Journal of *Cybersecurity and Privacy*

MDPI

Systematic Review

Systematic Artefact-Based Review of Government Digital Identity Programmes: Alignment, Maturity and Transparency

Matthew Comb * and Andrew Martin

Department of Computer Science, University of Oxford, Oxford OX1 3QD, UK; andrew.martin@cs.ox.ac.uk
* Correspondence: matthew.comb@cs.ox.ac.uk

Abstract

Digital identity is increasingly treated as foundational infrastructure for digital economies and public services, yet national approaches remain fragmented and difficult to compare. This study presents a PRISMA-guided systematic artefact-based review of government digital identity programmes, using programme-relevant government artefacts as the review corpus, including strategies, trust frameworks, guidance, service documentation, and identity-enabled public-service materials. Adapting an NLP pipeline for large-scale digital identity text analysis, the study identifies recurring themes, constructs comparative programme profiles, and operationalises three artefact-based measures: alignment, transparency, and maturity. Rather than assessing innovation performance or operational system quality directly, it examines the documentary layer through which programmes are described, justified, and made comparable. The analysis reveals substantial variation in how highly digitalised societies articulate governance, trust, interoperability, security, privacy, and service delivery. The review contributes a repeatable artefact-based framework for cross-jurisdictional comparison and provides a baseline for ontology development and future triangulation against citizen perception, expert assessment, and technical evaluation.

Keywords: digital identity; ecosystem; interoperability; ontology; natural language processing; identity management; socio-technical systems; governance; privacy; security; trust

1. Introduction

Digital identity has evolved rapidly as governments and organisations seek trusted, interoperable systems for an increasingly interconnected world [1–3]. Its significance extends beyond online verification: it underpins digital economies, modern governance, and inclusion for marginalised populations [4]. Yet, despite three decades of identity paradigms and implementation models, a universal and cohesive digital identity infrastructure remains elusive [5]. This fragmentation may be viewed either as a missed opportunity or, for those concerned with unintended consequences and data privacy, as a protective outcome [6,7].

Two notable advancements that promise stability in this dynamic environment are the Electronic Identification, Authentication and Trust Services (eIDAS) regulatory reform [8] and the Verifiable Credentials (VC) framework of the World Wide Web Consortium (W3C) [9]. eIDAS is a European Union (EU) regulation that standardises electronic identification and trust services for electronic transactions across member states, aiming to enhance trust in online services and transactions. This regulatory framework ensures that individuals and businesses can use their own national electronic identification schemes (eIDs) to access public services in EU countries where eIDAS is supported.

Received: 21 March 2026
Revised: 29 April 2026
Accepted: 15 May 2026
Published: 21 May 2026

W3C's VC standard promotes decentralized digital identity as a mechanism for creating, issuing, and verifying digital statements about an individual or entity with an emphasis on security, privacy, and user control. Together, these developments illustrate how centralised regulatory oversight and decentralised technical mechanisms may be combined. eIDAS provides a legal and regulatory framework for cross-border recognition of electronic identification and trust services, while W3C Verifiable Credentials provide a technical data model for issuing, holding, presenting, and verifying digital claims in a more decentralised manner. This combination offers one possible pathway toward a more cohesive and stable digital identity ecosystem [10].

The present study is positioned differently from macro-level digital government benchmarks. Indices such as the UN EGDI, OECD DGI, World Bank GTMI, and European eGovernment Benchmark assess broad digital government capability, readiness, service delivery, or transformation. By contrast, this article examines the public artefacts through which governments describe and evidence digital identity programmes. The focus is therefore not national digital government performance as a whole, but the documentary layer of digital identity programmes as a distinct object of comparison.

1.1. Research Motivation

A report commissioned by the Institute for Prospective Technological Studies (IPTS), part of the European Commission's Joint Research Centre (JRC), argues that cross-jurisdiction digital identity integration is constrained by socio-technical barriers extending beyond standards selection. These include distrust and anonymity concerns, high transition costs, legacy lock-in, limited organisational capacity, divergent legal requirements, usability barriers, digital divides and "digital drop-out", and tensions between interoperability-driven data sharing, privacy, data protection, and cross-border profiling risks [11].

Interoperability is further weakened by semantic ambiguity, misaligned trust and assurance expectations, fragile lifecycle controls such as enrolment and revocation, protocol and format translation challenges, and fragmented standards. Together, these factors increase programme risk and limit the transposability of solutions across institutional contexts, supporting the need for staged pilots and sustained attention to interoperability, enrolment, interface design, and legal certainty. Martin and Martinovic also identify the absence of national identifiers as a further barrier to national digital identity implementation [12].

Collectively, these barriers complicate the pursuit of a unified approach. Any interoperability agenda must also account for the practical challenges of bridging countries with largely centralised, government-based eID systems (often aligned with eIDAS-style trust frameworks) and countries adopting decentralised or wallet- and credential-mediated variants [13]. At the same time, standards and regulation must remain sufficiently open and flexible to accommodate diverse implementation paths—so that advanced ecosystems (e.g., Estonia) are not constrained by lowest-common-denominator requirements, while jurisdictions at earlier stages are still able to adopt a clear baseline and progressively catch up [14]. Against this backdrop, the following questions guide the study:

- Q1: Are governments aligned in their approach to digital identity solutions?
- Q2: What is the maturity of leading government digital identity programmes?
- Q3: What published information do governments prioritise in their digital identity approach?
- Q4: What patterns have emerged in these developments that could pave the way for a stable and universal digital identity?

To address these questions, we use a text-mining and NLP-based approach, we extract comparable patterns from programme artefacts (e.g., strategy, guidance, and supporting documentation) to support cross-country analysis based on what governments publish.

These aims are motivated by several persistent gaps in the digital identity literature, outlined below.

1.2. Research Gap

Digital identity is increasingly treated as foundational infrastructure for digital government and the digital economy. Yet, despite substantial investment and rapid technical innovation, the field remains fragmented and difficult to scale across jurisdictions and sectors. A universally accepted approach to interoperability remains elusive [15–17], and existing efforts are often shaped by localised policy constraints or sector-specific requirements, limiting portability and reuse. While ontologies and semantic alignment are frequently proposed as mechanisms for bridging divergent models, no single approach has yet gained sufficient traction to unify the domain.

At the same time, the literature remains unevenly distributed across technical, policy, and implementation concerns. Much of the existing work focuses either on architectures, standards, and enabling technologies, or on broad questions of digital government capability. Far less attention has been given to repeatable ways of comparing how governments publicly articulate, prioritise, and evidence digital identity programmes through their published artefacts. This matters because programme design is often distributed across strategies, trust frameworks, service guidance, regulatory materials, and implementation documents, yet these sources are rarely analysed together as a comparative evidence base.

Against this backdrop, several key gaps further motivate the research presented in this paper:

- **Decentralised identity:** SSI promises privacy and user empowerment, but evidence on integration into government systems and large-scale public-sector deployments remains limited [18,19].
- **Public-sector scraping:** Ethical frameworks for data mining and web-scraping are underdeveloped, particularly regarding privacy, consent, and potential misuse of government-linked data [20,21].
- **Ontologies in practice:** Ontologies are widely cited as enablers of interoperability, yet practical application and real-world evidence in cross-border or multi-agency settings remain sparse [22–24].
- **Programme comparison:** Comparative evidence linking national context, levels of digital maturity, and programme design remains limited, particularly where comparison depends on publicly available government artefacts rather than broad national indicators alone [21,25].
- **Integration of emerging technologies:** Blockchain, AI, and biometrics are frequently proposed, but guidance and empirical evaluation of integration and long-term implications remain limited [26,27].

Accordingly, the gap addressed in this paper is not the absence of another digital identity architecture or a new NLP algorithm. Rather, it is the absence of a digital-identity-specific, corpus-based comparative approach that can analyse government-published artefacts in a systematic and repeatable way, and use them to benchmark how programmes are described, evidenced, and aligned across jurisdictions.

1.3. Research Contribution

Motivated by the fragmentation of national digital identity efforts and the absence of a repeatable basis for programme-level cross-country comparison, this paper develops a systematic artefact-based comparative survey of government digital identity programmes. It treats government-published artefacts—including strategies, guidance, trust frameworks,

and service documentation—as the unit of analysis, using them to examine how programmes are described, justified, governed, and made externally inspectable.

Building on our prior NLP pipeline for large-scale analysis of digital identity corpora [5], we adapt a text-mining approach from patent analysis to government programme artefacts. The contribution is not a new NLP algorithm, a conventional systematic literature review, or a general-purpose maturity model, but a repeatable survey framework for analysing the public evidential layer of digital identity programmes. To support this, we define three artefact-based measures—alignment, transparency, and maturity—which capture documentary emphasis, breadth and balance of published coverage, and similarity between national programme profiles. These measures do not evaluate technical implementation, innovation performance, or social success directly; rather, they provide a structured basis for comparing how digital identity programmes are represented in public documentation.

The scientific value of this approach lies in treating public programme artefacts as an analytically meaningful layer of digital identity ecosystems. These artefacts shape what citizens, relying parties, vendors, researchers, and policymakers can know about a programme without privileged access to internal systems. They also help form the public basis on which claims about trust, legitimacy, accountability, usability, privacy, and security are communicated and scrutinised. Accordingly, this paper makes the following contributions:

1. **Public evidential layer:** We identify government-published digital identity artefacts as a distinct unit of analysis and argue that this layer shapes the public conditions under which programmes are explained, trusted, scrutinised, and compared.
2. **Survey corpus:** We curate a cross-country corpus of publicly available government digital identity materials and define a programme-oriented framing for systematic artefact-based comparative analysis.
3. **Artefact-based survey measures:** We define and operationalise *alignment*, *transparency*, and *maturity* as measures of published documentary emphasis, breadth, balance, and profile similarity, rather than as direct measures of operational performance.
4. **Cross-country documentary comparison:** We generate comparable national programme profiles and identify patterns of convergence and divergence in how digital identity is publicly described and evidenced.
5. **Basis for triangulation:** We establish a documentary baseline for future work comparing public trust claims with citizen perceptions of usability, privacy, security, and trustworthiness, as well as with technical or institutional assessments of deployed systems.

While this study focuses on government digital identity programmes and selected frameworks such as eIDAS and W3C Verifiable Credentials, digital identity encompasses a broader set of actors, technologies, and governance arrangements. The approach is intended as a complementary comparative layer that can be extended to additional jurisdictions, artefact types, and identity paradigms, and combined with deeper evaluation of operational deployments and user outcomes.

Practically, the framework can help policymakers, researchers, and programme evaluators inspect the public visibility of digital identity programmes. It can identify where documentation is broad and balanced, where particular themes dominate, and where public evidence is limited or uneven. In future work, the approach can be operationalised as a monitoring or audit-support tool by extending the corpus to additional jurisdictions, adding multilingual collection, refining indicator sets, and triangulating artefact-based outputs against citizen perception, expert review, technical architecture, and operational evidence.

1.4. Paper Structure

This article begins by providing background on the evolution of digital identity, from simple username–password systems to more advanced, decentralised, and regulatory-compliant frameworks. Section 3 presents the literature review, covering data mining, ontology construction, cross-country comparison, and comparative digital government benchmarking and maturity models. Section 4 outlines the study methodology, including corpus construction, data mining, natural language processing, clustering, artefact-based comparative measures, validation procedures, and the software tools used in the analysis. Section 5 presents the results and discussion, including country-level findings and comparative analysis of published digital identity programme artefacts. Section 6 discusses the limitations of the study, including issues of corpus scope, language, publication practices, and the interpretive limits of artefact-based analysis. Finally, Section 7 concludes the article by summarising the main findings, their implications, and directions for future research.

2. Background

Digital identity systems have evolved considerably over time, owing to technological advancements, changing user expectations, and shifting regulatory landscapes. Early implementations centred on the username-and-password model, pioneered by Fano and Corbato in the 1960s [28], which offered a basic—albeit vulnerable—means of authenticating users in an online environment. Over time, governments and corporations expanded upon this model by issuing credentials via identity providers (IdPs), thereby formalising processes for identity management [29].

2.1. Early Approaches to Digital Identity

Building on rudimentary username–password systems, centralised IdPs emerged to provide a single authentication point. IdPs simplified user experiences by offering single sign-on (SSO) functionality, enabling access to multiple services with one set of credentials [30]. While delivering efficiency and cost benefits [4], the centralised approach created new vulnerabilities: if an IdP fails or is compromised, trust in the system erodes [31]. Recognising these limitations, federated identity models introduced cross-domain authentication, enabling multiple identity providers to collaborate in a shared trust network [32,33]. Neither purely centralised systems (e.g., Microsoft Passport) nor solely federated models (e.g., the UK's Verify) succeeded without sufficient user adoption [34,35], spurring calls for interoperability standards [36].

Simultaneously, security measures advanced through two-factor authentication (2FA), which requires users to supply two distinct proofs of identity, thus significantly reducing the threats posed by stolen or weak passwords [37–39]. Governments worldwide soon realised that 2FA could bolster national e-services; for example, the United Arab Emirates leverages passcode-only, two-factor, and even three-factor authentication [40].

2.2. Strengthening Identity Security

To address the shortcomings of password-based approaches, Public Key Infrastructure (PKI) introduced robust cryptographic methods linking public and private keys to individuals [41,42]. Taiwan's early adoption of PKI demonstrated its potential for safeguarding electronic data exchange [43], and Estonia's longstanding e-governance success highlighted PKI's importance in ensuring trust and nonrepudiation [44,45]. However, PKI projects pose operational challenges, including user adoption hurdles in geographic regions such as the Middle East [46].

Biometric authentication has likewise gained popularity, propelled by the ubiquity of smartphones and the need for more secure, user-friendly solutions. Methods such as

fingerprint scanning, facial recognition, iris scans, and voice recognition leverage individuals' unique physiological features [47]. Governments in countries such as India [48], Nigeria [49], and Pakistan [50] have adopted biometric-based ID systems, aiming to reduce identity fraud, streamline services, and improve inclusivity. Despite these advantages, biometrics call for stringent safeguards to maintain user trust.

2.3. The Decentralised Shift

Shifting away from centralised systems, decentralised and self-sovereign identity (SSI) models empower users to control their data and mitigate risks associated with storing sensitive information in a single repository [51]. SSI emphasises principles like data minimisation and minimal disclosure, strengthening citizen trust and offering potential cross-border interoperability. Complementing these decentralised approaches are single sign-on solutions, which unify disparate government services under a single credential framework [52–55], reducing both user inconvenience and administrative overhead.

2.4. Regulatory Frameworks & Privacy-Focus

As digital identity practices matured, international bodies and national authorities developed standards and regulations to ensure security and interoperability. Notable initiatives include OpenID, VCs, and eIDAS which, as mentioned previously, harmonises digital identification and trust services across member states [8,56,57]. These measures define data schemas, set frameworks for mutual legal recognition of electronic signatures, and enforce interoperability requirements.

In particular, eIDAS mandates cross-border recognition of electronic identities within the EU and positions them as legally equivalent to traditional methods [8]. VCs augment these regulatory efforts by enabling granular, verifiable claims that are selectively disclosed depending on context [51,58], as illustrated by recent COVID-19 vaccination passports [9,10,59]. Despite these breakthroughs, ensuring robust key management and privacy protections remains paramount.

Finally, privacy-first strategies have become increasingly crucial amid high-profile data breaches and public concern about personal information misuse. These approaches prioritise user consent, data minimisation, anonymisation, and strong encryption—factors that ultimately fuel public acceptance and participation in e-government initiatives. As Cameron defined in 2005, digital identity fundamentally comprises "a set of claims made by one digital subject about itself or another," [60] and contemporary frameworks have evolved to uphold these claims while minimising the risk of misuse and maximising user autonomy.

3. Literature Review

The following literature review builds on the digital identity context established in the previous section by examining five key areas: (1) how government text and websites can be mined for data, (2) how such data inform the construction of ontologies, (3) why ontologies are valuable for knowledge representation in e-government contexts, (4) how these methods can support cross-country comparison, and (5) how the present study is positioned in relation to comparative digital government benchmarking and maturity-model literature. By exploring the existing body of research around these topics, we establish a foundation for understanding how public-sector information can be translated into structured, interoperable knowledge resources, and how government-published digital identity artefacts can be treated as a distinct unit of comparative analysis. This provides the foundation for analysing and comparing government services and digital identity initiatives across jurisdictions.

3.1. Data Mining of Government Resources

Data mining of government websites and related text corpora has become a pivotal aspect of modern e-governance, particularly in applications such as financial technology (fintech) and public health. By automating the extraction of structured data from websites and other digital sources, researchers and policymakers can rapidly evaluate public sentiment, track policy implementation, and support data-driven decision-making [15–17,21,26]. To begin we provide an overview of foundational techniques, tools, and ethical considerations in web scraping, followed by a discussion of the implications for e-government contexts and digital identity.

3.1.1. Overview of Data Mining Techniques

Data mining of government text and websites typically involves automated methods of extracting structured datasets from online sources [15–17,21,26]. Researchers often employ archival services such as the Wayback Machine (Wayback Machine URL = https://web.archive.org/) to capture historical snapshots of webpages, which allows for longitudinal analysis of policy changes, public discourse, or evolving government services [61]. Once collected, the data can be processed using machine learning techniques—including regression models, clustering methods, and deep learning—that reveal patterns and insights pertinent to policy-making and public-service improvements [17,23,27,62].

In practice, a variety of tools and technologies support data collection and analysis workflows, including programming libraries, statistical software, and cloud-based analytics platforms [16,17,26,27]. Legal and ethical considerations are also essential in guiding data collection efforts, with cases like *hiQ v. LinkedIn* illustrating the distinction between publicly accessible and private data under the United States' Computer Fraud and Abuse Act (CFAA) [20,63]. In addition, ethical concerns, such as privacy, data misuse, and data representativeness, are particularly significant when handling sensitive information, highlighting the importance of responsible and compliant research practices [16,21].

3.1.2. E-Government Context

In the public sector, e-government efforts face multiple additional challenges: inconsistent data formats, high data volume, and restricted website access can hinder straightforward data mining [25–27]. Government websites often use technical defenses like IP blocking and CAPTCHAs—tests to distinguish humans from bots—and frequently update their architectures. These measures demand highly adaptable scraping frameworks [16,61]. Despite these hurdles, data mining has proven valuable in diverse e-government applications, including the measurement of public sentiment during crises (e.g., COVID-19), the tracking of evolving governmental measures across different regions, and the prediction of resource needs [17,21,26]. For instance, analysing data from social media and official platforms provides insights into the adoption and adaptation of digital identity systems, informing strategies to improve authentication mechanisms and build citizen trust through metadata analysis and pattern recognition [15,16,26].

3.2. Knowledge Construction

Ontology construction from data-mined websites is crucial for transforming raw or semi-structured information into coherent semantic frameworks that define key concepts, relationships, and attributes. These frameworks enable a unified, semantically rich view of otherwise fragmented datasets, benefiting e-government by enhancing policy analysis, optimising resources, and improving service delivery [16,21,22,24,27]. Techniques for building such ontologies typically combine automated approaches—such as clustering,

topic modelling, and Named Entity Recognition (NER)—with domain-expert curation to ensure relevant and accurate structures [17,23,27,61].

Despite these methodological advances, challenges persist in aligning data across multilingual sources, inconsistent formats, and diverse governance structures [16,23]. Adhering to standards like the Simple Knowledge Organization System (SKOS) helps maintain semantic alignment, while privacy regulations such as the General Data Protection Regulation (GDPR) impose necessary constraints on data handling [20,63]. Consequently, constructing ontologies demands not only sophisticated technical solutions but also careful compliance with ethical and legal requirements, ensuring the protection of sensitive data [20,21].

Once established, ontology-based knowledge frameworks support both real-time and longitudinal analyses, offering decision-makers deeper insights into emerging issues and long-term trends [27,61]. Predictive models built atop these structures reveal policy gaps and inform targeted interventions, ultimately reducing risks and improving government services [16,17,26,64]. This integrated approach empowers governments to make timely, evidence-based decisions and fosters more transparent, efficient, and citizen-focused public administration.

3.3. Cross-Country Comparisons

Cross-country comparisons gleaned from government websites highlight both the potential synergies and persistent disparities in e-government initiatives. However, legal, cultural, linguistic, and technical differences across regions complicate uniform data collection and hamper straightforward benchmarking, while archived web data often lacks coverage or uniformity, limiting the scope of cross-country analyses [16,25,27,61]. Ontology-based approaches and text-mining methods can support more structured comparison by standardising concepts and indicators, but they do not by themselves eliminate differences in publication practice, corpus composition, or documentary scope [15–17]. By examining policy adoption, public sentiment, and citizen engagement in real time, researchers can identify patterns, best practices, and systemic challenges that transcend geographic boundaries, thereby informing more cohesive and data-driven policy strategies [21,26,61]. At the same time, such comparisons must be interpreted cautiously, as differences in what governments publish may reflect communication style, administrative structure, or document genre as much as underlying programme characteristics.

3.4. Comparative Digital Government Models

A parallel body of literature relevant to this study concerns comparative digital government benchmarking and maturity models. At a macro level, the United Nations E-Government Development Index (EGDI) provides a composite assessment of e-government development based on online services, telecommunications infrastructure, and human capital [65]. The OECD Digital Government Index (DGI) benchmarks the foundations required for coherent, human-centred digital transformation in the public sector [66]. The World Bank GovTech Maturity Index (GTMI) evaluates public-sector digital transformation across core government systems, public-service delivery, digital citizen engagement, and GovTech enablers [67]. The European Commission's eGovernment Benchmark assesses the digitalisation of public services, including dimensions such as user centricity, transparency, and key enablers [68]. Collectively, these frameworks provide important comparative baselines for digital government capability, service delivery, and public-sector digital transformation.

However, these frameworks operate at a different level of analysis from the present study. Their primary concern is broad digital government readiness, capability, service provision, or enabling conditions rather than the documentary footprint of national digital identity programmes. In the wider digital government literature, digital identity is com-

monly treated as one enabling component within digital public infrastructure rather than as the sole object of comparative assessment [69]. Consequently, while these indices are highly valuable, they do not directly measure how governments publicly describe, justify, govern, and evidence digital identity programmes through published artefacts.

Recent work on digital government maturity models also shows that the field is heterogeneous in scope, assumptions, and treatment of citizen centricity [70]. This heterogeneity leaves room for more focused, domain-specific benchmarking approaches that complement rather than replace macro-level frameworks. The present study is positioned in that complementary space. It does not seek to supplant established digital government indices, nor to claim that artefact-based measures are equivalent to operational performance. Instead, it introduces a digital-identity-specific, corpus-based comparative method centred on government-published artefacts, enabling structured comparison of published evidence across programmes.

4. Method

The research design follows a systematic artefact-based comparative survey approach. The method developed in this study is specific to government digital identity programmes and to the public artefacts through which those programmes are described, rather than to government digital programmes in general. The unit of analysis is the public documentary footprint of each government digital identity programme, rather than the deployed system itself or the wider national digital government ecosystem.

This review was reported in accordance with the PRISMA 2020 reporting framework where applicable to a systematic review of government-published programme artefacts rather than clinical trials or intervention studies [71]. The empirical object of the review is publicly available government digital identity programme documentation, not academic publications alone. The completed PRISMA 2020 checklist is provided as Supplementary Table S1. No prospective protocol registration was completed. The search strategy, source list, eligibility criteria, preprocessing approach, and analysis pipeline are reported in the Methods Section and Supplementary Materials to support reproducibility.

The method proceeds in five broad phases. First, potentially relevant government digital identity artefacts were identified from literature-derived sources, targeted web search, and official government websites. Second, records were screened against eligibility criteria and assigned to country-level strata. Third, retained artefacts were cleaned, normalised, and prepared for analysis. Fourth, frequency-based indicator mapping, cluster/topic analysis, and relationship extraction were applied. Fifth, the resulting outputs were synthesised through the artefact-based measures of alignment, transparency, and maturity.

In 2011 the UK Government initiated a project to deliver an identity assurance system called “GOV.UK Verify”. In later years, the scheme was heavily criticised due to its failure to meet projected targets and its delayed implementation—being four years behind schedule and heavily over budget. Many post-mortem reviews were conducted on the platform and summarised in a report by the Open Identity Exchange—’Digital Identity in the UK: The cost of doing nothing’—which described the key digital identity ecosystem indicators used to determine the ecosystem’s level of success.

While we have selected this report as a guiding reference for Table 1 because it is explicitly ecosystem-focused and provides a practical indicator set grounded in programme implementation experience, it should be acknowledged that other authoritative reports contribute additional points that are not always captured by ecosystem adoption indicators alone. The OpenID Foundation’s human-centric digital identity guidance [72] highlights values- and rights-based evaluation, privacy and security by design, and operational resilience expected of identity as critical infrastructure. The World Bank’s Principles on

Identification [73] further stress standards-based interoperability, openness and technology/vendor neutrality, long-term sustainability, and institutional governance requirements including accountability and grievance redress. Accordingly, we treat the selected ecosystem indicator set as a pragmatic baseline, and interpret it in conjunction with these complementary human-centric and governance-oriented perspectives when analysing national programmes.

Table 1. Success and failure indicators of a digital identity ecosystem [35].

	Success and Failure Indicators	Associated Words
(1)	Private sector involvement in design and delivery	partnership, provider, industry, sector, company, business, innovation, accreditation, institution
(2)	A shared vision involving government and industry	principle, interoperability, standardisation, collaboration, synergy, consensus, specification, team, cooperation, policy, participant, participation, protocol, convention, role
(3)	Wide range and availability of ID enabled services	accessibility, integration, diversity, ubiquity, comprehensiveness, reader, chip, welfare, expansion, ecosystem, child, family, age, tax, cargo, ship, disease, health, insurance, transportation, freight, passenger, airport
(4)	Banking services accessible using ID	bank, finance, transaction, order, account, wallet, signature
(5)	High frequency of use	application, app, service, user, consumer
(6)	Existence of a mandatory ID	compulsory, legal, obligatory, universal, enforcement, citizen, mitid, realme
(7)	An accepted history of national identity schemes	trust, acceptance, tradition, familiarity, confidence
(8)	A national residential register	centralisation, database, registry, recording, tracking, foreigner, immigration
(9)	Available to be used via a variety of channels	phone, multifunctional, versatile, flexible, multi-channel, accessibility, mobile, passport, online, voice
(10)	Comprehensive enrolment strategies using KYC data and passive approaches	KYC, streamlined, efficient, inclusive, comprehensive, process
(11)	Public trust in the scheme	security, privacy, reliability, confidence, credibility, transparency, assessment, proof, test, safety, certificate, authentication, risk, police, metadata, fraud, incident
(12)	Liability model and trust framework addressed	accountability, clarity, responsibility, trust, framework, authority, agency, commission, resolution, principal, oversight
(13)	A clear business case	cost, benefit, ROI, justification, value, future, economy
(14)	Regulatory clarity/confidence	compliance, legislation, confidence, certainty, regulation, parliament, governance, rule, law
(15)	Well-connected government IT and databases	integrated, connected, unified, streamlined, interoperable, architecture, soap, schema
(16)	Barriers to access an ID removed or addressed	inclusive, accessible, equal, barrier-free, facilitation
(17)	Strong public awareness and education	awareness, education, knowledge, outreach, information, environment

The selection of the indicator framework was treated as a construct-design decision rather than as a claim that one set of indicators is universally correct. Several alternative sources were considered. Macro-level digital government benchmarks, including the UN E-Government Development Index, OECD Digital Government Index, World Bank GovTech Maturity Index, and European Commission eGovernment Benchmark, provide valuable comparative measures of digital government capability, but they operate at the level of broad digital government readiness and service transformation rather than digital identity programme artefacts specifically. Regulatory and technical frameworks such as eIDAS and W3C Verifiable Credentials provide important requirements and interoperability models, but are either jurisdiction-specific or technical in orientation and do not by themselves provide a general ecosystem success/failure framework. Human-centric and rights-oriented sources, including OpenID Foundation guidance and the World Bank Principles on Identification, contribute important principles such as inclusion, accountability, privacy, security by design, openness, and sustainability, but these are primarily normative principles rather than directly operationalised artefact-scoring categories.

The UK Open Identity Exchange indicator set was therefore selected as the primary baseline because it is digital-identity-specific, ecosystem-oriented, and explicitly grounded in programme implementation experience. It includes indicators concerning public and private sector roles, shared vision, service availability, trust, regulatory clarity, liability, enrolment, public awareness, access barriers, and business case. These properties made it suitable for mapping observable language in government-published artefacts to a consistent comparative structure. The alternative frameworks were not dismissed as irrelevant; rather, they were treated as complementary interpretive sources. They were not merged into the primary scoring model because doing so would combine indicators operating at different levels of analysis and risk double-counting overlapping constructs such as trust, security, privacy, interoperability, and governance. Because the baseline indicator set is derived from a UK digital identity programme review, it may reflect assumptions, risks, and institutional priorities specific to that context. We therefore do not treat it as a universal definition of digital identity programme success. Instead, it is used as a pragmatic ecosystem-oriented baseline for a first comparative implementation of the artefact-based method. The resulting scores should be interpreted relative to this indicator set, and future work should test alternative or expanded frameworks derived from non-UK, multilingual, and developing-country contexts.

In this Methods Section of this academic article, we detail our approach to data mining government-published websites focusing on trust frameworks, digital identity, and digital onboarding content.

The research methodology involves utilising the list of identified success factors for digital identity, derived from the above-mentioned comprehensive UK report, as a foundational basis for our analysis. This analytical framework allows for a structured comparison of the data gathered across various countries. By systematically examining online governmental resources, this study aims to glean insights into the prevailing practices and strategies implemented in different nations, thus providing a greater understanding of the global landscape in digital identity and trust frameworks.

4.1. Data Sources and Search Strategy

The corpus was drawn from published government materials on digital identity, trust frameworks, digital onboarding, and digital citizenship, as summarised in Table 2. Relevant sources were identified through a two-stage process. First, the academic and practitioner literature was reviewed to identify referenced government programmes, public agencies, trust framework bodies, and official digital identity sources. Second, targeted Google searches were used to locate additional official websites and supporting documentation for each country stratum.

Searches combined country names with digital identity terms such as ”digital identity”, ”electronic identification”, ”trust framework”, ”digital identity wallet”, ”digital onboarding”, ”eID”, ”authentication”, and named national systems where known. Search results were screened manually for official status, programme relevance, text accessibility, and country assignment. Where programmes distributed material across multiple official domains, a manually verified allow-list of relevant hosts, derived from logged URLs, was supplied to the data-mining algorithm to keep collection within the intended source boundaries.

The full search strategy, including search strings, search dates, information sources, allowed hosts, and country-level inclusion notes, is provided in Supplementary Table S2.

Table 2. Government website data sources.

Country	Website	Keyword
Denmark	https://en.digst.dk/systems/mitid/ + https://www.mitid.dk/en-gb/	mitid
Finland	https://dvv.fi/en/digital-identity-reform/ + https://dvv.fi/en/european-digital-identity-wallet/	identity
South Korea	https://dgovkorea.go.kr/contents/blog + https://dgovkorea.go.kr/	government
New Zealand	https://www.digital.govt.nz/digital-government/programmes-and-projects/digital-identity-programme/	identity
Sweden	https://www.elegitimation.se/en + https://www.digg.se/en + https://docs.swedenconnect.se/technical-framework/	id
Iceland	https://island.is/en/electronic-id	identity
Australia	https://www.digitalidentity.gov.au/	identity
Estonia	https://e-estonia.com/solutions/	identity
UnitedKingdom	https://www.gov.uk/guidance/digital-identity/	identity
United ArabEmirates	https://u.ae/en/about-the-uae/digital-uae/regulatory-framework https://u.ae/en/about-the-uae/digital-uae/digital-transformation/platforms-and-apps/the-uae-pass-app	identity
Japan	https://www.kojinbango-card.go.jp/en/ + https://trustedweb.go.jp/en/	identity
Canada	https://www.canada.ca/en/government/system/digital-government/	identity
Singapore	https://www.smartnation.gov.sg/initiatives/strategic-national-projects/national-digital-identity/, https://api.singpass.gov.sg/	identity *
Netherlands	https://www.government.nl/topics/online-access-to-public-services-european-economic-area-eidas/everything-you-need-to-know-about-eidas	identity
Malta	https://mita.gov.mt	identity
United States	https://pages.nist.gov/800-63-4/+ https://www.login.gov/partners/our-services/	identity

Note: All URLs were accessed on 7 November 2023. * The Singapore endpoints have since been removed and are no longer accessible.

4.2. Data Sampling

The objective of data sampling was to construct a bounded, purposive comparative corpus of government-published digital identity material. The sample was not designed to provide an exhaustive global census of digital identity programmes, nor should countries outside the sample be interpreted as less important to the government digital identity landscape. Rather, the sample was selected to support a controlled first application of the artefact-based method across jurisdictions for which sufficiently public, text-accessible, and programme-relevant materials could be identified and associated with a defined national programme stratum.

1. **Sampling frame**—The sampling frame was limited to highly digitalised societies identified in the United Nations report *The Future of Digital Government* [74], reflecting advanced public-sector digitalisation.
2. **Country stratification**—The corpus was stratified by country to support comparison across distinct national digital identity landscapes.
3. **Source selection**—For each country, we identified the principal government website(s) publishing digital identity, trust framework, or identity-enabled e-government material, including relevant sources across administrative levels and public-sector agencies.

Country inclusion was guided by four criteria: (1) inclusion in, or close association with, highly digitalised government contexts; (2) identifiable national or national-level digital identity, trust framework, digital onboarding, or identity-enabled public-service material; (3) availability of public, text-minable artefacts, preferably in English; and (4) feasibility of assigning material to a coherent country stratum. Cases such as France, Germany, Italy, and India were not excluded for lack of significance, but because the study was deliberately bounded. Their inclusion would require additional corpus construction,

source delimitation, preprocessing, and recalculation of the comparative measures, and they are therefore treated as candidates for future extension and sensitivity testing.

4.3. Eligibility Criteria

Artefacts were eligible for inclusion where they met all of the following criteria: (1) they were publicly accessible without bypassing authentication, paywalls, or technical access controls; (2) they were published by, or clearly associated with, a national government, public agency, recognised trust framework body, or official digital identity programme; (3) they contained substantive material on digital identity, electronic identification, trust frameworks, digital onboarding, authentication, verification, credentials, identity-enabled public services, or related governance arrangements; and (4) they were sufficiently text-accessible for automated extraction and preprocessing.

Artefacts were excluded where they consisted primarily of navigation pages, generic news listings, media galleries, unrelated public-service material, pages outside the selected country stratum, or content that could not be reliably extracted as text. Video-only, image-only, heavily dynamic, or inaccessible pages were not included unless equivalent extractable text was available.

4.4. Selection Process

Source identification, screening, and corpus construction were conducted by the lead author. Eligibility decisions were documented through retained URL logs, allowed-host lists, and preprocessing outputs. Because the study is an artefact-based review of public programme documentation rather than a clinical evidence synthesis, formal duplicate independent screening was not undertaken.

Records were screened first at the level of URL, title, source domain, and metadata, and then, where necessary, by inspecting extracted page text. Borderline cases were resolved conservatively by applying the eligibility criteria above and retaining only material with a clear programme-level connection to the review scope. Duplicate and near-duplicate pages were removed before final preprocessing. Exclusion reasons were recorded at the level of URL or source category where available, including non-substantive content, unrelated administrative material, inaccessible or media-only content, duplicate content, and material outside the selected country stratum.

4.5. Data Items

For each retained artefact, the following data items were recorded where available: country stratum, URL, source domain, page title, retrieval status, source type, language status, extracted body text, preprocessing status, retained token set, word-frequency outputs, success-factor associations, cluster assignments, cluster labels, and extracted subject–predicate–object relationships. These data items formed the basis for the frequency, clustering, graph, and metric calculations reported below.

4.6. Corpus Bias and Source-Composition Assessment

Because this review analyses public programme artefacts rather than intervention studies, conventional study-level risk-of-bias assessment was not applicable. Instead, corpus-level bias was assessed qualitatively through source-composition checks. These considered whether each country corpus was dominated by policy, technical, regulatory, service, administrative, video-based, non-English, or otherwise unevenly distributed material. Source-composition effects are reported in the country-level discussion and limitations, particularly where they affect interpretation, as in the Estonia and Iceland cases.

4.7. Data Collection

In this research, we employed automated web scraping to gather textual data from selected government websites relevant to digital identity, trust frameworks, and identity-enabled public services. Data collection was restricted to websites and subdomains selected for substantive relevance to the study domain, with the aim of capturing material directly related to digital identity policy, governance, authentication, verification, credentials, and associated public-service access mechanisms.

To improve corpus relevance, pages were retained where the URL, title, metadata, or body text indicated a substantive connection to digital identity, trust frameworks, identity-enabled service access, or related government service functions. Pages consisting primarily of navigation material, news listings, media galleries, duplicate archives, or unrelated administrative content were excluded. For dynamically rendered pages, embedded applications, images, or video-led sources, only reliably extractable text was included. As a result, visually rich, application-based, or non-textual material may be underrepresented in the final corpus.

Data collection was automated in a controlled manner intended to avoid service disruption, excessive request load, or the bypassing of authentication barriers. No personal data were intentionally collected, and analysis was limited to publicly available programme documentation. Retained text was cleaned and structured by removing HTML and other non-substantive markup, with regard to relevant international data protection principles [75].

No systematic machine translation was applied. English-language programme pages or official English translations were preferred to support cross-country comparability. Where non-English terms remained in the extracted text, they were retained as corpus tokens rather than translated or inferred manually. This preserves traceability to the mined material, but may disadvantage jurisdictions whose most detailed documentation is unavailable in English or primarily embedded in images, videos, dynamic applications, or other non-textual formats. Variation in public availability, structure, language, and text accessibility across jurisdictions is therefore treated as a limitation of the study.

The associated words in Table 1 should be understood as indicative lexical cues rather than as a complete controlled vocabulary or formal ontology. Some terms are necessarily broad or context-dependent because government digital identity artefacts use heterogeneous policy, technical, legal, and service-delivery language. To reduce over-classification, ambiguous or generic terms were classified using the fixed prompt described in Section 4.9, which allowed the model to return "None" where no sufficiently clear association with a success factor was present. The resulting associations were used for descriptive benchmarking only and were checked against country-level word-frequency lists, cluster summaries, and interpretive discussion to identify obviously inconsistent mappings.

Figure 1 summarises the identification, screening, eligibility assessment, and inclusion process used to construct the artefact corpus. Sixteen country/source datasets were initially identified; four were either not mineable or retained for qualitative/contextual discussion only, leaving twelve datasets for crawling and metric analysis. Across these datasets, 15,344 pages were screened, 4192 met the screening criteria, 3855 were retrievable, and 3393 provided analysable data and were retained in the final corpus.

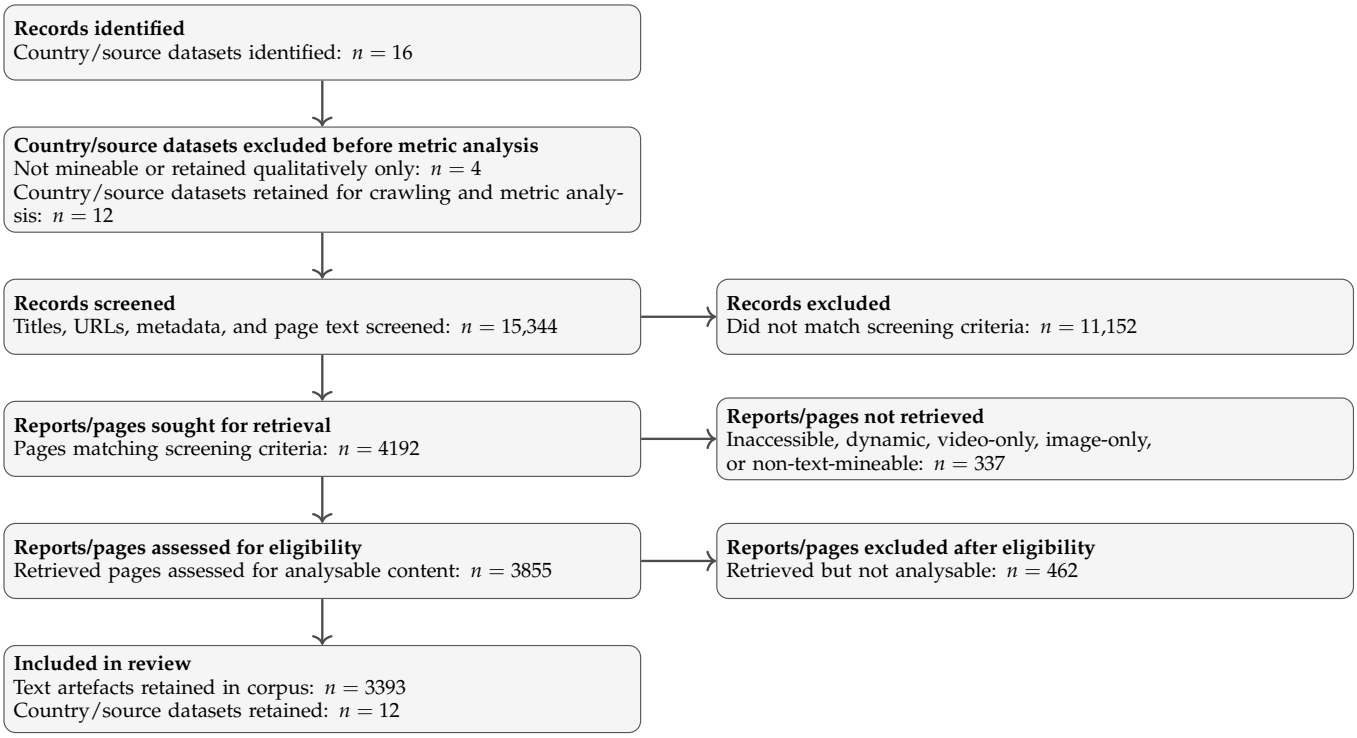

Figure 1. PRISMA 2020-guided method flow diagram.

4.8. Preprocessing

During this phase, a series of preprocessing steps were implemented to improve the consistency and reliability of subsequent analyses. These steps included systematic data cleaning, core NLP operations, and staged workflow validation, as detailed below.

1. **Data Cleaning**
   Preliminary efforts were made to refine the text data and prepare it for advanced processing:
   (a) *Text Normalisation*: All alphabetic characters were converted to lowercase to maintain uniformity and reduce complexity.
   (b) *Punctuation and Numeral Removal*: Punctuation marks, special characters, and numerals were systematically excluded, restricting the dataset to textual content.
   (c) *Redundant Successive Word Removal*: Consecutive duplicate words were identified and replaced with a single instance, minimising repetitive text noise.
   (d) *Duplicate and Near-Duplicate Reduction*: Where substantially identical textual content was encountered across multiple pages, duplicate or near-duplicate instances were removed to reduce corpus inflation and repeated signal.
2. **NLP Preprocessing**
   After cleaning, natural language processing techniques were applied to further structure the corpus:
   (a) *Tokenisation*: The text corpus was split into individual word tokens, forming the basis for subsequent NLP tasks.
   (b) *Lemmatisation*: Words were reduced to their dictionary forms, ensuring consistent representation of morphological variants.
   (c) *Stop Word and Redundant Word Removal*: High-frequency words with limited semantic value, as well as recurring superfluous terms, were removed to reduce noise.

(d) *Part-of-Speech (POS) Tagging*: Each token was tagged according to its grammatical category, aiding syntactic and semantic analysis.

3. **Workflow Validation Partitioning**
To support staged pipeline development and quality checking, the dataset was partitioned into an initial inspection subset and a retained analysis subset. The inspection subset comprised the initial n records from each designated stratum and was used to verify preprocessing behaviour, inspect intermediate outputs, and identify errors prior to full-scale analysis. After this validation step, the final workflow was applied to the retained corpus. Terms from the inspection subset were cross-referenced against final outputs to confirm that preprocessing, feature extraction, and aggregation behaved as expected. This partitioning step was used for workflow validation rather than supervised predictive training.

During cleaning, malformed character encodings were corrected where they occurred in the manuscript text or extracted corpus outputs. A small number of non-English or untranslated residual tokens remain in the country-level frequency tables where they formed part of the mined public corpus. These terms are retained as raw corpus evidence for transparency and are not treated as substantive findings unless they are explained in the relevant country discussion.

4.9. Data Analysis

The synthesis was descriptive and computational rather than meta-analytic. No effect sizes were pooled and no statistical meta-analysis was conducted. Instead, retained artefacts were synthesised through word-frequency analysis, success-factor association, K-Means clustering, LDA topic inspection, POS-based relationship extraction, graph-oriented exploration, and the artefact-based measures discussed above.

The assembled corpus of government website materials was analysed in Jupyter Notebook using a multi-stage pipeline designed to transform published artefacts into descriptive, comparative, and relational outputs. The purpose of the pipeline was to identify recurring terms and themes within the corpus and to support structured comparison across countries. The resulting outputs should therefore be interpreted as artefact-based analytical signals derived from published materials, rather than as direct measures of operational programme performance.

Taken as a reusable survey framework, the method comprises five high-level stages: (1) identification of programme-relevant public artefacts; (2) construction of country-level document strata; (3) preprocessing and linguistic normalisation; (4) extraction of frequency, cluster, and relationship features; and (5) comparison through artefact-based measures of alignment, transparency, and maturity. Within the fourth stage, the analysis branches into frequency-based indicator mapping, cluster/topic analysis, and relationship extraction. The framework is intended to be extensible to additional jurisdictions, artefact types, and indicator sets with resulting outputs interpreted as measures of published documentary evidence.

Figure 2 summarises the data-analysis workflow. After corpus construction and preprocessing, the analysis proceeds through three linked branches: a frequency and indicator branch used to construct the success-factor heatmap and derive alignment and transparency; a clustering branch used to summarise country-level themes and derive maturity; and a relationship branch used to generate graph-based interpretive aids. Model-assisted steps are used only for success-factor association and concise cluster labelling, and do not determine the underlying clustering, graph extraction, or mathematical form of the reported measures.

1. **Analysis of Word Frequencies:**
   For each national subset, the corpus was analysed to identify the 60 most frequently occurring terms. This descriptive step was used to establish a comparable lexical profile for each country and to identify recurring topics that warranted further analysis. The resulting frequency lists formed the input to subsequent stages of the pipeline.
2. **Success Factor Association:**
   The high-frequency terms were then associated with the success factors defined in Table 1. This step was intended to provide a structured mapping between observed terms and the study's indicator set. To support consistency in this mapping, each term was submitted to the OpenAI API using a fixed prompt template and model version GPT-4.1. The returned associations were aggregated to support the comparative analysis presented in Section 5.14.1.

   > You are assisting with a digital identity benchmarking study.
   >
   > **Task**: classify the keyword below into one and only one success factor from the predefined list in Table 1. Use the meaning most likely intended in a government digital identity context. If the keyword is too ambiguous, too generic, or unrelated to the listed success factors, return "None".
   >
   > **Rules**: 1. Select only one success factor from Table 1, or "None". 2. Do not invent new labels. 3. Base the classification on semantic relevance in the context of government digital identity programmes. 4. Return only the selected success factor name or "None".
   >
   > **Keyword**: [keyword]

   This step should be interpreted as a model-assisted classification procedure used to support descriptive benchmarking rather than as a fully validated semantic ground truth.
3. **Cluster Analysis via K-Means and LDA:**
   To identify themes beyond frequency counts, the corpus was analysed using K-Means clustering and Latent Dirichlet Allocation (LDA). These methods were applied in parallel because they capture complementary structures: K-Means groups documents by TF-IDF vector similarity, while LDA estimates latent topic distributions. Candidate cluster/topic counts were assessed using the Elbow Method and coherence testing, balancing statistical fit with interpretability.
   For LDA, each topic was summarised using the 10 highest-weighted terms in its topic-word distribution. For K-Means, each cluster was summarised using the 10 highest-weighted TF-IDF features nearest to the cluster centre. The final reported cluster set was selected based on quantitative fit and domain relevance, with both clustering outputs and algorithms retained in the code repository.
   To improve interpretability, the top terms for each final cluster were also submitted to the OpenAI API using the following prompt:

   > Identify the single most likely common theme represented by the following terms in a government digital identity context: [word1, word2, . . . , word10]. Return a concise descriptive label of no more than five words. Do not infer details not supported by the terms. Do not explain your answer. Return only the label.

   These labels were used as interpretive summaries only. They did not affect cluster formation, weighting, or scoring, and were included solely to support human-readable presentation of the results.
4. **POS Tagging and Triple Extraction:**

The final stage applied part-of-speech (POS) tagging and triple extraction to identify subject–predicate–object relationships within the corpus. This step was used to explore the relational structure of the mined text and to support graph-based visualisation. The extracted triples were published to a Neo4j database, where prominent nodes and relationships could be inspected visually for each country. These graph outputs were used as exploratory and interpretive aids rather than as direct inputs into the comparative scoring metrics.

These relationship outputs should be understood as shallow relational signals rather than as full semantic role representations of the text.

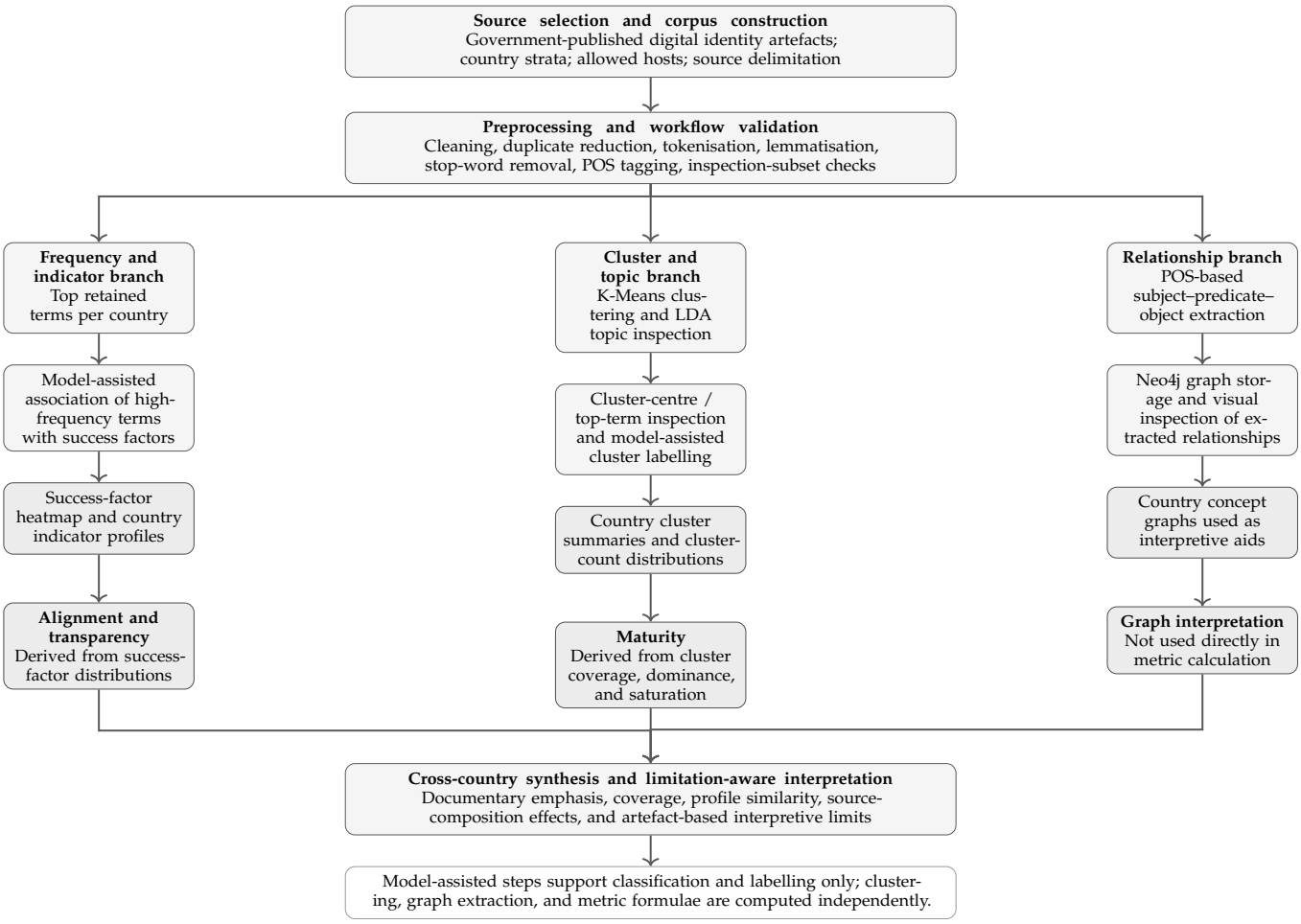

Figure 2. Artefact-based comparative survey pipeline for analysing government digital identity programme documentation.

4.10. Statistical Analysis

In this section, we define three artefact-based measures used to compare digital identity programmes on the basis of content published on national government websites. These measures are intended to characterise patterns in the published corpus rather than to provide direct measures of operational performance, citizen adoption, institutional capability, or wider societal outcomes. Accordingly, the terms *alignment*, *transparency*, and *maturity* are used here in a specific analytical sense tied to the distribution of published evidence within the selected indicator and cluster structures.

4.10.1. Alignment

In this study, alignment denotes the similarity between two national programmes in their score distributions across the success indicators in Table 1. It is an artefact-based

measure of published emphasis, not a direct measure of alignment with external standards, policy–implementation consistency, or operational interoperability. Drawing on principles from sequence alignment [76], ontology matching [77], and multi-criteria optimisation [78], the function compares indicator-level word scores using normalised similarity and aggregates them into a country-pair alignment score.

To derive a single overall alignment metric $A_{country-pair}$ between two national digital identity programmes, the following steps have been applied:

1. **Compute Pairwise Alignment Scores**
   For each pair of values that represent country-level scores against identified indicators as seen in Table 1, with $x \in D_A$ and $y \in D_B$, we calculated individual alignment scores S_i using the normalized absolute difference:

$$S_i = 1 - \frac{|x_i - y_i|}{\max(x_i, y_i)} \quad (1)$$

   This provides that the similarity score S_i lies within the range [0, 1], where higher values indicate closer similarity in published indicator scores. Where $x_i = y_i = 0$, S_i is treated as 1 with perfect alignment.
2. **Aggregate Scores Across Indicators** We then combine all pairwise scores to compute the country-pair alignment metric $A_{country-pair}$ using an average.

$$A_{\text{country}-\text{pair}} = \frac{\sum_{i=1}^{N} S_i}{N} \quad (2)$$

   where N is the total number of aligned variable pairs.
   Normalization is already inherent in the similarity calculation using the maximum of each pair, ensuring values remain within [0, 1].
3. **Metric Interpretation**

   $A_{country-pair} \approx 1$: High alignment
   $A_{country-pair} \approx 0$: Low alignment

   In plain terms, *alignment* measures how similar two countries appear in the way their published materials emphasise the selected digital identity success factors. A higher score indicates that two national corpora distribute attention across the indicators in similar ways, while a lower score indicates more divergent documentary profiles. It should not be interpreted as direct evidence of interoperability, policy agreement, or technical compatibility.

4.10.2. Transparency

In this study, transparency refers to the breadth and balance with which the selected features are represented in the published corpus. It should therefore be understood as an artefact-based measure of documentary visibility or communicative comprehensiveness, rather than as a direct measure of governance accountability, auditability, citizen oversight, source-code openness, or transparency of data processing practices. Our transparency metric references established methodologies from information theory and diversity measurement. It draws on Shannon's entropy [79] to quantify uniformity across feature distributions and integrates coverage metrics commonly used in data mining [80] to assess feature presence. By combining these approaches with weighted aggregation techniques from multi-criteria optimization [78], the transparency metric provides a systematic framework for evaluating the balance and completeness of feature mentions within the published dataset.

To evaluate the transparency of feature mentions, we consider the following metrics:

1. **Coverage (C)**
   Coverage measures the proportion of features that have been mentioned at least once:

$$C = \frac{\text{Number of Mentioned Features}}{\text{Total Features}} \tag{3}$$

2. **Uniformity (U)**
   Uniformity measures the balance of mentions across features using normalized entropy:

$$U = -\frac{1}{\log(K)} \sum_{i=1}^{K} p_i \log(p_i) \tag{4}$$

   where

   - K: Total number of features.
   - p_i: Proportion of mentions for feature i, calculated as

$$p_i = \frac{\text{Mentions for Feature } i}{\text{Total Mentions}} \tag{5}$$

3. **Transparency (T)**
   The transparency score combines coverage and uniformity, weighted by coefficients α and β:

$$T = \alpha C + \beta U \tag{6}$$

   - α: Weight assigned to coverage.
   - β: Weight assigned to uniformity.
   - For the present analysis, equal weights were adopted as a neutral baseline, such that $\alpha = 0.5, \beta = 0.5$.

   In plain terms, *transparency* measures how broadly and evenly the selected success factors are represented in a country's published corpus. A higher score indicates that more indicators are visible and that attention is distributed more evenly across them. A lower score indicates narrower or more concentrated coverage. It therefore captures documentary visibility and balance, not institutional openness or operational accountability.
   The equal weighting of coverage and uniformity is used as a neutral baseline rather than as an optimised policy judgement. We do not weight the transparency metric toward particular substantive dimensions such as security, privacy, or interoperability because the metric is intended to measure breadth and balance of documentary coverage across the selected feature set. Assigning higher weights to selected themes would change the metric from an artefact-based visibility measure into a normative priority-weighted evaluation. Nevertheless, the choice of weights may affect country ordering, and future work should report sensitivity analysis under alternative weighting schemes.

4.10.3. Maturity

In this study, maturity refers to documented programme maturity, namely the breadth and balance of evidence present across the cluster structure derived from the published corpus. It should not be interpreted as a direct measure of institutional capability, implementation success, service adoption, or trust outcomes. Our maturity metric, based on coverage, maximum value ratio, and saturation, integrates established methodologies from data analysis and distribution measurement. Coverage metrics [80] are employed to evaluate the completeness of non-empty clusters, while the maximum value ratio re-

flects the dominance of the largest feature relative to the total distribution, drawing on concepts from inequality measures such as the Gini coefficient [81]. Saturation, defined as the proportion of the mean cluster size to the maximum observed value, reflects the balance within the dataset and is inspired by proportional scoring approaches commonly used in statistics and decision science. By combining these components with weighted aggregation techniques from multi-criteria optimization [78], the maturity metric provides a framework for quantifying the balance and completeness of evidence within feature clusters.

To evaluate the maturity of clusters without relying on time, we use the following steps:

1. **Calculate Coverage Submetric (C).** Coverage measures the proportion of clusters that are non-empty:

$$C = \frac{\text{Number of Non-Empty Clusters}}{\text{Total Clusters}} \tag{7}$$

2. **Calculate Maximum Value Ratio (R).** This component captures the dominance of the maximum value relative to the total distribution of word counts. A higher R indicates a less balanced distribution:

$$R = \frac{\text{Maximum Value}}{\text{Sum of All Values}} \tag{8}$$

3. **Calculate Saturation (S).** Saturation measures how close the average cluster size is to the observed maximum value, reflecting the fullness of the clusters:

$$S = \frac{\text{Mean Value}}{\text{Maximum Value}} \tag{9}$$

4. **Maturity Score (M).** The maturity score combines these components into a weighted formula:

$$M = \alpha C + \beta(1 - R) + \gamma S \tag{10}$$

where

- α, β, γ: Weights assigned to each component, based on their importance in the present formulation.
- For the present analysis, fixed coefficients of $\alpha = 0.4, \beta = 0.3, \gamma = 0.3$ were used, placing slightly greater emphasis on cluster coverage while retaining contributions from dominance and saturation.

In plain terms, *maturity* measures how complete and balanced the derived cluster structure appears within the published corpus. A higher score indicates that evidence is spread across more non-empty clusters, that no single cluster dominates, and that the average cluster size is relatively close to the largest cluster. A lower score indicates a narrower or more uneven documentary profile. It should be interpreted as documented maturity within the corpus, not as direct evidence of implementation maturity or service quality.

For example, a government could score high on transparency if its website briefly mentions most success factors, such as trust, access, security, governance, and public awareness. But it might score lower on maturity if those mentions are shallow or concentrated in a small number of clusters.

Conversely, a country could have high maturity if its documents contain a rich and balanced set of themes, such as authentication, regulation, service access, technical standards, and assurance, but lower transparency if those themes do not map evenly onto the predefined success factors.

The maturity weights similarly represent a transparent baseline formulation rather than an empirically optimised weighting scheme. The slightly higher weight assigned to cluster coverage reflects the importance of breadth across the derived cluster structure, while the dominance and saturation terms capture balance. As with transparency, alternative weights could be used where a study has a validated external criterion or a policy reason for prioritising specific components. In the present article, the reported maturity scores should therefore be read as baseline artefact-based measures, with future work required to test robustness under alternative weighting assumptions.

Illustrative interpretation. The following simplified examples show how the three measures behave.

1. For *alignment*, suppose two countries have indicator-score vectors:

$$A = [4, 2, 0, 3], \qquad B = [2, 2, 0, 1].$$

Using Equation (1), the pairwise similarities are

$$\begin{aligned} S_1 &= 1 - \frac{|4-2|}{4} = 0.50, \\ S_2 &= 1 - \frac{|2-2|}{2} = 1.00, \\ S_3 &= 1.00 \quad \text{where } x_i = y_i = 0, \\ S_4 &= 1 - \frac{|3-1|}{3} = 0.33. \end{aligned}$$

The overall alignment score is therefore

$$A_{\text{country-pair}} = \frac{0.50 + 1.00 + 1.00 + 0.33}{4} \approx 0.71.$$

This indicates moderately high similarity in the documentary emphasis of the two countries.

2. For *transparency*, suppose a country has five possible success factors with mention counts:

$$[20, 10, 5, 0, 0].$$

Three of the five factors are mentioned, so coverage is

$$C = \frac{3}{5} = 0.60.$$

The non-zero mention proportions are

$$p = \left[\frac{20}{35}, \frac{10}{35}, \frac{5}{35}, 0, 0\right] = [0.571, 0.286, 0.143, 0, 0].$$

Using Equation (4), this gives

$$U = -\frac{1}{\log(5)}(0.571 \log(0.571) + 0.286 \log(0.286) + 0.143 \log(0.143)) \approx 0.59.$$

With equal weights, the transparency score is

$$T = 0.5(0.60) + 0.5(0.59) \approx 0.60.$$

This indicates partial but uneven coverage of the expected success factors.

3. For *maturity*, suppose five thematic clusters have the following sizes:

$$[30, 25, 20, 15, 10].$$

All five clusters are non-empty, so

$$C = \frac{5}{5} = 1.00.$$

The maximum value ratio is

$$R = \frac{30}{30 + 25 + 20 + 15 + 10} = \frac{30}{100} = 0.30.$$

The mean cluster size is

$$\bar{x} = \frac{100}{5} = 20,$$

and saturation is therefore

$$S = \frac{20}{30} \approx 0.67.$$

Using Equation (10),

$$M = 0.4(1.00) + 0.3(1 - 0.30) + 0.3(0.67) \approx 0.81.$$

This indicates a broad and relatively balanced thematic structure. By contrast, a more concentrated cluster distribution such as

$$[90, 10, 0, 0, 0]$$

would produce

$$C = \frac{2}{5} = 0.40, \qquad R = \frac{90}{100} = 0.90, \qquad S = \frac{20}{90} \approx 0.22,$$

and therefore

$$M = 0.4(0.40) + 0.3(1 - 0.90) + 0.3(0.22) \approx 0.26.$$

This lower score indicates a narrower documentary profile dominated by one cluster.

4.11. Tools and Software

The data mining pipeline was implemented using a combination of C# and Python (version 3.11). C# within Visual Studio was used for parts of the data collection and processing workflow, while Python and Jupyter Notebook (version 7.0) were used for corpus preparation, clustering, statistical analysis, and graph-oriented exploration of extracted relationships. The principal software components used in the study are summarised below.

1. **Visual Studio/C#** [82]: Visual Studio and C# were used to support the website data collection and pipeline orchestration components of the study, including data extraction and transformation steps prior to downstream NLP analysis.
2. **Jupyter Notebook** [83]:Jupyter Notebook was used as the main environment for exploratory analysis, iterative pipeline development, metric calculation, and result inspection.
3. **NumPy** [84]: NumPy was used for numerical operations on arrays and matrices, including the handling of intermediate numerical representations used in clustering and metric calculation.

4. **Pandas** [85]: Pandas was used for data cleaning, filtering, tabular transformation, and aggregation of corpus outputs into analysis-ready data structures.
5. **SpaCy** [86]: SpaCy was used for core natural language processing tasks, including tokenisation, part-of-speech tagging, and linguistic structuring of text for downstream analysis.
6. **Scikit-learn (sklearn)** [87]: Scikit-learn was used for TF–IDF vectorisation, K-Means clustering, and related analytical procedures used to identify patterns in the corpus.
7. **SciPy** [88]: SciPy supported numerical and statistical computations used in the analytical workflow.
8. **Gensim** [89]: Gensim was used for topic modelling, including the LDA-based analysis and associated topic inspection procedures.
9. **Natural Language Toolkit (NLTK)** [90]: NLTK was used in supplementary text-processing tasks, including token-level linguistic handling where required by the analysis workflow.
10. **Neo4j / Py2neo** [91]: Py2neo was used to interface with the Neo4j graph database, enabling storage and visual exploration of extracted subject–predicate–object relationships.
11. **OpenAI API**: The OpenAI API (GPT-4.1) was used in two bounded parts of the workflow: (1) associating high-frequency terms with the predefined success factors in Table 1, and (2) generating concise descriptive labels for final clusters from their top terms. These model-assisted steps supported interpretation and structured comparison, but did not determine the underlying clustering algorithms or the mathematical form of the reported metrics.

Where relevant, software versions, model settings, and prompt templates are reported in the accompanying repository and methodological documentation to support reproducibility.

5. Results and Discussion

This section presents the results of the artefact-based analysis of government-published digital identity materials. The results are organised in two stages. First, Sections 5.1–5.13 present country-level outputs, including word-frequency patterns, cluster summaries, and graph-based observations. These outputs show how digital identity is represented within each national corpus. Second, Sections 5.14–5.16 draw these outputs together through comparative measures of success-factor coverage, alignment, transparency, documented maturity, and the scientific contribution of analysing the public evidential layer of digital identity programmes. The purpose of the results section is therefore not only to describe national programmes individually, but also to identify cross-country patterns in how digital identity programmes are publicly articulated, evidenced, and documented.

Accordingly, the country-level sections report three linked outputs: top word frequencies mapped to the success-factor structure, POS-tagged graph visualisations, and K-Means cluster summaries. These outputs are retained in the main text because they provide the evidential trail from the mined national corpora to the subsequent cross-country comparison. They allow readers to inspect both representative patterns and exceptions before the results are aggregated into the success-factor heatmap and the alignment, transparency, and maturity measures.

These items contribute to constructing an aggregate of the identified terms cross-referenced against a list of success factors for comparison.

5.1. Denmark

NemID, introduced in 2010, was Denmark's electronic identification system, combining user IDs and passwords with a one-time password (OTP) generated from a physical key

card. It operated through a centralised server architecture secured by SSL/TLS encryption. As discussed in the published material, the system was eventually succeeded by a new approach that reduced reliance on physical key cards and responded to evolving security and usability requirements.

Table 3 highlights *MitID* as the most frequent term, appearing 3562 times. Developed as the successor to NemID, MitID is presented in the published material as a digitally oriented authentication solution that replaces the physical key-card model with app-based and alternative hardware-based authentication methods [92]. The material also describes the use of cryptographic controls, PKI, and broader integration capabilities, indicating that the documentary emphasis around MitID is centred on authentication, security, and service enablement.

Table 3. Denmark—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
mitid$^{(6)}$	3562	**service**$^{(5)}$	747	**phone**$^{(9)}$	438	**access**$^{(16)}$	360	**order**$^{(4)}$	305
data	2170	**legislation**$^{(14)}$	678	**strategy**$^{(2)}$	438	display	355	**process**$^{(10)}$	288
app$^{(5)}$	1280	impact	648	**user**$^{(5)}$	421	processing	350	**reader**$^{(3)}$	285
code	1227	digitisation	647	implementat.	413	eller	346	**chip**$^{(3)}$	280
business$^{(1)}$	1089	case	609	**policy**$^{(14)}$	412	**agency**$^{(12)}$	345	society	280
government$^{(2)}$	996	technology	589	**principle**$^{(2)}$	411	transition	342	skill	280
solution$^{(3)}$	959	de	532	man	401	initiative	337	audio	276
citizen$^{(6)}$	924	development	510	work	399	project	337	card	269
authority$^{(12)}$	891	intelligence	482	requirement	395	**passport**$^{(9)}$	328	country	268
sector$^{(1)}$	865	**system**$^{(15)}$	475	area	386	**protection**$^{(11)}$	317	level	262
security$^{(11)}$	855	bill	465	**rule**$^{(14)}$	382	part	315	password	256
information$^{(17)}$	802	time	455	number	365	et	313	**assessm.**$^{(11)}$	256

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Authentication Methods**—mitid, code, app, display, phone, passport, reader, chip, audio, card.
2. **Government Digital**—gov., digitisation, policy, agency, partnership, recommendation, area, initiative, procurement, value.
3. **Energy Trans.**—transition, energy, solution, potential, country, sector, scheme, technology, task, policy.
4. **Data Protection**—data, protection, authority, processing, information, process, infrastructure, solution, access, control.
5. **Business Technology**—business, sector, solution, technology, development, intelligence, skill, strategy, society, government.
6. **Civic Legislation**—citizen, impact, case, legislation, service, implement., rule, assessment, principle, time.
7. **App Usability**—mitid, app, man, code, user, et, eller, pin, screen, phone.
8. **Security Regulations**— security, business, information, authority, requirement, work, architecture, regulation, level, principle.

The word frequencies in Table 3 also suggest that the published Danish material spans several themes relevant to digital identity, including governance and policy (e.g., *principles*, *legislation*, and *policy*), authentication and delivery mechanisms (e.g., *chip*, *reader*, and *phone*), and associated use contexts such as *passport*. Within the present study, these terms are best interpreted as indicating the prominence of governance, implementation, and application-related themes in the published corpus, rather than as direct proof of programme maturity or completeness. The cluster analysis supports this interpretation by showing a diverse documentary profile across authentication, governance, security, usability, and technology-related themes.

Taken together, the clusters suggest that the Danish corpus reflects a broad documentary footprint spanning authentication mechanisms, public-sector digitalisation, data

protection, and regulatory structures. At the same time, the appearance of a distinct *Energy Transition* cluster indicates that the corpus is not limited exclusively to identity-specific material, and that some thematic spillover from adjacent areas of public-sector documentation may be present. Similarly, the prominence of security- and data-protection-related clusters indicates that these themes are clearly visible in the analysed material, but should not be interpreted on their own as definitive evidence that privacy by design is fully embedded in the programme. The POS-tagging graph in Figure 3 likewise shows multiple terms connected to MitID, suggesting a wide documentary range of applications and associations within the published corpus.

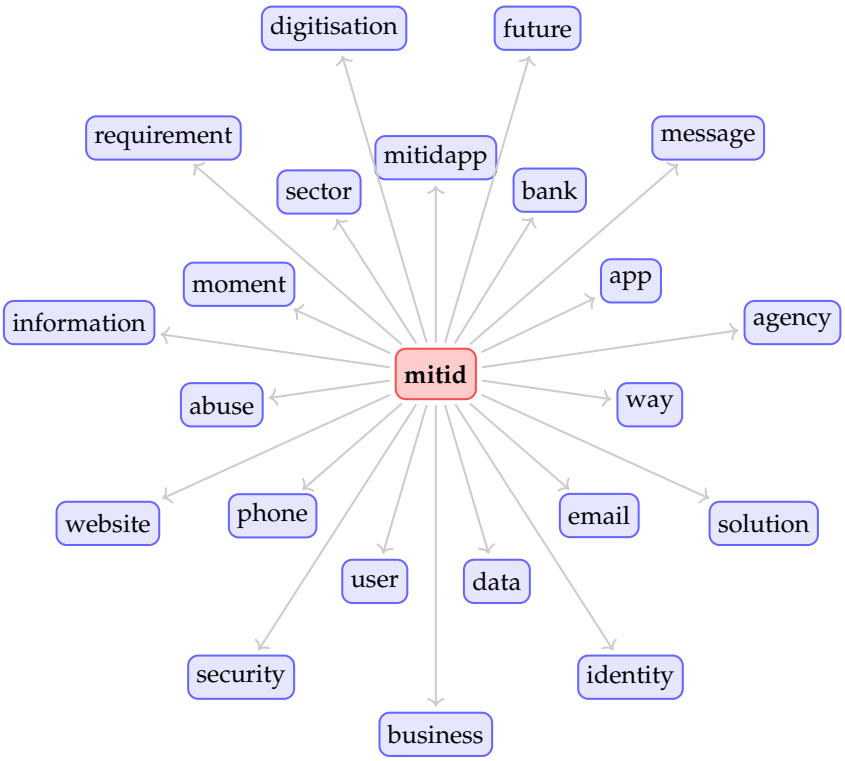

Figure 3. Denmark—MitID graph.

5.2. Finland

Suomi.fi e-Identification is Finland's national digital identity system, enabling secure electronic authentication for access to digital services [93]. Users are primarily authenticated using online banking credentials, mobile certificates, or smart cards. In the published material, the system is presented as an intermediary that verifies the user's identity through the chosen authentication method and transmits the authenticated identity to the service provider, with login details protected through encryption and related security controls.

Technically, Suomi.fi e-Identification is also described as interoperable with the eIDAS framework, allowing Finnish digital identities to be recognised across participating EU countries. In the documentary material, this positions the Finnish approach within a broader European interoperability context rather than as a purely domestic service mechanism.

Finland's primary digital identity graph is comparatively simple (Figure 4), and the word-frequency list in Table 4 gives prominence to terms such as identity, service, wallet, proof, and transaction. Within the present study, these patterns suggest that the published corpus places particular emphasis on identity-enabled transactions, verification mechanisms, and implementation-oriented aspects of digital identity. They do not, however,

by themselves establish that the programme is inherently simpler, cleaner, or more mature than those of other jurisdictions.

Table 4. Finland—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
identity$^{(6)}$	226	**mobile**$^{(9)}$	33	group	20	code	15	**architecture**$^{(15)}$	11
service$^{(5)}$	213	development	29	**user**$^{(5)}$	20	suomi	15	director	11
data	129	**foreigner**$^{(8)}$	27	organisation	18	actor	14	**system**$^{(15)}$	11
wallet$^{(4)}$	110	work	25	**security**$^{(11)}$	17	hacker	14	initiative	11
population	107	fi	25	**test**$^{(11)}$	17	session	13	**company**$^{(1)}$	11
agency$^{(12)}$	105	**ministry**$^{(14)}$	24	proposal	16	**device**$^{(9)}$	13	**team**$^{(2)}$	11
identification	69	**finance**$^{(4)}$	23	part	16	person	13	**participant**$^{(2)}$	11
solution$^{(3)}$	54	**transaction**$^{(4)}$	22	**future**$^{(13)}$	16	**country**$^{(3)}$	13	program	11
information$^{(17)}$	52	**specification**$^{(2)}$	22	consortium	16	implementat.	13	**commission**$^{(12)}$	10
project	46	**proof**$^{(11)}$	22	testing	16	**legislation**$^{(14)}$	12	progress	10
card	43	time	21	case	15	**police**$^{(11)}$	12	event	10
preparation	36	**parliament**$^{(14)}$	21	people	15	administration	12	way	10

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Identity Management**—identity, wallet, preparation, parliament, development, time, organisation, card, session, initiative.
2. **Administration**—solution, remote, case, ministry, finance, website, group, identification, representative, background.
3. **Identity Verification**—identity, card, service, solution, police, passport, test, finance, ministry, day.
4. **Mobile ID Solutions**—service, identification, mobile, fi, foreigner, transaction, identity, solution, device, alternative.
5. **Regulatory Proc.**—implement., wallet, regulation, market, proposal, legislation, project, government, decision, consultation.
6. **Digital Identity**—service, future, identity, code, student, idea, participant, architecture, comment, society.
7. **E-Governance**—service, data, agency, population, specification, wallet, solution, ministry, work, finance.
8. **Cybersecurity**—information, service, wallet, project, user, data, identity, security, identification, country.
9. **Software Testing**— testing, suomi, program, hacker, bug, bounty, vulnerability, potential, number, survey.

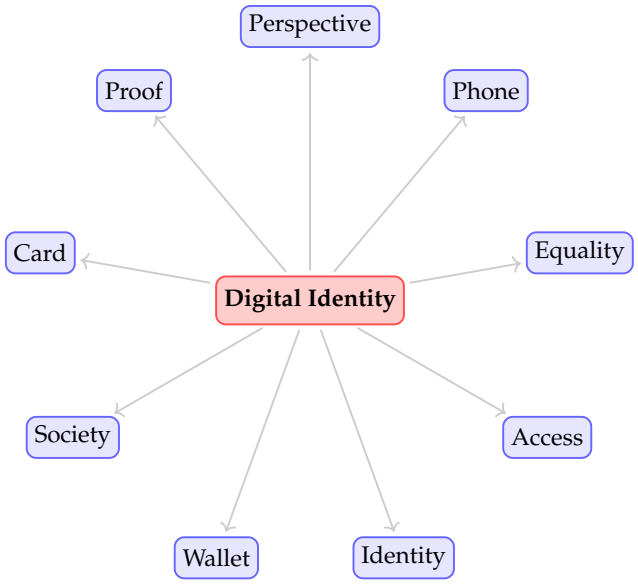

Figure 4. Finland-Digital Identity graph.

Cluster analysis indicates a diverse set of terms spanning identity management, administration, verification, mobile identity solutions, regulation, e-governance, cybersecurity, and software testing. Within the present framework, this suggests that the published Finnish corpus captures both policy and implementation-oriented aspects of digital identity, rather than reflecting only a single technical or legislative perspective.

POS-tagging analysis also indicates a distinct emphasis on implementation-related relationships, including connected terms associated with proof and transaction. This suggests that the published material gives visibility to operational and verification-oriented aspects of the ecosystem. A notable omission, however, is the term *privacy*, which does not appear in the top 60-word frequency list, and no clearly privacy-centred cluster is present in the final grouping. In addition, *security* appears only modestly in both the word and cluster analyses. These absences should be interpreted cautiously. They may reflect terminology choices, corpus composition, source selection, or the distribution of relevant material across other documents rather than an underlying lack of privacy or security provisions. Given Finland's stated alignment with eIDAS, it is plausible that some privacy- and security-related requirements are embedded in regulatory or technical materials not strongly represented in the analysed corpus, but this remains an inference rather than a direct finding from the text-mining results.

5.3. South Korea

PASS, launched by the Korea Financial Telecommunications and Clearings Institute (KFTC), is South Korea's most prominent mobile-based digital authentication system [94]. Utilised primarily for financial transactions, PASS offers multimodal authentication methods, including biometrics and PINs, while ensuring robust encryption and high-level security. In addition, it boasts of interoperability, allowing users to authenticate across multiple banks and services seamlessly.

Alongside PASS, there is growing interest in leveraging blockchain for decentralised digital identity solutions in South Korea, emphasising user-controlled personal data management.

In our analysis (see Table 5), the word "government" is most present in the word frequency analysis with additional governance terms such as ministry, policy, citizen, welfare, and agency. This theme is carried out through cluster analysis, where all clusters contain some form of government function supported by digital identity technology.

POS-tagging provides a visual representation of this in Figure 5, indicating the relationship between citizens and government and the services and use cases relating to identity. Communication with the public regarding security and privacy measures related to identity is minimal.

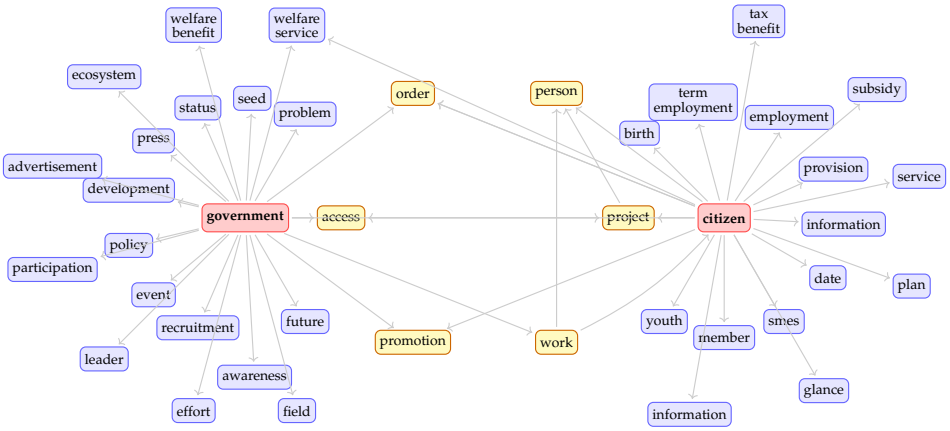

Figure 5. Korea—government/citizen graph.

Table 5. South Korea—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
government$^{(2)}$	563	**user**$^{(5)}$	84	**citizen**$^{(6)}$	53	year	41	content	36
service$^{(5)}$	513	official	74	**cooperation**$^{(2)}$	51	layer	41	website	36
system$^{(15)}$	213	**future**$^{(13)}$	72	center	49	**test**$^{(11)}$	40	component	36
information$^{(17)}$	206	**institution**$^{(1)}$	71	**welfare**$^{(3)}$	49	**process**$^{(10)}$	39	**business**$^{(1)}$	36
data	173	management	70	function	49	**standard**$^{(2)}$	39	**device**$^{(9)}$	35
ministry	154	document	69	time	47	**expansion**$^{(3)}$	38	practice	34
mobile$^{(9)}$	101	project	67	model	47	tool	38	program	34
development	94	platform	65	analysis	47	work	37	**identity**$^{(6)}$	32
safety$^{(11)}$	92	**sector**$^{(1)}$	64	status	44	**agency**$^{(12)}$	37	processing	32
online$^{(9)}$	89	**environment**$^{(17)}$	61	plan	43	technology	37	**security**$^{(11)}$	32
certificate$^{(11)}$	85	**access**$^{(16)}$	57	authenticat.	42	**history**$^{(7)}$	37	issue	32
policy$^{(2)}$	84	experience	57	**innovation**$^{(1)}$	42	type	37	**framework**$^{(12)}$	32

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Public Administration**—government, service, policy, official, experience, relationship, project, development, subsidy, year.
2. **Data Management**— layer, management, function, processing, data, file, reservation, type, configuration, business.
3. **Government Services**—information, data, service, government, citizen, management, operation, time, transport, agency.
4. **Digital Transformation**—government, future, service, online, history, expansion, content, innovation, status, map.
5. **Government Documentation**—ministry, safety, document, information, analysis, service, institution, model, government, data.
6. **Systems Development**—environment, mobile, service, framework, standard, component, runtime, job, development, egovernment.
7. **Social Services**—mobile, service, welfare, cooperation, device, card, index, nation, republic, plan.
8. **Digital Development**—government, test, platform, tool, practice, case, development, initiative, plan, field.
9. **Identity Verification**—service, user, certificate, authentication, sector, signature, access, institution, data, information.

5.4. New Zealand

New Zealand's digital identity framework, led by the Department of Internal Affairs, is presented as an effort to support secure and streamlined online identity verification across the public and private sectors. In the published material, this initiative is framed as part of the country's wider digital strategy, with emphasis on privacy, trust, interoperability, and the development of standards for digital identity services [95].

The framework places particular emphasis on trust, setting out rules and accreditation processes intended to support reliable digital identity services and the protection of personal data. The published policy material also indicates an intention to facilitate interoperability with international trust frameworks, including those of Australia, Canada, and the United Kingdom. Within this framing, the ecosystem encompasses a range of providers and participants, including individuals, businesses, and government entities, all of whom are positioned as potential beneficiaries of accreditation and greater trust in digital transactions.

The government has identified the following benefits of the trust framework [95]:

1. Enables individuals and organisations to safely and more easily transact on behalf of others.
2. Ensures that personal data are private and secure, enabling growth in the digital economy because citizens and businesses have confidence in the digital marketplace.
3. Improves protection against cyber threats and invasion of privacy.

4. This reduces compliance costs for citizens, businesses, and the government by locating compliance obligations in one place and developing regulations and policies with all parties in mind.

New Zealand initiated the formal process for a Digital Identity Trust Framework with the introduction of the Digital Identity Services Trust Framework Bill to Parliament in late 2021. In the published material, this legislative step is presented as establishing a regulated environment for digital identity services, with particular emphasis on trust, security, accreditation, and interoperability.

Within the analysed corpus, *framework* and *identity* are the two most prevalent terms across the sampled websites, with frequencies of 1359 and 1158 respectively (Table 6). Other prominent terms, including *government*, *ecosystem*, *bill*, *cabinet*, *governance*, *policy*, *legislation*, and *compliance*, indicate that the published material is strongly oriented toward regulatory design, institutional arrangements, and framework development. The prominence of *privacy* and the presence of *security* similarly suggest that these themes are given substantial visibility within the documentary corpus, although this should be interpreted as evidence of published emphasis rather than direct proof of implementation characteristics such as privacy by design.

Table 6. New Zealand—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
framework$^{(12)}$	1359	cabinet	242	people	217	**risk**$^{(11)}$	155	**user**$^{(5)}$	117
identity$^{(6)}$	1158	bill	241	**legislation**$^{(14)}$	206	data	153	establishment	116
service$^{(5)}$	894	**department**$^{(14)}$	241	**authority**$^{(12)}$	204	entity	151	official	116
information$^{(17)}$	607	**cost**$^{(13)}$	239	requirement	193	issue	145	power	108
government$^{(2)}$	606	**governance**$^{(14)}$	238	minister	187	**access**$^{(16)}$	141	**principle**$^{(2)}$	107
option	410	**privacy**$^{(11)}$	233	**security**$^{(11)}$	183	**policy**$^{(2)}$	134	**realme**$^{(6)}$	106
accreditation$^{(1)}$	386	board	230	**benefit**$^{(13)}$	177	**resolution**$^{(12)}$	131	**participation**$^{(2)}$	104
rule$^{(14)}$	347	impact	228	**provider**$^{(1)}$	176	work	129	purpose	104
ecosystem$^{(3)}$	323	**sector**$^{(1)}$	227	potential	164	group	128	affair	103
participant$^{(2)}$	322	**process**$^{(10)}$	225	**agency**$^{(12)}$	163	dispute	123	**economy**$^{(13)}$	103
standard$^{(2)}$	255	development	224	statement	162	regime	122	paper	102
system$^{(15)}$	251	**compliance**$^{(14)}$	224	organisation	159	testing	122	phase	99

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Identity Ecosystem**—identity, service, ecosystem, information, people, government, provider, benefit, sector, framework.
2. **Framework Principles**— framework, information, principle, purpose, authority, dev., cabinet, official, individ., page.
3. **Governmental Framework**—government, option, framework, cabinet, group, iwi, information, agency, status, chief.
4. **Accreditation Framework**—framework, accreditation, rule, participant, establishment, regime, potential, offence, mechanism, enforcement.
5. **Governance Impact**—impact, statement, cost, framework, board, governance, accred., template, government, standard.
6. **Data Security**—privacy, security, standard, data, identity, framework, information, service, rule, management.
7. **Dispute Resolution**—option, resolution, dispute, participant, process, value, framework, scheme, paper, user.
8. **Risk Assurance**—ministry, risk, department, level, information, assurance, assessment, requirement, business, confidence.
9. **Government Services**—service, framework, gov., minister, compliance, testing, board, phase, department, legislation.

POS-tagging visualisations (Figure 6) and cluster analysis reinforce the strongly governance- and regulation-oriented profile of the New Zealand corpus, while also identifying related emphases such as dispute resolution, assurance, risk management, and service delivery.

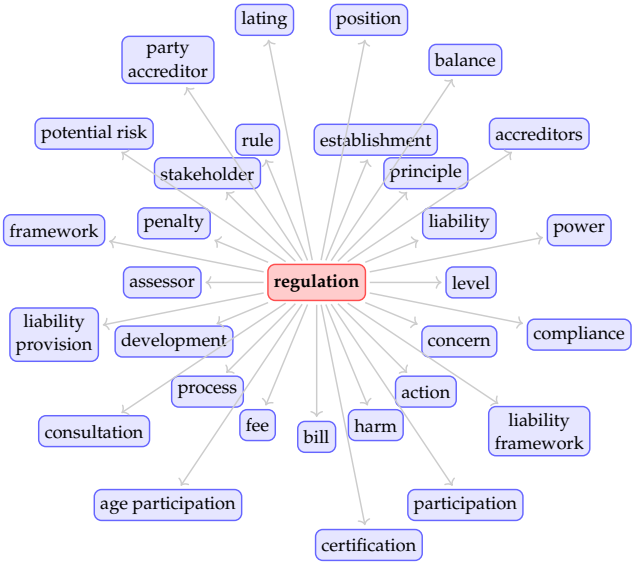

Figure 6. New Zealand—regulation graph.

Overall, the analysed material shows a strong focus on governance, accreditation, legislative design, privacy, and service regulation. Within the scope of this study, the New Zealand corpus therefore appears to emphasise the institutional and regulatory architecture of digital identity more strongly than downstream implementation or technical infrastructure. This profile is consistent with the published orientation of New Zealand's trust framework, while remaining subject to the general limitations of corpus composition and document scope discussed elsewhere in this paper.

5.5. Sweden

BankID, developed collaboratively by major Swedish banks, is Sweden's predominant electronic identification system. It supports digital authentication and transaction signing and is available in several forms, including computer files, smart cards, and, most prominently, the Mobile BankID application [96].

The published material describes BankID as relying on encryption and user authentication through personal security codes within the application. Given its wide acceptance by banks, government agencies, and other platforms, BankID occupies a central position in Sweden's digital authentication environment [97,98].

Sweden's eID implementation follows the eIDAS guidelines, and Sweden participates as a node in the eIDAS network. Because both BankID and eIDAS are externally defined frameworks, the primary website material mined in this study was heavily linked to the Sweden Connect technical framework. As a result, the Swedish corpus is more technical in character than several of the other national corpora analysed in this paper.

Consequently, the word frequencies and clusters are predominantly technical, and the outputs should be interpreted in that context. The corpus gives strong visibility to protocol, metadata, schema, certificate, signature, service-provider, and message-exchange concepts. Two of the identified clusters are explicitly security-oriented, indicating that security-related themes are prominent in the published technical material. The remaining clusters suggest substantial documentary emphasis on integration, interoperability, and message or service architecture, with protocols, service bindings, service messages, and data models all appearing in the mined text.

Within the analysed corpus, Table 7 reflects a highly technical documentary profile, with strong emphasis on *identity*, *service*, *provider*, *assertion*, *request*, *protocol*, *metadata*, *certificate*, and *security*. This suggests that the Swedish material captured in the study is oriented

more toward technical implementation frameworks and interoperability specifications than toward broader policy or public-facing programme communication. Accordingly, the results are better interpreted as evidence of the technical and standards-oriented emphasis of the published corpus than as a direct measure of overall programme maturity.

Although the term *proof* is not present, the prominence of terms such as *user*, *profile*, *signature*, *certificate*, and *person* suggests that the published material addresses multiple aspects of identity-related interaction beyond simple identification alone. At the same time, because the corpus is heavily technical and source-dependent, these findings should be understood as reflecting the character of the documentation captured for analysis rather than the full scope of Sweden's digital identity programme.

Table 7. Sweden—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
element	2792	party	840	content	564	**principal**$^{(12)}$	445	property	357
provider$^{(1)}$	2010	section	828	requirement	549	speci	441	format	357
service$^{(5)}$	1841	authentication	818	**metadata**$^{(11)}$	543	**authority**$^{(12)}$	441	artifact	349
message	1770	entity	815	**certificate**$^{(4)}$	525	session	433	profile	348
assertion	1606	specification	779	**security**$^{(11)}$	521	time	414	requester	341
identity$^{(6)}$	1546	attribute	736	version	520	extension	414	ion	336
request	1535	elem	721	**rule**$^{(14)}$	518	http	410	ing	326
value	1419	**information**$^{(17)}$	709	**schema**$^{(15)}$	506	oasis	408	namespace	323
signature	1329	data	706	document	485	identifier	394	**soap**$^{(15)}$	308
protocol	1297	**user**$^{(5)}$	647	case	467	processing	392	statement	307
type	1291	name	634	number	455	rtion	371	issue	302
xml	889	page	614	context	449	**system**$^{(15)}$	358	person	301

word $^{(successfactor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Identity Data**—person, number, information, address, member, condition, service, time, security, registration.
2. **Data Components**—type, section, element, data, value, extension, message, format, rule, endpoint.
3. **Security Elements**—party, assertion, certificate, value, information, page, code, status, framework, eid.
4. **XML Elements**—signature, xml, element, schema, namespace, value, property, data, document, object.
5. **Communication**—request, message, protocol, session, authority, statement, responder, http, authentication, requester.
6. **Protocol**—protocol, context, version, assertion, artifact, processing, message, speci, authentication, specification.
7. **Security Requirements**—security, oasis, specification, requirement, level, assertion, soap, time, profile, right.
8. **Entity Identifier**—element, elem, content, identifier, type, identification, specifies, entity, article, assertion.
9. **Service Provider**—provider, service, identity, user, entity, metadata, authentication, request, element, agent.

5.6. Iceland

Island.is is the official web portal for Icelandic government services. It provides access to information and services from multiple government agencies and functions as a centralised portal for a wide range of online public services, including social security benefits, licence renewal, tax filing, and business registration. In the published material, the platform is presented as part of Iceland's broader effort to simplify interactions with government agencies, improve administrative efficiency, and increase accessibility of public services [99].

Slykill (which translates to "password" or "key" in Icelandic) forms part of the authentication system used to access *Island.is* services securely. Managed by Iceland's Directorate of Information Technology and e-government, it is presented in the published material as a primary electronic authentication solution. Users are authenticated through a combination of usernames, passwords, and an additional layer of security via one-time passwords (OTPs) delivered through SMS or a dedicated OTP generator. As described in the source material, *Slykill* integrates with multiple online platforms across the public and private

sectors and is supported by encryption and security controls intended to protect data integrity and service access [100].

The published corpus associated with Icelandic digital identity-related services is extensive and is the largest national strata set in this research. As a result, the Icelandic material provides substantial visibility into the surrounding service environment in which digital identity operates. At the same time, the breadth of the source material means that the corpus appears to capture a wider administrative and regulatory domain than digital identity alone. Accordingly, the Icelandic corpus offers insight not only into authentication and service access, but also into the wider documentary context of government service delivery.

Word-frequency analysis reflects this breadth, with prominent terms including *health*, *child*, *disease*, *insurance*, and *family*, alongside administrative and regulatory terms. These patterns suggest that the published material places considerable emphasis on civic and service-oriented applications. However, they also indicate that the Icelandic corpus spans multiple service and regulatory contexts, which should be taken into account when interpreting the results.

POS-tagging of the Iceland data also revealed several notable term relationships, as shown in Figure 7. These relationships are exploratory in nature and suggest avenues for further investigation into how identity-related concepts are embedded within Iceland's wider public-service documentation.

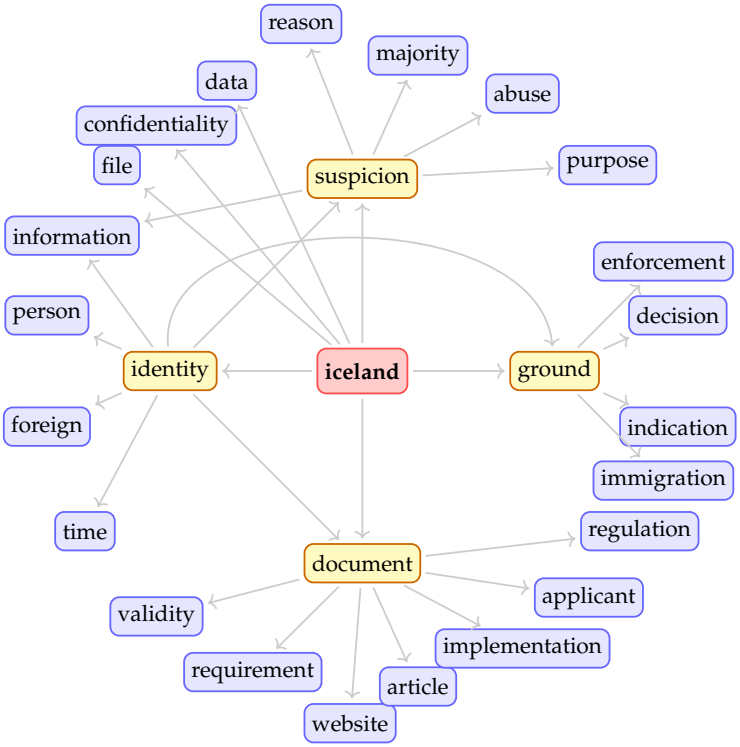

Figure 7. Iceland—identity graph.

The clustering results reinforce the impression that the Icelandic corpus extends well beyond narrowly defined identity architecture and includes substantial regulatory, immigration, workplace, and maritime material. Terms such as *ship*, *cargo*, *sea*, and *water* are especially prominent in several clusters, as shown in Table 8.

These cluster patterns suggest that the Icelandic corpus reflects a broad administrative and regulatory documentary environment in which digital identity-related services are situated, rather than a tightly bounded identity-only corpus. The prominence of maritime, immigration, health, and regulatory themes therefore indicates that the analytical outputs

are influenced not only by identity-programme content, but also by corpus scope and document genre.

Table 8. Iceland—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
iceland	18,110	**protection**$^{(11)}$	8439	**cargo**$^{(3)}$	6112	**family**$^{(3)}$	4468	**disease**$^{(3)}$	3975
residence$^{(8)}$	15,811	article	8209	**document**$^{(8)}$	6012	month	4420	**device**$^{(9)}$	3933
health$^{(11)}$	14,671	var	8180	**immigration**$^{(16)}$	5957	amendment	4412	einnig	3856
fyrir	13,694	hefur	7768	time	5883	number	4397	minister	3789
information$^{(17)}$	11,064	**system**$^{(15)}$	7621	**convention**$^{(2)}$	5756	**part**$^{(3)}$	4397	equipment	3785
case	10,315	**person**$^{(16)}$	7519	work	5672	period	4288	**insurance**$^{(3)}$	3714
ship	10,048	applicant	6959	condition	5623	paragraph	4258	**national**$^{(7)}$	3701
provision$^{(3)}$	9277	**authority**$^{(12)}$	6893	**certificate**$^{(4)}$	5291	fram	4185	board	3598
child	9275	**requirement**$^{(10)}$	6661	state	5094	code	4125	**risk**$^{(11)}$	3520
service$^{(5)}$	9067	**safety**$^{(11)}$	6408	decision	4795	**age**$^{(3)}$	4119	home	3501
year	8635	data	6274	member	4732	area	4052	**passenger**$^{(3)}$	3499
regulation$^{(14)}$	8466	**country**$^{(7)}$	6239	space	4623	**procedure**$^{(10)}$	4030	date	3452

word $^{(successfactor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Immigration Terms**—residence, iceland, case, child, applicant, immigration, year, protection, provision, document.
2. **Workplace Procedures**—service, training, assessment, risk, health, device, level, safety, management, work.
3. **Authority Entities**—health, iceland, service, country, information, authority, insurance, state, licence, arrival.
4. **Maritime Terms**—code, cargo, gas, test, bulk, ship, tank, requirement, death, infection.
5. **Residential Care**—group, status, nursing, home, par, ed, refugee, device, study, resident.
6. **Seafaring Terms**—ship, space, cargo, water, craft, deck, solution, area, control, passenger.
7. **Data Regulation**—data, information, article, state, provision, regulation, processing, authority, disease, treatment.
8. **Maritime Regulations**—convention, safety, amendment, ship, regulation, sea, life, article, certificate, organization.

Although some conventional identity-related terms such as *privacy* and *identity* are not prominent in Table 8, the corpus does show substantial emphasis on service delivery, regulation, authority, certificates, safety, and protection. These patterns suggest a pragmatic, service-oriented documentary footprint, but they should not be taken as a complete account of Iceland's digital identity architecture or governance arrangements. Rather, the Icelandic results illustrate both the value and the limitation of the present approach: a large and information-rich corpus can reveal extensive surrounding service context, while also introducing documentary noise from broader administrative domains. Further work using tighter source delimitation would help distinguish core digital identity material from adjacent service and regulatory content.

5.7. Australia

Australia's Trusted Digital Identity Framework (TDIF) is presented in the published material as a structured national policy framework for digital identity verification. It places emphasis on privacy, security, accreditation, interoperability, and data integrity across both government and commercial contexts, and is framed as part of Australia's broader effort to support a secure and trustworthy digital ecosystem [101].

At the operational level, myGovID, overseen by the Australian Digital Transformation Agency, is described as a principal mechanism for digital identity verification within the Australian government context. In the published material, it is presented as enabling individuals to verify their identities through a mobile application using accredited Australian identity documents, while integrating with wider government service architecture. The same material also emphasises privacy-related design choices, including limits on unnecessary data collection and handling of user information within the broader TDIF framework.

Critical commentary has, however, questioned the extent to which Australia's current approach reflects more recent developments in digital identity. One digital identity expert has argued that myGovID relies on comparatively dated technological assumptions and lacks some of the more advanced security and wallet-based capabilities increasingly associated with private-sector identity ecosystems [102]. The related academic literature has also raised concerns about security limitations within the TDIF framework [103,104]. These sources provide useful context for interpreting the documentary profile of the Australian corpus, particularly where governance and compliance language may be more visible than detailed technical implementation material.

With these perspectives in mind, the Australian corpus can be read as a strongly governance- and compliance-oriented documentary set. Cluster analysis indicates substantial emphasis on security, data protection, identity management, risk management, governance, and documentation, suggesting that the published material devotes considerable attention to the institutional and regulatory architecture of the TDIF framework.

Word-frequency analysis (Table 9) likewise shows a broad set of governance, privacy, accreditation, and service-related terms. User- and ecosystem-oriented terms such as *identity*, *user*, *party*, *person*, *provider*, and *consumer* are all present, while governance-related terms such as *government*, *legislation*, *oversight*, *risk*, *law*, and *policy* are also highly ranked. Within the present framework, this suggests that the published Australian material gives strong visibility to institutional design, compliance, accreditation, and user protection.

Table 9. Australia—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
identity$^{(6)}$	12,158	applicant	2714	**fraud**$^{(11)}$	1688	level	1302	management	978
system$^{(15)}$	5794	data	2504	applicability	1675	**participant**$^{(2)}$	1263	control	973
entity	5260	**rule**$^{(14)}$	2496	**oversight**$^{(12)}$	1672	consultation	1170	official	959
information$^{(17)}$	5227	user	2469	attribute	1668	**role**$^{(2)}$	1129	purpose	941
requirement$^{(10)}$	4926	party	2445	part	1630	individual	1118	impact	917
service$^{(5)}$	4488	assessment	2198	bill	1580	**incident**$^{(11)}$	1104	**law**$^{(14)}$	914
privacy$^{(11)}$	4390	document	2191	department	1533	**standard**$^{(2)}$	1080	paper	859
provider$^{(1)}$	3256	section	2165	credential	1498	testing	1072	**protection**$^{(11)}$	852
accreditation$^{(1)}$	3119	**finance**$^{(4)}$	2119	**access**$^{(16)}$	1450	**verification**$^{(10)}$	1027	name	841
security$^{(11)}$	3048	**risk**$^{(11)}$	2070	authenticat.	1423	time	998	person	809
government$^{(2)}$	2908	**process**$^{(10)}$	1834	**framework**$^{(12)}$	1360	**policy**$^{(2)}$	991	number	800
legislation$^{(14)}$	2730	**authority**$^{(12)}$	1734	request	1334	**business**$^{(1)}$	978	**consumer**$^{(5)}$	791

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Digital Identity Management**—identity, service, provider, party, user, attribute, government, request, legislation, level.
2. **Data Security**—information, identity, government, data, legislation, user, document, security, rule, applicant.
3. **Governance Framework**—oversight, authority, entity, identity, legislation, accreditation, power, function, information, participant.
4. **Compliance Requirements**—entity, information, rule, accreditation, government, identity, service, incident, requirement, security.
5. **Risk Management**—risk, security, assessment, fraud, applicant, identity, management, rating, information, finance.
6. **Privacy Protection**—privacy, information, protection, legislation, impact, safeguard, requirement, assessment, data, policy.
7. **Process Accreditation**—requirement, finance, accreditation, department, applicant, assessment, role, process, official, provider.
8. **Documentation Standards**—section, requirement, chapter, exposure, identity, information, rule, accreditation, paper, entity.

Australia also records one of the highest occurrences of the term *privacy*, ranked seventh with 4390 references, while *security* appears 3048 times. These figures suggest that privacy- and security-related themes are highly visible within the public documentary footprint of the framework. At the same time, the relative prominence of *credential* com-

pared with *authentication* is better interpreted as indicating a documentary emphasis on credential-related language than as evidence, on its own, of a more advanced technical model. Similarly, the absence of terms such as *encryption* and *certificate* from the most prominent terms may indicate that detailed technical implementation material is less visible in the sampled corpus than governance, compliance, and framework documentation.

Overall, the Australian corpus suggests a documentary profile centred on governance, accreditation, privacy, security, and risk management. Within the scope of this study, this indicates that Australia's published digital identity material places substantial emphasis on the policy and assurance foundations of the TDIF framework, while leaving some of the more detailed technical implementation issues less visible in the public corpus.

5.8. Estonia

Estonia's digital identity infrastructure revolves around the Estonian ID Card, a mandatory identity document embedded with a chip containing two certificate pairs for authentication and digital signatures. This card is a central component of Estonia's e-governance model, supporting legally binding digital signatures and PIN-secured access to a wide range of e-services, including the nation's i-voting system. Complementing the ID Card, Estonia has also introduced Mobile-ID and Smart-ID; the former enables mobile devices to function as digital identity tools, while the latter provides an app-based two-factor authentication solution that removes the need for a card reader and expands digital access [105,106].

Published descriptions of Estonia's digital identity system indicate that the Estonian ID Card, Mobile-ID, and Smart-ID are supported by the secure management of digital credentials. Two pairs of digital certificates and their respective private keys are embedded within the ID Card chip and are used for authentication and digital signing [107]. In the published material, these mechanisms are presented as part of a broader infrastructure for authentication, authorisation, and controlled access to digital services. The same material also refers to mechanisms for the renewal and revocation of credentials, reflecting an emphasis on continuity and trust within the digital identity ecosystem.

In this study, the extracted dataset for Estonia was more limited in size than initially expected. This was partly because the source, *e-estonia.com*, made extensive use of video content, which was not amenable to the text-mining methodology employed here. As a result, the analysed corpus should be understood as a partial documentary representation of Estonia's digital identity programme rather than an exhaustive account.

Within the corpus analysed here, the Estonian material reflects a broad documentary spread across infrastructure, administration, service access, civic information, and governance-related themes. At the same time, the top-word analysis in Table 10 shows that the published material also gives prominence to practical and sector-linked terms such as *school*, *transport*, *safety*, *education*, and *passenger*. This suggests that the documentary footprint captured in the corpus extends beyond core identity architecture into broader service and administrative contexts. Accordingly, the clustering results should be interpreted as reflecting the composition and emphasis of the analysed material, rather than as a complete representation of Estonia's digital identity programme.

There are also notable omissions in the analysed material, including relatively limited explicit coverage of credentials, data privacy, and specific security implementations. These absences should be interpreted cautiously. They may reflect source selection, publication style, the video-heavy nature of the source material, or the fact that some technical or regulatory detail is documented elsewhere rather than in the corpus captured for this study. Given Estonia's recognised maturity in other digital government contexts, these results are better understood as indicating limits in the documentary corpus available to this analysis

than as evidence of weakness in the underlying programme. Further research using a broader and more text-rich source base would help clarify this distinction [14].

Table 10. Estonia—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
data	244	embassy	72	tool	54	**passenger**$^{(3)}$	46	newsletter	36
service$^{(5)}$	135	event	72	**mobile**$^{(9)}$	54	niis	45	community	36
system$^{(15)}$	135	solution	70	**safety**$^{(11)}$	53	**provider**$^{(1)}$	45	news	36
government$^{(2)}$	122	infrastructure	69	**education**$^{(17)}$	53	hour	45	trend	36
centre	117	record	68	material	53	ease	40	medium	36
mobility	95	state	66	**information**$^{(17)}$	52	host	40	practice	36
online	91	**security**$^{(11)}$	60	**freight**$^{(3)}$	50	**institution**$^{(1)}$	40	link	36
city	78	management	59	skill	50	expert	40	**airport**$^{(3)}$	36
school	74	**company**$^{(1)}$	59	interoperab.	49	ground	40	floor	36
minute	74	transport	58	technology	49	field	39	statue	36
digitalisation	73	**business**$^{(1)}$	58	transportat.	48	plan	38	country	35
contact	73	**registry**$^{(8)}$	56	exercise	46	**access**$^{(16)}$	38	**tax**$^{(3)}$	35

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Digital Identity**—security, estonian, identity, tax, infrastructure, investment, ecosystem, model, data, component.
2. **Education**—school, technology, card, teaching, material, education, management, world, tool, skill.
3. **Administration**—centre, minute, city, contact, hour, airport, ecosystem, sector, identity, government.
4. **Online Service**—ground, floor, statue, service, online, year, day, road, hour, country.
5. **Civic Information**—information, data, citizen, doctor, interoperability, service, society, organisation, connection, state.
6. **System Management**—data, government, mobility, service, event, embassy, safety, registry, freight, transportation.
7. **Digital Governance**—state, plan, provider, expert, digitalisation, practice, link, service, tax, file.
8. **Online Accessibility**—access, vehicle, data, information, infrastructure, people, online, patient, bus, world.
9. **Digital Data Management**—record, online, data, patient, access, time, business, country, signature, platform.

5.9. United Kingdom

The UK's GOV.UK Verify was the initial government-backed digital identity solution designed to enable citizens to authenticate their identities online when accessing government services. Over time, however, its role diminished as the government moved away from continued funding and the wider UK digital identity landscape shifted toward a more private-sector-driven model [108].

In parallel, the UK introduced a digital identity and attributes trust framework to guide this evolving ecosystem. The published framework sets out technical standards, procedural requirements, and legal expectations for digital identity providers, with the aim of supporting secure, interoperable, and trustworthy identity services while encouraging private-sector participation and innovation [109].

Within the corpus analysed in this study (Table 11), the most frequent term is *user*, with 3222 occurrences. Other highly ranked terms include *service*, *framework*, *identity*, *access*, *requirement*, *security*, and *provider*. Taken together, these terms suggest that the published UK material places strong emphasis on user interaction, service access, framework design, compliance, and technical or organisational requirements. Although *privacy* does not appear in the top 60 terms, the wider corpus includes sufficient privacy-related language, including terms such as *ciphertext*, *privacy*, and *protection*, to support the formation of a distinct data-privacy cluster.

The range of clusters identified also indicates that the UK corpus spans more than identity policy alone. Alongside clusters focused on security, identity management, and data privacy, the corpus also includes substantial material relating to accessibility, user interfaces, web development, and content compliance. This suggests that the documen-

tary footprint captured for the UK reflects a broad public-service and standards-oriented environment in which digital identity is embedded.

Table 11. United Kingdom—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
user$^{(9)}$	3222	control	867	**assessment**$^{(11)}$	656	scheme	439	purpose	366
data	2078	web	845	attribute	588	communicat.	439	algorithm	360
service$^{(5)}$	1756	clause	810	**security**$^{(11)}$	571	inspection	436	case	359
document	1519	text	782	**system**$^{(15)}$	570	**provider**$^{(1)}$	427	element	351
software	1496	organisation	769	criterion	529	speech	423	**rule**$^{(14)}$	342
requirement$^{(10)}$	1413	audio	750	output	516	**signat.**$^{(4)}$	422	mechanism	334
access$^{(16)}$	1403	part	730	**policy**$^{(2)}$	514	video	421	operation	334
information$^{(17)}$	1393	**accessibility**$^{(3)}$	730	**device**$^{(9)}$	491	platform	414	input	330
identity$^{(6)}$	1160	type	726	etsi	491	product	394	**protection**$^{(11)}$	324
interface	1125	**framework**$^{(12)}$	720	time	479	functionality	393	alternative	319
technology	984	success	719	**standard**$^{(2)}$	457	network	391	**voice**$^{(9)}$	308
content	928	page	672	table	447	**process**$^{(10)}$	390	screen	297

word $^{(successfactor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Security**—access, control, policy, data, service, space, solution, device, user, case.
2. **Identity Management**—service, identity, user, type, assessment, requirement, framework, inspection, organisation, text.
3. **Content Management**—document, success, criterion, mobile, form, table, content, audio, requirement, medium.
4. **User Interface**—software, interface, user, technology, screen, platform, element, success, criterion, table.
5. **Compliance**—clause, requirement, performance, statement, user, conformance, functionality, content, interface, document.
6. **Web Development**—page, web, document, form, requirement, audio, content, software, conformance, error.
7. **Accessibility**—information, output, speech, user, content, accessibility, audio, text, change, service.
8. **Data Privacy**—data, protection, device, organisation, identity, ciphertext, processing, purpose, process, privacy.

Within the present framework, the UK results are therefore best interpreted as indicating a documentary profile that combines digital identity governance with broader concerns around service usability, accessibility, compliance, and technical implementation, making the UK corpus comparatively rich.

5.10. United Arab Emirates

UAE PASS is the cornerstone of the United Arab Emirates' digital identity and signature infrastructure, playing a pivotal role in its Smart Government strategy. Utilising the Emirates ID as a foundation for enrolment, this system grants individuals secure entry into a wide array of governmental services, enabling seamless transactions, electronic signing of documents, and verification processes. The security framework of the UAE PASS incorporates a combination of Personal Identification Numbers (PINs), biometric verification, and Quick Response (QR) codes [110,111].

Although it is primarily integrated with government agencies, there has been a concerted effort to extend its application to the private sector. This expansion aims to establish an all-encompassing digital identity network throughout the UAE.

Our research shows that government is the most significant term in the mined UAE corpus (see Table 12). This is supported by related terms such as policy and law, and the government constitutes a number of the analysed word clusters:

Future-centric words such as transformation, innovation, and the future are also present; however, modern digital identity terms such as identity, credentials, and verification are missing.

The terms were predominantly consumer-centric; therefore, technical terms were missing. The published content and corresponding word sets appear to focus on digital onboarding.

Table 12. United Arab Emirates—top word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
government(2)	719	business	117	**sector**(1)	84	emirate	60	programme	51
service(5)	653	**education**(17)	112	**environment**(17)	82	people	59	**team**(2)	50
data	403	law	104	transformation	78	**authority**(12)	58	**access**(16)	49
information(17)	238	technology	100	**innovation**(1)	74	statistic	58	accuracy	48
policy(14)	175	job	100	**future**(13)	72	site	57	city	48
entity	171	home	97	**strategy**(2)	72	right	56	society	47
participation	156	development	95	**system**(15)	70	project	55	challenge	47
initiative	145	**voice**(9)	94	resource	68	**user**(5)	54	copyright	47
community	135	visa	91	**process**(11)	66	term	53	faq	46
platform	135	beta	90	solution	63	charter	53		
medium	129	**citizen**(6)	88	**health**(3)	62	**mobile**(9)	52		
map	120	website	85	experience	62	level	51		

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Digital Services**—way, individual, edocuments, etransactions, service, reality, government, provision, development, imagination.
2. **Public Sector**—service, government, entity, customer, platform, experience, information, transaction, development, medium.
3. **Community Project**—initiative, project, mobile, government, community, line, hub, year, city, implementation.
4. **Government Services**—clock, visa, job, education, business, service, government, right, beta, voice.
5. **Social Innovation**—community, participation, map, government, policy, sector, service, technology, innovation, law.
6. **Government Transformation**—transformation, government, service, platform, policy, strategy, apps, experience, level, authority.
7. **eTransactions**—accuracy, information, zone, exam, etransactions, evaluation, event, everyday, everyones, evidence.
8. **Data Protection**—data, dissemination, law, protection, provider, entity, emirate, standard, policy, government.

5.11. Japan

Japan's "My Number" system, launched in 2015, provides every resident with a unique 12-digit number, laying the groundwork for an integrated digital identity and administrative system. This system, primarily aimed at unifying tax and social services, also offers a "My Number Card" with an IC chip that enables access to various e-services [112].

With the establishment of the Digital Agency in 2021, Japan underscored its resolution to enhance its digital infrastructure, with the My Number system at the forefront of this transformation. However, the push towards digitisation has raised data privacy and security concerns, prompting Japan to emphasise robust cybersecurity and data protection in the evolution of its digital identity framework [113].

Considering these concerns, the word cluster "online privacy" was detected within the dataset. However, privacy did not occur in the top-word frequencies (see Table 13). Another interesting observation was that address, resident, and residence appeared more frequently than identification, and several metadata items were mentioned in the word list, such as birth, name, and address. The limited prominence of privacy-related terms is noteworthy, while the presence of metadata terms such as *birth*, *name*, and *address* suggests a documentary emphasis on personal-information handling.

The other clusters identified below focused on the usability of the platform and "My Number" Card.

Table 13. Japan—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
number	1167	content	152	detail	101	**registration**(10)	82	braille	63
card(6)	1063	date	140	**status**(3)	100	**phone**(9)	80	box	63
certificate(4)	393	office	140	**resident**(8)	96	**identity**(6)	78	representat.	62
information(17)	352	point	134	**procedure**(10)	95	**process**(10)	77	code	62
name	272	online	133	**residence**(8)	94	store	77	type	61

Table 13. *Cont.*

Word	#	Word	#	Word	#	Word	#	Word	#
issuance	240	email	119	community	92	convenience	72	**authority**$^{(12)}$	60
address	232	photo	117	case	91	**agency**$^{(1)}$	68	pin	60
web	199	page	115	opinion	89	comment	66	center	59
website	193	municipality	114	contact	88	download	66	**user**$^{(5)}$	58
notification	190	document	112	**identification**$^{(6)}$	87	period	66	year	55
service$^{(5)}$	182	request	110	person	87	birth	66	mail	52
inquiry	166	**system**$^{(15)}$	102	data	83	issue	65	month	52

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Website Elements**—website, opinion, content, comment, structure, process, order, format, page, target.
2. **Personal Information**—number, card, certificate, date, residence, information, birth, pin, braille, address.
3. **Identification Process**—card, number, issuance, notification, online, office, procedure, municipality, website, insurance.
4. **Online Privacy**—web, request, photograph, community, background, policy, privacy, data, organization, address.
5. **Digital User Profile**—photo, signature, certificate, data, box, user, community, character, content, experience.
6. **Identity Verification**—certificate, service, information, document, identity, convenience, store, online, identification, verification.
7. **Information Request**—inquiry, number, point, information, content, agency, address, year, representative, authority.
8. **Web Communication**—email, case, information, web, page, address, issue, error, registration, demonstration.
9. **Contact Information**—number, contact, download, municipality, phone, center, office, card, material, code.

5.12. Canada

In analysing Canada's digital identity landscape, provincial initiatives such as British Columbia's BC Services Card illustrate the role of region-specific digital identity solutions, while the federal Pan-Canadian Trust Framework (PCTF) reflects a broader effort to support interoperability, privacy, and security across jurisdictions [114,115].

The Digital Identity and Authentication Council of Canada (DIACC), which brings together public- and private-sector participants, is a key source of published material relating to Canada's trust framework and wider digital identity ecosystem [116]. In the documentary corpus used in this study, the DIACC material provides substantial visibility into framework design, trust relationships, participant roles, credentials, assurance, and governance-related concepts.

The published Canadian corpus is comparatively rich in framework-oriented terminology. As in the New Zealand and Australian material, *identity* is the most frequent term, with 986 occurrences. Other contemporary digital identity terms, including *credential*, *issuer*, *provider*, and *verification*, are also present at meaningful levels, indicating that the corpus places emphasis on trust framework and credential ecosystem concepts.

Cluster analysis similarly shows distinct groupings around authentication, verification, proof, standardisation, participants, and regulation. Within the present study, this suggests that the Canadian corpus distinguishes between related but different documentary themes such as credential handling, authentication processes, proof or evidence, and regulatory context, rather than treating digital identity as a single undifferentiated topic.

The prominence of *privacy* and *security* in Table 14 suggests that both themes are clearly visible within the published corpus. Likewise, the presence of governance-related terms such as *policy*, *regulation*, *authority*, and *framework* indicates that institutional and administrative aspects of the trust framework are well represented in the documentary material. These results should, however, be interpreted as evidence of published emphasis rather than as direct proof that privacy, security, or governance are comprehensively realised in practice.

Table 14. Canada—top 60 word frequencies and cluster summary.

Word	#	Word	#	Word	#	Word	#	Word	#
identity$^{(6)}$	986	**framework**$^{(12)}$	312	evidence	197	management	131	technology	112
information$^{(17)}$	864	assurance	300	**authority**$^{(12)}$	184	**business**$^{(1)}$	131	**role**$^{(2)}$	107
process	698	**privacy**$^{(11)}$	269	**provider**$^{(1)}$	181	issuer	126	context	106
organization	668	**system**$^{(15)}$	242	term	162	control	125	**regulation**$^{(14)}$	105
conformance	645	document	241	loa	158	definition	119	**registry**$^{(8)}$	103
credential	638	ecosystem	239	entity	156	input	118	session	103
component	630	party	234	record	151	purpose	117	**assessment**$^{(11)}$	103
criterion	574	**requirement**$^{(10)}$	234	**risk**$^{(11)}$	148	**legislation**$^{(14)}$	116	type	102
authenticat.	444	relationship	220	data	148	**team**$^{(2)}$	116	**standard**$^{(2)}$	101
level	400	recommend.	219	model	147	source	116	authenticator	99
service$^{(5)}$	341	**policy**$^{(14)}$	217	**security**$^{(11)}$	142	case	114	**transaction**$^{(4)}$	95
person	329	**participant**$^{(2)}$	198	**access**$^{(16)}$	138	**user**$^{(5)}$	113	**verification**$^{(10)}$	94

word $^{(success factor)}$: Words are indexed as per Table 1.

Cluster summary

1. **Standardization**—conformance, criterion, component, level, process, assurance, requirement, person, specifies, loa.
2. **Authentication**—credential, authentication, process, service, level, framework, team, assurance, provider, document.
3. **System**—organization, information, person, processor, identity, process, authority, policy, plan, role.
4. **Verification**—identity, information, ecosystem, process, verification, validation, service, record, organization, resolution.
5. **Guidelines**—component, recommendation, framework, document, term, convention, authentication, person, privacy, vector.
6. **Participants**—information, relationship, issuer, credential, process, party, participant, entity, policy, attribute.
7. **Regulation**—legislation, regulation, policy, jurisdiction, ecosystem, organization, privacy, conformance, criterion, individual.
8. **Proof**—evidence, identity, source, organization, information, authority, person, record, assurance, validation.

Taken together, these clusters suggest that the Canadian corpus is strongly oriented toward trust framework design, participant relationships, assurance, verification, and regulation. Within the present framework, this indicates a documentary emphasis on the governance and operational architecture of digital identity rather than on a single implementation pathway. The Canadian results therefore support the view that publicly available material gives substantial visibility to framework construction and ecosystem roles, while remaining subject to the broader limitations of corpus composition and public-document scope discussed elsewhere in this paper.

5.13. Additional Nations

In a comprehensive analysis of global digital identity frameworks, certain nations presented unique challenges limiting their ability to thoroughly mine and evaluate their digital identity landscapes. Factors contributing to these limitations include the scarcity of publicly available material outlining national digital identity strategies, a significant portion of the content being published exclusively in languages other than English, and the absence of a centralised national identity or trust frameworks. These conditions create barriers to accessing and aggregating comprehensive data, affecting the depth and breadth of analysis within these national contexts. The following subsections summarise the four country/source datasets identified but not retained for metric calculation: Singapore, the Netherlands, Malta, and the United States. These cases are discussed qualitatively to provide contextual comparison and to indicate possible directions for future corpus expansion.

5.13.1. Singapore

Singapore developed SingPass, a comprehensive digital identity system managed by the Government Technology Agency (GovTech). The system provides access to over 400 digital services from the government and private sectors. SingPass ensures robust

security through encryption and multifactor authentication. Features such as biometric verification in its mobile application contribute to a secure yet user-friendly experience. Integrated with SingPass is MyInfo, a digital vault of personal data that enables automatic form-filling across services, exemplifying Singapore's commitment to a seamless and secure digital society.

5.13.2. Netherlands

In the Netherlands, the DigiD system is the primary digital identity solution for accessing government services by combining traditional passwords with a one-time code or biometric data for added security. This is part of a broader electronic identity (eID) scheme aimed at expanding the range of authentication options and improving the integration of public and private-sector services, highlighting the nation's efforts to create a versatile and comprehensive digital identity landscape.

5.13.3. Malta

Malta's eID system offers a secure SSO service for citizens and businesses to interact with government services online. With various levels of authentication to match the sensitivity of the service, the system is designed to be accessible and secure, reflecting Malta's commitment to enhancing digital operations while adhering to the European Union's eIDAS regulation for cross-border digital identity services.

5.13.4. United States

In the United States, the absence of a singular nationwide digital identity system has led to a mosaic of guidelines, initiatives, and solutions. NIST Special Publication 800-63-4 provides federal digital identity guidance for identity proofing, authentication, and federation [117]. States and private entities contribute to a patchwork of varying practices. This diversity poses challenges to interoperability, and any progression towards a unified system would require addressing the pivotal concerns of privacy, security, and the decentralised nature of U.S. governance.

Each of these countries demonstrates a unique path in the evolution of digital identity, balancing the UX, security, and interoperability within their strategies. Their experiences provided valuable insights into the multifaceted nature of digital identity systems worldwide.

5.14. Measurement and Comparison

This section examines and compares national digital identity programmes using the structured framework of success factors derived from Table 1. The resulting measures should be interpreted as artefact-based analytical outputs derived from government-published materials rather than as direct measures of programme performance, implementation success, or citizen outcomes.

5.14.1. Depth of Digital Identity Programmes

The heatmap presented in Table 15 illustrates the number of keywords each nation scored against the digital identity success factors outlined in Table 1. In this study, these scores indicate the relative prominence of the selected success-factor themes within the published corpus for each country, rather than the intrinsic quality or effectiveness of the underlying programme.

Iceland achieves the highest score of any nation on a single factor, most notably in "Wide range and availability of ID-enabled services". Within the corpus analysed here, this suggests a comparatively strong documentary emphasis on service availability and related themes. New Zealand, the overall top scorer across the indicator set, records

a particularly strong score for "vision", indicating that strategic and forward-looking language is especially prominent in its published material. Australia, which is establishing a new trust framework, records the highest score for "trust", the most widely scored indicator across all nations, suggesting that trust-related concepts occupy a central place in the published documentary footprint of its programme.

Table 15. Digital identity success factors (heatmap).

	Factor	DK	FI	KR	NZ	SE	IS	AU	EE	GB	AE	JP	CA
(1)	Private-sector involvement in design and delivery	2	1	4	3	1	0	3	4	1	2	1	2
(2)	A shared vision involving government and industry	3	3	4	6	0	1	5	1	2	3	0	4
(3)	Wide range and availability of ID-enabled services	3	2	2	1	0	8	0	4	1	1	1	0
(4)	Banking services accessible using ID	1	3	0	0	1	1	1	0	1	0	1	1
(5)	High frequency of use	3	2	2	2	2	1	2	1	1	2	2	2
(6)	Existence of a mandatory ID	2	1	2	2	1	0	1	0	1	1	3	1
(7)	An accepted history of national identity schemes	0	0	1	0	0	2	0	0	0	0	0	0
(8)	A national residential register	0	1	0	0	0	2	0	1	0	0	2	1
(9)	Available for use via a variety of channels	2	2	3	0	0	1	0	1	3	2	1	0
(10)	Comprehensive enrolment strategies	1	0	1	1	0	2	3	0	2	0	3	2
(11)	Public trust in the scheme	3	4	4	3	1	4	6	2	3	1	0	4
(12)	Liability model and trust framework addressed	2	2	2	4	1	1	3	0	1	1	1	2
(13)	A clear business case	0	1	1	3	0	0	0	0	0	1	0	0
(14)	Regulatory clarity/confidence	3	3	0	5	1	1	3	0	1	1	0	3
(15)	Well-connected government IT and databases	1	2	1	1	1	1	1	1	1	1	1	1
(16)	Barriers to access an ID removed or addressed	1	0	1	1	0	2	1	1	1	1	0	1
(17)	Strong public awareness and education	1	1	2	1	1	1	1	2	1	3	1	1

Conversely, "An accepted history of national identity schemes" emerges as the least represented feature, with only South Korea and Iceland showing any presence of this factor in the analysed corpus. This result suggests that historical continuity and long-established identity-scheme narratives are less visible within the published materials of the remaining countries, although this should not be interpreted as evidence that such histories are absent in practice.

5.14.2. Alignment Between Nations

Table 16 presents the alignment levels between nations. In this study, alignment refers to the similarity of published indicator-score distributions across countries, rather than to direct interoperability, policy compatibility, or alignment with external standards. The table therefore highlights the extent to which national programmes exhibit similar documentary emphasis across the selected success factors.

Table 16. Digital identity alignment.

	DK	FI	KR	NZ	SE	IS	AU	EE	GB	AE	JP	CA	...	$\bar{x}$
Denmark		0.58	0.62	0.66	0.52	0.48	0.70	0.45^{-}	0.73^{+}	0.59	0.50	0.72		0.59^{+}
Finland	0.58		0.48	0.44	0.52	0.34^{-}	0.48	0.39	0.51	0.58	0.48	0.59^{+}		0.49
Korea	0.62	0.48		0.65^{+}	0.29^{-}	0.34	0.48	0.48	0.50	0.61	0.36	0.47		0.48
New Zealand	0.66	0.44	0.65		0.45	0.29^{-}	0.68^{+}	0.40	0.57	0.60	0.39	0.60		0.52
Sweden	0.52	0.52	0.29^{-}	0.45		0.40	0.66^{+}	0.34	0.64	0.58	0.61	0.62		0.51
Iceland	0.48	0.34	0.34	0.29^{-}	0.40		0.42	0.50	0.60^{+}	0.33	0.49	0.50		0.43
Australia	0.70	0.48	0.48	0.68	0.66	0.42		0.37^{-}	0.65	0.50	0.47	0.85^{+}		0.57
Estonia	0.45	0.39	0.48	0.40	0.34^{-}	0.50^{+}	0.37		0.44	0.49	0.41	0.43		0.43
Great Britain	0.73^{+}	0.51	0.50	0.57	0.64	0.60	0.65	0.44^{-}		0.65	0.58	0.65		0.59^{+}
U.A.E.	0.59	0.58	0.61	0.60	0.58	0.33^{-}	0.50	0.49	0.65^{+}		0.39	0.48		0.53
Japan	0.50	0.48	0.36^{-}	0.39	0.61^{+}	0.49	0.47	0.41	0.58	0.39		0.50		0.47
Canada	0.72	0.59	0.47	0.60	0.62	0.50	0.85^{+}	0.43^{-}	0.65	0.48	0.50			0.58

Note: Dark gray (+) indicates the highest value in a row; light gray (−) indicates the lowest value in a row.

Great Britain demonstrates the highest overall alignment, indicating that its published materials distribute emphasis across the selected indicators in a way that is comparatively similar to the other countries in the sample. By contrast, Estonia records the lowest alignment score. Within the present framework, this suggests that Estonia's published indicator profile differs more substantially from the rest of the comparison set. Given Estonia's recognised strength in digital identity in other contexts, this result should be interpreted cautiously. It may reflect differences in publication style, document type, language availability, or corpus composition rather than weaker programme capability. Further research is required to distinguish between documentary effects and underlying programme characteristics.

5.14.3. Transparency of Digital Identity Programmes

Table 17 highlights the transparency levels of national digital identity programmes. In this study, transparency refers to the breadth and balance with which the selected success factors are represented in the published corpus. It should therefore be interpreted as an artefact-based measure of documentary visibility or communicative comprehensiveness, rather than as a direct measure of institutional openness or accountability.

Table 17. Digital identity transparency.

Country	Coverage	Uniformity	Transparency
GB	0.82	0.84	**0.83**
DK	0.82	0.76	**0.79**
FI	0.82	0.76	**0.79**
AE	0.76	0.82	**0.79**
KR	0.82	0.74	**0.78**
IS	0.82	0.70	**0.76**
CA	0.76	0.75	**0.76**
JP	0.65	0.81	**0.73**
SE	0.53	0.94	**0.73**
NZ	0.76	0.68	**0.72**
AU	0.71	0.67	**0.69**
EE	0.59	0.73	**0.66**

See Methods Section 4.10.2 for definitions of metrics.

Great Britain, Denmark, and Finland emerge as the most transparent within this analytical framework, indicating comparatively broad and balanced public coverage of the selected feature set. In contrast, New Zealand, Australia, and Estonia record the lowest transparency scores. Estonia's position is consistent with the earlier observation that relatively limited or uneven documentary coverage can affect comparative results. Notably, all transparency scores remain within a relatively high range, from 66% to 83%, suggesting that the sampled countries generally publish a moderate-to-high level of information across the selected factors, even where the distribution of coverage differs.

5.14.4. Programme Maturity

The analysis also evaluates documented programme maturity. In this study, maturity refers to the breadth and balance of evidence present across the cluster structure derived from the published corpus, rather than to a direct measure of operational maturity, institutional capability, implementation success, or citizen adoption. Table 18 highlights maturity levels of national digital identity programmes.

Denmark achieves the highest maturity score, followed closely by Finland, South Korea, and Great Britain. Within the present framework, these results indicate comparatively complete and balanced documentary coverage across the analysed clusters. Sweden records the lowest maturity score, with Estonia ranking second lowest. These results suggest that,

relative to the selected cluster structure, the published material for these countries is either narrower in scope or less evenly distributed. In Estonia's case, for example, the lower score may reflect the composition, accessibility, or selectivity of the published corpus rather than immaturity of the underlying programme. These findings should therefore be interpreted as measures of documentary balance and completeness, not as definitive assessments of real-world programme maturity.

Table 18. Digital identity maturity.

Country	Coverage	MVR	Saturation	Maturity
DK	0.82	0.11	0.55	**0.53**
FI	0.82	0.14	0.41	**0.50**
KR	0.82	0.13	0.44	**0.50**
GB	0.82	0.15	0.39	**0.49**
IS	0.82	0.29	0.21	**0.48**
AE	0.76	0.15	0.39	**0.47**
NZ	0.76	0.18	0.32	**0.46**
CA	0.76	0.16	0.37	**0.46**
AU	0.71	0.20	0.29	**0.43**
JP	0.65	0.18	0.33	**0.41**
EE	0.59	0.22	0.26	**0.38**
SE	0.53	0.20	0.29	**0.36**

See Methods Section 4.10.3 definitions of metrics.

5.15. Cross-Country Synthesis and Interpretation

Collectively, the country-level and comparative results indicate that the analysed government digital identity programmes differ not only in substantive programme design, but also in the way they are documented publicly. This distinction is important because the present study measures published evidence rather than operational performance. The results therefore reveal the documentary profile of each programme: what governments choose to describe, emphasise, and evidence in publicly accessible materials.

A first cross-country pattern is the emergence of distinct documentary archetypes. New Zealand, Australia, and Canada exhibit strongly governance- and trust framework-oriented profiles, with high visibility of terms related to frameworks, accreditation, regulation, privacy, security, assurance, and participant roles. Sweden, by contrast, presents a more technically oriented corpus, with strong emphasis on protocol, metadata, assertion, certificate, schema, service-provider, and message-exchange terminology. Iceland, Estonia, and the United Arab Emirates show more service-portal or wider digital-government-oriented profiles, where identity-related material is embedded within broader administrative, civic, sectoral, or onboarding documentation. Denmark, Finland, South Korea, and Great Britain occupy more mixed positions, combining authentication, service access, public administration, governance, and technical implementation themes.

A second pattern is that the comparative measures do not collapse into a single country ranking. Great Britain records the highest transparency score, Denmark records the highest documented maturity score, and different country pairs exhibit different alignment relationships. This suggests that the three measures capture different aspects of the published corpus. Alignment measures similarity in the distribution of published indicator emphasis. Transparency measures the breadth and balance of coverage across selected success factors. Maturity measures the breadth and balance of evidence across the derived cluster structure. A country may therefore be highly transparent in the sense of broad published coverage, but not necessarily highest in documented maturity, and a technically rich corpus may score differently from a governance-rich corpus.

A third pattern concerns the role of document genre and source composition. Some results that may appear intuitive at a high level become analytically useful because the method identifies how those differences appear in the corpus. For example, the Swedish results are shaped by the prominence of technical framework material, while the New Zealand and Australian results are shaped by trust framework and regulatory documentation. Estonia's lower comparative scores should not be read as evidence of weak digital identity capability; rather, they illustrate how a limited or less text-accessible corpus can affect artefact-based measurement. This reinforces the importance of interpreting the results as documentary evidence rather than as direct programme evaluation.

A fourth pattern is that common digital identity themes are visible across jurisdictions, but are not documented with equal emphasis. Trust, security, service access, regulation, interoperability, and user interaction recur across the corpus, supporting the view that these are central concerns in government digital identity programmes. However, other themes such as liability models, business case, historical identity-scheme continuity, and public education are less consistently represented. This unevenness is itself a finding: it suggests that while governments often publish material on trust, service enablement, and regulation, they do not consistently communicate the full ecosystem logic of digital identity programmes in a balanced way.

These findings place the study in a complementary position relative to the existing digital government benchmarking literature. Macro-level indices such as the UN EGDI, OECD DGI, World Bank GTMI, and European Commission eGovernment Benchmark assess broad digital government readiness, capability, service delivery, and transformation. The present study does not replicate those assessments. Instead, it provides a digital-identity-specific artefact-based lens that shows how national programmes are described and evidenced through public documentation. The contribution is therefore a systematic way of comparing documentary emphasis and published evidence across digital identity programmes, rather than a claim to measure operational success or institutional maturity directly.

Some of these findings may appear intuitive in the broad sense that governments differ in their approaches to digital identity. The contribution of the analysis is that these differences are not asserted impressionistically; they are derived from a repeatable corpus-based pipeline and expressed through comparable artefact-based measures.

5.16. Contribution of the Artefact-Based Findings

The contribution of the artefact-based findings is not that they provide a definitive ranking of national digital identity programmes, but that they make the public evidential layer of those programmes visible and comparable. This matters because digital identity systems are trust infrastructures: citizens, relying parties, policymakers, and researchers often encounter them first through public claims about security, privacy, usability, governance, accreditation, and accountability.

The results show why this public evidential layer is analytically meaningful. The same broad category of trust framework-oriented programme appears differently across jurisdictions: New Zealand is characterised by regulatory design and accreditation, Australia by compliance, privacy, risk, and oversight, and Canada by credentials, assurance, authentication, and verification. Sweden and Iceland illustrate a different contrast: Sweden exposes a more technical and standards-oriented documentation layer, while Iceland embeds identity-related evidence within wider service and administrative documentation. These examples are not introduced as operational judgements; they show that the method can distinguish what is publicly evidenced, how it is evidenced, and what is therefore available for external inspection.

This distinction is especially important in cases such as Estonia. Estonia is widely recognised as a mature digital-government context, yet it records lower comparative scores in this artefact-based analysis because the sampled public corpus is relatively limited and partly video-based. This does not weaken the value of the method; rather, it demonstrates what the method measures. It identifies the visibility, balance, and structure of published evidence, not the underlying technical capability of the programme.

The study therefore defines one side of a larger comparison: the government-published evidential record. Future work can use this baseline to compare public claims with citizen perceptions of usability, privacy, security, and trustworthiness, as well as with expert review, technical architecture analysis, operational performance, and independent security or privacy evaluation.

6. Limitations of the Study

This study has several limitations that should be considered when interpreting the findings. Most importantly, the measures reported in this paper are artefact-based measures derived from publicly available government materials. They therefore reflect the documentary footprint of national digital identity programmes rather than the full underlying programme in operation. As a result, the reported measures of alignment, transparency, and maturity should not be interpreted as direct proxies for implementation success, institutional capability, service quality, citizen adoption, or broader societal outcomes.

A further limitation concerns corpus scope and comparability. Digital identity programmes are documented unevenly across jurisdictions, and the analysed corpora were not fully homogeneous in type or emphasis. Some countries published material that was more policy- and regulation-oriented, while others provided more technical, service-oriented, or sector-specific documentation. In some cases, broader administrative content may also have been captured alongside digital identity material. These differences in document genre, publication strategy, and source composition may influence term frequencies, cluster structure, and comparative scores independently of the underlying programme itself.

The study is also affected by variation in disclosure and accessibility. The amount of publicly available information on digital identity programmes differs substantially across countries, with some governments providing extensive public documentation and others publishing only limited or selective material. This creates the risk that countries with richer or more text-accessible public documentation appear comparatively stronger on artefact-based measures, even where this does not necessarily reflect greater programme capability. Related to this, non-English-speaking jurisdictions may be disadvantaged where relevant materials are unavailable in English, only partially translated, or expressed using terminology that is not strongly represented in the analytical corpus.

Selection effects also remain. The study focuses on highly digitalised societies, which supports a more meaningful comparison among countries with established digital government activity, but also narrows the scope of inference. Even within this group, countries differ in administrative traditions, communication cultures, legal structures, publication practices, and the degree to which digital identity is centralised or distributed across institutions. These differences can affect the visibility and comparability of programme documentation.

The focus on highly digitalised societies also means that the findings should not be generalised to all national digital identity contexts. Developing countries and emerging digital identity programmes may exhibit different publication practices, institutional constraints, identity architectures, inclusion challenges, and levels of public documentation. The present sample therefore supports comparison within a bounded set of relatively advanced digital government contexts, but it does not provide a global assessment of digital identity programmes. Future work should extend the corpus to developing-country

contexts and compare whether the same artefact-based measures behave consistently across different administrative, linguistic, and infrastructural conditions.

The selection of countries and indicators also shapes the findings because both choices occur before data collection, classification, clustering, and metric calculation. Several significant digital identity programmes, including those of France, Germany, Italy, and India, are not included in the present corpus. Their exclusion should be understood as a limitation of scope rather than as a judgement about their relevance or maturity. Similarly, the use of the Open Identity Exchange success/failure indicators provides a coherent ecosystem-oriented baseline, but alternative indicator frameworks could produce different emphases and potentially different comparative results. Future work should therefore extend the corpus to additional countries and conduct sensitivity analysis using alternative or expanded indicator sets. The Estonia and Iceland cases illustrate this issue particularly clearly: Estonia's public corpus was affected by limited text-accessible material and video-heavy sources, while the Icelandic corpus captured a broad administrative and regulatory service environment that extended beyond narrowly defined digital identity content.

The regional origin of the selected success-factor framework is a further limitation. Although the Open Identity Exchange indicators provide a coherent and implementation-grounded baseline, their UK programme context may privilege some ecosystem assumptions over others. Alternative frameworks, especially those developed in non-European, developing-country, or rights-based contexts, may yield different mappings and different comparative scores.

Another limitation is the dynamic nature of digital identity ecosystems. Digital identity frameworks, trust models, regulations, and associated technologies evolve rapidly. The findings reported here therefore represent a time-bound view of the published materials available during the study period and may not fully reflect later policy changes, implementation developments, or newly published documentation.

The study is further limited by the absence of direct access to proprietary or internal systems. Because the analysis is based on published web material, it cannot independently verify technical implementations, internal controls, operational performance, or the practical effectiveness of the systems described. The study should therefore be understood as an analysis of public documentary evidence rather than a full technical or institutional audit.

The NLP methods used in this study are also intentionally limited. Word-frequency analysis, clustering, topic inspection, POS tagging, and triple extraction provide transparent and repeatable corpus-level signals, but they do not fully capture deeper semantic relations, rhetorical framing, sentiment, modality, or evaluative stance. Government documents are often formal and institutionally constrained, so sentiment analysis may not always be the most informative extension; however, semantic role labelling, discourse analysis, argument mining, and contextual embedding-based methods could provide richer insight into how governments frame trust, risk, privacy, accountability, and citizen agency. Future work should therefore extend the present artefact-based measures with deeper semantic analysis while preserving comparability across jurisdictions.

The weighting choices used in the transparency and maturity metrics are also a limitation. The present study uses transparent baseline weights rather than weights optimised against external validation data or policy priorities. This preserves interpretability but does not establish that the chosen weights are uniquely optimal. Future work should conduct sensitivity analysis by varying the transparency weights across the interval $\alpha + \beta = 1$ and varying the maturity weights across the simplex $\alpha + \beta + \gamma = 1$, then testing whether country ordering and substantive conclusions remain stable.

These limitations delimit, rather than negate, the contribution of the artefact-based results. The study contributes evidence about the public documentation and communication

of digital identity programmes, not about the complete technical or social reality of those programmes. This distinction is important because public artefacts are themselves part of the trust infrastructure of digital identity. They shape what can be externally inspected, what citizens and relying parties can understand, and what claims governments make visible about usability, privacy, security, governance, and trust.

Taken together, these limitations indicate that the results should be interpreted cautiously and comparatively. The study provides a structured way to analyse how governments publicly describe and evidence digital identity programmes, but it does not claim to provide a complete or definitive measure of programme quality. Future research could strengthen this line of inquiry by incorporating tighter source delimitation, genre-sensitive analysis, multilingual collection strategies, longitudinal sampling, and complementary expert or case-based validation.

7. Conclusions

This study examined governmental approaches to digital identity through the analysis of publicly available digital identity artefacts. Rather than evaluating programme performance directly, the study developed a digital-identity-specific, corpus-based comparative method for analysing how governments describe, structure, and evidence their digital identity programmes in published materials. The results show that the value of this approach lies not simply in confirming that national programmes differ, but in making those differences visible as measurable documentary patterns. Across the analysed corpus, countries varied in the extent to which they published material emphasising governance, trust frameworks, regulation, technical implementation, authentication, interoperability, service access, privacy, and user interaction.

With respect to Q1, the study finds that governmental approaches to digital identity, as reflected in published materials, vary substantially across jurisdictions. Some countries exhibit broad documentary coverage across governance, services, regulation, and technical implementation, while others present narrower or more selectively documented profiles. Although some convergence is visible in the emphasis placed on interoperability, trust frameworks, and digitally enabled public services, the comparative results do not indicate a single, unified documentary model of digital identity across countries.

Regarding Q2, the study does not measure operational maturity over time; instead, it compares documented programme maturity as reflected in the breadth and balance of the published corpus. Within this analytical framing, some jurisdictions display broader and more evenly distributed documentary coverage across the selected indicators and clusters, while others appear more concentrated in specific regulatory, technical, or service-oriented areas. These differences suggest variation in publication style, programme emphasis, and documentary scope, rather than providing definitive evidence of underlying programme capability.

For Q3, the published material indicates that governments commonly place emphasis on trust, usability, regulation, accreditation, interoperability, and service access. Security- and privacy-related themes are also visible in several jurisdictions, although their prominence varies across the corpus. At the same time, some topics appear less consistently represented, including liability models, shared business rationales, and public-facing educational material. This suggests that while trust and service delivery are well-established priorities in public documentation, other elements of digital identity ecosystems are communicated less consistently.

For Q4, the comparative analysis identifies several recurring documentary patterns, including the prominence of regulatory clarity, trust frameworks, identity-enabled service delivery, and connections to wider digital government infrastructure. These recurring themes do not amount to a universal digital identity model, but they do indicate areas of

common emphasis that may inform future comparative work, ontology development, and cross-jurisdictional analysis.

Placed in the context of the existing literature, the findings support the need for a digital-identity-specific layer of comparison alongside broader digital government benchmarks. Existing comparative frameworks provide valuable macro-level assessments of digital government capability, infrastructure, service delivery, and transformation. However, they do not directly show how digital identity programmes are publicly described, justified, governed, and evidenced through national artefacts. The present results therefore complement this literature by demonstrating that public documentation itself can be analysed as a comparative evidence base. At the same time, the findings reinforce cautions from the literature on cross-country comparison and data mining: differences in publication practice, language, source availability, and document genre can affect measured outputs and must be considered when interpreting comparative results.

Overall, the study contributes a structured artefact-based method for comparing the public evidential layer of government digital identity programmes. The contribution is not that the paper provides a definitive judgement on programme quality or implementation success, but that it makes the public documentation of such programmes measurable and comparable. This is relevant to the scientific discussion because public artefacts mediate how digital identity infrastructures are explained, trusted, scrutinised, and compared. They also provide the documentary baseline against which future studies can assess citizen perceptions of usability, privacy, security, and trustworthiness, and compare those perceptions with the technical and institutional reality of deployed systems.

A further extension would be to connect the artefact-based analysis developed here with technical validation methods. Robust feature-representation approaches, such as those developed for accurate human parsing in complex visual scenes, may inform future work on extracting stable features from heterogeneous or multimodal digital identity artefacts [118]. Similarly, proactive detection and watermarking methods developed for deepfake detection may provide useful reference points for future studies of anti-counterfeiting, credential authenticity, and security verification in deployed digital identity systems [119]. These technical extensions are outside the scope of the present artefact-based documentary analysis, but they are relevant to future triangulation between public claims and operational security evidence.

Future research should build on this artefact-based baseline through tighter source delimitation, expansion to additional countries such as France, Germany, Italy, and India, multilingual collection, longitudinal sampling, sensitivity testing against alternative indicator frameworks, and triangulation of government-published claims against citizen perceptions, expert review, technical architecture analysis, operational performance evidence, and independent security or privacy evaluation.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jcp6030093/s1, Table S1: PRISMA 2020 Checklist used in the current systematic review study. Table S2: Search strategy and source identification procedure.

Author Contributions: Conceptualization, M.C. and A.M.; methodology, M.C.; investigation, M.C.; formal analysis, M.C.; data curation, M.C.; writing—original draft preparation, M.C.; writing—review and editing, M.C. and A.M.; supervision, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Commonwealth Scholarship Commission in the UK (Reference: CSC CR-2019-67).

Institutional Review Board Statement: The study's ethical review and approval were provided by the Department of Computer Science Departmental Research Ethics Committee (DREC), University of Oxford (Reference: CS_C1A_021_025).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this article and in the accompanying public GitHub repository. The distributed client/server architecture used for extraction and mining of clustered terms is not included in the repository due to the complexity of its operational requirements. However, a Jupyter Notebook has been provided as a test harness for the NLP Python code and clustering algorithms, together with preprocessed data files generated during the research. These materials are publicly accessible at https://github.com/oxford-mc/government-digital-identity (accessed on 14 May 2026). Further information is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hert, P.D.; Papakonstantinou, V.; Malgieri, G.; Beslay, L.; Sanchez, I. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Comput. Law Secur. Rev.* **2018**, *34*, 193–203. [CrossRef]
2. Comb, M.; Martin, A. The Pervasiveness of Digital Identity: Surveying Themes, Trends, and Ontological Foundations. *Information* **2026**, *17*, 85. [CrossRef]
3. Comb, M.; Martin, A. Universal Digital Identity Stakeholder Alignment: Toward Context-Layered RAG Architectures for Ecosystem-Aware AI. *Digital* **2026**, *6*, 4. [CrossRef]
4. Laatikainen, G.; Kolehmainen, T.; Abrahamsson, P. Self-Sovereign Identity Ecosystems: Benefits and Challenges. Technical Report. 2021. Available online: https://jyx.jyu.fi/jyx/Record/jyx_123456789_77892 (accessed on 14 February 2026).
5. Comb, M.; Martin, A. Mining digital identity insights: Patent analysis using NLP. *EURASIP J. Inf. Secur.* **2024**, *2024*, 21. [CrossRef]
6. Masiero, S.; Arvidsson, V. Degenerative outcomes of digital identity platforms for development. *Inf. Syst. J.* **2021**, *31*, 903–928. [CrossRef]
7. Cheesman, M. Self-Sovereignty for Refugees? The Contested Horizons of Digital Identity. *Geopolitics* **2022**, *27*, 134–159. [CrossRef]
8. European Parliament and Council. Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on Electronic Identification and Trust Services for Electronic Transactions in the Internal Market and Repealing Directive 1999/93/EC (eIDAS Regulation). 2014. Available online: https://eur-lex.europa.eu/eli/reg/2014/910/oj/eng (accessed on 10 October 2023).
9. W3C. W3C Verifiable Credentials. 2019. Available online: https://www.bcs.org/media/4653/kent-w3c-verifiable-credentials-031019.pdf (accessed on 3 March 2026).
10. Loscio, B.F.; Burle, C.; Calegari, N. Data on the Web Best Practices. 2017. Available online: https://www.dublincore.org/webinars/2017/data_on_the_web_best_practices_challenges_and_benefits/slides.pdf (accessed on 24 January 2026).
11. Elliott, J.; Birch, D.; Ford, M.; Whitcombe, A. *Overcoming Barriers In the EU Digital Identity Sector*; European Commission: Brussels, Belgium, 2007.
12. Martin, A.; Martinovic, I. Security and Privacy Impacts of a Unique Personal Identifier. 2016. Available online: https://www.ctga.ox.ac.uk/publications/security-and-privacy-impacts-of-a-unique-personal-identifier (accessed on 14 May 2026).

13. Wang, F.; Filippi, P.D. Self-Sovereign Identity in a Globalized World: Credentials-Based Identity Systems as a Driver for Economic Inclusion. *Front. Blockchain* **2019**, *2*, 28. [CrossRef]
14. Vihma, P. The (Bumpy) Road to European Digital Identity-e-Estonia. 2022. Available online: https://e-estonia.com/the-bumpy-road-to-european-digital-identity/ (accessed on 10 December 2025).
15. Zhao, B. Web Scraping. In *Encyclopedia of Big Data*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 1–3. [CrossRef]
16. Luscombe, A.; Dick, K.; Walby, K. Algorithmic thinking in the public interest: Navigating technical, legal, and ethical hurdles to web scraping in the social sciences. *Qual. Quant.* **2022**, *56*, 1023–1044. [CrossRef]
17. V., D.S. Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics. *J. Soft Comput. Paradig.* **2020**, *2*, 101–110. [CrossRef]
18. Giannopoulou, A.; Wang, F. Self-sovereign identity. *Internet Policy Rev.* **2021**, *10*, 1–10. [CrossRef]
19. Feulner, S.; Guggenberger, T.; Lautenschlager, J.; Urbach, N.; Völter, F. Self-sovereign identity in the public sector: Affordances, experimentation, and actualization. *Gov. Inf. Q.* **2025**, *42*, 102052. [CrossRef]
20. Sobel, B.L.W. A New Common Law of Web Scraping. *Lewis Clark Law Rev.* **2020**, *25*, 147.
21. Alomari, E.; Katib, I.; Albeshri, A.; Mehmood, R. Covid-19: Detecting government pandemic measures and public concerns from twitter arabic data using distributed machine learning. *Int. J. Environ. Res. Public Health* **2021**, *18*, 282. [CrossRef]
22. Raj, J.S.; Iliyasu, A.M.; Bestak, R.; Baig, Z.A. (Eds.) *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2020*; Lecture Notes on Data Engineering and Communications Technologies; Springer: Singapore, 2021; Volume 59. [CrossRef]
23. Hassani, H.; Beneki, C.; Unger, S.; Mazinani, M.T.; Yeganegi, M.R. Text mining in big data analytics. *Big Data Cogn. Comput.* **2020**, *4*, 1. [CrossRef]
24. Morshedi, R.; Chu, B.; Huang, E. Web Scraping: Applications in Infrastructure Planning. In Proceedings of the 24th Association of Public Authority Surveyors Conference (APAS2019), Pokolbin, NSW, Australia, 1–3 April 2019; Technical Report.
25. Sellars, A. Twenty Years of Web Scraping and the Computer Fraud and Abuse Act. *BUJ Sci. Tech. L.* **2018**, *24*, 372.
26. Nguyen, H.A. Web scraping: A big data building tool and its status in the fintech sector in Viet Nam. *J. Sci. Technol. Inf. Commun.* **2023**, *2*, 41–54. Available online: https://tapchi.ptit.edu.vn/jstic-ptit/index.php/jstic/article/view/1423 (accessed on 14 May 2026).
27. Yan, H.; Yang, N.; Peng, Y.; Ren, Y. Data mining in the construction industry: Present status, opportunities, and future trends. *Autom. Constr.* **2020**, *119*, 103331. [CrossRef]
28. Fano, R.; Corbato, F. Time-sharing on computers. *Sci. Am.* **1966**, *215*, 128–143. [CrossRef]
29. Cao, Y.; Yang, L. A survey of Identity Management technology. In Proceedings of the 2010 IEEE International Conference on Information Theory and Information Security, ICITIS 2010, Beijing, China, 17–19 December 2010; pp. 287–293. [CrossRef]
30. Dib, O.; Toumi, K. Decentralized identity systems: Architecture, challenges, solutions and future directions. *Ann. Emerg. Technol. Comput. (AETiC)* **2020**, *4*, 19–40. [CrossRef]
31. Diebold, Z.; O'mahony, D. Self-Sovereign Identity using Smart Contracts on the Ethereum Blockchain. Master's Thesis, University of Dublin, Dublin, Ireland, 2017; Technical report.
32. Bramhall, P.; Hansen, M.; Rannenberg, K.; Roessler, T. User-centric identity management. *IEEE Secur. Priv.* **2007**, *5*, 84–87. [CrossRef]
33. Selvanathan, N.; Jayakody, D.; Damjanovic-Behrendt, V. Federated identity management and interoperability for heterogeneous cloud platform ecosystems. In Proceedings of the ARES '19: 14th International Conference on Availability, Reliability and Security, Canterbury, UK, 26–29 August 2019. [CrossRef]
34. Whitley, E.A. *Trusted Digital Identity Provision: GOV.UK Verify's Federated Approach*; CGD policy paper; Center for Global Development: Washington, DC, USA, 2018; pp. 94–120.
35. Open Identity Exchange. *Digital Identity in the UK: The Cost of Doing Nothing*; Open Identity Exchange: London, UK, 2018.
36. Fioravanti, F.; Nardelli, E. Chapter 17 Identity Management For E-Government Services Chapter Overview. Technical Report. 2008. Available online: https://www.mat.uniroma2.it/~nardelli/publications/Digital-Government-08.pdf (accessed on 2 January 2026).
37. Cristofaro, E.D.; Du, H.; Freudiger, J.; Norcie, G. A Comparative Usability Study of Two-Factor Authentication. *arXiv* **2013**, arXiv:1309.5344.
38. Wang, D.; Wang, P. Two Birds with One Stone: Two-Factor Authentication with Security beyond Conventional Bound. *IEEE Trans. Dependable Secur. Comput.* **2018**, *15*, 708–722. [CrossRef]
39. Jin, A.T.B.; Ling, D.N.C.; Goh, A. Biohashing: Two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognit.* **2004**, *37*, 2245–2255. [CrossRef]

40. Al-Khouri, A.; Bal, J. Electronic government in the GCC countries. *Int. J. Comput. Inf. Eng.* **2008**, *2*, 1614–1629.
41. Perlman, R. An Overview of PKI Trust Models. *IEEE Netw.* 1999, *13*, 38–43. [CrossRef]
42. Adams, C.; Lloyd, S. *Understanding PKI: Concepts, Standards, and Deployment Considerations*; Addison-Wesley Professional: Boston, MA, USA, 2003.
43. Shan, H.L. *Government PKI Deployment and Usage in Taiwan*; Technical report; Procon Ltd.: Marlborough, UK, 2004.
44. Kalja, A.; Robal, T.; Vallner, U. New generations of Estonian eGovernment components. In Proceedings of the 2015 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, 2–6 August 2015.
45. Lekkas, D.; Zissis, D. LNICST 99 - Leveraging the e-passport PKI to Achieve Interoperable Security for e-government Cross Border Services. In *International Conference on e-Democracy*; Technical report; Springer: Berlin/Heidelberg, Germany, 2011.
46. Al-Khouri, A.M. PKI in Government Identity Management Systems. *arXiv* **2011**, arXiv:1105.6357. [CrossRef]
47. Scott, M.; Acton, T.; Hughes, M. Title An assessment of biometric identities as a standard for e-government services. *Int. J. Serv. Stand.* **2005**, *1*, 271–286.
48. Singh, P. Aadhaar and data privacy: Biometric identification and anxieties of recognition in India. *Inf. Commun. Soc.* **2021**, *24*, 978–993. [CrossRef]
49. Ayamba, I.; Ekanem, O. National identity management in Nigeria: Policy dimensions and implementation. *Int. J. Humanit. Soc. Sci. Stud.* **2016**, *3*, 279–287.
50. Alimia, S. Performing the Afghanistan-Pakistan border through refugee ID cards. *Geopolitics* **2019**, *24*, 391–425. [CrossRef]
51. Toth, K.C.; Anderson-Priddy, A. Self-Sovereign Digital Identity: A Paradigm Shift for Identity. *IEEE Secur. Priv.* **2019**, *17*, 17–27. [CrossRef]
52. Clercq, J.D. Single Sign-On Architectures. In *International Conference on Infrastructure Security*; Technical report; Springer: Berlin/Heidelberg, Germany, 2002.
53. Soares, D.; Amaral, L. Information systems interoperability in public administration: Identifying the major acting forces through a Delphi study. *J. Theor. Appl. Electron. Commer. Res.* **2011**, *6*, 61–94. [CrossRef]
54. Mecca, G.; Santomauro, M.; Santoro, D.; Veltri, E. On federated single sign-on in e-government interoperability frameworks. *Int. J. Electron. Gov.* **2016**, *8*, 6–21. [CrossRef]
55. Alghamdi, I.A.; Goodwin, R.; Rampersad, G. E-Government Readiness Assessment for Government Organizations in Developing Countries. *Comput. Inf. Sci.* **2011**, *4*, 3–17. [CrossRef]
56. Brunner, C.; Gallersdorfer, U.; Knirsch, F.; Engel, D.; Matthes, F. DID and VC: Untangling decentralized identifiers and verifiable credentials for the web of trust. In Proceedings of the ICBTA 2020: 2020 the 3rd International Conference on Blockchain Technology and Applications, Xi'an, China, 14–16 December 2020; ACM International Conference Proceeding Series; pp. 61–66. [CrossRef]
57. European Parliament and Council. Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework. *Off. J. Eur. Union* **2024**, L 2024/1183. Available online: https://eur-lex.europa.eu/eli/reg/2024/1183/oj (accessed on 14 May 2026).
58. Bauer, D.; Blough, D.M.; Cash, D. Minimal information disclosure with efficiently verifiable credentials. In Proceedings of the ACM Conference on Computer and Communications Security, Alexandria, VI, USA, 27–31 October 2008; pp. 15–24. [CrossRef]
59. Laborde, R.; Oglaza, A.; Wazan, S. A User-Centric Identity Management Framework based on the W3C Verifiable Credentials and the FIDO Universal Authentication Framework. In Proceedings of the 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 10–13 January 2020.
60. Cameron, K. The Laws of Identity. Microsoft Corporation White Paper, May 2005. Available online: https://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf (accessed on 28 April 2026).
61. Arora, S.K.; Li, Y.; Youtie, J.; Shapira, P. Using the wayback machine to mine websites in the social sciences: A methodological resource. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 1904–1915. [CrossRef]
62. Ashari, I.F.; Banjarnahor, R.; Farida, D.R.; Aisyah, S.P.; Dewi, A.P.; Humaya, N. Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *J. Appl. Inform. Comput.* **2022**, *6*, 7–15. [CrossRef]
63. Parks, A.M. Unfair Collection: Reclaiming Control of Publicly Available Personal Information from Data Scrapers. *Mich. Law Rev.* **2022**, *120*, 913–945. [CrossRef]
64. Thota, P.; Ramez, E. Web Scraping of COVID-19 News Stories to Create Datasets for Sentiment and Emotion Analysis. In Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 29 June 2021–2 July 2021; pp. 306–314. [CrossRef]
65. United Nations Department of Economic and Affairs. *UN E-Government Survey 2024*, 13th ed.; Technical report; United Nations: New York, NY, USA, 2024.
66. OECD. *Digital Government Index and Open, Useful and Re-Usable Data Index: 2025 Results and Key Findings*; Technical report; OECD Publishing: Paris, France, 2026. [CrossRef]

67. World Bank. *GovTech Maturity Index 2025: Tracking Public Sector Digital Transformation Worldwide*; Technical report; GTMI 2025 Update brief; World Bank: Washington, DC, USA, 2025.
68. European Commission. *Digital Decade 2024: eGovernment Benchmark*; Technical report; European Commission: Brussels, Belgium, 2024.
69. OECD. *Digital Public Infrastructure for Digital Governments*; Technical report; OECD Publishing: Paris, France, 2024. [CrossRef]
70. Waara, Åsa. Examining Digital Government Maturity Models: Evaluating the Inclusion of Citizens. *Adm. Sci.* **2025**, *15*, 73. [CrossRef]
71. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef]
72. Garber, E.; Haine, M. (Eds.) *Human-Centric Digital Identity: For Government Officials*, version 1.1; OpenID Foundation: San Ramon, CA, USA, 2023. Available online: https://openid.net/wp-content/uploads/2023/10/Human-Centric_Digital_Identity_Final-v1.1.pdf (accessed on 14 May 2026).
73. World Bank. *Principles of Identification*; Technical report; World Bank: Washington, DC, USA, 2025.
74. Li, J. *E-Government Survey 2022*; Technical report; United Nations: New York, NY, USA, 2022.
75. European Union. General Data Protection Regulation (GDPR). 2016. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 12 March 2024).
76. Needleman, S.B.; Wunsch, C.D. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *J. Mol. Biol.* **1970**, *48*, 443–453. [CrossRef]
77. Shvaiko, P.; Euzenat, J. Ontology matching: State of the art and future challenges. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 158–176. [CrossRef]
78. Branke, J.; Deb, K.; Roman, S. *Multiobjective Optimization, Interactive and Evolutionary Approaches*; Technical report; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.
79. Shannon, C.E.; Weaver, W. The mathematical theory of communication. *Bell Syst. Tech. J.* **1949**, *27*, 379–423. [CrossRef]
80. Manning, C.D.; Raghavan, P.; Schutze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2009.
81. Cowell, F.A. Measurement of inequality. *Handb. Income Distrib.* **2000**, *1*, 87–166.
82. Microsoft. *Visual Studio IDE Software*; Microsoft: Redmond, WA, USA, 2026. Available online: https://visualstudio.microsoft.com/vs/ (accessed on 21 March 2026).
83. Project Jupyter. Jupyter Notebook. Software. 2026. Available online: https://jupyter.org/ (accessed on 12 March 2026).
84. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef] [PubMed]
85. McKinney, W. Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June 28–3 July 2010; Volume 445, pp. 51–56.
86. Honnibal, M.; Montani, I.; Van Landeghem, S.; Boyd, A. spaCy: Industrial-Strength Natural Language Processing in Python. In *Zenodo Software*; Zenodo: Honolulu, HI, USA, 2020. [CrossRef]
87. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
88. Gommers, R.; Virtanen, P.; Haberland, M.; Burovski, E.; Reddy, T.; Weckesser, W.; Oliphant, T.E.; Nelson, A.; Cournapeau, D.; Polat, I.; et al. *scipy/scipy: SciPy 1.17.0*, version v1.17.0; Zenodo: Geneva, Switzerland, 2026. [CrossRef]
89. Řehůřek, R.; Sojka, P. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, Valletta, Malta, 22 May 2010*; European Language Resources Association (ELRA): Valletta, Malta, 2010; pp. 45–50. Available online: https://is.muni.cz/publication/884893/en (accessed on 14 May 2026).
90. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O'Reilly Media, Inc.: Newton, MA, USA, 2009.
91. Robinson, I.; Webber, J.; Eifrem, E. *Graph Databases: New Opportunities for Connected Data*, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2015; pp. 27–46.
92. Kingo, T.; Aranha, D.F. User-centric security analysis of MitID: The Danish passwordless digital identity solution. *Comput. Secur.* **2023**, *132*, 103376. [CrossRef]
93. Yli-Huumo, J.; Paivarinta, T.; Rinne, J.; Smolander, K.; Suomi, K.S. Suomi.fi—Towards Government 3.0 with a National Service Platform. In *International Conference on Electronic Government*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 3–14. [CrossRef]

94. Feigenbaum, E.A.; Nelson, M.R. The Korean Way with Data: How the World's Most Wired Country Is Forging a Third Way. Carnegie Endowment for International Peace. 2021. Available online: https://carnegieendowment.org/research/2021/08/the-korean-way-with-data-how-the-worlds-most-wired-country-is-forging-a-third-way (accessed on 28 April 2026).
95. Department of Internal Affairs. *Regulatory Impact Statement: Progressing Digital Identity: Establishing a Trust Framework*; New Zealand Government: Wellington, New Zealand, 2020. Available online: https://www.regulation.govt.nz/assets/RIS-Documents/ris-dia-pdietf-jun20.pdf (accessed on 28 April 2026).
96. Bogusz, C.I.; Kyriakou, H. Digital Identity as a Platform of Platforms: Investigating BANKID'S Effect on Swedish Organizations Research in Progress. In Proceedings of the ECIS 2023: European Conference on Information Systems, Kristiansand, Norway, 11–16 June 2023; Technical report.
97. Husz, O. Bank Identity: Banks, ID Cards, and the Emergence of a Financial Identification Society in Sweden. *Enterp. Soc.* **2018**, *19*, 391–429. [CrossRef]
98. Liesbrock, P. The Giant is Lagging Behind: How the German Electronic ID Fails to Reap its Potential. Master's Thesis, Stockholm University, Stockholm, Sweden, 2022.
99. van Marion, L.; Hovland, J.H. *The Nordic Digital Ecosystem: Actors, Strategies, Opportunities*; Nordic Innovation: Oslo, Norway, 2015. Available online: https://www.norden.org/en/publication/nordic-digital-ecosystem-actors-strategies-opportunities (accessed on 28 April 2026).
100. Hansteen, K.; Ølnes, J.; Alvik, T. *Nordic Digital Identification (eID): Survey and Recommendations for Cross-Border Cooperation*; TemaNord 2016:508; Nordic Council of Ministers: Copenhagen, Denmark, 2016; p. 508. Available online: https://www.norden.org/en/publication/nordic-digital-identification-eid (accessed on 28 April 2026). [CrossRef]
101. Department of Finance. *Trusted Digital Identity Framework (TDIF): 02—Overview*; Release 4.8; Commonwealth of Australia, Department of Finance: Parkes, Australia, 2023. Available online: https://www.digitalidsystem.gov.au/sites/default/files/2023-07/tdif_02_overview_-_release_4.8_-_finance_1.pdf (accessed on 28 April 2026).
102. Burton, T. Government Smart Wallet Won't Work Without Overhaul: Digital Expert. *Australian Financial Review*, 7 August 2023. Available online: https://www.afr.com/politics/federal/government-smart-wallet-won-t-work-without-overhaul-digital-expert-20230801-p5dswv (accessed on 28 April 2026).
103. Frengley, B.; Teague, V. How Trustworthy is the Trusted Digital Identity Framework? Evaluating Security and Privacy in Australian Digital Identity. Doctoral Dissertation, University of Melbourne, Parkville, VIC, Australia, 2020.
104. Shah, R. *Policy Brief: The Future of Digital Identity in Australia*; Technical report; Australian Strategic Policy Institute: Canberra, ACT, Australia, 2022.
105. Margetts, H.; Naumann, A. Government as a platform: What can estonia show the world? *Res. Pap. Univ. Oxf.* **2017**, *1*, 1–41.
106. Heller, N. Estonia, the Digital Republic. *New Yorker* **2017**, *18*, 12.
107. Metcalf, K.N. How to build e-governance in a digital society: The case of Estonia. *Rev. Catalana Dret Public* **2019**, *2019*, 1–12. [CrossRef]
108. National Audit Office. *Investigation into Verify*; HC 1926, 2017–19; National Audit Office: London, UK, 2019. Available online: https://www.nao.org.uk/wp-content/uploads/2019/03/Investigation-into-verify.pdf (accessed on 14 May 2026).
109. UK Government. UK Digital Identity and Attributes Trust Framework Alpha v2. 2023. Available online: https://www.gov.uk/government/publications/uk-digital-identity-attributes-trust-framework-updated-version/uk-digital-identity-and-attributes-trust-framework-alpha-version-2 (accessed on 10 April 2026).
110. Al-Khouri, A.M. eGovernment Strategies: The Case of the United Arab Emirates (UAE). *Eur. J. ePractice* **2012**, *17*, 126–150. Available online: https://joinup.ec.europa.eu/sites/default/files/16/b2/39/ePractice%20Journal-Vol.%2017-September%202012.pdf (accessed on 14 May 2026).
111. Westland, D.; Al-Khouri, A.M. Supporting e-Government Progress in the United Arab Emirates. *J. e-Gov. Stud. Best Pract.* **2010**, *2010*, 897910. [CrossRef]
112. Samudio, R.E.R. *E-Government Challenges in Smart Societies: The Japanese Experience*; Würzburg University Press: Würzburg, Germany, 2023; pp. 71–88. [CrossRef]
113. Okazaki, S.; Li, H.; Hirose, M. Consumer privacy concerns and preference for degree of regulatory control: A study of mobile advertising in Japan. *J. Advert.* **2009**, *38*, 63–77. [CrossRef]
114. Milberry, K.; Parsons, C. A National ID Card by Stealth? Technical Report. 2013. Available online: https://www.bccla.org/wp-content/uploads/2013/09/BC-Services-Card.pdf (accessed on 14 May 2026).
115. Wolfond, G. A Blockchain Ecosystem for Digital Identity: Improving Service Delivery in Canada's Public and Private Sectors. *Technol. Innov. Manag. Rev.* **2017**, *7*, 35–40. [CrossRef] [PubMed]
116. Digital ID & Authentication Council of Canada (DIACC). Digital Trust and Identity Design Principles. 2021. Available online: https://diacc.ca/the-diacc/principles/ (accessed on 14 May 2026).
117. National Institute of Standards and Technology. *Digital Identity Guidelines*; NIST Special Publication 800-63-4; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2025. [CrossRef]

118. Liu, Y.; Wang, C.; Lu, M.; Yang, J.; Gui, J.; Zhang, S. From simple to complex scenes: Learning robust feature representations for accurate human parsing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 5449–5462. [CrossRef] [PubMed]
119. Wang, C.; Ma, W.; Zhang, S.; Gui, J.; Li, Q.; Liu, Y.; Xia, Z. Focus on finding deepfakes: A robust proactive detection method based on orthogonal moment watermarking. *IEEE Trans. Image Process.* **2026**, *35*, 3507–3521. [CrossRef]