOSE*.

[3] Mark Allman. 2018. Comments on DNS Robustness. In *Proc. of IMC*.

[4] Scott Anderson, Loqman Salamatian, Zachary S. Bischof, Alberto Dainotti, and Paul Barford. 2022. iGDB: connecting the physical and logical layers of the internet. In *Proceedings of the 22nd ACM Internet Measurement Conference*.

[5] APNIC. 2024. APNIC Service Region. https://www.apnic.net/about-apnic/corporate-documents/documents/corporate/apnic-service-region/.

[6] Calvin Ardi and Matt Calder. 2023. The Prevalence of Single Sign-On on the Web: Towards the Next Generation of Web Content Measurement. In *Proc. of IMC*.

[7] Jari Arkko. 2019. Centralised Architectures in Internet Infrastructure. *IETF Internet Draft* (2019).

[8] Jari Arkko. 2020. The influence of Internet architecture on centralised versus distributed Internet services. *Journal of Cyber Policy* (2020).

[9] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D Joseph, Randy H Katz, Andrew Konwinski, Gunho Lee, David A Patterson, Ariel Rabkin, Ion Stoica, et al. 2009. *Above the clouds: A berkeley view of cloud computing*. Technical Report.

[10] The World Bank. 2017. The world by region. https://datatopics.worldbank.org/sdgatlas/archive/2017/the-world-by-region.html.

[11] Fabián E. Bustamante, John Doyle, Walter Willinger, Marwan Fayed, David L. Alderson, Steven Low, Stefan Savage, and Henning Schulzrinne. 2024. Towards Re-Architecting Today's Internet for Survivability: NSF Workshop Report. *SIGCOMM Comput. Commun. Rev.* (2024).

[12] Michael Butkiewicz, Harsha V. Madhyastha, and Vyas Sekar. 2011. Understanding Website Complexity: Measurements, Metrics, and Implications. In *Proc. of ACM SIGCOMM*.

[13] Esteban Carisimo, Alexander Gamero-Garrido, Alex C. Snoeren, and Alberto Dainotti. 2021. Identifying ASes of State-Owned Internet Operators. In *Proc. of IMC*.

[14] Anupam Chander and Uyen P. Le. 2014. Breaking the Web: Data Localization vs. The Global Internet. *SSRN Electronic Journal* (2014).

[15] Intersoft Consulting. 2013. General Data Protection Regulation (GDPR).

[16] Luigi Coppolino, Salvatore D'Antonio, Giovanni Mazzeo, and Luigi Romano. 2017. Cloud security: Emerging threats and current solutions. *Computers & Electrical Engineering* (2017).

[17] Omar Darwich, Hugo Rimlinger, Milo Dreyfus, Matthieu Gouel, and Kevin Vermeulen. 2023. Replication: Towards a Publicly Available Internet Scale IP Geolocation Dataset. In *Proc. of IMC*.

[18] Chris Demchak and Peter Dombrowski. 2013. Cyber Westphalia: Asserting State Prerogatives in Cyberspace. *Georgetown Journal of International Affairs* (2013).

[19] David Dittrich, Erin Kenneally, and Michael Bailey. 2013. Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report. *SSRN Electronic Journal* (2013).

[20] Trinh Viet Doan, Roland van Rijswijk-Deij, Oliver Hohlfeld, and Vaibhav Bajpai. 2022. An Empirical View on Consolidation of the Web. *ACM Trans. Internet Technol.* (2022). https://doi.org/10.1145/3503158

[21] Craig Dobson. 2023. *Achieving Equity in Digital Government Services*.

[22] Frédérick Douzet, Louis Pétiniaud, Loqman Salamatian, Kevin Limonier, Kavé Salamatian, and Thibaut Alchus. 2020. Measuring the Fragmentation of the Internet: The Case of the Border Gateway Protocol (BGP) During the Ukrainian Crisis. In *2020 12th International Conference on Cyber Conflict (CyCon)*.

[23] Sean Fleming. 2021. Europe's digital sovereignty: From rulemaker to superpower in the age of US-China rivalry. *World Economic Forum* (2021).

[24] Romain Fontugne, Ksenia Ermoshina, and Emile Aben. 2020. The Internet in Crimea: a case study on routing interregnum. In *2020 IFIP Networking Conference (Networking)*.

[25] AWS Cloud Computing for Federal Government. 2023. Cloud Computing for Federal Government. https://aws.amazon.com/federal/

[26] Azure for US Government. 2023. Azure for US Government | Microsoft Azure. https://azure.microsoft.com/en-us/explore/global-infrastructure/government#options

[27] Josh Fruhlinger. 2020. The OPM hack explained: Bad security practices meet China's Captain America.

[28] International Monetary Fund. 2020. STATE-OWNED ENTERPRISES: THE OTHER GOVERNMENT. https://www.imf.org/~/media/Files/Publications/fiscal-monitor/2020/April/English/ch3.ashx.

[29] Petros Gigis, Matt Calder, Lefteris Manassakis, George Nomikos, Vasleois Kotronis, Xenofontas Dimitropoulos, Ethan Katz-Bassett, and Georgios Smaragdakis. 2021. Seven years in the life of Hypergiants' off-nets. In *Proc. of ACM SIGCOMM*.

[30] Matthias Gotze, Srdjan Matic, Costas Iordanou, Georgios Smaragdakis, and Nikolaos Laoutaris. 2022. Measuring Web Cookies in Governmental Websites. In *Proceedings of the 14th ACM Web Science Conference 2022*. https://doi.org/10.1145/3501247.3531545

[31] Robert L Grossman. 2009. The case for cloud computing. *IT professional* (2009).

[32] Rumaisa Habib, Aimen Inam, Ayesha Ali, Ihsan Ayyub Qazi, and Zafar Ayyub Qazi. 2023. A First Look at Public Service Websites from the Affordability Lens. In *Proc. of the WWW*.

[33] Heritage. 2023. Economic Freedom. https://www.heritage.org/index/explore?view=by-region-country-year&u=637879938517599600

[34] hotspotshield. 2023. HotSpotShield VPN. https://www.hotspotshield.com

[35] THE WHITE HOUSE. 2023. FACT SHEET: Building Digital Experiences for the American People | OMB.

[36] Rebekah Houser, Shuai Hao, Chase Cotton, and Haining Wang. 2022. A Comprehensive, Longitudinal Study of Government DNS Deployment at Global Scale. In *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. https://doi.org/10.1109/DSN53405.2022.00030

[37] Hsu-Chun Hsiao, Tiffany Hyun-Jin Kim, Yu-Ming Ku, Chun-Ming Chang, Hung-Fang Chen, Yu-Jen Chen, Chun-Wen Wang, and Wei Jeng. 2019. An Investigation of Cyber Autonomy on Government Websites. In *The World Wide Web Conference*.

[38] Christian Huitema, Geoff Huston, Emden Hs, Germany Leer, and Zhang. 2021. *Draft Report of DINRG Workshop on Centralization in the Internet*.

[39] Geoff Huston. 2018. DNS Resolver Centrality. https://labs.apnic.net/?p=1260

[40] Geoff Huston. 2021. CDN and centrality. APNIC Blog. https://blog.apnic.net/2021/07/02/opinion-cdns-and-centrality/

[41] Shadi Ibrahim, Bingsheng He, and Hai Jin. 2011. Towards pay-as-you-consume cloud computing. In *Proc. of Conference on Services Computing*.

[42] ICT Development Index. 2023. Ict Development Index by Country 2023. https://worldpopulationreview.com/country-rankings/ict-development-index-by-country

[43] India's MINISTRY OF LAW AND JUSTICE. 2023. India's Digital Personal Data Protection (DPDP) Act. https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf.

[44] Costas Iordanou, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris. 2018. Tracing Cross Border Web Tracking. In *Proc. of IMC*.

[45] ipInfo. 2023. ipInfo. https://ipinfo.io

[46] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*.

[47] Bernardus Jansen, Natalia Kadenko, Dirk Broeders, Michel van Eeten, Kevin Borgolte, and Tobias Fiebig. 2023. Pushing boundaries: An empirical view on the digital sovereignty of six governments in the midst of geopolitical tensions. *Government Information Quarterly* (2023).

[48] Wayne Jansen, Tim Grance, et al. 2011. Guidelines on security and privacy in public cloud computing. (2011).

[49] Mattijs Jonker, Gautam Akiwate, Antonia Affinito, kc Claffy, Alessio Botta, Geoffrey M. Voelker, Roland van Rijswijk-Deij, and Stefan Savage. 2022. Where .ru? assessing the impact of conflict on russian domain infrastructure. In *Proceedings of the 22nd ACM Internet Measurement Conference*.

[50] Aqsa Kashaf, Vyas Sekar, and Yuvraj Agarwal. 2020. Analyzing third party service dependencies in modern web services: Have we learned from the miraidyn incident?. In *Proc. of IMC*.

[51] Erin Kenneally and David Dittrich. 2012. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *SSRN Electronic Journal* (2012).

[52] Avery Koop. 2023. Mapped: GDP per Capita Worldwide. https://www.visualcapitalist.com/mapped-gdp-per-capita-worldwide/

[53] Rashna Kumar, Sana Asif, Elise Lee, and Fabian E. Bustamante. 2023. Each at Its Own Pace: Third-Party Dependency and Centralization Around the World. *Proc. ACM Meas. Anal. Comput. Syst.* (2023).

[54] Benjamin Cedric Larsen. 2022. The Geopolitics of AI and the Rise of Digital Sovereignty. *Brookings Institution* (2022).

[55] Brazil's General Data Protection Law. 2023. LGPD Brazil - General Personal Data Protection Act. https://lgpd-brazil.info

[56] Ton Viet Le. 2019. Vietnam - Data Protection Overview.

[57] Leonard Kleinman. 2020. *The Rise Of Third-Party Digital Risk*.

[58] J. Livingood, M. Antonakakis, B. Sleigh, and A. Winfield. 2019. Centralized DNS over HTTPS (DoH) implementation issues and risks.

[59] M. Luckie, B. Huffaker, A. Marder, Z. Bischof, M. Fletcher, and k. claffy. 2021. Learning to Extract Geographic Information from Internet Router Hostnames. In *Proc. of CoNEXT*.

[60] Ewen MacAskill, Gabriel Dance, Feilding Cage, Greg Chen, and Nadja Popovich. 2013. NSA files decoded: Edward Snowden's surveillance revelations explained. https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1

[61] James Manyika, Susan Lund, Jacques Bughin, Lola Woetzel, Kalin Stamenov, and Dhruv Dhingra. 2016. Digital Globalization: The New Era of Global Flows. *McKinsey & Company* (2016).

[62] Foivos Michelinakis, Hossein Doroud, Abbas Razaghpanah, Andra Lutu, Narseo Vallina-Rodriguez, Phillipa Gill, and Joerg Widmer. 2018. The Cloud that Runs the Mobile Internet: A Measurement Study of Mobile Cloud Services.

[63] Giovane Moura, Sebastian Castro, Wes Hardaker, Maarteb Wullink, and Cristian Hesselman. 2020. Clouding up the Internet: how centralized is DNS traffic becoming?. In *Proc. of IMC*.

[64] United Nations. 2023. EGOVKB | United Nations > About > Overview > E-Government Development Index. https://publicadministration.un.org/egovkb/en-us/About/Overview/-E-Government-Development-Index

[65] China News. 2022. Explainer: What's Behind Strained China-Japan Relations.

[66] nordvpn. 2023. Nord VPN. https://nordvpn.com

[67] NRI. 2023. Network Readiness Index. https://networkreadinessindex.org/countries/#ranking-wrapper

[68] State of California Department of Justice. 2023. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa

[69] U.S. Government Accountability Office. 2021. SolarWinds Cyberattack Demands Significant Federal and Private-Sector Response (infographic).

[70] Russia Data Protection Overview. 2020. Russia - Data Protection Overview.

[71] Norbert Pohlmann, Michael Sparenberg, Illya Siromaschenko, and Kilian Kilden. 2014. *Secure Communication and Digital Sovereignty in Europe.*

[72] United Nations Development Programme. 2023. Human Development Index. https://hdr.undp.org/data-center/human-development-index#/indicies/HDI

[73] S. A. Rhoades. 1993. The Herfindahl-Hirschman Index.

[74] RIPE NCC. 2024. IPmap. https://ipmap.ripe.net/.

[75] Nayanamana Samarasinghe, Aashish Adhikari, Mohammad Mannan, and Amr Youssef. 2022. Et tu, Brute? Privacy Analysis of Government Websites and Mobile Apps. In *Proceedings of the ACM Web Conference 2022.* https://doi.org/10.1145/3485447.3512223

[76] Marianne Schneider-Petsinger, Jue Wang, Yu Jie, and James Crabtree. 2019. *US-China Strategic Competition The Quest for Global Technological Leadership.*

[77] Selenium. 2024. SeleniumHQ Browser Automation.

[78] Dr. Muhammad Riaz Shad. 2018. Cyber Threat in Interstate Relations: Case of US-Russia Cyber Tensions. *Policy Perspectives* (2018).

[79] Sudheesh Singanamalla, Esther Han Beol Jang, Richard Anderson, Tadayoshi Kohno, and Kurtis Heimerl. 2020. Accept the risk and continue: Measuring the long tail of government https adoption. In *Proc. of IMC.*

[80] Raffaele Sommese, Leandro Bertholdo, Gautam Akiwate, Mattijs Jonker, Roland van Rijswijk-Deij, Alberto Dainotti, KC Claffy, and Anna Sperotto. 2020. Manycast2: Using anycast to measure anycast. In *Proc. of IMC.*

[81] Raffaele Sommese, Mattijs Jonker, Jeroen van der Ham, and Giovane C. M. Moura. 2022. Assessing e-Government DNS Resilience. In *2022 18th International Conference on Network and Service Management (CNSM).* https://doi.org/10.23919/CNSM55787.2022.9965155

[82] Ilona Stadnik. 2021. Control by infrastructure: Political ambitions meet technical implementations in RuNet. *First Monday* (2021).

[83] Internet World Stats. 2023. World Internet Users Statistics and 2019 World Population Stats. https://www.internetworldstats.com/stats.htm

[84] Bruce Stokes. 2016. Hostile Neighbors: China vs. Japan.

[85] surfshark. 2023. Surfshark VPN. https://surfshark.com

[86] The European Commission. 2024. Overseas Countries and Territories. https://international-partnerships.ec.europa.eu/countries/overseas-countries-and-territories_en.

[87] The French Republic. 2024. Accord sur la Nouvelle-Caledonie signe a Noumea le 5 mai 1998. https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000555817.

[88] The World Bank. 2024. *Brief: Digital Government for Development.*

[89] United Nations. 2022. UN E-Government Survey 2022 – The Future of Digital Government. (2022).

[90] United Nations. 2024. Non-Self-Governing Territories. https://www.un.org/dppa/decolonization/en/nsgt.

[91] US Department of State. 2024. A Guide to the United States' History of Recognition, Diplomatic, and Consular Relations, by Country, since 1776: Morocco. https://history.state.gov/countries/morocco.

[92] Luciano Zembruzki, Arthur Selle Jacobs, Gustavo Spier Landtreter, Lisandro Zambenedetti Granville, and Giovanne Moura. 2020. `dnstracker`: Measuring Centralization of DNS Infrastructure in the Wild. In *Proc. of AINA*.

Region	Country Code
North America (NA)	Canada United States
Latin America and the Caribbean (LAC)	Mexico Brazil
Europe and Central Asia (ECA)	France Bosnia
North Africa and the Middle East (MENA)	UAE Israel
Sub-Saharan Africa (SSA)	South Africa Egypt
South Asia (SA)	India Pakistan
East Asia and Pacific (EAP)	Japan New Zealand

Table 6: Two countries per region were selected to compare content delivery strategies between government websites and top sites. Our selection criteria focus on capturing countries with varying levels of digital development within each region.

A Ethical Considerations

Our work focuses on understanding how governments approach the delivery of digital services, delving into the strategy they employ to select providers and serving locations. While examining these mechanisms may raise concerns about potential vulnerability exposures, our work does not focus on highlighting security weaknesses or individual government strategies in detail. Instead, we focus on discerning whether governments predominantly rely on on-premises or third-party providers for their digital services. Our objective is to compare various strategies and identify global trends in government digital service delivery without emphasizing individual issues that could compromise the functioning of a government. This approach ensures a comprehensive understanding of the landscape without compromising the entities' security [19, 51].

Our work looks to empirically characterize the different governments' responses to the conflicting trends of Internet centralization and digital sovereignty. We aim to understand how governments approach the delivery of digital services, leveraging centralized, often global, third-party providers while maintaining sovereignty over their data and digital assets. By analyzing government strategies, we explore how different nations balance the benefits of third-party, their engagement with third-party providers registered abroad, and the prevalence of hosting resources beyond their limits.

B Country Selection Criteria

Table 9 presents our final country selection, encompassing 61 countries from across the globe. This selection includes 2 countries from North America, 8 from Latin America and the Caribbean, 29 from Europe and Central Asia, 5 from North Africa and the Middle East, 2 from Sub-Saharan Africa, 3 from South Asia, and 12 from East Asia and Pacific (EAP). These countries combined represent 82.70% of the global Internet population.

C Government Dataset Overview

Table 8 presents statistics on the number of government landing URLs, internal government URLs, and government hostnames collected for each country.

D Governments vs Topsites

In this analysis, we focus on determining whether *government content delivery strategies differ from those used by popular websites* To explore this, we compare government websites' hosting strategies with popular sites across all regions to detect patterns or contrasts in their digital infrastructure approaches. In this case, we explore discrepancies in government and non-government sites in a subset of countries of our dataset, consisting of two countries per region as indicated in Table 6.

D.1 A Comparing Methodology

The initial step of our comparison is to identify the serving infrastructure of popular websites in the studied countries by leveraging the methodology described in (§3.4) and (§3.5)

Our second step consists of identifying the fraction of non-government topsites that use either on-premise or third-party solutions to deliver content. This mirrors our government site analysis and redefines categories as (1) self-hosting, (2) global, (3) local, and (4) foreign providers. To identify self-hosted solutions, we use a heuristic from previous research [50, 53], which first checks for Canonical Name (CNAME) redirects. This heuristic involves comparing the Second-Level Domain (2LD[5]) of the CNAME with the site's 2LD; a match suggests a self-hosting. In cases of mismatch, and if the site uses HTTPS, we further check if the CNAME's 2LD appears in the site's Subject Alternative Names (SANs) List. A match here also indicates a private provider, helping us identify sites like img.youtube.com, which belong to the same entity despite different 2LDs in the CNAME and the site. For sites not using CNAME redirects or those not identifiable as CDN providers, we categorize them differently, as their hosting type remains to be determined.

Features	VIF Factor
internet_users	2.06
HDI	8.61
IDI	4.11
NRI	9.09
GDP	5.00
econ_freedom	3.71

Table 7: Variance Inflation Factor (VIF) for each feature.

E Explanatory Factors

To find possible mechanisms driving the hosting of government websites outside a country, we train an Ordinary Least Squares (OLS) regression model. To this end, we estimate the following equation:

$$Y_i = \alpha + \beta_1 I_i + \beta_2 E_i + \beta_3 G_i + \beta_4 H_i + \beta_5 N_i + \beta_6 U_i + \epsilon_i \qquad (1)$$

where Y_i is the outcome variable that refers to the percentage of government URLs that are served from outside the country i, I_i

[5] By 2LD, we refer to 2LD+TLD in this work

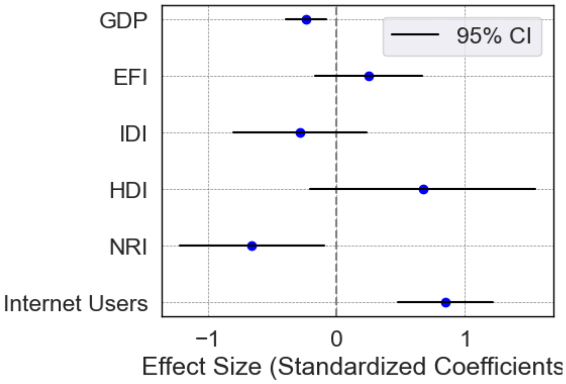


Figure 12: Correlates of the percentage of server IPs hosted outside a country. Shown are the estimated coefficient values. Error bars indicate 95% confidence intervals.

is the ICT Development Index (IDI) [42] in i, E_i is the Economic Freedom Index (EFI) [33], G_i is the GDP [52] of country i, H_i is the Human Development Index [72] of i, N_i is the Network Readiness Index [67], and U_i is the total number of Internet users [83] in i. All variables are standardized (i.e., transformed to have a mean of 0 and a standard deviation of 1).

We analyze the Variance Inflation Factor (VIF) to estimate how much the variance of an estimated regression coefficient increases if your predictors are correlated. If there is no multicollinearity, the VIFs will be close to 1, meaning the variance of the coefficient estimate is not inflated. As multicollinearity increases, so does the VIF, indicating a higher inflation level in the coefficient variance. Our analysis reveals that all the included explanatory variables had a Variance Inflation Factor (VIF) under 10, suggesting that multicollinearity may not be significant enough to warrant removing variables [46]; as shown in Table 7.

Figure 12 shows the values of the estimated coefficients along with 95 percent confidence intervals. We find that the coefficients of only three features are statistically significant at the 5% level: U_i (β = 0.845, 95% CI = [0.476, 1.214], $p < 0.001$), N_i (β = -0.660, 95% CI = [-1.225, -0.095], p = 0.022) and G_i (β = -0.239, 95% CI = [-0.399, -0.079], p = 0.003). Intuitively, this means that a one standard deviation change in GDP leads to a -0.239 standard deviation change in the outcome variable. Thus, countries with higher GDP and greater network readiness tend to place less of their services outside their own country. Moreover, countries with larger Internet populations tend to place more of their services outside the country. This could be suggestive of higher traffic to government websites in these countries.

North America

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
United States	US	1340	38702	2343	Canada	CA	216	6626	127

Europe and Central Asia

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
Germany	DE	777	28841	451	Russia	RU	106	5813	46
United Kingdom	GB	373	9005	320	France	FR	669	9705	238
Italy	IT	129	8518	123	Spain	ES	251	14602	175
Netherlands	NL	1293	39026	966	Poland	PL	594	29699	470
Sweden	SE	335	9110	285	Belgium	BE	994	217598	637
Greece	GR	91	6025	88	Switzerland	CH	83	3225	25
Turkey	TR	226	14817	228	Ukraine	UA	93	3928	98
Czechia	CZ	49	2153	46	Romania	RO	65	3427	49
Hungary	HU	109	204042	70	Portugal	PT	295	15809	253
Bulgaria	BG	144	5798	75	Kazakhstan	KZ	52	648	16
Serbia	RS	66	3295	67	Latvia	LV	291	13263	239
Estonia	EE	118	9871	119	Georgia	GE	73	2226	61
Bosnia	BA	59	2929	58	Albania	AL	80	5536	79
Moldova	MD	50	3464	24	Denmark	DK	110	2922	110
Norway	NO	162	4382	158					

East Asia and Pacific

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
China	CN	193	6195	190	Japan	JP	93	3635	75
Indonesia	ID	76	3690	79	Vietnam	VN	56	1642	54
Thailand	TH	81	3267	82	South Korea	KR	0	0	0
Malaysia	MY	261	20206	247	Australia	AU	708	6883	440
Singapore	SG	87	4368	90	New Zealand	NZ	251	7358	233
Taiwan	TW	58	2996	54	Hong Kong	HK	108	6857	92

South Asia

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
India	IN	207	13612	213	Bangladesh	BD	333	15757	329
Pakistan	PK	118	3133	108					

Middle East and North Africa

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
Egypt	EG	69	4683	66	Algeria	DZ	202	2231	184
Morocco	MA	144	8440	137	United Arab Emirates	AE	49	5277	50
Israel	IL	101	2994	98					

Sub-Saharan Africa

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
Nigeria	NG	189	11332	187	South Africa	ZA	189	11332	187

Latin America and Caribbean

Country	ccTLD	Landing URLs	Internal URLs	Hostnames	Country	ccTLD	Landing URLs	Internal URLs	Hostnames
Brazil	BR	272	15711	212	Mexico	MX	317	9418	140
Argentina	AR	201	6238	100	Chile	CL	448	24571	434
Bolivia	BO	194	12842	189	Paraguay	PY	146	6744	133
Costa Rica	CR	196	12231	176	Uruguay	UY	67	4322	27

Table 8: Summary of government dataset statistics grouped by region, including the number of landing government URLs, internal government URLs, and government hostnames collected for each country.

North America
99.98% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
United States	US	0.915	0.921	92	5.760	Nord	Canada	CA	0.851	0.936	93	0.685	Nord

Europe and Central Asia
87.98% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
Russia	RU	0.816	0.822	90	2.299	Hotspot Shield	Germany	DE	0.877	0.942	92	1.459	Nord
Turkey	TR	0.798	0.838	83	1.3371	Nord	United Kingdom	GB	0.914	0.929	97	1.200	Nord
France	FR	0.883	0.903	85	1.114	Nord	Italy	IT	0.838	0.895	85	1.011	Nord
Spain	ES	0.884	0.905	94	0.802	Nord	Ukraine	UA	0.803	0.773	79	0.7545	Nord
Poland	PL	0.844	0.876	87	0.640	Nord	Kazakhstan	KZ	0.863	0.811	92	0.304	Surfshark
Netherlands	NL	0.938	0.941	93	0.302	Nord	Romania	RO	0.762	0.821	86	0.2738	Nord
Belgium	BE	0.827	0.937	94	0.198	Nord	Sweden	SE	0.941	0.947	95	0.183	Nord
Czechia	CZ	0.809	0.889	85	0.1719	Nord	Portugal	PT	0.827	0.866	84	0.165	Nord
Hungary	HU	0.783	0.846	90	0.1584	Nord	Switzerland	CH	0.875	0.962	96	0.155	Nord
Greece	GR	0.846	0.887	83	0.150	Nord	Serbia	RS	0.824	0.802	84	0.125	Nord
Denmark	DK	0.972	0.948	98	0.105	Nord	Norway	NO	0.888	0.961	99	0.099	Nord
Bulgaria	BG	0.777	0.795	79	0.0886	Nord	Georgia	GE	0.750	0.802	79	0.0669	Nord
Moldova	MD	0.725	0.767	60	0.0566	Nord	Bosnia	BA	0.626	0.780	79	0.0522	Nord
Albania	AL	0.741	0.796	83	0.0404	Nord	Latvia	LV	0.860	0.863	91	0.031	Nord
Estonia	EE	0.939	0.890	91	0.024	Nord							

East Asia and Pacific
91.14% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
China	CN	0.812	0.768	76	18.6404	Hotspot Shield	Indonesia	ID	0.716	0.705	66	3.9163	Nord
Japan	JP	0.900	0.925	83	2.1878	Nord	Vietnam	VN	0.679	0.703	79	1.5661	Nord
Thailand	TH	0.766	0.800	88	1.1416	Nord	Korea	KR	0.953	0.925	97	0.9184	Nord
Malaysia	MY	0.774	0.803	97	0.5715	Nord	Australia	AU	0.941	0.951	96	0.4314	Nord
Taiwan	TW	-	-	-	0.4175	Nord	Hong Kong	HK	-	0.952	96	0.1234	Nord
Singapore	SG	0.913	0.939	96	0.1005	Nord	New Zealand	NZ	0.943	0.937	96	0.0841	Nord

South Asia
96.33% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
India	IN	0.588	0.633	46	15.376	Nord	Bangladesh	BD	0.563	0.661	39	2.3824	Surfshark
Pakistan	PK	0.424	0.544	21	2.1393	Surfshark							

Middle East and North Africa
40.22% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
Egypt	EG	0.590	0.731	72	1.0096	Surfshark	Algeria	DZ	0.561	0.745	71	0.698	Surfshark
Morocco	MA	0.592	0.683	88	0.4719	Surfshark	UAE	AE	0.901	0.911	100	0.2246	Nord
Israel	IL	0.889	0.919	90	0.1474	Nord							

Sub-Saharan Africa
40.43% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
Nigeria	NG	0.453	0.535	55	2.846	Surfshark	South Africa	ZA	0.736	0.713	72	0.6371	Nord

Latin America and Caribbean
68.59% Internet population coverage

Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN	Country	ccTLD	EDGI	HDI	IUI	% pop.	VPN
Brazil	BR	0.791	0.754	81	3.285	Nord	Mexico	MX	0.747	0.758	76	2.036	Nord
Argentina	AR	0.820	0.842	88	0.775	Nord	Chile	CL	0.838	0.855	90	0.347	Nord
Bolivia	BO	0.617	0.692	66	0.164	Surfshark	Paraguay	PY	0.633	0.717	76	0.1139	Surfshark
Costa Rica	CR	0.766	0.809	83	0.082	Nord	Uruguay	UY	0.839	0.809	90	0.0602	Surfshark

Table 9: The 61 countries chosen for our study represent a diverse range of digital developments across all regions. This selection is based on three key indicators: (1) the E-Government Development Index (EGDI), (2) the Human Development Index (HDI), and (3) the International Telecommunication Union/World Bank Internet Penetration rates (UIU). The table also displays the percentage of the world’s Internet population covered by our selection on a per-region basis, showing our study’s geographic and developmental coverage.