

SPECIAL ARTICLE

Locus-Specific Databases: From Ethical Principles to Practice

Richard G. H. Cotton,^{1,2*} Clémentine Sallée,³ and Bartha M. Knoppers³¹Genomic Disorders Research Centre, Melbourne, Australia; ²Department of Medicine, University of Melbourne, Australia;³Centre de Recherche en Droit Public Université de Montréal, Montréal, Quebec, Canada

Communicated by Alastair Brown

Locus-specific databases (LSDBs) play an essential role in clinical care and research. They differ from traditional genetic databases in that they propose to place the mutations of “anonymized” patients directly on the World Wide Web. The proliferation of ethical guidelines and legal requirements affects the rapid and free transmission of clinical data, which is vital for both the daily management of patients and research into better diagnostics and treatment. This paper proposes a review of ethical principles endorsed by international instruments that are of particular relevance to LSDBs. It aims to translate them into 12 proposed practical guidelines that LSDB curators can use in collecting data for clinical research. Perhaps these guideposts will serve as a first step toward translating principles into practice. *Hum Mutat* 26(5), 489–493, 2005. © 2005 Wiley-Liss, Inc.

KEY WORDS: mutation; database; ethics; guidelines

INTRODUCTION

At first glance it may appear that placing the mutation of a patient on a World Wide Web site is not much different from publishing the same mutation in the conventional scientific literature and then posting it on the Web. But is placing a mutation found in a sample taken for diagnostic purposes directly on the Web really similar to the conventional publication route? Perhaps, as we attempt to demonstrate here, the ethical principles governing locus-specific databases (LSDBs) are not unique. However, the need to distinguish such data from other genetic data requires refinement of the interpretation of such principles.

To assess the guidance that the then-current ethical statements gave to curators of LSDBs, a study was conducted in 2000 [Knoppers and Laberge, 2000] (see Table 1). That study examined several statements, such as the *Statement on the Principled Conduct of Genetic Research* [HUGO, 1996] (www.gene.ucl.ac.uk/hugo/conduct.htm), *Statement on DNA Sampling: Control and Access* [HUGO, 1999] (www.hugo-international.org/Statement_on_DNA_Sampling.htm), *Universal Declaration on the Human Genome and Human Rights* [UNESCO, 1997] (http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html), and the *Convention for the Protection of Human Rights and Dignity of the Human Being With Regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine* [Council of Europe, 1997] (<http://conventions.coe.int/treaty/en/treaties/html/164.htm>), as the possible basis of a framework for an allelic variation database.

LSDBs are collections of published and unpublished allelic variations and information. Established and used by both clinicians and researchers, these databases can be individually or centrally curated. Traditionally they have been kept in the private files of scientists, laboratories, and clinicians, but increasingly the mutations they contain are being placed on the Web. Deploing the absence of systematic and organized methods of collection and

official publication of molecular variations or mutations, Professor R. Cotton instigated discussions that in 1996 resulted in the creation of the Mutation Database Initiative (MDI). It was sponsored and coordinated by the Human Genome Organization (HUGO) and supported by both the American Society of Human Genetics and the March of Dimes. This consortium of database curators was composed of approximately 600 members from 34 countries. The initiative led to the creation of the Human Genome Variation Society and the adoption of a society journal (*Human Mutation*, www.hgvs.org). Members of the MDI are also kept informed by a mailing list. The project aims to collect up-to-date molecular variations in a centralized database and publish them in *Human Mutation*. These mutations will thus be available on the Web to all researchers and the public.

A distinctive feature of LSDBs, in comparison with traditional human genetic research databases, is that they are research infrastructures—tools to be used by scientists or clinicians undertaking studies in functional genomics. As a result, the benefit is not directly gained by individual participants, but more generally in the long term by the population as a whole. Furthermore, these LSDBs are to be posted on the Web in order to ensure proper documentation on mutations, and avoid losses, misnomers, or overlap. However, putting such publications in the public domain and providing access to authorized individuals that are not bound by the clinician/researcher–patient/research participant confidential relationship raise specific ethical issues.

Received 8 March 2004; accepted revised manuscript 22 June 2005.

*Correspondence to: Richard G.H. Cotton, Genomic Disorders Research Centre, Level 2, 161 Barry St., Carlton South, VIC 3053, Australia. E-mail: cotton@unimelb.edu.au

DOI 10.1002/humu.20245

Published online 19 September 2005 in Wiley InterScience (www.interscience.wiley.com).

TABLE 1. Practical Guidelines

- 1) Develop a common ethical framework.
- 2) Define ethical principles and scientific approach.
- 3) Clarify purpose and limits of LSDB (Locus Specific Database).
- 4) Create an ethics review board.
- 5) Foster international standardization.
- 6) Anonymize prior to transmission.
- 7) Protect confidentiality.
- 8) Inform donors of transmission to LSDB.
- 9) No disclosure without consent.
- 10) Take vulnerable persons into account.
- 11) Take specific communities/cultures into account.
- 12) No virtual “medical advice” relationship.

The purpose of this paper is to examine the normative developments that have occurred since the first paper was published in 2000, and to derive a set of practical guidelines therefrom.

I. NORMATIVE DEVELOPMENTS

Regulation of LSDBs

The last 5 years have seen the adoption of specific international instruments that govern genetic [UNESCO, 2003] (http://portal.unesco.org/shs/en/file_download.php/199979/a35f1645326a6511fd0dca29ccDeclaration_en.pdf), WHO [2003] (www.law.ed.ac.uk/ahrb/publications/online/whofinalreport.rtf); genomic [HUGO, 2002] (www.hugo-international.org/Statement_on_Human_Genomic_Databases.htm); or health [WMA, 2002a,b] (www.wma.net/e/policy/d1.htm) databases. They address the ethical, legal, and social issues with regard to genetic applications in health research and activities in general, and of genetic testing in particular [European Commission, 2004] (http://europa.eu.int/comm/research/conferences/2004/genetic/pdf/recommendations_en.pdf), or refine the ethical principles pertaining to biomedical research [CIOMS, 2002] (www.cioms.ch/frame_guidelines_nov_2002.htm), [WMA, 2002b] (www.wma.net/e/policy/b3.htm). While they do not establish a clear distinction between LSDBs and classic genetic databases, these normative documents can potentially affect LSDBs because genetic information sent to an LSDB could be collected in the course of research or during clinical practice. Although they are not legally binding, these instruments can urge states, national medical associations, and clinician-researchers to promote and adopt the ethical principles they prescribe in order to frame research, medical practice, and data management, and to foster international cooperation. The HUGO [2002], UNESCO [2003], WHO [2003], and European Commission [2004] have provided specific guidance for the collection and management of genetic data and genomic databases.

Setting the stage for our discussion and the guidelines we propose, the HUGO *Statement on Human Genomic Databases* [HUGO, 2002] maintained that genomic databases are global public goods for the benefit of humanity. The *International Declaration on Human Genetic Data* [UNESCO, 2003] aimed to “ensure the respect of human dignity and protection of human rights and fundamental freedoms in the collection, processing, use and storage of human genetic data, human proteomic data and of the biological samples from which they are derived “...”. The WHO report entitled *Genetic Databases—Assessing the Benefits and Impact on Human Rights and Patient Rights* [WHO, 2003] proposed international guidelines “to govern the establishment,

uses, and maintenance of these databases, and access thereto” and warned us “not to place too much emphasis on genetics in discussing ‘the human condition’” because “we are, each of us, more than the sum of our genetic component parts.” Similarly, the European Commission’s *25 Recommendations on the Ethical, Legal and Social Implications of Genetic Testing* [European Commission, 2004] proposed that “genetic exceptionalism” should be avoided, internationally, in the context of the EU and at the level of its member states since “all medical data, including genetic data” should “satisfy equally high standards of quality and confidentiality.”

While the various international documents considered above provide us with a multitude of rules, there are certain commonalities in the values they endorse. We selected a few principles that are of particular relevance for LSDBs, namely scientific and ethical responsibility and competence, confidentiality, ethical oversight, and protection of vulnerable individuals or disadvantaged individuals or groups. A review of these general principles in the first part of this paper will serve as the background for their translation into proposed practice guidelines as discussed in the second part.

A. Responsibility and Competence

In order for an LSDB to be scientifically valid and ethically legitimate, its curators must demonstrate responsibility and competence. Curators of LSDBs are responsible for data quality and accuracy, and their appointment should be based on competence, training, and skill. They must ensure not only that their data are up-to-date and validated, but also that they were obtained according to predetermined ethical and scientific standards. Indeed, “there is a scientific responsibility to ensure the professional competence of researchers working with data, as well as the quality and accuracy of the data” [HUGO, 2002]. Competence encompasses the ability to “understand the data” as well as “the context in which the information comes, and the relevance of other factors which further contextualize the information” [WHO, 2003]. Furthermore, “the persons and entities responsible for the processing of human genetic data, human proteomic data and biological samples should take the necessary measures to ensure the accuracy, reliability, quality, and security of these data and the processing of biological samples (...)” [UNESCO, 2003].

B. Confidentiality and Security of Data

In the field of genetics, particularly as regards genetic databases and data transfer, the issues of confidentiality and security are primary. Confidentiality has various applications. The data to be transmitted to LSDBs should not only be obtained with consent, but their confidential transfer to an LSDB should also be disclosed to the donor.

The extent of (comprehensible) information to be provided to patients, research participants, or donors when data are collected has increased over time. It now covers inter alia the nature of the research; foreseen as well as foreseeable uses, risks, and benefits; the security systems contemplated; foreseen and foreseeable users; and data-sharing. Moreover, this informed consent is to be obtained “based on adequate process and procedure (...) regardless of the purpose of the collection and the level of anonymity” (Paragraph 23) [European Commission, 2004]. Even when consent has been obtained, the privacy choices made by individuals must be respected to avoid violating the relationship of

trust that has been established between the individual and the clinician/researcher/investigator.

Furthermore, a donor should a fortiori be recontacted or at least informed (if so desired) should data collected for clinical or research purposes be anonymized for transmission to a LSDBs for secondary uses that were unforeseen at the time of collection. Ideally, consent for such future anonymized transmission (subject to ethical oversight) should be obtained when the sample is collected. If not, recontact and recontact may be necessary, although advance notification on the consent form is preferable. Indeed, as recommended by the WHO, “when obtaining informed consent to the provision of a sample or information for a genetic database, participants should be informed “...” of the uses to which data might be put, including potential use “...” of the possibility of future uses of data, beyond the limits of the present consent, and should be provided with an opportunity to withhold consent to such uses” (Recommendation 6). The World Health Organization foresees the possibility of obtaining blanket consent for future research “in circumstances where anonymity of future data can be guaranteed” [WHO, 2003]. Finally, it should be mentioned, as outlined by the European Commission Expert Group, that “[c]onsent issues for samples from individuals and groups entering new collections may differ from those applicable to existing collections [or when data are primarily collected for care purposes], where contact may not have been maintained with the source individuals over the years” (Paragraph 23) [European Commission, 2004].

Although limited exceptions to confidentiality have been contemplated by the above-mentioned international instruments (e.g., for public health purposes or specific research needs), data collected for research purposes and entered into LSDBs should be kept confidential. “Human genetic data, human proteomic data and biological samples collected for the purpose of scientific research should not normally be linked to an identifiable person. Even when such data or biological samples are unlinked to an identifiable person, the necessary precautions should be taken to ensure the security of the data or biological samples” [UNESCO, 2003]. Furthermore, as stated by HUGO, “participants should be informed of the degree of identifiability of their data (e.g., coded, anonymized, aggregate, etc.)”, and “of the security mechanisms in place to ensure confidentiality” [HUGO, 2002], as also mentioned by the WHO (Recommendation 6.5) [WHO, 2003].

It bears noting that the HUGO, WHO, UNESCO, and European Commission instruments specific to genetic data and databases use a different taxonomy to describe confidentiality mechanisms. UNESCO uses the term “unlinked to an identifiable person,” WHO speaks of absolute, proportional and reasonable anonymity; HUGO refers to coding, anonymization, or aggregation; and the European Commission refers to coding, double-coding, anonymized, and anonymous data or samples. Unless some common understanding is achieved, this lack of harmonization will impede international collaboration (see above).

C. Ethical Supervision

The *Statement on Genomic Databases* [HUGO, 2002] (referring to its 1996 *Statement on the Principled Conduct of Genetic Research* [HUGO, 1996]) recommended the “continual review, monitoring and oversight” of research projects. Similarly, the *International Declaration on Genetic Data* [UNESCO, 2003] mandated that ethical supervision should occur at international, national, local, and institutional levels, and the European Commission urged “Member States to ensure that approval by a competent review

committee [be] obtained before research is undertaken” (Paragraph 20) [European Commission, 2004]. More specifically, WHO recommended that “appropriate ethical approval mechanisms be established to oversee the creation and maintenance of genetic databases” (Recommendation 4) [WHO, 2003].

According to the principles upheld by UNESCO, each research protocol should be reviewed by an independent, multidisciplinary ethics committee. Such a committee should be established pursuant to article 16 of the 1997 *Universal Declaration on the Human Genome and Human Rights* (Article 6b) [UNESCO, 1997]. National ethics committees must also adopt ethical standards and intervene in the absence of legislation (Article 6b). State ethics committees should collaborate when projects involve more than one country (Article 6d). WHO also provided further guidance on the duties of an ethics review committee. The composition of such a committee will necessarily vary depending on whether the genetic data exist in the context of pure research, clinical research, or clinical practice (Recommendation 4) [WHO, 2003].

Currently, no international ethics committee is responsible for the scientific and ethical oversight of international projects or databases. HUGO's ethics committee neither reviews protocols nor monitors compliance.

D. Taking Into Account Vulnerable Persons As Well As Communities and Cultures

In conformity with the Nuremberg Code [1949], international guidelines or statements that have been adopted since 2000 reflect a protective attitude toward disadvantaged individuals or communities. As recommended by the WHO, “research using samples and information taken from vulnerable subjects, such as incapacitated adults or children, must be carried out in full conformity with internationally agreed principles and guidelines.” (Recommendation 11) [WHO, 2003]. Such research answers to more stringent conditions in terms of justification, ethical clearance, and consent because it “must be shown to hold the reasonable prospect of benefiting the class of persons to which the particular subject belongs, either in the immediate or foreseeable future” (Recommendation 11) [WHO, 2003] or, more generally, to serve the subject's interests (Recommendation 25a) [European Commission, 2004]. In particular, researchers should be sensitive to the possibility of discrimination and stigmatization from the results of genetic research and databases. Indeed, according to the 2002 *HUGO Statement on Human Genomic Databases*, “[i]ndividuals, families and communities should be protected from discrimination and stigmatization” [HUGO, 2002], and the European Commission [2004] stated that “healthcare policy making [should] be sensitive to the risks of stereotyping and stigmatization based on ethnic origin, and recognize and respect ethnic and cultural sensitivities” (Recommendation 12). Moreover, the selection of a particular group for research must be justified. An equitable distribution of burdens and benefits must be shown for both exclusion or inclusion.

II. PRACTICAL GUIDELINES

Taking then the “ethical guideposts” of both the original paper (Table 1) and the recent international normative principles just discussed, we posit the following 12 practical guidelines:

Develop a Common Ethical Framework

From the outset, it is important for mutation databases to develop a common set of overarching principles. This is especially important for international databases so that contributing national

or disease-specific efforts can follow such guidelines to ensure compatibility, as well as further collaboration and harmonization.

Define Ethical Principles and Scientific Approach

Defining the ethical principles found within the common framework does not entail abstract or static definitions of the meaning of terms such as “consent” or “confidentiality.” Rather, the approach or process to be followed by participating scientists should be delineated in practical terms. This allows for the actual realization of, for example, consent and confidentiality, in the day-to-day activities of a database.

Clarify the Purpose and Limits of the LSDB

Much depends on the clarification of the mission, objectives, and limits of an LSDB. Considering the possible confusion between such databases and the classic clinical genetic databases that are often the subject of the normative instruments described above, the practical guidance developed will depend on an accurate statement of the goals, and especially the limits, of the database.

Create an Ethics Review Board

The presence of an ethics review board and its multidisciplinary membership ensures accountability through the review of submissions, ongoing monitoring, and transparency. At a minimum, international databases should create such boards, while at the local level databases should be subject to institutional review and oversight.

Foster International Standardization

a) Even more problematic (if not fatal) to scientific collaboration than the current myriad of ethical rules is the lack of standardization. Since identified data are extremely rare in the research setting, we suggest that instead of the plethora of complex terms referred to in existing texts, two categories of data should be used: 1) coded data (or samples), further divided into single- and double-coded; and 2) anonymized data (or samples, i.e., unidentifiable).

However, this is more than just an issue of nomenclature. It goes to the very root of compatibility and exchange among researchers and among databases. As much as possible within and among databases, a set of agreed-upon standards should be adopted. Participating LSDBs should be strongly encouraged to adhere to these standards.

Anonymize Prior to Submission

All data in the LSDBs should be anonymized by the curator prior to submission. This means that enough identifiers should be removed to avoid the possible identifiability of the sample. The qualitative and quantitative nature of identifiers to be removed should be deemed sufficient as long as it makes it practically impossible for an individual using reasonable manpower and time to identify the sample source. For instance, a nonrandom code based on or derived from the sample source would not be sufficient to secure anonymization. The U.S. Health Insurance Portability and Accountability Act [HIPAA, 1996] (<http://aspe.hhs.gov/admsimp/pl104191.htm>) holds that in order for health information to be considered deidentified (i.e., anonymized), a minimum of 18 identifiers should be removed. These identifiers range from the individual's name to his/her date of hospital discharge or postal code's first three digits.

Protect Confidentiality

The finding of rare mutations together with the regional location of the submitting author could indirectly serve to identify a patient. Such rare occurrences are difficult to avoid. The fact of data transmission should not be recorded in the medical record so as to protect personal privacy and the confidentiality of the physician–patient relationship.

Inform Donors of Transmission to LSDBs

At the time consent is given, mention should be made that anonymized data on location-specific mutations are being sent to a central database. Such notification allows for queries or refusal to participate.

No Disclosure Without Consent

While LSDBs contain only anonymized data, clinician-researchers who submit data may be asked for further information by others. Unless the purpose is immediate treatment, no personal identification should be provided without consent. If the research project involves collaboration with other researchers and the exchange of data prior to anonymization, participants should provide specific consent to such exchange. In any event, only coded data should be exchanged.

Take Vulnerable Persons Into Account

Persons who do not have the capacity to consent can be vulnerable. The inclusion of data from such individuals should be subject to more stringent review.

Take Specific Communities/Cultures Into Account

If identifiable groups are involved, prior engagement with such subpopulations is a prerequisite to obtaining individual consent. Care should be taken in interpreting such data because there may be a difference of worldviews.

No Virtual “Medical Advice” Relationship

For counseling purposes, FAQs on LSDBs are acceptable provided they are within the framework of the ethical principles governing clinicians. However, LSDBs should not offer interactive clinical advice.

CONCLUSIONS

Our overview of normative developments reveals an emerging consensus on the ethical principles that guide genetic databases. There is also an obvious need for further international harmonization in the areas of genetic research and genetic database management to ensure transparency and accountability. While the *Statement on Human Genomic Databases* [HUGO, 2002] mentioned that “compatibility should be fostered through the use of common nomenclature ‘...’ (Recommendation 3c), such standardization should not be limited to the taxonomy used, and should also include the principles. As mentioned in the *Report of the Secretary-General on Information and Comments Received From Governments and Relevant International Organizations and Functional Commissions Pursuant to Council Resolution 2001/39* [ECOSOC, 2003], “despite the existence of numerous declarations, guiding principles and codes dealing with the issue of genetic data, the changing conditions of genetic research call for the establishment of an international instrument that would enable States to agree on ethical principles, which they would then have to transpose into their legislation.”

Recognizing that the justification for a database is more likely to be grounded in communal value rather than direct individual benefit, the 2003 WHO report on *Genetic Databases—Assessing the Benefits and Impact on Human Rights and Patient Rights* [WHO, 2003] adopted the following fundamental guiding values:

1. The pursuit of human well-being.
2. The quality of human dignity, including fundamental human rights and the principle of nondiscrimination.
3. The principle of respect for persons, including the imperatives of beneficence and nonmaleficence.
4. The principle of respect for individual autonomy.

Our adaptation (text in brackets) of the introductory comments of the Council for International Organizations of Medical Sciences (CIOMS) in its *International Ethical Guidelines for Biomedical Research Involving Human Subjects* [CIOMS, 2002] maintains that:

The challenge to international research ethics [and international flow of data] is to apply universal ethical principles [as well as common standards, measures, and taxonomy] to biomedical research [in genetics and data management] in a multicultural world with a multiplicity of health care [and legal] systems and considerable variation on standard of health [and of living].

LSDBs play an essential role in clinical care and research. The proliferation of ethical guideposts and requirements endangers the rapid and free transmission of clinical data, which is vital for both the daily management of patients and research into better diagnostics and treatment. Lack of knowledge regarding sound, approved guidelines can inhibit the entry of data into LSDBs and thereby deny patients of strategies derived from the experience of others worldwide. In that light, we propose 12 practical guidelines that LSDB curators can embrace when collecting data for use in clinical research. These guidelines are a first step toward translating principles into practice.

REFERENCES

- CIOMS. 2002. Council for International Organizations of Medical Sciences. International ethical guidelines for biomedical research involving human subjects; Geneva, www.cioms.ch/frame_guidelines_nov_2002.htm.
- Council of Europe. 1997. Convention for the protection of human rights and dignity of the human being with regard to the application of biology and medicine: convention on human rights and biomedicine; Oviedo. *Int Dig Health* Leg 48, 99, <http://conventions.coe.int/treaty/en/treaties/html/164.htm>.
- ECOSOC. 2003. United Nations Economic and Social Council. Report of the Secretary-General on information and comments received from governments and relevant international organizations and functional commissions pursuant to Council resolution 2001/39; Geneva. www.un.org/esa/coordination/ecosoc/Genetic_privacy.pdf.
- European Commission. 2004. Directorate-General for Research, European Commission, 25 recommendations on the ethical, legal and social implications of genetic testing, Brussels, http://europa.eu.int/comm/research/conferences/2004/genetic/pdf/recommendations_en.pdf.
- HIPAA. 1996. Health Insurance Portability and Accountability Act; Washington, <http://aspe.hhs.gov/admsimp/pl104191.htm>.
- HUGO. 1996. Human Genome Organization statement on the principled conduct of genetic research, 3:2 Genome Digest 2, www.gene.ucl.ac.uk/hugo/conduct.htm.
- HUGO. 1999. Human Genome Organization statement on DNA sampling: control and access, 6:1 Genome Digest 8, www.hugo-international.org/Statement_on_DNA_Sampling.htm.
- HUGO. 2002. Human Genome Organization. Statement on human genomic databases; London, www.hugo-international.org/Statement_on_Human_Genomic_Databases.htm.
- Knoppers BM, Laberge CM. 2000. Ethical guideposts for allelic variation databases. *Hum Mutat* 15:30–35.
- Nuremberg Code. 1949. Trials of war criminals before the Nuremberg military tribunals under control council law, no. 10, vol. 2. Washington, DC: U.S. Government Printing Office. p 181–182.
- UNESCO. 1997. United Nations Educational, Scientific and Cultural Organization. International Bioethics Committee, universal declaration on the human genome and human rights, http://portal.unesco.org/en/ev.php-URL_ID=13177&URL_DO=DO_TOPIC&URL_SECTION=201.html.
- UNESCO. 2003. General conference, international declaration on human genetic data, Geneva. http://portal.unesco.org/shs/en/file_download.php/199979a35f1645326a6511fd0dca29cc/Declaration_en.pdf.
- WHO. 2003. World Health Organization. Genetic databases—assessing the benefits and the impact on human rights and patient rights; Geneva, www.law.ed.ac.uk/ahrb/publications/online/whofinalreport.rtf.
- WMA. 2002a. World Medical Association. Declaration on ethical considerations regarding health databases; Washington, www.wma.net/e/policy/d1.htm.
- WMA. 2002b. World Medical Association. Declaration of Helsinki, www.wma.net/e/policy/b3.htm.