

Master's Programme in Computer, Communication and Information Sciences

# Detecting digital dependence

Inferring public-sector hosting arrangements from Internet infrastructural records

---

Jaakko Kilpi

Master's thesis  
2025

Copyright ©2025 Jaakko Kilpi

---

<b>Author</b>	Jaakko Kilpi	
<b>Title of thesis</b>	Detecting digital dependence: Inferring public-sector hosting arrangements from Internet infrastructural records	
<b>Programme</b>	Master's Programme in Computer, Communication and Information Sciences	
<b>Major</b>	Human-Computer Interaction	
<b>Thesis supervisor</b>	Prof. Vili Lehdonvirta	
<b>Thesis advisor(s)</b>	Prof. Vili Lehdonvirta and Dr. Otto Kässi	
<b>Date</b>	<b>Number of pages</b>	<b>Language</b>
5.10.2025	76 + 3	English

---

### Abstract

Governments increasingly rely on digital infrastructures provided by companies, raising concerns about digital sovereignty and dependence on a small set of global cloud providers. This thesis asks whether the hosting providers of public-sector digital services can be inferred from publicly observable infrastructural records, and what forms of reliance such analysis reveals.

A dataset of verified hosting arrangements was assembled through Freedom of Information (FOI) requests in the United Kingdom, Finland, and the Philippines, supplemented by confirmed cases of Chinese hyperscaler use. These disclosures provided a rare form of ground truth against which predictive models could be evaluated. Observable records, such as DNS records, were collected for each domain and transformed into categorical features. Whereas previous studies often relied on single-record heuristics to attribute hosting, this thesis evaluates predictive models trained with stratified cross-validation under different provider groupings.

The findings show clear patterns of reliance. The UK and Finland relied heavily on Amazon Web Services and Microsoft Azure, while the Philippines retained significant self-hosting. No FOI responses indicated use of Chinese hyperscalers. Predictive models reproduced provider classifications with substantially higher accuracy than trivial baselines. Feature importance analysis further showed that accurate predictions did not hinge on a single record but instead drew on a combination of technical records across record types.

The study demonstrates that public-sector hosting providers can be inferred from infrastructural records with reasonable reliability, though only under conditions of validated training data and carefully structured categories. Prediction cannot substitute institutional transparency, but it can complement it by offering systematic and scalable visibility into otherwise opaque dependencies.

---

**Keywords** Digital sovereignty, interdependence, Internet measurement, machine learning

---

---

**Tekijä** Jaakko Kilpi

---

**Työn nimi** Digitaalisen kytköksisyyden havaitseminen

---

**Koulutusohjelma** Master's Programme in Computer, Communication and Information Sciences

---

**Pääaine** Human-Computer Interaction

---

**Vastuuopettaja/valvoja** Prof. Vili Lehdonvirta

---

**Työn ohjaaja(t)** Prof. Vili Lehdonvirta ja Dr. Otto Kässi

---

**Päivämäärä** 5.10.2025    **Sivumäärä** 76 + 3    **Kieli** englanti

---

### **Tiivistelmä**

Hallinnot tukeutuvat enenevässä määrin yritysten tarjoamiin digitaalisiin infrastruktuureihin, mikä herättää huolia digitaalisesta suvereniteetista ja riippuvaisuudesta pieneen joukkoon globaaleja pilvipalveluntarjoajia. Tämä diplomityö tutkii, voidaanko julkisen sektorin digitaalisten palveluiden isännöintipalveluntarjoajat ennustaa julkisesti havaittavien infrastruktuuritietojen avulla ja millaisia riippuvaisuuden kaavoja analyysi voi paljastaa.

Aineisto varmistetuista isännöintijärjestelyistä koottiin tietopyyntöjen avulla Isossa-Britanniassa, Suomessa ja Filippiineillä. Sitä täydennettiin vahvistetuilla tapauksilla kiinalaisten hyperskaalareiden käytöstä. Nämä tiedot tarjosivat harvinaisen pohjatotuusaineiston, jota vasten ennustemalleja voitiin arvioida. Havaittavia tietueita kerättiin jokaisesta verkkotunnuksesta ja muunnettiin ennustemallien hyödyntämiseksi kategoriseksi piirteiksi. Aiemmat tutkimukset ovat usein nojautuneet yksittäisiin tietueisiin palveluntarjoajan määrittämisessä, mutta tässä työssä arvioitiin stratifoidulla ristiinvalidoinnilla koulutettuja ennustemalleja eri palveluntarjoajaryhmittelyillä.

Tulokset osoittavat selkeitä riippuvuuden kaavoja. Iso-Britannia ja Suomi olivat vahvasti keskittyneet Amazon Web Servicesiin ja Microsoft Azureen, kun taas Filippiineillä itse ylläpidetyt järjestelyt säilyivät merkittävänä. Yksikään tietopyyntövastaus ei osoittanut kiinalaisten hyperskaalareiden käyttöä. Ennustemallit kykenivät toistamaan palveluntarjoajaluokitukset huomattavasti korkeammalla tarkkuudella kuin triviaalit vertailumallit. Piirreanalyysi osoitti, että tarkkuus ei perustunut yksittäisiin tietueisiin, vaan useiden erilaisten teknisten tietueiden yhdistelmiin.

Tutkimus osoittaa, että julkisen sektorin isännöintipalveluntarjoajia voidaan päätellä infrastruktuuritietojen perusteella kohtuullisella luotettavuudella, mutta vain valikoidun opetusdatan ja huolellisesti jäsennehtyjen kategorioiden olosuhteissa. Ennustaminen ei voi korvata institutionaalista läpinäkyvyyttä, mutta se voi täydentää sitä tarjoamalla systemaattisen ja skaalautuvan näkyvyyden muuten läpinäkymättömiin ilmiöihin.

---

**Avainsanat** Digitaalinen suvereniteetti, keskinäisriippuvuus, Internetin mittaaminen, koneoppiminen

---

## Table of contents

Preface and acknowledgements .....	7
Abbreviations .....	8
1 Introduction .....	9
1.1 Background and motivation .....	9
1.2 Objectives and research question .....	10
1.3 Approach and structure .....	11
2 Literature review .....	13
2.1 Defining public sector organizations .....	13
2.1.1 Existing definitions and perspectives .....	13
2.1.2 Practical constraints .....	15
2.2 Public sector digital services .....	16
2.3 Hosting .....	18
2.3.1 Evolution of hosting infrastructure .....	19
2.3.2 Current public sector hosting landscape .....	20
2.3.3 Technical stack of web services .....	22
2.3.4 Attribution challenge .....	24
2.4 Methodological approaches .....	25
2.5 Summary and research motivation .....	29
3 Research methodology .....	30
3.1 Ground truth .....	30
3.1.1 FOI background and legal considerations .....	30
3.1.2 Scope .....	31
3.1.3 FOI request procedure .....	32
3.1.4 Response processing and label assignment .....	33
3.1.5 Supplementary data on Chinese hyperscalers .....	35
3.2 Observational data .....	36
3.2.1 Data sources and collection .....	36
3.2.2 Data cleaning .....	37
3.2.3 Class balance and baseline .....	39
3.3 Modelling approach .....	40
3.3.1 Feature engineering .....	40

3.3.2	Candidate models .....	41
3.3.3	Training and validation .....	43
3.3.4	Evaluation.....	44
3.4	Limitations .....	46
4	Results .....	49
4.1	Hosting provider ground truth .....	49
4.1.1	Hosting distribution in FOI responses .....	49
4.1.2	Hosting distribution in supplementary Chinese hyperscaler dataset	50
4.2	Predictive modelling results .....	51
4.2.1	Overall model performance.....	51
4.2.2	Misclassification and per-class performance.....	54
4.2.3	Interpretability and feature importance .....	56
5	Discussion.....	58
5.1	Patterns of infrastructural interdependence .....	58
5.2	Predictive feasibility.....	60
5.3	Limitations and future developments .....	61
5.3.1	Data-related constraints.....	61
5.3.2	Modelling constraints.....	62
5.3.3	Conceptual and definitional limits .....	64
5.3.4	Generalisability.....	65
5.3.5	Future research.....	65
6	Conclusions .....	67
	References.....	69
	A. Email templates for the UK and Finland .....	77
	B. Decision tree of depth 6.....	79

## **Preface and acknowledgements**

Writing this thesis has been quite the task. During the process, there have been times where progression felt smooth, steady, and even quick. And times where it came to a stressful, frustrating halt. Every single one of those experiences has taught me something important. When I first began working on the thesis, the 6-months allocated for completing it seemed quite long. Looking back at it, it feels like it all passed in the snap of a finger, a reminder that perhaps the negative times weren't so bad after all.

Finishing this thesis marks, in many ways, the closing chapter of my formal education thus far. Aptly, the feeling is surreal. However, I feel like the writing process has provided a valuable opportunity for me to build my confidence in moving forward. No longer was I just completing courses and doing what was strictly asked of me. Instead, I was creating a new body of work that stands on its own feet, showing in the process that I, too, can do the same.

I want to thank my supervisor Vili Lehdonvirta and advisor Otto Kässi for the opportunity to work on this at the right time. I would also like to thank you both for all the words of encouragement and guidance you've given to me during this process. I'm sure they'll stick with me for quite a while.

I also want to thank my family and friends. My family, for keeping me focused with their words of encouragement and support. And my friends for giving their support to do the polar opposite when needed.

Helsinki, 5 October 2025  
Jaakko Kilpi

## Abbreviations

FOI	Freedom of Information
DNS	Domain Name System
NS	Name Server
MX	Mail Exchanger
TLS	Transport Layer Security
ASN	Autonomous System Number
SVM	Support Vector Machine
CDN	Content Delivery Network
AWS	Amazon Web Services

# 1 Introduction

As governments expand the scope of their digital services (Margetts and Dunleavy, 2013; OECD, 2021), the infrastructure that underpins these platforms has become a matter of strategic concern. The choice of hosting provider carries significant political and economic weight. Relying on foreign firms can create long-term dependencies, leaving governments vulnerable to vendor lock-in and the political choices of the states in which those firms are based (Armbrust et al., 2010; Paquette et al., 2010). Such decisions influence the degree of autonomy that states retain over their digital infrastructure, with consequences for governance, accountability, and long-term policy flexibility (Farrell and Newman, 2019; Flyverbom et al., 2019). Yet for many public-facing services, the hosting arrangements in place remain opaque to external observers. This thesis examines whether these arrangements can be systematically predicted from publicly observable technical records, and what patterns of reliance and interdependence can be found in the public sector.

## 1.1 Background and motivation

Over the past decade, governments worldwide have increasingly adopted cloud infrastructure as the backbone of new and updated digital services. This shift has been promoted as offering greater scalability, lower costs, improved reliability, and streamlined service delivery. Digitalization is now considered a mandatory element of modernization strategies across the OECD and beyond, and public agencies are expected to move essential services onto scalable digital platforms (Margetts and Dunleavy, 2013; OECD, 2021).

This rapid adoption has occurred in a highly concentrated market. A handful of hyperscale providers account for the majority of global spending on infrastructure- and platform-as-a-service, together commanding roughly two-thirds of the market (OECD, 2025a). Outside North America and Europe, Chinese providers such as Alibaba Cloud and Huawei Cloud are gaining significant market share, particularly in Asia (Lehdonvirta et al., 2025; Yang and Li, 2025). In Europe, however, reliance on non-European providers has produced a marked asymmetry between consumption and supply. While there are smaller European providers, the commanding market shares tend to belong to non-European hyperscalers in many cases (OECD, 2025a).

Dependence on external cloud providers is not merely a technical or budgetary matter, as it also carries strategic implications. Once public bodies move workloads onto a given provider, they may face vendor lock-in, proprietary APIs, pricing models, and data transfer costs can make switching

prohibitively expensive (Armbrust et al., 2010). Beyond these economic dependencies, governments also expose themselves to political risks. Hyperscale providers are subject to the legal regimes of their home states, meaning that extraterritorial regulations such as the US CLOUD Act can affect how data is accessed and governed abroad (Rojszczak, 2020). These concerns have fuelled European initiatives such as GAIA-X (Braud et al., 2021), which explicitly frame cloud strategy in terms of “digital sovereignty” (Autolitano and Pawlowska, 2021; Blancato, 2024). Farrell and Newman’s (2019) concept of weaponised interdependence further underlines this point, as states in central network positions can exploit dependencies for coercive or surveillance advantage, and cloud infrastructure exemplifies such dynamics.

The implications of these risks are particularly acute for governments. Unlike private firms, public bodies provide services that are essential to democratic legitimacy and citizen trust, support governmental operations, and hold highly sensitive personal data (OECD, 2025b). Dependence on external providers therefore goes beyond questions of efficiency, affecting sovereignty, resilience against disruptions, and the capacity of states to govern their own digital infrastructures.

Despite the strategic stakes, it is often difficult to determine where government services are hosted. Procurement records may be incomplete or outdated (OECD, 2025c), and hosting arrangements are rarely disclosed in metadata or other public-facing documentation. Domain attribution studies show that WHOIS and registrar data are insufficient and require additional inference from auxiliary sources (Sebastián et al., 2023). This lack of visibility complicates efforts to assess the extent of public-sector reliance on hyperscale cloud providers.

This thesis addresses this challenge by asking whether the hosting provider of a public-sector website can be predicted from publicly observable technical records, and how reliable such estimates are when evaluated against verified ground-truth data. By doing so, it aims to contribute both a methodological evaluation to the field of Internet measurement and a perspective on the visibility of infrastructural interdependence in the public sector.

## **1.2 Objectives and research question**

The overarching aim of this thesis is to examine whether the hosting provider of a public-sector website can be predicted from publicly observable data. This question matters both for methodological reasons, as it evaluates the feasibility of predicting hosting providers, and for substantive reasons, as it relates to broader debates about digital sovereignty and infrastructural interdependence.

To pursue this aim, the thesis sets out four objectives:

1. To review and assess existing approaches for inferring hosting arrangements, and to identify their limitations.
2. To assemble a ground-truth dataset of government websites and their hosting providers, enabling systematic evaluation of methods that link observable data to hosting providers.
3. To design and evaluate a predictive model that tests how reliably publicly observable data can identify hosting providers, and to develop it in a way that allows extension with additional data or applications.
4. To reflect on the broader implications of such methods for digital sovereignty and infrastructural interdependence.

These objectives lead to the guiding research question of the thesis:

**RQ:** Can the underlying hosting provider of a public-sector digital service be predicted using publicly observable information?

### **1.3 Approach and structure**

This thesis adopts an exploratory approach to the question of whether the hosting provider of a public-sector digital service can be predicted from publicly observable information. The project is driven by two parallel aims. On one hand, it tests the feasibility and limitations of predictive methods for linking external technical records to verified hosting providers. On the other hand, it reflects on what these methods reveal about interdependence in public-sector digital infrastructure. The intention is not only to demonstrate predictive accuracy, but also to evaluate these methods with attention to interpretability, generalizability, and the limits of what they can reveal about infrastructural dependence.

The research proceeds in several stages. First, a literature review (chapter 2) establishes the conceptual and technical background. It surveys how “public sector” and “public-sector digital services” can be defined, examines the evolution of hosting technologies and the market dominance of hyperscale providers, and outlines the layers of hosting infrastructure that complicate direct attribution. It then evaluates prior attempts to infer hosting arrangements from observable data, highlighting their limitations. This review forms the basis for identifying what kinds of questions any new method should address and clarifies the scope of the technical records to be considered.

Building on this foundation, chapter 3 describes the methodology of the study. The study assembles a ground-truth dataset of government websites and their verified hosting providers, obtained through Freedom of Information requests. This dataset enables systematic validation of methods that attempt to map observable records to hosting providers. Predictive

models are then designed and evaluated to test how reliably these records can identify providers. The modelling strategy emphasises transparency, as models are selected and interpreted not only for their accuracy, but also for their capacity to reveal which features matter, and for their potential to be applied to new datasets or contexts.

Chapter 4 presents the empirical results, reporting overall performance, per-class outcomes, and the relative informativeness of different features. Chapter 5 discusses the broader implications of these findings for digital sovereignty, infrastructural transparency, and methodological generalisation. Finally, Chapter 6 concludes by summarising contributions, limitations, and avenues for further research.

## **2 Literature review**

This chapter reviews previous research on the hosting of public-sector digital services. It first considers how the boundaries of the public sector are defined and how digitalisation has transformed service delivery. It then examines the infrastructures that support these services, outlining their historical evolution, current hosting practices, and the layered technical stack that complicates attribution. Existing approaches to identifying hosting providers are assessed, drawing on both technical records and institutional data, before highlighting the methodological limits of these studies. The chapter concludes by identifying a gap in systematic, validated methods for assessing governments' infrastructural dependencies.

### **2.1 Defining public sector organizations**

What comprises the “public sector”? This question is central to any attempt at assessing how governments manage their digital infrastructure. However, each country maintains its own legal and administrative boundaries for the public sector. For example, some include partially state-owned enterprises to a certain ownership share, while others don't. These inconsistencies can introduce substantial bias into cross-country comparisons. An organization treated as a core part of government in one setting might be excluded entirely in another, even if their roles in digital service provision are functionally similar. For the purposes of this study, it is therefore necessary to adopt a clear and universal definition.

The following sections will examine existing public sector definitions in various contexts, outline the practical constraints that shape this study's scope, and present the operational definition used throughout the analysis.

#### **2.1.1 Existing definitions and perspectives**

As defined by Merriam-Webster (“Definition of PUBLIC SECTOR,” 2025), the public sector refers to “the part of an economy which is controlled or owned by the government”. However, various definitions exist that limit the scope of this in different ways. This subsection will explore these definitions and their contexts of use.

In the field of public administration, the public sector is typically understood as an encapsulation of the state and its effects. Definitions in this domain emphasize institutional form, public accountability, and the delivery of collective goods and services. According to Lane (2000), the public sector consists of “the institutions of politics, government and bureaux” (p. 1), which are distinguishable from the private sector by their orientation towards public interest as opposed to private interest. Lane also stresses that

“there is no single way to make the private-public distinction”, as interpretations tend to clash. In practice, a publicly owned corporation can thus simultaneously be considered a part of the public sector in terms of ownership, while also being considered a part of the private sector in terms of its private interest of monetary gain.

Bozeman and Bretschneider (1994) present the concept of “publicness”, a trait that “reflects the extent the organization is influenced by public authority”. This presents the public-private split not as a dichotomy but as a scale. While different from a binary classification, this reinforces the understanding of public sector being difficult to define perfectly, especially due to varying governance structures between countries.

The public administration perspective frames the public sector as an abstract and adaptive concept defined not only by funding or ownership, but also by, for example, service provision. However, difficulties arise in cross-national comparisons of public sector digital autonomy, as comparisons between countries would require further insight into the organization structure of each country selected for analysis. While the field of public administration offers theoretically rich definitions of the public sector, the definitions share ambiguous boundaries.

In contrast to public administration literature, legal definitions of the public sector focus on more tangible criteria such as ownership, control, or the legal responsibilities of organizations. These definitions vary between countries and legal systems.

Comparing two countries with different national legal definitions for the public sector can become difficult, especially when comparing public sectors at large. For example, in Finland, incorporated entities are officially not considered a part of the public sector (“Public sector,” n.d.) even if owned by the government. On the other hand, in the United Kingdom, this is not the case, and corporations controlled by the government, even if they don’t receive their funding from the government, are included in the local definition of the public sector (Oliver, 2024). When comparing the entire public sectors of countries, these differences in legal definitions can cause notable unintended differences to be included in the comparison.

Supranational definitions provide a more reliable basis for international comparisons of public sectors, though they also make generalisations. The United Nations (2009) defines its equivalent of the public sector as the “General Government”, which “consists of institutional units that, in addition to fulfilling their political responsibilities and their role of economic regulation, produce services (and possibly goods) for individual or collective consumption mainly on a non-market basis and redistribute income and wealth.” The European Union specific European System of Accounts (ESA) (Eurostat., 2013) is harmonized with the UN version, and the definition remains largely similar. The ESA definition goes further in explaining the role of public control and explains that it works as an additional identifier in

determining if an institution is part of the public sector or not. However, the exact definition for when something is controlled by the public is left to be defined by individual nations themselves. Both sources make cases for government-controlled corporations being a part of the public sector as public corporations.

While useful and less focused on abstract concepts as their public administration counterparts, legal (/statistical?) definitions of the public sector are not without their flaws. Country-specific ones, while specific, are not necessarily always comparable to each other. On the other hand, the internationally applicable definitions remain unclear regarding edge case definitions.

From an economic perspective, the public sector can be considered to exist for the purpose of providing services that markets are unable or unwilling to offer efficiently. A foundational concept to this is the public good, first introduced by Samuelson (Samuelson, 1954) as the “collective consumption good”. These public goods are defined by their non-rivalrous and non-excludable nature, which clearly distinguishes them from the private goods offered by the private sector. While this dichotomy of goods is useful in conceptualizing the purposes of the public and private sectors, it does not in and of itself serve as a clear dividing line between the two sectors due to observational uncertainty.

Defining the public sector is, as observed prior, difficult. With precise definitions lacking in extendibility and internationally applicable definitions remaining abstract to allow for different interpretations, international comparisons between the public sectors of multiple countries remains complex. While classifying some entities as part of the public sector, for example ministries, is obvious and universally agreed upon, problems arise in edge cases such as corporations, which act with a profit incentive.

### **2.1.2 Practical constraints**

For the purposes of this thesis, a definition of the public sector should be consistent across countries, unambiguous, and straightforward to apply. While international definitions such as those outlined in the SNA 2008 (United Nations et al., 2009) and ESA 2010 (Eurostat, 2013) are broadly comparable, they allow for enough interpretive flexibility that cross-country discrepancies can arise.

The most significant ambiguity concerns corporations. Including corporations in the public sector introduces substantial variation. First, corporations can substantially differ from other government entities with regards to their operational motives. Their services may not be of equal importance to government services, and their mode of operation may allow them to enact different digital infrastructure policies. Second, the degree to which corporations are included in the public sector differs between

countries. If one country includes for-profit entities in the public sector, while another doesn't, the results may become distorted and patterns in infrastructure decisions obscured.

The other notable constraint concerns the inclusion of subnational government entities such as municipal administrations. While these entities are universally recognized as part of the public sector in formal definitions (Eurostat., 2013; United Nations et al., 2009), their inclusion in this thesis would introduce significant complexity without improving analytical comparability. Furthermore, subnational government entities are often localized and not necessarily representative of nationwide infrastructural trends. Trends regarding the digital infrastructure of subnational government entities could be explored in a study conducted on a national scale, but for the purposes of this thesis, these are ruled out.

These considerations indicate the difficulty of applying a single, internationally comparable definition of the public sector. Corporations pose problems of intent and comparability, while subnational governments introduce complexity without improving national representativeness. For the purposes of this thesis, these constraints necessitate a narrower working definition. The precise operational definition of “public sector” adopted in this thesis is presented in Chapter 3 as part of the methodological design.

## **2.2 Public sector digital services**

Having defined the institutional boundaries of the public sector, this section turns to the services these organizations deliver. The shift from traditional to digital service delivery (Dunleavy et al., 2006; OECD, 2014) has fundamentally transformed how governments interact with citizens, driven by demands for greater efficiency, accessibility, and (Janowski, 2015; OECD, 2014). This section explores what constitutes a public sector digital service, examines the spectrum from purely informational to fully transactional services, and establishes a critical distinction between citizen-facing and internal services.

Public services represent the mechanism which governments fulfil their obligations to citizens, delivering everything from healthcare and education to social protection and regulatory oversight (OECD, 2017). Traditionally delivered through physical offices, paper forms, and face-to-face interactions, these services have undergone radical transformation in the digital age. The OECD (2014) defines digital government as “the use of digital technologies, as an integrated part of governments’ modernisation strategies, to create public value,” indicating a shift from merely digitizing existing processes to fundamentally reimagining governance and service delivery.

The drivers of this digital transformation are multifaceted. Dunleavy et al. (2006) identify the “digital-era governance” paradigm as a response to both the limitations of the New Public Management (Hood, 1991) and the

opportunities presented by technological advancement. Digital channels offer governments the ability to achieve seemingly contradictory goals of reduced costs and improved service quality and increased accessibility while maintaining security. Furthermore, these digital channels enable further personalization while standardizing processes. As noted by Janowski (2015), digital transformation in a government doesn't happen immediately, but instead progresses through different levels of adoption beginning within government agencies and eventually reaching the level of changing government-citizen interactions. Eventually, this transformation of service delivery would enable services that could not exist within non-digital delivery channels, such as Estonia's eResidency program (Margetts and Naumann, 2017).

Digital public services exist along a spectrum of complexity and interaction. At the most basic level, informational services provide citizens with access to government information, covering everything from public health guidelines to legal documents. These represent the foundational level of digital service provision, demonstrating that even static, non-interactive content constitutes a valuable public service. Moving along the spectrum, services become increasingly sophisticated, with some offering downloadable forms and searchable databases, while others enable two-way interaction between citizens and government through online applications and submissions. At the most advanced level, integrated services connect across departments and agencies to provide seamless, proactive service delivery that crosses bureaucratic boundaries. This spectrum reflects what Layne and Lee (2001) describe as the evolutionary stages of e-government development, though as Coursey and Norris (2008) observe, not all services follow a linear evolutionary progression.

A critical distinction must be made between citizen-facing and internal administrative services. Citizen-facing services, which are those accessible to individuals, businesses, and organizations outside of the government, represent the publicly visible interface of digital government. These include, for example, tax filing systems, benefit applications, business registrations, and legislative information portals. Carter and Bélanger (2005) identify trust and perceived ease of use as crucial factors in citizen adoption of e-government services, highlighting how these services must be designed to meet citizen expectations while maintaining security and integrity. Examples of citizen-facing services range from simple informational websites of government departments to complex transactional platforms like the UK's GOV.UK or Singapore's SingPass system.

By contrast, internal services support the administrative machinery of government itself. These include case management systems, interdepartmental communication platforms, and decision support tools. While not directly visible to citizens, these services are essential for efficient government operations and indirect service delivery. Importantly, these

services can still be considered public services in that they serve the public interest by enabling more effective governance and service delivery, even if their users are government employees rather than citizens directly. This distinction between internal and external services is not always clear-cut, as many services have both public-facing components and internal administrative interfaces. For instance, a social benefits system might provide a citizen portal for applications while simultaneously offering caseworker interfaces for processing and decision-making.

This distinction between citizen-facing and internal services has significant implications for infrastructure analysis and hosting provider identification. Citizen-facing services, by their nature, must be publicly accessible via the Internet, making their technical characteristics observable through publicly available data points such as DNS queries. Internal services, conversely, typically operate within government networks, behind firewalls, or on intranets, rendering them invisible to external observation. As Scholl and Klischewski (2007) note, e-government systems often struggle with fragmentation and siloed architectures, which reinforces separation between public-facing and internal systems. Furthermore, the invisibility of internal services makes it difficult to estimate their infrastructure, since an external observer cannot know which services exist.

The observability of citizen-facing services makes them particularly suitable for the type of hosting provider prediction this thesis undertakes. These services are publicly accessible, must maintain public DNS and other records, and respond to standard Internet protocol, which means that there exist plenty of records about their infrastructure. Furthermore, citizen-facing services often represent the most critical and politically sensitive digital infrastructure, as service outages or security breaches can erode trust and thus limit digital government service success (Alshibly and Chiong, 2015).

The landscape of public sector services continues to evolve rapidly. The COVID-19 pandemic accelerated digital service adoption, with many governments fast-tracking digital initiatives that could otherwise have taken much longer to implement (OECD, 2020). This rapid digitalization has intensified debates about digital sovereignty and data localization, directly related to the hosting decisions this thesis seeks to detect and analyse.

## **2.3 Hosting**

While the previous section examined the nature of public sector digital services, delivering these services depends on hosting infrastructure. This infrastructure comprises multiple technical layers and components that work together to make websites and services accessible on the Internet. This section explores the technical foundations of web hosting, examining the various hosting models, the layered architecture of modern web services, and

the technical complexity that makes identifying hosting providers a non-trivial challenge.

### **2.3.1 Evolution of hosting infrastructure**

In the early days of networked computing, organizations requiring computational resources typically purchased, installed, and maintained their own mainframes and servers within their facilities. While time-sharing services and bureau computing existed as early as the 1960s, allowing organizations to rent computing time on remote systems (Campbell-Kelly and Garcia-Swartz, 2013), these services were different from modern web hosting. This on-premises approach dominated enterprise computing through the 1980s, requiring hardware and supporting infrastructure such as dedicated server rooms with cooling systems, uninterruptible power supplies, network connectivity, and skilled personnel.

As the World Wide Web expanded rapidly in the 1990s, specialized hosting providers emerged to manage the technical complexities of maintaining Internet-connected servers. Data centers evolved from simple colocation facilities to sophisticated managed hosting providers, offering superior network services that most organizations could not economically replicate. This evolution accelerated with the introduction of virtualization technology in the early 2000s, which allowed providers to partition physical servers into multiple isolated virtual machines, improving both resource utilization and customer isolation (Barham et al., 2003).

The introduction of Amazon Web Services' Elastic Compute Cloud (EC2) in 2006 marked the beginning of the cloud computing era, presenting the concept of Infrastructure-as-a-Service (IaaS) and fundamentally changing infrastructure provisioning (Armbrust et al., 2010). Rather than renting specific servers, organizations could now provision computing resources on-demand through APIs, scaling within minutes rather than weeks. This model, adopted subsequently by Microsoft Azure and Google Cloud Platform, abstracted away the underlying physical infrastructure entirely. The unit of hosting progressively shifted from physical servers to virtual instances, then to containers with technologies like Docker (Pahl, 2015). Today, many workloads run as serverless functions that execute without visible server infrastructure (Castro et al., 2019).

While this historical trajectory outlines the general evolution of hosting infrastructure, a key question for this thesis is where public sector digital services are situated within it. Are government websites still predominantly maintained on dedicated, on-premises servers, or have they migrated to national hosting providers' data centres and global cloud platforms? Recent research has mapped this landscape, offering empirical insights into how governments host and deliver their digital services. The next section reviews this body of work in greater detail.

### 2.3.2 Current public sector hosting landscape

While the previous section traced the evolution of hosting infrastructure in general terms, a central motivating question for this thesis is where governments locate their digital services within this landscape. In recent years, researchers have produced a body of empirical work that seeks to map the hosting arrangements of public sector domains. These studies provide insight into the balance between domestic infrastructure and reliance on global providers, the degree of centralization at infrastructural layers, and the cross-border dependencies that result from government choices.

One of the most comprehensive comparative pictures comes from Kumar et al. (2024), who analyse over one million government URLs across 61 countries, representing more than 80% of the global Internet population. Their study finds that public sector websites are far from self-contained, with approximately 62% of government URLs and 53% of delivered bytes depending on third-party providers. Simultaneously, the authors claim that most services are still domestically delivered: around 87% of government URLs are served from servers within national borders, and 77% by domestically registered organisations (Kumar et al., 2024). Regional differences are pronounced, as governments in North America overwhelmingly keep their services within their borders (~98%), while in Sub-Saharan Africa only ~52% of services remain domestic (Kumar et al., 2024).

Other studies confirm that public sector reliance on foreign providers is widespread and not confined to less technologically advanced contexts. Jansen et al. (2023) examine six countries and conclude that none achieve full infrastructural autonomy. Even presumably highly developed digital governments such as the UK and the Netherlands depend heavily on U.S.-based providers, with the top hosting organisations in their datasets being overwhelmingly American (Jansen et al., 2023). Cross-border routing is common, as, for instance, almost one fifth (18.8%) of traceroutes to UK government services exited the country (Jansen et al., 2023). The authors argue that these dependencies represent trade-offs, with governments accepting foreign reliance in exchange for resilience, scalability, and security offered by large hyperscalers.

Beyond web servers, other layers of the hosting stack demonstrate similar consolidation. Houser et al. (2022) conducted a ten-year longitudinal study of authoritative DNS for government domains across 190 countries. They show that reliance on single third-party DNS providers increased by roughly 60% between 2011 and 2020, with the largest providers serving an ever-greater share of government domains. Sommese et al. (2022) compare DNS resilience across the Netherlands, Sweden, Switzerland, and the U.S. .gov domain space, finding that approximately 80% of .gov domains depend on a

single authoritative DNS provider. They also note that Microsoft is the dominant provider of e-government email (MX) across all four cases studied (Sommese et al., 2022). Together, Houser et al. (2022) and Sommese et al. (2022) show that consolidation extends beyond web hosting into critical supporting services.

Variations between governments is nevertheless significant. Singanamalla et al. (2020) analyse 135 000 government hostnames globally in a study of HTTPS adoption and highlight divergent hosting practices through case studies. In South Korea, for instance, almost all government websites were privately hosted at the time of measurement, with only approximately 0.2% making use of cloud or CDN services (Singanamalla et al., 2020). More recent work by Silva et al. (2023) on the security posture of 3,068 government domains shows similarly uneven outcomes, noting that while some governments have adopted modern TLS configurations and robust hosting practices, others remain exposed to outdated software and known vulnerabilities. Singanamalla et al. (2020) and Silva et al. (2023) emphasize that adoption of cloud and third-party services, although widespread, is not uniform and that some governments remain dependent on older infrastructure.

Procurement and contractual data corroborate the picture of concentrated markets. Ghezzi et al. (2022, 2023) examine the government contracts and invoices in Finland, finding that Microsoft and Amazon capture a large share of government spending on hosting and cloud. These findings indicate that reliance on a limited set of global firms is visible not only in technical measurements but also in the financial records of public administrations.

Geopolitical factors can also shape the public sector hosting landscape. Jonker et al. (2022) investigate the Russian .ru domain infrastructure before and after the 2022 invasion of Ukraine. They find evidence of repatriation in providers as international sanctions and domestic policy pressures reshaped hosting arrangements. Although not limited to government services, the study illustrates how political events can prompt sudden reconfigurations in national hosting ecosystems, with potential implications for government domains.

Across this literature, a recurring theme is hybridity. Even an apparently simple government website may be supported by multiple layers of providers, including on-premises systems, commercial data centres, public cloud platforms, and content delivery networks (Varghese and Buyya, 2018). For researchers, this raises a critical problem: when services are distributed across such layers, what does it mean to identify “the hosting provider”? Different studies have answered this question in different ways, often attributing services to whichever signal is most visible. To understand the limitations of such approaches, it is first necessary to examine how hosting works in practice, which is the subject of the next section.

### 2.3.3 Technical stack of web services

A modern web service operates through multiple technical layers, each potentially involving different infrastructure providers and leaving distinct digital fingerprints. Understanding these layers is key to comprehending how hosting works in practice and what information might indicate the underlying hosting provider.

At the foundation lies the Domain Name System (DNS), which translates human-readable domain names into IP addresses (Mockapetris, 1987). When a user visits a website, their request first queries DNS resolvers to find the authoritative nameservers for that domain. These nameservers, specified in the NS records, may be operated by dedicated DNS providers like Cloudflare or Route53, by the domain registrar's default nameservers, or by the organization's own infrastructure. The DNS response typically contains A records for IPv4 addresses or AAAA records for IPv6 addresses, though increasingly common CNAME records can create chains of redirections through multiple domains before reaching the final IP address (Liu and Albitz, 2006). Each step in this resolution process provides clues about the infrastructure behind a website.

The network layer determines how traffic routes across the Internet to reach the destination server. Every IP address belongs to an Autonomous System (AS), identified by a unique AS number (ASN) assigned to organizations that control blocks of IP addresses (Hawkinson and Bates, 1996). Large hosting providers like Amazon (AS16509) or Google (AS15169) operate their own autonomous systems, making ASN lookups valuable for attribution. BGP (Border Gateway Protocol) determines how traffic routes between autonomous systems on the Internet, with BGP announcements establishing the paths that packets should follow to reach specific IP ranges (Rekhter et al., 2006).

Between users and origin servers often sits a delivery layer of intermediary services. Content Delivery Networks (CDNs) cache static content at edge locations worldwide, serving requests from geographically close servers to improve performance (Pathan and Buyya, 2007). Reverse proxies and load balancers distribute incoming requests across multiple backend servers, providing redundancy and scalability. These services fundamentally alter the technical footprint of a website, as when a CDN acts as a reverse proxy, external queries resolve to the CDN's IP addresses and ASN rather than those of the actual hosting provider.

The application layer encompasses the actual web servers and associated infrastructure that generate responses. Web server software leaves characteristic signatures in HTTP headers (Fielding and Reschke, 2014). TLS certificates, required for HTTPS connections, contain information about the domain and the certificate authority (CA) that issued them. However, widespread adoption of free certificates from Let's Encrypt (Aas et al., 2019)

has made certificate issuers less useful as hosting indicators. The specific configurations of servers, revealed through response headers and behaviour patterns, can provide hints about the hosting environment.

Email services, while separate from web hosting, provide additional attribution signals through MX (Mail Exchanger) records in DNS. These records specify which servers handle email for a domain, often revealing use of services like Microsoft 365, Google Workspace, or a specialized email security provider (Klensin, 2008). While not directly indicative of underlying hosting infrastructure, MX records may reflect an organization's broader technology choices, as organizations may be more likely to adopt Azure if they are already using Microsoft for their email.

When a user requests a website, their query traverses these layers in sequence. The process begins with DNS resolution, where the domain name is translated to an IP address, potentially through multiple CNAME redirections. The resulting IP address determines the network path via BGP routing through various autonomous systems. If a CDN is employed, the request first reaches an edge server rather than the origin server, with the CDN subsequently fetching any uncached content from the actual hosting provider as needed. Finally, the application layer processes the request and generates the response that travels back through the same infrastructure. This layered architecture is fundamental to modern networking (Peterson and Davie, 2011). An illustration of this is provided in Figure 1.

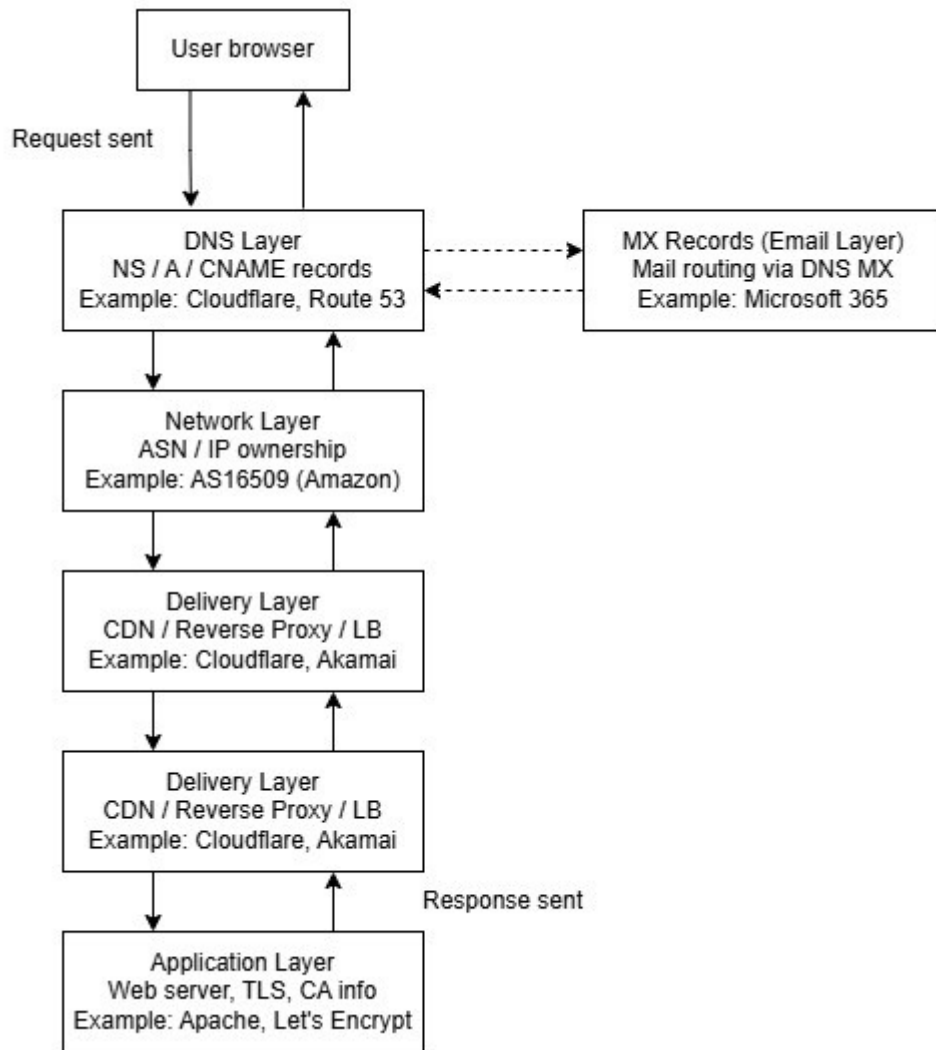


Figure 1: Illustration of the technical layers of a web service

### 2.3.4 Attribution challenge

The term “hosting provider” itself is not entirely straightforward. As infrastructures have grown more layered and modular, responsibilities for storage, computation, networking, and delivery may be distributed across multiple organisations (Plantin et al., 2018). In some cases, the term refers narrowly to the entity operating the physical servers, while in others it encompasses cloud platforms, managed service providers, or intermediary content delivery networks. This ambiguity complicates both empirical and conceptual efforts to define who “hosts” a given service. For practical purposes, many measurement studies focus on the infrastructure that ultimately serves the web application, that is, the “origin host”, while recognising that ancillary services such as DNS, CDN, or email provision contribute to the broader technical ecosystem (Ager et al., 2011; Jansen et al., 2023; Kumar et al., 2024).

The layered structure of web services described previously makes hosting attribution inherently difficult. Cloud platforms operate through distributed architectures spanning multiple data centres and availability zones, while even on-premises solutions might involve reverse proxies, load balancers, or security appliances that obscure the underlying servers. As a result, traffic to seemingly simple government websites may pass through several intermediary services before reaching the actual hosting infrastructure that contains the web servers delivering the service.

Compounding this complexity, no Internet standard requires websites to declare where they are hosted. Unlike domain registration, which is linked to WHOIS records, or TLS certificates, which identify their issuers, no protocol obliges organisations to disclose whether their services run on a public cloud, a government data centre, or some other arrangement. While some services may voluntarily expose such information, the practice is neither consistent nor reliable enough to serve as a basis for systematic attribution.

This opacity means that no single technical indicator can be taken as definitive evidence of a hosting provider. Various technical records may each point towards different organisations, depending on how intermediaries are configured. Individually, these records risk misattribution, but together they can form distinctive patterns. For example, a particular combination of DNS configurations and ASN associations may consistently correlate with a specific hosting provider, even when one of the indicators alone would be ambiguous.

These constraints carry direct methodological implications. Any viable approach to hosting attribution must integrate multiple observable records and interpret them in combination rather than isolation. Furthermore, such methods cannot be meaningfully evaluated without ground-truth data in the form of confirmed information on where a set of websites is hosted in practice. Validation against this kind of dataset is necessary to distinguish genuine attribution from coincidental correlation, and to assess how well combining different technical records can overcome the ambiguities inherent in the technical stack.

## **2.4 Methodological approaches**

While the previous section examined the technical infrastructure underlying web hosting, this section reviews methodological approaches to identifying or inferring hosting providers from observable data and institutional records. The challenge of determining which provider hosts a given website has received limited systematic attention in the academic literature. While not all of these studies directly address hosting provider identification, they offer relevant techniques and insights for understanding how infrastructure can be inferred from technical infrastructure records.

It is also important to distinguish between prediction and attribution. Existing studies typically aim at attribution in the strict sense, where the provider is inferred directly from a type of technical record. An alternative approach, explored in more recent work, treats the task as a matter of prediction under uncertainty, asking whether hosting providers can be inferred probabilistically from multiple indicators. This reframing emphasizes feasibility and reliability rather than definitive assignment and sets up the methodological discussion in the following chapter.

A notable study of service provider inference is ‘Who’s got your mail’ (Liu et al., 2021), which investigates how domains outsource their email infrastructure. By combining data from MX records, Autonomous System mappings, and SMTP/TLS observations, the authors inferred the mail service provider responsible for each domain. The study demonstrates the strength of a rule-based approach that integrates multiple heterogeneous signals, providing more robust results than relying on a single source such as IP address ownership. However, it leaves open questions about validation, since the authors did not compare their inferences to an independent ground-truth dataset. Moreover, while the rules appear well-suited to the email ecosystem, they are relatively rigid, which limits the approach’s ability to incorporate additional signals or adapt to other provider types.

Another example of inference from observable signals in the Internet measurement literature is Ager et al. (2011), who investigated the distribution of content delivery by resolving DNS queries for a large set of websites. Their “Web Content Cartograph” approach used resolution structures, including CNAME chains and redirection behaviour, to attribute websites to content delivery networks (CDNs). The study demonstrated that large-scale DNS measurements can be used to uncover the presence and prevalence of content delivery operators, showing DNS as a valuable signal for mapping web infrastructure. At the same time, the reliance on DNS alone limited the precision of their attribution, as shared infrastructures and complex delegation chains sometimes obscure the actual operator behind a service. This reinforces the potential of DNS as a signal for hosting provider inference, while also pointing to the limitations of using it in isolation.

TLS certificates have also been studied as a source of information about Internet infrastructure. Durumeric et al. (2013) conducted one of the first Internet-wide scans of HTTPS, collecting millions of X.509 certificates to examine the certificate authority (CA) ecosystem and deployment practices. Although the study did not attempt to identify hosting providers, it demonstrated that TLS certificates can serve as a useful infrastructural signal for exposing relationships between domains, hosting operators, and CAs. At the same time, the work showed the limits of certificates for attribution, showing how certificates alone are insufficient for reliable provider inference.

Hosting and cloud service provision can also be inferred through institutional procurement data rather than technical signals. At the macro

level, Appelt and Galindo-Rueda (2016) demonstrated how European procurement records, in particular data from the EU's Tenders Electronic Daily (TED), can be systematically mined to study supplier activity and innovation outcomes. Their analysis illustrates that procurement datasets contain structured information about contracts, vendors, and service categories, providing a path to identifying providers. At the organizational level, Jones (2015) presented a detailed case study of a United Kingdom public body's move to cloud services, showing how contract records and implementation documents revealed the specific vendors responsible for hosting and service delivery. In Finland, Ghezzi et al. (2022, 2023) extend this approach by analysing municipal procurement and invoice data, demonstrating a concentration of hosting and cloud contracts among Microsoft and Amazon. These studies confirm that institutional artifacts such as contracts and award notices can surface hosting and infrastructure providers. However, they also highlight limitations, as coverage can be fragmented, disclosures may lag behind current practice, and subcontracting arrangements often remain opaque.

While institutional datasets such as procurement records can reveal hosting relationships indirectly, the closest strand of research to systematic hosting provider inference has relied on technical signals. Tajalizadehkhoob et al. (2016) combined passive DNS data with WHOIS records to attribute domains to more than 45 000 distinct hosting organizations. This large-scale analysis enabled the authors to examine heterogeneity in the hosting market and to relate provider characteristics to security outcomes (Tajalizadehkhoob et al., 2016). The study demonstrated the feasibility of constructing a provider-level view of hosting from technical traces. It also revealed methodological challenges surrounding the often incomplete and inaccurate nature of WHOIS records, noting that the attribution process does not always resolve the underlying operator with precision (Tajalizadehkhoob et al., 2016). Furthermore, though the authors note that the method is inaccurate, they did not attempt to verify the results with any ground truth dataset. Thus, this approach remains limited in reliability for precise provider inference.

A more recent line of work has applied machine learning directly to hosting inference. Tarahomi et al. (2024) investigated whether reverse DNS (PTR) names can be used to identify cloud providers, training a Markov chain classifier on PTR token sequences to distinguish between major providers such as Amazon, Microsoft, and Google. By validating the classifier against known IP ranges published by the providers, the authors demonstrated that PTR records contain sufficient information to support accurate provider attribution (Tarahomi et al., 2024). However, the validation was necessarily limited in scope, as it covered only major cloud operators and relied on published ranges that may not capture all cases. The study highlights the promise of applying probabilistic methods to hosting inference, though the

validation process leaves uncertainty about the reliability of using a single technical signal.

Complementing this, Peng et al. (2025) studied authoritative name server (NS) records to study the extent of centralization across hosting providers. By collecting NS domains at scale and clustering them based on naming conventions, the authors were able to attribute large sets of domains to specific providers such as Amazon, Cloudflare, and GoDaddy (Peng et al., 2025). While the approach demonstrates the scalability of NS-based attribution, it relies on distinctive naming schemes and manual verification, which limits coverage for providers without recognizable patterns and for cases where domains delegate their NS service to third parties.

These methodological patterns are also evident in studies that focus specifically on government hosting. Kumar et al. (2024) present the largest comparative dataset to date, but their attribution method relies primarily on Autonomous System (AS) mappings. This approach produces broad coverage yet risks conflating network operators with service providers and does not attempt validation against ground truth. Jansen et al. (2023) similarly combine traceroutes with AS resolution to evaluate digital sovereignty across six countries. While effective at showing cross-border dependencies, their attribution of hosting providers remains indirect, and again validation is absent. Other government-focused studies have made use of DNS and mail records. Houser et al. (2022) conduct a longitudinal analysis of authoritative DNS, while Sommese et al. (2022) examine DNS resilience and MX records in four national contexts. Both illustrate the value of DNS-based approaches but also show their limitations, as inference often depends on naming conventions and visible delegation rather than contractual arrangements. Finally, Singanamalla et al. (2020) apply Internet-wide TLS scanning to 135 000 government hostnames to evaluate HTTPS adoption, indirectly revealing hosting practices. Their reliance on certificate metadata provides scale but only partial accuracy in attributing providers. These studies demonstrate that even government-specific work has tended to depend on straightforward applications of technical records, producing interesting descriptive findings but without systematic validation or integration of multiple technical records.

In summary, prior work shows that provider inference is possible both from technical infrastructural records and institutional records, but existing approaches remain partial. Most rely on a single type of technical record, leaving them vulnerable to incompleteness or noise. Furthermore, with methods that don't use institutional data, validation against ground truth has been limited or absent. Taken together, these limitations indicate a need for methods that combine multiple heterogeneous technical infrastructural records and are subject to systematic validation against independent sources of infrastructural transparency.

## 2.5 Summary and research motivation

Existing research consistently finds that governments do not host their digital services independently but instead depend on a mixture of national infrastructure and foreign providers. Comparative studies reveal that governments host many of their services domestically yet rely heavily on third-party providers such as Microsoft and Amazon, with considerable variation across regions and infrastructural layers. These findings show that governments rarely achieve infrastructural autonomy, as their services are embedded in a hybrid landscape of national and foreign dependencies. Here, infrastructural autonomy refers to the capacity of a government to operate and control the technical infrastructures behind its digital services without reliance on external providers. However, while such studies provide valuable descriptive evidence, their methodological underpinnings are limited. Attribution often depends on a single type of technical record that doesn't directly imply hosting, and validation against independent sources is rarely attempted.

Work on inference methods shows the potential of technical traces as well as institutional records such as procurement data. Yet the former are vulnerable to noise and ambiguity when used in isolation, while the latter are incomplete, fragmented, and difficult to acquire on a larger scale. Studies have demonstrated both the feasibility and the limitations of these approaches. While they produce important empirical insights, their reliance on simple attribution without systematic validation leaves open questions about accuracy.

Together, these strands reveal a clear research gap. To date, no study has systematically combined multiple heterogeneous technical records with institutional sources of ground truth to estimate who the hosting of public-sector websites. As a result, our understanding of governments' infrastructural dependencies remains partial: we know that foreign reliance exists, and that global providers dominate in many contexts, but we cannot reliably quantify or validate the extent of this dependence. This gap is not merely technical but also political, since without robust methods of hosting provider estimation it is difficult to assess questions of sovereignty, transparency, and resilience.

Addressing this gap motivates the present study, which asks:

**RQ:** Can the underlying hosting provider of a public-sector digital service be predicted using publicly observable information?

The following chapter outlines a novel methodology for addressing this problem that overcomes some of the main limitations of the methodologies used in previous research.

## **3 Research methodology**

This chapter sets out the methods used to evaluate whether publicly observable information can be employed to infer the hosting provider of public-sector websites. The aim is not to build a production-ready classifier, but to explore the feasibility and limitations of inference from observable technical records.

The methodology proceeds in three stages. First, a ground-truth dataset is assembled by submitting Freedom of Information (FOI) requests to public-sector bodies in three countries regarding the hosting of government websites. Second, an automated process collects publicly observable infrastructure signals for the same set of websites. Third, a predictive model is trained to predict provider labels from the collected signals, generating estimates that can be compared against the FOI ground truth. The remainder of this chapter describes each stage in detail: first the FOI dataset, then the collection and processing of technical signals, followed by the predictive modelling approach, and finally the evaluation design.

### **3.1 Ground truth**

The first stage of the methodology establishes a ground-truth dataset of hosting arrangements through Freedom of Information (FOI) requests. While observational infrastructure data can indicate likely providers, validation requires an authoritative reference point. FOI requests provide a mechanism to obtain such information directly from government organisations, making it a suitable means of establishing baseline labels for this study. The following subsections describe the FOI frameworks and procedures used, the scope of the targeted organisations, and the processing of responses into a structured dataset.

#### **3.1.1 FOI background and legal considerations**

The primary source of ground-truth data for this study was information released under Freedom of Information (FOI) legislation. FOI laws grant the public a legal right to request information from government bodies, subject to certain exemptions (Banisar, 2006). These laws vary in scope and practice across countries, but in general they provide a structured mechanism to obtain authoritative information about government activities (Hazell and Worthy, 2010). Compared to alternative sources such as procurement databases or public contracts, which can be inconsistent and difficult to interpret (Prier and McCue, 2009), FOI requests allow for more targeted and timely information to be retrieved.

The dataset was designed to cover three countries: the United Kingdom, Finland, and the Philippines. All of the three countries have established FOI systems (*Act on the Openness of Government Activities (621/1999)*, 1999; *Freedom of Information Act 2000*, 2000; *Executive Order No. 2*, 2016). The presence of established FOI systems was a prerequisite for country selection. Simultaneously, differences in FOI practices inevitably shaped the approach. Unlike the UK and Finland, where FOI requests may be filed by any individual, the Philippines' Executive Order No. 2 (2016) limits FOI requests to Filipino citizens. Consequently, requests in the Philippines were filed by a collaborator with citizenship.

From a legal and ethical standpoint, the use of FOI responses as data raises minimal concerns. Once released, FOI responses are public information by default and can be freely cited, reproduced, and shared (Banisar, 2006). The data did not include personal information about individuals, and all requests were made through official channels within the bounds of each country's legislation. Some organisation invoked FOI exemptions to withhold information, citing security or confidentiality grounds. These refusals were respected and are recorded in the dataset as "Withheld".

### **3.1.2 Scope**

For the purposes of this study, the operational definition of the public sector was chosen in a way that could be consistently applied across countries. The scope was defined as central government ministries, nationwide bureaus and agencies, and arm's-length public bodies under their control. Services directly operated by these organisations were also included. Publicly owned corporations were excluded, as their commercial interests and inconsistent classification across countries would have reduced comparability. Subnational governments, such as municipalities, were also excluded, since their inclusion would have introduced substantial complexity without improving representativeness at the national level. These exclusions were consistent with the constraints discussed in Chapter 2.

Countries were chosen based on two primary criteria: the availability of a consistent institutional framework for requests, and their potential to illustrate different hosting alignments. The UK was expected to be the most closely aligned with US-based providers, while Finland was expected to have a higher share of local hosting alongside international providers. Although the Philippines is a U.S. ally, the presence of Chinese infrastructure providers' data centres in the country (Lehdonvirta et al., 2025) suggested possible reliance on Chinese hosting. Limiting the scope to three countries was a pragmatic decision to ensure that the data collection could be completed within the timeframe of the project, while still allowing for comparison across distinctly different hosting environments.

Within each country, requests were directed to national-level organisations according to publicly available lists. For the UK, central departments, their associated domains, and any clearly linked subdomains were identified through the gov.uk portal. In Finland, ministries and their subordinate agencies were mapped through official government websites, as well as additional subdomains listed on the sites of the departments and agencies. In the Philippines, requests were targeted at the main government departments which were expected to provide the clearest responses. The number of target websites per country was set in the range of 100-200, balancing feasibility with the need for sufficient variation in the dataset. For the UK and Finland, websites were identified and requested individually, with requests sent to a government agency containing a single or multiple associated URLs. In the Philippines, requests were broader and asked departments to provide information on any information systems they would be willing to share information on. Although multiple requests were sometimes required to obtain clarification from a single organisation, only one answer was logged per website.

For the purposes of this study, “hosting provider” refers to the organisation responsible for operating the servers that deliver the web application’s content, corresponding to the application layer in the technical stack described earlier. FOI requests were phrased to capture this layer of infrastructure, though in practice some organisations interpreted the question more broadly, referring instead to service integrators or umbrella organisations managing multiple systems. Such cases were harmonised manually to ensure comparability. While this definition necessarily abstracts from hybrid or multi-provider arrangements, it reflects the most practical level at which hosting can be operationally identified for modelling purposes.

### **3.1.3 FOI request procedure**

The FOI requests were sent and received between May and August 2025. This period was initially intended to be shorter but was extended due to some cases where responses took longer. The results provide a snapshot of hosting arrangements during this period, though such arrangements may change over time.

Requests for the United Kingdom and Finland were sent directly by the author. The FOI data concerning the Philippines’ public sector organizations is discussed in Kilpi et al. (2025, unpublished manuscript). Requests were primarily sent from a role-specific email address under the aalto.fi domain, which was established for organisational purposes. A small number were inadvertently sent from a personal aalto.fi address, but this did not appear to affect response rates or outcomes, as all requests were signed with the author’s name.

Requests were structured around short, direct templates designed to elicit clear and comparable answers. One template was prepared for each country, though the overall style was uniform across countries. The purpose of the templates was to ensure that all requests would be formal, legally grounded, and phrased to reduce ambiguity. The UK and Finnish templates explicitly cited the relevant FOI legislation, requested information in a structured format, and included the URLs of the websites concerned. The Philippines requests were broader and requested departments to provide information on any of their information systems. Minor grammatical and terminological adjustments were made during the process, but these did not alter the content of the requests.

The information requested was limited to three elements:

1. identification of the hosting arrangement,
2. the commercial providers involved,
3. the year since when the arrangement had been in place.

Requests did not ask for contract documents, as provider names were considered sufficient for building the dataset. To reduce opportunities for vague replies, the templates were worded to require specific information rather than leaving room for a generic “not known” response.

Responses typically arrived within the legal timeframes for FOI requests, averaging two to three weeks. Some organisations withheld information on legal or security grounds, particularly for sensitive or critical systems. Others did not respond despite follow-ups.

The email templates are included in the appendix.

#### **3.1.4 Response processing and label assignment**

The status of all target URLs was tracked in a structured spreadsheet, with each entry corresponding to a specific website or domain. For each entry, fields included the name of the entry, the URL, whether a request had been sent and answered, notes on the request or the site, notable content of any response, the start date of the hosting arrangement (if provided), the reported hosting provider, the deduced hosting provider label, and any additional details. This ensured that every URL in the sample had a recorded status, regardless of whether or not a usable FOI response was obtained.

Responses were classified immediately upon receipt into four categories to describe the status of the requested information:

- Yes: responses that were considered complete even if provider was ultimately unknown.
- Maybe: incomplete responses for which further clarification could plausibly be obtained. These were reclassified as Yes if later clarified, and if not, they were excluded from further analysis.
- No: no response was received. No was also the default status until a response was received.
- Withheld: responses where the organisation explicitly declined to disclose the information, typically citing security concerns under FOI legislation as well as legal exemptions

Ambiguities were handled consistently. A common example was responses that identified only a developer or contractor (e.g. “the website was developed by Company X”) without mentioning hosting. These were coded under the developer’s name but later mapped to “Other” during analysis. In cases where the hosting was handled through a managed hosting company with a direct link to a bigger cloud provider, the result was mapped directly to the associated cloud provider. When organisations redirected requests between departments without resolution, the status was left as “Maybe”. Contradictory responses were not encountered.

The structured templates kept ambiguity relatively low, but not all requests yielded usable results. For the UK, 61 URLs received complete responses (56.5% of requests sent), with 22 incomplete (20.4%) and 16 withheld (14.8%). For Finland, 139 URLs received complete responses (81.3%), with 13 incomplete (7.6%) and 8 withheld (4.7%). In both cases, the remainder represented non-responses. Disclosures were withheld primarily in the areas of defence and security. For the Philippines, comparable statistics could not be calculated because the open-ended request format meant that only the final number of systems could be identified. However, response rates to overall requests were tracked so that a response rate of 61.4% and a withheld rate of 7.0% were identified. The results are shown in Table 1.

Table 1: Response rates per country

<b>Country</b>	<b>Targeted URLs</b>	<b>Complete responses</b>	<b>Incomplete responses</b>	<b>Withheld responses</b>	<b>No response</b>
UK	108	61 (56.5%)	22 (20.4%)	16 (14.8%)	9 (8.3%)
Finland	171	139 (81.3%)	13 (7.6%)	8 (4.7%)	11 (6.4%)
Philippines	n/a	66 identified	n/a	n/a	n/a

Only responses with identifiable assigned hosting provider labels were retained for further analysis. The classification process was carried out manually by the author, with decisions documented to enable reproducibility

in future work. All organisations were weighted equally in this set, meaning that small agencies and high-profile central portals contributed a single data point each. This ensured consistency for modelling purposes, even though it does not reflect differences in institutional size or visibility.

### **3.1.5 Supplementary data on Chinese hyperscalers**

While the ground-truth dataset constructed through FOI requests offered extensive coverage of the United Kingdom, Finland, and the Philippines, no examples of Chinese hyperscale providers appeared in those responses. This absence limited the ability of the study to assess whether predictive methods would meaningfully distinguish cases of Chinese cloud adoption, which is an increasingly important dimension of global infrastructure competition (Lehdonvirta et al., 2025). To address this limitation, a small supplementary dataset of government websites hosted on Alibaba Cloud, Huawei Cloud, or Tencent Cloud was assembled from alternative sources.

The supplementary cases were identified through a combination of procurement contracts, award notices, and provider press releases. Each case was only included where the hosting arrangement was declared explicitly in an official or otherwise authoritative document. To maintain relevance to the current market, only sources from 2022 onwards were considered, with the majority dating to 2025. This process yielded 19 confirmed domains, each linked to a corresponding public body and to one of the three major Chinese hyperscalers. While several of these failed to meet the criteria previously set as the operational definition of the public sector in Section 3.1.2, their inclusion was judged to be necessary given the scarcity of public evidence. Since the purpose of this dataset is primarily to enable training and evaluation of models rather than to characterise national-level adoption, the impact of this departure from the main scope is expected to be limited.

The supplementary data points were treated in the same way as the FOI responses. Each domain was processed using the same pipeline for collecting observable infrastructure records, and the full combined dataset was queried simultaneously to ensure comparability in timing and format. This integration ensured that the Chinese cases could be analysed using the same features as the FOI-derived domains.

Nevertheless, the supplementary dataset is inherently less reliable than the FOI responses. The number of cases is small, creating difficulties when they are treated as separate classes during model training and validation. In stratified cross-validation, for instance, classes with only a handful of examples may not appear in every fold, which reduces the stability of performance estimates. Furthermore, while FOI responses provide authoritative ground truth, contract announcements and press releases are one step removed, and thus carry a greater risk of inaccuracy or incompleteness. For these reasons, the supplementary data should be viewed

as an extension of the scope of the analysis rather than providing results of equal certainty.

## **3.2 Observational data**

In addition to the ground-truth data obtained through FOI requests, this study collected a parallel set of observational data from publicly visible infrastructure. This dataset consists of technical indicators associated with each target domain, such as DNS records, TLS certificate issuers, and related metadata. Unlike the FOI responses, which provide official disclosures, the observational data is derived entirely from automated external queries. It captures what can be learned about the hosting environment of a website from the “outside,” without privileged access. Whereas FOI responses provided authoritative baseline labels of hosting arrangements, the observational data offered complementary external traces that could be systematically compared against those labels and transformed into input features for predictive modelling.

### **3.2.1 Data sources and collection**

While several record types were collected during the process, four categories were ultimately retained as core features: NS records, MX records, TLS certificate issuers, and ASN descriptions. Other records, such as A, CNAME, and TXT, were gathered to enable or contextualize these features, but were not directly used in the following modelling stage.

NS records served as the primary DNS-based feature. They identify the authoritative name servers for a domain, which often map directly to specific infrastructure providers. To ensure that hidden dependencies were not missed, CNAME chains were followed to their terminal targets before querying for NS records. A records were also collected, but only as an intermediary step toward WHOIS and ASN lookups rather than as features themselves.

MX records were treated as a distinct signal because they reflect organisational choices about email hosting, which may diverge from website hosting. Queries were first attempted at the exact domain. If no record was found, queries were retried by progressively removing the leftmost subdomain label. This ensured that if email was handled at a higher-level organisational domain, that provider would still be captured as a feature. The process stopped before reaching the public suffix (e.g., gov.uk), since suffix-wide records reflect umbrella namespaces rather than the infrastructure of individual organisations.

TLS certificate issuers were extracted by attempting a TLS handshake (Rescorla, 2018) on port 443 and reading the organisation field of the returned certificate. Although issuers typically represent certificate

authorities rather than hosting providers, they provide an additional trace of infrastructural dependencies, and in some cases may correlate with service arrangements.

ASN descriptions were obtained from RDAP and WHOIS lookups of the IP addresses returned in A records. For each public IP, the corresponding Autonomous System Number and its description were recorded. In most cases, these values reflect the delivery network visible to the outside world rather than the ultimate hosting backend (Ager et al., 2011). This makes ASN descriptions especially useful for identifying reliance on intermediary infrastructure providers. When a domain resolved to multiple IPs, the resulting ASN descriptions were deduplicated. In practice, this collapsed to a single provider for nearly all domains in the dataset, though the method allowed for multiple values if a domain’s IPs spanned different ASNs. The next section describes the collection of observational records, followed by cleaning steps and the construction of label groupings for evaluation.

### **3.2.2 Data cleaning**

The raw outputs of the collection process required several preprocessing steps before they could be used as features for predictive modelling.

First, only domains with a valid ground-truth label were retained. Entries with ambiguous labels were excluded, including those marked “Unknown”, “DICT”, “Hybrid”, or “Multicloud”, as well as one case where the response listed both public sector and commercial providers (“Valtori, Azure, AWS”), leaving the actual hosting structure ambiguous. The label “DICT” was used by some respondents to refer generically to the Department of Information and Communications Technology rather than an actual hosting provider and was treated as ambiguous as the hosting provider used by DICT was left unclear. These exclusions removed labels that could not be reduced to a single clear provider. After filtering, the dataset contained 250 domains with primarily unambiguous hosting classifications. In some responses, organisations had mentioned only a software development company and not a hosting provider. While ambiguous, these were kept under the company’s name so that they could later be reassigned if information about the company’s infrastructure partnerships became available.

Second, signal values that could yield multiple raw results were simplified into a single representative feature per domain. For NS records, multiple authoritative name servers belonging to the same provider were collapsed into a single identifier. For MX records, domains could specify multiple mail servers. These were first mapped to provider identities and then collapsed into a single representative value, so that each domain retained only one MX provider. Except for one ASN description value, TLS certificate issuers and ASN descriptions resolved to one value per domain and required no further

simplification. The exception was treated similarly to the others, as it clearly implied a different hosting arrangement.

Third, string normalisation was applied where necessary to ensure consistency. In particular, “MX\_provider” values were lowercased so that functionally identical records (“google.com” vs. “GOOGLE.COM”) were treated as the same category. TLS issuers were also canonicalized, but only in one case: multiple variants of “DigiCert Inc.” were collapsed into a single consistent label. No other issuers showed comparable similarity with each other, so the remaining values were left unchanged. Other string-based features were preserved in their original form to reflect the reporting conventions of the respective authorities.

The dataset was structured from the outset with one row per domain. Preprocessing preserved this format while simplifying the technical records into four final features: NS provider, MX provider, ASN description, and TLS issuer. Supplementary fields collected during preprocessing were also retained but not used in the modelling stage. After preprocessing, the dataset contained 250 domains across three countries, each with a single unambiguous hosting label and four core features. These structured data points formed the input for the predictive modelling stage described in the next section. This figure refers to the FOI-derived dataset only. An additional 19 domains from a supplementary dataset on Chinese hyperscalers are introduced in Section 3.4, bringing the full set to 269.

As preparation for model training and evaluation, it was necessary to decide how the FOI-disclosed hosting providers would be mapped into classes for training. Without grouping, the dataset would contain dozens of individual provider labels, many with only a single occurrence, making systematic evaluation impractical. To address this, four alternative label schemes, referred to as ontologies, were defined, each collapsing providers into broader categories while preserving different analytical emphases

The order of these ontologies reflects the motivating concern of this thesis: the extent to which public sector hosting depends on either U.S.- or China-based hyperscalers. The first two groupings therefore distinguish these provider blocs explicitly, while the latter two test simplified variants that merge U.S. providers or collapse all foreign hyperscalers together.

1. Individual US and Chinese hyperscalers, Other (seven labels)
2. US hyperscalers grouped, Chinese hyperscalers grouped, Other (three labels)
3. Individual US hyperscalers vs. Other (AWS, Azure, Google Cloud, Other).
4. Binary US hyperscalers grouped vs. Other

The first ontology distinguishes between major providers in both the U.S. (AWS, Azure, Google Cloud) and China (Alibaba Cloud, Tencent Cloud,

Huawei Cloud), with all remaining providers grouped under “Other.” This ontology captures the most detailed geopolitical distinction. The second ontology collapses the individual hyperscalers into U.S. and Chinese blocs to test whether the model can still differentiate at this coarser geopolitical level. The third includes the three main U.S. hyperscalers as separate labels and merges all other providers into a single “Other” category, abstracting Chinese providers into the “Other” category, while providing details on specific U.S. providers market shares. Finally, the simplest binary case tests that U.S. hyperscalers as a group can be reliably separated from all other providers.

The models are trained and validated separately under each scheme, allowing for performance to be compared across alternative ways of structuring the label space. These label groupings are referred to as “ontologies.” The term is used here not in the strict philosophical sense, but in the practical sense common in information science, to denote an explicit scheme for categorising entities (Gruber, 1993). Each ontology represents a different way of mapping the raw provider labels into analytically tractable classes. Using multiple ontologies allows us to test whether results depend on the choice of grouping, and to balance between granularity and statistical stability.

### 3.2.3 Class balance and baseline

Table 2 summarises the number and percentage of domains in each class under the four ontologies.

Table 2: Domain distributions under four ontology schemes

<b>Ontology</b>	<b>Label</b>	<b>n</b>	<b>%</b>
<b>US + Chinese hyperscalers separate</b>	AWS	108	40.1%
	Azure	41	15.2%
	Google Cloud	3	1.1%
	Alibaba Cloud	8	3.0%
	Huawei Cloud	7	2.6%
	Tencent Cloud	4	1.5%
	Other	98	36.4%
<b>US + Chinese hyperscalers grouped</b>	Big US cloud	152	56.5%
	Big Chinese cloud	19	7.1%
	Other	98	36.4%
<b>US hyperscalers separate</b>	AWS	108	40.1%
	Azure	41	15.2%
	Google Cloud	3	1.1%
	Other	117	43.5%
<b>US hyperscalers grouped</b>	Big US cloud	152	56.5%
	Other	117	43.5%

The distributions show both the advantages and the limitations of each scheme. Ontology 1 (US and Chinese hyperscalers separate) offers the most fine-grained view, distinguishing seven classes. This maximises potential detail but also results in strong class imbalance, as AWS alone covers 40.1 percent of the dataset while several other classes have fewer than ten observations. Ontology 2 (US and Chinese hyperscalers grouped) merges these into broader blocs, enabling direct comparison between US- and Chinese-based hyperscaler cloud provision while improving balance.

Ontology 3 (US hyperscalers separate) focuses only on the three major US providers and merges all others into a single “Other” category. This retains some provider-level interpretability and is particularly useful for examining which of the US providers is dominant but sacrifices the US-China distinction. Finally, Ontology 4 (US hyperscalers grouped) represents the simplest configuration, collapsing all major US cloud platforms into one versus all others. This scheme provides the cleanest binary comparison but at the cost of losing intra-US differentiation. Among all the ontologies, it produces the highest class balance.

The majority-class baseline, which is the accuracy achieved by always predicting the largest class, varies across ontologies. It is 40.1% for Ontologies 1 and 3, and 56.5% for Ontologies 2 and 4. These figures establish reference points for later model evaluation, as meaningful predictive performance must substantially exceed these baselines.

### **3.3 Modelling approach**

The final stage of the methods involved building predictive models to predict hosting providers from the collected observational data. Whereas the FOI responses provided authoritative ground-truth labels, and the preprocessing pipeline structured the technical records into categorical features, the modelling stage operationalized this dataset into a supervised classification problem. The aim was to test whether observable infrastructure traces could reliably predict the hosting arrangements disclosed through FOI. To this end, a set of candidate models were trained and validated under multiple label groupings. This stage was designed not only to assess predictive accuracy, but also to compare different modelling strategies in terms of interpretability and their suitability for extension to new datasets.

#### **3.3.1 Feature engineering**

As described in Section 3.2, the cleaned dataset includes four categorical infrastructure features. Since these features were non-numeric and, in many cases, high-cardinality, they were transformed into machine-learning-ready vectors using one-hot encoding (Pedregosa et al., 2011). An example of this is provided in Table 3.

Table 3: Example of one-hot encoding

Domain	NS=Google	NS=AWS	TLS=Google	TLS=AWS
gov.uk	1	0	0	1
valtioneuvosto.fi	0	1	1	0

Before encoding, missing values were made explicit by assigning them to a single placeholder category “<NA>”. This ensured that domains without, for example, an MX provider or TLS certificate could still be represented consistently in the feature space. One-hot encoding (Pedregosa et al., 2011) was then applied separately to each feature column. Each unique value within a feature became a binary indicator column, with presence marked as 1 and absence as 0.

The encoder was configured to ignore unseen categories during the prediction phase. This design choice was necessary because new domains outside the training set may contain providers not previously observed; ignoring such categories prevents inference from failing while still retaining comparability for known features. Although this discards some information, it was considered preferable to misclassifying unseen providers into an inappropriate category, and alternatives would have likely produced less stable training and model results given the small and imbalanced dataset. To manage memory use while preserving interpretability, the encoded matrix was stored in sparse form (Virtanen et al., 2020).

Across the 269 labelled domains, one-hot encoding resulted in a feature space of 163 binary variables: 47 categories for NS providers, 45 for MX providers, 50 for ASN descriptions, and 21 for TLS issuers. 75 of these were “singletons”, which are categories that appeared for only one domain. While sparse, they were retained, as rare providers may still be characteristic of specific hosting arrangements.

### 3.3.2 Candidate models

To evaluate the predictive potential of the collected features, a range of modelling approaches was considered, spanning from simple baselines to more complex ensemble methods. The selection was guided by three priorities. First, the models needed to be suitable for sparse, high-cardinality categorical features. Second, they had to generalise well from a relatively small dataset of 269 domains. Finally, they needed to remain as interpretable as practically possible, since understanding which signals drive predictions is an important part of the analysis, even if some higher-performing ensembles are less transparent.

The simplest benchmark was a majority-class baseline (Pedregosa et al., 2011), which always predicts the most common hosting provider observed in the training data. Although trivial, this baseline establishes a meaningful lower bound: any useful model should be expected to perform substantially

better, and thus its inclusion provides context for interpreting accuracy scores. Building on this, multinomial logistic regression (Bishop, 2006, ch. 4) with one-hot encoded categorical features was chosen as the primary linear classifier. Logistic regression is well suited to sparse, high-dimensional data and provides interpretable estimates of class probabilities, which is valuable for uncertainty-aware comparisons (Bishop, 2006, ch. 4). Its main limitation lies in the difficulty of capturing complex, non-linear interactions between signals, such as cases where a CDN or intermediary network may front for an underlying cloud provide.

Tree-based methods (Quinlan, 1986) were also included, as they are a natural fit for categorical infrastructure features. Single decision trees offer transparent, rule-based classifications that can be directly inspected to understand which features drive predictions, though they are prone to overfitting in small datasets (Quinlan, 1986). Random forests mitigate this risk by averaging across multiple bootstrapped trees, typically improving predictive performance while offering interpretable feature importances (Breiman, 2001). Gradient boosted trees, such as XGBoost (Chen and Guestrin, 2016), extend this idea by sequentially building trees that correct the errors of earlier ones (Friedman, 2001). These models are often state-of-the-art in tasks involving structured categorical data, and while they sacrifice some interpretability compared to simpler tree models, they provide a useful benchmark for assessing the upper bound of achievable accuracy with the available dataset (Chen and Guestrin, 2016).

Several alternatives were considered but excluded as inappropriate for this task. Naïve Bayes classifiers (Domingos and Pazzani, 1997) were ruled out because their assumption of conditional independence between features does not hold in this dataset, where signals such as NS providers and TLS issuers can often be expected to co-occur. k-Nearest Neighbours (Cover and Hart, 1967) was rejected because distance metrics become less meaningful in high-dimensional one-hot spaces, and computational cost scales poorly. Support Vector Machines were also excluded, as they do not handle sparse, high-cardinality categorical data efficiently and provide limited interpretability (Cortes and Vapnik, 1995). Neural networks were set aside for similar reasons, as they are likely to overfit with only 269 labelled samples on top of not providing enough transparency needed for the analysis (Goodfellow et al., 2016, ch. 7). Table 4 summarises the considered models.

Table 4: Overview of candidate models

Model type	Included?	Justification
Majority-class baseline	yes	Establishes lower bound
Logistic regression	yes	Interpretable, handles sparse high-dim data
Decision tree	yes	Transparent rules
Random forest	yes	Less prone to overfitting than decision trees, feature importance
Gradient boosting (XGBoost)	yes	Handles complex patterns
Naïve Bayes	no	Assumes independence, unrealistic for this data
kNN	no	Poor in high-dim sparse data
SVM	no	Inefficient for sparse categorical data
Neural networks	no	Risk of overfitting, low interpretability, unsuited for small dataset

In summary, the final set of models consisted of a majority-class baseline, logistic regression, decision trees, random forests, and gradient-boosted trees. The next subsection outlines the training and validation procedures used to compare these models in a consistent and reliable way.

### 3.3.3 Training and validation

Reliable model evaluation depends on how the available data is partitioned into training and validation sets. In this study, this was particularly important given the dataset size of 269 labelled domains and the presence of class imbalance between frequent providers like AWS and rarer ones like Google Cloud.

The simplest approach is to split the dataset once into a training set and a held-out test set. This provides a clean estimate of generalisation but is highly sensitive to how the data happens to be divided, particularly in small samples (Kohavi, 1995). A single unlucky split could, for instance, leave a minority provider absent from the training set altogether, making the evaluation unrepresentative. Even if all classes are present, the resulting performance estimates may vary substantially depending on the random split.

At the other extreme, leave-one-out cross-validation (LOOCV) trains a separate model for each observation, using all remaining data for training and evaluating on the single left-out sample (Kohavi, 1995). LOOCV makes maximal use of the dataset, but with 269 folds it would be computationally expensive and prone to high variance in estimates, since each test set contains only one domain. The method also does little to address class imbalance, as rare providers may contribute only a handful of scattered test cases.

Between these extremes, k-fold cross-validation provides a balanced compromise. In the method, the data is partitioned into k folds of roughly equal size, with each fold used once for validation while the others serve for training (Kohavi, 1995). This provides multiple estimates that can be averaged, reducing sensitivity to any particular split. As Gorriz et al. (2024) note, k-fold cross-validation is the most widely adopted method for estimating generalisation performance. Common choices are k=10, which reduces bias produces relatively small folds and thus higher variance when minority classes are present, or k=5, which creates larger folds and thus more stable estimates.

For this study, stratified k-fold cross-validation (Pedregosa et al., 2011) with k=5 was adopted. (SHOW THROUGH EXAMPLE?) Stratification ensures that each fold maintains approximately the same class distribution as the full dataset, reducing the risk that rare providers disappear from folds. For example, if 40% of domains are hosted on AWS and 15% on Azure in the full dataset, stratified k-fold ensures that these proportions are preserved in each fold. Without stratification, some folds might be dominated by AWS domains while containing very few Azure cases, leading to unstable estimates of model performance. Using five folds strikes a balance between computational feasibility and robustness. Each model is trained on 80% of the data and validated on 20%, while the repetition across folds provides an averaged performance estimate that is less sensitive to any single split. This proportion is simply a consequence of the chosen fold count rather than a separate design choice.

Finally, to ensure reproducibility, random seeds were fixed for fold generation and model training. As Beam et al. (2020) emphasise, “the only way to ensure that the results of these models are reproducible is to set a quantity known as the random seed.” Fixing seeds guarantees that repeated runs yield identical splits and outcomes, enabling meaningful comparisons across models while still preserving the general principle of randomness.

This stratified 5-fold procedure provides a reliable and widely recognised framework for evaluation and forms the basis of the results reported in chapter 4.

### **3.3.4 Evaluation**

The performance of the models was evaluated using a set of metrics chosen to reflect both overall predictive accuracy and the ability to handle class imbalance. Accuracy, which measures the proportion of correctly classified classes, was included as the most intuitive and widely understood measure, providing a straightforward comparison to the majority-class baseline (Sokolova and Lapalme, 2009). Any useful model should substantially exceed this baseline, making accuracy a natural starting point for analysis.

However, accuracy alone risks obscuring whether the classifier is only performing well on the dominant class.

To address this, results were also reported in terms of precision, recall, and F1-score for each class (Powers, 2020; Sokolova and Lapalme, 2009). Precision measures how many of the predicted instances of a class were correct, while recall measures how many of the true instances of that class were successfully identified. The F1-score is the harmonic mean of precision and recall, defined as

$$F1 = 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}$$

Unlike accuracy, which may be inflated by dominant classes, the F1-score balances the trade-off between false positives and false negatives, providing a more reliable measure of performance when classes are imbalanced. These metrics provide a more granular view of how well the models distinguish between providers, especially less frequent ones such as Google cloud. For summary reporting across classes, macro-averaged F1 was adopted. Unlike weighted F1, which tends to mirror accuracy by emphasizing the majority class, macro F1 treats each class equally, ensuring that rare providers influence the score just as much as common ones (Sokolova and Lapalme, 2009). This choice reflects the study’s emphasis on capturing performance across all providers rather than optimising for the dominant category. In practical terms, macro-averaging rewards models that perform consistently across both common and rare classes, which enables us to detect hosting patterns beyond the largest providers.

Additional diagnostics were used to supplement these core metrics. Confusion matrices were inspected to reveal systematic patterns of misclassification particularly in cases where cloud providers are modelled as separate classes rather than collapsed into the residual “Other” (Stehman, 1997). Feature importance scores were also extracted, albeit through different mechanisms depending on the model. Logistic regression coefficients provide insight into which signals are most strongly associated with each provider (Bishop, 2006, ch. 4), while tree-based models allow the computation of feature importance through split-based or permutation-based measures (Breiman et al., 1984; Breiman, 2001). Together, these outputs support not only evaluation but also interpretability, clarifying which features carry the greatest weight in classification. Finally, cross-fold variance was recorded alongside average scores, providing an indication of how stable the models were across different splits of the data. Low variance suggests robustness, whereas high variance would signal sensitivity to the specific training and validation partitions, which is critical to assess given the small dataset size.

Several additional metrics were considered but excluded as inappropriate for this task. Receiver Operating Characteristic (ROC) and Precision-Recall Area Under the Curve (AUC) metrics are well suited to binary problems, but their multiclass generalisations require one-vs-rest aggregation, which complicates interpretation (Hand and Till, 2001). While such metrics may be beneficial in cases where the classification task is collapsed into a binary distinction between cloud hyperscalers and all other providers, they were left out to ensure that metrics remained directly comparable across experimental settings. Calibration curves, which test whether predicted probabilities correspond to true frequencies (Niculescu-Mizil and Caruana, 2005), were also deemed unnecessary, since the focus of this study lies in provider assignment rather than the calibration of predicted probabilities. More advanced interpretability techniques such as SHAP or LIME (Lundberg and Lee, 2017; Ribeiro et al., 2016) could in principle be applied, but these methods are most useful for explaining opaque models such as neural networks. Since the models chosen already provide direct or indirect measures of feature contribution, the added complexity of such tools was not justified. The considered evaluation metrics are summarised in Table 5.

Table 5: Considered evaluation metrics

<b>Metric</b>	<b>Included?</b>	<b>Justification</b>
Accuracy	yes	Simple, intuitive
Precision / Recall	yes	Capture class-specific correctness and completeness
F1-score (macro-avg)	yes	Balances precision and recall, weighs rare classes equally
Confusion matrix	yes	Reveals systematic misclassifications
Feature importance	yes	Identifies which signals contribute most
Cross-fold variance	yes	Assesses stability of model performance across folds
ROC / AUC	no	Suited to binary tasks
Calibration curves	no	Probability calibration is not the focus
SHAP / LIME	no	More useful for opaque models

Together, these metrics and diagnostics were chosen to balance simplicity, fairness across classes, and interpretability. They provide the basis for the comparisons reported in Chapter 4, where results are reported in terms of overall accuracy, class-specific performance, and the infrastructural features most influential in classification.

### 3.4 Limitations

In addition to the limitations noted in previous sections, several methodological constraints of this study deserve explicit recognition.

First, the collection of ground truth data through Freedom of Information (FOI) requests imposed structural constraints on the dataset. FOI processes are inherently labour-intensive and thus limit the attainable sample size. Moreover, organisations and countries differ in how they interpret their disclosure obligations under FOI law, leading to variation in the level of detail or clarity of responses. In some cases, organisations reported only umbrella organisations or intermediaries rather than hosting providers or gave vague answers such as “hybrid” or “unknown”, which had to be excluded prior to the modelling. These differences affect the comparability of the ground truth across countries. A further constraint arises from temporality, as FOI responses represent a snapshot in time, even though hosting arrangements are dynamic and subject to change.

Second, the transformation of raw observational signals into features for model training required simplification that may have obscured nuance. Collapsing multiple NS or MX records to a single provider ensured that the dataset remained tractable for modelling, but at the cost of discarding information about redundancy or multi-provider setups. Canonicalization was deliberately limited to avoid over-normalising distinct entities, yet even minimal adjustments, such as consolidating variants of “DigiCert Inc.”, risk masking subtle distinctions. The modelling pipeline also assumed that each domain corresponded to a single hosting provider label, whereas in practice some services may be split across multiple infrastructures. Finally, one-hot encoding of categorical features introduced sparsity, which can challenge certain algorithms and inflate the dimensionality of the feature space relative to the small sample size.

Third, while stratified k-fold cross-validation was chosen to provide robust estimates, it remains sensitive to small sample sizes and class imbalance. Some rare providers may only appear a handful of times across folds, which introduces instability in per-class metrics. Furthermore, the absence of an entirely external test set means that model evaluation rests entirely on resampling strategies, limiting claims about generalisation beyond the study’s sample. Creating a dedicated hold-out validation set would have further reduced the already limited training data, hence why resampling was preferred despite its constraints.

Lastly, the scope of the dataset reflects both geographic and conceptual choices. Only three countries were included, constraining the diversity of institutional and infrastructural arrangements represented. The observational features were restricted to publicly visible DNS, MX, TLS, and ASN signals, leaving aside other potentially informative but less accessible traces. Finally, the label ontologies emphasised dependence on the largest cloud providers, in line with the study’s motivation of assessing risks of concentrated reliance, but at the expense of granularity regarding alternative providers, which may still include significant foreign entities but not concentrated at the same scale as hyperscalers.

These limitations do not invalidate the study but highlight the conditions under which findings should be interpreted. They also point to avenues for future research, such as scaling data collection to additional jurisdictions, incorporating longitudinal measurements, or experimenting with richer feature representations.

## 4 Results

This chapter presents the empirical results of the study. It begins with results from the Freedom of Information (FOI) requests, which provide the verified ground-truth dataset of hosting providers used to evaluate subsequent methods. These results summarise the distribution of providers across three countries studied and reveal the extent of reliance on external providers. The chapter then turns to the predictive modelling results, evaluating how accurately publicly observable records can identify hosting providers. Model performance is reported overall and per class, followed by confusion matrices, feature importance analyses, and an assessment of stability across validation folds.

### 4.1 Hosting provider ground truth

The ground-truth dataset used in this study combines disclosures from multiple sources. The majority of entries come from FOI requests submitted in the United Kingdom, Finland, and the Philippines, which provided verified information about the hosting of government websites. Since no examples of Chinese hyperscalers appeared in those responses, additional confirmed cases of Alibaba Cloud, Huawei Cloud, and Tencent Cloud were gathered from procurement record, official announcements, and provider case studies. Together these form a single dataset of government websites with known hosting providers, which serves as the foundation for the analyses presented in the remainder of this chapter.

#### 4.1.1 Hosting distribution in FOI responses

After filtering as described in Section 3.2.2, the following results summarise the providers identified in the FOI responses.

In the United Kingdom, hyperscalers appeared prominently. A majority of complete responses named one of three major US providers, with AWS accounting for around 59.0%, Microsoft Azure for 18.0%, and Google Cloud for 0 cases. Alongside these, there were several smaller commercial providers, such as DXW and Civic Computing. Self-hosted services were notably at 0% alongside Google Cloud.

In Finland, hyperscalers appeared less prominently, although they still held a majority share of the market. AWS accounted for 44.5% of complete responses, Azure for 18.2%, and Google Cloud for 1.5%. As with the UK, several smaller commercial providers appeared as well, such as Seravo and Twoday. Self-hosting was more common than in the UK at 5.8%. In some cases, government-owned corporations operated their own datacentres, which blurred the boundary between commercial and self-hosted provision.

These were categorized under “Other”, as exact operations were not always clear, and they were not self-hosted by the agencies in question.

In the Philippines, the hyperscalers were present, but not dominant. AWS accounted for 21.2%, Azure for 9.6%, and Google Cloud for 1.9%. Self-hosting remained more popular than the large hyperscalers combined at 61.5%. This indicates a markedly different hosting landscape compared to the UK and Finland, with self-hosting playing a much more prominent role. The distributions across hyperscalers and hosting types are summarised in Tables 6-7.

Table 6: Services hosted by hyperscalers, per country

Country	Total (n)	AWS	Azure	Google Cloud
UK	61	36 (59.0%)	11 (18.0%)	0 (0%)
Finland	139	61 (44.5%)	25 (18.2%)	2 (1.5%)
Philippines	52	11 (21.2%)	5 (9.6%)	1 (1.9%)

Table 7: Types of hosting providers per country

Country	Total (n)	Hyperscalers	Self-hosted	Other
UK	61	47 (77.0%)	0 (0%)	14 (23.0%)
Finland	139	88 (64.2%)	8 (5.8%)	41 (29.9%)
Philippines	52	17 (32.7%)	32 (61.5%)	3 (5.8%)

The results highlight distinct national patterns. In the United Kingdom, reliance on hyperscalers was particularly pronounced, with AWS and Azure together accounting for over 75% of the measured domains. Finland showed a similar albeit not as strong of a reliance on AWS and Azure. The Philippines presented a more balanced picture, with hyperscalers not being as dominant as self-hosting and other non-hyperscaler hosting solutions. Notably, Chinese hyperscalers were not present in any of the countries, a gap addressed in Section 4.1.3.

#### 4.1.2 Hosting distribution in supplementary Chinese hyperscaler dataset

The supplementary dataset comprises 19 domains, which are displayed in Table 8.

Table 8: Services hosted by Chinese hyperscalers

Hosting provider	Count
Alibaba Cloud	8
Huawei Cloud	7
Tencent Cloud	4

These figures should not be read as indicative of broader market trends. The dataset is too small and uneven to allow statistical generalisation, and its scope diverges from the stricter focus of the FOI requests. Its value lies instead in extending the FOI dataset with concrete training examples that broaden the coverage of the combined dataset. By including these cases, the subsequent modelling stage can account for Chinese hyperscalers as distinct classes, even if their prevalence in the sample is limited.

## 4.2 Predictive modelling results

This section reports the performance of the predictive models trained on the combined dataset. The aim is not only to measure overall accuracy, but also to examine how performance varies depending on how hosting providers are grouped into classes. Results are presented in terms of accuracy, macro-F1, and confusion patterns. We first establish baseline performance, then compare models across the four label schemes, and finally examine the most complete scheme in greater depth.

### 4.2.1 Overall model performance

All models substantially outperformed the majority-class baselines across the four label ontologies, confirming that the approach captures meaningful variation in observable hosting records. Accuracy is the most intuitive starting point for analysis. In every case, models improved between 15-40 percentage points compared to always predicting the largest class. At the same time, macro-F1 reveals sharp contrasts between ontologies, as it penalises the models for ignoring rare providers. Together, these two metrics provide complementary perspectives: accuracy establishes that the classifiers are useful, while macro-f1 tests whether they perform fairly across classes and thus remain extendable.  $\Delta$ Accuracy and  $\Delta$ Macro-F1 indicate the performance gain compared to the majority-class baseline, averaged across validation folds.

Ontology 1 (US and Chinese hyperscalers separate) presented the hardest challenge (Table 9). Expanding to seven classes sharply increased imbalance: AWS retained 40% of cases while Alibaba, Tencent, and Huawei contributed fewer than ten each. Accuracy remained acceptable at 0.74-0.76 for the best performing models, but macro-F1 collapsed to 0.37-0.43, which reveals the models' subpar performance. Notably, balanced Logistic Regression traded lower accuracy for the highest macro-F1, reflecting its stronger weighting of minority classes. This ontology shows the limits of multi-class attribution with a small dataset, where rare categories overwhelm the models' capacity to generalise.

Table 9: Ontology 1 results

<b>Model</b>	<b>Accuracy</b>	<b><math>\Delta</math>Accuracy</b>	<b>Macro-F1</b>	<b><math>\Delta</math>Macro-F1</b>
Majority baseline	0.401 $\pm$ 0.009	+0.000	0.090 $\pm$ 0.009	+0.000
LogReg	0.755 $\pm$ 0.027	+0.353	0.379 $\pm$ 0.027	+0.289
LogReg (balanced)	0.725 $\pm$ 0.042	+0.323	0.425 $\pm$ 0.066	+0.335
DecisionTree (d=6, leaf=3)	0.751 $\pm$ 0.041	+0.349	0.374 $\pm$ 0.022	+0.284
DecisionTree (d=10, leaf=2)	0.740 $\pm$ 0.058	+0.338	0.371 $\pm$ 0.015	+0.281
RandomForest	0.747 $\pm$ 0.050	+0.346	0.375 $\pm$ 0.034	+0.285
XGBoost	0.725 $\pm$ 0.052	+0.323	0.362 $\pm$ 0.033	+0.272

Ontology 2 (US and Chinese hyperscalers grouped) offered a middle ground (Table 10). By collapsing both the three US providers and the three Chinese providers into singular classes, it reduced the task to three labels while retaining the analytical distinction between US and Chinese hyperscalers. Here, balanced Logistic Regression was the most consistent performer, combining high accuracy with strong class balance. These results suggest that this ontology best captures the study’s aims, as it distinguishes Chinese providers as a class without reducing them to noise, while remaining tractable for models.

Table 10: Ontology 2 results

<b>Model</b>	<b>Accuracy</b>	<b><math>\Delta</math>Accuracy</b>	<b>Macro-F1</b>	<b><math>\Delta</math>Macro-F1</b>
Majority baseline	0.565 $\pm$ 0.009	+0.000	0.241 $\pm$ 0.003	+0.000
LogReg	0.803 $\pm$ 0.028	+0.238	0.758 $\pm$ 0.052	+0.517
LogReg (balanced)	0.810 $\pm$ 0.047	+0.245	0.773 $\pm$ 0.083	+0.533
DecisionTree (d=6, leaf=3)	0.766 $\pm$ 0.039	+0.201	0.627 $\pm$ 0.048	+0.386
DecisionTree (d=10, leaf=2)	0.807 $\pm$ 0.035	+0.242	0.665 $\pm$ 0.058	+0.425
RandomForest	0.721 $\pm$ 0.027	+0.156	0.605 $\pm$ 0.077	+0.365
XGBoost	0.788 $\pm$ 0.063	+0.223	0.733 $\pm$ 0.060	+0.492

Ontology 3 (US hyperscalers separate) produced four classes (Table 11). Accuracy reached 0.82-0.83 for the best models, nearly doubling the 0.44 baseline. Macro-F1 scores, however, averaged lower at around 0.71, revealing that minority classes were not handled consistently.

Table 11: Ontology 3 results

<b>Model</b>	<b>Accuracy</b>	<b><math>\Delta</math>Accuracy</b>	<b>Macro-F1</b>	<b><math>\Delta</math>Macro-F1</b>
Majority baseline	0.435 $\pm$ 0.013	+0.000	0.172 $\pm$ 0.029	+0.000
LogReg	0.822 $\pm$ 0.040	+0.387	0.707 $\pm$ 0.133	+0.535
LogReg (balanced)	0.781 $\pm$ 0.049	+0.346	0.592 $\pm$ 0.045	+0.421
DecisionTree (d=6, leaf=3)	0.829 $\pm$ 0.035	+0.394	0.716 $\pm$ 0.134	+0.544
DecisionTree (d=10, leaf=2)	0.803 $\pm$ 0.016	+0.368	0.658 $\pm$ 0.097	+0.486
RandomForest	0.818 $\pm$ 0.035	+0.383	0.701 $\pm$ 0.130	+0.529
XGBoost	0.788 $\pm$ 0.062	+0.353	0.679 $\pm$ 0.141	+0.507

Ontology 4 (US hyperscalers grouped) simplified the task to two categories and delivered the strongest overall results (Table 12). Both accuracy and macro-F1 clustered around 0.80-0.83, demonstrating that models could not only beat the baseline by nearly 27 percentage points, but also classify both classes reliably. Here, the higher depth Decision Tree achieved the best scores in both accuracy and macro-F1, though multiple models were not far behind. This ontology reveals that collapsing providers can eliminate noise while preserving meaningful analytical distinctions, allowing for stable and interpretable results.

Table 12: Ontology 4 results

<b>Model</b>	<b>Accuracy</b>	<b><math>\Delta</math>Accuracy</b>	<b>Macro-F1</b>	<b><math>\Delta</math>Macro-F1</b>
Majority baseline	0.565 $\pm$ 0.009	+0.000	0.361 $\pm$ 0.004	+0.000
LogReg	0.818 $\pm$ 0.058	+0.253	0.815 $\pm$ 0.058	+0.454
LogReg (balanced)	0.829 $\pm$ 0.048	+0.264	0.827 $\pm$ 0.046	+0.466
DecisionTree (d=6, leaf=3)	0.803 $\pm$ 0.057	+0.238	0.800 $\pm$ 0.056	+0.439
DecisionTree (d=10, leaf=2)	0.836 $\pm$ 0.044	+0.271	0.834 $\pm$ 0.043	+0.473
RandomForest	0.795 $\pm$ 0.076	+0.230	0.793 $\pm$ 0.077	+0.432
XGBoost	0.832 $\pm$ 0.072	+0.267	0.831 $\pm$ 0.071	+0.470

Taken together, the findings show three consistent findings. First, all models beat trivial baselines by wide margins, confirming the predictive value of observable infrastructure records. Second, class definitions matter more than model choice. Grouping providers stabilises performance, while fine-grained distinctions quickly degrade macro-F1. Third, model performance reflects a trade-off between accuracy and balance. While some

models may at times edge it out in raw accuracy, balanced Logistic Regression is proved to be the most consistently balanced, especially in the analytically most important Ontology 2. Interestingly, the ensemble methods Random Forest and XGBoost did not deliver the gains that could have been expected of them, performing worse than simpler classifiers under the constraints of this dataset.

While these results establish the broad picture of overall accuracy and stability, they also average over substantial differences between classes. The next section examines confusion matrices and per-class metrics to show where models were most and least successful.

#### 4.2.2 Misclassification and per-class performance

While overall accuracy provides a useful benchmark, per-class metrics reveal how well the models handle both majority and minority providers. Both macro and weighted averages are reported, with the former reflecting performance across classes equally, while the latter accounts for their relative frequency. Under the three-class ontology (Ontology 2: US hyperscalers grouped, Chinese hyperscalers grouped, Other), Logistic Regression delivers the most balanced performance. In the unweighted variant, recall for the Chinese class reaches 0.526 (F1 = 0.69), while the US class achieves an F1 of 0.85 and the “Other” class 0.745. Introducing class balancing shifts emphasis towards the minority, lifting Chinese recall to 0.789. This improvement comes at the cost of precision, producing more false positives, but raising the macro-F1 to 0.772. The trade-off illustrates how reweighting improves minority coverage without substantially harming overall accuracy. Detailed per-class performance for Ontology 2 (grouped US and Chinese hyperscalers grouped) is shown in Tables 13-15.

Table 13: Logistic Regression metrics (Ontology 2)

	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
Big US cloud	0.826	0.875	0.85
Big Chinese cloud	1.0	0.526	0.69
Other	0.745	0.745	0.745
Macro avg	0.857	0.715	0.761
Weighted avg	0.809	0.803	0.8

Table 14: Balanced Logistic Regression metrics (Ontology 2)

	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
Big US cloud	0.885	0.809	0.845
Big Chinese cloud	0.6	0.789	0.682
Other	0.762	0.816	0.788
Macro avg	0.749	0.805	0.772
Weighted avg	0.82	0.81	0.813

Tree-based models, by contrast, struggle with the small Chinese hyperscaler classes. Decision trees frequently misclassify these domains into the US class, yielding recalls as low as 0.211. Random forests and XGBoost improve stability for the majority but rarely succeed in lifting the minority beyond chance level. Overall, the findings suggest that linear models, especially when balanced, remain better suited to handling skewed class distributions in this setting.

Table 15: Decision tree (d=6) metrics (Ontology 2)

	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
Big US cloud	0.85	0.822	0.836
Big Chinese cloud	1.0	0.211	0.348
Other	0.653	0.786	0.713
Macro avg	0.834	0.606	0.632
Weighted avg	0.789	0.766	0.757

When the label space is expanded to seven categories (Ontology 1), performance degrades sharply. AWS, Azure, and Other continue to be detected with reasonable accuracy, but all minority hyperscalers collapse to near-zero scores. In every model except Balanced Logistic Regression, Chinese providers and Google Cloud receive zero recall. This collapse reflects the scarcity of such cases in the dataset, where folds often lacked training examples for these providers. It also emphasizes the value of grouping, as fine-grained classes are too sparse to support systematic classification. With broader groupings, useful distinctions and reasonable accuracy can still be surfaced. Table 16 summarises the results for Ontology 1.

Table 16: Balanced Logistic Regression metrics (Ontology 1)

	<b>Precision</b>	<b>Recall</b>	<b>F1</b>	<b>n</b>
AWS	0.859	0.787	0.821	108
Azure	0.786	0.805	0.795	41
Google Cloud	0.0	0.0	0.0	3
Alibaba Cloud	0.1	0.125	0.111	8
Huawei Cloud	0.273	0.429	0.333	7
Tencent Cloud	0.25	0.25	0.25	4
Other	0.774	0.735	0.754	98
macro avg	0.434	0.447	0.438	269
weighted avg	0.76	0.725	0.741	269

Together these results show that observable infrastructure records do provide enough signal to separate US, Chinese, and non-hyperscaler hosting, but that success depends critically on how classes are defined. Balanced Logistic Regression stands out as the method most capable of capturing minority categories. The next section explores which features underpin these predictions.

### 4.2.3 Interpretability and feature importance

Beyond aggregate scores, it is useful to inspect which observable records the models relied most upon. For Balanced Logistic Regression under Ontology 2, the strongest predictors align well with expectations about provider infrastructure. Table 17 summarises the top positive signals for each class. It should be noted that signals associated with Chinese hyperscalers draw on very few confirmed cases, so their apparent importance reflects the specificity of these traces rather than robust statistical support.

Table 17: Strongest positive predictors per class (Ontology 2)

<b>Class</b>	<b>Strongest positive predictors</b>
Big US Cloud	Ambientia ASN, Microsoft ASN, Amazon ASN
Big Chinese Cloud	suremail.cn MX, alidns NS, Chinanet ASN
Other	google.com MX, datacenter.fi MX, Let's Encrypt TLS

These records highlight the mixture of global and local infrastructure records that distinguish providers. For US hyperscalers, autonomous system numbers (ASNs) directly associated with Amazon and Microsoft dominate, alongside smaller commercial intermediaries such as Ambientia. Chinese hyperscalers are recognised primarily through DNS and MX records tied to Chinese providers. The residual “Other” class often reflects domestic or regional providers, where local MX records on top of those of Google as well as the non-profit free certificate authority Let's Encrypt are most

informative, showing how services that don't rely on big cloud providers for hosting prefer alternative and accessible providers for their various technical layers.

To complement the regression view, the depth-6 decision tree trained under the same ontology was analysed. While its accuracy was lower than that of the Balanced Logistic Regression, it provides a valuable illustration of how classification proceeds hierarchically. The root node, for instance, splits first on a Google MX record, immediately distinguishing a large set of primarily non-hyperscaler cases. Following this, most early splits are formed by certain ASNs splitting US providers from the rest. Subsequent branches begin to use other types of technical records more as they begin to assign cases to the other categories. Unlike regression coefficients, which weight features globally, the tree shows how feature combinations, rather than individual features alone, guide classification. The full tree is presented in Figure 2 (Appendix).

Together, these interpretability exercises suggest that the models are not relying on spurious correlations but on meaningful infrastructural fingerprints. Both regression coefficients and tree structures confirm that the models draw on a combination of different types of technical records, each contributing predictive value in distinguishing providers. This supports the inclusion of diverse infrastructural signals in future modelling efforts.

## 5 Discussion

This chapter interprets the findings presented in Chapter 4 in light of the research questions, the broader literature, and methodological approach. The discussion proceeds in three stages. Section 5.1 considers what the FOI dataset reveals about reliance and interdependence in public-sector hosting. Section 5.2 then evaluates the predictive modelling results, assessing their methodological significance and what they suggest about the feasibility of moving from attribution to prediction. These reflections lead to Section 5.3, which considers the limitations of the study and outlines avenues for future development, before the concluding chapter summarises the thesis.

### 5.1 Patterns of infrastructural interdependence

The FOI dataset assembled for this study provides the first systematically validated views of how governments host their digital services. Previous research has typically relied on simple attribution methods for identifying the hosting providers, such as mapping Autonomous System Numbers, DNS delegations, or TLS certificates directly to hosting providers (Ager et al., 2011; Durumeric et al., 2013; Jansen et al., 2023; Kumar et al., 2024). While these approaches are straightforward and can be easily applied at scale, they rest on strong assumptions that risk conflating network operators, intermediaries, and service providers. Crucially, have generally not been validated against independent institutional data, leaving uncertainty about the accuracy of their attributions. FOI disclosures, by contrast, supply verified institutional statements about hosting arrangements, even if their coverage is uneven and subject to ambiguity. In this sense, the dataset serves a dual role. It is substantively informative in highlighting how reliance on providers varies across contexts, and methodologically significant as a baseline against which predictive models can later be tested.

The results from the United Kingdom and Finland indicate a strong concentration on two U.S. hyperscalers: Amazon Web Services (AWS) and Microsoft Azure. In the UK, AWS accounted for nearly 60 percent of complete responses, with Azure representing another 18 percent. Google Cloud did not appear at all, and self-hosting was likewise absent. Finland exhibited a similar distribution, with AWS hosting 44 percent of services and Azure 18 percent, while Google Cloud appeared only marginally. Taken together, these results point to a duopoly structure in which public-sector cloud reliance is effectively concentrated on two providers, with AWS as the clear leader. This finding aligns with prior studies documenting the predominance of U.S. firms in European government infrastructure (Ghezzi et al., 2022; Jansen et al., 2023). It also indicates the absence of strong European alternatives, underlining the asymmetry that scholars have

described in terms of “weaponised interdependence” (Farrell and Newman, 2019), where governments’ reliance on a small set of foreign providers creates vulnerabilities alongside operational benefits of scalability, cost efficiency, and resilience.

The Philippines presents a contrasting profile. Hyperscalers were present but accounted for only a quarter of the identified services, while nearly half of the dataset consisted of self-hosted arrangements. The persistence of self-hosting in the Philippines illustrates that cloud adoption is not universal and that alternative models of provision remain significant. Several factors may explain this divergence. Different modernization trajectories also matter, as governments with higher budgets may have been able to migrate rapidly to cloud infrastructures. Others may adopt incrementally or maintain legacy hosting. Geopolitical context may also play a role: although Chinese providers have a presence in the Philippines (Lehdonvirta et al., 2025), the country’s security alliance with the United States (Castro et al., 2019) may reduce the likelihood of reliance on Chinese firms. In this sense, hosting choices appear to mirror broader international alignments. Crucially, the Philippine case demonstrates that infrastructural self-reliance can persist alongside cloud adoption, offering governments reduced exposure to external dependencies but also raising potential trade-offs in terms of scalability and efficiency.

Equally notable across all three countries is the complete absence of Chinese hyperscalers from the FOI responses. Providers such as Alibaba Cloud and Huawei Cloud have grown substantially in global market share, particularly across Asia (Yang and Li, 2025). Yet none of the verified disclosures in the United Kingdom, Finland, or the Philippines indicated reliance on them. This absence is unlikely to be explained by incomplete reporting or limited sample alone. Rather, it suggests that public-sector adoption of Chinese hyperscalers remains limited outside their domestic markets. This may reflect geopolitical caution, with governments avoiding Chinese platforms for sovereignty or security reasons, or it may reflect market trajectories in which Chinese providers consolidate regionally before expanding into the public sector abroad, even as their global market share grows in other domains. Regardless of the reason, their absence is itself a meaningful finding, which shows the asymmetry of global hosting reliance, in which U.S. hyperscalers dominate internationally (Liukkonen, 2025). Due to their absence, supplementary data was required in the modelling stage.

Taken together, these findings reveal a spectrum of infrastructural interdependence across national contexts. The United Kingdom and Finland illustrate cases of heavy integration into global hyperscale ecosystems, trading sovereignty for the benefits associated with mature global cloud infrastructures. The Philippines illustrates that alternatives persist, with governments able to retain greater infrastructural self-reliance, whether by choice or constraint. No government achieves full autonomy, as earlier

studies have shown (Jansen et al., 2023), but the degree and nature of dependence vary. Sovereignty in this domain is therefore best understood as a continuum, with different governments positioned at different points depending on resources, policy priorities, and geopolitical alignments. Cloud adoption is not inevitable, and the persistence of self-hosting demonstrates that infrastructural arrangements continue to reflect national circumstances.

## **5.2 Predictive feasibility**

The predictive modelling results demonstrate the potential of shifting from rule-based attribution to systematic classification. Whereas Section 5.1 examined the hosting patterns revealed by FOI disclosures, the focus here is on how far observable technical records can reproduce those verified labels. Rather than assuming a direct link from one record to a provider (Ager et al., 2011; Durumeric et al., 2013), this study combines multiple records (ASN, NS, MX, TLS) within predictive models, trained and validated against FOI ground truth, to assess their reliability.

The results show that predictive models consistently outperformed trivial baselines across all label groupings. This confirms that observable records do contain meaningful information about hosting providers, and that systematic modelling can extract that information in a reliable way. Model comparisons indicate that simple classifiers such as logistic regression were sufficient to achieve strong results, balancing accuracy with interpretability. More complex ensembles such as random forests and gradient-boosted trees did not deliver the expected improvements, likely because the small, sparse dataset limited the advantages of ensemble methods. In this respect, the choice of how providers were grouped into categories mattered more than the choice of algorithm. Broad groupings, such as distinguishing U.S. hyperscalers, Chinese hyperscalers, and “Other,” produced stable results with balanced performance across classes. The configuration that distinguished U.S. hyperscalers, Chinese hyperscalers, and “Other” proved the most informative, balancing analytical relevance with stability. Fine-grained categories, by contrast, revealed the limits of the dataset, as minority providers collapsed to near-zero recall even when balanced training was applied.

These findings highlight a broader methodological implication. The success of prediction rests not only on the models themselves but on how the classification task is defined. Aggregated labels provide tractable distinctions that align with the thesis’ motivating questions of sovereignty and interdependence, whereas disaggregated classes push beyond what the available data can support. In other words, predictive accuracy is achievable, but only when the label space is carefully designed to reflect both analytical priorities and the constraints imposed by FOI-based ground truth.

Interpretability analysis reinforces this conclusion. Logistic regression coefficients and decision-tree splits revealed that providers were identified not by a single decisive feature, but by combinations of records spanning different infrastructural layers. For instance, U.S. hyperscalers were strongly associated with Amazon and Microsoft ASNs, while Chinese hyperscalers were distinguished through DNS and MX providers tied to Chinese networks. The residual “Other” class drew on a mix of local MX providers, regional ASNs, and widely used TLS issuers such as Let’s Encrypt. The variety of predictive features across categories shows that no single technical record is sufficient for reliable attribution. Instead, it is the layered combination of infrastructural traces that enables robust classification. This represents an advance over earlier approaches that treated individual records as definitive.

At the same time, the models’ misclassification patterns underline their dependence on validated training data. The difficulty of reliably detecting Chinese providers reflects not a lack of discriminative records but the scarcity of confirmed cases in the dataset. This illustrates a key methodological constraint, as predictive models can only generalize to providers that are adequately represented in training data. Ground-truth labels are therefore indispensable, not just for evaluation but for enabling prediction in the first place.

Taken together, these results suggest that machine-learning-based prediction offers a promising complement to traditional attribution methods. By integrating multiple technical records, the models capture information that single-record heuristics overlook. Yet their utility is conditional. They require validated ground truth for training, sufficient representation of providers to ensure stability, and carefully designed categories that balance granularity with statistical robustness. Under these conditions, predictive modelling can provide a useful systematic and scalable way of mapping public-sector hosting dependence.

### **5.3 Limitations and future developments**

The findings presented in this thesis should be interpreted with several constraints in mind. These relate to the construction of the dataset, the modelling pipeline, the definitions adopted for key concepts, and the scope within which results can be generalised.

#### **5.3.1 Data-related constraints**

A central limitation of this study lies in the way ground-truth data was collected. Freedom of Information (FOI) requests provided verified labels that were indispensable for training and validating the predictive models, but the process was constrained by uneven response rates, differences in disclosure practices, and occasional refusals or non-responses. For example,

while security-related agencies in both the United Kingdom and Finland withheld information, some UK agencies in different areas also withheld information for security reasons. As a result, the dataset offers a valuable snapshot but not a complete census of government hosting.

Ambiguity in responses further limited reliability. Some organizations conflated hosting with system administration or referred only to umbrella organisations such as Valtori, without clarifying the underlying infrastructure. When left unclear, these cases had to be excluded, which reduced coverage and highlighted how institutional arrangements can obscure technical reality. Exclusions were necessary for modelling, but they inevitably narrowed the dataset and may have removed edge cases that illustrate the diversity of public-sector hosting arrangements. Further issues arose when some agencies proved difficult to work with, such as the Finnish Valtori, which failed to respond even when other agencies instructed to redirect requests to them. This was not the case when some agencies requested the information from Valtori themselves.

FOI should also be understood as only one form of infrastructural transparency. It provides authoritative answers, but is time-consuming to obtain, fragmented across jurisdictions, and dependent on how public bodies interpret disclosure obligations. Moreover, not all countries maintain FOI legislation, which limits the universal applicability of this approach. Procurement records represent an alternative source, with the advantage of being structured and covering financial flows, but they too face lags, gaps, and opacity in subcontracting. This study's reliance on FOI means that its ground truth is robust but narrow, raising questions about how future research could combine multiple institutional sources to produce broader and timelier coverage.

The supplementary dataset of Chinese hyperscalers illustrates a further constraint. Because FOI responses in the three main countries yielded no examples of Alibaba, Huawei, or Tencent, additional confirmed cases had to be drawn from procurement notices and provider case studies. While this filled an important gap, the dataset was limited in size, less authoritative, and unstable under cross-validation. As such, results for Chinese providers should be interpreted cautiously, as the predictive models were trained on far weaker foundations for these cases than for U.S. or domestic providers.

### **5.3.2 Modelling constraints**

The predictive modelling stage faced significant constraints arising from dataset size and imbalance. With 269 domains in total, and some classes represented by only a handful of examples, the models struggled to generalise beyond the most common providers. In particular, rare providers such as Google Cloud or Tencent Cloud frequently collapsed to near-zero recall, even under balanced training. This instability reflects the structural challenges of

working with small samples in a multi-class classification task, and it shows how rare but politically salient providers can be the least reliably captured.

Feature engineering choices also imposed simplification. The study relied on four categories of observable records: authoritative name servers, mail exchangers, autonomous system numbers, and TLS certificate issuers. While these are meaningful infrastructural traces, they do not capture the full range of possible indicators. Other layers, such as HTTP headers or TXT records, were excluded either for practical reasons or because they are harder to collect consistently. Even within the selected features, collapsing multiple NS or MX records into a single provider, and using one-hot encoding to represent them, meant that information about redundancy or potential hybrid arrangements could have been lost. These design choices made the dataset tractable but at the cost of simplification.

Model selection reflected a further trade-off. Logistic regression, decision trees, random forests, and gradient-boosted trees were chosen to balance interpretability with predictive power. In practice, simple models performed as well as or better than more complex ensembles, likely because the dataset’s small size and sparsity limited the benefit of ensemble methods. More advanced methods such as neural networks or SVMs were deliberately excluded as inappropriate, but their absence also narrows the scope of conclusions. The study therefore demonstrates what can be achieved with transparent, conventional classifiers, but does not test whether more sophisticated models could overcome some of the limitations.

A further constraint lies how model performance was evaluated. Overall accuracy is dominated by the largest providers, so rare classes have virtually no impact on the score even if they are consistently misclassified. Macro-averaged F1, by contrast, weights all classes equally, which means that very small categories, such as Google Cloud, can disproportionately depress the overall results if they are not recognised. Neither measure fully captures the realities of the task. Accuracy overlooks rare providers, while macro-f1 arguably exaggerates their significance relative to their prevalence. This trade-off reflects a broader challenge in evaluating models under class imbalance, where conventional metrics often fail to capture minority-class performance (He and Garcia, 2009). In principle, an intermediate metric might provide a more proportionate balance, but within this study the choice was limited to reporting both perspectives side by side. In future research, the use of alternative evaluation metrics such as balanced accuracy (Brodersen et al., 2010) or micro-averaged recall (Sokolova and Lapalme, 2009) could provide a middle ground, better reflecting performance across both dominant and minority providers. More generally, refining evaluation strategies to account for class imbalance would improve the robustness of predictive claims in this domain. At the same time, the use of macro-F1 was well suited for this study’s aims, since its equal weighting across classes

emphasized whether smaller, politically salient providers could be recognised alongside the dominant hyperscalers.

### **5.3.3 Conceptual and definitional limits**

Another set of limitations arises from the conceptual choices made in defining the unit of analysis. The study treated each domain equally, whether it was a central government portal like “gov.uk” or the website of a small agency. This decision ensured compatibility for modelling but flattened the real-world importance of services, since some domains are vastly more significant for citizens and infrastructure. Weighting domains by traffic, function, or institutional role could produce a different picture of dependence and interdependence, though such features were out of scope for this thesis.

The definition of “hosting provider” also introduced conceptual complexity. As discussed in Chapter 2, modern infrastructures distribute responsibility across multiple layers of storage, computation, networking, and delivery. FOI responses sometimes reflected this hybridity, with organizations naming integrators or umbrella organisations rather than the entities operating the underlying servers. For modelling purposes, this study adopted an operational definition centred on the origin host, meaning the organisation responsible for running the servers that deliver the web application’s content. While this abstraction was necessary for tractability and comparability across countries, it inevitably simplifies the layered reality of contemporary hosting. Conceptually, this reflects a tension between the practical need to assign a single provider label and the theoretical difficulty of delineating responsibility in distributed infrastructures. Future work may benefit from revisiting this assumption, potentially moving beyond the traditional concept of a hosting provider toward frameworks that better capture multi-provider and hybrid arrangements.

Relatedly, the grouping of providers into categories shaped outcomes. Distinguishing AWS and Azure separately, for example, revealed their individual dominance, but grouping them together provided clearer and more stable predictive results. Similarly, grouping Chinese hyperscalers into a single category allowed meaningful evaluation, whereas separating them collapsed performance. These choices reflect a balance between granularity and statistical robustness, but they also mean that findings are contingent on how providers were categorised. Alternative groupings, such as separating regional European providers or distinguishing state-owned firms, might yield different insights.

Finally, the scope of what counts as the public sector could be contested. Many governments deliver services through state-owned enterprises, public corporations, or municipal governments, which were excluded from the FOI requests for consistency. Their inclusion might have expanded coverage and

revealed different dependency patterns, particularly where such entities manage critical infrastructure. In some jurisdictions, these entities may also be subject to FOI legislation, meaning that they could have been incorporated into the dataset. Future research could experiment with alternative definitions of the public sector, for instance training models on a broader set of domains while applying stricter criteria at the interpretation stage, testing how boundary choices influence the conclusions.

#### **5.3.4 Generalisability**

Beyond the modelling and conceptual constraints discussed above, the generalisability of the study is also limited. Only three countries were included, and while they were chosen for their diversity, they cannot represent the full range of global hosting practices. Extending the analysis to other contexts, such as larger economies and developing states, would likely reveal additional forms of reliance and interdependence. Similarly, the absence of Chinese providers from FOI data in the countries chosen for analysis does not imply that Chinese providers are absent from government infrastructure in other countries, particularly given their growing international market share (Yang and Li, 2025).

A further, core limitation is temporality. Hosting arrangements are dynamic, shifting with new contracts, migrations, and policy changes. FOI responses capture a snapshot that may quickly become outdated, and predictive models trained on such data risk lagging behind reality. Ideally, features for model training should be collected as close as possible in time to FOI responses, as previous work on internet infrastructure has also emphasized the volatility of such measurements (Durumeric et al., 2015).

Finally, the study focused exclusively on the public sector, which carries both advantages and limitations. Public-sector services are particularly important because of their political and sovereignty implications, but they are not solely representative of the broader Internet. Hosting in the private sector may follow different dynamics, shaped by market competition, profitability, and different transparency obligations (Varghese and Buyya, 2018). The models developed here may not transfer directly to those settings without adaptation, as hosting arrangements and market shares in the private sector may substantially differentiate from those in government contexts, meaning that predictive relationships learned here may not hold outside the public sector context.

#### **5.3.5 Future research**

While these limitations frame the boundaries of the present study, they also point to several directions for future work. The most immediate step is to expand the dataset. Collecting larger and more diverse samples across

additional jurisdictions would stabilise minority classes and improve generalisability. Such expansion would also enable systematic cross-national comparison, revealing how hosting dependence varies under contrasting institutional, legal, and geopolitical conditions. Repeating data collection over time, with technical records gathered simultaneously, would further support longitudinal analysis of hosting trends. In addition, combining FOI responses with procurement records or other institutional disclosures could broaden coverage and reduce reliance on a single, labour-intensive model.

Feature expansion is another promising avenue. The present study was restricted to four categories of publicly observable records. Additional infrastructural traces such as HTTP headers, CDN configurations, or reverse DNS could provide further discriminative power, particularly if combined with existing features in a multi-modal framework. More sophisticated feature representations, such as learned embeddings of provider names or network paths, may also reduce sparsity while retaining interpretability.

From a modelling perspective, alternative evaluation strategies could better capture performance under class imbalance. Metrics such as balanced accuracy (Brodersen et al., 2010), weighted F1 with adjusted class weights, or Bayesian uncertainty estimates (Gal and Ghahramani, 2016) could provide a more nuanced picture of how well models generalise to minority providers. At the same time, experiments with semi-supervised or weakly supervised approaches may allow models to leverage larger volumes of unlabelled domains while anchoring predictions in a smaller set of validated ground-truth cases.

Finally, future work should consider how predictive methods interact with broader questions of infrastructural transparency. Prediction cannot substitute disclosure, but it can complement existing mechanisms of accountability by providing systematic, replicable insights into hosting. Exploring how these methods might be embedded into monitoring frameworks represents an important step for both research and policy.

## 6 Conclusions

This thesis set out to answer the question:

**RQ:** Can the underlying hosting provider of a public-sector digital service be predicted using publicly observable information?

The motivation lay in growing concerns about the opacity of government reliance on a handful of global cloud providers and the implications this has for sovereignty, accountability, and infrastructural resilience. Hosting is a domain where dependence often remains hidden, and where conventional attribution methods are prone to both overreach and error. By combining institutional disclosures obtained through Freedom of Information (FOI) requests with predictive modelling on observable technical records, this study has sought to both document patterns of reliance and evaluate the feasibility of systematic prediction.

The findings can be summarized in two parts. First, the FOI dataset confirmed strong concentration on U.S. hyperscalers in the United Kingdom and Finland, with Amazon Web Services and Microsoft Azure dominating over European alternatives. The Philippines, by contrast, exhibited a more mixed landscape in which self-hosting remained prominent, underscoring that cloud adoption is not universal and that infrastructural self-reliance can persist. Across all three countries, the complete absence of Chinese hyperscalers revealed a notable asymmetry in global hosting markets, despite their growing international share elsewhere. Second, the predictive modelling results demonstrated that hosting providers can be inferred from observable records with substantially higher accuracy than trivial baselines. Logistic regression consistently outperformed expectations, balancing accuracy with interpretability, while more complex ensemble methods underperformed, likely due to the small and sparse dataset. Most importantly, the analysis showed that reliable classification does not rest on a single decisive record but emerges from the layered combination of infrastructural traces such as ASNs, DNS, MX, and TLS issuers.

Together, these results contribute to the study of government digital infrastructures in three respects. Empirically, the thesis provides one of the few systematically validated datasets on public-sector hosting, revealing both concentration and variation in patterns of reliance. Methodologically, it represents a structured test of predictive classification against verified ground truth, showing both the promise and the constraints of such approaches. Conceptually, it frames sovereignty not as a binary of autonomy or dependence but as a continuum shaped by resources, alignments, and infrastructural choices. Although institutional transparency provides the strongest form of verification, predictive modelling offers a practical

substitute for large-scale data collection in contexts where disclosure is unavailable or prohibitively difficult to obtain through direct requests.

The study also carries important limitations. FOI-based ground truth remains partial and uneven, and such requests are not universally available across jurisdictions. The dataset was small, with significant class imbalance that limited stability for minority providers, and the conceptual choice to reduce each domain to a single provider label inevitably obscured complex arrangements. These constraints underline that the findings should be seen as exploratory rather than definitive.

Looking ahead, future research could address these limits by expanding datasets across more jurisdictions, collecting technical records and institutional disclosures in parallel, and incorporating richer features beyond those tested here. Alternative evaluation metrics may also offer more proportionate ways of capturing performance under imbalance. Beyond research, predictive methods could play a role in monitoring frameworks, providing governments, auditors, or civil society with tools to track infrastructural dependencies more systematically.

In direct response to the research question, this thesis shows that public-sector hosting providers can be inferred from publicly observable information with reasonable reliability. However, this is only possible under conditions of validated training data, carefully structured class definitions, and deliberate methodological design. Prediction alone cannot resolve questions of accountability or sovereignty, but combined with institutional transparency it offers a promising path toward more systematic visibility into the infrastructures that underpin digital government.

## References

- Aas, J., Barnes, R., Case, B., Durumeric, Z., Eckersley, P., Flores-López, A., Halderman, J.A., Hoffman-Andrews, J., Kasten, J., Rescorla, E., Schoen, S., Warren, B., 2019. Let's Encrypt: An Automated Certificate Authority to Encrypt the Entire Web, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19. Association for Computing Machinery, New York, NY, USA, pp. 2473–2487. <https://doi.org/10.1145/3319535.3363192>
- Act on the Openness of Government Activities (621/1999), 1999.
- Ager, B., Mühlbauer, W., Smaragdakis, G., Uhlig, S., 2011. Web content cartography, in: Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11. Association for Computing Machinery, New York, NY, USA, pp. 585–600. <https://doi.org/10.1145/2068816.2068870>
- Alshibly, H., Chiong, R., 2015. Customer empowerment: Does it influence electronic government success? A citizen-centric perspective. *Electron. Commer. Res. Appl.* 14, 393–404. <https://doi.org/10.1016/j.elerap.2015.05.003>
- Appelt, S., Galindo-Rueda, F., 2016. Measuring the Link between Public Procurement and Innovation. OECD Sci. Technol. Ind. Work. Pap., OECD Science, Technology and Industry Working Papers.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M., 2010. A view of cloud computing. *Commun ACM* 53, 50–58. <https://doi.org/10.1145/1721654.1721672>
- Autolitano, S., Pawlowska, A., 2021. Europe's Quest for Digital Sovereignty: GAIA-X as a Case Study (No. IAI Papers 21, 14). Istituto Affari Internazionali (IAI), Rome.
- Banisar, D., 2006. Freedom of Information Around the World 2006: A Global Survey of Access to Government Information Laws. <https://doi.org/10.2139/ssrn.1707336>
- Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., Warfield, A., 2003. Xen and the art of virtualization. *SIGOPS Oper Syst Rev* 37, 164–177. <https://doi.org/10.1145/1165389.945462>
- Beam, A.L., Manrai, A.K., Ghassemi, M., 2020. Challenges to the Reproducibility of Machine Learning Models in Health Care. *JAMA* 323, 305–306. <https://doi.org/10.1001/jama.2019.20866>
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer, New York.
- Blancato, F.G., 2024. The cloud sovereignty nexus: How the European Union seeks to reverse strategic dependencies in its digital ecosystem. *Policy Internet* 16, 12–32. <https://doi.org/10.1002/poi3.358>
- Bozeman, B., Bretschneider, S., 1994. The “Publicness Puzzle” in Organization Theory: A Test of Alternative Explanations of

- Differences between Public and Private Organizations. *J. Public Adm. Res. Theory J-PART 4*, 197–223.
- Braud, A., Fromentoux, G., Radier, B., Le Grand, O., 2021. The Road to European Digital Sovereignty with Gaia-X and IDSA. *IEEE Netw.* 35, 4–5. <https://doi.org/10.1109/MNET.2021.9387709>
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.
- Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M., 2010. The Balanced Accuracy and Its Posterior Distribution, in: 2010 20th International Conference on Pattern Recognition. Presented at the 2010 20th International Conference on Pattern Recognition, pp. 3121–3124. <https://doi.org/10.1109/ICPR.2010.764>
- Campbell-Kelly, M., Garcia-Swartz, D.D., 2013. The history of the internet: the missing narratives. *J. Inf. Technol. Suppl Spec. Issue Hist. IS* 28, 18–33. <https://doi.org/10.1057/jit.2013.4>
- Carter, L., Bélanger, F., 2005. The utilization of e-government services: citizen trust, innovation and acceptance factors. *Inf. Syst. J.* 15, 5–25. <https://doi.org/10.1111/j.1365-2575.2005.00183.x>
- Castro, P., Ishakian, V., Muthusamy, V., Slominski, A., 2019. The rise of serverless computing. *Commun ACM* 62, 44–54. <https://doi.org/10.1145/3368454>
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. Association for Computing Machinery, New York, NY, USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297. <https://doi.org/10.1007/BF00994018>
- Coursey, D., Norris, D.F., 2008. Models of E-Government: Are They Correct? An Empirical Assessment. *Public Adm. Rev.* 68, 523–536.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Definition of PUBLIC SECTOR [WWW Document], 2025. URL <https://www.merriam-webster.com/dictionary/public+sector> (accessed 7.18.25).
- Domingos, P., Pazzani, M., 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Mach. Learn.* 29, 103–130. <https://doi.org/10.1023/A:1007413511361>
- Dunleavy, P., Margetts, H., Bastow, S., Tinkler, J., 2006. *Digital Era Governance: IT Corporations, the State, and E-Government*. Oxford University Press, Incorporated, Oxford, UNITED KINGDOM.
- Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., Halderman, J.A., 2015. A Search Engine Backed by Internet-Wide Scanning, in: *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, CCS '15*. Association for Computing

- Machinery, New York, NY, USA, pp. 542–553.  
<https://doi.org/10.1145/2810103.2813703>
- Durumeric, Z., Kasten, J., Bailey, M., Halderman, J.A., 2013. Analysis of the HTTPS certificate ecosystem, in: Proceedings of the 2013 Conference on Internet Measurement Conference, IMC '13. Association for Computing Machinery, New York, NY, USA, pp. 291–304.  
<https://doi.org/10.1145/2504730.2504755>
- Eurostat., 2013. European system of accounts :ESA 2010. Publications Office, LU.
- Farrell, H., Newman, A.L., 2019. Weaponized Interdependence: How Global Economic Networks Shape State Coercion. *Int. Secur.* 44, 42–79.  
[https://doi.org/10.1162/isec\\_a\\_00351](https://doi.org/10.1162/isec_a_00351)
- Fielding, R.T., Reschke, J., 2014. Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing. Request for Comments.
- Flyverbom, M., Deibert, R., Matten, D., 2019. The Governance of Digital Technology, Big Data, and the Internet: New Roles and Responsibilities for Business. *Bus. Soc.* 58, 3–19.  
<https://doi.org/10.1177/0007650317727540>
- Freedom of Information Act 2000, 2000.
- Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* 29, 1189–1232.
- Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, in: Balcan, M.F., Weinberger, K.Q. (Eds.), Proceedings of The 33rd International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, New York, New York, USA, pp. 1050–1059.
- Ghezzi, R., Kolehmainen, T., Setälä, M., Mikkonen, T., 2023. Enterprise Architecture as an Enabler for a Government Business Ecosystem: Experiences from Finland.  
<https://doi.org/10.48550/arXiv.2309.08266>
- Ghezzi, R.-K., Korhonen, M., Vilpponen, H., Mikkonen, T., 2022. The Role of In-House Procurement According to Finnish Municipalities' Purchase Invoice Data. <https://doi.org/10.48550/arXiv.2211.14570>
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
- Gorriz, J.M., Clemente, R.M., Segovia, F., Ramirez, J., Ortiz, A., Suckling, J., 2024. Is K-fold cross validation the best model selection method for Machine Learning? <https://doi.org/10.48550/arXiv.2401.16407>
- Gruber, T.R., 1993. A translation approach to portable ontology specifications. *Knowl. Acquis.* 5, 199–220.  
<https://doi.org/10.1006/knac.1993.1008>
- Hand, D.J., Till, R.J., 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Mach. Learn.* 45, 171–186. <https://doi.org/10.1023/A:1010920819831>
- Hawkinson, J.A., Bates, T.J., 1996. Guidelines for creation, selection, and registration of an Autonomous System (AS). Request for Comments.

- Hazell, R., Worthy, B., 2010. Assessing the performance of freedom of information. *Gov. Inf. Q.*, Special Issue: Open/Transparent Government 27, 352–359. <https://doi.org/10.1016/j.giq.2010.03.005>
- He, H., Garcia, E.A., 2009. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* 21, 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Hood, C., 1991. A Public Management for All Seasons? *Public Adm.* 69, 3–19. <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>
- Houser, R., Hao, S., Cotton, C., Wang, H., 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). Presented at the 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 193–204. <https://doi.org/10.1109/DSN53405.2022.00030>
- Janowski, T., 2015. Digital government evolution: From transformation to contextualization. *Gov. Inf. Q.* 32, 221–236. <https://doi.org/10.1016/j.giq.2015.07.001>
- Jansen, B., Kadenko, N., Broeders, D., van Eeten, M., Borgolte, K., Fiebig, T., 2023. Pushing boundaries: An empirical view on the digital sovereignty of six governments in the midst of geopolitical tensions. *Gov. Inf. Q.* 40, 101862. <https://doi.org/10.1016/j.giq.2023.101862>
- (Jochen) Scholl, H.J., Klischewski, R., 2007. E-Government Integration and Interoperability: Framing the Research Agenda. *Int. J. Public Adm.* 30, 889–920. <https://doi.org/10.1080/01900690701402668>
- Jones, S., 2015. Cloud computing procurement and implementation: Lessons learnt from a United Kingdom case study. *Int. J. Inf. Manag.* 35, 712–716. <https://doi.org/10.1016/j.ijinfomgt.2015.07.007>
- Jonker, M., Akiwate, G., Affinito, A., Claffy, kc, Botta, A., Voelker, G.M., van Rijswijk-Deij, R., Savage, S., 2022. Where .ru? assessing the impact of conflict on russian domain infrastructure, in: Proceedings of the 22nd ACM Internet Measurement Conference, IMC '22. Association for Computing Machinery, New York, NY, USA, pp. 159–165. <https://doi.org/10.1145/3517745.3561423>
- Kilpi, J., Ramizo, G.Jr., Kässi, O., Lehdonvirta, V., 2025. Inferring hosting providers with predictive modelling [working title].
- Klensin, D.J.C., 2008. Simple Mail Transfer Protocol. Request for Comments.
- Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1137–1143.
- Kumar, R., Carisimo, E., Riva, L.D.A., Buzzzone, M., Bustamante, F.E., Qazi, I.A., Beiró, M.G., 2024. Of Choices and Control - A Comparative Analysis of Government Hosting, in: Proceedings of the 2024 ACM on Internet Measurement Conference, IMC '24. Association for

- Computing Machinery, New York, NY, USA, pp. 462–479. <https://doi.org/10.1145/3646547.3688447>
- Lane, J.-E., 2000. *The Public Sector: Concepts, Models and Approaches*. SAGE Publications, Limited, London, UNITED KINGDOM.
- Layne, K., Lee, J., 2001. Developing fully functional E-government: A four stage model. *Gov. Inf. Q.* 18, 122. [https://doi.org/10.1016/S0740-624X\(01\)00066-1](https://doi.org/10.1016/S0740-624X(01)00066-1)
- Lehdonvirta, V., Wú, B., Hawkins, Z., 2025. Weaponised interdependence in a bipolar world: how economic forces and security interests shape the global reach of US and Chinese cloud data centres. *Rev. Int. Polit. Econ.* 32, 1442–1467. <https://doi.org/10.1080/09692290.2025.2489077>
- Liu, C., Albitz, P., 2006. *DNS and BIND*, 5th ed. O'Reilly Media.
- Liu, E., Akiwate, G., Jonker, M., Mirian, A., Savage, S., Voelker, G.M., 2021. Who's got your mail? characterizing mail service provider usage, in: *Proceedings of the 21st ACM Internet Measurement Conference, IMC '21*. Association for Computing Machinery, New York, NY, USA, pp. 122–136. <https://doi.org/10.1145/3487552.3487820>
- Liukkonen, K., 2025. *Competing clouds: U.S.-China tech rivalry and the geoeconomic turn*. Aalto University, Espoo, Finland.
- Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Margetts, H., Dunleavy, P., 2013. The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Philos. Trans. Math. Phys. Eng. Sci.* 371, 1–17.
- Margetts, H., Naumann, A., 2017. Government as a platform: What can Estonia show the world. *Res. Pap. Univ. Oxf.* 1, 1–41.
- Mockapetris, P., 1987. RFC 1034: Domain names - concepts and facilities [WWW Document]. URL <https://www.rfc-editor.org/rfc/rfc1034> (accessed 8.18.25).
- Niculescu-Mizil, A., Caruana, R., 2005. Predicting good probabilities with supervised learning, in: *Proceedings of the 22nd International Conference on Machine Learning, ICML '05*. Association for Computing Machinery, New York, NY, USA, pp. 625–632. <https://doi.org/10.1145/1102351.1102430>
- OECD, 2025a. *Competition in the Provision of Cloud Computing Services (No. DAF/COMP(2025)8)*. OECD.
- OECD, 2025b. *Government at a Glance 2025: Drivers of Trust in Public Institutions*. OECD Publishing, Paris. <https://doi.org/10.1787/oefdbcd-en>
- OECD, 2025c. *Digital Transformation of Public Procurement: Good Practice Report (No. OECD Public Governance Policy Papers No. 77)*. OECD Publishing, Paris. <https://doi.org/10.1787/79651651-en>
- OECD, 2021. *The E-Leaders Handbook on the Governance of Digital Government* [WWW Document]. OECD. URL

- [https://www.oecd.org/en/publications/the-e-leaders-handbook-on-the-governance-of-digital-government\\_ac7f2531-en.html](https://www.oecd.org/en/publications/the-e-leaders-handbook-on-the-governance-of-digital-government_ac7f2531-en.html) (accessed 7.1.25).
- OECD, 2017. Government at a Glance 2017. OECD Publishing, Paris. [https://doi.org/10.1787/gov\\_glance-2017-en](https://doi.org/10.1787/gov_glance-2017-en)
- OECD, 2014. Recommendation of the Council on Digital Government Strategies.
- Oliver, R., 2024. Public sector or private sector: How the ONS decides – and why it matters [WWW Document]. Natl. Stat. URL <https://blog.ons.gov.uk/2024/08/20/public-sector-or-private-sector-how-the-ons-decides-and-why-it-matters/> (accessed 7.24.25).
- Pahl, C., 2015. Containerization and the PaaS Cloud. *IEEE Cloud Comput.* 2, 24–31. <https://doi.org/10.1109/MCC.2015.51>
- Paquette, S., Jaeger, P.T., Wilson, S.C., 2010. Identifying the security risks associated with governmental use of cloud computing. *Gov. Inf. Q.* 27, 245–253. <https://doi.org/10.1016/j.giq.2010.01.002>
- Pathan, A.-M.K., Buyya, R., 2007. A Taxonomy and Survey of Content Delivery Networks.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peng, Q., Zhang, M., Chang, D., Zhang, J., Liu, B., Duan, H., 2025. Decoding DNS Centralization: Measuring and Identifying NS Domains Across Hosting Providers, in: 2025 55th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Presented at the 2025 55th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 266–278. <https://doi.org/10.1109/DSN64029.2025.00037>
- Peterson, L.L., Davie, B.S., 2011. *Computer Networks: A Systems Approach*, 5th ed. Morgan Kaufmann, Burlington, MA.
- Philippines, O. of the P. of the, 2016. Executive Order No. 2, Series of 2016.
- Plantin, J.-C., Lagoze, C., Edwards, P.N., Sandvig, C., 2018. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media Soc.* 20, 293–310. <https://doi.org/10.1177/1461444816661553>
- Powers, D.M.W., 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. <https://doi.org/10.48550/arXiv.2010.16061>
- Prier, E., McCue, C.P., 2009. The Implications of a Muddled Definition of Public Procurement. *J. Public Procure.* 9, 326–370. <https://doi.org/10.1108/JOPP-09-03-04-2009-002>
- Public sector [WWW Document], n.d. . Stat. Finl. URL [https://stat.fi/meta/kas/julkinen\\_sektor\\_en.html](https://stat.fi/meta/kas/julkinen_sektor_en.html) (accessed 7.24.25).
- Quinlan, J.R., 1986. Induction of Decision Trees. *Mach. Learn.* 1, 81–106. <https://doi.org/10.1007/BF00116251>

- Rekhter, Y., Hares, S., Li, T., 2006. A Border Gateway Protocol 4 (BGP-4). Request for Comments.
- Rescorla, E., 2018. The Transport Layer Security (TLS) Protocol Version 1.3. <https://doi.org/10.17487/RFC8446>
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Model-Agnostic Interpretability of Machine Learning. <https://doi.org/10.48550/arXiv.1606.05386>
- Rojszczak, M., 2020. CLOUD act agreements from an EU perspective. *Comput. Law Secur. Rev.* 38, 105442. <https://doi.org/10.1016/j.clsr.2020.105442>
- Samuelson, P.A., 1954. The Pure Theory of Public Expenditure. *Rev. Econ. Stat.* 36, 387–389. <https://doi.org/10.2307/1925895>
- Sebastián, S., Diugan, R.-G., Caballero, J., Sanchez-Rola, I., Bilge, L., 2023. Domain and Website Attribution beyond WHOIS, in: Proceedings of the 39th Annual Computer Security Applications Conference, ACSAC '23. Association for Computing Machinery, New York, NY, USA, pp. 124–137. <https://doi.org/10.1145/3627106.3627190>
- Silva, J.M., Ribeiro, D., Ramos, L.F., Fonte, V., 2023. A worldwide overview on the information security posture of online public services. <https://doi.org/10.48550/arXiv.2310.01200>
- Singanamalla, S., Jang, E.H.B., Anderson, R., Kohno, T., Heimerl, K., 2020. Accept the Risk and Continue: Measuring the Long Tail of Government https Adoption, in: Proceedings of the ACM Internet Measurement Conference, IMC '20. Association for Computing Machinery, New York, NY, USA, pp. 577–597. <https://doi.org/10.1145/3419394.3423645>
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* 45, 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Sommese, R., Jonker, M., van der Ham, J., Moura, G.C.M., 2022. Assessing e-Government DNS Resilience, in: 2022 18th International Conference on Network and Service Management (CNSM). Presented at the 2022 18th International Conference on Network and Service Management (CNSM), pp. 118–126. <https://doi.org/10.23919/CNSM55787.2022.9965155>
- Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* 62, 77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- Tajalizadehkhoo, S., Korczyński, M., Noroozian, A., Gañán, C., van Eeten, M., 2016. Apples, oranges and hosting providers: Heterogeneity and security in the hosting market, in: NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium. Presented at the NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium, pp. 289–297. <https://doi.org/10.1109/NOMS.2016.7502824>
- Tarahomi, S., Sommese, R., de Boer, P.-T., Linssen, J., Holz, R., Sperotto, A., 2024. Is a Name Enough? A First Look into Detecting Clouds Using DNS Pointer Records, in: 2024 20th International Conference on

- Network and Service Management (CNSM). Presented at the 2024 20th International Conference on Network and Service Management (CNSM), pp. 1–5. <https://doi.org/10.23919/CNSM62983.2024.10814277>
- The Covid-19 crisis: A catalyst for government transformation? [WWW Document], 2020. OECD. URL [https://www.oecd.org/en/publications/the-covid-19-crisis-a-catalyst-for-government-transformation\\_1doc0788-en.html](https://www.oecd.org/en/publications/the-covid-19-crisis-a-catalyst-for-government-transformation_1doc0788-en.html) (accessed 8.14.25).
- United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, World Bank (Eds.), 2009. System of national accounts 2008. United Nations, New York.
- Varghese, B., Buyya, R., 2018. Next generation cloud computing: New trends and research directions. *Future Gener. Comput. Syst.* 79, 849–861. <https://doi.org/10.1016/j.future.2017.09.020>
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Yang, B., Li, M., 2025. Offshore Embeddedness Beyond the Wall: Chinese Cloud Providers in Southeast Asia’s Data Governance Landscape. *Polit. Gov.* 13. <https://doi.org/10.17645/pag.10437>

## A. Email templates for the UK and Finland

Template for the United Kingdom:

I am writing to make a request under the Freedom of Information Act 2000. I am seeking information from the [???] about how their information systems are hosted. I am specifically referring to [?].

For [the/each] information system, I would like to know the following details on the [?].

1. Is the system hosted on-premises by the [?], by a commercial hosting provider (e.g., a cloud provider such as Amazon Web Services), a combination of the two, or via some other arrangement?
2. If a commercial hosting provider is involved, please supply the name of the provider(s).
3. Since when have the current hosting arrangements been in place?

Please provide this information by email.

Jaakko Kilpi, Research Assistant, Aalto University  
Department of Computer Science, Konemiehentie 2, 02150 Espoo, Finland

Vili Lehdonvirta, Professor, University of Oxford  
Oxford Internet Institute, 1 St Giles, Oxford OX1 3JS

Template for Finland:

Pyydän [viranomaista x] julkisuuslain 621/1999 nojalla luovuttamaan tietoa sen tietojärjestelmien ylläpidosta. Viittaa tietojärjestelmillä [palveluun/palveluihin] [URL].

Pyydän [jokaisen tietojärjestelmän osalta/tietojärjestelmästä] seuraavia tietoja:

1. Onko tietojärjestelmän ylläpito ja isännöinti järjestetty [viraston x] omasta toimesta, kaupallisen palveluntarjoajan kautta (esim. Amazon Web Services), näiden kahden vaihtoehdon välimuotona vai jonkin erilaisen järjestelyn kautta?
2. Jos jokin kaupallinen palveluntarjoaja on mukana palvelun ylläpidossa, selvitä kaikkien näiden kaupallisten toimijoiden nimet.
3. Mistä lähtien nykyiset ylläpidon käytännöt ovat olleet voimassa?

Vastausta pyydetään sähköisessä formaatissa mielellään kahden viikon kuluessa. Jos vastaus ei ole mahdollinen, pyydämme perustelevaan, miksi selvitykseen vaaditaan pidempi kuukauden toimitusaika.

Jaakko Kilpi, Research Assistant, Aalto-yliopisto  
Vili Lehdonvirta, Professor, Aalto-yliopisto  
Konemiehentie 2, 02150 Espoo, Finland

