Cyber Studies PROGRAMME

UNIVERSITY OF OXFORD

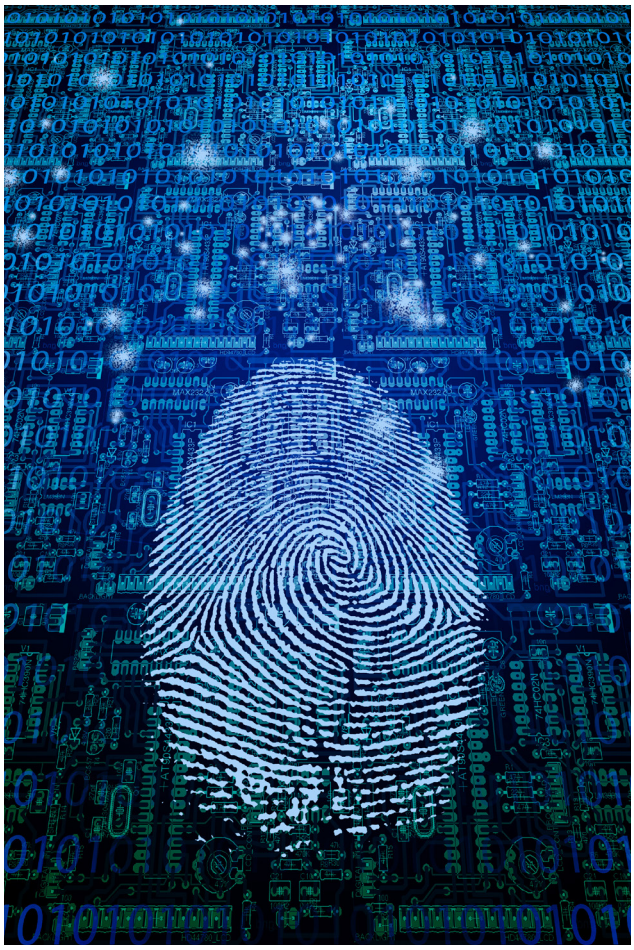
www.politics.ox.ac.uk/centre/cyber-studies-programme.html

Working Paper Series – No. 4 April 2016

# Security and Privacy Impacts of a Unique Personal Identifier

**Andrew Martin**
andrew.martin@cs.ox.ac.uk
**Professor of Systems Security**
**University of Oxford**

**Ivan Martinovic**
ivan.martinovic@cs.ox.ac.uk
**Associate Professor of Computer Science**
**University of Oxford**



## ABSTRACT

National Identity Schemes are highly complex socio-technical systems in which many competing requirements from diverse stakeholders must be balanced. From a technical systems perspective, we review the objectives of such schemes, along with the resulting requirements, particularly the strong need to ensure appropriate levels of privacy and security. These objectives and requirements are addressed within the context of the broad range of threats against the system. Considering available technologies and ideas, we explore the design choices that must therefore be made in designing such a scheme and illustrate these choices by discussing a number of existing identity schemes. The Estonian scheme is considered in a greater level of detail, evaluating it against the requirements and design choices of other nations as well as drawing on empirical data where possible to explore whether the issues of theoretical or hypothetical importance emerge as realistic concerns in a large deployed system.

European Union
European Social Fund

Investing
in your future

This publication is funded by the European Social Fund and the Estonian Government

# INTRODUCTION

There are a variety of reasons for establishing identity schemes, whether in the public sector (for government purposes) or the commercial sector, and whether exclusively in the digital domain (sometimes called "eID") or also in the analog embodiment (as with conventional ID cards and passports).[1] The development of generic schemes that are capable of being adapted to many purposes and used in many services brings many benefits to the citizen and the customer. If the schemes are well designed, an open environment is created where additional services and capabilities can be added as needs and opportunities arise. Such schemes are inherently accompanied by privacy concerns; designs must account for these concerns while also balancing the privacy benefits that emerge, for example, from avoiding the existence of multiple copies of personal data in diverse systems.

In the creation of distributed systems—increasingly not just computer systems but also complex cyber-physical systems—many requirements and design decisions need to be coordinated. This is particularly true with regard to those components which address issues of security and privacy. It is well known that a particular sub-system or component, taken in isolation, may exhibit excellent (or terrible) security characteristics, yet may still contribute to a larger system with much poorer (or better) security than expected.

This study addresses these issues with particular reference to national-scale identity schemes. States have long issued identity tokens and used unique identifiers to identify tax payers, social security recipients, voters, and more—but the scale of modern e-government, the scope for efficiency and citizen benefits from linking diverse systems together, and the desirability of strong identity technologies for many kinds of commerce mean that these schemes are under fresh scrutiny. Hitherto, many schemes have had narrow purposes; with an increasingly diverse range of online services in the public and private sectors, the main benefit comes from schemes that permit open integration of diverse applications. Meanwhile, renewed fears about domestic and foreign surveillance as well as intrusive data-mining of customer profiles in the commercial sector act as countervailing forces against a tendency to try to connect everything to everything else.

The challenge of designing identity systems that offer maximal usability and utility, yet also display adequate security in the long term and respect citizens' rights to privacy, is an enormous one. It would be easy for a scheme to founder on a mistaken perception or assumption, on a subsystem weakness that is actually insignificant at the large scale, or due to a collection of excellent components being assembled in an unfortunate way. Such concerns affect both policymakers and technologists, all of whom need to understand the consequences of their design decisions and how those consequences vary across schemes in different countries.

## IDENTITY, IDENTIFICATION, AND AUTHENTICATION

Identity and a person's sense of self are properly the subject of philosophy, psychology, or perhaps sociology. Our purpose here is to explore the engineering challenges that arise in building identity schemes; thus our scope can be limited to largely technical questions. We recognise, however, that these questions are themselves bearers of self-hood, and hence give rise to strong visceral reactions, particularly when threatened as in the case of identity theft or various invasions of privacy. We are interested here in schemes for *identity* (some would say *identity management*) and the *identifiers* that accompany them. *Authentication*—i.e., proving that a particular individual is the owner of a particular identity—is largely out of our scope, although the schemes we discuss must of course facilitate it. The *tokens* that facilitate authentication (be they plastic/printed cards, smart cards, active electronic devices, "virtual" online tokens, etc.) and that give rise to *identity assurance* are also broadly within the scope for this study. Crosby perhaps overstates the distinction in writing:

> "ID assurance is not ID management, in which an organisation keeps a close track of people and their movement. The distinction between the two is fundamental. ID management is designed to benefit the holder of the information. ID assurance is focused on bringing benefits to the consumer."[2]

1 In this paper, the term "eID" denotes an identifier for online use and "electronic ID (card)" refers to a physical card with a chip (in contrast to a traditional ID card comprising just paper or cardboard). The term "digital ID" encompasses both features.

2 James Crosby, "Challenges and Opportunities in Identity Assurance," Policy Review, HM Treasury, United Kingdom, 2008.

Our purpose is the study of schemes that underlie both ID management and assurance. We take it as a given that some form of identity scheme is necessary; societies have developed these schemes over centuries and would struggle without them. In absolute and idealistic terms, it may be that little identity is needed. Chaum, for example, describes schemes to allow society and commerce to function using purely anonymous transactions.[3] Similar principles underlie the design of Bitcoin, which enjoys strong audit properties to protect the integrity of the "currency," but spenders are untraceable—and therefore so are thieves, although Androulaki et al. cast some doubt on this point.[4] Blue-sky thinking about identity should not rule out such schemes, but the checks and balances developed over long periods (in this case, societal mitigations against theft) should not be jettisoned lightly, either.

Identity for us, then, is largely bound up with a demonstration of continuity—that the person withdrawing the money is the one who caused it to be deposited; that the person exercising the right to vote is the person who was born or naturalised into that right; and so on. To extend those examples, the critical question from a privacy perspective is whether anyone other than the person concerned should know that the person who withdrew a particular sum of money is also the person who exercised the right to vote. Clearly, in some contexts (not least, voting, in most democracies) it is possible to have too much identity; sometimes we need anonymity or at the very least pseudonymity. In some cases, identity also relies upon uniqueness: for instance, an individual may be permitted multiple independent bank accounts, but is not permitted two votes in the same election, nor allowed to draw upon public assistance twice for the same need.

## DOMAINS OF APPLICATION

Many ID schemes have historically arisen in the physical domain as identity documents, passports, etc. Some schemes are strictly designed for online use; these are sometimes called eID schemes. The cyber domain and the physical domain, however, have long since ceased to be separate. Many commercial entities have business interests in identity management and, in some cases, states outsource elements of national schemes to one or more such businesses.

3 David Chaum, "Security without Identification: Transaction Systems to Make Big Brother Obsolete," *Communications of the ACM*, Vol. 28, No. 10 (October 1985), pp. 1030–1044.

4 Elli Androulaki, Ghassan O. Karame, Marc Roeschlin, Tobias Scherer, and Srdjan Capkun, *Financial Cryptography and Data Security: 17th International Conference, FC 2013, Okinawa, Japan, April 1–5, 2013, Revised Selected Papers* (Berlin: Springer, 2013), especially pp. 34–51.

Domains of interest (online or offline) for a national identity scheme typically include:

- voting
- border control and right of abode
- tax registration
- long-term identity (e.g., for property title)
- proof of age (access to alcohol, tobacco, sexually explicit materials, etc.)
- licences (driver's licence, etc.)
- law enforcement and justice system and proof of identity in miscellaneous legal circumstances (witnesses, defendants, litigants)
- social security registration
- healthcare (proof of entitlement/insurance and/or access token for medical records)
- commercial applications, including:
  - banking (online and in person)
  - identity in high-value transactions and contracts
  - facilitating and validating digital signatures
  - online identity in e-commerce
  - employee identities and payroll numbers
  - the management of identity in ownership of shares, directing companies, etc.

The identity assurance possible through such schemes varies across a spectrum. At one extreme, the purpose may simply be instantaneous identification of an individual (e.g., at a border crossing). At the other, a unique identifier may be associated with an individual throughout the duration of one's life (e.g., a taxpayer ID). Digital signatures are a special case of the latter because they strongly bind the individual to the signed document long after the signing event.

Where an identity scheme incorporates a card carried by the citizen, this card may also be used as an ancillary token (e.g., to cross-reference the holder of a particular transport ticket). The purpose of a generic identity scheme is to avoid building each of these applications separately by developing an open architecture into which all solutions can readily be added with a minimum of extra effort, either in systems development or on the part of the users and citizens. Moreover, these solutions need not be single-point solutions linked only by a minimal identifier; instead, they can be integrated, cross-referenced ecosystems joining several solutions together.

### RESEARCH QUESTIONS AND ARGUMENT

The underlying motivation for our study can be summarised in another quote from Crosby:

> "In the absence of a universal ID assurance system, I believe consumers will have to grapple with an increasingly complex array of identity assurance processes of uncertain quality. As a result, [the nation] will fail to secure the economic and social advantage achievable at the forefront of ID assurance systems and processes. In a competitive world, any failure to secure advantage quickly becomes tantamount to locking in disadvantage. In other words, the opportunities inherent in ID assurance will not have been grasped but the challenges will remain."[5]

More specifically, we pursue three sets of research questions:

1 What are the main design decisions and alternatives for the basis of an identity scheme? What are the impacts of these design decisions upon the functionality and usability of the system built from any given scheme? What are the threats to security and privacy arising from such schemes?

2 Are the threats hypothesised in the first question set seen in the actual deployment of a given system? Does Estonian government data confirm the analysis of the first question?

3 Can other nations adopt all or some technical elements of the Estonian scheme in designing their own identity assurance systems?

The next section explores high-level requirements for identity schemes, giving particular importance to privacy, and considering a model of the threats to security and privacy that arise as a result of ID schemes. The following section develops an account of relevant design decisions to be made in delivering systems that meet those requirements and illustrates these choices with accounts of diverse schemes from a number of countries. We then discuss the experiences of developing and operating the Estonian ID scheme, which is distinctive and mature. We argue that one of the strongest features of the Estonian approach is the expectation that an identity is "public"—and thus users can share it freely with both government and private actors. The scheme's openness facilitates great utility and simple integration of databases. Such integration is only safe in the context of the scheme's strong audit requirements and controls over the creation of new databases and linkages. Finally, we conclude the paper with some broad observations on interoperability among national identity schemes.

## REQUIREMENTS OF IDENTITY SCHEMES

Viewed abstractly, identity schemes intending to deliver the mix of services described above must meet a range of technical and organisational requirements. These requirements are subject to subtle interplays and balances, both within the technological design choices made and with respect to the wider societal acceptance and utility of the services provided.

### HIGH-LEVEL REQUIREMENTS AND OBJECTIVES

Any practical scheme for identity will necessarily have as its foundation some form of *unique identifiers*. In the pre-digital age (and still today in some small communities), ordinary names served this purpose, but the range of possible names relative to the size of populations makes this impractical on the national scale, and even the use of birth dates to disambiguate identical names is insufficient. From a technological perspective, a unique identifier—"unique" in the sense that no one has the same identifier, not necessarily that each person has only one such identifier (see below)—is simply an additional, unambiguous personal name.

For some identity purposes, it is necessary to ensure existence of *just one name per person*. Certain state functions such as tax allowances, social security entitlements, or voter registrations require this property. Again, this does not require that the *same* identity is used for *all* purposes. Some systems may work best when an individual has a single identity, but this is not always an absolute requirement. In the provision of healthcare, for example, the best treatment may flow from all medical records being connected, but this need not be an absolute requirement if the patient chooses to isolate conversations with one clinician from those with another. Even where medical insurance is involved, it may not be strictly necessary to link different treatments, provided it is clear that the individual concerned is indeed insured.

5 Crosby, "Challenges and Opportunities in Identity Assurance."

A national-level scheme must avoid some simplifications that other schemes may employ. There is a strong requirement of *universality*: it must cover all citizens; it will probably cover all residents; it might cover some visitors (Estonia employs an additional category of "e-residents"). Similarly, a comprehensive scheme must include a capability to support identity management for the following categories: minors; those temporarily or permanently incapable; those who (through infirmity, for example) have vested authority in another person through a power of attorney; and those responsible for winding up the estate of a deceased person. In a related way, the scheme must offer *long-term stability*: it must work for individuals through all stages of life. Technology refreshes will be needed, but will be costly if introduced too frequently. Nevertheless, any such scheme must stay ahead of major exploitable vulnerabilities in the chosen technology. Finally, the scheme needs *durable processes for enrolment*, commensurate with the specific uses placed on the identities.[6]

One of the main uses for a unique identifier is in the construction of databases. Clearly, any design should support the use of the identifier as a *database index term*. It also then has value as a *database join term*—or in common terms, a means to join data sets together, that is, to link up data and individuals across diverse systems. This is good for utility; however, it may be bad for privacy (see below). Conversely, an identifier need not be directly associated with a database at all; an identity scheme may begin with, or continue to incorporate, paper-based records.

More prosaic but equally crucial concerns relate to economics: the cost of delivering the scheme relative to the cost of not doing so (or of delivering a poorer scheme) is a concern to the state and the citizen, but the balance of cost and utility will be different for different citizens and different contexts of use. Cost will be one factor that affects public and hence political acceptability; many other factors regarding usability and durability also play a part in decisions about which scheme to implement. Such issues matter because of network effects: the overall value and utility of a scheme grow somewhat faster than linearly in the number of everyday users and applications. Our concerns relate to the technical feasibility and design of schemes; thus we restrict ourselves to thinking about what comes at reasonable cost without making detailed models of the economic context.[7]

Other societal and legal norms also contribute to the design of an identity scheme. Most of our analysis here relates to the European context, as particular European Union (EU) law relates to privacy (though particular concerns vary by country) and to the development of EU electronic ID cards. It is also worthwhile to note that Common Law countries such as the United Kingdom have normally allowed individuals to assume any name—and hence, in some sense, identity—that they choose, and even to use more than one name concurrently, whereas in most other jurisdictions, an individual must have a single fixed name and undertake a substantial process in order to change it.[8] Perhaps this underlying philosophical difference helps to explain the general lack of unified identity schemes in Common Law countries, although contemporary constraints of banking, driving licence, and passport offices tend even in these countries to result in the use of a single *de facto* "official name," even where this concept does not exist in law.

## PRIVACY

Privacy is perhaps the greatest objection or concern that may arise when considering pervasive, universal, or state-sponsored identity schemes. Real or imagined threats to privacy (see section Attack Motives below) give rise to extensive debates and careful design. If overlooked, these threats can fatally undermine confidence in a scheme, such as in the cancellation of the United Kingdom's National Identity Card Scheme in 2011.

Privacy tends to defy definitive description.[9] Whittman observes that even across Western societies sharing many values, expectations and cultural mores around privacy differ widely.[10] Thus it is little wonder that differing national schemes rest on very different design decisions.[11]

6 William E. Burr, Donna F. Dodson, Elaine M. Newton, Ray A. Perlner, W. Timothy Polk, Sarbari Gupta, and Emad A. Nabbus, "Electronic Authentication Guideline," Special Publication SP 800-63-1 (Gaithersburg, Md.: National Institute of Standards and Technology, 2011).

7 Another area in which the reasonable management of the scheme(s) and associated databases, cards, card readers, and other technologies matters is in the management of the supply chain for the delivery of those elements. A nation's capability is potentially put at risk by vulnerable or actively subverted components.

8 Jane Caplan, "This or That Particular Person: Protocols of Identification in Nineteenth-Century Europe," *Documenting Individual Identity: The Development of State Practices in the Modern World*, Vol. 63 (2001); G.S. Arnold, "Personal Names," *Yale Law Review*, No. 5 (1906), p. 234.

9 Daniel J. Solove, *Understanding Privacy* (Cambridge, Mass.: Harvard University Press, 2008).

10 James Q. Whitman, "The Two Western Cultures of Privacy: Dignity versus Liberty," *Yale Law Journal* (2004), pp. 1151–1221.

11 A useful survey of concepts of privacy from an Estonian perspective comes from Maria Murumaa-Mengel, Pille

A valuable and extensive EU report explores social attitudes to privacy and electronic identity and how these attitudes vary across the Union. For example:

> "More than six respondents out of ten (63%) say that disclosing personal information is a big issue for them.… The highest percentages of respondents saying that it is not a big issue are found in Denmark (51%), Estonia (47%), Lithuania (46%), Sweden (45%) and Poland (44%). Conversely, the lowest percentages are found in France, Greece (both 23%), Malta and Slovenia (both 24%)."[12]

A general right to privacy is widely upheld, deriving in part from the Universal Declaration of Human Rights ("No one shall be subjected to arbitrary inference with his privacy")[13] and the European Convention on Human rights ("right to respect for private and family life").[14] A common working definition of privacy, at least in the information or data domain, is the following: "a person's right to control access to his or her personal information."

The means by which control is exercised, and its extent, are of particular relevance. Some especially relevant principles are embodied in the European Data Protection Directive 95/46/EC and other legislation. They may be summarised as follows:

**Transparency.** People should be able to know what data is held on them, to correct it, and to know how it is used in reaching decisions.

**Legitimate purpose.** Personal data to be obtained only for specified purposes and not further processed for an incompatible purpose. This might be taken also to imply a prohibition on linking data sets which were not designed to be connected.

**Proportionality.** A requirement that data be adequate and not excessive for their purpose(s).

Some consequences and further principles include obligations on the means of disposal (i.e., that data no longer needed should not be preserved) as well as anonymity and pseudonymity, which may be used to prevent the linking of particular data to particular individuals, but are often an illusionary protection because many surprisingly small data sets in fact permit re-identification of specific individuals with high degrees of accuracy. In the latter case, perhaps a better framework for evaluation relates not to the binary decision of *whether* re-identification is possible, but *how much effort* is required to do so, with a set level of confidence.

In designing systems to meet such goals, a number of principles of privacy by design have gained a great deal of traction and provide a good framework against which to evaluate a privacy-sensitive system, such as the identity schemes described here.[15]

1 Proactive not Reactive
2 Preventative not Remedial
3 Privacy as the Default
4 Privacy Embedded into Design
5 Full Functionality—Positive Sum not Zero Sum
6 End-to-end Security—Lifecycle Protection
7 Visibility and Transparency
8 Respect for User Privacy

There is, of course, much debate about how to adopt such principles in practice. Can transparency be used as a means to ensure privacy? Could the disposal principle above be fulfilled by a citizen's right to audit use of his or her personal data? Such a right is not preventative in the technological sense, but is broadly so in the wider context if misuse of personal data routinely and automatically results in meaningful punishment or sanction.

**Illustration: Privacy of ePassports**

Conventionally, identification and ID cards have been closely related to travel documents, which have intrinsic security and privacy requirements. These requirements are established by a number of international authorities and standards, most notably the International Civil Aviation Organization (ICAO), which is a specialised agency of the United Nations. The ICAO, in cooperation with the International Standard Organization (ISO), has

Pruulmann-Vengerfeld, and Katrin Laas-Mikko, "The Right to Privacy as a Human Right and Everyday Technologies," Estonian Institute of Human Rights (2014).

12 *Attitudes on Data Protection and Electronic Identity in the European Union*, Special Eurobarometer 359 (Brussels: European Commission, 2011).

13 United Nations General Assembly, *Universal Declaration of Human Rights*, 217 A (III), 10 December 1948, http://www.refworld.org/docid/3ae6b3712c.html, accessed 12 April 2016.

14 Council of Europe, *European Convention for the Protection of Human Rights and Fundamental Freedoms* [as amended by Protocols Nos. 11 and 14], 4 November 1950, ETS 5, http://www.refworld.org/docid/3ae6b3b04.html, accessed 12 April 2016.

15 Ann Cavoukian, "Privacy by Design: The Seven Foundational Principles," Information and Privacy Commissioner of Ontario, Canada, rev. ed., January 2011.

standardised ePassports based on contactless cards (ISO 14443) using Radio-Frequency Identification (RFID) and biometrics (currently standardised biometrics are facial recognition, fingerprint recognition, and iris recognition).[16] The biometrics of the passport holder are included in a chip that is embedded in the passport. The data from the chip are communicated wirelessly to the reader during the identity verification process. Similar to many modern ID cards, the security mechanisms of ePassports are based on strong cryptographic techniques and Public Key Infrastructure (PKI).[17]

The main privacy-related threat in this context is "skimming" of the ePassports, i.e., an illegitimate or unauthorised reader could collect personal information by actively querying the chip without user consent or by eavesdropping on legitimate communication, such as by passively intercepting the communication between a valid reader and a passport.[18] The ICAO has recognised that the security and privacy laws of issuing states may require that such privacy-sensitive data stored in ePassports (name, date of birth, gender, passport number, certain biometric data, etc.) should only be accessible to authorised persons (readers) and with the user's consent.

Take, for example, the Basic Access Control (BAC) protocol. It ensures that only authorised readers can wirelessly access personal information from ePassports. The protocol is based on the assumption that physical access to an ePassport is needed before the user data can be read; i.e., the ePassport holder knows when the reader is querying the RFID chip. Another privacy-related countermeasure implemented in ePassports is the concept of random unique IDs (UID), by which the ePassport transmits a random ID when it is activated for the first time, which should eliminate the threat of the ePassport's traceability.[19] Many studies on the security and privacy challenges of ePassports concluded that most of the safety mechanisms are insufficiently resilient against realistic adversaries.[20] In particular, it has been demonstrated that the BAC protocol is not resilient against many privacy-related attacks. Even with the BAC protocol, an attacker can successfully intercept and collect privacy-sensitive user data without the user's consent because the secret keys are guessable (i.e., they have a low information entropy). To mitigate this problem, the entropy of the keys must be increased, but this would affect many high-level design decisions, potentially degrading the usage and overall system performance of ePassports.

The example of ePassports shows that protecting privacy through merely technical means such as cryptographic encryption is a difficult, often elusive task in the real world. The ID ecosystem is governed by many *partially* orthogonal objectives, which results in many trade-offs, most prominently concerning usability, security, and system performance. Striking a good balance between them is difficult to achieve because it involves assumptions about threat actors' motives and capabilities, which are difficult to ascertain or predict. In the next section, we identify and discuss some of the main threats related to the ID ecosystem.

## THREAT MODEL

We limit our attention to threats against the identity scheme itself. Concerns about threats that apply directly to government and private functions that rely upon the scheme are outside our scope. For example, public divulgence of anonymised data sets for the purposes of transparency or research may raise legitimate concerns about re-identification and privacy (is it worthwhile to be concerned about a unique identifier when the individual is effectively identifiable from all kinds of other data anyway?)[21]—but these concerns are distinct from threats to the ID scheme itself. Some threats to the scheme necessarily encompass these broader concerns, because certain technological choices are finely balanced. For instance, unambiguity through a unique identifier helps to identify a required record with good precision. Without it, a database user may need to access multiple records, and

16 International Civil Aviation Organization, "Machine Readable Travel Documents," Document Series 9303, http://www.icao.int/publications/pages/publication.aspx?docnum=9303, accessed on 29 March 2016.

17 For further details, see ICAO, DOC 9303, Part 11 and Part 12.

18 For an overview of such attacks, see, for instance, Jaap-Henk Hoepman, Engelbert Hubbers, Bart Jacobs, Martijn Oostdijk, and Ronny Wichers Schreur, "Crossing Borders: Security and Privacy Issues of the European e-Passport," *Advances in Information and Computer Security: First International Workshop on Security, IWSEC 2006, Kyoto, Japan, October 23–24, 2006. Proceedings* (Berlin: Springer, 2006), pp. 152–167; and Ari Juels, David Molnar, and David Wagner, "Security and Privacy Issues in e-Passports," in *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks, SECURECOMM 2005, Athens, Greece, 5-9 September 2005*, pp. 74–88.

19 This protective measure, however, highly depends on the security of the underlying BAC protocol.

20 For a systematic survey on security and privacy of the ePassport see, for example, Gildas Avoine, Antonin Beaujeant, Julio Hernandez-Castro, Louis Demay, and Philippe Teuwen, "A Survey of Security and Privacy Issues in ePassport Protocols," *ACM Computing Surveys*, Vol. 48, No. 3 (February 2016), pp. 1–47.

21 Arvind Narayanan and Vitaly Shmatikov, "Robust De-Anonymization of Large Sparse Datasets," in *IEEE Symposium on Security and Privacy, 2008. SP 2008.* IEEE, (2008), pp. 111–125; Andrew C. Simpson, "On Privacy and Public Data: A Study of data.gov.uk," *Journal of Privacy and Confidentiality*, Vol. 3, No. 1 (2011), pp. 51–65.

thus view personal data he has no need to see, in order to find the relevant record. Likewise, there may appear to be a benefit from using different identity schemes for different purposes (social security, tax, immigration, voting), but inevitably databases will exist which correlate two or more of these identifiers. Ultimately, the database separation is perhaps illusionary, introducing potential error and possible association of records with the wrong individuals, without materially diminishing the ease of correlation across schemes.

Notwithstanding threats that arise from *not* having a sufficient identity scheme, our focus here is on threats against or because of the scheme itself.

**Adversaries and Their Capabilities**

Citizens may have a motive to hide their identities or perhaps to impersonate other identities. They may be able to steal or replace physical tokens or to discover ID numbers of those in their social circle, together with names, dates of birth, etc. They might be able to "shoulder surf" the entry of PIN codes by those close to them or learn these codes through informal sharing ("Please borrow my card and do this for me").

*Petty criminals* will have capabilities similar to average citizens regarding others who are known to them and may be able to socially engineer information from strangers as well.

*Organised criminals* might be able to produce visually convincing fake cards, install fake card readers (relay devices or data capture devices), and coerce government employees.

*Insiders (identity management authorities)*. Insiders may have the opportunity to violate procedural norms, create fake identities, or block the issue of legitimate ones. Most procedures will have exception mechanisms, which may be invoked inappropriately—that is, invoked as if at the behest of the citizen, but without the individual's actual knowledge or consent. Insiders also have scope to compromise privacy and security, in the form of confidentiality and integrity, by accessing individual records without cause or by attempting unauthorised bulk export of data.

*Unethical organisations* might have similar capabilities as organised criminals—but with the potential also to pose as legitimate actors with access to official databases.

*Online hackers and cyber militia* may have the skills to compromise the security of online services through the kind of attacks deployed against diverse organisations. They may also launch denial of service attacks (including Distributed Denial of Service attacks): if multiple state and private sector functions rely on an online service related to identity, such attacks have the potential to paralyse the whole operation of society.

*Intelligence agencies and offensive cyber security units*, at the very high end, may be able to clone cards, disrupt supply chains, steal bulk lists of card IDs and keys, etc. Such adversaries may also recruit or coerce any of the other players described above in order to achieve their own ends.

**Attack Motives**

**Perspective of the state.** The highest-level category of threats to the state relate to state security and integrity. In the management of any identity scheme, these threats will include:

- Immigration and Border Control threats.
- Attacks upon the Identity System:
  - denial of service
  - reputational damage through misuse or scandal
  - perceived failure of audit/control
- Attacks leading to a loss of confidence, economic impact, and so on, potentially affecting the viability of the scheme itself.
- Disruption of law enforcement (failure to identify suspects, criminals, or witnesses).
- Enumeration of citizens: in practice, some citizens' personal data is more interesting than others' (e.g., law enforcement officers, intelligence personnel, particular public servants or politicians). Enumeration of all subjects within a particular scheme could reveal some individuals whose identities require additional protection.
- Large-scale data theft: whether or not identifiers are public, the data associated with them will typically not be public, including address information and metadata associated with the card issuance. This could apply to any database associated with the identity scheme in the public or private sector. The use of unique identifiers may help the adversary by enabling the linking of stolen databases, perhaps in combination with public domain information.

A separate category of threats relates to the state functions and interaction with the citizen:

- social security fraud (individual or large-scale)
- tax fraud (individual or large scale)
- voting fraud

A third category concerns management of the scheme in the medium or long term:

- Legal equivalence of digital and physical documents may be desirable in theory, but requires much effort to ensure in practice, and includes risks relating to whichever is given pre-eminence in the case of a discrepancy.
- Political risks have the potential to add complexity to the design of the scheme, particularly political influences which aim (knowingly or not) to undermine the scheme.
- Long-term decision-making is generally needed in ID schemes: security flaws may show up much later than they are introduced, and the context or expectations may change, altering the security analysis.
- Compliance with external regulations and international interoperability (such as the EU's electronic identification and trust services—eIDAS—regulation) may undermine assumptions made in the initial conception of the scheme; issues of technical design harmonisation can give rise to unexpected vulnerabilities (particularly with regard to logging and privacy).

**Perspective of the citizen.** It is reasonable to assume that citizens are concerned about threats to individual identity, such as:

- threats to privacy:
  - unwanted disclosure of personal information
  - unwanted linking/correlation (whether by officially sanctioned authorities, or by individuals with malign intentions)
  - discrimination which arises as a result of over-identification
- threat of theft of digital assets
- threat of being defrauded (e.g., phishing and identity theft)
- threat of impersonation (overlaps with fraud)

In the case of "over-identification" we refer, for example, to a situation in a commercial context where pricing can be dynamic. To illustrate, an online retailer is able to set prices based on what it estimates the customer is willing to pay. If the retailer has too much information about the customer's identity and associated profile, this will work against the customer's interests. Such issues are also challenging the foundations of some forms of insurance, particularly health insurance, because with enhanced personal information, the notion of shared risk in the face of uncertain outcomes arguably becomes redundant.

Where different schemes are integrated, such as citizens of one country using their ID in another country, threats arise from the accidental clash of identifiers, or from record fragmentation if identifiers are not uniformly mapped by different functions and databases.[22]

**Perspective of other relying parties.** Some identity schemes seek to support both private commercial transactions and state functions. In such instances, threats to the scheme become threats to the commercial parties relying upon the identities it manages. The chief danger is one of fraud arising from inadequate authentication. If the scheme's design is inadequate, unauthorised parties may gain access to too much data about individuals, producing problems of liability and data management. In practice, schemes designed for government needs will probably exceed the level of assurance needed for most such commercial activities.

## DESIGN CHOICES AND ILLUSTRATIONS

This section identifies the main design choices underlying various national ID systems and defines a framework for a comparative analysis among them. We illustrate the abstract design options by reference to the national schemes of Austria, Belgium, Germany, Spain, the United States, Japan, and the United Kingdom, before undertaking a more detailed study of Estonia's scheme in the following section.

### CHOICES AND IMPLICATIONS

We list here some of the major design choices pertaining to the creation of a national identity scheme, drawing on the successes and failures of known systems to date.

22 Kjell Hansteen, Jon Ølnes, and Tor Alvik, "Nordic Digital Identification (eID): Survey and Recommendations for Cross Border Cooperation," Report 2016:508 (Copenhagen: TemaNord and Nordic Council of Ministers, 2016).

**Secret, Non-Secret, or Published ID**

Is the identifier intended to be a "secret," i.e., known only to the individual and authorised parties? If this is assumed, how is it enforced? Because it becomes a shared secret, relying parties necessarily learn it—and thus secrecy is compromised. How is this problem addressed? Here, we might consider the U.S. Social Security Number system or the Austrian approach, both described below. Alternatively, the identifier may be "non-secret" — i.e., by default known only to the possessor but safe to be shared with trusted and un-trusted services because by itself it is of little use; it does not authenticate the user and it conveys no personal information. A particularly strong variant of non-secrecy is for the ID to be published and stored in a public directory or to be somehow capable of being generated from other public data about the individual without that person's knowledge.

**Syntax and Semantics of IDs: Systematic or Random**

The creation (and syntax) of the identities themselves — represented as strings of digits — is subtle but critical. Usability is enhanced if they are systematic (the citizen can recall facts such as his or her date of birth and use facts these to reconstruct the ID). Conversely, the inclusion of the date of birth within the ID has proved problematic for reasons of privacy, because it tends to make it desirable to move the ID from the "non-secret" to the "secret" category. If there is little randomness (entropy) within the ID, a third party can easily guess or reconstruct the ID from limited information about the holder; hence the ID is effectively in the "published" category. If every transaction involves the use of a token, memorability is probably less of an issue. A related issue of formatting is the need for redundancy within the identifier (error-detecting or even error-correcting codes). Although redundancy adds overhead (e.g., making the ID string longer or reducing memorability), experience suggests that it is essential because it enhances privacy by reducing the risk of misidentification (mistyped identifiers are unlikely to resolve to a valid identity).

**Single or Multiple Identities**

Privacy can be enhanced by multiple pseudonymous identities, while systems based on a single identity are much easier to build and to protect and much easier for the citizen to manage. Multiple identities can prevent data from one context being used in another (tax records cannot be cross-referenced with health records, for example). If implemented well, this gives the citizen considerable control over how his or her data is used—a strong embodiment of the privacy principles described above. Conversely, multiple identities diminish many of the database benefits for the state, for private parties, and possibly for the citizen in cases where benefits to the individual rely on linking disparate data sources. In the case of multiple identities, the design of the scheme is of interest: are all of a citizen's identities tied to a single root or to a single token? If so, the scheme must ensure that the apparent multiplicity and unlinkability is not merely an illusion. A related question concerns who controls the mapping from the root ID to the multiple identities. Privacy is greatest if control is solely the ID owner's, but this presents a substantial usability challenge. If a separate agency helps to manage the mapping, then this agency is able to link the records the user wished to keep separate. If the user holds the only mapping, what happens if it is lost or destroyed? In addition, mechanisms are required to prevent multiple registrations in cases such as voter ID, where the citizen must clearly enrol only once.

**Mandatory or Optional**

Is use of the ID scheme compulsory for every citizen and in every circumstance? If so, this maximises the utility of the scheme, but possibly compromises the privacy of the individual and thus may invoke resistance (more so in some countries than others). A related non-technical issue is the cost to the citizen: a compulsory scheme with a non-trivial cost is typically unwelcome.

**Token-Based or Virtual**

Identities may be asserted by a number of means. Authentication may come from the ID itself, together with a password or PIN, for example. But the problems of password theft and guessing are well-known today, making such authentication unsuitable for high-grade transactions. Notwithstanding these problems, some identity schemes such as the current UK scheme ("Gov. UK Verify") are designed solely for online use and use online logins as primary authentication. Stronger schemes require a smart card or other token, together with a card reader deployed in every authentication context.

**Physical or Digital Documents**

In situations where a "mixed economy" exists, with some documents and records in the physical domain and others in the digital domain, are records intended to be interchangeable? The answer to this question affects the construction of conventional paper-based contracts. Should they be required to carry ID values as well as the parties' names? If not, how is a document in one domain unambiguously tied to a document or identity in the other?

**Authentication and other functions**

An ID scheme may aim primarily to assure or manage the citizen's identity. Distributing advanced infrastructure (smart cards, readers, etc.), however, creates the opportunity to add other functions, such as digital signatures for documents. This entails a more complex cryptographic scheme: it is wise to separate the cryptography used for authentication from the cryptography used for signing. Otherwise, in a severe attack, the former could be used to achieve the latter without the citizen's consent. Complexity increases overheads and presents usability challenges because the citizen must know whether he or she is authenticating or signing, which probably requires separate PINs, but enhances the utility of the overall scheme at relatively little additional cost.

**Interoperability**

Can other national identity schemes work with each other? How easily could a citizen from one country use his or her identity in another country? Such questions relate to many of the scheme's technical details; some overlap with political issues: for example, who will control the "root" signatures and certificate authorities of a cryptographic scheme? Standards such as NIST 800-63 help to clarify whether different issuers are following equivalent processes and addressing similar goals.[23] Matching procedural norms surrounding identities is particularly important if, for example, one country is given access to another country's primary identity system, perhaps in order to use official identities of one country on the official documents of another.

**Transparency**

Who is able to observe logs of ID-based systems access? Clearly, this question relates to the privacy issues raised above. Some observers hold that if the citizen has the right to audit accesses, this mitigates most risks associated with the misuse of private data. Unique identifiers enhance transparency in a different sense; they help to ensure that where records are linked, this is done accurately, whereas in matches involving the use of a name, the linking will necessarily be imprecise, raising the possibility that the wrong record will be accessed by accident.

Nations seldom have the opportunity to introduce a new scheme without regard to previously deployed "legacy" systems as well as constitutional and legal norms. Estonia, for instance, had fewer legacy systems at the outset than most other nations in this study (more on this below). The design choices listed above will seldom be made solely on the basis of which choice is "best"; rather, all systems make compromises for various reasons. Engineering is often the art of the possible. If one could discard one system and start from scratch, one might make different choices—but this avenue is generally impractical for many reasons.

## ILLUSTRATIONS: NATIONAL IDENTITY SCHEMES

Here we discuss concrete examples of different states and how they have approached digital identities and design choices identified above. Our aim is not to provide a full in-depth analysis of each scheme, but rather to illustrate core design elements. Some schemes have grown out of the state's traditional function of issuing travel and identity documents, strongly influencing their design. Other schemes have a very different primary goal: providing a trustworthy identifier to be used in the digital domain (an "eID"). These and other purposes are gradually converging, giving rise to interoperability challenges (both technical and procedural/organisational) and potentially complex threats, because a comprehensive scheme unifies all such fragments into a unified identity management regime.

Recently, the European Union established several relevant standards and requirements covering electronic identities in particular as well as various aspects of identity cards and travel documents. In particular, it is adopting a scheme to ensure interoperability of eIDs across member states. Its main goals are (a) that individuals and businesses can use eIDs issued under their own national scheme to access public services in another EU country; and (b) to create a Europe-wide market and ecosystem for "electronic trust services" (eTS)—that is, support of electronic signatures as well as the controls and systems that surround them.

**Austria**

Austria has been using its Social Security Number (SSN) as a national identifier for many years. The SSN includes a three-digit serial number, a checksum, and the user's date of birth in DDMMYY-notation. Some problems regarding the SSN have been reported, mainly related to SSN allocation, such as having duplicates or missing SSNs and an insufficient number of available SSNs because of limitations in format.[24]

Because the SSN contains the user's birth date, it is

23 Burr et al., "Electronic Authentication Guideline."

24 Sozialversicherung, "eCard: English Information (Austria)," https://www.sozialversicherung.at/portal27/sec/portal/ecardportal/content/contentWindow?contentid=10007.678587&action=2, accessed on 24 February 2016. As a practical workaround, some SSNs included additional months, such as 13, 14, or 15.

considered a privacy-sensitive number and its use is limited by law to specific areas, such as healthcare, taxation, and education. This renders the SSN unsuitable as a general identifier. In 2005, Austria introduced an eID to mitigate the problems of the SSN and to facilitate their transition to digital identities, including schemes to guarantee authenticity, uniqueness, and privacy.[25] Each citizen has been allocated a meaningless identification number called a "SourcePIN." One of the scheme's main design decisions is to keep the SourcePIN secret and to use it only as a "seed" for generating multiple pseudonyms called *sector-specific identifiers*. To achieve this goal, SourcePIN is cryptographically bound to a specific sector of governmental activity that results in an identity pseudonym called a *sector-specific PIN* (ssPIN). Consequently, within the functions of a particular government activity, the citizen can only be identified by the ssPIN given to that specific service or sector. Because the ssPIN is generated using cryptographic one-way functions and symmetric ciphers, it is not possible to work back from the ssPIN and calculate the source PIN. Nor is it possible to generate any other ssPINs from an existing ssPIN. An additional advantage of this scheme is its high degree of interoperability: foreign IDs can be created easily by generating substitutional sourcePINs and can therefore be integrated into Austrian eGovernment services.[26] While this scheme enjoys many benefits involving the protection of citizen data, particular disadvantages might arise, such as inefficiency of sector-specific PINs in terms of the complexity of data exchange, communication between different governmental institutions, and other problems related to the mapping and sharing of different identifiers.[27]

**Belgium**

In Belgium, every citizen has a single national identification number called the National Registry Number (NRN). The NRN is generated using the citizen's date of birth as well as a three-digit serial number denoting gender and a checksum. The government considers this number to be privacy-sensitive because of the inclusion of the user's birthdate (which is also considered privacy-sensitive by the EC directives); use of the number is thus restricted. Because there are no multiple identities involved, protecting the number has focused on authorising its usage—that is, the authorisation to request the citizen's NRN is approved by a special committee under the Belgian Data Protection Authority. It seems, however, that the widespread use of this number has affected its privacy. For instance, the new electronic identity card currently employs this number in a citizen's X.509 digital certificate information, which means that every time the ID is used to authenticate the citizen, the NRN is leaked, because digital certificates are considered public information and are thus commonly exchanged during an execution of cryptographic protocols used for authentication and key exchange. The Belgian electronic ID card holds three different 1024-bit RSA private signing keys: one to authenticate the citizen, one for non-repudiation signatures, and one to identify the card itself to the Belgian government.[28]

**Germany**

Germany does not utilise a unique national identification number. Yet the country has a strong tradition of ID card usage: citizens are obliged to possess either an ID card or a passport. Since 2010, Germany has provided an identity card with sector-specific pseudonyms similar to the Austrian ssPIN approach. Similar to ePassports, the identity card is based on smartcard and RFID technology and implements cryptographic protocols for authentication and digital signatures.[29] It offers cryptographic mechanisms such as Basic Access Control to ensure that wireless reading of personal information requires physical possession of the document. It also offers Extended Access Control (EAC) to prove that the chip on the identity card is genuine—a cryptographic countermeasure that can forestall counterfeiting—and to provide a means to generate strong cryptographic secrets for establishing secure communication channels between the card and the reader. Interestingly, while the German and Austrian identity schemes share many commonalities, their adoption rates are very different because possession of the identity card is mandatory in Germany but not in Austria. In addition, the German constitutional ban on national identification numbers (related to their abuse during World War II) assists in preventing the use of any sector-specific PIN as a general and single national identification number.[30]

25 Georg Aichholzer and Stefan Strauss, "The Austrian Case: Multi-card Concept and the Relationship between Citizen ID and Social Security Cards," *Identity in the Information Society*, Vol. 3, No. 1 (2010), pp. 65–85.

26 European Commission Interoperable Delivery of European eGovernment Services to Public Administrations, Businesses and Citizens (IDABC), "eID Case Study: Austria," *SYNeRGY—The IDABC Quarterly*, Issue 3 (July 2005).

27 For a good overview, see for example Niels Vandezande, "Identification Numbers as Pseudonyms in the EU Public Sector," *European Journal of Law and Technology*, Vol. 2, No. 2 (2011).

28 Ilse Mariën and Leo Van Audenhove, "The Belgian e-ID and Its Complex Path to Implementation and Innovational Change," *Identity in the Information Society*, Vol. 3, No. 1 (2010), pp. 27–41.

29 Bundesamt für Sicherheit in der Informationstechnik, "Electronic Identity Card (ID_Card PP): Common Criteria Protection Profile BSI-CC-PP-0061," Bonn, Germany (December 2009).

30 Vandezande, "Identification Numbers as Pseudonyms in the EU Public Sector."

**Spain**

In Spain, all citizens over the age of fourteen must possess a National Identity Document with a unique number in the following format: 00000000-A, where 0 is a digit and A is a checksum letter. The number is randomly generated; it is intrinsically meaningless and is used as identification for almost all purposes, such as taxation and driver's licencing as well as public- and private-sector transactions. The number is not considered privacy-sensitive because it does not contain any personal information. In a similar way to Germany, Spain is presently introducing a digital identity scheme based on smartcard and RFID technology, which in addition to a national identification number also contains biometric data such as fingerprints. Access to the biometric data is protected using a PIN and similar access control protocols to the German identity card. Moreover, the Spanish identity card utilises cryptographic keys and digital certificates to support digital signatures.[31]

**The United States**

The U.S. Social Security Number (SSN), a unique personal identifier, was introduced in the 1930s as part of the New Deal programme. A nine-digit number, it has suffered from an overload of uses, being both a personal identifier and an authentication "secret" (despite not being intended for the purpose). Garfinkel describes many problems arising from this dual use.[32] Berghel notes the particular dangers of identity theft because the SSN has become a commonplace identifier in many contexts (not just within the Social Security Department) and thus offers an opportunity to commit fraud to numerous individuals with access to databases managed by diverse organisations.[33] These problems are compounded by the limited entropy of the identifier and lack of any check-digits, such that simple data entry errors can lead to misidentification. Notwithstanding these critical analyses of the SSN system, Mercuri argues that the impact of identity theft may be decidedly overstated.[34]

**Japan**

In 2002, Japan introduced the national registry of Japanese citizens, or Basic Resident Registers Network. The registry content consists of name, address, date of birth, gender, and an eleven-digit individual number. In 2015, Japan introduced a National ID system based on a randomly generated twelve-digit individual identification number—the so-called My Number—similar to the U.S. Social Security Number. The number is issued to every person holding a resident record.[35] It can also be issued on a cryptography enabled smart-card (i.e., a digital ID card, called an Individual Number Card), which holds the digital certificates for the purpose of digital authentication. Initial planned uses of the card include social security, tax returns, and disaster relief assistance.[36] The government's general plan is to extend the use of My Number and the digital ID card as a basis for identification and authorisation in accessing public and private services, such as a health insurance, credit and debit cards, and employee ID cards. As a part of the My Number system, the Japanese government plans to provide an online service that will enable individuals to check all records of personal information exchange involving their My Number. As a countermeasure against identity theft, face-to-face identity confirmation will be necessary when revealing one's My Number to a third party. This scheme is still in its early deployment phase; not much information is available on the effectiveness and scalability of such countermeasures. Overall, the Japanese national ID system will be very similar to the Estonian ID, with mechanisms that will implement authorisation of requests to access personal information, involve dedicated committees to oversee the process, and strengthen the personal information protection laws. It seems, however, that the deployment of services based on My Number and the Individual Number Card has been delayed due to Japanese citizens' privacy concerns related to the tax system and government control.

**The United Kingdom**

During the 2000s, the UK government developed a National Identity Register with associated National Identity Cards. The scheme used a unique identifier; authentication was backed by biometrics.[37] The scheme proved highly unpopular with many sectors of the community because its objectives were not entirely clear to the population at large. The Conservative government abandoned the scheme following the 2010 elections. One of the underlying reasons for the old scheme's failure is

31 Vendezande, "Identification Numbers as Pseudonyms in the EU Public Sector."

32 Simson L. Garfinkel, "Risks of Social Security Numbers," *Communications of the ACM*, Vol. 38, No. 10 (October 1995), p. 146.

33 Hal Berghel, "Identity Theft, Social Security Numbers, and the Web," *Communications of the ACM*, Vol. 43, No. 2 (February 2000), pp. 17–21.

34 Rebecca T. Mercuri, "Scoping Identity Theft," *Communications of the ACM*, Vol. 49, No. 5 (May 2006), pp. 17–21.

35 A good overview of the individual number card can be found in the website of the Japan Agency for Local Authority Information Systems, https://www.kojinbango-card.go.jp/en/kojinbango/index.html, accessed on 14 March 2016.

36 Tomohiro Osaki, "Ready or Not, Government Will Soon Have Your My Number," *Japan Times* (20 September 2015).

37 Many of the scheme's principles are documented in Crosby, "Challenges and Opportunities in Identity Assurance."

the lack of an underlying framework and legal norms for identity as well as the distrust of any government that attempts to organise the collection and use of such information. Consequently, the scope of the government's new and more modest Gov.UK Verify scheme is limited to the management of online identity and transactions. Furthermore, it involves outsourced verification of identity, giving the citizen a choice of several private organisations with which to register.[38]

## ESTONIA

For many years Estonia has been laying the foundations for a digital society, providing e-services and implementing technical and legal means to support widespread usage of Digital IDs. In 2000, the Estonian parliament passed the Digital Signature Act, which made a digital signature equivalent to a hand-written signature; since then, all Estonian authorities have been obliged to accept digitally signed documents. That same year, the government also began to develop a scheme to launch Identity Cards implementing smart-card technology. The card stored digital certificates of public keys used for authentication and digital signatures.[39] Since their inception, the use of both capabilities has grown substantially (see Figure 1).

In 2005, Estonia was the first country to hold legally binding municipal elections over the Internet. Two years later, the country implemented the world's first online parliamentary elections. In 2015, approximately 30 percent of voters used online voting systems.[40] Since 2014, Estonia has been running an "e-Residency" project, which allows non-residents to obtain a digital ID (digital residency) similar to the ID cards of ordinary residents. E-residents are able to use digital signatures for tax declarations, online banking, and to establish enterprises in Estonia.[41] To support such a broad range of public and private services, the Estonian identity scheme had to be developed and deployed within a comprehensive

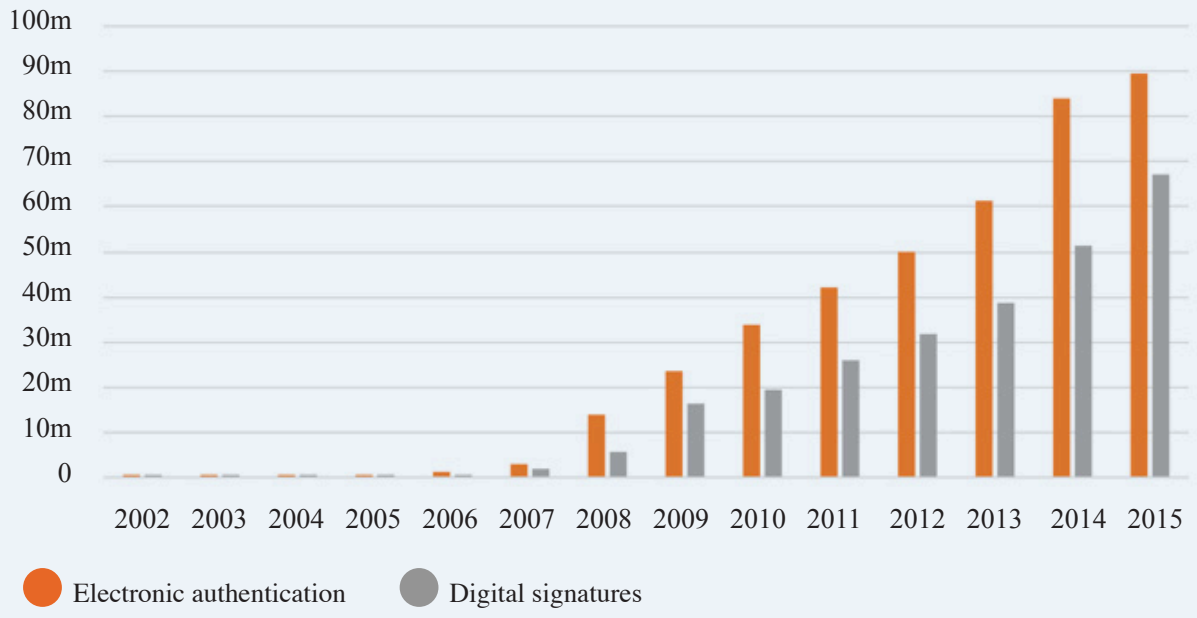
38 UK Cabinet Office, "Introducing GOV.UK Verify," Guidance, Government Digital Service, United Kingdom (February 2016).

39 Tarvi Martens, "Electronic Identity Management in Estonia between Market and State Governance," *Identity in the Information Society*, Vol. 3, No. 1 (2010), pp. 213–233.

40 Estonian National Electoral Committee (Vabariigi Valimiskomisjon), "Statistics about Internet Voting in Estonia," Tallinn, http://www.vvk.ee/voting-methods-in-estonia/engindex/statistics, accessed on 14 March 2016.

41 Taavi Kotka, Carlos Ivan Vargas Alvarez del Castillo, and Kaspar Korjus, "Estonian e-Residency: Redefining the Nation-State in the Digital Era," *Cyber Studies Working Paper Series*, No. 3, University of Oxford (September 2015).

**Figure 1: Growth in the Usage of Estonian ID Cards, 2002–2015.**




Source: Zuzana Pruuli, AS Sertifitseerimiskeskus, 26 January 2016.

identity system. The system relies on a single *unique national identifier* (i.e., the national ID), which is kept in a population database for all citizens and which includes name, date of birth, gender, address history, citizenship, and legal relationships. The national ID is not random; it is generated systematically in the form of GYYMMDDSSSC, which denotes gender (G), date of birth (YYMMDD), a serial number (SSS) to separate persons born on the same date, and a checksum (C).

Next we discuss the main design choices of Estonia's national ID and its usage as a physical and digital identifier within the design choices introduced above.

## ESTONIAN DESIGN CHOICES

In general, the Estonian ID can be categorised as a *public* and *single* identifier. Being public, it is printed on the front of ID cards together with other personal information such as date of birth, gender, and citizenship status. The national ID is also used as a foundation for many public- and private-sector services. In particular, its use within the eID ecosystem supports the implementation of many new and innovative services, such as access to health care (replacing separate health cards), an alternative to the driving licence, or as a public transportation ticket. In contrast to many other countries' systems, possession of the Estonian electronic ID card is mandatory for all permanent residents. The card issuance initially occurred in close private-public partnership between the Estonian Citizenship and Migration Board (CMB), which is the government organisation responsible for issuing identification documents, and AS Sertifitseerimiskeskus, which serves as a PKI Certification Authority (CA), and TRÜB Baltic AS, a subsidiary of Swiss TRÜB AG, which personalises the card.[42] In 2010, the CMB functions were handed over to the Estonian Police and Border Guard.

The diversity of physical and digital scenarios in which the Estonian ID card can be used requires design flexibility, but also strong security mechanisms to resist both physical and digital attacks. For this reason, the Estonian national ID is issued on a smart-card basis; the card consists of an embedded microcontroller and protected memory that implement cryptographic primitives used in authentication and authorisation protocols. Importantly, each ID card contains two digital certificates (i.e., two pairs of public and private keys protected with different PINs), a design that was selected to clearly separate authentication from signing. Because digital signatures in Estonia are equivalent to handwritten ones, assigning a separate signature key increases resilience against many potential attacks, such as attacks that seek to mislead users into signing documents instead of authenticating themselves. To facilitate the use of digital IDs and digital signatures, various services and frameworks have been introduced, for example, DigiDoc, a set of document formats that are based on digital signatures and allow documents to be digitally signed and encrypted using national ID cards.[43]

Another important aspect in the design of Estonia's ID system concerns its *interoperability*. This complex socio-technical matter raises a couple of challenges. First, on the social plane, the national ID includes personal information that in other countries might be considered private and which, therefore, might pose challenges to interoperability between the Estonian ID and other countries. Second, on a technical plane, the technical features of the Estonian ID ecosystem strongly depend on cryptographic protocols that are implemented in both hardware and software, which despite various benefits raises the potential for conventional security problems ranging from faulty implementation to intentional weakening of underlying cryptographic primitives. For example, according to the Estonian ID Card Principles and Solutions documentation,[44] the reliance on foreign software providers has been seen as a strategic national risk; consequently, the digital signature architecture has been designed from scratch (at the centre of this architecture is DigiDoc, which handles the verification of digital signatures). This last point helpfully illustrates the challenge of integration that exists across diverse design constraints.

## IMPLICATIONS OF ESTONIA'S DESIGN CHOICES

In this section we discuss various implications of the Estonian ID's underlying design choices.[45] The focus is on the main design decisions related to the use of the National ID as a public identifier; the use of the date of birth in the

42 Because the card's possession is mandatory, it raises questions about the costs incurred to citizens. Currently, the card application cost is €7 for persons under 15 years, €25 for those who are older. See Estonian Police and Border Guard Board, "Identity Documents," https://www.politsei.ee/en/teenused/riigiloivud/riigiloivu-maarad/isikut-toendavad-dokumendid/index.dot, accessed on 14 March 2016.

43 AS Sertifitseerimiskeskus, "Digidoc Format Specification," Version 1.3, Tallinn, May 2004, http://www.id.ee/public/DigiDoc_format_1.3.pdf, accessed on 30 March 2016.

44 AS Sertifitseerimiskeskus, "The Estonian ID Card and Digital Signature Concept: Principles and Solutions," Version 20030307, Tallinn, March 2003.

45 The data used in this section were collected in interviews of specialists involved in the design and deployment of the Estonian ID scheme. For more information, see the Acknowledgments below.

ID number and its implications on individual privacy; and technical concepts concerning the key management of the ID cards.

**The National ID as Public Information**

The main part of the national ID is generated using low entropy information, such as date of birth and gender. This implies that the information can be easily collected or guessed and and ID generated by an untrusted party. One of the main reasons behind this design decision seems to have been the objective of simplicity with respect to both technical design and human factors (e.g., education). Because the ID number is public, there is no need to implement measures that would be needed if the number were private or secret; such numbers are usually difficult to enforce and their effectiveness difficult to evaluate. A good comparative example is the Japanese approach, which mandates that revelation of the ID number to a third party requires face-to-face confirmation. This cumbersome requirement may result in scalability problems; furthermore, it could incentivise various social engineering attacks. Another important implication of simplicity in design is memorability. In the Estonian case, a person only needs to remember the three digit serial number, because the other ID components (birthdate and gender) are easily known to the user.

**Benefits of the National ID as a Unique Identifier**

A unique identifier guarantees uniqueness across the population. It significantly simplifies the linking of records and offers the possibility of an exact search.[46] Given a specific setting, the unique identifier may actually enhance transparency, because it helps to ensure that where records are linked, they are linked accurately. For example, if a name is used for a search instead of a unique ID, the results will necessarily be imprecise, raising the possibility that the incorrect record will be accidentally retrieved. Assuming a hypothetical scenario where a police officer was to search for a generic name—say, Tom Smith—the officer would potentially review data on many Tom Smiths, including sensitive information, before arriving at the right person. When the unique ID number is searched, however, it appears less intrusive because only the correct and relevant agency- or department-specific information is viewed. Another example in which the unique ID offers an additional advantage may particularly interest private firms: it enables citizens to avoid the use of employee ID cards. In other words, the Estonian-style national ID card easily renders such cards (and their associated bureaucracy) unnecessary, reducing a corporation's expenses in administrating such a programme. Moreover, the ID card system keeps aspects of an individual's data independent from other actors' data. Citizens associated with their employers can transact and sign documents commercially using a personal identity. Within this context it is important to mention that the Estonian law mandates a general rule of non-duplication for databases, meaning that data caching is allowed, but updates must pass up to the master record—in short, no information is stored twice. As previously mentioned, in the Estonian ID system, the Population Registry is the main database, but the overall state information system architecture is decentralised (i.e., citizen data are stored within the primary source database). Such an architecture allows for fine-grade logging and auditing of accesses and queries of individuals' records.

Overall, the Estonian scheme's defining principle of *referencing rather than storing*—i.e., storing data in a single, well-administrated database and using referencing instead of copying—was a crucial design decision. It has a potentially far-reaching consequence: it alleviates the tension that often exists between the usage of a unique ID and concerns about citizens' privacy.

**Privacy Implications of the Unique Public Identifier**

Privacy concerns in Estonian ID system nevertheless exist, as with any other identification scheme. Probably the main concern is the public nature of the national ID number's personal information (date of birth and gender). Within a broader European context, this design choice is probably the most controversial. Some countries may view such a design choice critically, but in others, including the Nordic countries, the privacy concerns are less severe because the local cultures prize openness and transparency.

In general, privacy concerns can be analysed in a cultural and historical context that also considers transparency. When information is private or sensitive, agencies that handle data are encouraged to implement effective auditing procedures such as the extensive logging and auditing mechanisms of the Estonian ID ecosystem, which are governed by different Data Protection Acts and the Estonian Data Protection Inspectorate. Concretely, the Estonian Personal Data Protection Act, the Public Information Act, and the Electronic Communication Act assist in protecting individuals' constitutional rights, which in this context include the right to obtain information about the activities of public authorities; the right to inviolability of private and family life in the use of personal data; and the right to access data gathered

46 In the Estonian case, the use of a single unique identifier was borrowed from the Nordic countries. See Hansteen et al., "Nordic Digital Identification (eID)."

in regard to oneself. Together with the national ID system, this legal framework allows Estonians to trace who accesses their data, when, and for what purposes. For example, it is possible to see which doctors have accessed one's personal data, or if policemen access data illegitimately (e.g., for personal reasons).[47] In Estonia, the creation of new databases is tightly controlled; it requires approval from the inspectorate that assesses privacy and transparency matters.

In this system, accountability rests not on a cadre of internal auditors, but on the self-policing and challenges of the citizens themselves. This approach has strengths and weaknesses. On the one hand, citizens who are able to monitor their own privacy will tend to review their data in much greater depth than one would receive from random audits. On the other, many citizens will take no interest at all because the costs of inspection are born personally (these citizens will derive only the minimal benefit of being within a population where some other citizens bear the costs).

Our assertion that the individual's ID is truly public is somewhat questionable, as one case illustrates. The Estonian Information Commissioner judged that a Lightweight Directory Access Protocol (LDAP) server established to deliver identity certificates for citizen IDs (thereby facilitating strongly secured integrity-protected email) breached the right to privacy because it allowed easy recovery of anyone's date of birth from the ID. As a result, the LDAP server was shut down.[48] This outcome is unfortunate: a more privacy-preserving ID might have enabled this useful service to continue. In our conclusions we revisit the role of the birth date in establishing an ID.

**Threats**

Although the Estonian scheme is quite mature, little information is publicly available regarding the observed prevalence of attacks upon the scheme. We infer that the Estonian scheme (and other modern identity schemes) has witnessed few significant attacks. This observation is potentially instructive: it may vindicate the system's overall design. If no attacks are observed, then either they are successful but stealthy or attackers have found no significant systemic vulnerabilities to exploit. Based on available evidence, we believe that the latter scenario is more likely. This state of affairs should inform any risk analysis constructed as a development of our threat model; in the absence of conclusive empirical data, any prioritisation of risks must be suspect, amounting to little more than a deductive guess. Notwithstanding the presumed robustness of the Estonian scheme against attacks, it is important to consider the scheme's longevity. In the broad design, the present scheme will presumably persist for decades to come; it would be premature to imagine that no significant attacks will occur in that time. The proactive exploration and mitigation of threats, therefore, must carry on, particularly as the context of the scheme's use continues to change and expand.

47 In the latter case, if the offending policeman does not share the data in question, he is merely fired; but if he passes the data on to third parties, he goes to jail.

48 James Sermersheim, "Lightweight Directory Access Protocol (LDAP): The Protocol," Network Working Group Request for Comments 4511, The Internet Society, June 2006; Estonia Ombudsman, "ID Card Holders Personal Code on the Internet (id-kaardi omanike isikukoodide avaldamine internetis)," in *Õiguskantsleri 2006. Aasta Tegevuse Ülevaade (Ombudsman 2006 Overview of activities)*, No. 6-8/061188, Tallinn, 2007, pp. 269–271.

## CONCLUSION

National identity schemes operate in a complex and dynamic socio-technical situation. The requirements that are known to designers when the scheme is first designed may change substantially during its lifetime. No one, for example, could have predicted how the U.S. SSN would be used in databases decades after its introduction for quite a different purpose. Some observers have argued that we may nevertheless learn lessons from what has happened in the past[49]—but we must avoid over-generalisation where circumstances subtly differ.

The requirements of an identity scheme are also subtly shaped by other external factors. Whitely and Hosein argue that the (subsequently abandoned) UK national identity scheme, which was designed around a single unique identifier per person, was shaped heavily by the choice of which government department would implement it (i.e., the Home Office also has responsibility for policing and border control). They also argue that allowing greater transparency (e.g., of credit ratings and mandatory reporting of data losses) would facilitate better protection against identity theft than a scheme solely reliant upon strong biometrics, as the UK scheme had, because the quality and durability of biometric tools continues to evolve.[50]

In the introduction we set out three research questions. A large part of our paper has been taken up with the first set of questions: What are the main design decisions and alternatives for the basis of an identity scheme? What are the impacts of these design decisions upon the functionality and usability of the system built from any given scheme? What are the threats to security and privacy arising from such schemes? We observed that the issues are complex and intertwined, and thus have resisted the temptation to align them into a simple cause-and-effect matrix.

We stated the second set of questions as follows: Are the threats hypothesised in the first question set seen in the actual deployment of a given system? Does Estonian government data confirm the analysis of the first question? These questions have a broadly negative answer; in any security analysis, many more threats are considered than are actually seen in practise. In fact, very few threats appear to have materialised; and few, if any, have arisen that might not have been hypothesised in advance.

Regarding the third question, about whether other nations can adopt all or some technical elements of the Estonian scheme in designing their own identity assurance systems, we conclude that one of the strongest elements of the Estonian scheme is its expectation that an identity is "public" and so may be shared freely with both governments and private industry. This facilitates great utility and simple integration of databases. Such integration is only safe in the context of the strong audit requirements and controls over the creation of new databases, new linkages, etc., and yet many privacy advocates would argue that such transparency is still an insufficient protection of privacy. The choice of an identifier that incorporates a date of birth greatly aids memorability, but may perhaps be judged as a mistake, since sensitivity around dates of birth is unavoidable even where they are not used as authentication secrets. A randomly chosen, public identifier appears to offer the greatest simplicity and usability, with arguably adequate privacy. It remains to be seen whether the next step in privacy, i.e., having multiple identities for different contexts, truly enhances privacy in practice and whether it introduces too many opportunities for errors and mistakes. Presently, a handful of other schemes are pursuing this course; their long-term evolution will be interesting to observe.

The privacy question reveals an important insight: no scheme operates in isolation and, therefore, one country's design decisions inadvertently affect unrelated schemes. For example, one's date of birth may not be regarded as private in some contexts, but it is certainly used as a partially identifying secret in some online services; thus its disclosure through an unrelated system is unwelcome.

What of the international dimension of national identity schemes? Although we have briefly mentioned interoperability between national schemes, it is clear that the international impacts of schemes, and the international context of their use, are of considerable and growing importance. Interoperability of schemes can be managed by standards and harmonised laws, to a large extent; here regional and international bodies such as the European Union have a major role to play.

49 Eleni Gessiou, Alexandros Labrinidis, and Sotiris Ioannidis, "Greek (Privacy) Tragedy: The Introduction of Social Security Numbers in Greece," in *Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society, New York, 2009*, WPES 2009, ACM, New York, pp. 101–104; Zacharias Tzermias, Vassilis Prevelakis, and Sotiris Ioannidis, "Privacy Risks from Public Data Sources," in Nora Cuppens-Boulahia, Frederic Cuppens, Sushil Jajodia, Anas Abou El Kalam, and Thierry Sans, eds., *ICT Systems Security and Privacy Protection* (Berlin: Springer, 2014), pp. 156–168.

50 Edgar A. Whitley and Ian R. Hosein, "Departmental Influences on Policy Design," *Communications of the ACM*, Vol. 51, No. 5 (May 2008), pp. 98–100.

Wider cultural and contextual issues such as approaches to privacy, expectations of government services, and norms around sharing of personal information, however, crucially impact the practical sharing of schemes and data. Alignment of objectives, expectations of relying parties, and clarification of subtle but related design decisions can help to develop such sharing practices. A recent report highlights some of the difficulties in achieving these goals, however, even in the relatively homogeneous context of the Nordic countries.[51] We hope that the conceptual and empirical framework we have begun to develop in this paper can orient and facilitate further investigation of cross-border alignment of national identity schemes.

51 Hansteen et al., "Nordic Digital Identification (eID)."

## About the Cyber Studies Programme

The Cyber Studies Programme seeks to create a new body of knowledge that clarifies the consequences of information technology for the structures and processes of political systems.

Our research mission is (a) to produce scholarly works that contribute to major academic debates and opinions; and (b) to apply these new understandings in the analysis of major policy problems affecting the security and welfare of states and citizens.

Our teaching mission is (a) to support, guide, and train students and researchers in Oxford and beyond in the work and methods of cyber studies within the subdisciplines of political science; and (b) to foster understanding across technical and non-technical communities to promote the development of this new field of study more broadly.

The Cyber Studies Programme is sponsored by the Centre for International Studies in the Department of Politics and International Relations, University of Oxford.

## Acknowledgements

The authors would like to thank Jamie Collier, Lucas Kello, Innar Liiv, and Kris Wilson for assistance in some of our research. We are particularly grateful to interviewees and experts who provided extensive valuable information, including Tarvi Martens, Viljar Peep, and Jaan Priisalu. We are also grateful for the comments of three anonymous expert reviewers.

European Union
European Social Fund

Investing
in your future

This publication is funded by the European Social Fund and the Estonian Government