Using a branch specific pseudonymous personal identification number for a register based census

Eva-Maria Asamer, Statistics Austria

Introduction

With the new General Data Protection Regulation (GDPR), awareness of data protection of individuals rose all over Europe. At the same time, administrative data is being used more and more to reduce costs and the respondent burden. In Austria, data protection regulations have already been strict before the new GDPR went into force. So in the eGovernment Act, a way of exchanging data within public administration using a pseudonymisation procedure was established. For a register based census, it is necessary to link data on micro level, but personal identifiers like names are not needed at any stage of the process.

E-government in Austria

The legal basis for electronic government processes in Austria is the eGovernment Act, came into effect on 1 March 2004. This law serves as the legal basis for eGovernment instruments and components. The Austrian eGovernment strategy is based on the principals Proximity to citizens, Convenience through efficiency, Trust and security, Transparency, Accessibility, Usability, Data security, Cooperation, Sustainability, Interoperability and Technological neutrality.[1] To meet the principal of Data Security in accordance with the other principals, the system of branch (or sector) specific personal identifiers was developed. This ensures that only authorised persons within the administration have access to personal data.

This system of eGovernment and secure pseudonym data transfer is also used in laws concerning statistics like the Federal Statistics Act and the Register-based census Act. While other governmental organisations use pseudonymised data for re-identifying the subjects in their register, these keys are used by Statistics Austria to match data on a personal level. This way there is no need for using personal identifiers such as names, which are always of concern of data protection.

Source PIN

The concept of an electronic ID (eID) starts with identification attributes, with which a person can be found in a base or source register. The base register for persons is the Central Register of Residents (CRR). For persons without residence in Austria the base register is the Supplementary Register for persons, as shown in figure 1. From the CRR number a so called source PIN is calculated using the Triple-DES method. This method is a symmetric encryption algorithm. The triple designation means, in essence, that the DES encryption process is repeated three times. The abbreviation DES stands for Data Encryption Standard. With this encryption method, the CRR number cannot be derived from the source PIN, apart from the Source PIN Register Authority.

[1] E-Government ABC, Digitales Österreich

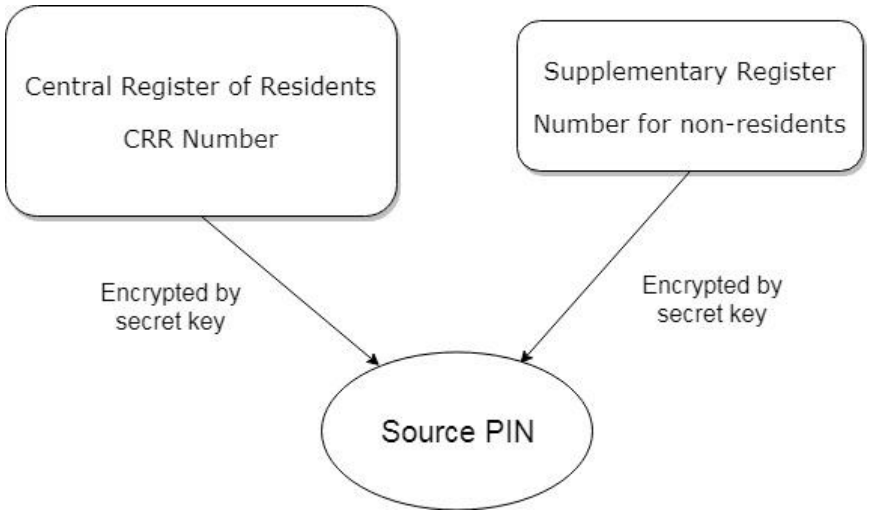

Figure 1 Base Registers for Source PIN

This source PIN is stored on the citizen card of a person, but nowhere else. Otherwise, it is computed each time it is needed, so the Source Register is a virtual register. For legal persons, the entry number in the Commercial Register or the Central Register of Associations or the registration number in the Supplementary Register is used as the SourcePIN, so this system can be used even further. The rest of the paper will concentrate only on SourcePINs (and branch specific PINs) for natural persons, as these are being used for the register based census in Austria.

Citizen card and Mobile phone signature

The Citizen card is a concept which was originally developed for physical cards like the e-card (card for the Austrian health system). It is a form of electronic identification for using online. People can use it to identify themselves by digital means to a public authority. A fundamental characteristic of the citizen card is a qualified electronic signature that can be generated with it and that makes it possible to sign forms or contracts which normally require a handwritten signature. The citizen card is available in many different formats, since it does not depend on a particular type of technology and does not necessarily have to be a "card". In many cases, the carrier medium is a chip card. Using a card reader (and additionally entering a signature PIN) one can use this electronic identity to identify himself for administrative processes on the internet.

In 2009 a second option was established. The so-called mobile phone signature can be used with any mobile phone. In contrast to the card-based citizen card, installing software and additional hardware (card reader) is no longer necessary. The mobile phone signature functions are similar to the solution banks use for e-banking. After successfully logging in with the access code (mobile phone number) and a password, a TAN code is sent via SMS to the activated mobile phone. When the TAN code is entered in the respective application, a qualified electronic signature is created. The mobile phone signature offers a user-friendly alternative to the well-known card-based citizen card, which is especially attractive for users who use it infrequently.

In 2016 an app was developed which sends the TAN automatically to make electronic administrative processes even more convenient for citizens while keeping high security standards at the same time.[2]

[2] Background Information, Citizen Card

Branch specific PIN

For data exchange within public administration in Austria the source PIN is not used directly, but for each public sector or branch, a specific PIN (bPIN) is calculated. This is done by encrypting the source PIN with a one-way function, a secure hash algorithm. This algorithm takes the value of the branch into account as well. So for the same branch always the same bPIN is calculated, whereas for two different branches two different bPINs are calculated for the same person. This relationship is shown in Figure 2.

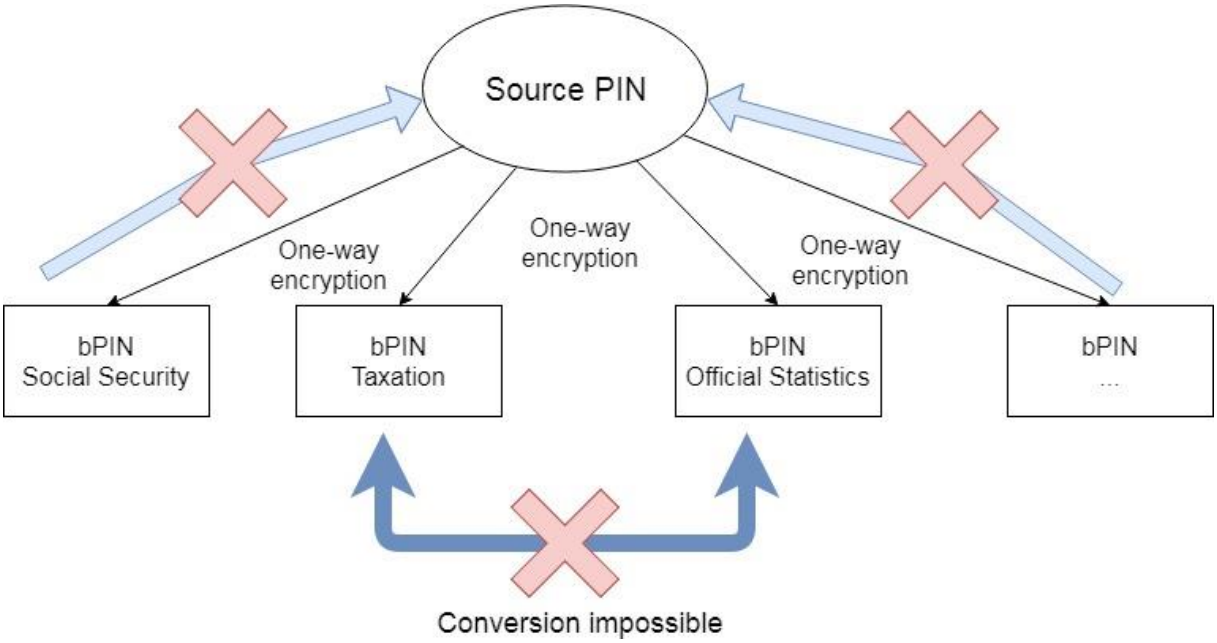

Figure 2 - Calculation of branch specific PINs

To exchange data between authorities the bPIN is encrypted further using a RSA encryption method with a key length of at least 1024 bit. The decrypted bPIN has 28 alphanumeric digits, whereas the encrypted one has 172 alphanumeric digits. This process is shown in Figure 3. As one can see, each time the bPIN is encrypted, a different value is calculated, as a time-stamp is included in this procedure. For data transfers, only these encrypted bPINs are used.

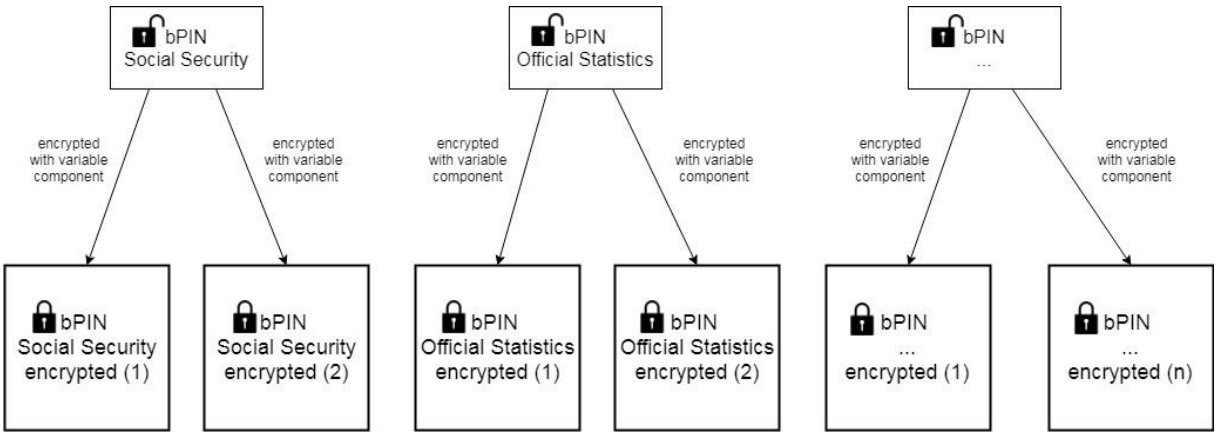

Figure 3 Encrypted bPINs

This encrypted bPIN can only be decrypted by the authority in the branch this bPIN belongs to. There the key is stored in a safe environment. The decrypted bPIN is then the same within the same sector. This step is illustrated in Figure 4.

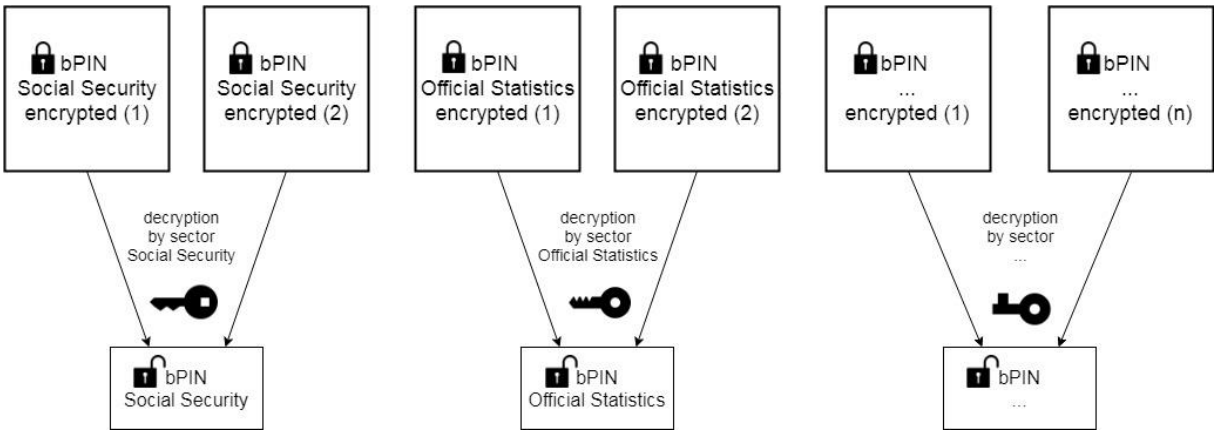

Figure 4 Decryption of encrypted bPINs

To illustrate how the codes in the Austrian system look like, an example for those PINs is given in Table 1.

Base Number	000247681888
SourcePIN (base 64)	Qq03dPrgcHsx3G0IKSH6SQ== (24 digits)
bPIN	j/NxdRQhp+tNyE9WhHdBSYuy3hA= (28 digits)
Encrypted bPIN	qX4/Mf2bMeop0/8tjHqS+OWox03/TViPmP6DoB+Z/ h2gDtMQE99xuBhfzyCy6jXgVEbuFGIqYSU1qxMeRe Qd4bbJzhekXvcrFAAn6mO1ZClokZnmRekidHI6bHnmR0cQjUyw gHjnpbGJIzqBOOXmdFEi2mZ59yKKdMW7yfwQviAs (172 digits)

Table 1 Example for PINs

According to the eGovernment Act the methods which are used to derive a Source PIN and a bPIN must be published, only the values for the seed for encrypting must be confidential to guarantee security.[3] So transparency is guaranteed while at the same time personal data is protected.

Register based Census

The Census in Austria is a register based one where the bPIN system is very important. Here we have to combine the information from more than 50 different data holders. Therefore we need a unique key for every individual, the bPIN_OS. This is accomplished by the Source Register Authority due to name, date of birth and other personal information via the Central Register of Residents as descripted in detail in the section before. The register authority has to order those bPINs in an encrypted form, which have to be transmitted to Statistics Austria. Statistics Austria decrypts the bPIN_OS and uses this key as a common matching variable.

[3] Bildung von Stammzahlen und davon abgeleiteten Personenkennzeichen, Bundesministerium für Digitalisierung und Wirtschaftsstandort

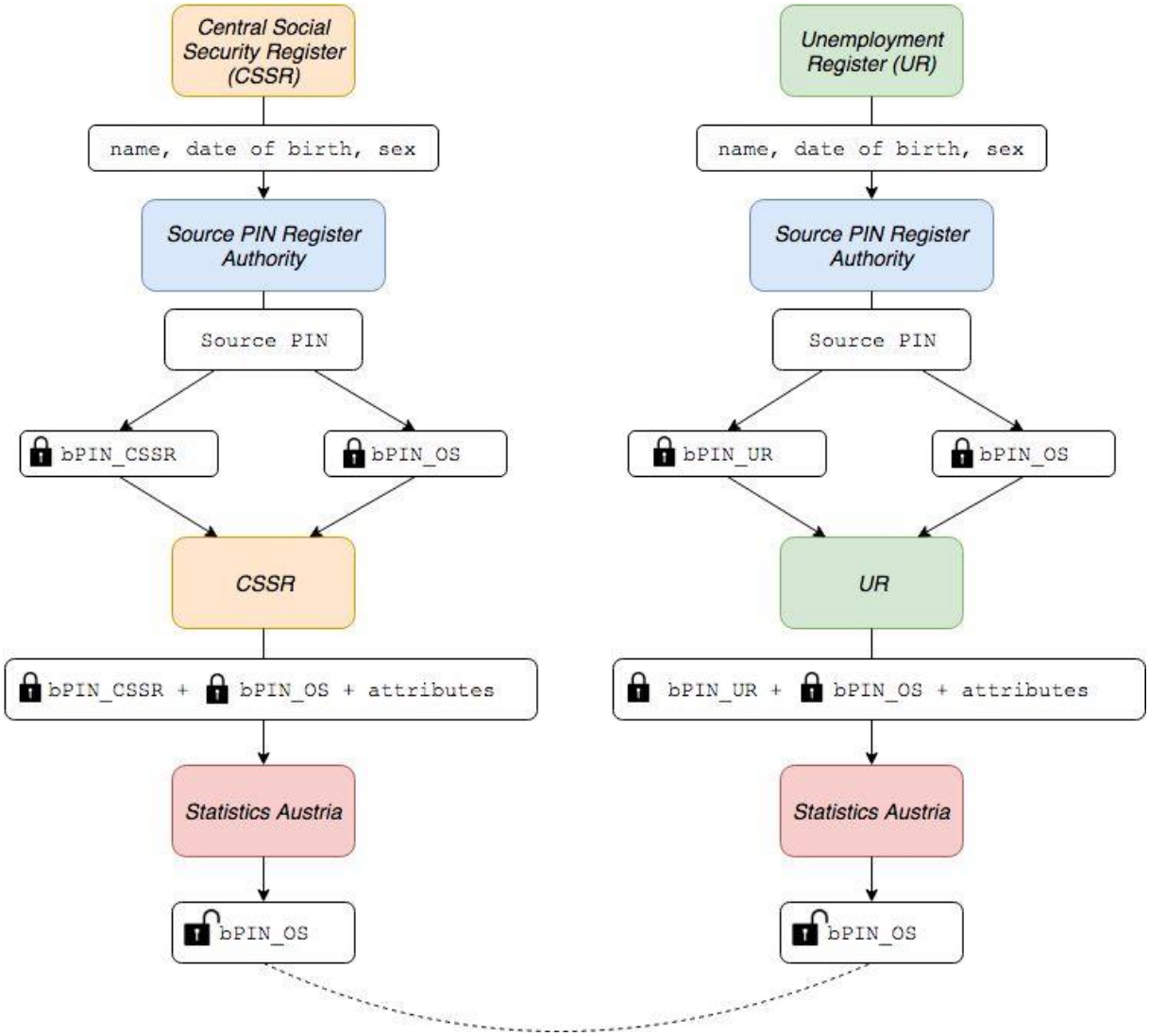

Figure 5 - Process of data delivery to Statistics Austria

In Figure 5 you can see an example of this process. The data holders – for instance the Central Social Security Register and the Unemployment Register - send the name, date of birth and sex of an individual to the Source-PIN Register Authority. The authorities identify the person in the SourcePIN Register and derive the bPIN. Every individual has two of the 172-digit bPINs: one for the data holder and the other one for official statistics - which is called bPIN_OS. Both bPINs are sent back to the data holder, who provides the bPINs as well as the required attributes of the individual to Statistics Austria. Only Statisics Austria is able to decrypt the bPIN_OS into a 28-digit bPIN and use them for merging the data of different data holders. We collect the data only with this outlined bPIN, without names.

Due to the fact that the registers in Austria were not connected with each other in the past and that the data has been collected independently, they often contain different values for the same variables of the same person. Therefore, collecting data only from one register is done to improve the quality of the variables. To get data of a satisfactory quality, the principle of redundancy has been used: Data on sex, date of birth, nationality, place of residence - to mention just a few examples - is collected from as many registers as possible. Inconsistent data has to be checked and converted to plausible values, if necessary, also with the help of the registers concerned.

The following bullet points outline the most important principles of the concept in case of data linking:

- Eight "base registers" are used. If more than one register contains this variable, a decision was made about which register to use as the base register for each variable. Many variables are also collected from several "comparison registers", which are used to confirm the values in the base registers (principle of redundancy).

- The data will be collected without the registered names and without the social security PIN, which both have to be replaced by the branch specific personal identification number (bPIN_OS) in the data deliveries to Statistics Austria for data protection reasons.

- The data from the registers is matched using the bPIN_OS and afterwards checked for consistency and adjusted according to plausibility rules.

In practice, there are some data sets where the quality of the identifying variables is very low, for example the names and dates of birth in the registers are sometimes inaccurate. "Data-twins" (especially when there is no place of birth in a data source) and other people who cannot be found by the procedure mentioned before and therefore cannot be linked, pose problems. Statistics Austria solved those problems with data linking procedures using identification variables like date of birth (also only the month or year), sex and address, but without names within the meaning of the data protection processes mentioned above.

Conclusions

The Austrian bPIN system proved to be very useful for the register based census and our annual population statistics for the fiscal equalisation as well as the annual register based labour market statistics. Data owners are now used to the system. Important data holders like the main association of social security funds or tax authorities are using bPINs now for other applications, too. In their own interest they clean their data, leading to better bPIN_OS at the same time.

The most important advantage of the bPIN is that the data protection is always guaranteed. Moreover, the communication between citizen and administrative authorities as well as between register authorities is very easy. Data matching is possible, which is the core of the register based census. The implementation of the bPIN system took some time because the register authorities had to implement new processes and databases. They had to find a routine because it was something brand-new. The register based census pushed the implementation of the bPIN and now it’s a well-known procedure.

Literature

E-Government ABC, Digitales Österreich (2017), https://www.digitales.oesterreich.gv.at/documents/22124/30428/E-Government-ABC.pdf/b552f453-7ae9-4d12-9608-30da166d710b (in German).

Back Ground Information, Citizen Card, https://www.buergerkarte.at/en/background-information.html

Digitalisation for citizens, Federal Ministry for DIGITAL and ECONOMIC AFFAIRS (2019), https://www.en.bmdw.gv.at/Digitalisation/Digitalisation_for_citizens/Seiten/default.aspx

Bildung von Stammzahlen und davon abgeleiteten Personenkennzeichen (bPK), Bundesministerium für Digitalisierung und Wirtschaftsstandort, https://www.bmdw.gv.at/DigitalisierungundEGovernment/Stammzahlenregisterbehoerde/Veroeffentlichungen/Seiten/Bildung-von-Stammzahlen-und-davon-abgeleiteten-Personenkennzeichen-%28bPK%29.aspx (in German)

Lenk, M. (2009), Methods of register based census in Austria, Vienna, https://unstats.un.org/unsd/statcom/statcom_09/seminars/innovation/Innovation%20Seminar/StatisticsAustria_register%20based%20census.pdf

Hackl, P. (2010) Using Administrative Data at Statistics Austria, Legal Provisions https://ec.europa.eu/eurostat/documents/1001617/4339944/OSTAT-migration.pdf/d62db930-a7f6-4b6a-94c4-f133aa8cb2e1