Journal of Information, Law and Technology

Personal Identifiability in the Icelandic Health Sector Database

Professor Einar Árnason
Professor of Evolutionary Biology and Population Genetics
Institute of Biology, University of Iceland
einar@lif.hi.is

This is a **refereed** article published on: 3 September 2002

Citation: Árnason E, 'Personal Identifiability in the Icelandic Health Sector Database', 2002 (2) *The Journal of Information, Law and Technology (JILT)*.
<http://elj.warwick.ac.uk/jilt/02-2/arnason.html>

Abstract

Personal identifiability is a fundamental question in the ongoing debate about the Icelandic Bill and Act on the Health Sector Database (HSD). If the data are personally identifiable, Iceland’s international legal commitments indicate that a priori consent must be obtained from patients for the use of their personal medical information. The HSD Act presumes that one-way coding of personal identifiers renders the data non-personally identifiable and that therefore a priori consent is not required.

The history of the debate on the HSD shows that the concept of personal identifiability was initially based on a notion of ‘considerable amount of time and manpower’ as a criterion for defining personal identifiability. This definition comes from Recommendation R(97)5 of the Committee of Ministers of the Council of Europe on Medical Data. As a result of the Icelandic Data Protection Commission's opinion on the HSD, that concept was rejected and the resulting Bill and HSD Act adopted a definition from the European Data Protection Directive (95/46). The rejected concept, however, reentered with the idea that one-way coding of personal identifiers means there is no key that can be used to trace the identity of a person in the database.

The question of what constitutes a key in this context is of fundamental importance. The database will collect and link data from different sources on individuals over time and therefore the method of coding must remain stable. It is possible therefore to construct a look-up table, which constitutes a key. Keys can also be built from comparisons of patterns of family trees as well as by putting generally available information into context.

The information in the Health Sector Database is personal information. Therefore reason and justice require that a priori consent be obtained from patients for the transfer of their health data to the database as Iceland's international legal obligations stipulate. Anything less is unreasonable and unjust.

Keywords: Personal identification, Icelandic Health Sector Database, EU Data Protection Directive, health records data, one-way coding, keys, genealogy, information context, privacy

1. Introduction

The debate continues on the merits of the Icelandic Act on a Health Sector Database (HSD) and the plans for its construction. Lawsuits have already been filed that challenge both the constitutionality of the Act and whether Iceland's commitments under international law are violated[1]. The exclusive license to establish and operate the Icelandic Health Sector Database has been given to a private American company deCODE genetics. The Health Sector Database will contain the medical records of the whole population of Iceland but it also will be a structure through which a genealogical database and a DNA database can be linked to the medical records. The intention of the company is to exploit the information for commercial profit by selling access to the database which can be used as a research tool for research in epidemiology and in genetic research as well as in studies of how to maintain health systems.

A fundamental assumption of the Health Sector Database Act is that the data on individuals are not personally identifiable because the personal identifiers will be coded with one-way methods. The Act presumes that one-way coding effectively renders the data anonymous. If, however, one-way coding is found not to be qualitatively different from coding with a key, the data would be personally identifiable. In that case Iceland is bound by its international commitments to obtain a priori consent of the patients for the use of their data for a purpose other than that for which they were originally gathered. In most circumstances the physicians receive or obtain the information from the patients under an ethical and legal duty of confidentiality that can only be lifted with the consent of the patients or by a legal obligation (such as specific legislation or a court order).

In this paper I trace the history of the concepts of personal identifiability and keys during the debate on the Icelandic Health Sector Database. Originally the definitions used were derived from the Recommendation No R(97)5 of the Committee of Ministers to Member States of the Council of Europe on the Protection of Medical Data[2]. In response to criticism they were replaced with definitions from the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995[3] on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data. The question of the existence of keys is fundamental and I ask what constitutes a key and describe ways of making keys to open up the database with a look-up table, by comparisons with genealogies, and from the context of general information. I will argue that the sort of 'one-way encryption' called for by the Health Sector Database Act will not render the data 'anonymous'. I will also argue that the de facto existence of a coding key to link new information on individuals to their previous information in the Health Sector Database and to link information in the DNA and genealogical databases to information in the Health Sector Database, as well as the fact that the data themselves allow the identification of individuals means that the data are identifiable. They, therefore, come under the provisions of the Directive 95/46/EC that is now legally binding on Iceland. In contrast, the Recommendation is merely meant as an interpretive aid to the Directive and it has no legal force. Therefore, the consent of individuals in the Icelandic population should be sought before personal medical data are entered into the Health Sector Database.

2. Overview of Database Plan

Based on the Act on a Health Sector Database, no. 139/1998[4] the government of Iceland has given a license to a private for-profit corporation, deCODE genetics of Delaware USA, to create and operate a database of the medical records on the entire population of Iceland. The information in the medical records database will be cross-referenced with a genealogical database of the entire nation and with a genetic database covering a large number of individuals, both of which are in the possession of the licensee, to make one interactive database (Figure 1) - referred to as the GGPR database of Genotypes, Genealogy, health and disease Phenotypes, and Resource use in the Icelandic health care system. DeCODE is permitted to operate the database for commercial profit. The database will allow subscribers to perform in silico disease gene mapping, following pathogenesis of disease and complications and response to treatment, provide information for management of health and disease and health care resources. The prospective customers are pharmaceutical and biotechnology companies, HMO's, insurance companies and public health organisations and deCODE itself. The Act stipulates in addition that the Ministry of Health and the Director General of Public Health shall have free access to statistical data from the database for compiling health reports, planning, policy-making and other projects that they specify.

The database is not in operation yet. DeCODE has been in operation since 1996 and in the past Icelandic health authorities have been making health reports and health policy without access to such a database. The database, therefore, is not crucial for either deCODE business or public policy. The assumption, however, is that the database would become a profitable venture for deCODE and that access to it would facilitate public health policy (and see accounts by 5, 6).

On ownership the license states that all information transferred to the database is and shall remain the 'common property of the Icelandic nation' under the protection and rule of the Minister of Health and Social Security. The license, issued for 12 years at a time, authorises deCODE to create and operate the database for financial profit. At the termination of the license or if the license is revoked deCODE shall hand over the database and all software and software rights for its operation to the ministry of health. If, on the termination of the license, the ministry operates the database for profit it must pay deCODE a fee for software and intellectual property rights. However, if the ministry operates the database not for profit solely in the interest of the public health system deCODE will not receive payments for software or intellectual property rights. Thus, deCODE seemingly retains some commercial rights. In a contract made in connection with the issuing of the license deCODE agreed to indemnify the state of Iceland against any and all claims that could be made if the Act and regulations are found not to be in compliance with rules of the European Economic Area or other international rules and agreements that Iceland is or will become party to. deCODE also agrees to pay all fines and financial costs levied against the state of Iceland due to such non-compliance. The private corporate interests of deCODE genetics and the public interests of Iceland and Icelanders thus are mingled. However, it is difficult to discern where public interests end and private corporate interests start and vice versa.

DeCODE shall pay a fee for the issuing of the license and the costs incurred by the various public regulatory bodies monitoring the operation of the database. DeCODE also pays an annual license fee to Iceland of 70 million Icelandic kronur (IKR; approximately USD 820,000). If deCODE turns a profit in operating the database Iceland gets a share in the profits up to a maximum of 70 million IKR.

Large amounts of information from medical records on each individual will be transferred to the database (Table 1). They exist in two forms, as hand- or typewritten information that will be digitised and already computerised information and information in a planned countrywide fully standardised electronic medical records system. More detailed information will be available for transfer from the latter system with a long list of items to be transferred (Table 1).

The information from medical records will be transferred to the database and used under presumed consent. Persons will not be asked to give their prior affirmative consent to participation. Instead, people are given the opportunity to opt-out of the database by registering their intention with the Director General of Public Health. Those who opt-out before information actually starts flowing to the database will have all their medical records data excluded. Those who opt-out after the database is already in operation can only exclude medical records data that is generated subsequent to their opt-out.

The procedure for transfer of information to the database is as follows (4, 7). A health institution will one-way code a person's identity number or social security number[i] into an encrypted personal number (PN). Workers at a health institute gather the medical records information on an individual that is permitted for transfer (Table 1) into a package. To protect the data during transfer they encrypt the package using a public/private key issued by deCODE, the licensee. This package is then attached to the encrypted personal number instead of to the identity number. This is then sent to the Identity Encryption Service, a department of the Data Protection Commission. The Director General of Public Health maintains an opt-out database of those who have opted out of the Health Sector Database, with either all their health records data or specified parts of their health records. The Director General using the same function will one-way encrypt the identity number of individuals registered in the opt-out database and transmit that to the IES. The IES uses the encrypted opt-out list to filter out data on those who have opted out (Figure 1). The IES may re-encrypt the already one-way encrypted identity number (SSN) and transmit this along with the respective data package to the database where it forms the first sub-database of medical records with encrypted personal identifiers. The encrypted SSN becomes the final personal number (PN) that is associated with the health data of an individual in the database. It is thus clear that there will be many holders of the one-way encryption function, including deCODE.

The licensee, deCODE, has built a genealogical database of the entire Icelandic population and some of the ancestors of most of the families. This database has about six to seven hundred thousand individuals with names and identity numbers. DeCODE will encrypt the identity numbers using the same one-way function as above and transmit it via the IES to form a second sub-database of genealogies with encrypted identity numbers forming the same personal numbers (PN) as for the health data (Figure 1). DeCODE and various physicians collaborating on individual research project on the genetics of various diseases have collectively amassed a large amount of information into a genotypic database. This database also may contain some information from medical records that pertain to the diseases involved as well as molecular genetic information. The collaborating physicians know the identity (names and kennitala) of the participants. The data from this database will be transferred via the same mechanisms to form a third sub-database of genotypic data (Figure 1) associated with the respective PN.

The three sub-databases of medical records, genealogies and genotypes are cross-link-able by the personal numbers (PN, which are the one-way encrypted and re-encrypted identity numbers or kennitala of Icelanders) and together form the GGPR database (Figure 1). They contain micro-data that are database records on individual subjects associated with the personal number. End users, deCODE employees, query the database via a query layer to produce intermediate results. Customers query the intermediate results and final results are delivered to them as macro-data. Macro-data refer to statistical results calculated from micro-data, such as the mean age of a group of individuals.

3 History of the Concept of Personal Identifiability During the HSD Debate

3.1 Definitions of the HSD Act

The Act on a Health Sector Database, no. 139/1998 [4] has these definitions:

> 2. Personal data: all data on a personally identified or personally identifiable individual. An individual shall be counted as personally identifiable if he or she can be identified, directly or indirectly, especially by reference to an identity number, or one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.
>
> 3. Non-personally identifiable data: data on a person who is not personally identifiable as defined in clause 2.
>
> 4. Coding: the transformation of words or numbers into an incomprehensible series of symbols.
>
> 5. One-way coding: the transformation of words or series of digits into an incomprehensible series of symbols which cannot be traced by means of a decoding key.

According to these definitions personal identifiability is very broad in scope and, conversely, non-identifiability is very limited, being defined as the complement of identifiability. One-way coding is defined as a method that is supposed to eliminate the possibility of identifying a person using a key. The definitions make clear that mere coding is not enough. For even though it produces an incomprehensible series of symbols it still could be ‘comprehended’ by the use of a key. The essential issue here is that the coding is one-way. Unidirectional one-way coding is supposed to be some kind of technical method that eliminates the key.

These definitions are now law in Iceland. Starting in 1997 from A draft Bill written by the current license holder five major steps can be identified in the debate on the Health Sector Database (Table 2). The changes made to the definitions of the various terms are contrasted in Table 2. I now discuss these steps and changes in definitions and concepts during the debate on the Health Sector Database.

3.2 First Draft of a Bill in July 1997

Dr. Kááí Stefánsson, CEO of deCODE genetics which is the license holder for creating and operating the Health Sector Database, had the Lawyers at Skólavörðustígur 12 draft a First Draft of Bill on Health Sector Databases dated July 14, 1997. He presented the draft Bill to the Ministry of Health as a fax on September 3, 1997[8]. The aim of the authors of the draft was that the Bill be passed through Parliament during the fall of 1997 and that the Act take effect on January 1, 1998. Article 2 of the draft had these definitions:

> '3. Personal information: Information on private matters, health matters, finances or other matters of a named or nameable individual, which it is reasonable and natural to treat as confidential. An individual shall not be counted as nameable if a considerable amount of time and manpower would be required in order to name him/her. When an individual is not nameable the information about him/her shall not be considered to be personal information' (EÁ translated from the Icelandic).

3.3 Bill and Draft of a Bill in the Spring and Summer of 1998

When the Bill on Health Sector Databases (notice the plural) was presented to the 122nd session of Parliament in the Spring of 1998[9] it contained the definition of the draft Bill. However, one addition was made. It was stated that even if there exists a key to the data, an individual shall not be considered personally identifiable if the entity in possession of the data does not have access to the key:

> '2. Personal data: data regarding personal matters, including health information, finance or other items regarding a personally identified or identifiable individual, which it is reasonable and natural to treat as confidential. A person shall not be counted as personally identifiable if a considerable amount of time and manpower would be required in order to identify him/her. The same applies if the identification could only take place through use of a decoding key, not available to the person having the information. When an individual is not personally identifiable information about him/her shall not be considered personal information under the meaning of this law' (EÁ translated from the Icelandic).

The argument that an individual shall not be considered personally identifiable `if a considerable amount of time and manpower would be required in order to identify him/her' comes from the Recommendation No R(97)5 of the Committee of Ministers to Member States of the Council of Europe on the Protection of Medical Data [2]. In explanatory notes the authors of the Bill further state that provisions in the Bill regarding the use of a key are based on `procedural rules that the Data Protection Commission, which operates by the Act on Processing and Handling of Personal Information No. 121/1989, has recently made for scientific research in the health sector. The rules specify that data shall be coded with a key before they are handed over to the researcher and that the key will then be kept by special guardians appointed by the Data Protection Commission.' These statements imply that the definitions in the Bill are in accordance with the Act on Processing and Handling of Personal Information No. 121/1989[10]. They also imply that the definitions conform to the procedures on the use of a key already established by the Data Protection Commission.

The Bill on Health Sector Databases met extensive opposition in Parliament and by the Icelandic public. It was withdrawn in the late Spring of 1998, rewritten by a working group in the Ministry of Health and a new draft version[11] sent out for comments to various bodies including the National Bio-ethics Committee and the Data Protection Commission in July 1998. This was the first time that these regulatory bodies were formally asked to review the Bill. The draft contained the same definition that ‘a considerable amount of time and manpower would be required in order to identify’ a person from the Recommendation No R(97)5 of the Committee of Ministers of the Council of Europe [2].

3.4 Data Protection Commission's Opinions on Bill

In its letter to the Ministry of Health dated September 4, 1998[12] the Icelandic Data Protection Commission overturned the draft Bill’s definition of personal identifiability and the reliance on the Recommendation No R(97)5 of the Committee of Ministers of the Council of Europe.

The Data Protection Commission’s opinion[12] was that the HSD Act should be in accordance with EU Directive 95/46/EU on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, which was to be ratified by Iceland as part of its obligations as a member of European Free Trade Association (EFTA) and the agreement on the European Economic Area (EEA) between EFTA and the EU. This means, in the Data Protection Commission’s opinion, that the EU Directive will have to be adopted as law in Iceland and that both general and specialised legislation, such as the HSD Act, must be consistent with the Directive.

Furthermore,

> ‘the Data Protection Commission maintains that in the definition of the concept of personal data in the database Bill, the definition of the above mentioned EU Directive 95/46/EC appears to be totally disregarded; this states in clause (a) Art. 2 that data on individuals are personal data, if a decoding key exists for the coded data. The directive makes no distinction as to whether the identification would require considerable time and manpower’ [12].

In fact the concept of considerable time and manpower is not found at all in the EU Directive 95/46/EC [3] but instead is derived from the Recommendation No R(97)5 of the Committee of Ministers Council of Europe, as already mentioned.

The Data Protection Commission also questioned the assertion that the Bill was in accordance with Iceland's Data Protection Act[ii]. The Commission also reccomended that the ‘definition of personal data in the Bill not be ambiguous’[12] and in particular that it follow the EU Directive that was to become binding on Iceland[iii].

This crystal clear statements by the Data Protection Commission overturned the definitions in the Bill as to what would constitute personal data. The foremost experts of the State of Iceland on personal identifiability and data protection had spoken loud and clear. The response of the Bill's authors in the working group of the Ministry of Health was to eliminate all terms based on the Recommendation No. R(97)5 and the terms about keys that would be in the possession of someone other than the researchers. Instead they adopted a direct translation into the Icelandic of the definition from the Directive (95/46/EC) that the Data Protection Commission said would `become binding under international law on Iceland's behalf.' The Directive states (Art. 2):

> 'For the purposes of this Directive
>
> (a) personal data shall mean any information relating to an identified or identifiable natural person (data subject); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;' [3]

When the Bill on a Health Sector Database (notice the singular) was submitted to the 123rd session of Parliament in October 1998 [13] the definition had been changed and now was based on the definition from the Directive.

> '2. Personal data: all data on a personally identified or personally identifiable individual. An individual shall be counted as personally identifiable if he can be identified, directly or indirectly, especially by reference to an identity number, or one or more factors specific to his physical, physiological, mental, economic, cultural or social identity'[13].

However, the translation of the definition of the Directive to the Icelandic as part of the definition of the Bill was imprecise. The Directive's identification number was translated into kennitala, which is the term used for the national identity number of everyone in Iceland. In the translation of the Icelandic Bill back to the English in the official version of the Bill[13] it became an identity number. Under the Directive the term identification number is a broad concept encompassing any kind of an identification or personal number. The Directive is not limited to a specific identity number such as the kennitala of Iceland. Thus, when the Bill speaks of an identity number it is narrower than the identification number of the Directive.

The Data Protection Commission reiterated its position, this time with its comments to the permanent Health and Insurance Committee of the Parliament dated October 26, 1998 [14]. The Commission tried to explain the difference between disconnecting personal identifiers from the health data (de-identified data) and the method of coding the personal identifiers with some encryption function. Coding produces a new personal number (PN) but the health data are still link-able to a particular person and thus they remain personal data. De-identified data (disconnected from personal identifiers) are regarded as anonymous unless the data were of such a nature or quantity that the individual can be identified without access to a personal identifier by reference to certain factors specific to the data subject's physical, physiological, mental, economic, cultural or social identity. If that was possible the data would not be regarded anonymous[iv]. In conclusion the Data Protection Commission said that 'the Bill's assertion that the database will contain non-personally identifiable health data, does not hold'. The Commission recommended that the definition be dropped from the Bill.

The Health and Insurance Committee did not heed the recommendation to drop the word `non-personally identifiable.' The definition thus based on the Directive became law with the passage of the Act in December 1998. As it is not workable to have a key for a database that is supposed to be anonymous, no matter who holds it, as the Data Protection Commission pointed out, one-way coding was adopted. Following that it was claimed that a key does not exist because it is not possible to trace directly back one-way coding of names or identity numbers. One-way coding of personal identifiers is thus the essential feature that is meant to ensure that the Bill and Act abide by the Directive [15, 16].

3.5 Admissions That Keys Exist

Both Dr. Kári Stefánsson and deCODE's department of database have recently admitted that keys exist. In an interview in the New Scientist July 15, 2000 Stefánsson states regarding the interconnection of health information and genetic information:

> Stefánsson: 'Once we have identified a family with one of these diseases, what we will do is to go to those people and ask them to give us blood so that we can isolate DNA. ...When we do this, we will ask for their permission to cross-reference their names with the help of the health-care database. But in order to do this, we will have to get their explicit, signed consent'.
>
> New Scientist: 'Does this mean that you can identify individuals from the database?'
>
> Stefánsson: 'No. The information in the database will be encrypted and the keys will be kept by the Data Protection Commission of Iceland'.

The fact that there are keys means that under the Directive and in the opinion of the Data Protection Commission, the data are personally identifiable and not anonymous. According to the Commission it does not make `any difference whether the person having the information has access to the decoding key or not.' This had been accepted by the Ministry of Health and the Parliament when changes were made to the Bill[13] both in response to the Data Protection Commission's opinion to the Ministry[12] and its opinion to the permanent Health and Social Security Committee of Parliament (see above).

In an article in the Icelandic newspaper Morgunblað February 27, 2001[17] the deCODE department of database stated that information will be rendered non-personally identifiable using special encrypting key that fulfils very strict technical security measures [v]. DeCODE's database department admitted that keys exist and that it is possible to personally identify individuals by applying the keys. To say that the keys fulfil 'strict technical security measures' presumably means that 'considerable time and manpower' would be required in order to break the keys. Be that as it may, it is irrelevant in this context. The Data Protection Commission already pointed out that that arrangement is not mentioned by the Directive and in response that language had been removed from the Bill presented to Parliament in the Fall of 1998[13].

3.6 Genealogy and Genetics Databases

In the third and final round of Parliamentary discussion on the HSD Bill a change was introduced (Art. 10) permitting the interconnection of medical records in the HSD database with a database of genelogical information and with a database of genetic information. Similarly, during the debate the definitions of what constitutes genetic data were also changed (Table 2). The Bill introduced to Parliament in the Spring of 1998[9] defined genetic information as information on individuals as well as information on groups of related individuals and information both on health and disease. This definition was removed in the Draft Bill circulated in the Summer of 1998[11] as well as in the Bill introduced to Parliament in the Fall of 1998[13]. During this time, which was the major period for debate on the Health Sector Database both in society at large and in Parliament, the definition of genetic data referred only to information about individuals (Table 2). In early December 1998, late in the Parliamentary debate, the definition from the Bill of Spring 1998[9] was reintroduced verbatim in a motion to change the Bill[13] and this definition became law.

These changes and the resulting definition mean that genetic information covers a wide field including information on inheritance of traits in groups of related individuals. The definition also means that it is easier to recognise individuals based on genetic information than if the more limited definition had been kept because the information in the database refers to inheritable traits of individuals as well as of groups of related individuals.

3.7 Opt-out Database

Another change made to the Bill in the fall of 1998 was an introduction of Art. 8 on the Rights of Patients. This specified that a patient could at any time request that his/her information not be entered onto the Health Sector Database by filling out a form and filing it with the Director General of Public Health. The Director General would enter those individuals on a coded registry or onto the opt-out database.

The opt-out database must be kept up to date and is required for the day to day transfer of data to the Health Sector Database. The opt-out database will provide the means for filtering out the medical information on those who have opted out from the stream of data being transferred to the Health Sector Database[7]. These individuals, however, will not be filtered out from the genealogical database that exist at the licensee and will be transferred via the same transfer layer as medical information to the Health Sector Database [7]. The opt-out database makes it more likely that individuals can be identified under the Health Sector Database scheme.

4. Building a Key with a Look-up Table

The claim that one-way coding means that it is impossible to trace back with a key only holds in a narrow technical sense. If a personal identification such as the name John Doe (or his identity number 010476-4878) is sent through coding, using for example a one-way hash function [18, 19], the outcome would be '6cad0ac09e9c602a6477db4247bdeed1', a new invented and unique personal number (PN). Similarly if the name Jane Doe (or 020587-5988) was one-way coded using the same method the outcome would be the new invented personal number '73c01bf88feb18695bd65e611ef1cf26'. If we only had access to the invented numbers '6cad0ac09e9c602a6477db4247bdeed1' or '73c01bf88feb18695bd65e611ef1cf26', it would be very difficult to find out that one of them represented the name John Doe and the other Jane Doe. If this was all, the individuals could be considered to be non-personally identifiable, because it would not be reasonably possible to go from the one-way encrypted personal number directly back to the name. The individuals, however, would only be non-personally identifiable in the narrow sense of going directly back from the code to the name.

During the operation of the HSD database, however, there will be a key in operation. The HSD database is a long-time and longitudinal data gathering and interconnection of previous, current and future data on each individual [20, 21]. The database will be updated regularly and when new data are generated (for example during a person's visit to a physician) they must find their way to the right place in the database and be connected to other data on that particular individual. (The same applies to updating of the genealogy and genetic databases. For the genealogy database for example, children are born and linkages among families are formed and broken with marriage and divorce). Therefore, there exists `knowledge' of who the individual is and where he/she can be found in the Health Sector Database or for that matter in any of the three databases that will be interconnected (Figure 1). That knowledge resides in the method used for coding. In order to update the database the method must remain the same, stable in time. The method, therefore, is a key because with access to the method a look-up table connecting the names or identity numbers with the encrypted personal numbers or vice versa can be made effortlessly.

4.1 Coding, a Transformation of Names

Coding is no more than a transformation of a name or identity number to another form. With one-way coding an individual gets a new and invented personal identity instead of his/her identity number - a so-called personal number or PN number [7]. Several documents on the database refer to a hash function as a method for such a one-way transformation [e.g. 7, 19]. A hash function is a transformation of an input m to an output string of a fixed length the hash value h, or H(m) ® h.

Cryptology basically requires a hash function i) to accept an input of any length, ii) to give an output of fixed length, iii) that it be easy to calculate H(x) for a given input x, iv) that H(x) be one-way , and v) that H(x) be collision free [18].

A hash function is one-way if the function is hard to invert, which means that given some hash value h it is very difficult to find some input x that will yield that hash value, H(x) ® h. If given some input x, and if it is computationally very difficult to find some other input y that is not the same as x such that $H(x) = H(y)$ (i.e. two different inputs that yield the same hash value) then the hash function is said to be collision free[18]. Sometimes hash functions may allow collisions that have to be dealt with in a special manner.

When the Act on a Health Sector Database refers to one-way coding as the transformation of words or series of digits into an incomprehensible series of symbols which cannot be traced by means of a decoding key, it seems to be based on a protocol such as this hash function. A repeated one-way coding would take the output of the first hash function as an input for the second and so on. One can take MD5 (Message Digest 5) as an example of a hash function for such one-way coding. MD5 will take a message of any length and `digest' it to produce a 128 bit `fingerprint.' Functions such as MD5 are generally used for electronic signatures of documents. I shall use it here to make an example look-up table, a key made with one-way coding.

4.2 Look-up Table

Even though one cannot directly break the key (e.g. through factoring; [18]) the function used for the database must remain stable in order to update the database. Therefore, anyone with access to the function (or functions) can easily make a table that contains side by side the input and the output of the function [20].

A look-up table of names or identifying numbers and coded (encrypted) names or personal numbers is a table (Table 3) that contains side by side the names and the coded names. One can look up in the table to find the encrypted name corresponding to a real name or to find the real name corresponding to an encrypted name. Such a look-up table is a key [22, 20, 23]. This was known during the debate on the Health Sector Database because the method is described in Appendix VI to the Bill[19]: feed the Icelandic National Registry of names or identity numbers through the function and make a dictionary or a table of the input and output. A table can also be made for a more limited group. If a decision was made to go back and open the database, for example if the Parliament passed a law to that effect or if a court of law ordered the opening up of the database, it would only take a moment of computer time to make a look-up table and open the database with a key. One would only have to bring together the holders of the function or functions and feed the National Registry through. Similarly, anyone who knows that a particular name or identity number is being transferred from a health institution to the database and can observe its encrypted personal number appear at the database can make a similar inference [21].

4.3 Personal Identifiability During Preparation for Transfer

In order to transfer data to the Health Sector Database health records must be opened, read and digitised. At this stage the data are fully personally identifiable. This is true for all current data that are destined to be included in the database. It is also true for the data of the more than 20,000 people who have already rejected participation by sending an opt-out form to the Director General of Public Health because the people who prepare the data for transfer are not supposed to know who has opted out. Data on everyone will be read, digitised and sent towards the database. This is also true for all deceased people. Their records will be opened, read and digitised. This examination of all health records is done for a purpose other than that for which they were gathered. Also contrary to the wishes of those who have opted out, their data will be examined for a purpose other than that which they were gathered, prepared for transfer and sent off in the direction of the database. If the Identity Encryption Service makes mistakes these data may end up in the database even if there is a specific ban against their use. The Data Protection Commission operates the Identity Encryption Service and oversees its work, thus in effect overseeing itself. Thus issues of privacy are raised for the preparation and transfer as well as for data already stored on the database.

5. Building a Key from the Shapes of Genealogies

The Act on a Health Sector Database permits the interconnection of the Health Sector Database of medical records with a genealogical database. According to the Security Target for an Icelandic Health Database made for the Data Protection Commission by Admiral Management Services Limited [7, 24] the genealogical database of the licensee (deCODE genetics) will be one-way coded in the same way as the Health Sector Database. The same also applies to a database of genetic information that the licensee has made through collaborative research on various diseases. The three databases must use the same encrypted personal numbers (or be related in a unique manner) in order for the interconnection of these three databases to be possible.

The genealogical database, however, also exists at the licensee using names and/or identity numbers as the personal identifiers. The licensee has announced a gift to the Icelandic nation in the form of open web access to its un-encrypted genealogical database [25]. The genealogical data also exists elsewhere in the society. Since the same database and same genealogical information exists using both encrypted and un-encrypted names anyone with access to both databases can build a key by comparing and matching the shapes of the patterns of family relationships in the two database versions[vi].

Theoretically there exists an enormous number of possible family trees connecting individuals in some group (the number is a power function of the number of individuals). The real family tree of a particular group of individuals, therefore, is likely to be unique and different in shape from the family tree of another group of the same number of individuals. The number of children and their gender and the connections of one family to another through marriage and childbirth form a pattern that can in most cases be used to recognise families. There are about 2,500 families with six children, and somewhat less than 20,000 families with two children in Iceland; the number of other common family patterns lie in between these numbers. It is easier to recognise a particular six-children family than a particular two-children one. The former are fewer and their potential theoretical patterns are more numerous. However, the interconnection of families makes them unique and thus recognisable. The families of John and Jane Doe have a unique pattern, as all other families in the country. They are recognisable by the unique shape of the family tree whether the individuals are referred to by name or by an encrypted personal number.

Figure 2 shows an example of two families and their interconnections. The first pattern is from a genealogical database that identifies individuals by name or by identity number. The second figure is from the genealogical database that identifies individuals by their encrypted personal number. The method used for encryption is a very safe triple encryption that is supposed to be very difficult to break. Nevertheless personal identification is possible, and relatively easy, because the family patterns are unique. The observed patterns (Figure 2) are the only family patterns in the two databases that match. One can thus make a key by reading directly from the figure.

6. Building a Key from the Context

In the familiar radio or TV game 'Name that person', someone appears in disguise and changes his/her voice in replying. The participants can ask: 'Are you a man (or a woman?'); 'Do you play the piano or do you play football?', and so on. The person in disguise replies truthfully yes or no. Finally the participants figure out from the context of the questions and answers who that person is and name him or her. This is an example of building a key from the context.

Even if one did not build a key using a look-up table or from comparing genealogies it would nevertheless be possible to build a key by putting general information in similar context as is done in the game[26]. If personal identifiers, such as name, identity number or cell-phone number, have been irreversibly stripped and replaced with a one-time-only (disposable) encrypted personal number one can speak of dis-connected[15, 16, 27] or de-identified data[26]. General information, demographic or health data can be attached to such an encrypted personal number. As the number of information bits thus attached to the encrypted personal number are increased the circle is narrowed until the combination of information bits becomes unique. Such a combination can be used to point to the individual as if it was a fingerprint. This amounts to making a key even if the personal identifiers have been stripped from the data.

This is called re-identifying[26] the individual based on information that is generally available. This is much easier in a small nation such as Iceland than among a more populous nation. Technology also has changed everything in this respect. With internet access in the current age of information there is more and more general information on a person available to almost anyone[26]. Such general information can be used to form a combination that uniquely identifies an individual. With that, one can put into context other information that accompanies the general information and thus pinpoint to whom sensitive personal information belongs.

As an example take the identity numbers of individuals that have been coded either with a one-time-only encryption function or a one-way hash function as described above. Attached to this personal number is general information such as gender, birth date, year of birth, height, town of residence as well as health information of varying sensitivity.

Examples could be operation for appendicitis, cancer of the colon or cancer of the breast, or diabetes (Table 4). More sensitive information, such as on venereal or mental disease, also might be included.

The yearly average number of births in Iceland is about 4,200, or 11--12 births per day. Few days have more than 20 births. Having information about birth date and year thus narrows the circle down to about 20 people at most[22]. By including information on gender the number is halved: on average six girls and boys are born per day and very seldom are there more than ten boys and girls born per day. By adding height, township or eye colour one can without doubt recognise most if not all people. Therefore general information comparable to that required for a passport application is sufficient to recognise an individual[22] without a name or identity number. From that one can identify what individual has which disease if given information such as that in Table 4. Individuals who have even a more ‘sensitive’ disease are recognisable in a similar manner. A male born February 2, 1979 is one of (on average) six males born that day. He is 176 cm high and lives in Dalvík. That must be Helgi. He has diabetes. One does not need a key, a family tree or personal identifier for that.

The various bits of information that will be transferred to the database (Table 1) are of a similar nature as in the above example. There are for example many dates and times of visits and other bits of information that are innocuous by themselves (Table 1). They can be combined in a similar manner to make a unique personal identifier without recourse to genealogy or personal identity number.

7. Discussion

The history of the concept of personal identifiability of the Bill and the Act on a Health Sector Database [4] was initially based on the premise that individuals would not be regarded personally identifiable if considerable time and manpower was required for identification. The criticisms levelled at the Bill, as well as the changes made to it in response to criticisms, show that the initial plan was based on false premises. The Data Protection Commission made this evident in its opinion and during the debate on the Bill. In spite of this, that claim is still being promoted by the main proponents of the database. This basic premise of the Bill was partly reintroduced and became law with the definition of Act that one-way coding of personal identifiers renders the individual's health data non-personally identifiable because there is no key. A question therefore arises of what a key is, what personally identifiable means, and, how keys can be built to open up the database.

7.1 Personal Data

During the debate on the Bill on a Health Sector Database it was claimed that the Bill might fulfil the requirements of international legal instruments [15] if the technical premise of the Bill was correct that one-way coding entails a final and complete unlinking of data and personal identifiers. However, the authors [15] also acknowledged the possibility that coding and one-way coding would not be considered qualitatively different and that therefore one-way coding would not be considered to be a final and complete unlinking of data and personal identifiers[16].

The premise of the Act does not hold up under close scrutiny. One-way coding does not mean that a key does not exist. One-way coding only means that it is difficult or computationally intensive to trace back directly from the encrypted personal number to the identity number or the name. By adopting this definition the Bill reintroduced the concept of 'considerable amount of time and manpower' that the Data Protection Commission had already rejected as it is not part of the Directive. Contrary to this, according to the Directive account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person[3]. The data are therefore personal data both under the Health Sector Database Act[4] and the Data Protection Act[28] that both are based on the Directive 95/46/EC[3]. Having ratified the Directive Iceland now is bound by it. If Iceland is going to fulfil its international commitments it is necessary to get consent of the individual before transfer of data to the Health Sector Database. Multiple one-way coding makes no difference.

Some of the officials of the State of Iceland who are supposed to enforce this Act claim that within the meaning of this law the data are non-personally identifiable. Some of their critics have called this the flat-earth-theory-of-law: if a legal text asserts that the earth is flat then it is flat in the meaning of that law even if it flies in the face of physical reality. Because the database Act asserts that one-way coding is the transformation of words or series of digits into an incomprehensible series of symbols which cannot be traced by means of a decoding key, these Icelandic officials have argued, it therefore means that a key does not exist in the meaning of that law.

Equality before the law is a basic rule of law [29]. The definition in the first versions of the Bill on Health Sector Database was based on a concept from the Recommendation R (97)5 of the Council of Europe Committee of Ministers stating that not being personally identifiable refers to methods that would require 'considerable time and manpower' to break. In my view this cannot be a foundation for a legislation. If this concept is used as a basis for law it means that those who have access to considerable time and manpower are above the law which is contrary to equality. In that case the foundation of law would also be dependent on the status of technology, which also is doubtful. For as Art. 27 of the Preamble of the Directive[3] states the scope of protection 'must not in effect depend on the techniques used, otherwise this would create a serious risk of circumvention'.

There is a basic difference between definitions in the Directive (95/46)[3] and the Recommendation R(97)5[2]. The Directive, which is now legally binding on Iceland and has been entered into Icelandic law with the Act on Personal Protection (no. 77/2000) [28], gave rise to the definition of personal data in the Act on the Health Sector Database. The difference between the Directive and the Recommendation is that the Directive defines in very broad terms what personal data are but does not discuss or define what non-personally identifiable data are. The Recommendation, however, defines non-personally identifiable data. The effect of this is that the Directive puts the burden of proof on anyone who claims that he or she is working with non-personally identifiable data. In addition the Recommendation is merely meant as an interpretive aid to the Directive and it has no legal force.

7.2 Building Keys

One-way coding of personal identifiers does not equal de-identifying data because the database is a longitudinal collection and linkage of data on an individual. Because the database is longitudinal the method of coding must remain stable in time or else the database could not be updated. Coding of the same identity number will therefore always produce the same personal number. Anyone who can send an identity number through the coding process and observe the outcome can thus make a look-up table, which is a key [21].

Even if it was not possible to make a look-up table, identifying a person is nevertheless possible by inference. The data will also be interconnected with genealogical data that also will be longitudinal data as are the health data. The shapes of family trees will change with birth of new children thus making it easier to recognise families with each updating of the genealogical database. Family trees soon become unique when the number of individuals in a group is increased. Comparisons of the patterns of family trees from a genealogical database containing one-way encrypted personal numbers as identifiers with the same genealogical database containing names or identity numbers as identifiers is a method for making a key.

As already discussed the Health Sector Database Act [4] as well as the Explanatory Notes to the Bill claim that one-way coding renders information on a personally identifiable individual non-identifiable because there is no key. The definitions also claim that it is not possible to identify an individual with reference to any factors specific to his/her physical, physiological, mental, economic, cultural or social identity. This is questionable. Enough general information (passport information) is publicly available to identify most individuals from the context [22]. Keys can be made this way.

In this paper I have discussed examples of methods that would be reasonably used to build keys to open up the Icelandic Health Sector Database. Personal identifiability is not a distant, theoretical possibility[22]. It is a real possibility and a problem easy to solve. It is possible to build keys to open the database and multiple one-way coding does not alter that in any way. If the Parliament changed the law and permitted the opening of the database or if a court of law ruled that it should be opened there is no technical hindrance to do so in an instant. The main premise of the Icelandic Act on a Health Sector Database therefore does not hold up to scrutiny. Also various entities involved in the preparation of data for and in the operation of the Health Sector Database could use methods of this nature to identify individuals in the database. These entities are identified as threats to the security of the database [7, 23, 22].

7.3 Lessons and Ramifications

The Icelandic case has legal and ethical ramifications and various lessons can be learned from it [e.g. see 30, 5, 6, 31, 32, 33]

The EU Data Protection Directive applies to personal data but does not apply to data that are truly anonymous. The Icelandic HSD Act argues that the technical solution of one-way coding of personal identifiers renders the data anonymous. Some have argued that the HSD Act is legal under Icelandic constitution and Iceland's international commitments [16, 34] while others have reached the opposite conclusion[35, 5]. The views that one-way coding can achieve this have been expressed[36] although doubts have also been expressed about whether one-way coding really is different from coding in this respect [16]. I have argued here that coding, be it one-way or multiple, is largely irrelevant. In reality anonymity does not exist in databases, such as the Icelandic HSD, that have large amounts of information from which contextual inferences about personal identity can be drawn.

The opt-out clause is another issue that has been reviewed both favourably[37] and unfavourably[5]. I have previously argued that it represents a totalitarian aspect of the Act because in reality an Icelandic citizen has no choice[38]. The individual is given two alternatives: to either belong to the health sector database via presumed consent (and thus directly be part of the business plan of a private corporation) or to register with the government and enter the opt-out database. However, the opt-out database, kept up to data, is required for the normal transfer of data to the health sector database. Thus one way or another the database plan involves everyone, no one is left alone.

The exclusive license issued to deCODE has ramifications for scientific freedom as well as European competition rules[16, 5]. Also Iceland has adopted European Directive 69/9/EC on Protection of Databases[39] that applies sui generis rights to databases. This may have ramifications when the license expires and the state takes over the database. There are provisions in the act that are meant to ensure these rights for the state[16].

8. Conclusions

One can reasonably expect that methods such as the ones described in this paper would be used to identify persons in the Icelandic Health Sector Database. The individuals are personally identifiable both in the preparation of data for transfer, in the opt-out database and in the Health Sector Database. Therefore, it is both right and reasonable to require the a priori consent of the individuals for inclusion of their data on the database and their use for a purpose other than what they were gathered for, as Iceland's constitution and international commitments dictate[12, 3]. Anything less is unreasonable and unjust.

References

1. Mannvernd (2002), Status of Lawsuits against the Icelandic Health Sector Database Act and related matters
<http://www.mannvernd.is/english/lawsuit.html>.

2. Council of Europe, Committee of Ministers (1997), Recommendation No. R (97) 5 of the Committee of Ministers to Member States on the Protection of Medical Data <http://www.coe.fr/cm/ta/rec/1997/97r5.html>

3. European Parliament and the Council (1995), Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. <http://www.privacy.org/pi/intl_orgs/ec/final_EU_Data_Protection.html>.

4. Alþingi, Icelandic Parliament (1998), Act on a Health Sector Database no. 139/1998. Passed by Alþingi December 17, 1998 <http://brunnur.stjr.is/interpro/htr/htr.nsf/pages/gagnagr-ensk>.

5. Greely H T (2000), Iceland's Plan for Genomics Research: Facts and Implications, Jurimetrics, 40, 153-191. <http://www.mannvernd.is/english/articles/HTGreely_Jurimetrics_2000.html>.

6. Rose H (2001), The Commodification of Bioinformation: The Icelandic Health Sector Database, Tech. rep., The Wellcome Trust, London. <http://www.mannvernd.is/greinar/hilaryrose1_3975.PDF>.

7. Admiral Management Services Limited (2001), Security Target for an Icelandic Health Database, Admiral Management Services Limited. Made for Icelandic Data Protection Authority. <http://www.personuvernd.is/tolvunefnd.nsf/Files/SecurityTarget/$file/SecurityTarget.pdf>.

8. Ministry of Health and Social Security, Iceland (1997), First draft of Bill on Health Sector Databases. Presented to Icelandic Ministry of Health and Social Security by K. Stefánsson of deCODE genetics, September 3, 1997. <http://www.mannvernd.is/english/laws/HSDbill_english_firstdraft_140797.html>.

9. Alþingi, Icelandic Parliament (1998), Bill on Health Sector Databases. 122 session of alþingi, spring 1998. Submitted to Alþingi, Icelandic Parliament, at 122nd session, Spring of 1998. <http://www.mannvernd.is/english/laws/HSDbill_english_122session1998.html>

10. Alþingi, Icelandic Parliament (1989), Act on the Recording and Handling of Personal Information No. 121/1989. Superceded by Act on the Protection of Individuals with Regard to the Processing of Personal Data No. 77/2000.

11. Ministry of Health and Social Security, Iceland (1998), Draft - Bill on a Health Sector Database. Circulated for comments in the Summer of 1998. <http://www.mannvernd.is/english/laws/HSDbill_english_summer1998.html>.

12. Data Protection Commission (1998), Data Protection Commission's opinion on the

draft Bill on a health-sector database. Letter from Data Protection Commission to Minister of Health, Ingibjörg Pálmadóttir, September 4, 1998. <http://www.mannvernd.is/english/news/Data_Protection_Commission_040998.html>.

13. Alþingi, Icelandic Parliament (1998), Bill on a Health Sector Database. Submitted to Alþingi, Icelandic Parliament, at 123rd session, Fall of 1998. <http://www.mannvernd.is/english/laws/HSD.bill.html>.

14. Tölvunefnd (1998), Umsögn tölvunefndar um frumvarp til laga um gagnagrunn á heilbrigðissviði. Beint til heilbrigðis- og trygginganefndar Alþingis (in icelandic).

15. Björgvinsson D Þ, Arnardóttir O M and Matthíasson V M (1998), Álitsgerð um ýmis lögfræðileg efni í frumvarpi til laga um gagnagrunn á heilbrigðissviði. Institute of Law, University of Iceland, opinion on legal aspects of Bill on a Health Sector Database. Requested and paid for by deCODE genetics which presented the opinion to Members of Alþing, Icelandic Parliament, on October 28, 1998 (in Icelandic).

16. Arnardóttir O M, Björgvinsson D Þ and Matthíasson V M (1999), The Icelandic Health Sector Database, European Journal of Health Law, 6 (307-362).

17. Heilbrigðishópur gagnagrunnsdeildar Íslenskrar erfðagreiningar (2001), Ópersónugreinanleg gagnasöfnun til ábyrgra vísindarannsókna, Morgunblaðið, February 21. DeCODE's department of database article in the Icelandic newspaper Morgunblað (in Icelandic), <http://www.mbl.is>.

18. RSA Laboratories (2000), RSA Laboratories' Frequently Asked Questions About Today's Cryptography, Version 4.1, RSA Security Inc. <http://204.167.114.22/rsalabs/faq/index.html>.

19. Sigurðsson G, Björnsdóttir S H and Björnsson B r (1998), Fylgiskjal VI. með frumvarpi til laga um gagnagrunn á heilbrigðissviði, Þskj. 109. Stiki ehf.: Minnisblað um feril heilsufarsupplýsinga frá heilbrigðisstofnun í miðlægan gagnagrunn. Bill on a Health Sector Database, Appendix VI. Memorandum from Stiki ehf. on process of health data from a health institution to a centralised database. (in Icelandic). <http://www.mannvernd.is/english/laws/HSD.bill.html>.

20. Anderson R (1998), The DeCODE Proposal for an Icelandic Health Database. Evaluation of the privacy aspects of DeCODE's proposal for a central database of Icelanders' medical records at the invitation of the Icelandic Medical Association. <http://www.cl.cam.ac.uk/users/rja14/iceland/iceland.html>.

21. Anderson R (1999), Iceland's Medical Database is Insecure, British Medical Journal, 319, 59, <http://bmj.com/cgi/content/full/319/7201/59/b>.

22. Benediktsson O (2000), Persónugreinanleiki í gagnagrunni á heilbrigðissviði. Personal identifiability in the Icelandic Health Sector Database. Made at the request of

Ragnar Aðalsteinsson, September 13, 2001 (in Icelandic).
<http://www.mannvernd.is/greinar/OBgreinanleikiMV.html>.

23. Anderson R (1999), Comments on the Security Targets for the Icelandic Health Database. Comments requested by Icelandic Medical Association on two documents written by Admiral Management Services Ltd. for the Data Protection Authority.
<http://www.cl.cam.ac.uk/ftp/users/rja14/iceland-admiral.pdf>.

24. Admiral Management Services Limited (2000), Approval Process Methodology. Icelandic Health Database, Admiral Management Services Limited.
<http://www.personuvernd.is/tolvunefnd.nsf/Files/7163.method.../
$file/7163.method.pdf>.

25. DeCODE genetics (2000), deCODE genetics and Frisk Software donate access to the genealogy database to the Icelandic nation.
<http://www.decode.com/news/releases/older/item.ehtm?id=1382>.

26. Sweeney L (1998), Re-identification of de-identified medical data, National Committee on Vital and Health Statistics Subcommittee on Privacy and Confidentiality.
<http://ncvhs.hhs.gov/980128tr.htm>.

27. Tölvunefnd (1998), Umsögn tölvunefndar um drög að frumvarpi til laga um gagnagrunn á heilbrigðissviði. Beint til Ingibjargar Pálmadóttur, heilbrigðisráðherra. Data Protection Commission opinion on Bill on a Health Sector Database presented to the Health and Social Security Committee of Alþing, the Icelandic Parliament, on October 26, 1998 (in Icelandic). <http://www.mannvernd.is/login/ums_tolvunefnd_261098.html>.

28. Alþingi, Icelandic Parliament (2000), Act on the Protection of Individuals with Regard to the Processing of Personal Data No. 77/2000.
<http://www.personuvernd.is/tolvunefnd.nsf/pages/1E685B166D04084D0025692200474
4AE>.

29. Aðalsteinsson R (2000), '...einungis eftir lögunum', Úlfljótur, 2000 (4), 1-32. '...only according to the law' (in Icelandic).

30. Annas G (2000), Rules for Research on Human Genetic Variation - Lessons from Iceland, The New England Journal of Medicine, 342, 1830-1833.

31. Winickoff D E (2000), Rhetoric equals cold cash in Iceland: The Biobank Act and deCODE genetics, GeneWatch, 13, 5-6.
<http://www.gene-watch.org/magazine/vol13/13-5decode.html>.

32. Winickoff D E (2000), Biosamples, Genomics and Human Rights: Context and Content of Iceland's Biobanks Act, Journal of Biolaw and Business, 4, 11-17.

33. Sigurdsson S (2001), Bibliography/Self-help Kit for Studying the HSD deCODE

Controversy <http://www.raunvis.hi.is/~sksi/kit.html>.

34. Jónatansson H (2000), Iceland's Health Sector Database: A Significant Head Start in the Search for the Biological Holy Grail or an Irreversible Error, American Journal of Law and Medicine, 26, 31-67.

35. Roscam Abbing H D C (1999), Central Health Database in Iceland and Patients' Rights, European Journal of Health Law, 6, 363-371.

36. Nielsen K K and Waaben H (2001), Lov om Behandling af Personoplysninger (Copenhagen, Denmark: Jurist- og Økonomforbundets Forlag).

37. Laurie G (2002), Genetic Privacy (Cambridge: Cambridge University Press).

38. Árnason E (2000), The Icelandic Healthcare Database, The New England Journal of Medicine, 343, 1734 <http://content.nejm.org/cgi/content/short/343/23/1734>.

39. European Parliament and the Council (1996), Directive 96/6/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. <http://europa.eu.int/ISPO/infosoc/legreg/docs/969ec.html>.

1.0

Table 1

I. From already existing medical records
A. Information from National Registry: Identity number or SSN, one-way encrypted. Gender and age, residence (county and mail code) and marriage status at the time of the recording of the information.
B. Coded and quantitative information: Disease diagnosis according to ICD-9/ICD-10 system. Operation number. Date of arrival and discharge. X-ray, CT, MR analysis. Research results. Physiological measurements. Coded drug treatment.
II. From standardised electronic system
1. Health Institute: The institutes identity number. Department. Medical speciality.
2. Patient identification: Type of patient. Identity number or SSN, one-way encrypted. Gender. Marriage status. County of residence. Employment. Education.
3. Arrival at health institute: Date that a patient enters a waiting list. Date, time and method of arrival. Where from the patient comes. Reason for hospitalisation.
4. Discharge from health institute: Date and time of discharge. Date of termination of active treatment. Date of arrival to walk-in clinic. Repeated visits to walk-in clinic. Where the patients goes after treatment.
5. Reason for arrival.
6. Physician's examination at arrival: Date and time of examination.
7. Drugs given at arrival: Drug type, unit, number, concentration, quantity and frequency of administering.
8. Allergy: Date of recording. Drug allergy. Other allergy.
9. Specialist's treatment plan.
10. Informed decision on treatment.
11. Physicians instructions.
12. Instructions for drug treatment: Date of instructions. Type and number of drug. Type of treatment. Concentration, unit, amount, frequency, how often, method of administering

(subcutaneous, etc). Date of termination.

13. Administering of drug according to instructions: Date and time. Drug type and number. Method of administering. Effects. Side effects.

14. Specialist's evaluation of treatment.

15. Drugs at discharge. Date of instructions. Drug type and number. Type of drug use. Concentration. Unit. Amount. Frequency. Method of administering. Date of termination.

16. Diaries.

17. Consultations. Date and time of request. Reason for request. Date and time of reply. Result/analysis.

18. Notes of physician at walk-in clinic. Date and time of notes. Diagnosis. Procedure number. Procedure code of physician. Treatment.

19. Information gathering by nurses. Date and time. Examination and measurements at arrival (e.g. temperature, pulse, breathing). Gordon’s health keys. Nourishment, metabolism and skin. Excretion. ADL. Movement and activity. Cognitive status, sensation.

20. Nursing process. Dates. Goal. Plan. Progress and evaluation. Nurse’s diagnoses.

21. Disease diagnosis. Date of diagnosis. Disease diagnosis by coding table. Disease diagnosis, physician’s text.

22. Operations. Operation number. Physician’s operation title. Date of operation.

23. Reports of vital signs. Date and time. Blood pressure. Pulse. Breathing.

24. Notes of walk-in clinic nurse. Date and time. Reason for arrival. Analysis and treatment (coded).

25. Immunisation. Date. Vaccination ICD-10 code. Vaccine. Side effects.

26. Reporting by other health workers. Date and analysis made by work therapists, physiotherapist, social workers, speech therapists, psychologists, neuro psychologists, pastors and deacons.

27. Scientific research connected to medical records.

28. Requests for tests and results.

29. Lifestyle. Smoking.

30. Coded social information.

31. Genetic information. Disease diagnosis obtained by examination of genetic material (e.g. analysis of inherited disease) and diagnosis based on chromosomal analysis, e.g. on inborn disease or malignant disease.

Table 1: *Information from medical records that can be transferred to the Health Sector Database from a health institute. Based on Appendix B of the license.*

July 1997	April 1998	July 1998	October 1998	December 1998
First Draft of Bill [8] written by deCODE in July 1997 and presented to Ministry of Health on September 3, 1997.	Bill on Health Sector Databases [9] submitted to 122nd session Parliament in April 1998.	Draft --- Bill on a Health Sector Database [11]. New version of the Bill rewritten by an *ad hoc* committee in the Ministry of Health and circulated for comments in July 1998.	Bill on a Health Sector Database [13] submitted to 123rd session Parliament in October 1998.	Act on a Health Sector Database [4] voted into law on December 17, 1998.

Definition of Health Sector Database

1. Health sector database: A collection of independent work, data, or other material containing information on health, arranged in an organised and systematic fashion and that can be accessed electronically or in other ways. *1. Health sector database:* A collection of independent work, data, or other material containing information on health and other related information, arranged in an organised and systematic fashion and that can be accessed electronically or in other ways. Health records that are kept according to law, other records that individual health institutions or research institutions keep on the individuals that they provide health service to, and records that official government health and insurance bodies keep on the users of the health service and on the operation of the health service are not considered a health sector database in

the meaning of this law. *1. Health sector database:* A collection of data
containing information on health and other related information, recorded in a standardised
systematic fashion in a single centralised database, intended to be a source of information.
1. Health sector database: A collection of data containing information on
health and other related information, recorded in a standardised systematic fashion on a single
centralised database, intended for processing and as a source of information. *1. Health
sector database:* A collection of data containing information on health and other related
information, recorded in a standardised systematic fashion on a single centralised database,
intended for processing and as a source of information.

Definition of health information

2. Health information: information on the health of individuals, other information regarding health
and genetic information. *3. Health information:* information on the health of
individuals, other information regarding health and genetic information. *3. Health
information:* information on the health of individuals and groups, including genetic information.
6. Health data: information on the health of individuals, including genetic
information. *6. Health data:* information on the health of individuals, including genetic
information.

Definition of personal information

3. Personal information: Information on private matters, health matters, finances or other matters
of a named or nameable individual, which it is reasonable and natural to treat as confidential.
2. Personal data: data regarding personal matters, including health
information, finance or other items regarding a personally identified or identifiable individual, which
it is reasonable and natural to treat as confidential. *2. Personal data:* data regarding
personal matters, including health information, finance or other items regarding a personally
identified or identifiable individual, which it is reasonable and natural to treat as confidential.
2. Personal data: all data on a personally identified or personally identifiable
individual. An individual shall be counted as personally identifiable if he can be identified, directly
or indirectly, especially by reference to an identity number, or one or more factors specific to his
physical, physiological, mental, economic, cultural or social identity. *2. Personal
data*: all data on a personally identified or personally identifiable individual. An individual shall be
counted as personally identifiable if he or she can be identified, directly or indirectly, especially by
reference to an identity number, or one or more factors specific to his physical, physiological,
mental, economic, cultural or social identity.

Definition of non-personal information

3. An individual shall not be counted as nameable if a considerable amount of time and manpower
would be required in order to name him/her. When an individual is not nameable the information
about him/her shall not be considered to be personal information. *2.* A person
shall not be counted as personally identifiable if a considerable amount of time and manpower
would be required in order to identify him/her. The same applies if the identification could only take
place through use of a decoding key, not available to the person having the information. When an
individual is not personally identifiable information about him/her shall not be considered personal
information under the meaning of this law.' *2.* A person shall not be counted
as personally identifiable if a considerable amount of time and manpower would be required in
order to identify him/her. The same applies if the identification could only take place through use
of a decoding key, not available to the person having the information. *3. Non-personally identifiable data:* data on a person who is not personally identifiable as defined in
clause 2. *3. Non-personally identifiable data*: data on a person who is not personally
identifiable as defined in clause 2.

Definition of coding

4. Coding: the transformation of
words or numbers into an incomprehensible series of symbols. *4. Coding:* the transformation of
words or numbers into an incomprehensible series of symbols.

Definition of one-way coding

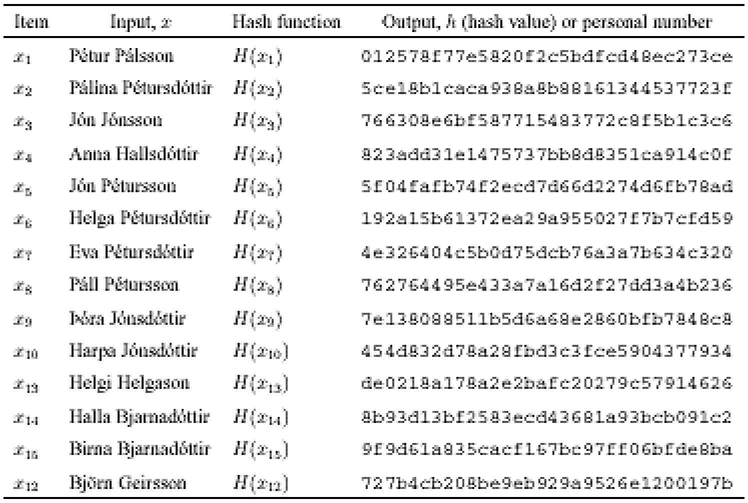
5. One-way coding: the
transformation of words or series of digits into an incomprehensible series of symbols which
cannot be traced by means of a decoding key. *5. One-way coding:* the

transformation of words or series of digits into an incomprehensible series of symbols which cannot be traced by means of a decoding key.

Definition of genetic information

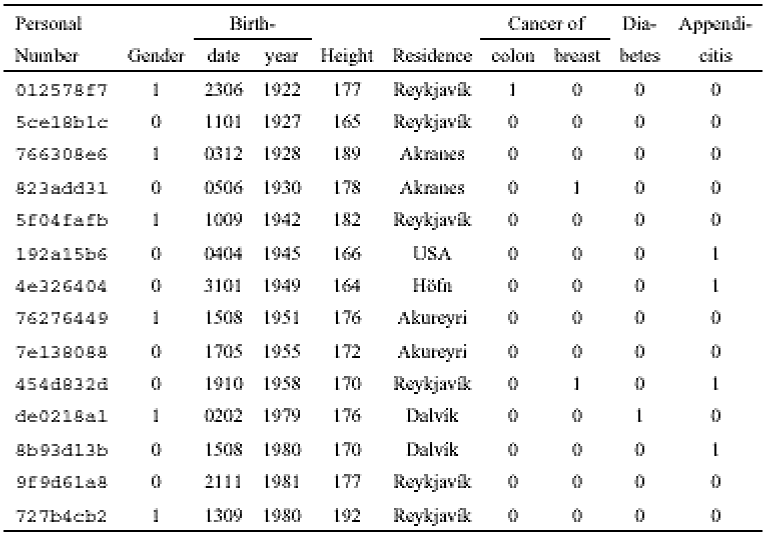
4. Genetic information: any kind of information regarding the inheritable characteristics of an individual or information that concerns the pattern of inheritance of such characteristics within a group of related individuals. *5. Genetic information:* any kind of information regarding the inheritable characteristics of an individual or information that concerns the pattern of inheritance of such characteristics within a group of related individuals, furthermore all information that concern the transfer of genetic information (genes) that relate to characteristics of disease or health of individuals or groups of related individuals irrespective of whether it is possible to diagnose these characteristics or not. *4. Genetic information:* any kind of information regarding the inheritable characteristics of an individual. *7. Genetic data:* any kind of information regarding the inheritable characteristics of an individual. *7. Genetic information:* any kind of information regarding the inheritable characteristics of an individual or information that concerns the pattern of inheritance of such characteristics within a group of related individuals, furthermore all information that concern the transfer of genetic information (genes) that relate to characteristics of disease or health of individuals or groups of related individuals irrespective of whether it is possible to diagnose these characteristics or not.

Table 2: *Chronology of HSD Bills and Act and definitions and changes of definitions of various terms during the debate*



Item	Input, x	Hash function	Output, h (hash value) or personal number
x_1	Pétur Pálsson	$H(x_1)$	012578f77e5820f2c5bdfcd48ec273ce
x_2	Pálina Pétursdóttir	$H(x_2)$	5ce18b1caca938a8b88161344537723f
x_3	Jón Jónsson	$H(x_3)$	766308e6bf587715483772c8f5b1c3c6
x_4	Anna Hallsdóttir	$H(x_4)$	823add31e1475737bb8d8351ca914c0f
x_5	Jón Pétursson	$H(x_5)$	5f04fafb74f2ecd7d66d2274d6fb78ad
x_6	Helga Pétursdóttir	$H(x_6)$	192a15b61372ea29a955027f7b7cfd59
x_7	Eva Pétursdóttir	$H(x_7)$	4e326404c5b0d75dcb76a3a7b634c320
x_8	Páll Pétursson	$H(x_8)$	762764495e433a7a16d2f27dd3a4b236
x_9	Þóra Jónsdóttir	$H(x_9)$	7e138088511b5d6a68e2860bfb7848c8
x_{10}	Harpa Jónsdóttir	$H(x_{10})$	454d832d78a28fbd3c3fce5904377934
x_{13}	Helgi Helgason	$H(x_{13})$	de0218a178a2e2bafc20279c57914626
x_{14}	Halla Bjarnadóttir	$H(x_{14})$	8b93d13bf2583ecd43681a93bcb091c2
x_{15}	Birna Bjarnadóttir	$H(x_{15})$	9f9d61a835cacf167bc97ff06bfde8ba
x_{12}	Björn Geirsson	$H(x_{12})$	727b4cb208be9eb929a9526e1200197b

Table 3: *A look-up table of names of individuals and their one-way coded personal numbers made with the MD5 one-way hash function*



Personal		Birth-				Cancer of		Dia-	Appendi-
Number	Gender	date	year	Height	Residence	colon	breast	betes	citis
012578f7	1	2306	1922	177	Reykjavik	1	0	0	0
5ce18b1c	0	1101	1927	165	Reykjavik	0	0	0	0
766308e6	1	0312	1928	189	Akranes	0	0	0	0
823add31	0	0506	1930	178	Akranes	0	1	0	0
5f04fafb	1	1009	1942	182	Reykjavik	0	0	0	0
192a15b6	0	0404	1945	166	USA	0	0	0	1
4e326404	0	3101	1949	164	Höfn	0	0	0	1
76276449	1	1508	1951	176	Akureyri	0	0	0	0
7e138088	0	1705	1955	172	Akureyri	0	0	0	0
454d832d	0	1910	1958	170	Reykjavik	0	1	0	1
de0218a1	1	0202	1979	176	Dalvik	0	0	1	0
8b93d13b	0	1508	1980	170	Dalvik	0	0	0	1
9f9d61a8	0	2111	1981	177	Reykjavik	0	0	0	0
727b4cb2	1	1309	1980	192	Reykjavik	0	0	0	0

Table 4: *Personal identification from the context of general information.*

Gender, birth date and year, height and township of residence are general data sufficient for identifying an encrypted individual without recourse to a key or family tree. Sensitive health information that accompanies the encrypted personal number in the table can thus be assigned to an individual. Only the first eight characters of the personal number are given to save space.

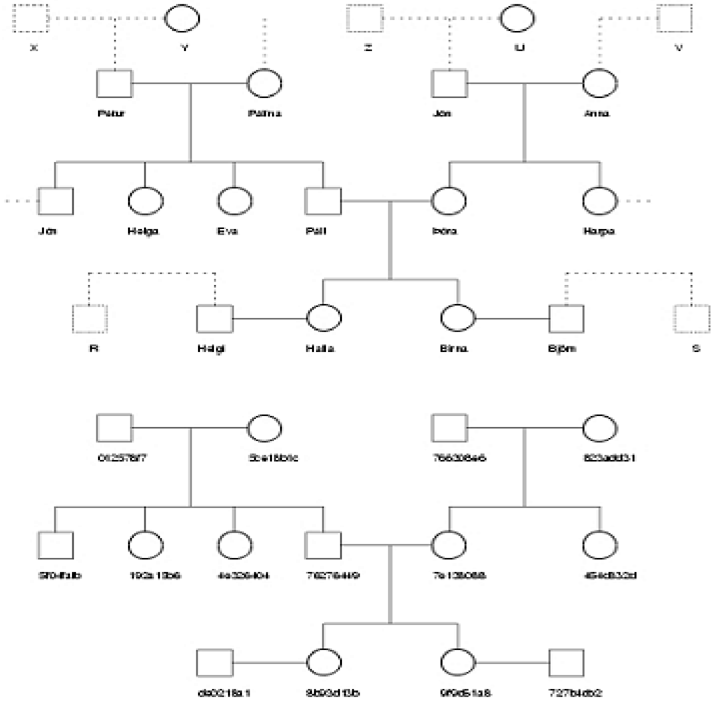


Figure 1: Organisation and flow of health data, genealogical and genetic information and opt-out list in the making of the GGPR database. Based partly on Figure 3.1 in Security Target. [7]

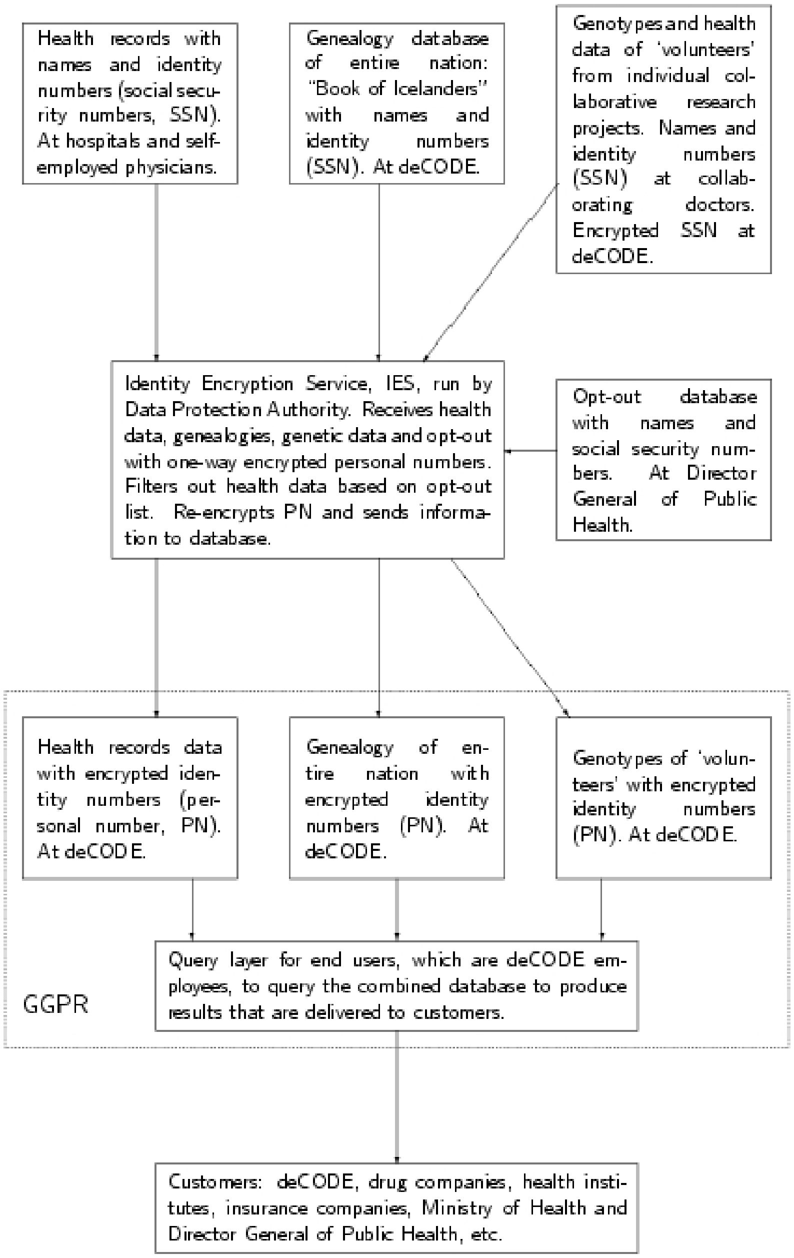

Health records with names and identity numbers (social security numbers, SSN). At hospitals and self-employed physicians.
Genealogy database of entire nation: "Book of Icelanders" with names and identity numbers (SSN). At deCODE.
Genotypes and health data of 'volunteers' from individual collaborative research projects. Names and identity numbers (SSN) at collaborating doctors. Encrypted SSN at deCODE.
Identity Encryption Service, IES, run by Data Protection Authority. Receives health data, genealogies, genetic data and opt-out with one-way encrypted personal numbers. Filters out health data based on opt-out list. Re-encrypts PN and sends information to database.
Opt-out database with names and social security numbers. At Director General of Public Health.
Health records data with encrypted identity numbers (personal number, PN). At deCODE.
Genealogy of entire nation with encrypted identity numbers (PN). At deCODE.
Genotypes of 'volunteers' with encrypted identity numbers (PN). At deCODE.
Query layer for end users, which are deCODE employees, to query the combined database to produce results that are delivered to customers.
GGPR
Customers: deCODE, drug companies, health institutes, insurance companies, Ministry of Health and Director General of Public Health, etc.

Figure 2: Comparison and pattern matching of a family tree from a genealogical database containing encrypted personal numbers and a database containing names. Only first names are given to conserve space; see Table 3 for family names. Broken lines are connections to close relatives and from there on to more distant relationships of the entire genealogy. Only the first eight characters of the personal number are given to save space.

i. The Icelandic identity number or Social Security Number, SSN, or kennitala as it is called in Icelandic, is a person's birth date with the addition of three random digits and a fourth digit indicating the century. Thus for each birth date there exist 1000 potential SSNs.

ii. The Data Protection Commission stated: 'it is questionable to maintain that the Bill's definition is based upon the definition of Act no. 121/1989. The terms of paras. 3 and 4 of Art. 1 of this Act entail that data on individuals are personal data within the meaning of the Act, even if the individuals in question are not identified by name, ID number or other form of identification, which can be linked to a person with or without a decoding key. By the terms of Act no. 121/1989, data are thus normally personal data, if a decoding key exists for coded data ...and the Commission does not believe that it makes any difference whether the person having the information has access to the decoding key or not.' [12].

iii. The Data Protection Commission further stated: 'wishes to emphasise that it is necessary for both the terms of general legislation on registration here in Iceland (now Act no. 121/1989) and the terms of special legislation (e.g. the prospective Act on a health-sector database) to fulfil the conditions of the EU Directive, after it has become binding under international law on Iceland's behalf. It is also important that legislation in Iceland should be consistent regarding such important factors as the definition of the concept of personal data.' [12]

iv. The Data Protection Commission stated: 'Under the Directive of the European Community the concept of personal information is broad and encompasses all information, opinions, or comments that can be connected directly or indirectly to a particular individual, i.e. it refers to all information that are personally identified or personally identifiable. It follows from clause (a) in Art. 2 of the Directive that information is considered personally identifiable if the information can be personally identified on the basis of any characteristic, directly or indirectly, by reference to an identity number or other characteristic, with or without an identifying key. Article 26 of the Directive's Preamble states that the main principles of protection must apply to any information concerning a personally identified or identifiable person and that in order to determine whether a person is identifiable (traceable), account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person. From this it follows that the main principles of protection shall not apply to data that have been completely disconnected from the individual such that it is impossible to trace the information to particular persons.

In principle there are two major ways for ensuring personal protection in such a database. One is to 'disconnect' the personal information from the identity of the person and the other is to 'code' the information as it is called. The Bill on a Health Sector Database states that information on individuals will be coded before transfer to the database. It assumes that the information in the database will be updated on a regular basis when new information is added. In order to do that it is necessary to be able to find older information on that same individual in the database and therefore the information in the database will only be coded and not disconnected. The difference between these two methods, coding and disconnection, is mainly the following. When personal data are coded, the individual is assigned a new, invented registration or personal code while a decoding key exists, by which individuals may be identified. On the other hand, when data are disconnected from personal identifiers, the individual is assigned an invented registration or personal code, as before, but this code has no decoding key. In this case the information is considered not to be personally identifiable unless the information can be personally identified by resorting to other means, such as by reference to certain factors specific to the data subject's physical, physiological, mental, economic, cultural or social identity as per clause (a) of Art. 2 of the Directive.

By reference to all of this, the Data Protection Commission considers that the Bill's assertion that the database will contain non-personally identifiable health data, does not hold. The Data Protection Commission therefore recommends that that word [non-personally identifiable] be dropped from Art. 1 of the Bill.' (EÁ translated from the Icelandic.).

v. The deCODE department of database stated: `Highly advanced technical solutions have been designed and will be used to three-times one-way code the individual's identity number. Each one-way coding will be done with a special encrypting key that fulfils very strict technical security measures. In order to decode the identity number and thus to personally identify the health information one would have to use the three keys in the right order. To ensure that that does not happen it is assumed that the three keys will be held by three separate bodies (the health institutions themselves, The Data Protection Commission, and by deCODE). This automatic triple coding will scramble the identity number in such a way that it will be possible to update the information on a particular individual when new data arrive at the HSD without the data ever becoming personally identifiable after they have been copied from the health records. In spite of such security measures, which we assert are unique in the history of Icelandic scientific research, even persons who are familiar with scientific research of this nature doubt that the data will be truly non-personally identifiable.' (EÁ translated from the Icelandic).

vi. One can argue that deCODE could possibly make a look-up table or dictionary of the encoded identifiers of all Icelanders directly using the genealogical database. As part of the process of building the database, as described in Figure 1, deCODE will submit the genealogical database with identity numbers to the IES and then receive it back with the encoded personal numbers (PN). If the order of the submitted records is the same as the records received back a look-up table can be built directly. If not, the matching of family patterns could be used.

This document was translated from LATEX by HEVEA.