THE HUMAN VARIOME PROJECT
an NGO official partner of UNESCO
sharing data • reducing disease

HVP/DA/004-01/EN

2017-04-05

WG08

G/DSDBAC

DRAFT FOR APPROVAL

CHECKLIST FOR GENE/DISEASE SPECIFIC DATABASE CURATORS TO ENABLE ETHICAL DATA MANAGEMENT

Authors

Rosemary Ekong;
Mauno Vihinen;
Michael Parker

Editor

Timothy D. Smith

Published by:
Global Variome
International Centre for Life, Times Square, Newcastle-Upon-Tyne, NE1 4EP, UNITED KINGDOM

Published 2017-04-05. Printed in the United Kingdom.

PDF: **ISBN <<PDF_ISBN>>**
Print: **ISBN <<PRINT_ISBN>>**

Contents

Foreword

The Human Variome Project is an international consortium of researchers, policy makers and healthcare professionals committed to the free and open collection, curation, interpretation, and sharing of genomic knowledge.

The Human Variome Project Consortium envisions a world where the availability of and access to genetic variation information is not an impediment to diagnosis and treatment; where the burden of genetic disease on the human population is significantly decreased; and where the sharing of genetic variation information is standard clinical practice.

To facilitate worldwide and interoperable sharing of genomic knowledge, the Human Variome Project Consortium produces Standards and Guidelines. HVP Standards are those systems, procedures and technologies that the Human Variome Project Consortium has determined shall be used by the community. These carry more weight than the less prescriptive HVP Guidelines, which cover those systems, procedures and technologies that the Human Variome Project Consortium has determined would be beneficial for the community to adopt.

HVP Standards and Guidelines are central to supporting the work of the Human Variome Project Consortium and cover a wide range of fields and disciplines, from ethics to nomenclature, data transfer protocols to collection protocols for clinical data. They can be thought of as both technical manuals and scientific documents, and while the impact of HVP Standards and Guidelines differ, they are both generated in a similar fashion.

HVP Standards and Guidelines make the collection, curation and sharing of information more efficient and reliable by establishing consistent protocols that can be universally understood. They facilitate interconnection of and interoperability between different systems.

HVP Standards and Guidelines represent a consensus of the Human Variome Project Consortium, each member of which has had the opportunity to participate in the development and review of each standard and guideline. In addition, as every effort is made to include all interests in the activity, HVP Standards and Guidelines can be considered to be representative of all interests concerned within the scope of each Standard or Guideline.

The Human Variome Project defines consensus as significant agreement between all affected parties covered by the scope of the standard or guideline. Consensus requires that all views and objections be considered, and that a concerted effort be made toward their resolution.

More information on the Human Variome Project is available at the Project's website (http://www.humanvariomeproject.org/). Procedures for the development of HVP Standards and Guidelines can be found in *PD06-2011: Standards Development Process*, available at http://short.variome.org/PD06-2011.

This Document

This document has been prepared by the HVP Working Group WG08: Ethics Checklist for Gene/Disease Specific Database Curators and Submitters. The Gene/Disease Specific Database Advisory Council acted as Sponsoring Council.

An Exposure Draft of this Document was released to the Human Variome Project Consortium on 2016-07-08.

Important Notice

HVP Standards and Guidelines are not intended to replace or substitute for any applicable legislation or regulation in any jurisdiction, or any institutional policy or funding agreement that a genetic variation information resource is operating under. Implementers of HVP Standards and Guidelines are responsible for determining and complying with all appropriate ethical and cultural protection practices and all applicable laws, regulations, policies and agreements.

Introduction

The Povey et al. (2010) guidelines were published to help curators of web-based locus-specific variation databases (LSDBs) make information within their databases accessible where these can be used for clinical and research purposes, while safeguarding the privacy of individuals. When looking at the guidelines to apply, curators found that some of these were difficult to achieve in practice. At the Human Variome Project meeting in Paris (May 2014) it was agreed that a more practical checklist was required for curators.

The Gene/Disease Database Advisory Council sponsored the formation of WG08, which was charged with drawing up "A checklist of actions and processes related to the ethical management of data in a genetic variation database that curators of gene/disease specific databases should consider when establishing and curating their database." A survey of curators showed that each point in the Povey et al. (2010) guidelines had been implemented by some curators. However, some of the points were considered unnecessary or not applicable and therefore not implemented by most curators that responded to the survey.

The checklist provided below includes information gained from the analysis of the curators' survey and from scenarios presented to the working group. Some of the 'practical' guidelines in Povey et al. (2010) have been retained and information previously published in other articles have also been included (Celli et al. 2012; Mascalzoni et al. 2014).

1 Scope

The purpose of this document is to provide practical steps that should enable LSDB curators collect and share data, whilst at the same time operating within acceptable ethical standards.

Implementation of the checklist will depend upon what is suited to the content of a database and what the local ethical and legal requirements are.

2 Terms and Definitions

De-identified data
De-identified data are data that have had features removed or replaced such that it is highly unlikely to identify an individual from the data alone. This includes names, dates, and other identifiers, some of which have been defined in the United States Health Insurance Portability and Accountability Act of 1996. It should be noted that the term 'de-identified' is often defined differently between different legal jurisdictions.

3 Checklist

1) Define the **purpose of your database.**

   a. Include the scope and type of information in database.

2) Define the database **policy governing data collection** (example in Vihinen et al. 2012; Appendix 1).

   a. Provide lay information for patients wishing to submit their data.

3) **Attribution**: To encourage the submission of unpublished data and as some recognition for their contribution, offer submitters co-authorship on publications (authored by the curators) that make use of data from submitters.

4) Establish an **Oversight Committee** (OC). In the curation of unpublished data practical ethical questions can arise. Therefore, it is necessary to have an independent body (the oversight committee) where the curator can direct questions to be discussed and addressed. This is essential where unpublished data is accepted into databases but may not be necessary where all data come exclusively from publications. It should be noted that an Oversight Committee differs from Ethics Committees (e.g. Institutional Review Boards (IRBs), Independent Ethics Committees (IECs) and Research Ethics Committees (RECs)) that are charged with ensuring high standards in the ethical conduct of research involving human subjects.

   a. Purpose of OC

      i. To act as an independent forum for the consideration of practical ethical issues arising in the day-to-day work of the database.

      ii. To consider any other matters relating to sharing of unpublished data submitted to the database, in line with local regulations/requirements and recommendations in the field.

   b. Guidance on composition of OC

      i. Members should be independent of the database, but knowledgeable about the condition and represent the different groups involved, e.g. clinicians, researchers, and lay persons from patient groups.

      ii. At least one member of the OC should have ethics training, e.g. a short course on research ethics (online options available).

      iii. The OC should not include curators.

5) **Data collection**

   a. Consented data

      i. Inform submitters of their responsibility to ensure that valid consent has been obtained and that only de-identified (coded) patient IDs are submitted. De-identified IDs allow submitters to respond to queries from the curator or to update new information about a particular case.

      ii. Note that completely anonymising patient IDs makes it virtually impossible to update valuable information that subsequently becomes available, either by the submitter or curator.

   b. **Unpublished data:** Submissions from diagnostic labs (health service labs and commercial sources), clinics/clinicians and sometimes from patients will mostly be unpublished.

      i. Ensure de-identified IDs are submitted.

6) **Curation of unpublished data**

a. **Unpublished data**: Received as a query or submitted for inclusion in the database. This data may come from a clinician, genetic counsellor, diagnostic labs or a patient.

i. If the data is from a query, inform the enquirer that the variant will be included in the database.

ii. Assign a de-identified code to each entry, if there is none already.

iii. Keep sensitive personal data non-public. This refers to information that is of a private nature that could be used in a discriminatory manner.

iv. In linking entries to details of the submitter, curators should implement what their local regulation permits.

b. Publicly viewable data (from submitted unpublished data)

i. Summarise publicly viewable data to ensure clarity on family relationships.

ii. Curate submitted data to ensure personal details do not identify individuals.

iii. Phenotype information is important for clinical diagnosis. Where this is available and efforts have gone into protecting the identity of the individual, clinical details should be displayed.

c. Non-public data

i. This section of the database is reserved for confidential information that curators will need to refer to.

7) Permitting the use of non-public data for scientific/clinical purposes

a. **Request from clinician or diagnostic lab**: Curators may receive requests to share non-public information from bona fide clinicians/diagnostic labs who need the information for patient care/diagnostic report. An example may be a new variant with the associated clinical data, segregation information and pathogenicity, which a submitter has requested that these should not be made public until after their impending publication (see point 8).

i. Forward request to the submitter.

b. Request from researcher

i. Forward request to the submitter.

8) Request to keep submitted data non-public: Some submitters request that data be kept non-public until they are published.

a. Make the submitter aware that publishing the variant in the database does not result in the rejection of a subsequent manuscript that mentions the data.

b. Note that searches, e.g. in LOVD, returns a message which indicates a variant at that nucleotide position is in the database, but the nucleotide information is not given. There is also a suggestion to contact the curator.

c. The following options may be adopted:

i. Enter data but make the entire entry ‘non-public’. Note point 8b above; or

ii. Enter data but make variant public and associated information ‘non-public’. This option should be discussed with the submitter.

d. Any request received should be forwarded to the submitter.

9) **Request for submitter’s details**: Some LSDBs do not link submitter details to unpublished data.

a. Any request for submitter details should be forwarded to the submitter allowing them to respond directly with the requester.

10) **Giving your opinion**: As a curator you will be considered as an ‘expert’ and will be asked your opinion on the consequences of an identified variant or other aspects of the disease.

a. If you have a team (clinical and scientific) that is qualified and knowledgeable about the disease, an opinion on the potential consequence of a variant may be given, especially when you (as the curator) have assigned “concluded pathogenicity” to variants listed in your database.

b. If you do not have a team and you do not have in-depth knowledge about the disease, refrain from giving any opinion.

11) **Sharing information with genome browsers**: This increases visibility for your database and should be encouraged.

3.1 Checklist in brief

1. Define the purpose of your database.
2. Define the database policy governing data collection.
3. Offer attribution to submitters.
4. Establish an oversight committee.
5. Data collected with valid consent and de-identified (responsibility of submitters).
6. Curate unpublished data to protect patient privacy whilst remaining useful.
7. Requests for non-public data should be forwarded to the submitter.
8. Requests to keep submitted data non-public can be honoured.
9. Requests for submitter’s details should be forwarded to the submitter.
10. Giving your opinion may be considered if you have a team qualified and knowledgeable about the disease**.**
11. Information can be shared with genome browsers.

4 Bibliography

Celli, Jacopo, Raymond Dalgleish, Mauno Vihinen, Peter E M Taschner, and Johan T. den Dunnen. 2012. “Curating Gene Variant Databases (LSDBs): Toward a Universal Standard.” Human Mutation 33 (2): 291–97. doi:10.1002/humu.21626.

Mascalzoni, Deborah, Edward S Dove, Yaffa Rubinstein, Hugh J S Dawkins, Anna Kole, Pauline McCormack, Simon Woods, et al. 2014. “International Charter of Principles for Sharing Bio-Specimens and Data.” European Journal of Human Genetics : EJHG, no. August: 1–8. doi:10.1038/ejhg.2014.197.

Povey, Sue, Aida I. Al Aqeel, Anne Cambon-Thomsen, Raymond Dalgleish, Johan T. den Dunnen, Helen V. Firth, Marc S. Greenblatt, et al. 2010. “Practical Guidelines Addressing Ethical Issues Pertaining to the Curation of Human Locus-Specific Variation Databases (LSDBs).” Human Mutation 31 (11): 1179–84. doi:10.1002/humu.21339.

Vihinen, Mauno, Johan T. den Dunnen, Raymond Dalgleish, and Richard G. H. Cotton. 2012. “Guidelines for Establishing Locus Specific Databases.” Human Mutation 33 (2): 298–305. doi:10.1002/humu.21646.

Appendix 1

Example of database policy from ORAI1base (Variation registry for Severe combined immunodeficiency) at http://structure.bmc.lu.se/idbase/ORAI1base/?content=_db_policy/IDbases.

DATABASE POLICY

The ImmunoDeficiency Variation Databases (IDbases) and other variation databases maintained at the Protein Structure and Bioinformatics Group (PSB), Lund University, are maintained and provided as a public service for academic community.

Individuals submitting data to and using the variation databases managed by the PSB should be aware of the following:

1. The PSB has a uniform policy of free and unrestricted access for academic community to all of the data records their databases contain. Scientists worldwide can access these records to plan experiments or publish any analysis or critique. Appropriate credit is given by citing the database. Instructions for citing are provided in each individual database.

2. The databases are intellectual property of the PSB. Details are available for Copyright and Liability.

3. Corrections of errors and update of the records by authors are welcome and erroneous records may be removed from the next database release.

4. Submitters are advised that the information displayed on the Web sites maintained by the PSB is fully disclosed to the public. It is the responsibility of the submitters to ascertain that they have the right to submit the data. This applies also the appropriate consent from the patient and/or family.

5. Beyond limited editorial control and some internal integrity checks, the quality and accuracy of the record are the responsibility of the submitting author, not of the database. The databases will work with submitters and users of the database to achieve the best quality resource possible.

6. Data in the PSB mutation databases may be shared with central repositories according to published Human Genome Variation Society guidelines.

7. The information provided on this site is designed to support, not replace, the relationship that exists between a patient/site visitor and his/her existing physician.

8. We keep the confidentiality of the data relating to individual patients and visitors to the web site, including their identity. No data is collected that would allow identification of the patients for whom information is stored and distributed in the database. We do not share any information about database visitors with third parties. As database curators and owners we undertake to honour or exceed the legal requirements of medical/health information privacy that apply in Sweden.

9. The database does not host any advertisements.