Using Administrative Data at Statistics Austria: Legal Provisions

Peter Hackl
Statistics Austria

Abstract: The paper describes a pseudonymisation technique that has been developed for e-government processes in Austria and is used in the context of official statistics (i) for delivering person-related data from administrative sources and (ii) for storing and matching person-related data within Statistics Austria. The register-based census is used for illustrating the technique and its application.

1. Introduction

At Statistics Austria, the pseudonymisation of statistical units is an important prerequisite for the statistical work. The two reasons for this situation are:

1 §15 of the Federal Personal Statistics Act requires that Statistics Austria purges – for the sake of data protection – all person-related data that could allow backtracking from statistical data sets as soon as possible; this applies to both natural and legal persons. Pseudonymisation makes the storing – and matching – of such micro-data legal and feasible.

2 Again for the sake of data protection, personal data from administrative sources can legally be provided to Statistics Austria only after pseudonymisation of the identity of the units.

In both situations, the pseudonymisation allows for substantial cost-reducing effects. The pseudonymisation procedure has been devised in the course of establishing e-government as personal data protection safeguard. The paper describes in section 2 the pseudonymisation concept and the procedure of its application. Section 3 illustrates the application of pseudonymisation in the register-based census approach of the Austrian 2011 census.

2. Pseudonymisation of data

Pseudonymisation is a procedure by which all person-related data within a data record are replaced by an artificial personal identifier that maps one-to-one to the respective person; the pseudonym allows tracking back of data to its origins. In contrast, anonymized data are purged from all person-related data that could allow backtracking; tracking back of data is not possible. In this context, person-related data may refer to persons and enterprises.

In Austria, the concept of pseudonymisation of person-related data is used for the communication between authorities and between authorities and citizens or enterprises; the pseudonymisation process is integrated in the e-government strategy [1]. The system of pseudonymized data transfer is also prescribed in regulations concerning statistics, e.g., in the Register-based Census Act [2]. For Statistics Austria, the artificial personal identifier is used to allow storing and matching data on a personal level and is a possibility to receive person-related data from administrative sources in a way that is compatible with the legal prescriptions of person-related data protection. The use of personal identifiers such as names is dispensable in compliance with §15 of the Federal Personal Statistics Act and with expectations of the Austrian data protection agency.

2.1 Sector-specific personal identifiers

As mentioned above, pseudonymisation means replacing all person-related data within a data record by an artificial personal identifier. For individuals, the artificial personal identifier is derived by applying cryptographic one-way functions on the personal identification number, the sourcePIN, which corresponds to the identifier of the person in the Central Register of Residents; input to these functions are the name, the date of birth, and the sex of the person. The cryptographic function is customized to the government body or sector for which the data are pseudonymized; hence, the artificial identifier is called a sector-specific personal identifier and is denoted, e.g., by bPK-as (bereichsspezifisches Personenkennzeichen, Amtliche Statistik) if it is calculated for Statistics Austria. Various bPKs can be calculated for a person such as bPK-health, bPK-tax, etc. Similarly, bPKs are designed and obtained for enterprises. The bPKs are calculated by the E-gov Authority (Stammzahlenregisterbehörde) established in the Chancellery as a part of the Data Protection Commission (DPC). The E-gov Authority is the only place where a bPK can be tracked back to the person or enterprise.

The bPK has the following characteristics:

1 The bPK guarantees that the person-related data correspond to a uniquely identified individual. As a consequence, registers that are based on bPKs show only minor error rates with respect to the correspondence to individuals.

2 For a specific individual, always the same bPK is obtained; hence, the bPKs allow matching of data – from different sources – that refer to a certain individual.

Statistics Austria stores person-related data for each person together with the bPK-AS that corresponds to the respective person or enterprise. The data can be result from a survey or can be transmitted from an administrative body. If data of a specific individual are contained in several data sets, the bPK-AS allows matching the data. E.g., data of a person from the labor force survey can be matched with income data of the same person that are provided by the tax office.

2.2 The transaction of person-related data

If a file of person-related data is to be transmitted from one administrative body, say the tax authority, to another body, say Statistics Austria, the sender has to ask the E-gov Authority to provide for all individuals in the file the bPK that corresponds to the receiving body, here the bPK-as. To calculate these bPKs, the E-gov Authority needs for each person the name, the date of birth, and the sex. The E-gov Authority provides the owner of the data, the tax authority, with the bPKs in encrypted form; the tax authority sends the data of each individual together with the encrypted bPK-as and a the encrypted bPK that corresponds to the tax authority to Statistics Austria. The latter bPK serves to identify the respective record if further inquiries of Statistics Austria are necessary. Only the receiving body is able to decrypt (decipher) its "own" bPKs and to use these bPKs for the data matching.

As a general rule, administrative bodies may store the bPK of another body only in encrypted form.

3. An illustration: The register-based census

In the year 2011 Austria will conduct a register-based census. The legal basis is the Register-based Census Act [2] of 16 March 2006 which in particular defines all sources of administrative data that has to be made available to Statistics Austria for that purpose. This act also stipulates

that a test census was to be carried out with the reference date 31 October 2006 together with a 0.3% sample census that allowed assessing the concepts and implementation of the register-based census.

For most of the variables that are subject of the census, data are available in public registers each of them being established by one of a number of administrative, mostly governmental bodies. These registers, however, have been established over that past by legal regulations in a widely uncoordinated way; the responsibilities for the content and in particular for the quality of the data are not always clearly defined. To a certain degree the contents of the registers are overlapping which gives the chance to apply consistency and plausibility checks and to develop strategies for identifying the most plausible values.

The following points outline the central principles of the register-based census; see [1].

1 The backbone of the census are eight "base registers": For each census variable, a base register has been selected based on criteria of data quality. If the variable is also available in one or several other registers, these registers are adopted as "comparison registers" for that variable; the comparison registers are used for consistency checks of the values in the base registers, for confirmation and in case of divergences for conversion to plausible values (principle of redundancy).

2 The Register-based Census Act stipulates that the owner of the various registers deliver their data to Statistics Austria, the person-related data like names and social security PIN replaced by the bPK-AS in encrypted form. The data from the various registers are matched and checked for consistency and in case of divergences adjusted according to plausibility rules.

For the test census the data linking procedure was used as described in section 2. Figure 1 illustrates the generation of the bPKs corresponding to three administrative registers on the basis of the sourcePIN by the E-gov Authority of the Data Protection Commission; the locks attached to the bPKs indicate that the bPK is encrypted.

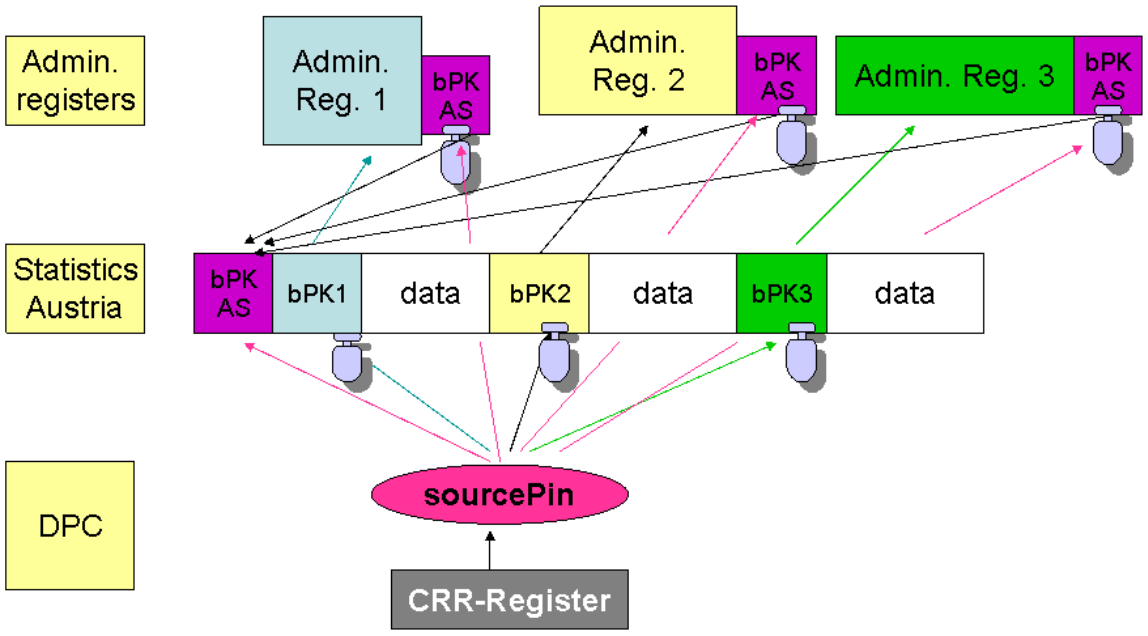

Figure 1: Data collection and matching for census

As an illustration the census results for number of persons are given: The Central Register of Residents (CRR), run by the Ministry of Interior, is the base register for the population number. From 8,323.733 registered units in the register, 77.813 or 0,9% are not found in any other register. Matching the Central Register of Residents with that of the social security authority confirmed 98,8%; together with 14 other comparison registers, among them the tax register, plausibility checks and adaptations increased the rate of evidence to 99,1%. By means of the encrypted bPKs of the delivering administrative bodies, decisions for the open cases were reached by asking persons and communities of their residence for further information. In general, the percentage of cases where the information of the base register is confirmed by the comparison registers is not as high as in the example shown above. However, the concepts of the register-based census turned out to work well. Moreover, the delivery of data including bPKs and the co-

operation in this context with Statistics Austria was a strong first step for many of the delivering administrative bodies towards improved quality of their registers.

4. Conclusions

The pseudonymisation technique that has been developed for e-government purposes has proved an important element of the new Austrian register-based census concept but also for applications in the statistical work in general. It allows storing and matching person-related data activities which are not legal without bPK due to §15 of the Federal Personal Statistics Act that requires for the sake of data protection the purging of all person-related data that could allow backtracking from statistical data sets as soon as possible. In addition, this pseudonymisation technique opens a way for efficient transactions of data from administrative sources like governmental registers to Statistics Austria. A first full-range application was the test census with the reference date 31 October 2006 that showed that the 2011 census should work well. The delivery of data including bPKs and the co-operation in this context with Statistics Austria was a strong first step for many of the delivering administrative bodies towards improved quality of their registers.

5. References

[1] Bundesgesetz über Regelungen zur Erleichterung des elektronischen Verkehrs mit öffentlichen Stellen (E-Government-Gesetz), BGBl I Nr 10/2004, www.ris.bka.gv.at/

[2] Bundesgesetz über die Durchführung von Volks-, Arbeitsstätten-, Gebäude- und Wohnungszählungen (Registerzählungsgesetz), BGBl. I Nr. 33/2006, www.ris.bka.gv.at/

[3] Lenk, M. (2009), Methods of register based census in Austria, Vienna.