

Rosemary Ekong      ORCID iD: 0000-0002-5984-7893

Mauno Vihinen      ORCID iD: 0000-0002-9614-7976

## **CHECKLIST FOR GENE/DISEASE-SPECIFIC VARIATION DATABASE CURATORS TO ENABLE ETHICAL DATA MANAGEMENT<sup>‡</sup>**

Rosemary Ekong<sup>1</sup>, Mauno Vihinen<sup>2</sup>

<sup>1</sup>Department of Genetics, Evolution & Environment, University College London, London, UK.

<sup>2</sup>Department of Experimental Medical Science, BMC B13, Lund University, SE-22184 Lund, Sweden.

**Corresponding author:** Rosemary Ekong, Department of Genetics, Evolution & Environment, University College London, London WC1E 6BT, UK.

Email: [r.ekong@ucl.ac.uk](mailto:r.ekong@ucl.ac.uk)

### **Grant numbers**

The Tuberous Sclerosis Alliance (03-17) and the Tuberous Sclerosis Association (2017-P01) support the LSDB work of R. Ekong.

The Swedish Research Council (Vetenskapsrådet, VR 2015-02510) and the Swedish Cancer Society (grant number CAN 2017/699) support M.Vihinen.

### **Abstract**

Databases with variant and phenotype information are essential for advancing research and improving the health and welfare of individuals. These resources require data to be collected, curated and shared among relevant specialties to maximise impact. The increasing generation

<sup>‡</sup> This paper is dedicated to the memory of Professor Sue Povey who advocated sharing variant data for the benefit of patients and their families.

of data which must be shared both nationally and globally for maximal effect presents important ethical and privacy concerns. Database curators need to ensure that their work conform to acceptable ethical standards. A Working Group of the Human Variome Project had the task of updating and streamlining ethical guidelines for locus-specific/gene variant database curators. In this article, we present practical and achievable steps which should assist database curators in carrying out their responsibilities within acceptable ethical norms.

**Key words:** Practical checklist, database curators, genetic variation databases, data collection, data curation, sharing variant data, ethical data management

## **Introduction**

We are now in an era where the genetic basis of clinical conditions is sought for targeted clinical care and research, especially where the gene(s) involved have been identified. Analysis of several genes at once or the entire human genome have generated huge amounts of genetic data which need to be carefully assessed to determine their involvement in a patient's phenotype. There are now several publicly available, web-based, locus-specific variation databases (LSDBs) or gene variant databases (GVDBs) (e.g. <https://databases.lovd.nl/shared/genes>). These databases can only be helpful as guides for both clinical and research purposes if they hold useful information. In addition, sharing data assists with patient care, as well as advancing research. However, collecting and sharing data require database curators to operate within national and international standards to ensure the confidentiality of patients and family members (Takashima et al., 2018) is not breached.

Guidelines were published in 2010 (Povey et al., 2010) to help curators of web-based LSDBs or GVDBs to make information within their databases accessible where these can be used for

clinical and research purposes, while safeguarding the confidentiality and privacy of individuals. With increasing interest in the curation of variant and phenotype information, several curators raised concerns at the 2014 Biennial Meeting of the Human Variome Project (HVP, <http://www.humanvariomeproject.org/>), about the application of existing guidelines. Curators indicated that they found some of the existing guidelines difficult to follow in practice, e.g. “Limit links to other LSDBs” when some diseases have two genes involved and linking variants found in both genes enables the interpretation of results. In response to this, an HVP working group was established to develop “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”. Although this checklist was primarily intended for curators of web-based LSDBs or GVDBs, extended consultation showed that its content is relevant to others in related areas, including national/population variation databases and biobanks. Therefore, we encourage the adoption of this Checklist by other database curators with publicly accessible data, where it is judged to be applicable. As new developments emerge, it is envisaged that this Checklist will be adapted to accommodate any relevant changes.

The purpose of the Checklist is to provide “practical steps” to enable LSDB and GVDB curators to collect and share data in a manner that ensures and promotes acceptable ethical standards. Not only should the work of curators meet international requirements, but individual database curators should be aware of, and act within, relevant local and national regulatory frameworks that govern their operations when sharing data both nationally and internationally. Requirements may vary in different countries (Ludvigsson et al., 2015; Stoeklé et al., 2019) or cultures (Al Aqeel, 2007), and this might require aspects of the

Checklist to be modified for implementation. On the other hand, the Checklist could present an opportunity to improve requirements where this had not been updated.

### **Survey of Curators and Other Material**

As a first step, a survey of curators was conducted to gain a better understanding of current practices and to determine the extent to which the existing guidelines (Povey et al., 2010) had been implemented. The Checklist below considers the results of this survey, as well as the views of experts in database curation from within HVP, members of HVP's various councils and committees, and inputs from the broader HVP membership. In addition, 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, Dalglish, Vihinen, Taschner & den Dunnen, 2012; Vihinen, den Dunnen, Dalglish & Cotton, 2012; Mascalzoni et al., 2015).

In implementing this Checklist, it is important that gene/disease database curators also refer to relevant international, national and regional documents. Some useful resources include:

- (1) "Ethical and Privacy Principles in relation to Responsible Sharing of Genomic and Health Related Data" produced by the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) (Appendix 1);
- (2) A broader "Framework for Responsible Sharing of Genomic and Health-related Data" (Version 10, Sept 2014) by the Global Alliance for Genomics and Health (GA4GH) (Knoppers, 2014), also available at <https://www.ga4gh.org/ga4gh/toolkit/regulatoryandethics/framework-for-responsible-sharing-genomic-and-health-related-data/>;

- (3) The “WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks” (World Medical Association, 2016), which covers additional ethical principles for the collection, storage and use of identifiable data will be relevant for curators with access to confidential/sensitive data.
- (4) The EU General Data Protection Regulation 2016/679 (GDPR) (2016) which came into effect on 25 May 2018.

### **Terms and Definitions**

For clarity and consistency several terms used in the Checklist are defined below.

- **Coded data**

Coded data refers to data that have undergone pseudonymisation as defined in Article 4(5) of the EU General Data Protection Regulation 2016/679 (GDPR) (2016) which came into effect on 25 May 2018.

- **Pseudonymisation**

This means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person". GDPR document is available at [http://ec.europa.eu/justice/data-protection/reform/files/regulation\\_oj\\_en.pdf](http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf). A coded ID permits re-identification by the submitter.

A simplified definition can be found in the GA4GH Data Sharing Lexicon (Global Alliance for Genomics and Health, 2016) where ‘pseudonymisation or coding’ is defined as “The act

of replacing an identifier with a code for the purpose of avoiding direct identification of the participant, except by persons holding the key linking the code and identifier”.

- **Database curator**

Database curator, as used here, refers to a person or persons who is/are “...responsible for assuring the quality, integrity, and access arrangements of data and metadata in a manner that is consistent with applicable law, institutional policy, and individual permissions” (Global Alliance for Genomics and Health, 2016). The activities of a database curator include data extraction, integration, presentation, publication, management, monitoring and reducing redundant information, thereby resulting in up-to-date information that is as accurate as possible. A Curator also oversees who has access to the information and under what circumstances. In HVP terminology, the database curator is responsible for ensuring that the database policy is kept current and useful. Some organizations use the term ‘data steward’ (e.g. Global Alliance for Genomics and Health, 2016) to equate with database curator, but this term is not widely used.

- **Personal data or Identifiable data**

The Checklist adopts the definition in Article 4(1) of the EU General Data Protection Regulation 2016/679 (2016) that defines ‘personal data’ as “any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”; and in alignment with the GA4GH Data Sharing Lexicon (Global

Alliance for Genomics and Health, 2016), personal/identifiable data refers to “data that alone or in combination with other data may reasonably be expected to identify an individual”.

- **Personal data breach**

This is defined in Article 4(12) of the EU General Data Protection Regulation 2016/679 (2016) as “a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed”.

### **Details of the Checklist**

**Define the purpose of the database:** This should include the scope and type of information in database. Further information can be found in Povey et al., (2010).

**Define the database policy:** In accordance with the agreed purpose of a curator’s database, database policy governing data collection, display, access, variant interpretation and classification, and corrections should be defined. The policy should also include clauses on data correction and withdrawal/erasure. Examples of database policies are in Vihinen et al., 2012 (also Appendix 2) and Knoppers, 2014 (GA4GH Proposed Policy Template). The database policy should also include information suitable for patients with limited technical knowledge who may wish to submit their data.

**Attribution:** Time constraint is one of the reasons large amounts of data remain unpublished and are held by individuals, research groups or clinical and commercial laboratories. To ensure such data contribute to advances in science and medicine, it is

important to get them into a single curated resource (widely accessible databases) where they can contribute towards discoveries (Ekong et al., 2016) that would otherwise not have been possible at the time.

- In recognition for the submission of unpublished data, submitter names should be listed along with their data (unless not permitted by database regulations). This form of microattribution (Patrinós et al., 2012) has been demonstrated by Giardine et al., (2011).
- Where local database regulations do not permit listing submitter names, consider offering co-authorship to submitters where their data are used in publications authored by curators, and the submitters have made significant scientific contributions. What constitutes ‘significant contribution’ can be found in the document published by the ICMJE (International Committee of Medical Journal Editors, 2017).
- If the options in the first two points are not possible, acknowledge submitters in publications.

**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(s) can direct questions to be discussed and addressed, including requests for large scale access. This is essential where unpublished data is accepted into databases but may not be necessary where all data come exclusively from publications. The OC should be independent of the curation and funding of the database that they oversee. 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.

- The purpose of the Oversight Committee is:
  - To act as an independent forum for the consideration of practical ethical issues arising in the day-to-day work of the database;
  - 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.
- Guidance on the composition of Oversight Committee
  - Members of the OC should be independent of the curation and funding of the database they have oversight of, but knowledgeable about the condition and represent the different groups involved, e.g. clinicians, researchers, database curators and lay persons from patient groups.
  - At least one member of the OC should have ethics training.

**Data collection:** This includes both published and unpublished data.

Published data: Privacy concerns discussed in this article do not apply to data obtained from publications.

Consented unpublished data

- Submitters should be informed of their responsibility to ensure that valid consent has been obtained and that only coded patient IDs are submitted. Coded IDs allow submitters to respond to queries from the curator or to update new information about a specific case.
- Note that completely anonymising patient IDs makes it virtually impossible to update valuable information that subsequently becomes available, either by the submitter or curator.

- Ensure coded IDs are used for submissions that are not linked to any publicly available source, e.g. data from diagnostic labs (health service labs and commercial sources), clinics/clinicians and sometimes from patients.

### **Curation of unpublished data with consent**

Unpublished data: This may be received as a query or submitted for inclusion in the database. For example, this data may come from a clinician, genetic counsellor, diagnostic laboratories, patient or researcher.

- If the data is from a query, it is important to inform the enquirer that the variant will be included in the database.
- Assign a coded ID to each entry, if there is none already.
- Keep sensitive personal data non-public. This refers to information that is of a private nature that could be used in a discriminatory manner.
- In linking entries to details of the submitter, curators should abide by relevant applicable regulations.
- Ensure that variant nomenclature adheres to HGVS standards (<http://www.HGVS.org/varnomen>). The original variant description should be kept in a separate column. If a variant is ambiguous or does not match the reference sequence, consult with the submitter. Incorrect variant descriptions should not be included in the database. If the problem with a variant description cannot be resolved, exclude the case from the database.
- Publicly viewable, unpublished data: Before submitted data can be made publicly available, the following should be done:
  - Information that will be publicly viewable should be summarised to ensure clarity on family relationships.

- Ensure that personal details in the submitted data are curated such that individuals cannot be identified.
- Phenotype information is important for clinical diagnosis. Where phenotype information is available, and efforts have been made to protect the identity of the individual, details on the phenotype should be displayed.
- Non-public data
  - This section of the database is reserved for sensitive information that curators will need to refer to.
  - Requests to share non-public data should be forwarded to the data submitter (see the following section on ‘Permitting the use of non-public data for scientific/clinical purposes’).

### **Permitting the use of non-public data for scientific/clinical purposes**

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 as a guide 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 the following section on ‘Request to keep submitted data non-public’). Such requests should be forwarded to the submitter.

Request from researcher should also be forwarded to the submitter.

**Request to keep submitted data non-public:** Some submitters request that data be kept non-public until they are published. It is important to make the submitter aware that:

- DNA diagnosis is improved by sharing data on genes, variants and phenotypes; and that publicly sharing data offers optimal care to patients and their families;
- Publishing the variant in the database does not result in the rejection of a subsequent manuscript that mentions the data;
- Searches in publicly available variant databases may return a message indicating a non-public record with a variant at that position is in the database, with the suggestion to contact the curator to receive more details.
- The following options may be adopted for requests to keep submitted data out of public view:
  - Enter data but make the entire entry ‘non-public’. Note the point above which is on ‘Searches in publicly available variant databases.....’; or
  - Enter data but make the variant public and associated information ‘non-public’. This option should be discussed with the submitter.
- Any request for information about the variant should be forwarded to the submitter.

**Request for submitter’s details:** Some LSDBs do not link submitter details to unpublished variant data, especially where most of these are rare variants. Any request for submitter details should be forwarded to the submitter allowing them to respond directly to the requester.

**Giving your opinion:** A curator will be considered an ‘expert’ and may be asked for an opinion on the consequences of an identified variant, or other aspects of the disease.

- If you, as a curator, have a team (both clinical and scientific) that is qualified and knowledgeable about the disease, the opinion of the team on the potential

consequence of a variant may be given, especially when the curator has assigned “concluded pathogenicity” to variants listed in their database.

- If a curator does not have a team and does not have in-depth knowledge about the disease, it is best to refrain from giving any opinion.

**Sharing variant information with genome browsers:** This increases visibility for LSDBs or GVDBs and should be encouraged.

**Managing personal data breach:** Curators may handle sensitive personal data. Where this happens or is suspected, steps should be taken to ensure personal data breaches are dealt with immediately incidents are identified.

- Prior to data curation, curators should identify who to notify in the event of a personal data breach and obtain further procedures on reporting.
- On discovering an incident involving personal data breach, a report of the incident should be filed with the appropriate contact.
- Establishments and/or infrastructure providers will have procedures that should be followed in the event of an incident involving personal data breach.

**Data erasure:** Complete erasure of publicly available data, presented in web-based databases, is impossible as the data may be distributed to cyberspace. However, where database curators have knowingly shared data, e.g. with genome browsers, and submitters have requested data erasure, those recipients should be informed of the request for erasure. The request for data erasure may be due to the withdrawal of consent. Further details on the specific grounds for request of data erasure can be found in the EU General Data Protection Regulation 2016/679 (2016).

When a request for erasure is received from a data submitter it is important that:

- The curator is provided with details that permit the identification of the data to be erased, e.g. the coded ID and a specific factor such as the genetic variant, to ensure the correct data is erased.
- The data is identified and deleted from the database.
- Where the data in question has been shared, e.g. with genome browsers, the recipient is contacted and informed of the request for erasure. However, it may not be possible for the curator to ensure that all the instances in third party systems are erased.

A brief description of the detailed Checklist is presented in Box 1.

In conclusion, database curators should maintain an ethical approach to their work and encourage their collaborators (i.e. data submitters) to adhere to the same standards. Ethical concerns in data sharing will not go away and database curators should acquaint themselves with requirements that enable ethical data presentation and sharing. A version of this Checklist is also available on the HVP website at

[http://www.humanvariomeproject.org/images/documents/HVP -  
Ethical Data Management- Checklist for Gene Disease Specific Database Curators -  
FINAL.pdf](http://www.humanvariomeproject.org/images/documents/HVP_-_Ethical_Data_Management-Checklist_for_Gene_Disease_Specific_Database_Curators_-_FINAL.pdf).

## **Acknowledgments**

We thank Professor Michael Parker for his participation in the working group, his suggestions and contributions; Helen Robinson and Amy McAllister of the HVP International Coordinating Offices (ICO), and Timothy Smith (formerly of the HVP ICO) for organising various stages of consultations, collating feedbacks and edits to the document.

Thanks also to all who made helpful suggestions for improvements to the Checklist: members of the HVP Gene/Disease Specific Database Advisory Council, the HVP International Confederation of Countries Advisory Council, the HVP International Scientific Advisory Council, the pan-European Research Infrastructure Consortium (ERIC) for Biobanking and BioMolecular resources Research Infrastructure (BBMRI-ERIC), and Professor Bartha Knoppers, Director of the Centre of Genomics and Policy, McGill University.

### **Funding sources**

RE is supported by The Tuberous Sclerosis Alliance (TS Alliance) and the Tuberous Sclerosis Association (TSA) for her LSDB work; MV is supported by the Swedish Research Council (Vetenskapsrådet) and the Swedish Cancer Society.

### **Conflict of interest**

The authors have no conflict of interest to declare.

### **Data Availability Statement**

Data sharing is not applicable to this article as no new data were created or analysed in this study.

### **References**

Al Aqeel, A.I. (2007). Islamic ethical framework for research into and prevention of genetic diseases. *Nat Genet.* 39:1293-8.

Celli, J., Dagleish, R., Vihinen, M., Taschner, P.E., den Dunnen, J.T. (2012). Curating gene variant databases (LSDBs): toward a universal standard. *Hum Mutat.* 33:291-7. doi: 10.1002/humu.21626.

Ekong, R., Nellist, M., Hoogeveen-Westerveld, M., Wentink, M., Panzer, J., Sparagana, S., Emmett, W., Dawson, N.L., Malinge, M.C., Nabbout, R., Carbonara, C., Barberis, M., Padovan, S., Futema, M., Plagnol, V., Humphries, S.E., Migone, N., Povey, S. (2016). Variants Within TSC2 Exons 25 and 31 Are Very Unlikely to Cause Clinically Diagnosable Tuberous Sclerosis. *Hum Mutat.* 37:364-70. doi: 10.1002/humu.22951.

EU General Data Protection Regulation 2016/679 (GDPR). (2016).

[http://ec.europa.eu/justice/data-protection/reform/files/regulation\\_oj\\_en.pdf](http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf).

Giardine, B., Borg, J., Higgs, D.R., Peterson, K.R., Philipson, S., Maglott, D., Singleton, B.K., Anstee, D.J., Basak, A.N., Clark, B., Costa, F.C., Faustino, P., Fedosyuk, H., Felice, A.E., Francina, A., Galanello, R., Gallivan, M.V., Georgitsi, M., Gibbons, R.J., Giordano, P.C., Harteveld, C.L., Hoyer, J.D., Jarvis, M., Joly, P., Kanavakis, E., Kollia, P., Menzel, S., Miller, W., Moradkhani, K., Old, J., Papachatzopoulou, A., Papadakis, M.N., Papadopoulos, P., Pavlovic, S., Perseu, L., Radmilovic, M., Riemer, C., Satta, S., Schrijver, I., Stojiljkovic, M., Thein, S.L., Traeger-Synodinos, J., Tully, R., Wada, T., Waye, J.S., Wiemann, C., Zukic, B., Chui, D.H., Wajcman, H., Hardison, R.C., Patrinos, G.P. (2011). Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. *Nat Genet.* 43(4):295-301. doi: 10.1038/ng.785

Global Alliance for Genomics and Health. (2016). GA4GH Data Sharing Lexicon.

[https://www.ga4gh.org/docs/ga4gh toolkit/regulatoryandethics/GA4GH\\_Data\\_Sharing\\_Lexicon\\_Mar15.pdf](https://www.ga4gh.org/docs/ga4gh toolkit/regulatoryandethics/GA4GH_Data_Sharing_Lexicon_Mar15.pdf).

International Committee of Medical Journal Editors. (2017). Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

<http://www.icmje.org/icmje-recommendations.pdf>.

Knoppers, B.M. (2014). Framework for responsible sharing of genomic and health-related data. *Hugo J.* 8(1):3. doi: 10.1186/s11568-014-0003-1.

Ludvigsson, J.F., Håberg, S.E., Knudsen, G.P., Lafolie, P., Zoega, H., Sarkkola, C., von Kraemer, S., Weiderpass, E., Nørgaard, M. (2015). Ethical aspects of registry-based research in the Nordic countries. *Clin Epidemiol.* 7:491-508

Mascalzoni, D., Dove, E.S., Rubinstein, Y., Dawkins, H.J., Kole, A., McCormack, P., Woods, S., Riess, O., Schaefer, F., Lochmüller, H., Knoppers, B.M., Hansson, M. (2015). International Charter of principles for sharing bio-specimens and data. *Eur J Hum Genet.* 23(6):721-8. doi: 10.1038/ejhg.2014.

Patrinos, G.P., Cooper, D.N., van Mulligen, E., Gkantouna, V., Tzimas, G., Tatum, Z., Schultes, E., Roos, M., Mons, B. (2012). Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain. *Hum Mutat.* 33(11):1503-12. doi: 10.1002/humu.22144.

Povey, S., Al Aqeel, A.I., Cambon-Thomsen, A., Dalglish, R., den Dunnen, J.T., Firth, H.V., Greenblatt, M.S., Barash, C.I., Parker, M., Patrinos, G.P., Savige, J., Sobrido, M.J., Winship, I., Cotton, R.G., Ethics Committee of the Human Genome Organization (HUGO). (2010). Practical guidelines addressing ethical issues pertaining to the curation of human locus-specific variation databases (LSDBs). *Hum Mutat.* 31(11):1179-84. doi: 10.1002/humu.21339.

Stoeklé, H.C., Turrini, M., Charlier, P., Deleuze, J.F., Hervé, C., Vogt, G. (2019). Genetic data, two-sided markets and dynamic consent: United States versus France. *Sci Eng Ethics*. Mar 12. doi: 10.1007/s11948-019-00085-4. [Epub ahead of print]

Takashima, K., Maru, Y., Mori, S., Mano, H., Noda, T., Muto, K. (2018). Ethical concerns on sharing genomic data including patients' family members. *BMC Med Ethics.* 19(1):61. doi: 10.1186/s12910-018-0310-5.

Vihinen, M., den Dunnen, J.T., Dalglish, R., Cotton, R.G. (2012). Guidelines for establishing locus specific databases. *Hum Mutat.* 33(2):298-305. doi: 10.1002/humu.21646

World Medical Association. (2016). WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks. <https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/>.

**BOX 1: The Checklist for LSDB/GVDB curators, in brief.**

1. Define the purpose of the database.
2. Define the database policy.
3. Offer attribution to submitters.
4. Establish an ethics oversight committee.
5. Collect data that has valid consent and coded ID.
6. Curate unpublished data to protect patient privacy.
7. Requests for non-public data should be forwarded to the data submitter.
8. Requests to keep submitted data non-public should be honoured.
9. Requests for submitter's details should be forwarded to the submitter.
10. You could consider giving your opinion if you have a team qualified and knowledgeable about the disease.
11. Information on variants can be shared with genome browsers.
12. Personal data breaches should be dealt with immediately.
13. Data erasure requests should be honoured.

**Appendix 1 – InSiGHT - *Ethical and Privacy Principles in relation to Responsible Sharing of Genomic and Health-Related Data (2018)***

An example from InSiGHT – how a framework approach can be helpful. A flexible approach assists database curators and their collaborators to address changing local and international changes in regulations, as well as changing community expectations. While policies in this area benefit from regular discussion and review for continuous improvement, this should not be an onerous task; a framework approach helps to keep the focus on improving outcomes.

<https://www.insight-group.org/content/uploads/2018/06/EthicsFramework.pdf>.

**Appendix 2 – An example of database policy** from ORAI1base (Variation registry for Severe combined immunodeficiency) at [http://structure.bmc.lu.se/idbase/ORAI1base/?content=db\\_policy/IDbases](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.*