Risk Analysis on Different Usages of the Swiss AHV Number

Evaluation on behalf of
the Federal Office of Justice and the Federal Data
Protection and Information Commissioner

Prof. Dr. David Basin
September 27, 2017

Contents

Zusammenfassung (Deutsch)

Die aktuelle Organisation und Verarbeitung personenspezifischer Daten in verschiedenen administrativen Registern der Schweiz ist organisch gewachsen, in einer Weise, die aus der Sicht des Datenschutzes als problematisch bezeichnet werden muss. Persönliche, oftmals sensitive, Daten sind in über 14'000 administrativen und organisatorischen Registern gespeichert und mit einem einheitlichen Identifikator, der AHVN13, indexiert. Die entsprechenden Computersysteme und die darin gespeicherten Daten sind anfällig für Attacken durch interne und externe Angreifer, welche sich unberechtigten Zugriff auf die Registerdaten verschaffen können. Dieses Risiko ist nicht unerheblich, da viele der Systeme, welche diese Register speichern und verarbeiten, von Organisationen wie Gemeindeverwaltungen, Schulen und Krankenhäusern verwaltet werden, welche nicht den gleich hohen Sicherheitsanforderungen wie die IT-Systeme des Bundes unterliegen.

Sowohl die AHVN13 als auch Identitätsattribute wie Vorname, Nachname und Geburtsdatum werden in diesen Registern verwendet um Personen mit Daten zu verknüpfen. Falls Daten entwendet werden, sind die zugehörigen Personen deshalb identifizierbar. Darüber hinaus macht es die Verwendung der AHVN13 als einheithlicher Personenidentifikator einfach, Daten aus verschiedenen Registern direkt zu verknüpfen. Dies ermöglicht es Angreifern, umfangreiche Informationsprofile der betroffenen Personen zu erstellen. Diese Datenschutzrisiken werden in Zukunft weiter zunehmen, da einerseits immer mehr Organisationen die AHVN13 für die Datenverarbeitung nutzen und andererseits immer mehr Daten gesammelt, gespeichert und verarbeitet werden, insbesondere in relativ unsicheren IT-Systemen von Kantons- und Gemeindeverwaltungen sowie Nichtregierungsorganisationen.

Da die Registerdaten zusammen mit den zugehörigen Identitätsattributen gespeichert sind, würde das alleinige Ersetzen der AHVN13-Nummern in einem Register durch sektorspezifische Identifikatoren oder andere Pseudonyme die Datenschutzrisiken nicht wesentlich reduzieren. Die Registerdaten könnten durch die Verwendung der in den Registern gespeicherten Identitätsattribute weiterhin mit relativ hoher Präzision mit Personen verknüpft werden. Allerdings gibt es Alternativen zum gegenwärtigen Ansatz, welche Datenschutzrisiken erheblich reduzieren. Diese beinhalten eine Umstrukturierung der Verarbeitung, Speicherung und Absicherung der Registerdaten.

Die nachfolgend aufgeführten Massnahmen würden die aktuellen Datenschutzrisiken erheblich reduzieren, insbesondere jene Risiken, welche die kontinuierliche Ausweitung der gegenwärtigen Verwendungweise der AHVN13 mit sich bringen.

- Einführung von nichtsprechenden Pseudonymen (wie Steuer- oder Krankenkassen-Identifikationsnummern) in einer angemessenen Art und Weise. Hierfür können sektorspezifische Identifikatoren in verschiedenen Varianten verwendet werden. Es ist wichtig zu beachten, dass die Speicherung dieser Identifikatoren direkt zusammen mit anderen Identitätsattributen im gleichen Register minimiert wird.

- Pseudonyme müssen für verschiedene Verwaltungs- und Geschäftsprozesse mit Personen verknüpft werden können. Daher ist es notwendig, sowohl technisch als auch organisatorisch sorgfältig zu regulieren und zu kontrollieren, wie und wann diese Verknüpfungen erfolgen.

- Die Erhöhung der Anforderungen an die Sicherheit und die Qualitätssicherung *aller* Systeme, welche sensitive Daten verarbeiten, die mit der AHVN13 oder sektorspezifischen Identifikatoren indexiert sind. Beides muss über die aktuellen Regelungen des EDI hinausgehen.

Diese Massnahmen sind mit Aufwand verbunden. Insbesondere bietet die Verwendung von sektorspezifischen Identifikatoren nicht die gleiche Bequemlichkeit und Einfachheit, wie dies bei der Nutzung der AHVN13 als ein einheitlicher Identifikator der Fall ist. Ausserdem verursacht der Einsatz sektorspezifischer Identifikatoren in unterschiedlichen Kontexten zusätzliche Kosten bei der Systemanalyse, dem Design und der Implementierung, da berücksichtigt werden muss, wie Identitäten und persönliche Daten in administrativen und anderen organisatorischen Prozessen verwendet werden und dafür angemessene Identifikationskonzepte entwickelt werden müssen. Eine solche Umstellung verändert auch die Art und Weise wie Organisationen mit ihren IT-Systemen arbeiten und wie sie mit ihren Kunden interagieren. Die Entscheidung, wo und wie diese Massnahmen eingesetzt werden, erfordert daher eine Kosten-Nutzen-Analyse, welche ausserhalb des Aufgabenbereichs dieser Studie liegt.

In der Schweiz verfügt der Bund bereits über Erfahrungen mit solchen Massnahmen in unterschiedlichen Kontexten wo sensitive, persönlich identifizierbare Informationen verarbeitet werden, oder er führt entsprechende Projekte durch. Beispiele hierfür sind die Pseudonymisierung von Daten bei statistischen Analysen im Bundesamt für Statistik und aktuelle Projekte zur Einführung sektorspezifischer Identifikatoren für elektronische Patientendossiers und für das Handelsregister. Die Ausweitung dieser Aktivitäten in andere Bereiche wäre aus der Sicht der Risikominderung wünschenswert.

Abstract (English)

The current organization and processing of person-specific data in different administrative registers in Switzerland has evolved historically in a way that is problematic from the data-protection perspective. Personal data, often sensitive, is stored in over 14,000 administrative and organizational registers, indexed with a unique identifier, the AHVN13. The associated computer systems, and the data they store, are subject to attacks, from both internal and external attackers, resulting in unauthorized access to the register data. This risk is non-negligible as many of the systems that store and process these registers are maintained by organizations, such as municipality adminstrations, schools, and hospitals, that are not subject to the same high security requirements as federal IT systems.

Both the AHVN13 and identity attributes, like first name, last name, and birthday, are used in these registers to associate individuals with data. Hence, if data is exposed, then it is personally identifiable. Moreover, the use of the AHVN13 as a standard identifier (where its systematic usage has been approved) makes it simple to directly link data from different registers, enabling attackers to build up extensive information profiles on data subjects. These privacy risks will further increase as more organizations use the AHVN13 for data processing and as more data is collected, stored, and processed, in particular, in relatively insecure, non-federal IT systems within the state administrations, municipalities, and non-government organizations.

Since register data is stored together with associated identity attributes, replacing AHVN13 numbers in a register with sector-specific identifiers or other pseudonyms would, alone, not substantially reduce privacy risks. The register data could still be linked to individuals, with relatively high precision, using the identity attributes stored in the registers. Nevertheless there are alternatives to the current approach that substantially reduce privacy risks, which involve restructuring how register data is stored, processed, and secured.

The following measures would substantially reduce privacy risks compared to the status quo and, in particular, the risks arising from the continued expansion of usage of the AHVN13 in its current form.

- Introduce non-descriptive pseudonyms (like taxpayer or e-health identification numbers) in a principled way and minimize their storage in registers together with other identity attributes. This can be done by using sector-specific identifiers in different variants.
- Pseudonyms will need to be linked to individuals for different administrative and business processes. Hence, it is necessary to carefully regulate, both technically and organizationally, how and when this linkage occurs.
- Strengthen the security requirements and assurance processes, beyond those of the current regulations of the Department of Home Affairs, for *all* systems that process sensitive data indexed with AHVN13 or sector-specific identifiers.

These measures do not come for free. In particular, the use of sector-specific identifiers does not offer the same convenience and simplicity that the use of the AHVN13 offers. Moreover, applying sector-specific identifiers in different contexts incurs additional costs in system analysis, design, and implementation, to account for how identities and personal data are used in administrative and business processes, and to design appropriate identification schemes. It also changes how organizations work with their IT systems and interact with their customers. The decision of where and how to employ these measures therefore necessitates a cost-benefit analysis for the individual scenarios, which is outside of the scope of this study.

Within Switzerland, the federal government has experience, or is running projects, incorporating all of these measures, in different contexts where sensitive personally identifiable information is processed. Examples include the pseudonymization of data during statistical analyses within the Federal Office of Statistics and the current projects introducing sector-specific identifiers for electronic patient records and for the commercial register. The expansion of these activities in other sectors would be desirable from the risk-reduction perspective.

1 Introductory Remarks

1.1 Context

The 13 digit AHV number (abbreviated AHVN13) is an administrative identifier for individual persons. The Bundesrat supports the systematic use of the AHVN13 by the federal, state, and municipal governments to improve the efficiency of IT supported administrative processes [2]. In contrast, the Swiss Federal Data Protection and Information Commissioner supports the introduction of sector-specific identifiers such as, for example, patient numbers provided by the federal legislation for electronic patient records. The argument given in support of sector-specific identifiers is that they are less risky from the privacy perspective. The objective of this study is to independently evaluate the risks associated with these different alternatives.

1.2 Project Scope

The project's scope is focused on the two tasks: summarizing usage scenarios and carrying out an associated risk analysis.

1.2.1 Usage Scenarios

The first task is to determine

1. how, and to what extent, the AHVN13 is used today and
2. how it might be used if its usage were substantially extended beyond that of social insurance, within the federal government, state governments, and municipalities. This includes non-person specific and anonymized usages, which are detached from personal details like name and birthday.

1.2.2 Risk Analysis

Based on the usage scenarios, the second task is to analyze the privacy risks based on the following considerations:

1. The kind of identifier used:
   (a) AHVN13 as the identifier;
   (b) other non-descriptive identifiers, such as E-Health numbers; and
   (c) the combination of AHVN13 internally, with other identifiers used externally, as with the commercial register.
2. With the usage of the identifiers:
   (a) for all areas or restricted to sectors
   (b) by the administration at all levels (federal, state, and municipal governments), possibly with restrictions.

Note that this is a *risk analysis*, as opposed to a *cost-benefit* analysis. The financial costs or efficiency advantages of the different alternatives are outside of the scope of this study.

1.3 Resources

For this analysis, laws and documents were used, both in and outside the public domain, some of which are cited in the bibliography at the end of this report. Interviews were also carried out on the usage of the AHVN13 with six different federal administrative agencies, namely: the Federal IT Steering Unit (FITSU), the Federal Social Insurance Office (FSIO), the Federal Office of Public Health (FOPH), the Federal Statistics Office (FSO), the Central Compensation Office CCO, and the Swiss Federal Commercial Registry Office (EHRA).

AHVN13	Last Name	First Name	Birthday	Nationality	Sex
756.1234 56789.5	Furrer	Luca	29.12.1980	Swiss	M
756.1231 23123.8	Smith	Betty	5.4.1966	USA	F

Figure 1: AHVN13 register except (fictive, omitting optional attributes)

2 Background on the AHVN13

In the next two sections we provide background on the AHVN13 and different usage scenarios. To illustrate how data is organized and used in association with AHVN13 numbers we give fictive examples of registers and other data tables.

2.1 Introduction and Expansion of the AHVN13

The AHVN13 number was introduced in 2008 for the social insurance. In contrast to the previously used 11 digit number, the AHVN13 is 13 digits and is non-descriptive (German "nicht sprechend") in that one cannot link AHVN13 numbers to individuals (e.g., their name, birthday, or other attribute) and vice versa.

Shortly after its introduction [1, 12] the use of the AHVN13 expanded to new domains. In 2009, the AHVN13 was introduced in the resident registers of the municipalities and states to aid in the collection of population statistics, which requires the harmonization of different registers. Since 2012, all education institutions also use AHVN13 in their records to simplify statistics tracking education. Moreover, since 2009, the AHVN13 can be used in explicitly approved administrative and other domains, provided the use is approved by federal or state law. Currently there are over 14,000 organizations approved for the "systematic use" of the AHVN13, see [18]. This means that they can link individuals from clearly defined groups with their AHVN13 and, moreover, they are obliged to follow the regulations laid out by the AHV laws (AHVG, Article 50). For institutions outside of social insurance, this requires explicit application and approval.

2.2 AHVN13 Register

2.2.1 UPI Database

The register linking AHVN13 numbers and individuals is maintained by the Central Compensation Office and stored in their UPI ("Unique Person Identification") database [13]. An individual *identity* is associated with the following five attributes: last name, first name, birth date, sex, and nationality. Moreover, any of the following optional attributes may be additionally be present: single name,

place of birth, and the last and first name of the parents. These attributes are optional in that they may be given to help improve the quality of the register. A fictive example of an excerpt from the AHVN13 table is given in Figure 1, where all optional attributes have been omitted.

The AHVN13 register is built from identity information provided by various sources (Infostar, ZEMIS, Vera, etc.) and rules are used to decide which identity data is to be used to determine a person's "administrative identity" when that person appears in multiple data sources. This administrative identity is the master record ("UPI-Referenzeintrag") identifying the person associated with an AHVN13.

The UPI system supports different operations for assigning AHVN13 numbers to identities, administrating assignments, and querying assignments, using UPIViewer and UPIServices. Querying allows users to retrieve AHVN13 numbers and the identity attributes associated with an identity by providing a subset of the identity's attribute values.

2.2.2 Security

As the CCO is a federal administration, its IT systems, including UPI, fall under the domain the Federal IT Steering Unit, which is responsible for the implementation and security of the information and communication technologies strategy for the federal administration. This means that a multifaceted protection strategy [9] is designed and applied to protect CCO systems. This includes baseline protection, access provisions, networks security, protection requirements analysis, and an information security and data protection concept.

Access to the UPI database, both within and outside the CCO, is restricted to authorized users and administrators. All interfaces to UPI have access control. Within the government network, access to computers requires that users authenticate themselves using a smart card (two factor authentication); to use UPI, a user ID and password are additionally required. Outside organizations approved for the systematic use of the AHVN13 also require public key certificates for web services. The services of the Federal Office of Information Technology, Systems and Telecommunication are used for this purpose, e.g., its public-key infrastructure.

2.2.3 Technical Considerations

We briefly describe several technical considerations related to the AHVN13 that are relevant for the subsequent sections.

The AHVN13 is an example of a *pseudonym*: AHVN13 numbers are values that say nothing about the identities they represent. These numbers are generated using a random number generator, as indicated in Figure 2.

In database terminology, the AHVN13 serves as a primary key for indexing identity records in the UPI database. In general, a *primary key* is an attribute (or attributes) that uniquely identifies each record in a database.

1. Generate a random number between 0 and 999,999,999.
2. Put 756 in front of this number and calculate the EAN13 check digit to obtain a AHVN13.
3. Check if this ANHVN13 already exits in the database. If it already exists, go to step 1 and repeat the procedure until a non-existing AHVN13 is generated.
4. Add the new AHVN13 to the database of numbers.

Figure 2: AHVN13 generation.

The AHVN13 is stored in the UPI database, and also in administrative and organizational registers, with additional identity attributes. These identity attributes can also identify individuals, but not always unambiguously. A *quasi-identifier* is a set of attributes that can uniquely identify (or reduce uncertainty about) most records in a database table. Different subsets of identity attributes may serve as quasi-identifiers. Based on information received from the CCO, the combination of first name, last name, and date of birth is a very high quality quasi-identifier, capable of uniquely reidentifying over 99.98% of the individuals that have been announced to UPI by Infostar (9,291,852 persons). In contrast, the first name and last name alone is a quasi-identifier of poorer quality, that is never-the-less able to uniquely reidentify approximately 75.89% of the individuals from the same population.

Data from multiple database tables can be linked by using the *join* operator of relational databases. This operator combines two tables by combining records in them that are equal on their common attributes. If two tables share a primary key like the AHVN13 then joining them on this key results in a table that unambiguously combines the information from both tables. The same holds if the primary key of the first table is directly associated with the primary key of the second table (known as a *foreign key*, from the perspective of the first table).

Shared quasi-identifiers also allow two tables to be linked, but the quality of the linkage (i.e., whether data from one table is associated with the correct individual in the other) depends on the quality of the quasi-identifier, i.e., how unambiguous the identification is. For example, if the two tables are registers that each have copies from UPI of the first name, last name, and birthdate of an individual, then the resulting quality will be extremely high.

3 Usage Scenarios

Here we describe how, and to what extent, the AHVN13 is used today, and alternative usage scenarios including non-person specific and anonymized usages.

3.1 Administrating Registers directly with the AHVN13

3.1.1 Database Organization

Organizations (at all levels: federal, state, municipal, and non-government organizations), which have been approved for the systematic use of the AHVN13, directly use AHVN13 numbers to administer their own registers. For example, each municipality has its own resident register, and each school has its own student register. The UPI database is then used to identify each individual in the register and associate them with a unique identity: their AHVN13 along with the associated attribute values retrieved from the UPI database.

The CCO recommends that when the AHVN13 is used, directly or indirectly over UPI, for an administrative register and stored in an information system, then it must be archived *with all identifying attributes received from the UPI register*, including the five mandatory attributes listed in §2.2.1 [17, page 3]. This means that copies of records from the UPI database are duplicated when they are stored locally at each administration.

3.1.2 Impact of Direct Disclosure

Organizations approved for the systematic use of the AHVN13 are required to control access to their registers [16]. However, a successful attack on an organization's information system exposes *both* its register-specific data as well as its local copy of (parts of) the UPI database. This means that when considering privacy risks for different usages of the AHVN13, one must consider *all* usages in an overarching way across organizations approved for systematic use. The compromise of any one of these organization's administrative register may not only reveal sensitive personal information, it also provides (partial) information on the contents of the UPI database.

3.1.3 Linkage with other Registers or Data Sources

Figure 3 gives an example of a fictive data set that follows the CCO's recommendation. This data set might, for example, be stored in a hospital database for processing insurance claims. The two tables above the horizontal line comprise the local register. The *Local AHV+ID Table* is the local copy of an AHVN13 table, augmented with a column containing *IDs*. These *IDs* are foreign keys, which serve as primary keys for second table, the *ID Insurer Table*, which contains domain-specific attribute values. Information from these two tables can be joined using the *ID* attribute. Moreover, the *Local AHV+ID Table* can be joined with any other register indexed with the AHVN13, by joining on the AHVN13 attribute.

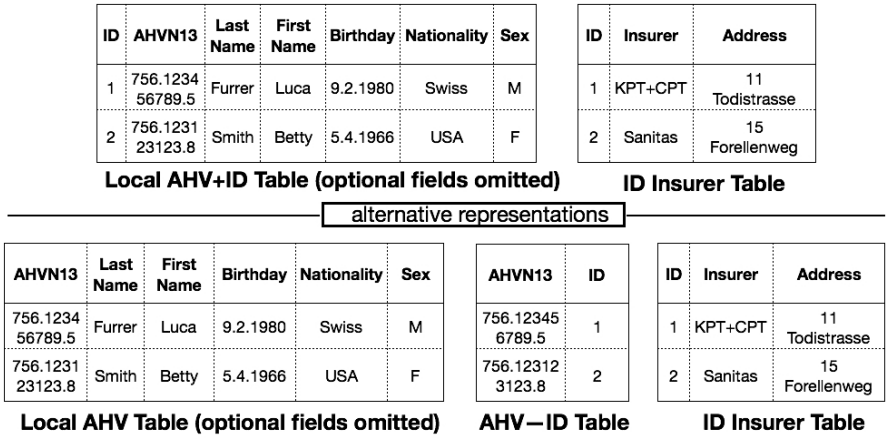


Figure 3: Two equivalent representations of an administrative register (omitting additional attributes).

Technical remark on equivalent representations. The following is relevant for subsequent analysis. There are different logically equivalent ways to represent the same information in a database system. One could represent the information by just one table, that is the join of the two tables above the line. Alternatively, one could factor the *Local AHV+ID Table* into the first two tables showed below the line, whereby the tables above the line represent the same information as the tables below the line. It is up to the local database administrator who sets up the data model to choose between representations.

Below the line, the *AHV-ID* table serves to directly link the *Local AHV Table* and the *ID Insurer Table* via the join operator. We will call such tables *linkage tables*; there are many equivalent names in the database literature, e.g., *associative tables*, *join tables*, etc.

3.2 Registers using sector-specific Identifiers

Sector-specific identifiers are currently being implemented for two sectors: healthcare and the commercial register. These sectors follow different approaches and illustrate how sector-specific identifiers can be used in sectors with very different data protection requirements. Patient healthcare data is extremely sensitive and should be kept private. In contrast, commercial register data should be made public, but in a privacy-respecting manner. In the following, we examine sector-specific identifiers generally and then describe these two particular examples.

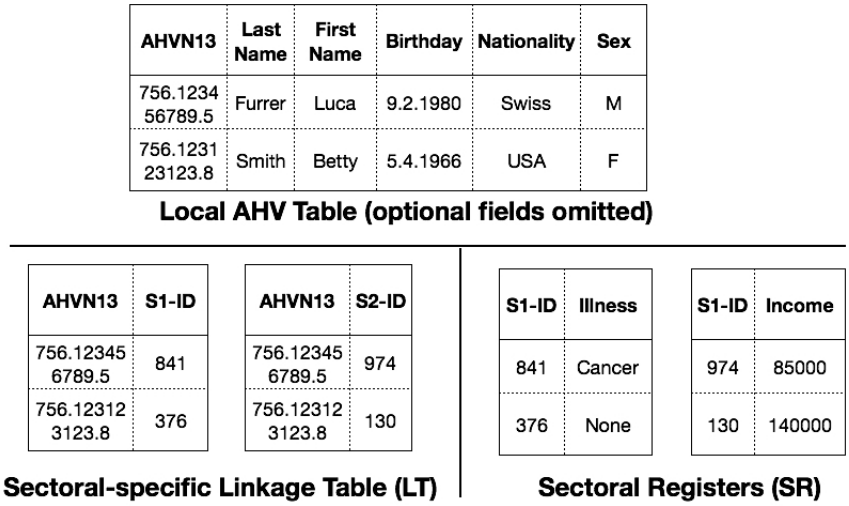


Figure 4: Sector-specific identifiers (omitting additional attributes)

3.2.1 General Setup

The idea behind sector-specific identifiers is that, rather than using the AHVN13 to identify individuals (where its systematic usage has been approved), one instead uses sector-specific identifiers. Records within a sector are then stored locally with the sector-specific identifiers, rather than the AHVN13 numbers. This idea is general and can be used for any notion of a sector. For example, one could have sectors corresponding to taxpayer information (with taxpayer identifiers), student (with student identifiers), etc.

Figure 4 illustrates the basic idea in a scenario with two sectors. The table above the line is an excerpt of the AHVN13 register administered by the CCO. Bottom-left, we have examples of the two linkage tables, one for each of the two sectors, that associate AHVN13 numbers with sector-specific identifiers. Finally, bottom-right, we have two sectoral registers, which are stored and managed in the information system of organizations in the respective sectors. Note, in general, within each sector there may be many such sectoral registers, one for each organization that maintains its own register.

In this scenario, an individual may seek the services of an organization within the sector, identifying himself, for example, with a sector-specific identity card that provides his sector-specific identifier and additional attributes like a picture and his first and last name (which need not be stored in the sector-specific database). His data can then be accessed, without reference to his AHVN13 or the need to convert his sector-specific identity to an AHVN13.

At a high-level, there are some similarities between this approach and previously described approach where the AHVN13 was used across all records. In particular, Figure 3 (below the line) resembles Figure 4 where *ID*s are replaced

with sector-specific identifiers. In both cases, AHVN13 numbers are linked to *ID*s, which are linked to domain-specific register data. However, there are several significant differences, which support data protection.

1. Sector-specific identifiers are intended to be shared only within a sector, but not more widely. As they are shared within a sector, one can link (accessible) data tables across the sector, indexed by sector-specific identifiers. But, as they are *not* shared across sectors. one cannot directly combine data tables from different sectors. This would require access to the sector-specific linkage tables, which is restricted. Moreover, one cannot link data with other registers that use the AHVN13 number. This makes it more difficult to obtain comprehensive profiles of individuals.

2. An AHVN13 number is associated to an individual a lifetime long. This need not be the case for a sector-specific identifier. When data tables do not contain quasi-identifiers like identity attributes, the use of sector-specific identifiers supports "the right to be forgotten"; the link between data and individuals can be deleted simply by deleting the association between a sector-specific identifier and the corresponding AHVN13 number.

3. The linkage tables can be stored in information systems with higher security requirements than the sectoral data itself. Storing these independently makes it more difficult for an attacker to compromise both the sectoral registers and the linkage table together.

3.2.2 Generation and Storage

Sector-specific identifiers are constructed to be both non-descriptive and non-invertible. Noninvertible means that given a sector-specific identifier without a linkage table, it is not possible to determine the corresponding AHVN13. Moreover, given sector-specific identifiers from different sectors (e.g., IDs 841 and 974 in the above example), it should be impossible to determine if they are associated to the same AHVN13 number.

For both health data and the commercial register, sector-specific identifiers are generated randomly, when needed, and stored in a linkage table associated with an AHVN13. The generation uses a procedure similar to that used to generate AHVN13 numbers, described in Figure 2, differenting essentially only in the identifiers' format. Sector-specific identifiers can also be generated from identifiers such as AHVN13 numbers using standard cryptographic building blocks, such as hash functions or block ciphers, provided the properties mentioned above hold.

As indicated above, a linkage table is produced which must be stored and protected. From the database perspective, this table could be stored anywhere, e.g., by the CCO or by some organization in the sector.

3.2.3 Example: E-Health

[7] describes the recent federal law on electronic patient records [4] and the associated requirements. Electronic patient records are virtual records containing

patient-specific medical data. Patients administer the access rights to their own records and can make their records available during treatment at hospitals and other healthcare institutions. The data stored in electronic patient records is, not surprising, highly sensitive. The data is administered by non-governmental health-service communities (German: *Gemeinschaften* and *Stammgemeinschaften*) responsible for providing electronic services, as well as local healthcare providers such as doctors, hospitals, and pharmacies.

There are, in essence, three levels of organizations, three different kinds of identifiers, and two different kinds of linkages.

Level 1: The CCO generates sector-specific identifiers called *PINs* (Patient identification numbers, also known as Pat.-IDs or EPD-PIDs). It generates them randomly, as described previously, and stores them in a linkage table, linking them to the AHVN13. As an additional security measure, the PINs stored in the linkage table are encrypted.

Level 2: There are 4-8 health-service communities. Each of these communities maintains its own *Master Patient Index (MPI)* table. This table (local to the health-service community) links both:

- PINs to the MPI's primary key, called the MPI-ID, and
- the MPI-ID to the different "local identifiers" used by the different healthcare providers to index patients in their own local databases.

Level 3: This consists of the different healthcare providers (doctors, pharmacies, etc.), within communities. Each healthcare provider stores medical data on persons they serve. The primary key for this data is a provider-specific "local identifier", which is again a pseudonym.

This schema provides flexibility in how data is linked. The PIN, but not the AHVN13 (which is not stored by the communities), is used to link data between different health-service communities. This is necessary for cross-communication queries to gather data from other communities. The PIN–AHVN13 linkage is used to ensure that individuals are consistently named inside the communities and to link the person to other IDs (e.g., local IDs) inside the MPI. It can also be used if data is approved to be linked outside of the healthcare sector, for example by the Federal Office of Statistics.

Security and Privacy The objective of using different identifiers at different levels is to limit the impact of a cyber-attack: if an attacker breaks into a hospital's information system, he should not learn identifiers that he can use throughout Switzerland to identify patient-specific data. Moreover, he should not be able to link these documents with data from other areas of the patient's life. This is unfortunately only partially the case because personal attributes are stored in the MPI and local healthcare registers, which serve as quasi-identifiers. Moreover, the patient data itself is sometimes labelled with these identifiers or even the patient's AHVN13.

The use of the PIN support the right to be forgotten since deleting the linkage between an AHVN13 and a PIN deletes the association with an individual. Unfortunately the effectiveness of this is also limited by quasi-identifiers at the level of the MPI and the local healthcare providers.

The security of the databases depends, of course on the measures taken within the health-service communities and by the healthcare providers. The security standards and assurance measures for the communities are regulated by the law and associated requirements [4]. The security of the IT systems used by the healthcare providers (e.g., hospitals) are to be checked by the health-service communities; it is unclear how well this will work in practice.

3.2.4 Example: Commercial Register

The commercial register is a public source of information about registered businesses in Switzerland. It is decentrally organized at the state level. Register entries are created by state administrations (KHRA), which contain information on companies registered in Switzerland and associated individuals, e.g., the owners and those persons with signature authority. The Eidgenössisches Amt für das Handelsregister (EHRA) oversees the KHRAs. Moreover, since 1990, it uses the data from the state registers to administer ZEFIX, which is a support infrastructure that consolidates information from the state registers and enables a Swiss-wide Internet search for information on Swiss companies.

The commercial register is currently undergoing modernization [6, 11] so that individuals associated with companies can be systematically identified. This will support administrative processes that are currently difficult because the comercial register information is decentralized and individuals in the registers are not named in a uniform way. Specifically, the EHRA will maintain a *Reference Person Register (RPR)* that will contain information on individuals registered in Switzerland (and are in UPI) as well as on foreign residents with businesses in Switzerland. The EHRA will introduce a sector-specific identifier, the *RprPersonId*, which will be the primary key for identifying individuals in the RPR.

The EHRA will provide IT infrastructure whereby KHRA employees can work with the RPR when creating their register entries to ensure that individuals are consistently named. For data protection, the AHVN13 will not appear in comercial register entries. Instead the sector-specific identifier RprPersonID will be used. The linkage between the RprPersonId and the AHVN13 will be maintained separately by the EHRA. The EHRA is considering publishing the RprPersonID in the SHAB (Schweizerisches Handelsamtsblatt) entries and in the register excerpts.

Security and Privacy The commercial register illustrates an alternative way of achieving data protection using sector-specific identifiers. In this case, some personal attributes, such as the names of individuals associated with Swiss companies, must be made public. While these could be directly associated with the AHVN13 internally for disambiguation, the use of a sector-specific

identifier reduces the risk of linkage outside the sector. Nevertheless, depending on which public attributes are used in commercial registry entries, it may still be possible to unambiguously reidentify individuals and perform reasonably high quality linkages with other registers and data sources, using these attributes as quasi-identifiers. There is no simple solution here, as there is a tradeoff between utility and privacy.

3.3 Cross-Register Analysis

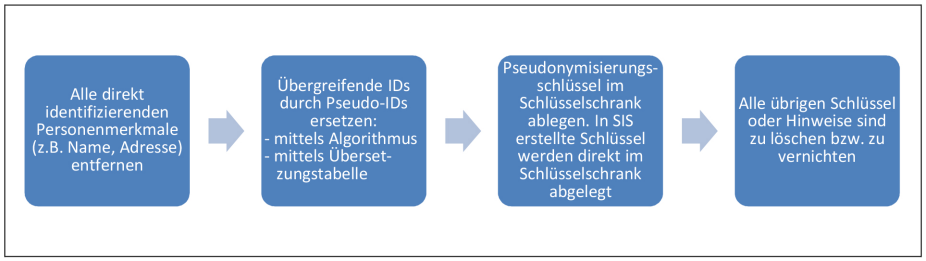
The previous sections indicate that care must be taken, even when replacing AHVN13 numbers with other pseudonyms: it is necessary to replace not only the AHVN13, but also all quasi-identifiers. The Federal Statistics Office, which combines and processes different registers for statistical purposes, provides a good example of how this can be done.

The harmonization of different registers using the AHVN13 makes it easier to check the consistency of registers and to carry out cross-register statistics. The Federal Statistics Office carries out both of these activities, with mandates given by the Register Harmonization Laws of 2006 and 2012. In particular, (1) the FSO offers services for the states and municipalities to check the consistency of their registers with the AHVN13 register. Moreover, (2) the FSO supports different agencies in performing cross-register statistics that require combining different registers and data sources.

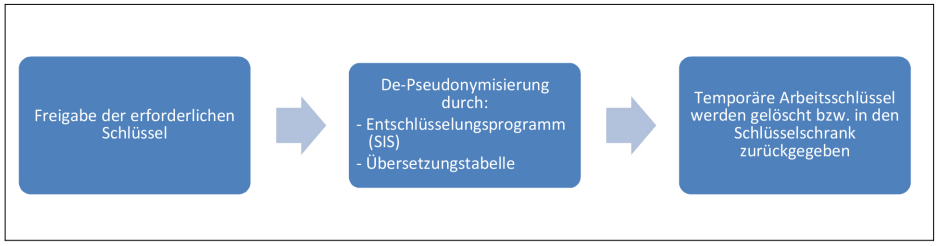
Cross-register analysis within the FSO is strictly regulated with respect to both data protection and data security [15]. All analysis must be approved and afterwards this analysis takes place using pseudonymized data. Combining data requires an appropriate key, and is logged and audited. In particularly sensitive cases, additional measures are required.

The FSO has very high standards for security, enforcing best security practices. This includes strong authentication, fine-grained, role-based access control, key management, clearly stipulated data anonymization and pseudonymization procedures, and cryptographic protection of data in storage and transit. Furthermore standard data protection principles are respected concerning purpose-based processing, transparency, minimality, etc. Particularly noteworthy is that the sensitivity of the processed data is accounted for during processing and more sensitive data (e.g., information on religion, political viewpoints, race, health, criminal record) has higher protection requirements. In this sense, processing is risk-based.

As an example, to illustrate the care taken in processing sensitive data, we describe in more detail how the FSO pseudonymizes data and how this pseudonymous data is processed [15, §5.3]. For pseudonymization, all direct personal identifiers in a data set (e.g., name, address) are replaced with pseudo-identifiers that reveal no information on the original identifiers. These pseudo-identifiers are cryptographically generated in a two step process: the first step uses an FSO-cryptographic key (or a translation table) and the second step uses a sector-specific key. This is also done in a way such that in different studies, different pseudo-identifiers are generated, so that one cannot link data from



Pseudonymization Process



De-pseudonymization Process

Figure 5: Processing Pseudonyms (from [15])

different studies. After generation, the keys are stored in protected storage, and deleted from the production environment.

De-pseudonymization is needed if data from different data sets or sectors must be combined for (approved) analyses. The details depend on the IT environment, but essentially the cryptographic keys must be retrieved from their secure storage and used to link the original *IDs* with the pseudo-identifiers, at which point the pseudo-identifiers can be replaced with the original *IDs*. Any external data sets must be imported into the FSO for this processing.

4 Risk Analysis Methodology

There is no universally agreed upon standard methodology for privacy risk assessments. Moreover, in this report we are not analyzing the privacy risks for a particular system, but rather different ways of setting up data processing ecosystems sharing common features in how they manage identities in privacy preserving ways. Hence we will focus on the risks associated with these different features. For this, we will use risk analysis principles based on the United States National Institute of Standards report on privacy engineering and risk management in federal systems [3].

4.1 Risk and Privacy Objectives

"Risk is a measure of the extent to which an entity is threatened by a potential circumstance or event, and is typically a function of: (i) the adverse impacts that would arise if the circumstance or event occurs; and (ii) the likelihood of occurrence." [3, §3.2]

Privacy risks arise from *problematic data actions* that cause adverse effects on individuals. These actions are carried out by *threat agents*, also called *attackers*. The actions can arise from unauthorized system usage, in particular the *unauthorized* access to data. Hence security risks can be privacy risks and a (classical) security risk analysis is part of a privacy risk analysis. In addition, privacy risks can arise during the *authorized* processing of personally identifiable information (PII).

Problematic data actions can violate an organization's privacy objectives or conflict with legal or regulatory requirements. Privacy researchers have put forth a list of general privacy objectives in the form of "Fair Information Practice Principles" [3], which we list in Appendix B. Principles particularly relevant for this report are the *minimization* of collected and processed PII and that data usage should be *purpose-specific and use-limited*.

4.2 Analysis based on Engineering Characteristics

A privacy risk analysis usually requires examining at a *particular* system from the technical and organizational perspective and analyzing the different ways that threat agents can carry out problematic data actions, the likelihood of this occurring, and the resulting impact. This is not possible in our study since there is no way to practically analyze the systems of all existing organizations approved for the systematic use of the AHVN13 currently, and in different extended usage scenarios.

The ability to link information between registers also has implications for our analysis. Namely, when considering privacy risks, one cannot consider the risks to individual systems in isolation. Instead, one must consider privacy risks of the entire administrative ecosystem as a whole. In particular, how are individuals identified, and how their identity information is processed together with their data, in both governmental and non-governmental systems.

The approach we therefore take is to broadly consider which system characteristics increase (or decrease) overall privacy risks. We can then coarsely estimate the privacy risk of different approaches to organizing identity information within systems, or compare their risk levels, by considering their adherence to these principles. [3, §3.1] provides a reference point for our analysis, identifying the following three general *privacy engineering objectives*, which stipulate core characteristics of privacy-preserving systems.

Predictability: A predictable system enables reliable assumptions by individuals, owners, and operators about personally identifiable information and its processing by an information system.

This is the core for building trust and accountability. De-identification techniques play a role in predictability by providing evidence that the information disclosed by systems is in line with individuals' expectations.

Manageability: A manageable system provides the capability for granular administration of PII including alteration, deletion, and selective disclosure.

This is necessary for administrations to be able to identify and correct inaccurate information, dispose of obsolete information, and ensure that only the necessary information is collected or disclosed.

Disassociability (alias unlinkability): This enables the processing of PII or events without association to individuals or devices beyond the operational requirements of the system.

As explained in [3, §3.1.1.3]: "Disassociability captures one of the elements of privacy-preserving systems — that the system actively protects or 'blinds' an individual's identity or associated activities from exposure. Unlike confidentiality, which is focused on preventing unauthorized access to information, disassociability recognizes that privacy risks can result from exposures even within an authorized perimeter. Disassociability advances the capabilities of a privacy-preserving system by engaging system designers and engineers in a deliberate consideration of points of exposure that are not essential for the operation of the system. In this way, it is most closely associated with capabilities that could be used to implement [data minimization]."

Information systems that violate these principles have a higher risk of privacy breaches.

Summarizing, our analysis methodology is to consider the different usage scenarios described in §1.2.1 and analyze their effects on these privacy engineering objectives. In particular, we will focus on the *security* of systems, since vulnerabilities can undermine all three of the above objectives, and on *linkability*. We will use this analysis to produce a rough qualitative estimation of the resulting privacy risks of the different alternatives.

5 Risk Analysis

We proceed by first highlighting relevant classes of motivated and skilled threat agents. Afterwards we examine system characteristics relevant for our scenarios and their effects on security and privacy risks. Finally, we analyze the privacy risks associated with the different scenarios.

5.1 Who might attack Swiss Registers?

As explained in §4.1, privacy risks arise from actions taken by attackers. In this section, we document that there are indeed attackers with sufficient motives and capabilities to access data in different Swiss registers.

For data registers, *internal users* may be the source of problematic data actions. In general, attacks and malfeasance by internal users is a serious and often underestimated problem [14]. An internal user may use his authorized access for unauthorized purposes. For example, he may be curious about the status of a given individual or wish to help a colleague in another agency. Disgruntled employees may seek to cause harm, e.g., to harm the reputation of their employer for revenge. Internal users with criminal intent may use their authorized access to acquire data to sell on the blackmarket, where personal data can be sold for use in identity theft or targeted advertising. Finally internal users may be negligent. While this itself may not be a problematic data action, users' failure to properly secure their systems, minimize the storage of sensitive data, delete data, etc., can make it easier for external attackers to access it.

Attackers may also be external to the system. Hackers are interested in the challenge of breaking into systems, and publishing sensitive data online may be seen as a badge of success. Politically motivated hactivists may carry out problematic actions to humiliate their opponents. Criminals break into systems, with the motive of selling stolen data. Nation-state adversaries may be interested in sensitive data including financial data, health data, or criminal records, which could be used to acquire intelligence on individuals, e.g., to catch tax evaders or influence elections. An example of the skill and effort nation-state attackers successfully apply to break into systems is provided by the recent Melanie report on the espionage case at RUAG [8].

We provide two concrete examples below that illustrate the incentives for collecting and aggregating personal data (first example) and the kinds of damage that can arise from the compromise of even a single system (second example).

Moneyhouse.ch The first example, taken from [5], illustrates how personal data may be monetized. It also shows how individual data sources can be linked together to provide a personal profile, where the data, in sum, reveals a significant part of a person's life. Finally, it illustrates that these risks can arise even without actively breaking into systems, but simply by linking publicly available non-pseudonymized data.

The Swiss company itonex AG ran the platform moneyhouse.ch, which collects and sells data about companies and private persons. At the time of

the report [5], the company offered a range of both free and paid information services, whose value is derived from linking different data sources, including data from administrative registers such as the commercial register. By linking data, itonex produces profiles of individuals that are available over their platform or as the result of google searches. Paying customers could access data on individuals including their first and last name, place of residence, zip code, age, birthday, job, household members (including children), and neighbors. The housing data included links to aerial photographs and Google streetview pictures of the property as well as information on the property's type, the number of households in the building, building costs and period, and the number of floors. Overall, the linked data provided a composite picture of the living situation of identifiable individuals, which is an important and sensitive aspect of their personal profile.

Equifax The second example comes from the United States. In September this year Equifax, one of the largest credit bureaus in the US, reported a data breach due to a vulnerability in a web application [19, 10]. Hackers exploited this vulnerability to access the names, addresses, Social Security numbers, credit card numbers, drivers licenses, and "dispute documents with personally identifying information" for 143 million Americans, almost half of the US population.

The impact of this breach is high. The data will circulate on the blackmarket for decades to come, enabling different forms of identity theft. As explained in [19]: "Typically, fraudsters mix stolen Social Security numbers, and potentially other information from the owners, with a borrowed mailing address and apply for new credit cards that they control. Some patient con artists even use the new personas to seek additional credit cards or loans, then max them all out at once, potentially making off with tens of thousands of dollars."

This example again shows the risk of aggregating personally identifiable information. There are motivated attackers with strong incentives to steal and monetize the data. Moreover, system security is imperfect and access control and other mechanisms can be defeated. Equifax is a company whose mission is to be a trusted authority on consumer credit information and presumably has high security standards. Finally, the impact will be long term: social security numbers (like AHVN13) are assigned permanently, and other personal attributes are difficult to change. This compromise will be a long-term source of problems for all affected.

5.2 System Characteristics affecting Privacy Risks

Here we make some general observations that we will subsequently leverage for our scenario comparison.

5.2.1 Perfect Security is an Illusion

Computer systems are never 100% secure in practice. Security is a prerequisite for data protection as one must prevent unauthorized access to data. Unfortunately,

there is always a risk of successful attacks exploiting weaknesses in hardware, operating system, applications, people, processes, etc. Hence systems may be compromised and the data they store may be read by attackers, even when systems are protected by state-of-the-art security measures.

5.2.2 Not all Systems are equally Secure

A system's security depends on the effort spent to secure it and the effectiveness of the measures taken, both technical and organizational. Assurance activities are also important to validate that these measures actually work as intended. Federal systems (see the discussion in §2.2.2), usually enjoy a relatively high level of security due to the high standards and assurance processes of organizations like the FITSU, which have responsibility for these systems.

The situation is different for non-federal systems administered by organizations approved for the systematic use of the AHVN13 at the level of states and municipalities, or non-government organizations like hospitals and schools. The federal government stipulates lower protection requirements and weaker assurance processes for these non-federal systems [16] (although these requirements may be supplemented by additional requirements from other administrations). These systems tend, in general, to be substantially less secure.

5.2.3 Handling of Identity Information

There are different ways of organizing register data in information systems, as explained in §3.1. In general, it is a privacy risk to make local copies of parts of the UPI register. The compromise of sufficiently many (less well protected) systems suffices for the attacker to reconstruct large parts of the UPI database. Moreover, it becomes a burden to keep redundant copies consistent.

Another privacy risk stems from the fact that UPI data is generally stored in administrative registers together with identity attributes (see §2.2.3) and the identity attributes alone are enough to both reidentify data and link tables with high precision (see §2.2.3). Hence removing just AHVN13 attributes or quasi-identifiers alone from administrative registers will not substantially reduce privacy risks. **Data protection is improved by decoupling administrative and organizational records from *both* the AHVN13 *and* quasi-identifiers capable of reidentifying large parts of the population.** Note that this requires a different approach to organizing data then is currently recommended to organizations approved for the systematic use of the AHVN13 (see §3.1.1).

5.2.4 Where and How Data is Stored

This point is closely related to the last one. Not only are there different ways to *organize* data in an information system, the data tables can also be *distributed* across separate databases and platforms in a distributed information system. Distribution, when done properly, can offer the same functionality as centralized storage (e.g., records can be linked as needed for business purposes), but has better security and privacy properties then centrally stored data.

We will illustrate this with an example based on Figure 4. Consider the following two scenarios:

Scenario 1: For each sector, the sectoral register is stored in the *same* database as the sector's linkage table.

Scenario 2: For each sector, these two tables are stored in different databases, on different computers. To support business processes, distributed query processing is supported, e.g, one may map a sector-specific identifier into an AHVN13 by combining data from the two databases, when this is (absolutely) necessary for business purposes. To support security, all communication is appropriately authenticated.

In the first scenario, a software vulnerability in the database storing a sectoral register could be exploited by an attacker to compromise *both* data tables for that sector. This would enable the attacker to read out all their contents and join the two tables, thereby linking the sectoral register to AHVN13 numbers. In the second case, the compromise of the sector's register would not allow this reidentification or further linkage outside of the sector. In particular, that would *additionally* require the adversary to compromise either (i) the database containing the linkage table or (ii) to compromise the (authenticated) communication channel used for distributed query processing to spoof queries within the distributed system. Hence reidentification or combining data from different registers is more difficult in the second scenario and the security and privacy risks are therefore correspondingly lower.

Note that one can make the difference between the two scenarios even more pronounced by considering additional security measures in the second scenario, e.g., hardened systems, encrypted storage, key management using hardware security modules, etc. In a suitably designed system, such measures can reduce the risk of certain kinds of attacks and hence the overall privacy risks; the flip side is an increase in the system's complexity and cost.

5.3 Scenario-specific Comparison

We return to the scenarios listed in §1.2 and compare their associated risks.

5.3.1 Extended Usage of the AHVN13

There are currently over 14,000 organizations approved for the systematic use of the AHVN13, many with relatively insecure systems. Hence, even in the status quo, there is a high risk that attackers can compromise systems and extract data sets containing both AHVN13 numbers and personal attributes. Hence, the leakage of even a single register may expose sensitive data that is personally identifiable. Moreover, given the use of the AHVN13, if multiple registers are compromised, their data can be more easily combined.

These risks will increase as more organizations use the AHVN13 in their information systems. This increased risk stems from two factors:

1. As more organization collect, store, and process personal data associated with AHVN13, the likelihood further increases that some of their systems will be compromised.

2. As more organizations outside of the federal government are approved to systematically use the AHVN13, then more data, which is directly and unambiguously linkable to individuals with the AHVN13 number, will be collected, stored, and processed in relatively insecure, non-federal IT systems. The increased risk stems from the lower security standards and weaker assurance processes for these systems and that they were typically not designed with security as a priority (see §5.2.2).

With respect to both of these factors, note that for those organizations that currently are already are collecting and processing personal data stored with identity attributes that serve as high-quality quasi-identifiers (e.g., first name, last name, and birthday), their data is already at risk, both for reidentification and linkage using the quasi-identifiers. In this case, the additional risk from *additionally storing* the AHVN13 is marginal (see §2.2.3 and §5.2.3), simply because reidentification and linkage can already take place.

By the above reasoning, one can also argue that there is no reason *not to store* the AHVN13 in registers, as this will have minimal impact on privacy risks. This argument has been put forth to justify the current approach taken and its expansion. The conclusion is basically correct *but only for the current setup* where AHVN13 numbers are always accompanied by identity attributes. But storing these identity attributes in the same register is not a necessity. The redundant storage of identity information is neither good database design nor good privacy practice.

The important question then is how the privacy risks associated with the status quo, and its expansion, compare to alternative scenarios that can potentially reduce the risks by incorporating higher standards (§5.2.2) and reorganizing, minimizing (§5.2.3), and distributing (§5.2.4) the storage of identity attributes and the associated possibly sensitive information. This is what we consider in the next two cases.

5.3.2 Other non-descriptive, sector-specific Identifiers

The AHVN13 is a non-descriptive identifier used across different sectors. In contrast, sector-specific identifiers are also non-descriptive identifiers whose usage is restricted to be *local* to a sector. When a sector-specific identifier is used, then registers in that sector cannot be directly linked to registers in other sectors. Hence, when implemented in a principled way, this measure reduces the risk of an attacker aggregating large quantities of data about individuals across sectors, revealing significant parts of their life.

We again emphasize that given the *current approach* to storing register data with identity attributes, simply replacing AHVN13 numbers with sector-specific identifiers would not substantially reduce privacy risks. The register data can still be linked to individuals using the identity attributes as quasi-identifiers.

These identity attributes also enable database tables to be linked with high precision.

It is possible and desirable to introduce sector-specific identifiers as an alternative to the AHVN13, but this must be done in a principled way. In particular, it is necessary both to minimize the storage of identity attributes and adequately protect all linkage tables (see §5.2.2–§5.2.4). **Introducing sector-specific identifiers in this way would thereby reduce the security risks with respect to usage of the AHVN13 in the status quo and its continued expansion.**

Doing this may require changing existing software architectures. Scenario 2 in §5.2.4 gives one example, suggesting that distributed query processing may be needed for some business cases. As another example, consider the healthcare domain, discussed in §3.2.3. There identity attributes should not be stored in the master patient indices of the health-service communities and their storage should be minimized in the databases of the healthcare providers and in healthcare documents themselves. One would instead primarily work with the associated pseudonym. Patients, in turn, would have to identify themselves with their pseudonym (for example on an e-health card); other identity attributes could also be presented, but these additional attributes should not be stored.

When sector-specific identifiers must be linked to identity information for administrative processes, this can still be done using (distributed) query processing using the linkage tables. These tables should additionally have very high protection requirements and extra measures should be taken to secure them and ensure they are only used for appropriate purposes. There are numerous design options here. For example, the linkage tables can be stored by the CCO, or within a sector-specific administration like the EHRA, or within a private organization like a healthcare community. It is possible to realize all these options with substantially lower privacy risks than the direct use of the AHVN13, provided the linkage tables are appropriately protected.

5.3.3 Combination of AHVN13 Internally with other Identifiers used Externally

The commercial register is an example of such a combination. In principle, the only difference between a sector-specific identifier used internally versus externally, is that externally available data requires no effort for an attacker to acquire. For less sensitive data, like commercial register data, this design can make sense, in particular when that data should be in the public domain. In this case, **the use of the external sector-specific identifier lowers the risk of attackers aggregating data from registers outside of the sector over the direct use of the AHVN13.** However, the level of risk-reduction depends on how easily and effectively individuals can be reidentified from other identity attributes also published.

References

[1] AHVN13 Information from Central Compensation Office. https://www.zas.admin.ch/zas/de/home/partenaires-et-institutions-/navs13.html and pages and documents accessible from there.

[2] Breitere Verwendung der AHV-Nummer. https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen/bundesrat.msg-id-65458.html.

[3] Sean Brooks, Michael Garcia, Naomi Lefkovitz, Suzanne Lightman, and Ellen Nadeau. Introduction to Privacy Engineering and Risk Management in Federal Sytems. Technical report, National Institute of Standards and Technoogy Internal Report 8062, January 2017.

[4] Bundesgesetz über das elektronische Patientendossier. https://www.admin.ch/opc/de/classified-compilation/20111795/index.html, 2015.

[5] Empfehlung des Eidgenössischen Datenschutz- und Öffentlichkeitsbeauftragten (EDÖB) gemäss Art. 29 des Bundesgesetzes vom 19. Juni 1992 über den Datenschutz (DSG, SR 235.1) betreffend die von itonex AG angebotenen Dienstleistungen unter www.moneyhouse.ch. https://www.edoeb.admin.ch/datenschutz/00626/00747/01245/index.html, january 2014.

[6] Erläuternder Bericht zur Änderung des Obligationenrechts (Handelsregisterrecht und Anpassungen im Aktien-, GmbH- und Genossenschaftsrecht) sowie des Revisionsaufsichtsrechts. https://www.bj.admin.ch/dam/data/bj/wirtschaft/gesetzgebung/handelsregister/vn-ber-d.pdf, 2012.

[7] Gesetzgebung Elektronisches Patientendossier. https://www.bag.admin.ch/bag/de/home/service/gesetzgebung/gesetzgebung-mensch-gesundheit/gesetzgebung-elektronisches-patientendossier.html.

[8] GovCERT.ch. APT Case RUAG, Technical Report. https://www.melani.admin.ch/melani/en/home/dokumentation/reports/technical-reports/technical-report_apt_case_ruag.html, May 2016.

[9] Informatiksteuerung des Bundes: Sicherheit. https://www.isb.admin.ch/isb/de/home/ikt-vorgaben/sicherheit.html and pages and documents accessible from there.

[10] Brian Krebs. Breach at Equifax May Impact 143M Americans. https://krebsonsecurity.com/2017/09/breach-at-equifax-may-impact-143m-americans, 2017.

[11] Modernisierung des Handelsregisters. https://www.bj.admin.ch/bj/de/home/wirtschaft/gesetzgebung/handelsregister.html.

[12] Brian Olivier, Jerome Brugger, Angelina Dungga, Esther Hefti, Thomas Selzam, Andreas Spichiger, and Katinka Weissenfeld. Gutachten: AHV-Nummer als einheitlicher organisationsübergreifender Personenidentifikator. Technical report, Berner Fachhochschule BFH, E-Government-Institut, February 2002.

[13] UPI–Benutzerhandbuch. Zentrale Ausgleichsstelle. https://www.zas.admin.ch/dam/zas/de/dokumente/Partenaires%20et%20institutions/UPI/UPI-Handbook-1-07.pdf.download.pdf/UPI-handbook-v1-07D.pdf, April 2017.

[14] David M. Upton and Sadie Creese. The Danger from Within. Harvard Business Review. https://hbr.org/2014/09/the-danger-from-within, September 2014.

[15] Verknüpfungsrichtlinien (Bundesamt für Statistik). https://www.bfs.admin.ch/bfsstatic/dam/assets/2760554/master, February 2017.

[16] Verordnung des EDI über die Mindeststandards der technischen und organisatorischen Massnahmen bei der systematischen Verwendung der AHV-Versichertennummer ausserhalb der AHV Das Eidgenössische Departement des Innern. https://www.admin.ch/opc/de/classified-compilation/20071554/index.html, 2009.

[17] Verwaltung der AHVN13 in Drittregistern. Zentrale Ausgleichsstelle. https://www.zas.admin.ch/dam/zas/de/dokumente/Partenaires%20et%20institutions/NAVS13/Regles-de-gestion-du%20NAVS13.pdf.download.pdf/RÃÍgles_gestion_V_1.3_D.pdf, May 2012.

[18] Verzeichnis der systematischen Benutzer der AHVN13. VerzeichnisdersystematischenBenutzerderAHVN13.

[19] Brian Womack. Equifax's Historic Hack May Have Exposed Almost Half of U.S. Bloomberg. https://www.bloomberg.com/news/articles/2017-09-08/equifax-s-historic-hack-may-have-exposed-almost-half-of-u-s, September 2017.

A Abbreviations

AHVN13: 13 digit AHV (*Alters- und Hinterlassenenversicherung*) number

AHVG: AHV Law

CCO: Central Compensation Office

EHRA: Swiss Federal Commercial Registry Office

FITSU: Federal IT Steering Unit

FOPH: Federal Office of Public Health

FSIO: Federal Social Insurance Office

FSO: Federal Statistics Office

ISC:

KHRA: State-level commercial registry administrations

MPI: Master Patient Index

PIN: Patient Identification Numbers (aliases: Pat.-IDs, EPD-PIDs)

RPR: Reference Person Register

RprPersonID: Sector-specific identifier, serving as primary key for the RPR

SHAB: Schweizerisches Handelsamtsblatt

UPI: Unique Person Identification database

<

B Fair Information Practice Principles

Access and Amendment: Agencies should provide individuals with appropriate access to PII and appropriate opportunity to correct or amend PII.

Accountability: Agencies should be accountable for complying with these principles and applicable privacy requirements, and should appropriately monitor, audit, and document compliance. Agencies should also clearly define the roles and responsibilities with respect to PII for all employees and contractors, and should provide appropriate training to all employees and contractors who have access to PII.

Authority: Agencies should only create, collect, use, process, store, maintain, disseminate, or disclose PII if they have authority to do so, and should identify this authority in the appropriate notice.

Minimization: Agencies should only create, collect, use, process, store, maintain, disseminate, or disclose PII that is directly relevant and necessary to accomplish a legally authorized purpose, and should only maintain PII for as long as is necessary to accomplish the purpose.

Quality and Integrity: Agencies should create, collect, use, process, store, maintain, disseminate, or disclose PII with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to ensure fairness to the individual.

Individual Participation: Agencies should involve the individual in the process of using PII and, to the extent practicable, seek individual consent for the creation, collection, use, processing, storage, maintenance, dissemination, or disclosure of PII. Agencies should also establish procedures to receive and address individuals' privacy-related complaints and inquiries.

Purpose Specification and Use Limitation: Agencies should provide notice of the specific purpose for which PII is collected and should only use, process, store, maintain, disseminate, or disclose PII for a purpose that is explained in the notice and is compatible with the purpose for which the PII was collected, or that is otherwise legally authorized.

Security: Agencies should establish administrative, technical, and physical safeguards to protect PII commensurate with the risk and magnitude of the harm that would result from its unauthorized access, use, modification, loss, destruction, dissemination, or disclosure.

Transparency: Agencies should be transparent about information policies and practices with respect to PII, and should provide clear and accessible

notice regarding creation, collection, use, processing, storage, maintenance, dissemination, and disclosure of PII