How to Protect Patients' Rights to Medical Secrets in Official Statistics

a report by

Drs David-Olivier Jaquet-Chiffelle *and* **Jean-Paul Jeanneret**

University of Applied Sciences of Bern and Swiss Federal Statistical Office (SFSO)

Dr David-Olivier Jaquet-Chiffelle

Dr Jean-Paul Jeanneret

Since 1997, Dr David-Olivier Jaquet-Chiffelle has been a full-time Professor of Mathematics and Cryptology at the University of Applied Sciences of Bern, Switzerland. In 1996, in collaboration with the SFSO, he developed the concept that is described in this article, after joineing the Swiss Federal Section of Cryptology (Swiss Government) as a scientific expert in 1994. From 1992 to 1994, he was a scientific associate at the University of Neuchâtel and worked in collaboration with the University of Bordeaux. Following his degree, he went to Harvard University to pursue his research. Dr Jaquet-Chiffelle was also a lecturer in the Department of Mathematics, having received his PhD degree in Mathematics in 1991.

Dr Jean-Paul Jeanneret is responsible for national healthcare statistics in Switzerland. He has been working for the Swiss Federal Statistical Office (SFSO) since 1996 and was responsible for the development of new statistics on hospitalisations. Dr Jeanneret received his PhD degree in Immunology and Epidemiology in 1990.

In Switzerland, the total cost for hospitals amounts to more than €8.5 billion a year. Questions regarding how this money is spent and the potential for cost reductions have led the Swiss government to order an exhaustive set of statistics concerning all hospitalised patients.

The Swiss Federal Office for Statistics (SFSO) is responsible for collecting medical data on all individuals that are hospitalised in Switzerland. Information on the diagnoses and on the corresponding treatments are given for all patients.

The first solution that was proposed by the SFSO was to hide (not to encrypt) the identifying data. For example, the name of the patient was replaced by its Soundex code. Although the Vereinigung Schweizerischer Krankenhäuser (VESKA) – Swiss Hospitals Association – had been compiling an internal nominative set of statistics for more than 20 years, the members of the Swiss Society for Medical Informatics (SSIM) reacted negatively to this first project. They argued that the new statistics would create a large database which would not preserve the confidentiality of the patients' medical records.

From a legal point of view, there are exceptions to what is deemed a medical secret when federal statistics are involved. The SFSO could also have forced healthcare providers to participate in the statistical survey. However, this would have led to an open conflict, so the SFSO contacted the Swiss Federal Section of Cryptology to seek a cryptographic expert who was capable of finding a solution to this problem.

From a statistical point of view, it is not necessary to know to whom a medical record belongs, but the SFSO needs to recognise that two different records actually belong to the same person. This is crucial in order to follow the history of the patients. At first, both conditions seem incompatible. The solution that has been developed solves this paradox. Basically, the data can be split into two categories:

- medical data – diagnosis and treatment, etc.; and
- non-medical data – last name, first name, date of birth and domicile, etc.

Some non-medical data is also useful for identification purposes. The epidemiological data forms the row data on which all statistical studies will be based. It contains the medical data, but also some non-medical data. As long as it does not allow the identification of the patient, it is not considered to be sensitive.

For this reason, in order to preserve the anonymity of the patient, the level of precision of non-medical data has been reduced to the minimum that is needed for the statistics. For example, the age is used instead of the date of birth, or the region instead of the domicile.

The non-medical data that is particularly useful for identification is not used directly in the statistics. Essentially, the solution involves replacing this identifying data with a calculated personal code – uniform linking code – which characterises the patient without revealing his or her identity. This satisfies the following properties.

- The identifying data allows the easy calculation of the personal code of a patient.
- The personal code of a patient does not allow his or her identification.
- The same person always receives the same personal code.
- Two different people always receive two different personal codes (no collision).

These properties are similar to those of a cryptographic, one-way hash function.

Initially, it must be decided on which identifying data the calculations will be based. This data should always be available and should stay constant over time. If there is too much restriction in the choice, there will be many collisions. If too much data is collected, it will not always be available and could change over time. Eventually, the decision was made to restrict the identifying data to the minimal set of identifying data – date of birth, sex, last name and first name.

In order for the personal code to remain the same for a given patient, it is crucial to minimise the consequences of spelling mistakes. Therefore, the identifying data is passed through a robust compression transformation. However, if the rate of compression is too high, there is the risk of the introduction of collisions – two different patients who are not distinguished from each other.

In 1997, the robust compression transformation was tested on the database of the University Hospitals of Geneva, which contains more than 222,000 records. The results were extremely encouraging – the rate of collision was only 0.3%. In addition, the transformation detected several doubles – two seemingly different patients who are actually the same person – in the hospital database. These results could help to ensure the accuracy of the database of the University Hospitals of Geneva.

In order to identify multiple hospitalisations, the cryptographic transformation that is applied by the hospitals on the minimal set of identifying data has to be the same in all hospitals and remain consistent over time.

It would not be reasonable to make the security of this transformation dependent on a secret key. A long-term secret key that is distributed to about 400 hospitals cannot be trusted. However, if this transformation is public, the resulting linking code is not completely resistant against a dictionary attack. As a consequence, the linking code has to be encrypted during the transmission from the hospitals to the SFSO and then in the SFSO database.

The session key that is used to encrypt the linking code during transmission is generated in the background by hospital computers. Entropy is given by measuring, for example, the time in milliseconds between two keystrokes/the acceleration of the mouse.

A public key cryptosystem – Rivest, Shamir and Adleman (RSA) – is then used to transmit the value of the session key to the SFSO. Some redundancy is introduced in order to control the origin of the session key and to test the specific implementation in the hospital of both the robust compression transformation and the encryption algorithm.

After reception, the encrypted linking codes are first decrypted and then re-encrypted directly and uniformly by the SFSO and thereby become the uniform anonymous linking codes that are used as personal codes.

The RSA private key of the SFSO and the master key that is used to encrypt the linking codes uniformly are both sensitive secret keys. Using a protocol that is based on a secret sharing scheme, those two keys form a secret that is shared between three independent individuals.

Table 1: Multiple Hospitalisations in Switzerland

1998			
No. Stays	**No. Patients**	**Multiple Hospital Rate**	**No. Cases**
1	333,274	85%	333,274
2	44,730		89,460
3	10,132		30,396
4	3,352		13,408
5	1,220		6,100
6	581		3,486
7	298	15%	2,086
8	149		1,192
9	90		810
10	47		470
11	32		352
12+	69		1,055
Total	**393,974**	**100%**	**482,089**

Source: SFSO, http://www.statistik.admin.ch/stat_ch/ber14/statsant/statsante1_2001.pdf. This statistic for 1998 takes into consideration only 38% of all cases of hospitalisations – those that have been received with a valid linking code.

An information technology (IT) system has been developed that is now in use in Switzerland on a large scale. During the first year, it was used for exactly 482,089 cases of hospitalisations – about 38% of all hospitalisations in Switzerland during 1998. The uniform linking code showed that these 482,089 cases concerned only 393,974 patients. It describes precisely the distribution of multiple hospitalisations without revealing the identity of the patients (see *Table 1*).

For 1999, more than 700,000 records have been transmitted with their encrypted linking codes. Data for 2000 will be sent to the SFSO at the end of summer 2001. More than 80% of hospitals have already validated their cryptographic module. The SFSO expects to receive more than one million hospitalisations with a linking code this year. By the end of the year, all hospitals should use the system.

Since this concept has been developed, several other institutions have been interested in adapting the system/protocols to their own needs in Switzerland, in other European countries, as well as in Canada. ■

Additional Information

This article has been written as part of a series for Information Security Solutions Europe 2001 (ISSE 2001) from 26–28 September 2001 at QEII Conference Centre, London (http://www.eema.org/isse).