Chapter 14
Anonymity and Privacy in Biobanking

Judit Sándor and Petra Bárd

14.1 The Concept of Anonymity

Anonymity[1] is often seen as the best way to protect individual privacy in the biomedical context. The usual reasoning provides the following logic: data protection norms apply in case of identifiable data, however if the identity cannot be revealed no harm is done to anyone. With this connection of contentions anonymity becomes a defensive rather than protective element of personal privacy. However, the rationale behind data protection requires much more than simple escaping of legal problems. Furthermore, anonymity may often do harm, when an individual may no longer benefit of the findings of the research.

As early as 2001 Ellis and Mannion stated that the key to the permissibility of the use of genetic samples for research without consent is the anonymisation of genetic samples.[2]

But in many cases, by ensuring anonymity, privacy is excluded, as privacy is no longer there once no one can identify the origin of a biological sample or data. This is so despite the fact that before the concerns for genetic information in medical law (such as in the field of transplantation, blood donation), anonymity served entirely different functions. For instance, numerous cases in biomedical law prove that anonymity can serve as a guarantee for altruism, and in addition it is also seen as a safeguard against scientific and other biases in research. Donating blood or tissue to an identified person and to receive blood, tissue from an identified person creates a special and difficult interpersonal relationship between the donor and the recipient. The preferred method is an ultimate form of altruism, i.e. the total absence of a personal relationship, as in organ donation between complete strangers, where the identities of the donor and recipient are hidden through anonymisation.

[1] Anonymity is derived from the Greek word ἀνωνυμία, *anonymia*, meaning "without a name" or "namelessness". In colloquial use, anonymous typically refers to a person, and often means that the personal identity, or personal identifiable information of the given person is not known.

[2] Ellis and Mannion, (2001, 1).

J. Sándor (✉)
Central European University, 1051 Budapest, Hungary
e-mail: sandorj@ceu.hu

C. Lenk et al. (eds.), *Biobanks and Tissue Research*, The International Library of Ethics, Law and Technology 8, DOI 10.1007/978-94-007-1673-5_14,

In another context anonymity protects the individual against disadvantageous social consequences, such as exclusion from insurance. Therefore, while epidemiological data are necessary to plan health insurance schemes, the use of identifiable data is limited, for instance, for the purposes of quality assurance within health care or is justified in the field of private and commercial insurance.

In case of genetic and biobank research, anonymity has only a limited use as it is necessary to accompany DNA analysis with health care data to provide a meaningful conclusion. Therefore in most cases coded personal and health care data should be stored, not only physical samples. Furthermore, anonymity does not serve the interest of the donors as participants increasingly demand feedback on the findings relevant to their health care in biobank research.[3]

Anonymity as such is not the main focus of European data protection norms as it is interpreted and questioned only to see the limitation of data protection provisions but it is not defined positively. Still in the medical, scientific, and ethics literature anonymity has a principal role. Graeme Laurie distinguishes between "absolute" and "proportional" anonymity. While absolute anonymity is "achieved when no means are available to link data to an identifiable individual", we can talk of proportional anonymity whenever "no reasonable means of identification of specific individuals is possible,"[4] meaning linked or coded information, when the access to the link or the code is appropriately defined and restricted. Absolute anonymity should not be overestimated, since it might deprive researchers and donors from important values, e.g. it prevents longitudinal research being conducted or the feedback of results to research participants, still, the growing tendency to data mining at the same time shows the importance of at least some form of anonymisation. Many and often contradictory functions are associated with anonymity that should be re-examined in the case of biobanks.

In the following we will analyse anonymity in connection to privacy, confidentiality, health care and genetic research before we come to the functions of both anonymity and privacy in the legal framework of biobanks. After having highlighted the heterogeneity of norms and having offered certain technical solutions for anonymisation in a biobank context, we will explore two viable models: double coded samples in Estonia and the three-tier Hungarian solution. In the last part we will summarize our conclusions.

[3]See the findings of the second international workshop within the Tiss.EU project organized by Judit Sándor and Petra Bárd from the Center for Ethics and Law in Biomedicine (CELAB) at the Central European University (CEU), Budapest, Hungary. About 30 persons, speakers included, participated at the workshop that took place at CEU on 6–8 April 2009. The workshop made a major contribution to one of the four Focal Themes of the Tiss.EU project by addressing questions of "Anonymisation and Pseudonymisation as Means of Privacy Protection" (Focal Theme C) in relatively unexplored jurisdictions of Central and Eastern Europe, such as the Czech Republic, Hungary, Slovakia, and Romania. Due to the interdisciplinary nature of the workshop's subject, invited speakers represented a wide range of disciplines, such as law, medicine, philosophy, and information technology.

[4]Laurie (2002, 295).

14.2 Anonymity and Privacy

In comparison to anonymity, *privacy* is a much more dynamic notion and it serves slightly different functions. The core element of privacy is to maintain control over personal information which would be impossible once the data have been anonymized. Genetic data is never collected alone; in some jurisdictions dozens of pages long questionnaires need to be filled out by the patients or donors who often have to disclose special or sensitive information. On the one hand, these data are a treasure for researchers, on the other hand, they pave the way towards genetic or other types of discrimination. Should we attempt to overcome the drawbacks of deleting the link between the individual and his or her data, alternative means of privacy protection have to be found.

Personality can be protected by law in many different ways. Protection of dignity, liberty, autonomy, self-determination, privacy, or the right not to be discriminated against – these all serve diverse elements of personhood. While some rights such as the right to have someone's paternity acknowledged, or authorship, do require the use of the name and identity, there are some instances where personality is better protected by anonymity. For instance, *freedom of anonymous speech* is an important value in democracy: a person should be able to disseminate an opinion freely without disclosing his or her identity and without fear of retribution. The use of *pseudonyms* also protects personality. Women writers in Victorian times found it necessary to use male pen names to be taken seriously. For literature lovers, Mary Ann Evans is known to the world as George Eliot. Only a few people would know who Amandine Lucile Aurore Dupin was – but her pseudonym, George Sand, the first French female novelist of great reputation, is recognized by everyone.

In some cases, however, anonymity or anonymisation can be a harsh violation of rights – for instance, in the case of unrecognized authorship, when references are not used, or whenever paternity and identity are denied.

Issues of anonymity on the one hand and identity disclosure on the other hand are highly relevant in the contemporary debates of medical law as well. The anonymity of research subjects; the identity of gametes, tissues, and organ donors in cases of transplantation; the identity of gene donors in a biobank pose relatively novel concerns for bioethicists and scholars of biomedical law.

14.3 Anonymity and Confidentiality

Confidentiality "is the respectful handling of information disclosed within relationships of trust, such as healthcare relationships, especially as regards further disclosure. Confidentiality serves privacy. Researchers invariably promise to respect data-subjects' privacy, either by de-identifying the data to make them impersonal or by handling them securely."[5] While confidentiality originates from the

[5]Lowrance (2002, 8).

deontological norms of medical ethics, anonymity refers to a technical handling of data. The two notions are therefore strongly related to one another. While privacy has also an active dimension (control of personal information), confidentiality protects the doctor – patient relationship. The primary risks of a classical biobank to donors are related to the loss of confidentiality between either the doctor and the patient or the researcher and the donor. With time passing the nature of the danger normally diminishes to some extent, since research participants pass away and materials become archived ones. In biomedical research, in a biobank operated for therapeutic purposes or a population biobank, however the case is somewhat different: disclosing data against the data subjects' will, or accidentally identifying former donors (if the target group is small and research participants' identities are disclosed incidentally on the basis of circumstances even if information has been stripped of personal data that are mostly believed to contribute to identification) this may intrude well into the rights of persons other than the research subjects, first and foremost their relatives. This potential risk may be reduced by criminal[6] law sanctions, civil law sanctions,[7] or different forms of anonymisation. In order to avoid a breach of confidentiality, the same data protection rules and confidentiality standards have to apply for re-users of data, i.e. researchers in the original and in third countries. Due to the differences in the legal systems and the consequences attached to a breach of confidentiality, the safest way to transfer data to third countries is in anonymized format. As the Hungarian law of 2008 on the protection of human genetic data and the regulation of human genetic studies, research and biobanks (discussed below) prescribes, for the purposes of human genetic research, only anonymized, encoded or pseudonymised genetic samples or data may be transmitted to third countries, and only if the law of the given country provides for data protection corresponding to that under the Act No. LXIII of 1992 on the protection of personal data and the publicity of data of public interest. During the transmission of encoded genetic samples and data into third countries, the code key necessary for personal identification may not be transmitted.[8] An element of trust can be traced in the European – including the mentioned Hungarian – model,[9] since despite the discrepancies of the legal families of the European Economic Area, transfer of data to EEA countries shall be deemed as transfer within Hungary and the same confidentiality requirements are being presumed.[10]

[6]See for example Article 321 of the Swiss Criminal Code.

[7]See for example Article 33 of the Lithuanian Act on Data Protection on pecuniary and non-pecuniary damages.

[8]Article 28 (2).

[9]Article 29 Data Protection Working Party (2004, 19).

[10]Article 28 (1).

14.4 Functions of Anonymity in Health Care Law

As we saw one cannot tell whether anonymity is a positive value in itself in biomedical law: in some cases it has an important function, while in other fields anonymity can be a violation of important rights and interests. Ghost surgery (when surgery is conducted by someone who was not known by the patient prior to surgery) is a violation of informed consent. Publication of research results by disclosing relevant identifiers is a violation of privacy rights.

In the field of organ and tissue transplantation the name of the donor should be known to the medical staff. Moreover, a clear and accurate medical examination of compatibility is also an inevitable condition for donation while the recipient as a rule should not be connected to the (deceased) donors' family.

Clinical establishments involved in organ transplantations and coordinating organizations of transplantations do not disclose data and information in relation to donors and recipients, i.e. to those who are provided with organs. Under the current state of law organ transplantation is completely anonymous, and therefore all inquiries in relation to names or relatives are turned down. However, as in Hungary just a few such interventions are performed, if the donor's family declares that the organs of their relative will be used, the recipient may identify the donor from the scheduled date of the surgery. Furthermore, there exists no legal obstacle for the donor or the recipient to reveal this information.

The other field in medical law that kept anonymity as a main rule is the field of assisted reproduction when gametes are originated not from the social parents but from the donors. For a long time, in order to protect the integrity of the legal family, anonymity seemed to be a rule without exception. But if someone looks at the most recent changes in the field of the offspring's right to identity, it often seems to prevail over the donors' interests of anonymity. Of course, in the first countries where the laws on anonymity were changed, such as Sweden and United Kingdom, the legislative power did not adopt laws with retroactive effect. After the entry into force of the relevant legal instruments, the donors must be informed in advance on the possibility to reveal their identity in front of the child when it reaches maturity.

Having highlighted the meaning and importance of anonymity in related fields, in the following we will focus on anonymity in genetic research and anonymisation of data in genetic biobanks.

14.5 Genetic Research

Issues of anonymisation came into the frontline of the literature with the spread of large scale genetic tests and human genetic research. Human genetic research, being engaged with the structure of genetic material (genes and chromosomes), their disorders and the appearance in physical, intellectual, and behavioural features of the genetically encoded programme and the regular features of the transmission from the parents to the offspring of the genetically encoded programme and the

exploration of the disorders of these processes, has an overwhelming scientific significance. The laws on genetic research lay down the framework of the use for research of samples and data, data protection guarantees necessary for use, rules on genetic research on the population and the conditions of samples in the archived collection for a new research.

If the genetic sample taken in the context of genetic testing is intended to be used for research purposes, a repeated consent procedure is required by the law.

Genetic research on human behaviour should be conducted in a fashion that respects the dignity of the research participant by taking into account not only genetic but also the extra-genetic features of the personality. At the moment there are no detailed provisions how to ensure this. One solution could be if a social scientist having knowledge in psychology or/and sociology were involved in the study in order to avoid stigmatisation of the research participant and with the aim to help in developing a more balanced assessment of personality. By this method the danger of genetic determinism and reductionism could be more easily avoided.

During the research, the person concerned may request the encoding, pseudonymising or anonymising of the genetic sample intended for research purposes and that of the derived genetic data. The fate of the data in a biobank, its form of anonymisation, or possible destruction thus depends on the data subject – at least until a link can be established between him or her and the information or the sample. However simple that may sound, the complexity of the issue is highlighted by the fact that we do not possess a common definition of crucial terms, such as 'biobank' and 'anonymisation'. In the following we will give a brief overview of these notions.

14.6 Anonymity in Biobanks

DNA sampling, data collection, sharing and exchange of information are all important for genetic research, clinical care, and future treatments. However, the corresponding ethical and legal framework is not well defined. Most health care institutions have no written policies or agreements regarding this activity, and even if there was a willingness on the side of hospitals, clinics, and research institutes to adjust their practice to some general norms, researchers or drafters of these internal guidelines are in an extremely difficult position due to the large number of international, national, and professional guidelines that contain different, sometimes even contradicting recommendations relevant for biobanks.

A fundamental underlying question is how we define biobanks. Repositories of human samples and related data can be grouped along the stored material, which can be organs, tissue, blood, cells or other materials, such as urine or liquor. Biobanks can also be distinguished according to their sizes: these repositories may vary from population biobanks to three samples in a pathologist's refrigerator. A biobank does not only contain human biological samples, data are also stored there. Robert F. Weir and Robert S. Olick define biomedical classical (clinical) research databases as follows: a database that is developed and restricted to authorized clinical investigations (e.g. oncology, pathology, etc.) in several academic medical centres.[11] These databases contain genetic and other biomedical information about individual patients, derived from their clinically collected tissues, with the electronic data sometimes being transmitted to a central database. The above mentioned two scholars differentiate between commercial databases, which are human tissue databases that are restricted to scientists willing to pay to have access to DNA sequences and the databases that include other protected information.

A population biobank, based on the definition of the Council of Europe[12], is a collection of biological materials that has the following characteristics: (i) the collection has a population basis; (ii) it is established, or has been converted, to supply biological materials or data derived there from for multiple future research projects; (iii) it contains biological materials and associated personal data, which may include or be linked to genealogical, medical and lifestyle data, and which may be regularly updated; and (iv) it receives and supplies materials in an organized manner.

Forensic databases greatly differ in nature from the above classical and population biobanks. In the broad sense forensic databases are DNA databanks held by authorized laboratories of police and official forensic institutions for criminal and other legal procedures, such as the identification of victims, missing persons, perpetrators, the establishment or rejection of paternity, etc.

One may think of other divisions of biobanks as well, but the crucial point for our current discussion is the double nature of these databases, i.e. the fact that they contain both tissues and data, that is information on the donated human biological material and the donor linked to these tissues. Therefore the question arises of whether traditional data protection rules are an effective tool in the fight against the misuse of information, and whether anonymisation of samples is the best safeguard, or on the contrary, whether it limits the autonomy of research participants in a biobank – a question very much related to the issue of genetic exceptionalism.[13] Some authors[14] state that since tissues and data are different, they raise different issues. Therefore different regulations are said to be needed. At the same time, there

[11] Weir and Olick (2004, 294).

[12] Recommendation Rec(2006)4 of the Committee of Ministers to member states on research on biological materials of human origin, Article 17.

[13] The *Genomics Law Report* defines genetic exceptionalism in the following way: "Genetic exceptionalism is the concept that genetic information is inherently unique and should be treated differently in law than other forms of personal or medical information. There are several reasons for such special consideration: genetic information can predict disease occurrence in a person and their blood relatives; it uniquely identifies a person; and it can be used to discriminate and stigmatize individuals. While these issues deserve attention and steps should be taken to protect people, over-regulation could limit our ability to investigate how genetic information predicts disease and improve medical outcomes." Available at www.genomicslawreport.com/index.php/2009/10/06/genetic-exceptionalism-and-the-precautionary-principle (accessed March 11, 2011).

[14] Trouet and Sprumont (2002, 3–19).

is a trend for unified regulations as well. If DNA represents special human rights questions, its protection should reflect these corresponding concerns.

Many scholars and researchers consider tissue research a form of medical research, or at least realize the similarities between the two, and therefore propose that the protection of personal medical data shall cover this field. Confidentiality can be ensured through various legal means: antidiscrimination laws prohibit discrimination, while criminal laws may also sanction discriminatory behaviour. Confidentiality may also be ensured through anonymisation. This is the point where scientists' interests may clash with legal requirements. Based on the study "Ethical and regulatory aspects of biobanks: global consensus and controversies", Bernice Elger summarized the literature and regulatory frameworks on confidentiality, anonymisation and consent.[15] Elger and her colleagues interviewed persons from all related disciplines, such as scientists, biobankers, physicians, lawyers, and ethicists from different parts of the world and from different types of institutions. Experts and researchers agreed on only a few issues: first, they are opposing irreversible anonymisation of samples at the time of collection and storage; second, in their view, researchers have to be tightly controlled; third, a distinction needs to be drawn between publicly and privately funded projects; fourth, it is strongly advisable to place research data and results in the public domain within a reasonable time-frame in order to stimulate scientific progress; and finally, they call for a unified definition of anonymisation and establishing common conditions under which material and data are shared with others. The last issue is especially topical for our analysis, since data sharing and transnational research are hampered by the differing understandings of anonymisation and pseudonymization.

Pseudonymisation refers to a technique of data processing in which anonymity is assured assure while keeping a link to be able to update information and to re-contact participants whenever information valuable to the donors is discovered. The next logical step is to determine what kinds of pseudonymisation techniques are adequate: double coding, single coding or some other method. Even if one term refers to a certain technique method, a lack of consensus on the normative definition prevents researchers from agreeing on standardisation.

14.6.1 Heterogeneity of Norms

In the myriad of terms one can find references to samples that are anonymous, anonymised, anonymously coded, coded, unidentified, de-linked, permanently de-linked, not traceable, unlinked, identifiably linked, pseudonymised, encoded, encrypted, directly identified, confidential, identifiable, not traceable, or in the

[15]Ibid. Other collaborators in the research were: Nikola Biller-Andorno, University of Zurich, Switzerland; Agomoni Ganguli-Mitra, University of Zurich, Switzerland; Andrea Boggio, Bryant University, USA; Alexander Mauron, University of Geneva, Switzerland; and Alexander M. Capron, University of Southern California, USA.

UNESCO terminology[16]: data linked and unlinked to an identifiable person, furthermore, data irretrievably unlinked to an identifiable person.[17] Data unlinked to an identifiable person means data replaced by or separated from all identifying information about that person by use of a code. It can be applied to a biological, whereas data irretrievably unlinked to an identifiable person is data that cannot be linked to an identifiable person, because the link to any identifying information has been destroyed.[18]

Different legal families adhere to distinct legal traditions, and prefer one or another term over others for legal historical reasons. Sometimes even the same term is used with a different meaning, like the words "anonymised" and "coded" which are filled with different content in continental and common law jurisdictions.

In the European setting the right to privacy is laid down in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms. The European Court of Human Rights deducted the right to data protection from that provision. It regularly refers to the 1981 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) and the Additional Protocol to Convention 108 regarding supervisory authorities and trans-border data flows (ETS No. 181), also adopted in the framework of the Council of Europe. Among the European standards, Recommendation Rec (2006) 4 of the Committee of Ministers of the Council of Europe to member states on research on biological materials of human origin can be referred to first. The instrument[19] distinguishes between non-identifiable and identifiable samples. The former are unlinked samples, while the latter are linked anonymised and coded samples.

As to European standards on anonymisation specifically, Recommendation No. R (97) 5 of the Committee of Ministers to Member States on the Protection of Medical Data may be helpful: "the expression 'personal data' covers any information relating to an identified or identifiable individual. An individual shall not be regarded as 'identifiable' if identification requires an unreasonable amount of time and manpower. In cases where the individual is not identifiable, the data are referred to as anonymous" (Principle 1).

Apart from the Charter of Fundamental Rights, and more specifically its Articles 7 and 8 on privacy and data protection, the main European Union data protection instrument is Directive 95/46/EC. This document, however, does not speak of an "unreasonable amount of time and manpower," but states in Article 2a that

> "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.

[16]International Declaration on Human Genetic Data, 16 October 2003, Article 2, Points (ix) and (x).

[17]Elger and Caplan (2006, 661–66).

[18]UNESCO International Declaration on Human Genetic Data, 2003, Article 2, (x) and (xi).

[19]See Article 3 on the identifiability of biological materials.

In the context of biobanks, especially when it comes to information security, Article 17 of the Directive is worth deeper exploration. According to this provision the controller is obliged to

> implement appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing.

Article 29 of the EU Data Protection Directive 95/46/EC establishes a "Working Party on the Protection of Individuals with regard to the processing of Personal Data" (hereinafter referred to as the "Article 29 Working Party"). According to the Article 29 Working Party, electronic health records create a new risk scenario,[20] and acknowledging that genetic data may pose special risks even among sensitive data, the data controller can be requested to carry out risk assessment, establish security policies and provide ongoing training for staff.[21]

Bernice Elger and Arthur Caplan summarize the European approach to distinguishing levels of anonymisation in the following.[22] The first category of *anonymous* DNA samples does not exist, only for "archaeological" tissue for which no material for comparison to an identified person exists. The second type of samples are *anonymised* ones, which are stored alongside certain information which is crucial for research, but information that would allow the identification of the donor is all stripped. Depending on whether the latter information can be restored or not, anonymised samples can be divided into *irreversibly anonymised* (unlinked) and *reversibly anonymised* (linked) ones. In the latter case identification is possible via a code (pseudonym), but researchers or users of the material do not have access to the code. *Coded* samples are like linked (reversibly) anonymised ones, with the difference that researchers or users do have access to the code. One has to be cautious with the terminology, as in the US "anonymised" means irreversibly unlinked or reversibly linked, but the researchers do not have access to the code, while in Europe the word "coded" means reversibly linked, where researchers have or do not have access to the code. The last category is that of *identified samples*, where the information stored along the samples permits the direct identification of the donor, such as when the name and birthday are indicated on a tube.

Putting these terminological discrepancies apart, the main controversy has evolved around the question of how to assure adequate anonymisation – be it linked or unlinked. Who shall decide which degree of anonymisation is adequate? How many characteristics must be stripped to obtain truly irreversible or reversible economisation? What are the standards for technical questions of security?

[20] Article 29 Data Protection Working Party (2007a, 11).

[21] Article 29 Data Protection Working Party (2004, 11–12).

[22] Elger and Caplan, *op. cit.*

14.6.2 Technical Solutions of Data Security

Addressing technical questions of security, the need for standards and for cooperation with IT institutions has been stressed. Again Article 29 Working Party may give some guidance as to the preferred standards: in its Opinion 4/2007 the Working Party stated that "Pseudonyms should be random and unpredictable. The number of pseudonyms possible should be so large that the same pseudonym is never randomly selected twice. If a high level of security is required, the set of potential pseudonyms must be at least equal to the range of values of secure cryptographic hash functions."[23] In its previously drafted working document on the processing of personal data relating to health in electronic health records[24] Article 29 Working Party promoted, among diverse technical solutions, Privacy Enhancing Technologies PETs. PETs are IT solutions that mitigate the drawbacks of technological development in personal data management, so that donors (or data subjects in general) regain influence over information about themselves. The Working Party also proposed that legal safeguards refer, among others, to the development of a reliable and effective system of electronic identification and authentication as well as constantly updated registers for authorized persons who can access databases; documentation of all processing steps which have taken place within the system; and preventing unauthorized access or alteration of data.[25]

When searching for effective anonymisation in case of biobanks, one may borrow solutions from other fields where data protection is a concern, such as protocols in securing internet communications, emails, online purchase, etc. A viable solution is the anonymous tracking model for individual minority subsidies, where the state wishes to reduce or eliminate the disadvantages suffered by certain minorities by positive discrimination or affirmative action programs. An interesting field for comparison is the case of minority subsidies. The question is how to subsidise disadvantaged minorities if we cannot identify them, because the law prohibits having certain characteristics (e.g. ethnicity) registered. On the one hand, these pieces of sensitive information enjoy special protection, and we wish to deny authorities' access to it. On the other hand, this information would be crucial in order to have an effective and correct system of subsidy, free from abuse and embezzlement. In order to overcome the problem, several authors propose the use of modern information technology, unidirectional data transformation procedures, and emphasize the importance of trusted third parties. The technique can be adapted to different settings, like genetic databanks, as well, especially since in both cases the handling of sensitive, special, classified information is at stake. In such a sensitive setting one may make use of the available modern information technology which offers data

[23] Article 29 Data Protection Working Party (2007b, 18).

[24] Article 29 Data Protection Working Party (2007a, 11).

[25] Ibid., 19–20.

management technology, that allows to make the link between the data and the data subject, in our case the patient, the donor, or the suspect, unidirectional.[26]

A probabilistic distortion method was suggested by Johannes Gehrke from Cornell University, which would totally disable re-identification of donors. According to this method an extremely small probabilistic number (that might be positive, negative, or zero) is added to the values in the database, thereby distorting the original figures so they can never be traced back. This number can be mathematically tailored, customized, so that the statistical properties of the attributes may be the same, i.e. the probabilistic number is small enough not to modify the outcome of the research, but is large enough to ensure that data cannot be joined by simply testing equality attributes. Of course, this system, just as any other alternatives, may have some drawbacks: it may lead to distorted results in case specific attributes are being compared, and it clearly introduces an uncertainty element.[27]

The fact that a sample is seen as being anonymised has vital consequences from the point of view of obligations of acquiring informed consent. Should the definition of anonymisation be broadened to too many types of pseudonymised samples, or, if the definition is softened by reference to a minimal risk of identification, or to reasonable effort, or reasonable amount of time and manpower needed for identification, data protection rules do not apply any more. More specifically, in the EU context we can derive from Recital 26 of Directive 95/46/EC that data collected in an anonymous way, or data that have been rendered anonymous at a later point in time, are outside the scope of the Data Protection Directive 95/46/EC, since "the principles of protection must apply to any information concerning an identified or identifiable person" only. Therefore, the British approach prescribes safeguards even for anonymisation, since thereafter, on the one hand, the data subject will not be able to influence the fate of the data relating to him or her, and on the other hand, legal guarantees will not apply. Therefore, the Office of the Information Commissioner issued a legal guidance on the Data Protection Act of 1998, which states that "in anonymising personal data the data controller will be processing such data and, in respect of such processing, will still need to comply with the provisions of the Act."[28] Thereafter however – in line with the provisions of the Directive – the data fall outside the scope of the law.

14.6.3 The Case of Estonia: Double Coded Samples

Current Estonian informed consent rules have their roots in the European Union Data Protection Directive 95/46/EC. However, the European Union rules do not include extensive informing requirements; consent sometimes is not even necessary. Informed consent in data protection differs greatly from informed consent in medicine. Informed consent in medicine (i) aims at the protection of life and health, (ii) extensive information is required, (iii) it is project specific, and (iv) consent is almost always necessary. In comparison, informed consent in data protection (i) aims to protect privacy, (ii) less information is required, (iii) specification of one field of use is enough, and (iv) there are a number of cases where consent is not necessary.[29]

In the Estonian Genome Project data is stored in a coded form, where persons are identifiable. In case of a consent withdrawal, the code will be erased; however, erasure of all data can also be applied for. Whenever needed, data is issued in pseudonymised form from which data subjects cannot be identified – neither directly, nor indirectly. This is realised through the so-called five donors rule, which ensures that every data in the database matches at least five persons. For each and every research use the ethics committee's approval is needed.

The previous consent form is not suitable for the Estonian Genome Project, since it enables future research with yet unknown project goals. Theoretically one could opt for asking for a specific consent at a later stage after data collection. However, according to the Estonian expert, for practical reasons, taking into account the nature and level of risks, considering autonomy as empowerment rather than as a disempowerment, bearing in mind the value of biological samples, population biobanks deserve a new type of informed consent. Therefore, in the Estonian Genome Project an open consent requirement has been adopted. Open consent is consent to participate in a population biobank and in research projects utilising data and/or samples from a biobank. The consent is open, i.e. not limited in respect of time, projects, researchers, etc. It justifies interference with bodily integrity and data privacy. It should be noted, however, that even open consent does not give authorization for everything. Neither is it an indication for commitment to future participation, i.e. withdrawal of consent is still possible. Further, the open consent system relies on certain conditions, like public control.

Concerning the crucial issue of withdrawal of the gene donor's consent, Article 12 Section (4) point (7) of the Human Genes Research Act (HGRA)[30] sets forth that gene donors have the right to withdraw their consent until tissue samples or the descriptions of state of health are coded, and in such case the gathered information and blood sample shall be destroyed. Afterwards, a gene donor has the right to apply, at any time, to the chief processor for the destruction of data that enables decoding.[31] However, destruction of data that enable decoding (i.e. anonymization) does not mean also the destruction of biological material or other data.[32] In line

[26] See, for example, Claerhout and DeMoor (2005, 257–65); Székely (2009, 27–62).

[27] Referred to by Zoltán Alexin at the second international workshop organised in Budapest within the Tiss.EU project. See also Xiao, Wang, and Gehrke (2009).

[28] Data Protection Act 1998 (2002, 15).

[29] See the findings of Ants Nõmper at the second international workshop organised in Budapest within the Tiss.EU project.

[30] Human Genes Research Act (passed by the Riigikogu) (December 13, 2000, 104, 685).

[31] Article 20 Section (1) HGRA.

[32] Nõmper and Kruuv (2003, 213–24).

with Article 10 Section (2) of the HGRA, a gene donor has the right to request termination of biological material and other data available in the genebank, if the identity of a gene donor is unlawfully disclosed. After the data that allows decoding is destroyed, the health data and the tissue samples of a gene donor, stored in the genome bank, are anonymous. Thus, the regulation of personal data protection does not apply, since it does not extend to data processing performed with anonymised data in line with Article 7 Section (2).

14.6.4 The Case of Hungary: Three Options to Grant Confidentiality of Genetic Samples

The Hungarian law adopted in 2008 on the protection of human genetic data and the regulation of human genetic studies, research and biobanks presents a unique solution on the European continent, therefore, we shall elaborate the details of the regulation. First, the debates surrounding lawmaking will be presented and second, we will discuss the rules on anonymisation in greater detail.

Hungary became Party to the Oviedo Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine of 4 April 1997, and to its Additional Protocol on the Prohibition of Cloning Human Beings by Act VI of 2002. Thereby Hungary undertook to monitor the regulatory range of bioethics and medical-biological research continuously and to prepare legislation in this subject, and that is what the Act aims to comply with.

As to European Union legislation, the following was taken into account by the lawmaker: 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 and Directive 2004/23/EC of the European Parliament and of the Council of 31 March 2004 on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage, and distribution of human tissues and cells.

The Parliamentary Assembly of the Council of Europe discussed the draft Additional Protocol on genetic testing for health purposes to the Oviedo Convention, covering the field related to the Act, with the exception of genetic research, at its session in Strasbourg between 21 and 25 January 2008. During the preparation of the Hungarian Act on biobanks, the draft of this Protocol was also reviewed.

As a result of the legislative process, Act No XXI of 2008 on the protection of human genetic data and the regulation of human genetic studies, research and biobanks entered into force on 1 July 2008. The law has become shorter and simpler than it was foreseen in the original policy paper in a desperate effort to avoid sensitive issues. Thus, the law addresses the use of genetic information only in a very narrow biomedical sector: in the fields of genetic testing, screening, and research.[33]

[33]In Article 1 of the Act the purpose of the law is stated as "to lay down rules on human genetic tests and screening (studies) and human genetic research, the conditions and purposes of the treatment of genetic data and rules on biobanks." The Act applies to genetic sampling for human genetic study and human genetic research performed under this Act in the territory of the Republic of Hungary, the processing of genetic data irrespective of the place of sampling, and to genetic testing and screening and human genetic research and to biobanks.

The law restricts the use of genetic data only in this biomedical context, so even in the lack of regulation of the broader use of genetic data based on the Act genetic data processed for diagnostic or research purposes cannot be disseminated for the purposes of insurance. Despite the intended laconic law, the mere word "genetics" was an invitation for a vehement debate by various political actors. Fears of genetic discrimination, exploitation or trafficking data to foreign countries were the major concerns in the political debate.

Even earlier in the course of the legislative debate lawyers, data protection activists were mobilized and advocated for newer and newer guarantees for the protection of genetic data.[34] The issues of data protection were so dominant in the debate that some other, broader human rights aspects were entirely left out from the final version of the law.[35]

By focusing on data protection questions, such as how to store genetic data (should it be stored as anonymous data, coded, single or double coded, and who should get access to the code or who should keep the code?) and creating a stronger protection for the genetic data, some other elements of the ethical-legal framework were sacrificed, such as the prohibition of discrimination based on genetic characteristics; they were referred to general laws. As relatively little public consultations were conducted on biobanks in Hungary, various data protection rules, including the protection of the health care data are fragmented and dispersed in various norms.

The consequences of the careful approach towards data protection make the Hungarian solution unique. The specificity of the Hungarian Parliamentary Act in its final form is that it regulates three different levels of coding and anonymity:

(a) *the encoded genetic sample or data* means genetic sample or data regarding which all the personal identification data relating to the person giving the sample are replaced by a code;
(b) *pseudonym genetic sample or data* means encoded genetic sample or data regarding which the code replacing the personal identification data was provided to the person concerned;
(c) *anonymised genetic sample or data* means genetic sample or data regarding which all the personal identification data relating to the person giving the sample was made incapable of identifying the person.[36]

[34]In order to understand the main focus of the debate, it should be mentioned that in the Hungarian law, the most considerable field within the right to privacy is the protection of personal data. The classical concepts of inviolability of domicile and secrecy of correspondence are also important subjects to be protected, but there is a much higher uncertainty in the abstract fields of privacy e.g. concerning the right to disposal of someone's personal body.

[35]Such as the right not to be discriminated against in the field of public health insurance and education.

[36]See Article 3 Points (d), (e), and (f).

For a long time it was believed, also in Hungary, that anonymous data could guarantee the highest level of protection for the genetic data. However, many problems were identified in respect of systematic anonymization of genetic samples and data. First of all, anonymous data cannot be matched with other health data, and as such the relevance of the data is reduced for scientific research. Anonymisation is also contestable, taking into account the participants' interests, since a further feedback, based on the request of the owner of the sample, would be usually impossible.

The third part of the law needs greater elaboration, as it provides specifically for rules on the operation of *biobanks*. By legislating on the operational rules of biobanks, the conditions of the operation of collections containing human biological material samples shall be established. Accordingly, the genetic samples and data shall be stored only in biobanks and, as a general rule, in a format determined by the declaration of consent of the person concerned. There is a safeguard provision providing for the conditions of the storage of the genetic sample or data in a way that allows personal identification and states the prohibition of a register involving information that contains personal identification data. It is stated that a biobank may be established and maintained by a health service provider authorised to conduct genetic studies and certain medical researchers and another institution entitled to conduct human genetic research only.[37] Larger scale population based study is also mentioned in the Act. Under Article 17 human genetic research on the population may be conducted for the determination of the distribution of genetic variants between individuals within a given group or between individuals belonging to different groups, and to the exploration of the nature and consequences of the latter. The law provides for the tasks of the responsible person being employed in the biobank, the keeping of data stored in the biobank and the forwarding of data as well as the register of biobanks.

During the storage of the genetic sample or data, the protection of these shall be ensured against destruction, termination, change, injury, publication or access by unauthorised persons.[38] Unless provided otherwise by this Act, genetic samples and data shall be stored in an encoded format. Encoded genetic samples, data and code keys shall be stored separately, both physically and electronically. Access to the code key shall be authorised to a person being responsible[39] within the framework of the Act. During the separate storage of the code key, it shall be ensured that no other person may access it apart from the person entitled thereto. The code of the pseudonym sample or data shall be put at the exclusive disposal of the person providing the sample. Storage of genetic sample or data together with personal identification data is

[37] Article 23 (1).

[38] Article 24.

[39] Within the biobank, the person responsible for the protection of genetic samples and genetic data, the registering of genetic samples and data and the keeping of the register shall be the head of the institution maintaining the biobank and the person designated by the latter for the supervision of the operation of the biobank. Article 26 (1).

subject to the consent of the person concerned.[40] A register containing genetic samples and data stored together with personal identification data or encoded genetic samples and data may not be linked to a register containing personal identification data.[41]

Every genetic sample and data stored in the biobank and all related procedures and activities and the forwarding of the genetic sample and data shall be registered for at least 30 years following the recording of the data, except when the person concerned withdraws his or her consent to the treatment of genetic data. In such a case, every register relating to genetic data shall be destroyed following the information of the person concerned. The register shall contain the types, quantities, origins and destination of collected, studied, stored, processed and distributed or otherwise used genetic samples and the genetic data derived from these. After expiry of the mandatory registration period, the data shall be subject to treatment under the Act XLVII of 1997 on Health Care Data.

14.7 Conclusions: End of Anonymity?

In case of contemporary biobanks legislations it seems that we have much less emphasis now on anonymity than at the dawn of the first biobanks. Anonymity seems to be an illusion that we might never be able to achieve in case of genetic samples, and perhaps it no longer serves the interests of research participants. Anonymisation was seen as an attractive tool in the securing privacy and autonomy, in the prevention of harm that the leaking out of information to unauthorized persons – such as insurers or employers – may mean. Securing privacy, confidentiality, data protection or autonomy does not however mean full anonymization. First, because anonymous DNA samples do not exist, since theoretically one might always derive samples from living donors and archived materials, and compare them to a sample in a biobank. Second, and more importantly, anonymisation does not necessarily serve the interest of donors. Donors shall be aware of the details of data protection, such as: who has access to their data? Under what conditions do they have access, and what is the level of security?[42] Patients cannot determine the destination of their samples. Furthermore, feedback of research results is the most desired outcome of such research, which is also disabled by delinking samples and data. Thirdly, stripping the code or delinking information from samples prevents researchers from going back to patients and conduct longitudinal research. Ultimately, by slowing down research, anonymisation may harm donors and non-donor patients more than a secure system of pseudoanonymisation. Such a system, of course, cannot operate without an element of trust, which can only be established if the necessary institutions, safeguards (including confidentiality

[40] Article 25 (1).

[41] Article 25 (2).

[42] Chadwick (2001, 203–10, at 207).

requirements, respecting donors' and their relatives' privacy and the right not to know, etc.), and procedures are created – not without establishing a corresponding legislative framework.

References

Article 29 Data Protection Working Party. 2004. Working Document on Genetic Data of 17 March 2004, 12178/03/EN, WP 91.

Article 29 Data Protection Working Party. 2007a. Working Document on the Processing of Personal Data Relating to Health in Electronic Health Records (EHR) of 15 February 2007, 00323/07/EN WP 131.

Article 29 Data Protection Working Party. 2007b. Opinion 4/2007 on the Concept of Personal Data of 20 June 2007, 01248/07/EN WP 136.

Chadwick, R. 2001. "Informed Consent in Genetic Research." In *Informed Consent in Medical Research*, edited by L. Doyal and J. S. Tobias, 203–10. London: BMJ Books.

Claerhout, B., and G. J. E. DeMoor. March 2005. "Privacy Protection for Clinical and Genomic Data: The Use of Privacy-Enhancing Techniques in Medicine." *International Journal of Medical Informatics* 74 (2): 257–65.

Elger, B. S., and A. L. Caplan. 2006. "Consent and Anonymization in Research Involving Biobanks: Differing Terms and Norms Present Serious Barriers to an International Framework." *European Molecular Biology Organization Reports* 7 (7): 661–66.

Ellis, I., and G. Mannion. 2001. "Humanity Versus Utility in the Ethics of Research on Human Genetic Material." *Genetics Law Monitor* 1 (5): 1.

Human Genes Research Act (passed by the Riigikogu). December 13, 2000. (as RT I 2000, 104, 685), entered into force on January 8, 2001 (RT is *Riigi Teataja* or *State Gazette*). In the original language: *Inimgeeniuuringute seadus*, official English translation is available at http://www.legaltext.ee/text/en/X50010.htm. Accessed March 11, 2011.

Laurie, G. T. 2002. *Genetic Privacy: A Challenge to Medico-Legal Norms*. Cambridge: Cambridge University Press.

Data Protection Act of 1998. 2002. *Legal Guidance*. London: Office of the Information Commissioner.

Lowrance, W. W. 2002. *Learning from Experience: Privacy and the Secondary Use of Data in Health Research*. London: Nuffield Trust.

Nõmper, A. and K. Kruuv. (2003) "The Estonian Genome Project." In *Society and Genetic Information: Codes and Laws in the Genetic Era*, edited by J. Sándor, 213–24. Budapest: CEU Press.

Recommendation Rec (2006) 4 of the Committee of Ministers to member states on research on biological materials of human origin.

Székely, I. 2009. "Positive Discrimination and Data Protection: A Typology of Solutions and the Use of Modern Information Technologies." In *Privacy Protection and Minority Rights*, edited by M. D. Szabó, 27–62. Budapest: Eötvös Károly Policy Institute. Accessed March 11, 2011. http://www.ekint.org/ekint_files/File/kiadvanyok/privacy_minority.pdf.

Trouet, C., and D. Sprumont. 2002. "Biobanks: Investing in Regulation." In *Baltic Yearbook of International Law*, edited by I. Ziemele, 3–19, vol. 2. Leiden: Brill/Martinus Nijhoff Publishers.

UNESCO International Declaration on Human Genetic Data, 16 October 2003.

Weir, R. F., and R. S. Olick. 2004. *The Stored Tissue Issue*. Oxford: Oxford University Press.

Xiao, X., G. Wang, and J. Gehrke. 2009. Interactive Anonymization of Sensitive Data, *Proceedings of the 35th SIGMOD International Conference on Management of Data, 2009*, 1051–1054. Accessed March 11, 2011. www.cs.cornell.edu/bigreddata/publications/2009/sigmod2009-p1051-xiao.pdf.

Epilogue

As we have seen in cases of tissue collections we witness a heterogeneity of various ethical and legal terms and approaches. In many jurisdictions tissue collections and biobanks exist under different names. Various overlapping terms have been in use, including the terms "registries, repositories, biological archives, pathological sample-collections, genome databases, gene-banks, population biobanks" etc. While some years ago the term "biobank" was almost unheard of, it seems to be used too broadly now. Therefore, the crucial moment for adopting a European-wide definition for biobanks potentially has already passed. Not just the wide application of the catchy term "biobanks" but also the controversies in the social meaning of genetic data and some countries' rejection of the term "bank" (because of the strong connection with the commercial applications) seem to be obstacles to the late harmonisation of the term "biobank". Furthermore, it seems that while biobanks for genetic studies were the centre of attention about four or five years ago, now new forms of biological collections also have to be taken into account. Particularly in the field of regenerative medicine, tissue and cell collections have been developed for purposes of research and therapy that aren't genetic in nature.

Framing and reframing new technologies under the EU Tissue Directive[1] and the EU Regulation on Advance Therapies[2] seems to even further complicate the question. Clinicians often misunderstand the scope and applicability of these norms, in addition to the fact that there are countries that tend to apply the norms of organ and tissue transplantation even in the field of tissue research in biobanks. In general, it can be noted that while the subject matter – such as human tissue – involves various activities for clinicians and researchers, from harvesting and collecting, to therapeutic use and research, in law, the function of the tissue determines the legal framework. It follows that the use of tissue for human therapy or in vivo research requires strict safety measures, while doing in vitro research on human tissues raises

[1] Directive 200423/EC of the European Parliament and of the Council of 31 March 2004 on setting standards of quality and safety for donation procurement, testing, processing, preservation, storage and distribution of human tissues and cells.

[2] Regulation (EC) No 1394/2007 of the European Parliament and of the Council of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004.

C. Lenk et al. (eds.), *Biobanks and Tissue Research*, The International Library of Ethics, Law and Technology 8, DOI 10.1007/978-94-007-1673-5,