



HAL
open science

Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs).

Sue Povey, Aida I Al Aqeel, Anne Cambon-Thomsen, Raymond Dalglish, Johan T. den Dunnen, Helen V Firth, Marc S Greenblatt, Carol Isaacson Barash, Michael Parker, George P. Patrinos, et al.

► To cite this version:

Sue Povey, Aida I Al Aqeel, Anne Cambon-Thomsen, Raymond Dalglish, Johan T. den Dunnen, et al.. Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs).. *Human Mutation*, 2010, 31 (11), pp.1179. 10.1002/humu.21339 . hal-00589432

HAL Id: hal-00589432

<https://hal.science/hal-00589432>

Submitted on 29 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs).

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0116.R1
Wiley - Manuscript type:	Special Article
Date Submitted by the Author:	13-Jul-2010
Complete List of Authors:	<p>Povey, Sue; University College London, Wolfson House 4 Stephenson Way NW1 2HE, Dept of Genetics Evolution and Environment</p> <p>AL Aqeel, Aida; Riyadh Military Hospital, Dept of paediatrics; King Faisal Specialist Hospital and Research Centre, Dept of Genetics and Stem Cell Therapy Program</p> <p>Cambon-Thomsen, Anne; Faculté de Medecine, 37 allées Jules Guesde,, UMR Inserm, Université Paul Sabatier - Toulouse III, U 558 Epidémiologie et analyses en santé publique : Département d'épidémiologie et de santé publique</p> <p>Dalgleish, Raymond; University of Leicester, University Road LE1 7RH, Dept of Genetics</p> <p>den Dunnen, Johan; Leiden University Medical Center, Human & Clinical Genetics</p> <p>Firth, Helen; Addenbrooke's Hospital NHS Trust, CB2 2QQ, Department of Medical Genetics</p> <p>Greenblatt, Marc; University of Vermont College of Medicine</p> <p>Isaacson Barash, Carol; 317 Lamartine Street, Suite 200, 8Genetics, Ethics & Policy Consulting</p> <p>Parker, Michael; University of Oxford, Badenoch Building, Old Road Campus, Headington,, Ethox Centre, Department of Public Health and Primary Health Care</p> <p>Patrinou, George; University of Patras, School of Health Sciences, Department of Pharmacy</p> <p>Savige, Judy; University of Melbourne, Department of Medicine</p> <p>Sobrido, Maria-Jesus; Fundacion Publica Galega de Medicina Xenomica; Institute of Health Carlos III, Center for Network Biomedical Research on Rare Diseases</p> <p>Winship, Ingrid; University of Melbourne, Royal Melbourne Hospital, 2nd floor, 300 Grattan Street, Parkville ,Victoria 3050, Clinical Genetics</p> <p>Cotton, Richard; The University of Melbourne, Royal Melbourne Hospital, Parkville VIC 3010, Faculty of Medicine, Dentistry and</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	Health Sciences; Howard Florey Institute, Carlton South VIC 3053, Genomic Disorders Research Centre
Key Words:	phenotype, LSDB, data-sharing, ethics, diagnostic test, consent, curators, guidelines



For Peer Review

1
2
3
4 **Practical guidelines addressing ethical issues pertaining to the curation of**
5
6
7 **human locus-specific variation databases (LSDBs).**
8
9

10
11 Sue Povey¹, Aida I. Al Aqeel², Anne Cambon-Thomsen³, Raymond Dalgleish⁴, Johan T. den
12 Dunnen⁵, Helen V. Firth⁶, Marc S Greenblatt⁷, Carol Isaacson Barash⁸, Michael Parker⁹,
13
14 George P. Patrinos¹⁰, Judith Savige¹¹, Maria-Jesus Sobrido^{12,13}, Ingrid Winship¹⁴ and Richard
15
16 G. H. Cotton^{15,16}
17
18
19
20
21
22
23

24 ¹Dept of Genetics, Evolution and Environment, University College London, Wolfson House,
25
26 4 Stephenson Way, London, NW1 2HE, UK
27

28 ²Dept of Paediatrics, Riyadh Military Hospital, Department of Genetics and Stem Cell
29
30 Therapy Program, King Faisal Specialist Hospital and Research Centre, Kingdom of Saudi
31
32 Arabia
33
34

35 ³UMR Inserm, Université Paul Sabatier - Toulouse III, U 558 Epidémiologie et analyses en
36
37 santé publique : Département d'épidémiologie et de santé publique Faculté de Medecine, 37
38
39 allées Jules Guesde, F-31073 Toulouse, France
40
41
42

43 ⁴Dept of Genetics, University of Leicester, University Road, Leicester, LE1 7RH, UK
44

45 ⁵Human and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands
46

47 ⁶Dept of Medical Genetics, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK
48
49

50 ⁷ University of Vermont College of Medicine, Burlington, Vermont, USA
51

52 ⁸Genetics, Ethics & Policy Consulting, 317 Lamartine Street, Suite 200, Boston, MA 02130,
53
54 USA
55
56

57 ⁹Ethox Centre, Department of Public Health and Primary Health Care, University of Oxford,
58
59 Badenoch Building, Old Road Campus, Headington, Oxford, OX3 7LF, UK
60

1
2
3
4 ¹⁰University of Patras, School of Health Sciences, Department of Pharmacy, Patras, Greece

5
6
7 ¹¹Dept of Medicine, University of Melbourne, Epping, Australia

8
9 ¹²Fundacion Publica Galega de Medicina Xenomica, Santiago de Compostela. Spain,

10
11 ¹³Center for Network Biomedical Research on Rare Diseases, Institute of Health Carlos III,
12
13 Madrid, Spain

14
15 ¹⁴. Dept of Clinical Genetics, University of Melbourne, Royal Melbourne Hospital, 2nd
16
17 floor, 300 Grattan Street, Parkville, Victoria 3050, Melbourne, Victoria, Australia

18
19
20 ¹⁵Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Royal
21
22 Melbourne Hospital, Parkville VIC 3010, Australia.

23
24
25 ¹⁶ Genomic Disorder Research Centre, Howard Florey Institute, Carlton South VIC 3053,
26
27
28 Australia

29
30
31
32
33
34
35 **Key words PHENOTYPE, LSDB, DATA-SHARING, ETHICS, CURATORS,**
36
37 **GUIDELINES, DIAGNOSTIC TEST, CONSENT**

Abstract

More than 1000 web-based locus-specific variation databases (LSDBs) are listed on the website of the Human Genetic Variation Society (HGVS). These individual efforts which often relate phenotype to genotype are a valuable source of information for clinicians, patients and their families as well as for basic research. The initiators of the Human Variome Project recently recognised that having access to some of the immense resources of unpublished information already present in diagnostic laboratories would provide critical data to help manage genetic disorders. However, there are significant ethical issues involved in sharing these data worldwide. An international working group presents second-generation guidelines addressing ethical issues relating to the curation of human LSDBs which provide information via a web-based interface. It is intended that these should help current and future curators and may also inform the future decisions of ethics committees and legislators. These guidelines have been reviewed by the Ethics Committee of the Human Genome Organisation (HUGO).

Introduction

This document is designed to assist a curator who intends to provide access to the information contained in a human Locus Specific variation Database (LSDB). For this purpose, an LSDB is defined as a listing of known sequence variants in a specific human gene together with some assessment of the effects of these variants on the phenotype. It may also highlight the frequency of both common and rare variants (*e.g.* single nucleotide polymorphisms) prevalent in particular populations groups. Although ethical issues arise in a database of any format, currently the access is nearly always provided via a web interface, usually available to everyone but occasionally restricted to selected professional groups. An example of a well-known LSDB which can be accessed by anyone is that describing the mutations in the gene DMD, deficient in Duchenne Muscular Dystrophy, curated by one of the authors (JdD) in Leiden (<http://www.dmd.nl/nmdb2/home.php>). This format is available in open-source software and is now used in many other LSDBs. The need for these guidelines has been highlighted by the recognition by initiators of the Human Variome Project (HVP) of the immense unpublished and inaccessible resource of information existing in diagnostic laboratories and the significant clinical need to have access to this information (Cotton et al., 2007; Kaput et al., 2009).

These guidelines are largely an expansion in detail of the first generation guidelines proposed by Cotton and co-authors in 2005 (Cotton et al., 2005) which were rooted in the principles described by Knoppers and Laberge in 2000 (Knoppers and Laberge, 2000). They were discussed and modified as a result of the international HVP planning meeting in Spain May 2008 attended by participants from a wide range of developed and emerging countries (Kaput et al., 2009). Details of this can be found in the published meeting report and its

1
2
3
4 supplementary information (Kaput et al., 2009). The content and order of headings in the
5
6
7 current version differs slightly from that of Cotton et al 2005 and is shown in **Box 1**.

8
9 **Box 1:12 major points to consider pertaining to ethical issues arising in the curation of**
10 **human Locus Specific variation Databases (LSDBs)**

- 11 1. Clarify the main purpose of the particular database,
- 12 2. Define database policy with respect to sources of data.
- 13 3. Take Specific communities/cultures into account.
- 14 4. Take vulnerable persons into account.
- 15 5. Create an ethics oversight committee
- 16 6. Remove identifying information before submission to the database
- 17 7. Add further protection of confidentiality if needed.
- 18 8. Allow no further disclosure without consent.
- 19 9. Provide provision for removal of data from the database.
- 20 10. Be cautious in response to requests to an LSDB curator for a private opinion
- 21 11. Limit links to other LSDBs
- 22 12. Consider carefully the transfer of publicly available data from LSDBs to genome
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
browsers

36
37 **Background: Develop a common ethical framework**

38
39 The goal of all such databases is the sharing of genomic and phenotypic information for the
40
41 benefit of humanity. This requires the protection of privacy, which in this context is the right
42
43 of the individual and members of their family to be protected against intrusion into their
44
45 personal information and further intrusions ensuing from access to this, by publication of
46
47 information. The balance between the public's interest in the value of the shared information
48
49 and its interest in the strict protection of privacy has been widely discussed,(see footnote¹
50
51
52
53
54
55
56
57
58
59
60
which should read as follows:-

(for example by the Academy of Medical Sciences UK in 2006
www.acmedsci.ac.uk/download.php?file=/images/project/Personal.pdf adverse comment in (Matthews, 2007)
UK government report in 2009 on genomic medicine
(<http://www.publications.parliament.uk/pa/ld200809/ldselect/ldscitech/107/10702.htm>), in a commentary from
an Islamic perspective (Al Aqeel, 2007), from the USA (Taylor, 2008) and in 2003 from the French National
Bioethics advisory Committee (<http://www.ccne-ethique.fr/docs/en/avis076.pdf>) end of footnote

1
2
3
4 This balance will be viewed differently in different cultures (Al Aqeel, 2007) and so
5
6 international input into detailed guidelines is essential to ensure collective agreement which is
7
8 requisite to effective collaboration. Harmonisation of standards will be a challenge. While the
9
10 development of a common ethical framework must be nurtured by culture and country-
11
12 specific input, the converse also holds true: the guidelines will serve as reference for the
13
14 developers of national laws and local ethics committees.
15
16
17
18
19

20
21 For many of the genes and for most of the issues dealt with below it seems likely that an
22
23 independent group of well-informed individuals to oversee specific LSDBs not only at their
24
25 initiation but on an ongoing basis will be essential. This general need is underlined by the
26
27 2008 revision of the Declaration of Helsinki

28
29
30 <http://www.wma.net/en/30publications/10policies/b3/17c.pdf>

31
32 which states that monitoring of ongoing studies must be put in place in addition to the initial
33
34 approval by an ethics committee. Governance is thus necessary as new issues may appear in
35
36 the course of a project or activity.
37
38

39
40 Inevitably in these guidelines there is a strong emphasis on the validity and complexities of
41
42 consent and the increasing difficulty of guaranteeing privacy in an era of electronic
43
44 publishing and growing internet use. The authors would not wish to discourage the curation
45
46 of LSDBs on this account. Current experience of curators indicates that the majority of
47
48 patients and research participants are likely to be happy to share their data, although
49
50 inevitably there will be exceptions, and participants' preference may change based on new
51
52 understanding of clinical significance. So far curators have been as likely to receive
53
54 complaints about the omission of a personal unique variant from the relevant database as its
55
56 unexpected inclusion.
57
58
59
60

1
2
3
4 The guidelines are presented in approximately the order in which the issues are encountered
5
6
7 by the prospective curator:-
8
9

10
11 **1. Clarify the main purpose of the particular database, recognising that this may change**
12 **over time.** Who does the curator expect will use this LSDB and why?
13
14

15
16
17
18 This will allow evaluation of the exact information required or desirable and whether
19
20 compliance with the remaining guidelines will be possible with a database open to the public.

21
22 It may be necessary to decide that at least part of the information should be restricted to
23
24 identified persons. For discussion of robust methods for validation of identities of enquirers
25
26 see the GEN2PHEN Knowledge Centre
27

28
29
30 <http://www.gen2phen.org/researcher-identification-primer> .
31
32

33
34
35 Many LSDBs are used as a tool by diagnostic laboratories assessing the likelihood that the
36
37 DNA change that they have found is the necessary and sufficient cause of a serious disease
38
39 and can be used to inform treatment and/or prevention, including preimplantation and
40
41 prenatal diagnosis and neonatal screening. However the data needed will vary in different
42
43 diseases so even for this use the ethical issues will be slightly different for each database.
44
45

46
47 Some examples of questions which must be considered by curators at this stage are shown in
48

49 **Box 2** .The answers will allow the generation of a list of ethical requirements which any
50
51 submitter must fulfil. They will inform decisions about any need for control of access and
52
53 may also help in the determination of appropriate members of an ethical oversight committee
54
55 as described later.
56
57
58
59
60

1
2
3
4
5
6
7 Both those LSDBs which catalogue very rare or even unique changes relating to serious
8
9 disease and those which deal with common variants of small individual effect should adhere
10
11 to stringent rules of data standardization, validation, quantification, and transparency of
12
13 sources, as described by participants at the HVP planning meeting (Kaput et al., 2009). These
14
15 aims should be clarified and explained in terms understandable by non specialists on the
16
17 public part of the LSDB website.
18
19

20 **Box 2 :Questions about aims and data required**

21
22
23 How much detailed clinical data will be needed and will this be in the form of a link to
24
25 another database?

26
27 Is any family information needed, *e.g.* to support conclusions on pathogenicity?

28
29 Will an attempt be made to record every apparently unrelated case with the same mutation?

30
31 What ethnic and geographic origin data will be needed and for what purpose?

32
33 Will an attempt be made to record all known 'neutral' ('normal') variation?

34
35 Is the aim to evaluate the contribution of common variants to common diseases?

36
37 Is the goal to inform basic research into the mechanism of disease, *e.g.* modifier genes?

38
39 Is the aim to evaluate genetic variation in response to therapy in individual or populations?

40
41 Will the LSDB collect results of *in vitro* functional analyses?

42
43 Will the data include results from a cell or tissue culture of patient/participant material?

44
45 Will the database be used to assemble volunteers for new therapies such as mutation-specific
46
47 strategies?

48
49 Is the interest mainly from an evolutionary perspective?
50
51
52

53 **2. Define database policy with respect to sources of data.** This will be dealt with in two
54
55 main sections:- existing data and future data, and within each there is a section on data
56
57 collected for research and that collected in a diagnostic setting. At the end of this section we
58
59
60

1
2
3
4 suggest a possible consent form which we propose might be appropriate in future,
5
6 particularly for results obtained in the course of clinical testing.
7
8
9

10 11 **EXISTING DATA**

12 13 **A. Published data**

14
15
16 Usually, information which is already published and available electronically is assumed to
17
18 have been collected by the appropriate standards that existed at the time of collection. By
19
20 virtue of their public availability, these data are generally assumed to be usable in an LSDB.
21
22

23 Although it is possible that the person giving the consent for the research and publication
24
25 may not have foreseen the full implications of web-based sharing of information, in most
26
27 situations the likelihood of re-identification and/or misuse of data is considered low.
28
29

30 Whenever feasible, the LSDB curator is encouraged to inform the original data producer
31
32 (defined as corresponding author for the publication) in order to explore whether any
33
34 restrictions or modifications of data are appropriate. This recommendation is more for
35
36 respect of free will and autonomy of patients than for fear of the risk of re-identification. The
37
38 specifics of a particular rare disease may warrant more stringent monitoring of data; this
39
40 should be addressed by the ethics oversight committee of the specific LSDB. The main
41
42 obligation on the curator is to check the scientific accuracy as far as possible, including
43
44 writing to authors when necessary. Curators should keep in mind that integration of
45
46 published data may on occasion give rise to conclusions with serious implications for
47
48 individuals or groups. Similar concerns have been discussed previously, for example, relating
49
50 to accumulated data on CGH microarrays (Tabor and Cho, 2007). Occasionally discussion
51
52 with the oversight group might be needed before full public release of the integrated data.
53
54
55
56
57
58
59
60

B Data existing in diagnostic laboratories or as clinical reports

Available but unpublished diagnostic data are a major problem. This is both because of limited clinical data but also because current practice does not usually ensure that those consenting to genetic testing have given permission for sharing (including scientific publication) of these data beyond the laboratory and clinical team undertaking the analysis.

LSDBs must be cautious in accepting unpublished data from any investigators or from accredited diagnostic laboratories, and consider issues that could limit the clinical accuracy of unpublished submissions, including standardization of clinical language, source of data, individual identification and consent.

It is suggested that where the quality of the data appears to be high and of significant clinical value but it is not feasible to obtain explicit consent, the decisions about which data should be uploaded and also which should be publicly displayed, protected by controlled access or displayed only in summary form should be made by the independent LSDB ethical oversight committee (see point 5). This committee must also be sensitive to different cultural views. In many cases it may be appropriate for these data to be anonymized, *i.e.* made 'not identifiable' (see point 6 for explanation of ICH sample coding terminology). Note that the current version of LSDB software LOVD (Fokkema et al., 2005) has the option to store data that are not public but that can be queried. The result of a query hitting non-public data is a notification that there is such information in the database but that the curator needs to be contacted to get more information.

In some cases patients/families already report the data themselves (often with a copy of the lab result they obtained) and this can be encouraged with appropriate further information requested if needed. LSDBs should then have a consent form that should be signed by the self-submitter and by all relatives whose results the submitter forwards to the LSDB.

FUTURE DATA

A Research data

In future research projects, consent forms should specifically indicate what data will be included in a publicly available database and describe their possible intended uses. The clinical significance of published data will continue to change as new findings emerge and the ethical repercussions will depend on many different variables. It may be appropriate to include agreement in the original consent about the need for re-contact or for delegation of decision-making to an ethical committee for future unforeseen uses and implications

B Diagnostic Data

With regard to consent, we would strongly recommend that informing donors of the possibility of transmission of data to an LSDB should in the future be part of the consent form for all genetic testing, together with an explanation of why this is useful, and how their privacy will be secured. Refusal to allow inclusion of data in an LSDB should not affect genetic testing since this would contravene the traditional commitment of medicine as exemplified in the UK General Medical Council guidelines 2006 'Make the care of your patient your first concern'

http://www.gmc-uk.org/Good_Medical_Practice_0510.pdf_32611016.pdf

This should be made clear to the person being asked for consent. It is essential that this is done without coercion in order to preserve the freedom of the consent. However the information provided to the patients and families should clearly explain the value of gathering such data and mention that in the long term, if data cannot be collected, interpretation of testing results may be less reliable or even impossible and development of future possibilities for treatment might be compromised. Although the curator should require a statement that the submitter has obtained appropriate consent in whatever way is acceptable in the country of origin of the data, the primary responsibility is that of the submitter. The

1
2
3
4 curator should supply to the testing laboratory a clear explanation about the LSDB on an
5
6 information sheet that the laboratory can provide to clinicians and patients.
7
8
9

10
11 A suggested form of wording as an addition to the consent for diagnostic testing (and which
12
13 may also be appropriate for testing as part of a research project) is as follows:
14
15
16
17

- 18
19 i. I understand that the interpretation of DNA test results, including my own is based
20
21 mainly on publicly available data from others who have been tested before me.
22
23 ii. I agree that the results of my DNA test and clinical examination may be added to
24
25 these public data sets, in a manner which does not disclose my personal identity and
26
27 which is in agreement with data protection law in my country.
28
29 iii. This information will then be available to help the diagnosis of others, and to further
30
31 understanding about the disease. Improved understanding of the molecular
32
33 mechanisms of disease may be important in developing new treatments and/or
34
35 prevention.
36
37 iv. Any information which could identify me or members of my family may only be
38
39 stored when a high standard of privacy and confidentiality (as defined and in
40
41 accordance with national standards for health data) is maintained. However,
42
43 unintentional third party cross examination of stored non-identified data, might
44
45 indicate, but not prove, identity. Should this happen, third party users of the database
46
47 will undertake not to explore this information further or to contact me
48
49
50
51
52
53
54 v. I understand that I will not receive any payment for this.
55
56
57
58
59
60

1
2
3
4 This wording will need modification in certain circumstances and if any part of the
5
6 interpretation of the result to be shown on the database depends on family history or testing
7
8 of other members of the family, this should be considered and consent sought if appropriate.
9
10
11
12
13
14 .

15
16
17
18 **3. Take Specific communities/cultures into account.** Identifiable groups such as Ashkenazi
19
20 Jews or Roma (Gypsies) may be particularly affected by a specific disease and thence
21
22 become a major part of the relevant LSDB. Following consultation with the community,
23
24 every effort must be made to take this into account and to provide privacy protection and
25
26 respect cultural sensitivity ensuring that high ethical standards are maintained. A small
27
28 specialised database gives the greatest chance of the unintended identifiability of one of the
29
30 subjects. It may occasionally be necessary to store data only at summary level to preserve
31
32 anonymity, as has been done in the Israeli and other National/Ethnic Mutation Databases
33
34 (NEMDBs) (Patrinos, 2006; Zlotogora et al., 2007). The cultural sensitivity of particular
35
36 groups such as the Maori of New Zealand will need a step of local consultation before any
37
38 sharing of DNA data, even for disorders not especially prevalent in that group.
39
40
41
42
43
44
45
46
47
48

49
50 **4. Take vulnerable persons into account.** Persons who do not have the capacity to consent
51
52 either because of disability or young age are especially vulnerable. In some disorders, this
53
54 will apply to many of the patients/ participants and regular external review of procedures for
55
56 obtaining consent from appropriate relatives/ representatives or other suitable authority
57
58 should be in place. Usually this will be part of the remit of the LSDB oversight committee.
59
60

1
2
3
4 **5. Create an ethics oversight committee.** A variation database which accepts genotype and
5
6 phenotype data not already in the public domain (or that are in the public domain but whose
7
8 combination and integration are foreseen to change the degree of identifiability of persons)
9
10 and makes them widely available, should have an independent and well-informed oversight
11
12 group. This should be drawn from several disciplines and from relevant society stakeholders,
13
14 including patient groups, in order to review the particular ethical issues arising in relation to
15
16 that LSDB. They should recommend constraints needed in uploading and displaying data and
17
18 decide on any requirement for control of access or for anonymization. The delegation of such
19
20 decisions to a committee with a long term remit will balance the difficulty of having truly
21
22 informed consent in such a fast moving field. The decisions of this committee may require
23
24 formal ethical approval either from their own institution or a national body.
25
26
27
28
29
30
31
32

33 A database which only accepts publicly available data would still benefit from some
34
35 independent ethical review and this is strongly recommended. This could be provided either
36
37 by a specific oversight committee as described above, or perhaps by an international HGVS
38
39 or HVP group who could advise a number of databases
40
41
42
43

44 **6. Remove identifying information before submission to database.** Every effort must be
45
46 made to ensure that the individuals whose DNA variation is displayed in an LSDB are not
47
48 individually identifiable. With increasing availability of total genomic sequence and the
49
50 enormous amount of personal information retrievable from websites, absolute certainty of
51
52 non identifiability is no longer guaranteed (Barash, 2007; Homer et al., 2008; Lowrance and
53
54 Collins, 2007; Walter, 2007). However, with care, the risk of identification from an LSDB
55
56 will be very low in almost every case, particularly if data from genome-wide analyses such as
57
58 SNP genotyping data are not associated with the mutation. (see point 7 for possible
59
60

1
2
3
4 exceptions such as unique variants). There is now a set of definitions including sample
5
6 coding terminology agreed internationally and recognised by all constituents of the
7
8 International Conference on Harmonisation (ICH) that has become official in 2008
9
10
11 <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm129296.pdf>
12

13 These definitions will be used here and are explained below:-
14
15
16
17

18 Decide whether data should be **'coded'** (also called **'re-identifiable'**) or **'anonymized'** (also
19
20 called **'de-identified'**). For the purposes of these guidelines, 'coded' is taken to mean
21
22 removal of all existing identifiers as far as is compatible with usefulness of the data and
23
24 substitution by proxy identifiers which are used in the database. The link between the existing
25
26 identifier and the proxy identifier could be maintained either by the submitter acting as the
27
28 'honest broker' between the hospital records and the LSDB or within a securely non-public
29
30 area of the LSDB. It would be desirable that each proxy identifier be unique and generated by
31
32 a standard coding mechanism, perhaps by some national or international body. This would
33
34 avoid inadvertent duplication of identifiers that might arise if the process was left to
35
36 individual LSDB curators. For published data, and especially for recent publications, many
37
38 cases are already coded and then classified as 're-identifiable' specifically for publication.
39
40 These codes might be acceptable if not recorded in hospital notes. However, new coding for
41
42 the database would be safer. Unpublished data, unless anonymized as explained below,
43
44 should always be re-coded to make them re-identifiable and not directly identifiable. In the
45
46 case of unpublished data on a rare disorder it is advisable that no link between a particular
47
48 entry and the submitter should be displayed, and that ethnic origin and geographic data on the
49
50 donor are not visible. This is recommended even if the clinical data displayed are minimal.
51
52
53
54
55
56
57
58 **Anonymized (or 'de-identified')** here means that identifiers and any information which
59
60 might be used as clues to identity through other links are permanently removed and the link

1
2
3
4 to the ID used by the submitter is **destroyed**. This would limit the usefulness of the data,
5
6 particularly with regard to long term phenotype follow-up data and any late correction of
7
8 wrong information but also possibly in ways that cannot currently be predicted. It also makes
9
10 withdrawal of consent impossible and is not the approach of choice.
11
12

13
14
15
16 If the decision is to proceed with **coded** or in other words '**re-identifiable**' data, although the
17
18 original identifier is replaced by a code, many other pieces of information give possible clues
19
20 to identity and will need to be removed to avoid unauthorised re-identification and, in many
21
22 countries, to obey privacy laws. In most published cases, all data given in the publication
23
24 would be acceptable to include in the database, including geographical location and ethnic
25
26 group of individuals studied (both for cases and for population controls) and clinical details
27
28 of patients. Limited family details present in the original report which help in the
29
30 interpretation of the mutation may also be included. A link to the original publication is
31
32 acceptable and useful. Note that data on frequency and population/ethnicity can be very
33
34 helpful in the design of cost-effective targeted diagnostics and/or treatment protocols. For
35
36 unpublished data, the specific nature of the LSDB will be considered by the ethical oversight
37
38 committee in recommending which data can be collected and displayed. Special
39
40 consideration should be given to rare mutations (see point 7).
41
42
43
44
45
46
47
48

49 **7. Add further protection of confidentiality if needed.** This may be necessary in the case of
50
51 rare or unique mutations in rare diseases, unique combinations of clinical features or where
52
53 higher protection is required for some other reason. The oversight committee will be valuable
54
55 here. The database will have limited usefulness if important genotype phenotype data cannot
56
57 be released because this information alone might allow identification of the individual. In
58
59 diseases where very detailed clinical data have been collected (especially clinical
60

1
2
3
4 photographs or detailed pedigrees) access to these data may have to be restricted by
5
6 appropriate registration and approval for access. A possible solution here which could also
7
8 be applied to any unpublished data for which explicit consent is not certain would be to
9
10 display the mutation in the database with no other data at all. Someone with a genuine reason
11
12 for wanting to know if this mutation causes disease could then click on a tool which would
13
14 send an email to the curator to explore the possibility of finding more information (see
15
16 section 8). There are several variations to this approach and the display of every variant
17
18 recorded including those for which no other information can be displayed publicly makes
19
20 accurate enquiries easier for the user than requiring the enquirer to specify in correct format a
21
22 variant which is not displayed As mentioned in point 2, one of us (JdD) has already provided
23
24 a similar facility for LOVD databases (Fokkema et al., 2005).
25
26
27
28
29
30
31
32

8. Allow no further disclosure without consent. Information about a particular entry
33
34 beyond what is publicly viewable in the database should not be supplied by a curator to an
35
36 enquirer unless consent for this has been explicitly provided. The request should be referred
37
38 to the submitter who will use using professional judgement in their response. This will usually
39
40 require seeking further explicit consent from the patient. If there is any doubt the independent
41
42 oversight committee should be consulted .
43
44
45
46
47
48

9. Provide provision for removal of data from database. The parents or guardians who
49
50 have given consent for a child's or incompetent adult's information to be included in the
51
52 LSDB should be made aware of their right to withdraw this information at any time (unless
53
54 data are truly not identifiable). The LSDB should make available information in order to
55
56 facilitate this task. If a child reaches the age of consent (16 in many but not all countries)
57
58 and is capable of making a decision, those who previously authorised data sharing should
59
60

1
2
3
4 ensure that he/she is aware of his/her LSDB entry and has the right to withdraw it. This should
5
6 also be the case if an adult previously judged incompetent becomes competent. However, it
7
8 should be made clear that whilst it will be possible to eradicate information which was
9
10 originally displayed from the database, it may not be possible to eradicate it from other
11
12 sources which have used this information for example in an overview publication.
13
14
15
16
17

18 **10. Be cautious in response to requests to an LSDB curator for a private opinion on**
19
20 whether a particular variant is pathogenic, especially if any of the information used is not
21
22 published. Add a disclaimer about responsibility for the clinical use of the opinion and be
23
24 cautious of a 'virtual medical advisor relationship'. From the medico-legal point of view, it is
25
26 safest to obtain clinical interpretation from published data. If unpublished information is
27
28 used, a careful record should be kept. Recommendations of the IARC Working Group on
29
30 Unclassified Genetic Variants encourage that classification of pathogenicity be carried out
31
32 not by individuals but by teams of experts that can carefully evaluate all lines of evidence
33
34 (Greenblatt et al., 2008; Tavtigian et al., 2008).
35
36
37
38
39
40
41

42 **11. Limit links to other LSDBs.** If mutations in more than one gene are relevant to a
43
44 particular disease, it may be useful to record the variation of both genes in the same
45
46 individual and link the entries so that the fact that it is one person is recorded. This facility is
47
48 already available on at least one LSDB platform (LOVD) and can be of great value in the
49
50 interpretation of results. However, logically it may eventually extend to enough genes to
51
52 allow identification of the individual. At this point, the considerations of the ethics of large
53
54 scale resequencing will be relevant (see Lowrance and Collins 2007). For example, a recent
55
56 investigation into DNA variants causing X-linked mental retardation included substantial
57
58 amounts of sequence information on the coding regions of X-linked genes (Tarpey et al.,
59
60

1
2
3
4 2009). These data were regarded as too sensitive for the full set of variants for each patient to
5
6 be entered onto the LSDBs in the most informative way. See <http://www.LOVD.nl/MR> for
7
8 the summary data submitted.
9

10
11
12
13 **12. Consider carefully the transfer of publicly available data from LSDBs to genome**
14 **browsers.** Some of the genetic variations collected and displayed by curators of LSDBs are
15
16 now visible also in one or several of the main tools used by scientists worldwide analysing
17
18 the human genome, for example at the National Center for Bioinformatic Information (NCBI)
19
20 at Bethesda and the UCSC Genome Browser at Santa Cruz. This does not raise entirely new
21
22 ethical issues except in the need for adequate recognition of the work of the LSDB curator. It
23
24 makes the misuse of data for re-identification slightly more likely and may also increase the
25
26 chance of a mistake being widely disseminated in a short time. Further discussion of the
27
28 sharing of data with genome browsers can be found elsewhere (den Dunnen et al., 2009).
29
30
31
32
33
34
35
36

37 **Acknowledgements**

38
39 We would like to thank Professor Ruth Chadwick, chair of the HUGO Ethics Committee and
40
41 other members of that committee in particular Dr Thomas H. Murray, President of The
42
43 Hastings Center in New York for reading the manuscript and providing valuable criticism.
44
45

46 We also thank an anonymous referee for constructive comments. Before this stage many
47
48 people contributed helpful suggestions for the improvement on the manuscript. Particular
49
50 thanks are due to Professor Bartha Knoppers, PhD, Director of the Centre of Genomics and
51
52 Policy, Faculty of Medicine, Department of Human Genetics, McGill and former Chair of
53
54 the International Ethics committee of the Human Genome Organization (HUGO). Thanks also
55
56 to Professor Finlay Macrae, at the Royal Melbourne Hospital, for constructive comments on
57
58 the manuscript and to members of the working group on ethics at the HVP planning meeting
59
60

1
2
3
4 for lively and helpful discussion. We are grateful to Ms Lauren Martin for help with the
5
6 manuscript.
7

8
9 SP's LSDB is supported by the Tuberous Sclerosis Alliance. AC-T's, RD's, JTdD's and
10
11 GPP's contributions have received funding from the European Community's Seventh
12
13 Framework Programme (FP7/2007-2013) under grant agreement number 200754 — the
14
15 GEN2PHEN project.
16

17
18 Participation of AC-T in the work on these guidelines is also supported by European funded
19
20 research projects TECHGENE, FP7 Grant agreement no. 223143, and the EU Public Health
21
22 project PHGEN 2, Grant agreement N° 2008302. MJS's participation is supported by funds
23
24 from the REGENPSI network Consellería de Educación, Xunta de Galicia.
25
26
27
28
29
30
31
32

33 **References**

34
35 Al Aqeel AI. 2007. Islamic ethical framework for research into and prevention of genetic
36 diseases. *Nat Genet* 39(11):1293-8.
37

38
39 Barash CI. 2007. Threats to privacy protection. *Science* 318(5852):913-4.
40

41 Cotton RG, Appelbe W, Auerbach AD, Becker K, Bodmer W, Boone DJ, Boulyjenkov V,
42 Brahmachari S, Brody L, Brookes A and others. 2007. Recommendations of the 2006 Human
43 Variome Project meeting. *Nat Genet* 39(4):433-6.
44

45 Cotton RG, Sallee C, Knoppers BM. 2005. Locus-specific databases: from ethical principles
46 to practice. *Hum Mutat* 26(5):489-93.
47
48

49 den Dunnen JT, Sijmons RH, Andersen PS, Vihinen M, Beckmann JS, Rossetti S, Talbot CC,
50 Jr., Hardison RC, Povey S, Cotton RG. 2009. Sharing data between LSDBs and central
51 repositories. *Hum Mutat* 30(4):493-5.
52

53
54 Fokkema IF, den Dunnen JT, Taschner PE. 2005. LOVD: easy creation of a locus-specific
55 sequence variation database using an "LSDB-in-a-box" approach. *Hum Mutat* 26(2):63-8.
56 Greenblatt MS, Brody LC, Foulkes WD, Genuardi M, Hofstra RM, Olivier M, Plon SE,
57 Sijmons RH, Sinilnikova O, Spurdle AB. 2008. Locus-specific databases and
58 recommendations to strengthen their contribution to the classification of variants in cancer
59 susceptibility genes. *Hum Mutat* 29(11):1273-81.
60

1
2
3
4 Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan
5 DA, Nelson SF, Craig DW. 2008. Resolving individuals contributing trace amounts of DNA
6 to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet
7 4(8):e1000167.
8

9
10 Kaput J, Cotton RG, Hardman L, Watson M, Al Aqeel AI, Al-Aama JY, Al-Mulla F, Alonso
11 S, Aretz S, Auerbach AD and others. 2009. Planning the Human Variome Project: The Spain
12 report. Hum Mutat 30(4):496-510.

13 Knoppers BM, Laberge CM. 2000. Ethical guideposts for allelic variation databases. Hum
14 Mutat 15(1):30-5.
15

16
17 Lowrance WW, Collins FS. 2007. Ethics. Identifiability in genomic research. Science
18 317(5838):600-2.
19

20
21 Matthews R. 2007. Comment: Consent is crucial to medical research. New Sci 195(2615):18.
22

23 Patrinos GP. 2006. National and ethnic mutation databases: recording populations'
24 genography. Hum Mutat 27(9):879-87.
25

26
27 Tabor HK, Cho MK. 2007. Ethical implications of array comparative genomic hybridization
28 in complex phenotypes: points to consider in research. Genet Med 9(9):626-31.
29

30 Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C,
31 Dicks E, Menzies A and others. 2009. A systematic, large-scale resequencing screen of X-
32 chromosome coding exons in mental retardation. Nat Genet 41(5):535-43.
33

34
35
36 Tavigian SV, Greenblatt MS, Goldgar DE, Boffetta P, 2008 for the IARC Unclassified
37 Genetic Variants Working Group, Assessing pathogenicity: overview of results from the
38 IARC Unclassified Genetic Variants Working Group, Hum Mut 28:1261-1264
39

40
41 Taylor P. 2008. Personal genomes: when consent gets in the way. Nature 456(7218):32-3.
42

43 Walter C. 2007. A little privacy, please. Computer scientist latanya sweeney helps to save
44 confidentiality with "anonymizing" programs, "deidentifiers" and other clever algorithms.
45 Whether they are enough, however, is another question. Sci Am 297(1):92, 94-5.
46
47

48 Zlotogora J, van Baal S, Patrinos GP. 2007. Documentation of inherited disorders and
49 mutation frequencies in the different religious communities in Israel in the Israeli National
50 Genetic Database. Hum Mutat 28(10):944-9.
51
52
53
54
55
56
57
58
59
60