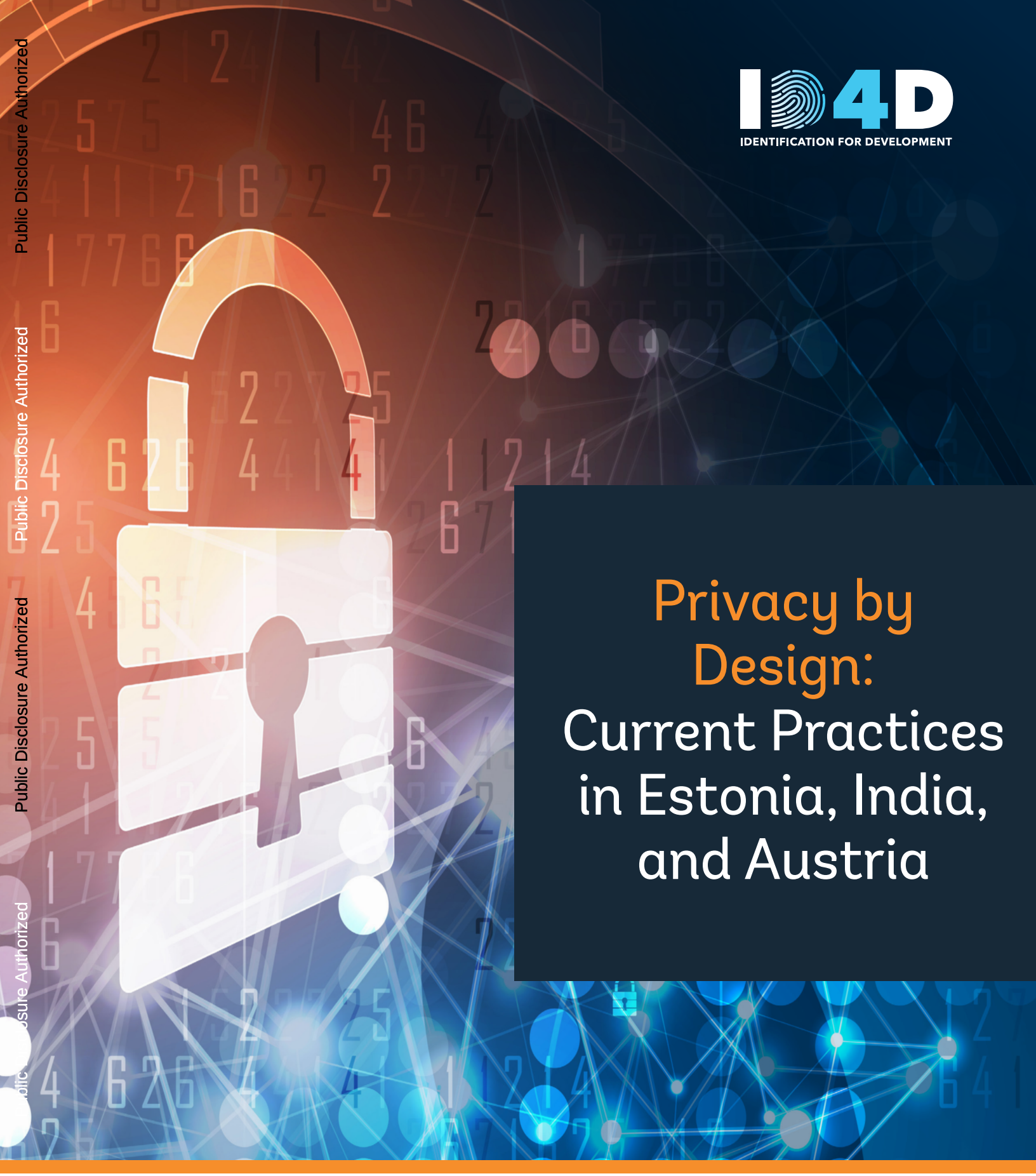
ID4D
IDENTIFICATION FOR DEVELOPMENT



Privacy by Design: Current Practices in Estonia, India, and Austria

WORLD BANK GROUP

1818 H Street, NW, Washington, D.C., 20433
Telephone: 202-473-1000; Internet: www.worldbank.org

Cover images: vs148/Shutterstock.com

Contents

About ID4D

The World Bank Group's Identification for Development (ID4D) Initiative uses global knowledge and expertise across sectors to help countries realize the transformational potential of digital identification systems to achieve the Sustainable Development Goals. It operates across the World Bank Group with global practices and units working on digital development, social protection, health, financial inclusion, governance, gender, legal, and among others.

The mission of ID4D is to enable all people to access services and exercise their rights by increasing the number of people who have an official form of identification. ID4D makes this happen through its three pillars of work: thought leadership and analytics to generate evidence and fill knowledge gaps; global platforms and convening to amplify good practices, collaborate, and raise awareness; and country and regional engagement to provide financial and technical assistance for the implementation of robust, inclusive, and responsible digital identification systems that are integrated with civil registration.

The work of ID4D is made possible through support from the World Bank Group, the Bill & Melinda Gates Foundation, the UK Government, the Australian Government and the Omidyar Network.

To learn more about ID4D, visit id4d.worldbank.org.

Acknowledgments

This report was prepared by Anita Mittal and Rridhee Malhotra with contributions from Maja Andjelkovic, Hannes Astok, Luda Bujoreanu, Julia Clark, Vyjayanti Desai, Sanjay Jain, Jonathan Marskell, Shamus Ozmen, Robert Palacios, Mari Pedak, Ziad Reslan, David Satola, Frank Leyman, Uuno Vallner, Ott Köstner, Siddharth Shetty and Sarah Wolfe.

Abbreviations

API	Application Program Interface
ASA	Authentication Service Agency
AUA	Authentication User Agency
CC	Citizen Card
CCE	Citizen Card Environment
CRR	Central Register of Residents
DPA	Data Protection Authority
GDPR	General Data Protection Regulation
IDEEA	ID Enabling Environment Assessment
MOA	Online Application Module
OECD	Organisation for Economic Co-operation and Development
PbD	Privacy By Design
PET	Privacy Enhancing Technology
PIN	Personal Identification Number
RIHA	Administration system for State Information System
ssPIN	Sector Specific PIN
UIDAI	Unique Identification Authority of India
UIN	Unique Identification Number
UIN	Universal Identification number

Introduction

Digital identification systems, integrated with civil registration, can play a transformational role across many development areas, such as financial inclusion, expanding access to services and social safety nets, and effective humanitarian response. But while the opportunity is great, so are the risks. One set of risks results from collecting, using, and managing personal data, which creates serious privacy challenges. Risks also include:

- Incorrect or inaccurate data collection, leading to mistaken identity or unjust treatment;
- Data collected for one purpose being used for another purpose without the user's consent; and
- Unauthorized or inappropriate transfer of data between government agencies, governments, and even with third non-governmental parties.

The importance of data privacy in building digital ID systems is highlighted in the *Principles on Identification* (Figure 1). These principles have been signed onto by more than 20 international organizations and development partners as being fundamental to maximizing the benefits of identification systems for sustainable development.[1]

Figure 1. Principles on Identification

Inclusiveness: Universal Coverage and Accessibility

1. Ensuring universal coverage for individuals from birth to death, free from discrimination.
2. Removing barriers to access and usage and disparities in the availability of information and technology.

Design: Robust, Secure, Responsive, and Sustainable

3. Establishing a robust—unique, secure, and accurate—identity.
4. Creating a platform that is interoperable and responsive to the needs of various users.
5. Using open standards and ensuring vendor and technology neutrality.
6. ***Protecting user privacy and control through system design.***
7. Planning for financial and operational sustainability without compromising accessibility.

Governance: Trust, Privacy, and User Rights

8. ***Safeguarding data privacy, security, and user rights through a comprehensive legal and regulatory framework.***
9. Establishing clear institutional mandates and accountability.
10. Enforcing legal and trust frameworks though independent oversight and adjudication of grievances.

In particular, as Principles #6 and #8 highlight, there is a need to (i) protect user privacy and control through the design of the digital ID system itself and (ii) ensure the right legal and regulatory framework is in force to enable that protection and control.

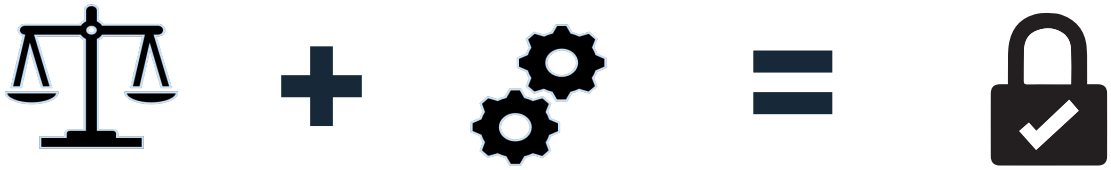
1 Principles on Identification, World Bank Group, 2017.

In order to pursue these goals and get the benefits of digital ID systems while mitigating the privacy risks, the World Bank Group's ID4D Initiative is pursuing a two-pronged approach (Figure 2):

1. **Conduct an ID Enabling Environment Assessment (IDEEA):** IDEEA studies a country's existing administrative and legal framework to identify areas where it can be improved, with a focus on data protection, privacy, and non-discrimination.

2. **Incentivize Privacy by Design (PbD) features:** By signaling to countries and designers of ID systems alike the importance of data privacy, ID4D aims to incentivize incorporating privacy into digital ID systems by default.

Figure 2. Two-Pronged Approach



IDEEA	PbD	
Legal and Regulatory Framework to protect user rights and privacy, e.g., Privacy and Data protection law, enforcement authority	Privacy by Design principles and design techniques and technology, e.g., user consent, authentication, encryption, tamper-proof audit logs, and authorization	A **robust system** that allows sharing of data as per user consent and control ensuring user privacy and data protection

Through the combination of (1) the right enabling legal and regulatory framework and (2) the inclusion of privacy by design as a default feature of digital ID systems, we can obtain a robust system that enables sharing and using user data safely.

But to be able to get to that point, we need to first understand how far countries have already come in enabling privacy by design in their digital ID systems. This note thus is intended to explore the examples of 'privacy by design' features already being deployed in digital ID systems in a few jurisdictions globally, including minimal data collection, randomized unique identity numbers allotment, and tokenization in Estonia, India, and Austria.

The privacy and data protection features in these three countries have been studied through desk research and discussions with technical experts from Estonia and India.

In drafting this report, we also made use of the following frameworks/standards:

i. OECD Guidelines on the Protection of Privacy [2]

ii. Privacy by Design by Ann Cavoukian[3]

2 http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm
3 https://www.ryerson.ca/content/ryerson/pbdce/about/ann-cavoukian.html

iii. ISO/IEC 29100 privacy framework[4]

iv. The privacy principles under General Data Protection Regulation (GDPR)[5]

Finally, it is important to stress that the privacy-by-design features of Austria, Estonia, and India included in this note are not being presented as benchmarks. Instead, the exploration of these country examples is intended to ensure a level of understanding of what currently exists at the country level so that innovation in this field does not end up re-inventing the wheel.

What is Personal Data?

Personal data are any information that relates to an **identified or identifiable living individual.** Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data, for example, a name and surname; a home address; an e-mail address; an identification number; location data (for example the location data function on a mobile phone); and an Internet Protocol (IP) address.

Personal data that has been de-identified, encrypted, or **pseudonymized** but can be used to re-identify a person remains personal data.

Personal data that has been rendered **anonymous** in such a way that the individual is not or no longer identifiable is no longer considered personal data.

Source: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en

What is Data Privacy versus Data Protection?

- **Data privacy** is appropriate use and governance of personal data—things like putting policies and processes in place to ensure that consumers' personal information is being collected, shared, and used in appropriate ways. Data privacy is the right to have control over how your personal information is collected and used, and a system's design should ensure to build privacy into the design of the system.
- **Data protection,** also known as information security, focuses on protecting data from malicious attacks and the exploitation of stolen data for profit. While security is necessary for protecting data, it's not sufficient for addressing privacy. Data protection is an essential aspect of Information Technology (IT) for organizations of every size and type.[6] Various technologies applied to ensure data security include encryption, digital signing, backups, data-masking, and data erasure. Data security also protects data from corruption.

4 http://standards.iso.org/ittf/PubliclyAvailableStandards/c045123_ISO_IEC_29100_2011.zip
5 http://www.privacy-regulation.eu/en/article-5-principles-relating-to-processing-of-personal-data-GDPR.htm
6 https://www.techopedia.com/definition/26464/data-security

How Did the Principles of Privacy by Design Come About?

The exponential growth of the digital economy was accompanied by an increasing awareness that personal data can be misused for nefarious reasons and requires better protection. Among the earliest attempts to do so was a legislative proposal in 1973 by the U.S. Department of Health, Education and Welfare (HEW) that resulted in the issuing of the Fair Information Practices (FIPS).

In 1980, the Organisation for Economic Co-operation and Development (OECD) expanded the principles underpinning FIPS by drafting a set of eight Fair Information Practices codified in the OECD "Guidelines on the Protection of Privacy and Transborder Flows of Personal Data." In the following decade, Dr. Anne Cavoukian of Ryerson University in Canada coined the term 'Privacy by Design,' which she explained through seven core principles and eleven connected practices (Table 1).

The first attempt at a more global set of standards came in the form of the ISO 29100 Privacy Framework Principles, which were built off of existing principles developed by a number of states and international organizations. This was followed by the first set of rules with "teeth" in the form of Article 5 of the General Data Protection Regulation (GDPR) (EU) 2016/679, which was enacted in 2016. This regulation set out principles on privacy and the processing of personal data, and penalties for failing to comply – albeit with applicability only within the European Union. See Figure 3 to view the timeline.

Table 1. Cavoukian's Eleven Fair Information Practices

S.No	PbD principle	Description
1	**Purpose specification**	The purposes for which personal information is collected, used, retained, and disclosed shall be communicated to the individual (data subject) at or before the time the information is collected. Specified purposes should be clear, limited, and relevant to the circumstances.
2	**Collection limitation**	The collection of personal information must be fair, lawful, and limited to that which is necessary for the specified purposes.
3	**Data minimization**	The collection of personally identifiable information should be kept to a strict minimum. The design of programs, information and communications technologies, and systems should begin with non-identifiable interactions and transactions, as the default. Wherever possible, identifiability, observability, and linkability of personal information should be minimized.
4	**Use, retention, and disclosure limitation**	The use, retention, and disclosure of personal information shall be limited to the relevant purposes identified to the individual, for which he or she has consented, except where otherwise required by law. Personal information shall be retained only as long as necessary to fulfill the stated purposes, and then securely destroyed.
5	**Security**	Entities must assume responsibility for the security of personal information (generally commensurate with the degree of sensitivity) throughout its entire life cycle, consistent with standards that have been developed by recognized standards development bodies.

S.No	PbD principle	Description
6	**Accountability**	The collection of personal information entails a duty of care for its protection. Responsibility for all privacy-related policies and procedures shall be documented and communicated as appropriate, and assigned to a specified individual. When transferring personal information to third parties, equivalent privacy protection through contractual or other means shall be secured.
7	**Openness**	Openness and transparency are key to accountability. Information about the policies and practices relating to the management of personal information shall be made readily available to individuals.
8	**Consent**	The individual's free and specific consent is required for the collection, use, or disclosure of personal information, except where otherwise permitted by law. The greater the sensitivity of the data, the clearer and more specific the quality of the consent required. Consent may be withdrawn at a later date.
9	**Accuracy**	Personal information shall be as accurate, complete, and up-to-date as is necessary to fulfill the specified purposes.
10	**Access**	Individuals shall be provided access to their personal information and informed of its uses and disclosures. Individuals shall be able to challenge the accuracy and completeness of the information and have it amended as appropriate.
11	**Compliance**	Organizations must establish complaint and redress mechanisms, and communicate information about them to the public, including how to access the next level of appeal.

Source: https://www.ryerson.ca/content/ryerson/pbdce/about/ann-cavoukian.htm

Figure 3. Privacy Principles Timeline 1973–2016

- Fair Information Practices (FIPS) drafted by HEW
- OECD built upon FIPS to create eight Fair Information Practices
- "Privacy by Design" coined by Dr. Ann Cavoukian 7 principles and 11 fair information practices
- Standard derived from existing principles developed by countries and international organizations
- Article 5 of GDPR on principles for processing personal data

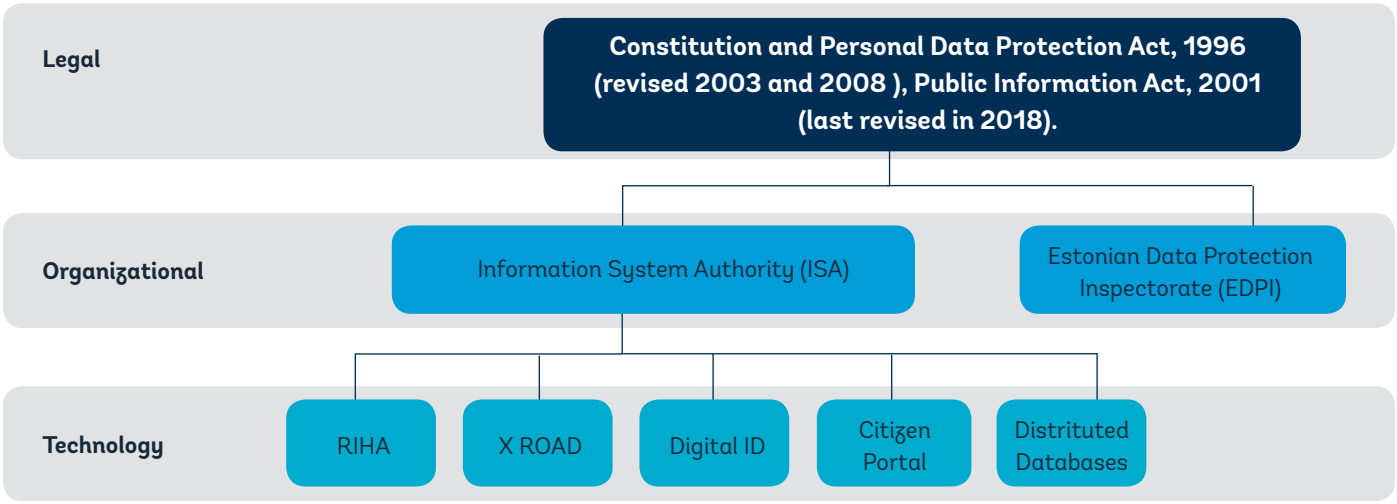
Annex I sets out each of the FIPS, the OECD Guidelines, Cavoukian's principles, the ISO standards and Article 5 of the GDPR. None of the five sets of principles has emerged as the definitive guide followed globally. However, given the overlap and similarities between them, this report will adopt one of the more accessible set of principles—Cavoukian's articulation of eleven fair information practices in assessing country practices—to avoid the repetition inherent in assessing each country practice against all five.

Country Case Studies—Estonia, India, and Austria

Estonia

With a population of 1.3 million, Estonia is one of the most digitally integrated societies in the world, offering its residents 99 percent of public services online, including the ability to vote. Enabling this digital integration in Estonia is a strong legal and regulatory framework supported by robust technology, all of which engenders a high level of confidence among the Estonian public in the country's e-governance systems.

Figure 4. Components of Estonian Privacy Ecosystem



Legal and Institutional

Legal: Privacy in Estonia is recognized under the country's constitution, with three sections specifically setting out (i) the right to privacy, (ii) the right to free self-realization, and (iii) a data subject's right to request information about him-/herself.

The Public Information Act (*Avaliku teabe seadus*) operationalizes the constitutional right to privacy by setting out: i) the conditions for accessing and refusing to grant access to public information; ii) public information for which access is restricted; iii) the conditions for establishing and administering databases; and iv) the mechanism for state and administrative supervision of organization access to public information.

Institutional: The Estonian Data Protection Inspectorate, founded in 1999, is a supervisory authority, empowered by the Data Protection Act, Public Information Act and Electronic Communication Act. The inspectorate's mandate is to protect the following right enshrined under the Estonian Constitution:

- ✓ Right to obtain information about the activities of public authorities;
- ✓ Right to inviolability of private and family life in the use of personal data; and

✓ Right to access data gathered in regard to yourself.

ICT systems

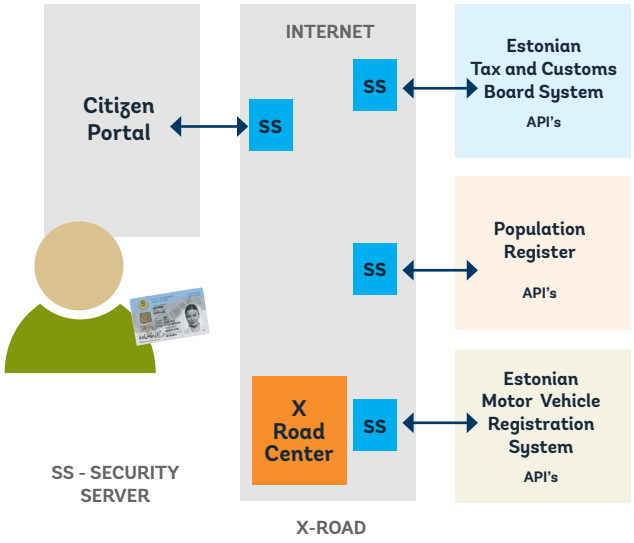
Three ICT building blocks underpin the Estonian e-government system.

1. **Estonia Digital ID:** The key to accessing all public and private services in Estonia is the Estonian Digital Identity card launched in 2002. It is mandatory for every Estonian citizen above the age of 15 and every European citizen residing in Estonia to obtain this identity card. The Estonian digital ID is also available as a mobile ID on mobile devices (special SIM with digital certificates) or as a smart -ID[7] which can be accessed through Android and iOS smartphones.

   Whether you decide to use a physical ID card, a mobile ID, or a smart ID, your credential has two digital certificates—one for authentication of the user and one for digitally signing documents. Access to these certificates on the card or mobile device is secured by a Personal Identification Number (PIN). So even if the card or mobile device is lost, it cannot be used by another user without having the PIN.

2. **X-Road:** if the digital ID is your key, then the X-Road is the vehicle through which your data in different databases can be securely transferred. The "X-Road" is basically a data exchange platform that allows the nation's various databases (236 as of 2017), both in the public and private sector, to securely exchange data. The data is distributed across these databases linked by the Unique Identification Number (UIN), called Personal Identification Code in Estonia. The process is explained in Figure 5.

Figure 5. Estonian e-governance Systems Architecture



1. A user wanting to use an online service "XYZ" of Department "ABC" authenticates their identity by using the citizen portal using their digital ID (smart card or mobile ID). (Single sign -on solution enables the user to request service from any department seamlessly.)
2. Using X-Road, the service being accessed itself obtains the data needed to process the service request from other databases.
3. The Security Server component of the requesting system encrypts the data and sends it to the system (database) from which data are desired over the Internet.
4. The security server at the data provider system end authenticates the requesting system, and if the authorization check succeeds, forwards the request to the system.
5. The security server time stamps, digitally signs, and logs the transaction and sends an encrypted response, provided by, the data provider system to the requestor system.
6. The security server decrypts the response, and then the service processes the request based on data fetched in real-time and returns the response to the user.

7 Citizens' portal does not allow to enter with smart ID directly.

3. **RIHA (Administration system for State Information System):** RIHA serves as a catalogue for the state's databases and gives information on the following:

   - The information systems and databases that make up the state's information system;
   - Data collected and processed by these information systems;
   - Services, including X-Road services, provided by these information systems and the list of users (organizations) of these services;
   - Responsible and authorized processors of the information systems and databases and services and contact details of these individuals;
   - Legal basis for the database operations and processing; and
   - The reusable components that ensure the interoperability of information systems (XML assets, classifications, dictionaries).

Table 2. Evaluation of Estonian Privacy by Design Features Based on Cavoukian's Eleven Fair Information Practices

S. No	Privacy principle	Estonian system
1	**Purpose specification**	• A central authority, the Estonia Information Authority, explicitly defines the scope and data sharing rules from different systems/databases that are required to fulfill the data requirements of a given service. • Only the **minimal data** required for the purpose of delivering a service is permitted to be shared by the systems with the requesting system. • X-Road, the data exchange platform, ensures that only authenticated and authorized systems can access the data from another system.
2	**Collection limitation**	• **Minimal data are collected for ID** issuance, namely. full name, contact details (address, phone, e-mail), date and place of birth, personal identification number, citizenship, gender, and photograph. Biometric data are only captured if a passport is also issued along with ID, but biometric data are not used for identification or authentication. • Estonia follows the **"once only principle,"** which requires that the citizen not be required to share the same data twice with authorities. The data of the citizen are referenced/ accessed in real-time for the different systems for service delivery.
3	**Data minimization**	• **UIN (Universal Identification Number)** an 11-digit number, referred to as a **Personal Identification Code in Estonia**, consists of gender, century of birth, date of birth, serial number separating persons born on a same date, and a checksum. It is simple, both from technical design (easy to generate) and usability perspective (easy to remember), but it does **reveal gender and date of birth** information of the individual. This is in contrast to random number UINs adopted by other countries. • **Link**—The data of a citizen is **linked** across systems/**databases through UIN** which provides complete 360 views of the user. Estonia has leveraged this knowledge for efficiencies in service delivery with due regard to user privacy and data protection. The privacy and data protection risks are mitigated through strong technology-enabled checks and controls, and a strong regulatory environment.

S. No	Privacy principle	Estonian system
		• **Access control (authentication and authorization):** Data sharing is allowed after authentication and authorization of the requesting entity and is encrypted in transit to protect against attacks. For example, if a citizen accesses a parental benefit service which needs data from the social insurance system, the social insurance system will only share data fields which have been authorized by the central authority for the parental benefit service. • While logging the transactions, **pseudonymizing** of personal data is done to protect privacy. • When the data are shared/used for analytics, **anonymization** of data is done to protect privacy.
4	**Use, retention and disclosure limitation**	• Data that can be accessed from other databases is reviewed and approved by the Central Estonian Information Authority. All data access is based on authorization granted by that authority. The automated logging of transactions ensures that no user, can access data without leaving their footprint in the logs.
5	**Security**	• All the data are **encrypted** and **digitally signed** and transmitted over a secure communication channel when sharing data between systems. • The transaction logs are time stamped and digitally signed making them **tamper proof.** Any change would make the digital signature invalid. • The logs are **hash chained** to further enhance the immutability of the logs and make changing logs at a later date an extremely tedious and difficult task. • **Block chain technology** has also been deployed for enhancing the integrity of transaction logs. The log is pseudonymized to protect privacy but can trace the user in the event a transaction is contested. • As the authentication holds the key to all data of the user, strong **multi-factor authentication** using the smart card with digital certificates (something you have) + PIN (something you know) ensures that someone else cannot impersonate you. • The **machine to machine interaction** involving data access by one system from another system or user-initiated data access is allowed only after **authentication and authorization** of systems using digital certificates and access rules (done by the security servers of X-Road).
6	**Accountability**	• The **automated tamper proof logging** of transactions which are performed after successful authentication of users/systems holds people/systems accountable for data access. Internal system users also continuously monitor these logs/reports for violations. • The **data protection laws** stipulate heavy penalties for unauthorized access to data. There have been reported cases of punishment of law enforcement officers for using unauthorized data access for personal gains. There is also a **dedicated data protection authority** to handle grievances and complaints. • The users are **notified through registered e-mail/SMS** of authentication **attempts and data breaches, if any.**

S. No	Privacy principle	Estonian system
7	**Openness**	• Access to the **time stamped and digitally signed** logs of all transactions where the user data were accessed by the X-Road system (which is a third-party system) ensures authenticity, integrity, and non-repudiation of transactions. The system/database owners cannot choose which transactions to log or hide/delete transactions. • **Personal data monitor,** an Artificial Intelligence (AI) enabled software that filters and logs transactions containing personal data, is used at the exit points of information systems from where the data flows to other systems. (see box on *Personal Data Usage Monitor* for details). • This independent software can capture transactions containing personal data based on rules defined to identify personal data in data traffic flowing out of a system and log them with time stamp and digital signature. These logs are then accessible to the users via the citizen portal. • **Block chain technology** has also been deployed for integrity of transaction logs.
8	**Consent**	• User **authentication** while accessing a service serves as consent to the service provider to access data from other databases. • The authentication and transaction logs at each system are time stamped to match the consent of the user with data access.
9	**Accuracy**	• Users can **view and update** their information on the citizen portal to keep it accurate. • "Once only principle" is the definition of data ownership for each data element, **real-time sharing of data** across systems instead of each system maintaining their copy facilitates consistency/accuracy of data across systems.
10	**Access**	• On the citizen portal, the users can **access history** of when and where his/her personal data was accessed. Users can also contest any data access that they did not authorize.
11	**Compliance**	• Tamper proof **transaction logs,** independent (certified) personal data monitor software, internal monitoring by system users of the systems, and availability of access history to the citizens helps in maintaining and **demonstrating** compliance with the data protection and privacy laws.

User Consent and Choice—eHealth System

The core of Estonia eHealth is the Digital Health Record system, using HL7 and DICOM message formats for interconnection. The data transport and security layer are provided for by the "X-Road" middleware software.

Patients can view all their health care data through the Estonian eHealth Patient Portal by using their digital ID to authenticate their identity. By default, medical specialists can access data, but any patient can choose to deny access to care providers, including one's own general practitioner/family physician.

Others, such as pharmacists and insurance agents, can get access to a patient's medical records, but only with the patient's explicit knowledge and consent. All data access requests within the system are recorded, and patients can on request access this record.

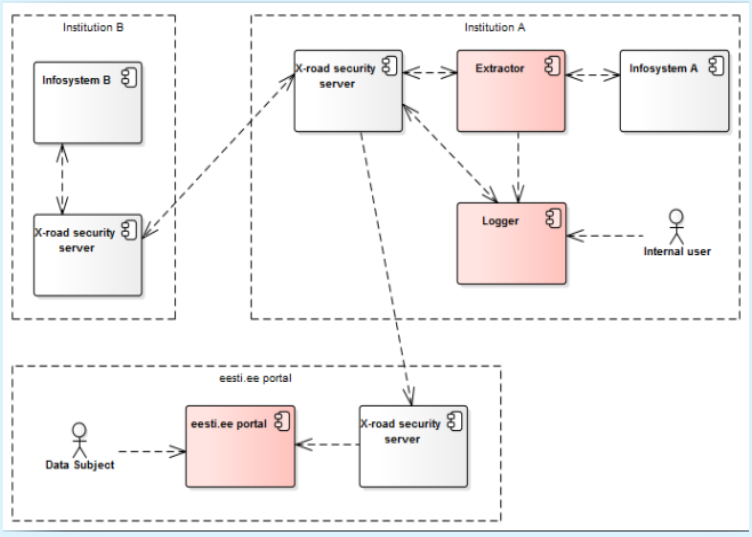
(*Source:* https://e-estonia.com/solutions/healthcare/e-health-record/)

Personal Data Usage Monitor

Estonian citizens and residents have the ability to monitor how their data have been used by the government. Personal Data Usage Monitor (open source software on Github) offers residents a comprehensive view of how his or her personal data has been used by the government. There is a component (extractor) which automatically flags outgoing messages which contain personal data.

A personal data usage log record is created for each time a resident's data are accessed. The log record contains metadata about personal data usage. The logger component logs this event in time stamped, digitally signed tamper proof logs. These logs can then be accessed by the user (via the citizen portal) to check for any unauthorized usage of personal data. Internal system users also check the logs to monitor the activity and flag anomalous behavior for preventive and corrective measures.

(*Source:* https://github.com/e-gov/AJ/blob/master/preliminary/Overview.md)



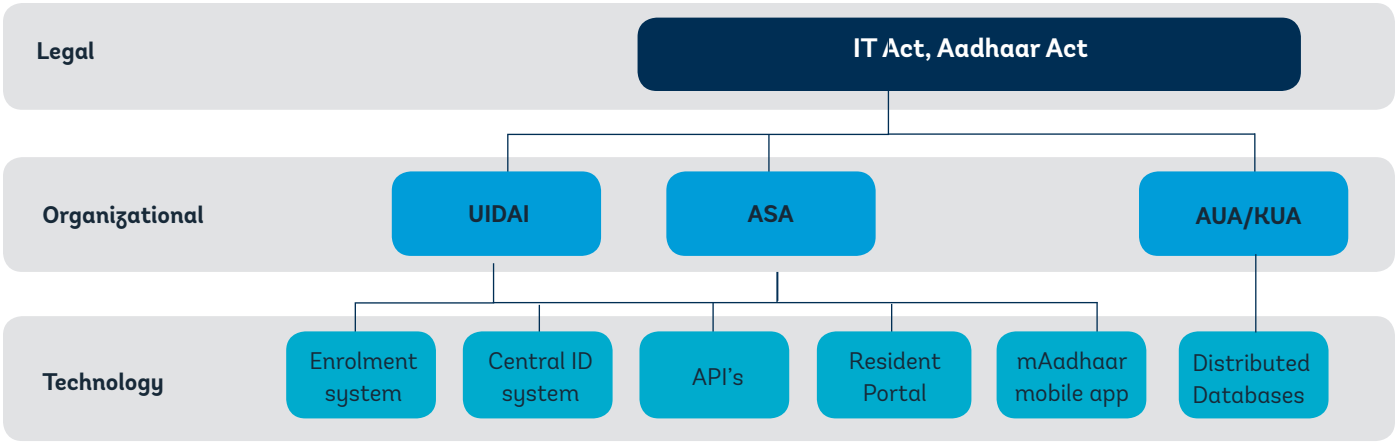
India

With a population 1,000 times the size of Estonia's, India can lay claim to having the world's single largest biometric-based digital identification system called Aadhaar (which roughly translates to "Foundation" in English). Enrolment in Aadhaar requires the collection of ten fingerprints, two iris images, and a digital facial photograph, along with basic biographic data (name, date of birth, sex, and address). Once these four requirements have been met, a randomly generated number is given to the user. Unlike other systems, including Estonia's, there is no physical credential that is provided as a result of signing up for Aadhaar. Instead, all authentication is done using biometrics.

One important distinction to note about India's digital ID system is that it was introduced *before* a legal and regulatory framework was enacted. This has led to intense debate and culminated in legal challenges to the constitutionality of the system altogether that have gone all the way up to the Indian Supreme Court.

The Data Empowerment and Protection Architecture for sharing data was formulated in the end of 2017. At the time of this writing, a draft Personal Data Protection Bill has been published (July 2018). These two pieces of legislation aim at addressing the legal and technical aspects of data protection and privacy, not only for the identification system but for all types of personal data. Because of the state of flux of the regulatory environment underpinning Aadhar in India, this note will focus solely on the technical aspects of the system itself.

Figure 6. Components of Indian Privacy Ecosystem



Legal and Institutional

Legal: The Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act 2016,[8] aims to provide a legal basis for Aadhaar, the unique identification number project which was launched in 2009. It was passed on 11 March 2016 by the Lok Sabha, the federal Indian Parliament.

Institutional: The Unique Identification Authority of India (UIDAI) was created with the objective of issuing Unique Identification numbers (UIDs), called "Aadhaar" numbers, to all residents of India. The unique number and its ecosystem were aimed at eliminating duplicate and fake identities and allowing verification and authentication in an easy, cost-effective way. UIDAI started as an attached office of the Planning Commission in 2009 and later in 2015 was attached to the Ministry of Electronics and Information Technology.

8 https://uidai.gov.in/images/the_aadhaar_act_2016.pdf

The Aadhaar ecosystem includes the following entities and actors:

1) **ASA:** Authentication Service Agency (ASAs) offer their UIDAI-compliant network connectivity as a service to requesting entities to transmit their requests to web services of the ID system Agencies interested in providing ASA services have to fulfill requirements as laid out by UIDAI and get approval from UIDAI. (UIDAI, n.d.)

2) **AUA/KUA:**[9] Authentication User Agency (AUA), a government/public/private legal agency registered in India, engaged in providing Aadhaar Enabled Services to Aadhaar number holders using authentication services of UIDAI. An AUA/KUA connects to the Central Identity Repository (CIDR) system through an ASA. AUA/KUA fulfill the laid out procedural requirements of UIDAI.

ICT System

The key components of the Aadhaar system include the following:

1) **Enrolments Software:** The enrolment software, owned by UIDAI, captures demographic information and biometric data with the consent of the user obtained at registration. The software then securely transmits that information to the Aadhaar system.

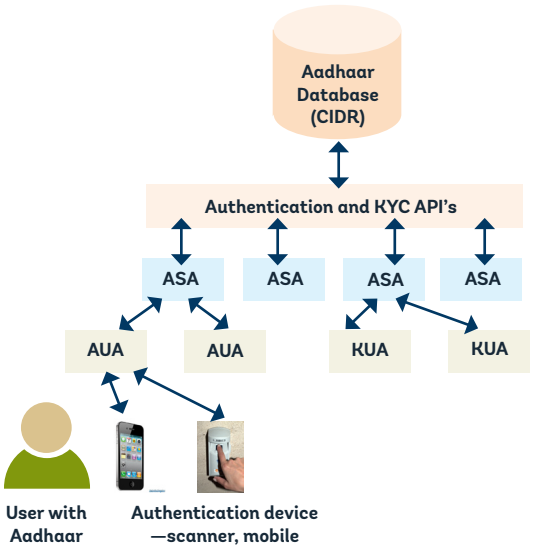
2) **CIDR:** The Central Identity Repository system stores the demographic and biometric data after issuance of the Unique ID number (Aadhaar number).

3) **Aadhaar services/APIss:** UIDAI has **open APIs** to allow service providers in the public and private sector to authenticate users based on one or more of the following: biometrics, demographics, and One Time Password (OTP) on registered mobile phones. The service providers must register as **AUA/sub AUA** with UIDAI and access the APIs via the **ASA.**

   eKYC service shares the demographic data and the photograph of the user with the service provider when the user provides consent. This enables the onboarding of users for services such as opening bank accounts, getting up a SIM card, etc. Recently, UIDAI has announced that a limited eKYC API would also be made available for a category of service providers with limited KYC data.

9 KUA–KYC User Agency can invoke Know Your Customer webservice which returns demographic data and photo.

Figure 7. Aadhaar Authentication Ecosystem



1. The user with an Aadhaar number presents the Aadhaar number or Virtual ID number and a biometric or OTP to the service provider (AUA).
2. The encrypted biometric from the UIDAI certified biometric device is packaged by the AUA as per the API specification and sent to ASA.
3. ASA transmits this packet over a leased line and invokes the authentication API of the Aadhaar system.
4. The API checks the incoming data against the CIDR and returns a YES/NO response based on the result of the match.
5. This response is conveyed by ASA to AUA and onwards to the user. AUA provides the service when the response is YES.

Table 3. Evaluation of India's Aadhaar Privacy by Design Features Based on Cavoukian's Eleven Fair Information Practices

S. No	Privacy principle	Indian system
1	**Purpose specification**	During the enrollment process, **the purposes** for which the data collected may be shared by UIDAI are explained and user consent and choice for sharing data captured. Demographic data sharing for electronic KYC is allowed for customer onboarding purposes only and requires user consent.
2	**Collection limitation**	**Minimal data are collected for enrolment for ID, namely** full name, address, date of birth and gender. The system also collects **multimodal biometrics**, namely. photograph, two irises, and ten fingerprints.
3	**Data minimization**	**Zero Semantics UIN**—The UIN (Aadhaar number) is a **random number** and does not on its own convey any meaning/information about the user. Linking of data across various systems/databases up until early 2018 was based on the Aadhaar number. In the light of increasing privacy and surveillance concerns, UIDAI recently launched the **virtual ID and tokenization features** discussed further in Box 1. • With the **tokenization** feature, instead of a UIN, a token which is calculated based on service provider code and Aadhaar number is used to identify the user in the service provider database, thus avoiding linkability of data across databases. • A **virtual ID** is a temporary ID number mapped to the UIN, that can be generated and used by a user instead of exposing the Aadhaar number. **Fingerprint and iris data are never shared.** **UIDAI certified biometric devices** used by service providers for authentication encrypt **the biometric** data captured from the user in the device itself before it reaches the service provider system, thus securing it. Data, when used for analytics purposes, are **anonymized** before sharing.

S. No	Privacy principle	Indian system
4	**Use, retention and disclosure limitation**	The biometric data are never shared by UIDAI with anyone. The biometric data provided by the user for authentication are not available to the service provider application as it is encrypted in the certified biometric devices before reaching the service provider application. KYC data consisting of demographic data and a photograph are shared by UIDAI with registered service providers only with user consent. No data are disclosed during authentication of users by UIDAI. Only a YES or NO response on an authentication request is given to service providers. The electronic Aadhaar, which is digitally signed by UIDAI, can be generated by the users on the portal. This can be used for offline authentication and KYC. The users can choose the demographic fields to be included in electronic Aadhaar, which enables them to limit the data that will be visible and shared with the service provider.
5	**Security**	The digital signature on the electronic Aadhaar ensures integrity and authenticity of the electronic Aadhaar document—enabling detection of any forgery of the Aadhaar document. This security feature enables it to be used for offline authentication and identity verification with a higher level of assurance. The Aadhaar also has two digitally signed Quick Response (QR) codes, one with photograph and demographic data and the other with demographic data only. The QR code, in both electronic Aadhaar or printed Aadhaar document, can be used for electronic capture of demographic data during offline KYC/authentication. The QR code prevents the data to be read visually without a QR code reader and the digital signature validation of the QR code enables detection of fake /forged Aadhaar's. All the data are **encrypted** and **digitally signed** and transmitted over a secure communication channel when sharing data between systems. Data is stored in **encrypted f**ormat and not exposed/available even for admin user or other type of user in plain text format The transaction logs are time stamped and digitally signed making them **tamper proof.** Any change would make the digital signature invalid. Users and systems are **authenticated and authorization** rules enforced before providing access to services (API's) or administrative functions. **Data tampering** is prevented by ensuring that data updates can only be done by authorized applications and not through command line queries/scripts. **Data is partitioned** and held in multiple database systems, with a random alias being the only link, which ensures that there is no centralised data table where all resident data is available. Access to the API's and hence to the CIDR is through a **network of trusted service providers (AUA and ASA)** only. Users can **lock/unlock** the use of **biometrics** to disable/enable biometric-based authentication.

S. No	Privacy principle	Indian system
6	**Accountability**	The **automated tamper proof logging** of transactions performed after successful authentication of users/systems holds people/systems accountable for data access. Internal system users also continuously monitor these logs/reports for violations. Users are **notified through registered e-mail/SMS of authentication attempts,** though access to e-mail/phones limits the universality of this feature. The **Aadhaar Act and IT Act** stipulates heavy penalties for unauthorized access to data. AUA and ASA sign agreements with UIDAI to access UIDAI services.
7	**Openness**	Access to the **time stamped and digitally signed** logs of all transactions where the user data was accessed ensures authenticity, integrity, and non-repudiation of transactions. The **resident portal** provides information regarding policies and procedures of data sharing and other information to the citizens.
8	**Consent**	User **authentication** while accessing a service serves as consent to the service provider to access data from Aadhaar. Only trusted registered services (AUA/KUA) can access the Aadhaar system API's and can access it only through a trusted secure network of service agencies (ASA). User consent is captured during registration for digital ID on paper forms.
9	**Accuracy**	Users can **view and update** their information on the Aadhaar resident portal and through various other channels to maintain its currency/accuracy.
10	**Access**	On the resident portal, citizens can **access authentication history** of when and where authentication was attempted. If they find authentication attempts to which they had not consented, they are able to contest these occurrences.
11	**Compliance**	Tamper proof **transaction logs** and availability of access history to the users help in maintaining and **demonstrating** compliance with the data protection and privacy laws.

Box 1. Virtual ID and Tokenization

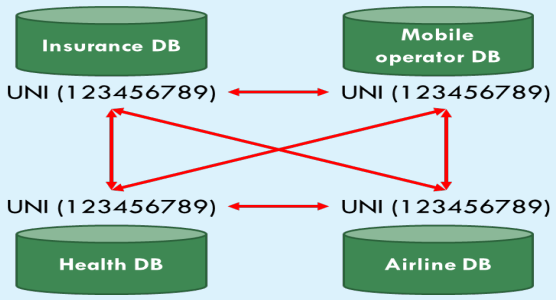
Virtual ID: A 16-digit random number is mapped to an Aadhaar number. Once you have generated a Virtual ID, you can provide that 16-digit number, instead of your Aadhaar number, to any agency seeking to use your Aadhaar number for authentication.

A key privacy-enhancing aspect is that the Virtual ID is temporary and revocable. This means that service providers cannot rely on it or use it for correlation across databases. The users can change their Virtual ID at will just as one would reset their computer password/PIN.

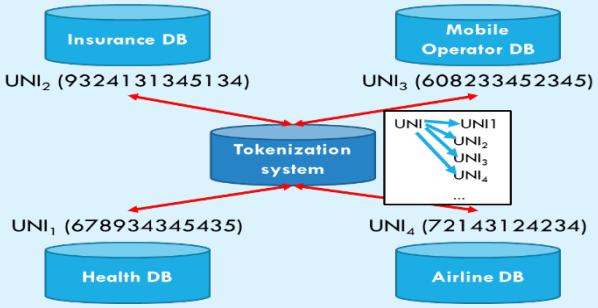
Tokenization: When a user gives Aadhaar/Virtual ID for authentication, the ID system generates a unique **token** (72 char alphanumeric code) which is specific to that agency and Aadhaar number. Different agencies will be given different tokens to identify the same person in their system, thereby eliminating the linkability of information in the databases based on an Aadhaar number. Only the Aadhaar system knows the mapping between the Aadhaar number and the tokens provided to the service providers.

Service providers or AUA's will be categorized as **global AUA** or **local AUA.** Global AUAs are allowed to store and use Aadhaar numbers and use **full** eKYC API which returns an Aadhaar number along with the token. On the other hand, local AUAs are only allowed to use **limited** eKYC API and use the token to identify the user instead.

(*Source:* https://uidai.gov.in/resources/uidai-documents/circulars,-notifications-office-memorandums.html)



Databases using UIN

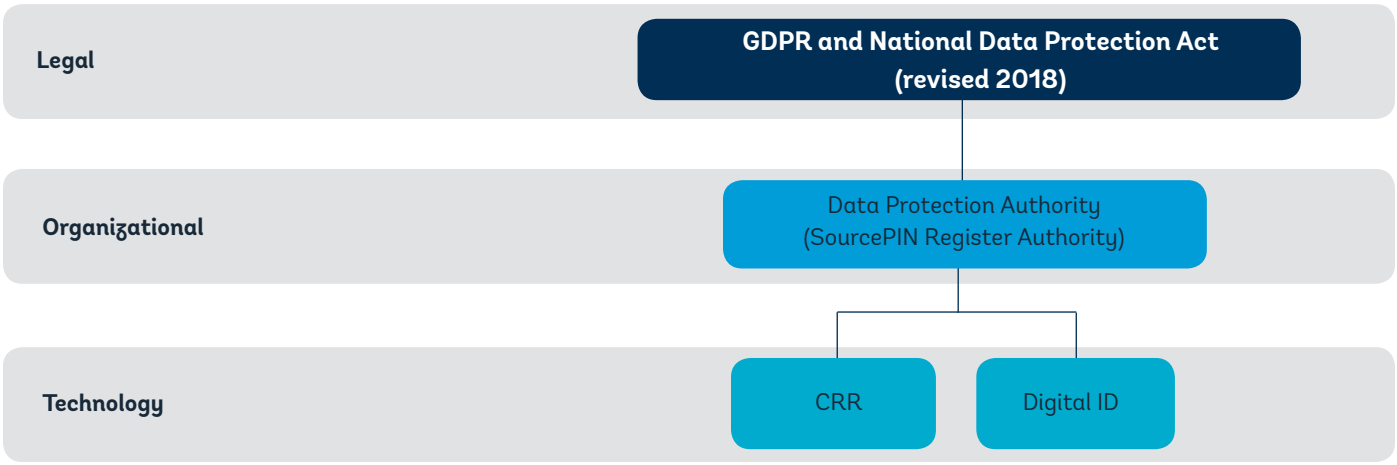


Databases using tokenization

Austria

With a population of 8.7 million, Austria was one of the first countries to implement a national ID system that enables residents to access public services online using an electronic ID card. Tokenization is the focus of study in the country case study of Austria and hence a detailed assessment is not presented with reference to the Ann Cavoukian's Fair information practices. Austria's tokenization - privacy by design features is analogous to the Indian one which uses virtual IDs and tokenization. However, unlike India's centralized ID authentication system, Austria's system is decentralized.

Figure 8. Components of Austrian Privacy Ecosystem



Legal and Institutional

Legal: As an EU member, Austria is subject to Article 5 of the GDPR discussed above. The GDPR is supplemented and locally implemented through the Austrian Data protection Act (Datenschutzgesetz, short DSG). Rounding up the legal framework for privacy in Austria is the SourcePIN Register Authority Regulation passed in 2009, which sets out the role and responsibilities of the SourcePIN Register Authority responsible for citizen ID cards and cooperation with service providers.

Institutional: The Austrian Data Protection Authority (DPA) is an independent authority entrusted with protecting individual rights and interests in the privacy of personal data. The functions of the SourcePIN Register Authority set out in the 2009 Regulation are carried out by the Data Protection Authority.

ICT systems

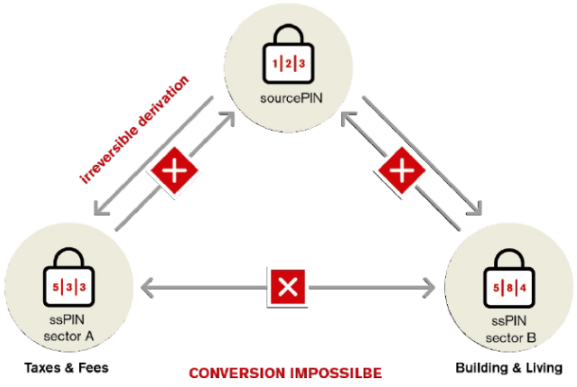
The **Central Register of Residents (CRR)** is a national information system that contains data about every resident of Austria (citizen and non-citizen alike). Austria mandates that all residents register their presence in the country, and the CRR contains the records of all these registrations. Each data record in the CRR has a unique identifier—a 12-digit CRR number, resident's full name, sex, date of birth, citizenship, and full address. Records of foreigners additionally contain passport data.

While registration is mandatory, there is no equivalent requirement that every resident obtain a physical "ID card." Instead, Austria has a virtual **Citizen Card (CC)** which can be installed on different devices (e.g., smart cards, USB devices, mobile phones), with smart cards and mobile signature[10] being the two most prevalent interfaces used.

10 https://www.digital.austria.gv.at/mobile-phone-signature

The data contained on a CC is called **Identity Link** and consists of full name, date of birth, the source PIN (as unique identifier created by strong encryption of the CRR number), and the cryptographic keys required for e-signature and encryption. To ensure integrity and authenticity, the Identity Link data structure is digitally signed by the SourcePIN Register Authority at issuance. Access to SourcePIN and cryptographic keys on CC is protected by a PIN.

The Austrian public administration is divided into 26 sectors, including tax, health, education, etc. To safeguard user privacy, the eGovernment Act stipulates that different identifiers be used for each sector data system that a user accesses. A sector-specific personal identifier (ssPIN) is created from the SourcePIN using one-way derivation, a method through which a sector specific-PIN is algorithmically computed from the SourcePIN.

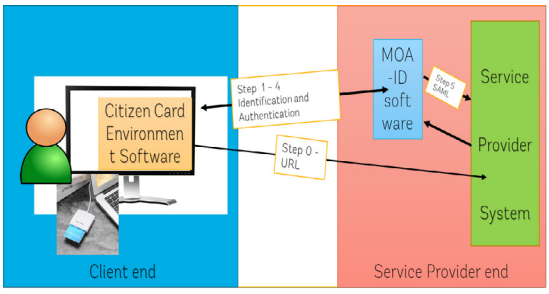


Source: Digital Austria, Federal Chancellery

Handling and usage of the e-ID

In order for a resident to use a CC, they need the activated CC, a card reader, a PC connected to the Internet, and special software (Citizen Card Environment [CCE]) at the user end. Special software "MOA-ID" is needed at the service provider end that helps with authentication.[11]

Figure 9. Process Flow for Authentication



1. When a user requests a service, MOA-ID checks the integrity and authenticity of the CC.
2. MOA-ID calculates the ssPIN by applying a cryptographic hash function H (SHA-1) to the concatenation of the SourcePIN and the sector-specific identifier of the service provider.
3. MOA-ID requests consent of the user for the service by requesting electronic signature of the user.
4. MOA-ID verifies the citizen's qualified signature.
5. The user can now avail the service (MOA-ID sends SAML response to service provider system).

11 Slamanig, 2013.

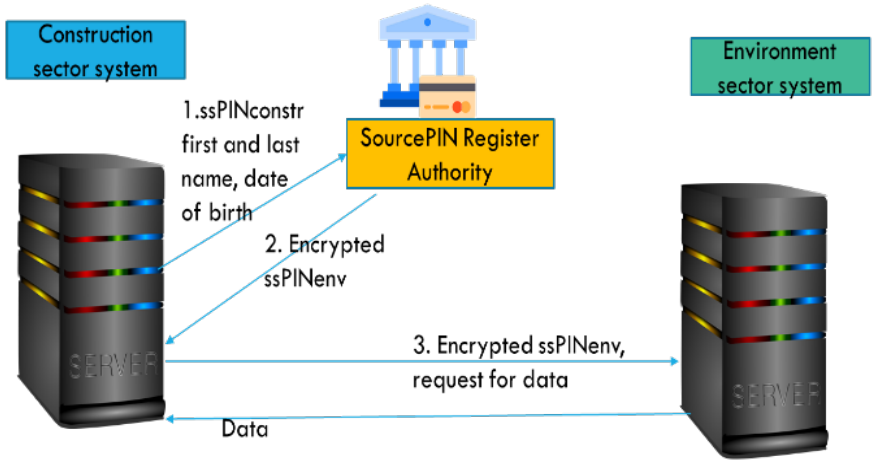
Data Sharing within the sector (Barotanyi, 2017)

Unlike the SourcePIN, the ssPIN can be stored in administrative procedures. Public authorities can use the same ssPIN to retrieve a citizen's data stored within the same procedural sector, for example, if they need to view the citizen's records or use it to pre-fill forms. However, authorities do not have access to ssPINs from other sectors.

Data Sharing across sectors

Administrative procedures often require authorities from different sectors to work together. For example. If an authority requires a sector-specific person identifier from another procedural sector in order to identify a natural person, they can request it from the SourcePIN Register Authority by providing the ssPIN from their own procedural sector, the person's first and last name, and their date of birth. The SourcePIN Register Authority sends the desired ssPIN to the authority that requested it in encrypted form, and the ssPIN can only be decrypted by the public authority that is responsible for the other procedural sector.

Figure 10. Cross Sector Data Exchange - Austria



Concluding Thoughts

Data privacy is top priority, as the extent to which personal data are driving growth in the digital economy becomes clearer. With the rollout of digital identification systems, there is a unique opportunity to ensure that privacy is embedded at the onset into these systems, as opposed to having it be an afterthought, as has been the case in many developed countries.

Given that context, this note is intended to present the most pertinent privacy by design principles and standards—and then to look at the actual implementation of these principles at the country level in models as diverse as Estonia's and India's. It is our hope that the juxtaposing of privacy principles and actual privacy by design features/regulatory frameworks already implemented globally will spur further conversations about developing these systems. Privacy by design is ripe for innovation and extrapolation to all digital ID systems.

That said, there is still a number of areas that merit further exploration, including:

a. Assessing/evaluating different approaches to privacy by design;

b. Assessing design techniques from the perspective of usability, maturity, ease of implementation, etc.; and

c. Studying existing and emerging privacy enhancing technology (pet) such as homomorphic encryption, attribute based credentials, and hardware storage modules.[12]

12 Two good sources for additional information on Privacy Enhancing Technologies are (a) technologies for enhancing security and privacy in Aadhaar and (b) privacy enhancing strategy, design and technology report.

References

Barotanyi, B. E. 2017. *The ABC guide of eGovernment in Austria.* Vienna: Austrian Federal Chancellery, Federal Platform Digital Austria.

ID4D WBG. 2016, January. http://documents.worldbank.org/curated/en/213581486378184357/pdf/112614-WP-REVISED-PUBLIC.pdf. Retrieved from World Bank.

Slamanig, B. Z. 2013. On Privacy-Preserving Ways to Porting the Australian eID System to the Public Cloud. *FIP Advances in Information and Communication Technology*, (pp. 300–314).

UIDAI. (n.d.). uidai.gov.in. Retrieved from UIDAI.

WBG and CGD. 2018, February. http://documents.worldbank.org/curated/en/213581486378184357/pdf/112614-WP-REVISED-PUBLIC.pdf. Retrieved from World Bank.

Annex I

Annex Table 1. Principles Mapped across Four Frameworks

S.No	OECD—Guidelines on the protection of Privacy -1980	Ann Cavoukian—Fair information practices 1990's	ISO/IEC 29100 Privacy Framework—2011	GDPR (General Data Protection Regulation)—2016
1	Purpose specification	Purpose specification	Purpose legitimacy, and specification	Purpose limitation
2	Collection limitation	Collection limitation	Collection limitation	
3		Data minimization	Data minimization	Data minimization
4	Use limitation	Use, retention, and disclosure limitation	Use, retention, and disclosure limitation	Storage limitation
5		Consent	Consent, and choice	Lawfulness, fairness, and transparency
6	Data quality	Accuracy	Accuracy, and quality	Accuracy
7	Individual participation	Access	Individual participation, and access	
8	Openness	Openness	Openness, transparency, and notice	Lawfulness, fairness, and transparency
9	Accountability	Accountability	Accountability	Accountability
10		Privacy compliance	Privacy compliance	
11	Security safeguards	Applied security	Information security	Integrity, and confidentiality

id4d.worldbank.org