

Collaborative projects involving research biobanks – ethical and legal aspects of data protection

Bernice S. Elger^a, Eva De Clercq^b

^a Associate professor (health law, ethics, humanitarian medicine), Centre for legal medicine, University of Geneva

^b Senior researcher, PhD (bioethics), Institute for biomedical ethics, University of Basel

Abstract

Collaborative projects involving research biobanks raise various ethical and legal issues as insufficient data protection and/or inappropriate data management may have harmful effects for sample donors. The idea of data protection follows from the principle of non-maleficence and from the respect for autonomy of donors. The idea of patient (or donor) protection, by contrast, is based on the concept of beneficence and can be in conflict with those autonomy rights. In the present article we focus on this often tense relationship between data protection and patient (or donor) well-being. For this purpose, we first address the different levels of anonymity and data protection. Next we identify the advantages and disadvantages and examine the ethical acceptability of the various possible solutions to this problem. Our main aim is to raise awareness about the ethical issues related to data protection in biobanking, including knowledge about the different levels of anonymity, the careful use of appropriate terminology, and the identification and acknowledgement of various and possibly conflicting interests.

Introduction

Biobanks have become an indispensable tool of translational research in many countries and most areas of medicine [1–6]. The scientific value of biobank research depends on both procedural and ethical factors. The first aspect is associated with the ‘technical’ quality of samples [7]: the tissue has to be taken from the correct location, e.g. the tumor tissue and not adjacent healthy areas; the samples have to be processed rapidly to preserve the quality of the DNA; and the storage facilities have to ensure the best possible long term quality of the tissue or extracted DNA [8]. The ethical dimension is related to the twin-respect for autonomy and non-maleficence. The principle of autonomy demands that future use of human biological samples is granted through informed consent of the research participants. Further, if collaborative research is planned in different countries, informed consent forms should fulfill the legal and ethical requirements of various domestic frameworks. Protection from harm or non-maleficence is another major issue in human subject research. However, since biobank research is not carried out directly on persons, the principal mechanism of harm is not directly physical, but is intrinsically linked to appropriate data protection. The risk of violation of confidentiality can in fact lead to adverse effects such as stigmatization, abuse of

sensitive information against the interests of individuals and groups by insurance companies or employers or other third parties. The principles of non-maleficence and autonomy are closely related to the following two issues: (1) the access of researchers to the identity of so-called “interesting” donor individuals or groups in order to request further information and new samples and maybe more importantly (2) questions related to the right to know *and* not to know the results of research. The latter right has been developed mainly in response to the expansion of predictive genetic testing in the 1990s. Insufficient data protection and/or inappropriate data management may have harmful effects for sample donors. Participants might, in fact, against their interest and autonomous wish, be informed about research results that will have a negative impact on their lives. However, the desire to reveal these results is consistent with the principle of beneficence as they might save participants’ life or significantly improve their health. The problem of return of results confronts us with the difficulty of balancing different ethical considerations: while the tracing of participants must be possible in order to reveal clinically valuable and beneficial results [9], caution should be exercised to prevent unauthorized identification of sample donors.

In the present article we want to focus upon this often tense relationship between data protection and patient well-being. For this purpose, we first address the different levels of anonymity and data protection in the era of research biobanking. Next we identify the advantages and disadvantages and examine the ethical acceptability of the various possible solutions to this problem while keeping in mind that the autonomy rights of the involved stakeholders, i.e. patients and researchers, might clash. Our main aim is to raise awareness about the ethical issues related to data protection in biobanking, including knowledge about the different levels of anonymity, the careful use of appropriate terminology, and the identification and acknowledgement of various and possibly conflicting interests. Hence, we will not discuss the procedural factors linked to collecting, processing and storing such tissue or DNA samples.

Different levels of anonymity

If we examine the idea of “data protection” from a quantitative point of view, the greater the anonymity the less

likely it is that the donor of a biological sample or source of data will be identifiable. That means that protection is at its highest possible level if anonymity is maximized and identifiability minimized. Various classification systems exist concerning anonymity. The use of multiple and contradictory terminologies has been compared by some to a Babylonian confusion [10, 11]. In the present article we will use the terminology proposed in the guidelines published by the Council of Europe with the aim to harmonize practices and terminology in Europe [12]. Five levels of anonymity are distinguished:

Identified samples and data: researchers and biobank personnel have direct access to the identity of the sample and data sources, i.e. through a label that mentions the patient's or research subject's name or other identifiers such as a social security number.

Reversibly anonymized samples and data: all details that will permit the donor identification are stripped reversibly and a distinct pseudonym is used to refer to each donor. Research and biobank personnel who use data or samples do not have access to the pseudonym which links the data to the identity of the donor, only an independent secret holder or the treating physician has. **Coded samples and data:** the same definition as reversible anonymisation with the exception that researchers and biobank personnel have access to the pseudonym or code and are able to find out the identity of the sample donor.

Irreversibly anonymised samples and data: all details that will permit donor identification are stripped irreversibly.

Anonymous samples and data: Full anonymity is difficult to achieve for genetic material as long as family members are alive and DNA finger printing permits to identify the donor.

In some national classification systems several of the five categories are merged. For example, in the Swiss federal law [13], as well as in the classification used by the US Office of Human Research Protection [14], both coded and reversible anonymization are combined into one category that is called "coded". Hence, it is important to be critical and not to overlook the different levels of protection that are possible within the Swiss federal category of "coded" depending on who has access to the code.

The most secure solution – which level of anonymity is appropriate?

Discussions have been ongoing for at least a decade as to which level of anonymity is the most appropriate [15–18]. It is important to notice that the highest possible anonymity of data is not necessarily the most ethical solution. The above described five levels of anonymity all have distinct advantages and disadvantages. This may explain why various research participants, researchers and biobank personnel disagree on which values or principles should be given priority. While on

the one hand, the most secure irreversible anonymisation makes it more difficult to find out the identity of data sources, the risk of data mismanagement is increased as this type of data protection renders it more difficult to prevent multiple entries of the data of the same participant. It also precludes to add new data and samples in the future and to return clinically useful results to participants. The use of anonymized and coded data by contrast permits to obtain significant benefits, including the addition of new data and samples and the return of clinically useful results to participants. It is evident that reversible anonymization and the use of an independent secret holder provides more secure protection than the use of coded data where researchers and biobankers have access to the code. On the other hand, as reported by experts in an international study [17], the risk of errors and data management increases with the complexity of coding systems. Using identified data does not significantly increase advantages, but decreases data protection to a high extent. Further, despite the pro and cons of these various types of data protection, we should not forget that while protective technical means are constantly being improved [18–20], hacking and other ways of unlawful data breaches are growing in parallel [21, 22]. But more importantly, information technology advances as such do not offer a solution to the fact that as long as genetic information is provided and comparative DNA from other data basis or from family members are available, data protection will never be complete [11, 23]. Therefore the only truly effective "data protection" is the one that protects research participants from unethical and abusive use of information through policy that prohibits efficiently data abuse by employers, insurance companies or other stakeholders who could have a vicious interest in the data.

Data protection, patient protection, or both?

The frequent use of the term "data protection" could lead to the assumption that the ultimate goal of protection is to keep data secret by all means. But how does data protection actually relate to the ethical concept of patient or research subject protection? A closer look at both the concepts of "data protection" and "patient protection" shows that they may be in conflict.

The idea of data protection follows from the principle of non-maleficence and from the respect for autonomy of sample donors, in particular then from the right for respect of privacy and the right to control the data use [23–25]. This means that patients or research subjects might want to balance stricter privacy protection, for example through irreversible anonymization, with their interests to be recontacted and informed of results which would require less strict privacy protection, i.e. the use of reversibly anonymized data by researchers and biobank managers.

The idea of patient protection, if it is based on the concept of beneficence, can be, at least partially, in conflict with those autonomy rights. Research ethics committees, for example can restrict autonomy rights for the benefit of increased patient protection. However, it is not clear in which direction this protective restriction might go, i.e. (1) towards greater (for example irreversible) anonymity which would ensure protection from discrimination or (2) towards lower levels of anonymity which would allow the return and communication of clinically useful results.

What is new: ethical and legal advances in data protection?

In the past, different definitions of personal data co-existed. Personal data means data that can be linked to a person, i.e. the term refers to protected data where the data source (person) is identifiable. The more recent guidance refers to the need of “considerable amount of time and manpower” as a criterion of defining personal “identifiability”. That latter definition is found in Recommendation R(97)5 of the Committee of Ministers of the Council of Europe on Medical Data [26]. A different definition is found in the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. That Directive [27] provides guidance on the protection of individuals with regard to the processing of personal data and on the free movement of such data, but it is not targeted specifically towards genetic databases [26, 28]. The Directive states that data on individuals are personal data, if a decoding key exists for the coded data. It is of no importance whether this identification would require considerable time and manpower [27, 29, 30]. The working party of the relevant article 29 of the Directive 95/46/EC of the European Parliament tends to achieve some harmonization between the two conflicting guidance documents and aims at pragmatic solutions to define non-identifiable data based on current realistic and available information technology [31, 32]. But despite the many years of ongoing discussion, the advance seems very limited, except for the recognition that although there are new technical protective means, the definition will need to remain pragmatic.

Different ethically acceptable solutions: minimal or maximal ethics?

From an ethical point of view, there have been some remarkable developments. Radiologists and geneticists increasingly seem to insist on informing research participants about life saving and otherwise important medical results [33, 34]. As a result current thinking trends tend to recognize the return of (clinically relevant) results as a right of research participants and a duty of researchers [35]. That does not mean, how-

ever, that results should be returned at any cost. A way has to be found to balance data protection with patient protection and this is not always an easy task. The ethical obligation to prevent voluntarily abuse (minimal ethics), in fact, does not yet define the appropriate amount of resources and manpower that must be invested to *actively* protect research participants (maximum ethics) [36]. The search for an ethically acceptable solution to this problem is further complicated by the existence of conflicting interests of different stakeholders with regard to data protection. Some researchers, for instance, might be reluctant to recontact the participants since it is very time consuming and costly and thus they might prefer irreversible anonymization. Others researchers might want to have access to further information and thus prefer reversible anonymization, while research subjects do not want to be recontacted.

At present, depending on the biobank research project, several ethically acceptable solutions to the issue of data protection have been proposed. Irreversible anonymization is rarely sufficiently justified and should remain the exception as this type of data protection excludes the return results to participants. Further, while it does increase confidentiality, ultimately the risk of identification through data base matching and DNA fingerprinting can never be fully eliminated. However, irreversible anonymization remains an ethically acceptable solution in the case of biobank research studies where it is highly unlikely that clinically relevant results will ever be generated. The use of reversibly anonymized and coded data, allows the return of results and could thus lead to improved patient protection. As such, this type of data protection entails a positive action which is consistent with the principle of beneficence. In this case the “violation” of the patient’s autonomy rights (not to know) might be justified if its aim is to make a greater good (health improvement or lifesaving). Respect for research participants’ autonomy in fact does not necessarily mean that they should have the right to choose from a “menu” of different options concerning data protection. Research ethics is the domain where medical paternalism is still authorized [37]. That means that research ethics committees and not research participants decide at present which is the most appropriate balance between risks and benefits, including the field of data protection. However, the use of reversible anonymization can be very costly, especially if – as legally required in Switzerland – information about genetic results requires the availability of genetic counselling [38]. Further, given the federal system of research ethics committees in Switzerland, it is possible, that some cantons will opt for minimal ethics, while others prefer more costly solutions in line with more maximal ethics and this in turn might raise issues of justice and inequality among research participants across Switzerland. Therefore it is of utmost importance that cantonal research ethics committees keep up

to date with recent developments concerning international medical ethics, and consider appropriately the more recent recognition of ethical duties and rights in the field of return of research results.

Conclusions

The ethical and legal issues of data protection that are raised by collaborative projects involving research biobanks need to be addressed appropriately by researchers, biobank personnel and research ethics committees. Although there is no solution that will maximize the interests of all stakeholders, a variety of distinct practically possible and ethically acceptable solutions to data protection exist. As data protection is not always the same as participant protection, research ethics committees have to discuss carefully each individual research project involving biobanks. The autonomy rights of researchers also need to be protected and could outweigh autonomy rights of participants, especially if the latter conflict with their professional duties to prevent harm and promote health. Such a situation presents itself whenever researchers and ethics committees rely on their right and duty to restrict available choices of data protection by using reversible anonymization against the participants' wishes in order to enable routinely the return of clinically relevant results.

Ethically appropriate data protection remains a crucial and complex goal. While many ways can lead to Rome, each biobank project has to explain and justify carefully the chosen level of anonymity and to obtain appropriate participant consent for the involved specific risks and benefits. Further, a minimal standard of data protection should exist in order to avoid the use of identified samples or data, and research ethics committee members should be carefully trained to enable them to balance the risks and benefits of data protection levels in each individual biobank project.

Zusammenfassung

Kollaborative Forschungsprojekte mit Biobanken – ethische und rechtliche Aspekte des Datenschutzes

Kollaborative Forschungsprojekte mit Biobanken werfen verschiedene ethische und rechtliche Fragen auf. Denn ungenügender Datenschutz und/oder unangemessenes Datenmanagement können den Spendern von biologischem Material schaden. Das Konzept des Datenschutzes leitet sich aus dem Prinzip des *nihil nocere* und aus dem Respekt vor der Autonomie der Spender des biologischen Materials ab. Das Konzept des Patienten- (oder Spender)schutzes basiert dagegen auf dem Prinzip des Wohltuns und kann mit Autonomierechten in Konflikt stehen. In diesem Artikel untersuchen wir die oft schwierige Beziehung zwischen Datenschutz und

Wohlergehen von Patienten (oder Spendern). Dazu nehmen wir zunächst die verschiedenen Niveaus von Anonymität und Datenschutz unter die Lupe, benennen ihre Vor- und Nachteile und analysieren dann die ethische Akzeptanz verschiedener Lösungen der aufgeworfenen Probleme. Unsere Hauptziele sind, das Bewusstsein für ethische Fragen, die sich im Zusammenhang mit Datenschutz und Biobanken stellen, zu schärfen, Wissen über die verschiedenen Niveaus von Anonymität zu vermitteln, den sorgfältigen und angemessenen Gebrauch verschiedener Terminologien zur Anonymität anzuregen und anzumahnen, dass mögliche Interessenkonflikte korrekt identifiziert und benannt werden.

Résumé

Projet de recherche en collaboration avec des biobanques – Aspects éthiques et juridiques de la protection des données

Les projets de recherche collaborative impliquant les biobanques soulèvent des questions éthiques et juridiques telles que la protection insuffisante des données et/ou l'usage inapproprié des données qui peuvent avoir des effets nocifs pour les donateurs de l'échantillon. L'idée de la protection des données se base sur le principe de non-malfaisance et sur le respect de l'autonomie des donneurs. L'idée de la protection des patients (ou donneurs), en revanche, est basée sur le concept de bienfaisance et peut être en conflit avec les droits à l'autonomie. Dans le présent article, nous nous concentrons sur cette relation souvent tendue entre la protection des données et le bien-être du patient (donneur). A cette fin, nous examinons d'abord les différents niveaux et procédures d'anonymisation et de la protection des données. Ensuite, nous identifions les avantages et les inconvénients et nous examinons l'acceptabilité éthique des différentes solutions possibles à ce problème. Notre objectif principal est de sensibiliser sur les questions éthiques liées à la protection des données dans le cas des biobanques, y compris les connaissances sur les différents niveaux d'anonymisation, l'utilisation prudente de la terminologie appropriée, et l'identification et la reconnaissance des intérêts différents et potentiellement contradictoires.

Correspondence

Bernice S. Elger
Centre universitaire romand de médecine légale
Centre médical universitaire
1, rue Michel Servet
CH-1211 Genève 4

E-mail: Bernice.Elger[at]unige.ch

References

1. Battay TW, Valant V, Kassis SB, et al. Recommendations from the International Stroke Genetics Consortium, part 2: biological sample collection and storage. *Stroke* 2015;46(1):285–290.
2. Vaithilingam RD, Safii SH, Baharuddin NA, et al. Establishing and managing a periodontal biobank for research: the sharing of experience. *Oral Diseases* 2015;21(1):e62–69.

3. van Leeuwen EM, Karssen LC, Deelen J, et al. Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. *Nature communications* 2015;6: 6065.
4. Vaught J. Biobanking in china. *Biopreservation and biobanking* 2015;13(1):1.
5. Wild CP, Bucher JR, de Jong BW, et al. Translational cancer research: balancing prevention and treatment to combat cancer globally. *Journal of the National Cancer Institute* 2015;107(1):353.
6. Brisson AR, Matsui D, Rieder MJ, et al. Translational research in pediatrics: tissue sampling and biobanking. *Pediatrics* 2012;129(1): 153–162.
7. Betsou F, Gunter E, Clements J, et al. Identification of evidence-based biospecimen quality-control tools: a report of the International Society for Biological and Environmental Repositories (ISBER) Biospecimen Science Working Group. *The Journal of molecular diagnostics*: JMD 2013;15(1):3–16.
8. Botling J, Edlund K, Segersten U, et al. Impact of thawing on RNA integrity and gene expression analysis in fresh frozen tissue. *Diagn Mol Pathol* 2009;18(1):44–52.
9. Cervo S, De Paoli P, Perin T, et al. Cost-effective organization of an institutional human cancer biobank in a clinical setting: CRO-Biobank experience toward harmonization. *The International journal of biological markers* 2015:0.
10. Knoppers BM, Saginur M. The Babel of genetic data terminology. *Nat Biotechnol* 2005;23(8):925–927.
11. Elger BS, Caplan AL. Consent and anonymization in research involving biobanks: differing terms and norms present serious barriers to an international framework. *EMBO Rep* 2006;7(7):661–666.
12. COE. Council of Europe. Rec(2006)4 of the Committee of Ministers to member states on research on biological materials of human origin: <https://wcd.coe.int/ViewDoc.jsp?id=977859> (accessed July 30, 2009).
13. Swiss federal law. Law About Human Subject Research (Loi fédérale relative à la recherche sur l'être humain): www.admin.ch/opc/fr/classified-compilation/20061313/index.html (accessed September 5, 2014).
14. OHRP. Guidance on Research Involving Coded Private Information or Biological Specimens. (Rockville, MD, USA: Office for Human Research Protections). 2004.
15. Elger B. Kritische Analyse der Reglementierung von Biobanken im Vorentwurf des Bundesgesetzes über die Forschung am Menschen. *Bioethica Forum* 2006;49:42–45.
16. Elger B. Ethical Issues of Human Genetic Databases. A Challenge to Classical Health Research Ethics? London/New-York: Ashgate; 2010.
17. Elger B, Biller-Andorno N, Mauron A, et al. (eds). *Ethical Issues in Governing Biobanks: Global Perspectives*. London: Ashgate; 2008.
18. Elger B, Lavindrasana J, Iacono LL, et al. Health Data Depersonalisation for Prospective Research in the Life Sciences. *Computer Methods and Programs in Biomedicine* 2009;submitted.
19. Nietfeld JJ, Litton JE. Bio-PIN could cut biobanking risks. *Nature* 2011;472(7344):418.
20. Nietfeld JJ, Sugarman J, Litton JE. The Bio-PIN: a concept to improve biobanking. *Nature reviews. Cancer* 2011;11(4):303–308.
21. Wikina SB. What caused the breach? An examination of use of information technology and health data breaches. *Perspectives in health information management / AHIMA, American Health Information Management Association* 2014;11:1h.
22. Crowcroft J. On the duality of resilience and privacy. *Proceedings. Mathematical, physical, and engineering sciences / the Royal Society* 2015;471(2175):20140862.
23. Elger B. Anonymization and coding. In: Elger B, Biller-Andorno N, Mauron A, et al. (eds). *Ethical Issues in Governing Biobanks: Global Perspectives*. London: Ashgate;2008:167–88.
24. Border P. Genomics: Sequence sharing: www.nature.com/nature/journal/v470/n7333/full/470169a.html (accessed March 10, 2015). *Nature* 2011;470:169–170.
25. Gaskell G, Gottweis H. Biobanks need publicity. *Nature* 2011;471(7337):159–60.
26. COE. 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: www1.umn.edu/humanrts/instree/coerecr97-5.html (accessed July 30, 2009).
27. European Parliament and the Council. 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: www.dataprotection.ie/viewdoc.aspx?DocID=89 (accessed July 30, 2009) or <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML> (accessed March 3, 2015).
28. Elger BS. *Ethical Issues of Human Genetic Databases. A Challenge to Classical Health Research Ethics?* London: Ashgate;2010.
29. Data Protection Commission. 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: www.mannvernd.is/english/news/Data_Protection_Commission_040998.html (accessed January 2004).
30. Arnason E. Personal identifiability in the Icelandic Health Sector Database. *Journal of Information, Law & Technology* 2002: www2.warwick.ac.uk/fac/soc/law/elj/jilt/2002_2/arnason/ (accessed July 30, 2009).
31. European Commission. Reform of the data protection legal framework: http://ec.europa.eu/justice/data-protection/review/index_en.htm (accessed March 3, 2015). Secondary Reform of the data protection legal framework: http://ec.europa.eu/justice/data-protection/review/index_en.htm (accessed March 3, 2015).
32. European Commission. Article 29 Working Party (Directive 95/46/EC): http://ec.europa.eu/justice/data-protection/article-29/index_en.htm (accessed March 3, 2015). Secondary Article 29 Working Party (Directive 95/46/EC): http://ec.europa.eu/justice/data-protection/article-29/index_en.htm (accessed March 3, 2015).
33. Brown SD. Professional norms regarding how radiologists handle incidental findings. *Journal of the American College of Radiology: JACR* 2013;10(4):253–257.
34. Pullman D, Hodgkinson K. Genetic knowledge and moral responsibility: ambiguity at the interface of genetic research and clinical practice. *Clin Genet* 2006;69(3):199–203.
35. Knoppers BM, Joly Y, Simard J, et al. The emergence of an ethical duty to disclose genetic research results: international perspectives. *Eur J Hum Genet* 2006;14(11):1170–1178.
36. Ogien R. Qu'est-ce que l'Étique minimale? *Revue de Théologie et de Philosophie* 2008;RThPh 2008/II-III:107–118.
37. Elger B. Le Paternalisme médical: mythe ou réalité? *Aspects philosophiques et empiriques d'un phénomène persistant*. Geneva: Editions Médecine&Hygiène; 2010.
38. Swiss Federal Law. Law About Genetic Analysis of Human Beings (Loi fédérale sur l'analyse génétique humaine): www.admin.ch/ch/f/rs/810_12/index.html (accessed July 12, 2009).