

## Executive Summary

The development of biotechnology and bioinformatics affords the opportunity to store and analyse increasingly large amounts of genetic data. Genetic research involving the use of databases containing human genetic and genomic information, sometimes alone or in combination with other personal or medical information, has thus become increasingly important. More recently, the databases contemplated and being developed for genetic research are quite different in nature and larger in magnitude. Many of these emerging databases focus on and include data, information and biological samples from populations. These population databases, also referred to as human genetic research databases (HGRDs), may contribute significantly to science's understanding of the complex multi-factorial basis of diseases (genetic and non-genetic components) and therefore to improvements in detection, prevention, diagnosis, treatment and cure. Such databases may contribute significantly to the identification of genes associated with disease, an understanding of the frequency of genetic variants in particular populations, and to a better understanding of the reasons for drug reactions (both positive and negative) and reactions to other environmental factors.

These databases also raise a number of issues and concerns. While some of these are not new, the increasing breadth and scope of such databases amplifies them. Moreover, the combination of a broader set of genetic data and personal information in these databases raises new issues about the use of such information, especially in a non-clinical or non-research context. In addition, such databases are increasingly international in scope, covering populations from numerous countries, which raises new sets of issues.

Despite pressing concerns, there is limited international guidance on the establishment, governance and management of human genetic research databases. While certain institutions, such as UNESCO and the Council of Europe, have developed instruments which focus on the use of genetic data, these instruments do not address the multitude of issues raised by such databases. The underlying motivation for the OECD's involvement in this field is the need to address the development, use and access to population databases containing genetic data and personal/medical information. The aim of this Report is to begin the process of considering, at the international level, the policy challenges associated with the broader subject of the establishment, management and governance of human genetic research databases. Ultimately, this Report aims to provide a summary of the complex issues that need to be considered or addressed, in recognition of the significant contribution that human genetic research databases could play in translating scientific advances into innovation in health.

---

*Human genetic research databases*

---

The linchpin question is the determination of which databases ought to be considered a human genetic research database. There are numerous types of different databases including nucleotide sequences, sequences variations, mutation sequences, gene expression, gene loci, protein structures, as well as model organisms and diseases (*i.e.* pathology databases). Many of the databases currently being undertaken share the characteristic of collecting data, information and biological samples with the aim of allowing research for one or multiple diseases and conditions. However, the linchpin question is much broader. The consideration is whether only databases that contain data, information and biological samples for one or more “populations” or for large subsets of one or more populations ought to be considered HGRDs.

There are diverse other dimensions to the linchpin question. One dimension is consideration of whether biobanks/tissue banks that are being amassed, especially within the private sector, should be included in the definition of HGRDs. Another key dimension is whether “human genetic research databases” should be considered to include databases that also contain personal, medical or other data and information. This determination includes consideration not only of the type of biological samples or information that will be collected and stored within the database but also the determination of the source of that information. Genetic data have been broadly defined as “all data, of whatever type, concerning the hereditary characteristics of an individual or concerning the pattern of inheritance of such characteristics within a related group of individuals.” Genetic data do not necessarily include information derived from DNA or RNA specimens; they can be inferred from family history, medical records or phenotype. Genetic data may be distinguished from personal genomic data which have been defined as “detailed personal data derived from analysis of DNA specimens.”

Moreover, consideration should be given to whether HGRDs should include databases that contain personally identifiable information or only databases that contain data and information that cannot be associated with or result in the identification of individuals. Personal data have been defined as “any information relating to an identified or identifiable individual”. Genetic data may be collected from individuals in different manners, each associated with varying degrees of identifiability. This complex issue is related to those raised by anonymisation and linkability of data, which are considered separately.

---

*Establishment of a human genetic research database*

---

At the inception of an HGRD, a crucial issue is the nature and scope of the database. Also critical to the scientific legitimacy of such a database is ensuring that the sample population chosen is genetically representative of the population it is to serve. Ideally, all groups should be included in a given study. However, owing to financial and practical constraints, this is not always feasible. An important consideration is the selection of criteria that will ensure a careful and rigorous selection process which results in a representative database.

Another issue pertinent to the nature and scope of the database is whether or not children should be allowed to participate in genetic studies. Some have argued that young children should be excluded from genetic studies. This, however, would severely obstruct research on genetic diseases which occur early in life. Others have argued in favour of including children, but have often tied their support for inclusion to the issue of consent. Consideration should be given to whether or not children should be included in population database studies, and if so what are the appropriate safeguards.

The intended uses of the population database should also be determined at inception. This determination is especially important in order to ensure that appropriate information can be provided to the participants in the project. One of the obvious objectives in establishing HGRDs will be to carry out research. However, one issue will be whether the specific nature of the intended research may be determined or determinable at the time that the HGRD is established or at the time of the collection of biological samples, data and information. The degree to which this determination may be made will have implications for issues pertaining to consent, to communication with the community and to building public trust. Another set of determinations pertain to whether the biological samples, data and information collected in a database could or should be allowed to be employed for other purposes. Therein, a key consideration is whether HGRDs should be allowed to be used for non-scientific/medical research purposes. Examples of other purposes for which the contents of a database could be employed include the delivery of clinical genetic services, law enforcement, insurance, legal actions and identification purposes (*e.g.* for military or civil).

Different funding structures are available for the establishment of HGRDs: for-profit (*i.e.* private undertaking), not-for-profit (*i.e.* public undertaking) or mixed model (*i.e.* public-private partnerships). For example, the Icelandic Health Sector Database was envisaged as a for-profit endeavour, the Estonian Genomic Database was conceived as a mixed model, and the United Kingdom Biobank database was established as a not-for-profit undertaking. What criteria should be employed to determine whether an HGRD should be a public, private or mixed- model undertaking?

The collection of a large number of data and information about a given individual raises numerous privacy and confidentiality issues. Privacy is generally considered to mean the right to be left alone. In the context of genetic research, it could also mean the right not to know genetic information. Genetic information obtained in the research context raises unusual privacy concerns because it has the potential to generate information and knowledge beyond that which was originally sought, and because it raises the possibility that researchers obtaining the information might be obligated, in some situations, to provide that information back to the patients who contributed DNA samples.

Confidentiality connotes the notion of a professional keeping information private once it has been revealed by a client, and is part of a fiduciary relationship. In the context of genetic research, the concern is most often related to keeping genetic information that is collected in a research setting from third parties such as health insurers or employers. However, in the computer age, there are now also concerns about violation of databases or the sale of genetic information for marketing purposes.

Privacy and confidentiality issues may vary with the type of database. Database systems built for the purpose of finding answers to only one or a limited number of research problems may raise different considerations from broad population databases. In the case of specific research databases, the manner in which data are collected, stored and

accessed will be more targeted and more limited. In such cases, the storage of the information may be simpler (*e.g.* not connected to an external network). However, in the case of an HGRD, the information collected and stored in the database may be more broadly accessed. Conversely, the information contained in specifically-targeted databases may be more easily linked back to identifying information and therefore enable identification. Large population databases, by their very size, may reduce this type of risk. An important consideration is the principles that should be established to ensure that the participant's privacy and confidentiality are respected.

Another factor that influences the issue of privacy and confidentiality is the nature of the biological samples, data and information collected. Identified samples pose the most direct challenges to privacy and confidentiality. Thus, many researchers have chosen, or have been required, to use coded, unlinked or unidentified samples. Even though direct identifiers have been removed from coded samples, they may remain identifiable and thus still present privacy and confidentiality concerns. Researchers often find coded samples more valuable (even critically so, in certain types of research) than unlinked or unidentified samples because the linkage with identity provides a way to follow up with individuals in longitudinal studies. Therefore, it has been advocated that this type of linkage cannot, and should not, be completely prohibited. To ensure privacy and confidentiality of databases with coded material, technical and procedural computer security (including control and monitoring of access to, and transport of, data) may be essential. An important element is the measures that should be undertaken to ensure the protection of the data and information contained within databases.

In establishing a population database, public support is essential, given that participation is voluntary. This implies that the collection of biological samples, data and information and the inclusion in the HGRD of these depend on the consent of the donor. It may be difficult to estimate participation rates in projects where benefit is indirect, long-term and at the population level, especially in resource-poor communities or populations with different beliefs, cultures or languages. In order to build public trust, it will be important to bridge the distance between the research community and participants. However, the manner in which public support is elicited may vary considerably. In seeking to establish an HGRD, the methods employed to engage the public, the information that should be provided, and the manner for communicating it most effectively are important factors.

---

### *Data and sample collection and management*

---

The cornerstone of a human genetic research database is the data, information and biological samples collected and stored therein. At the outset, one important consideration is whether the data, information and biological samples should be unidentified (*i.e.* anonymous), unlinked (*i.e.* anonymised), coded (*i.e.* linkable or identifiable), or identified. Such considerations may have implications for issues of privacy and confidentiality as well as participants' involvement. Each of these approaches has advantages and disadvantages which must be assessed in light of the HGRD's objectives and purpose. For example, anonymous data may raise the least risk for breach of privacy but may not be as valuable for researchers, especially for longitudinal studies.

HGRDs raise issues with respect to ownership of the data, information and the biological samples collected. Irrespective of whether the database is a private, public or a

mixed-model endeavour, the issue of who should be able to claim title to the data, information and biological sample arises. The ownership issue includes not only consideration of the immediate question of proprietorship but also longer-term implications, for example for the database's functioning or any commercialisation being contemplated. Equally important is the issue of remuneration for the provision of the biological samples, data and information. Whether or not to remunerate participants, beyond simple reimbursement of basic expenses, is another important issue. It includes consideration of whether such activity is permitted pursuant to the applicable national or regional law, as well as whether it may affect the credibility of the HGRD, the researchers, and collectors as well as its representativeness.

Informed consent is one of the most complex issues for human genetic research databases. Informed consent has become the pillar for protecting autonomy in research involving human subjects. Within the medical/scientific field, informed consent generally presumes the ability to indicate clearly to the participant the use and purpose of the particular research activity. While this is feasible for purpose-specific research, the very nature of HGRDs renders the provision of this type of information difficult. Therefore, the issue becomes what constitutes informed consent within the context of an HGRD given that the purposes for which the data, information and biological samples are collected and the uses for which they may be employed, usually, may be described only in a general manner. Many have queried whether the traditional model of informed consent is applicable in the context of HGRDs or whether a new model/paradigm should be developed. For example, some authors have advocated blanket consent. Others have advocated a general consent for limited purposes, with the undertaking to return to the participant should the proposed uses go beyond those limited purposes.

Additional consent questions raised by HGRDs include children's consent and renewed consent. In situations where a determination has been made about the involvement of children in genetic studies and HGRDs, consideration of the consequences of their involvement and the manner in which to obtain consent is primordial. For young children, obtaining consent may involve obtaining the consent of the parents. For older children, consent may involve a variety of approaches depending on their level of development and understanding. Issues that arise in the context of renewed consent include returning to the participant to obtain consent for new uses in the context of the HGRD. For example, the question arises of whether a new consent is required when existing databases are converted to HGRDs.

Re-contacting participants, in any of the above-mentioned situations, raises practical difficulties (*e.g.* the person may have died or moved away) but also more complex issues, such as whether or not the person wished to be re-contacted. This question involves consideration of a principled approach to re-contacting participants. As well, consideration should be given to whether participants ought to be provided with information pertaining to the issue of re-contacting prior to obtaining their initial consent.

Having voluntarily contributed to an HGRD, participants may, at some point wish to withdraw from the study and have their biological samples, information and data destroyed. One issue is whether HGRDs' should adopt a policy pertaining to participants' right to withdraw their data, information and samples and related issues, including whether this is possible and if, so under what circumstances. In some cases, it may be possible for participants to withdraw their data, information and biological samples throughout the duration of the HGRD. In others, depending on the manner in which the HGRD is established, it may be possible to withdraw data, information and biological

samples only prior to their anonymisation. Moreover, the right of withdrawal may entail various options. For example, if the participant's data have been included in information provided to a third party, it may not be possible for these to be selected or withdrawn.

In many epidemiological studies, information or results are provided back to either the database and/or the participant. However, given the breadth of an HGRD, the question is whether or not the providing back of results is feasible or desirable. First, there is the issue of whether or not results from users of the data and samples should be returned to the database. While the providing of results back to the database may enrich it overall, quality assurance of the results provided and included in the database is an important consideration. A second set of issues pertain to the providing back to the participants results derived from the use of the data and biological samples contained in the HGRD. Given the breadth and purpose of HGRDs, the issue is whether it is feasible to contemplate a policy of providing results back to participants and the value of doing so, especially if provided outside of a clinical context.

The training of health-care professionals and researchers will be important to the success of an HGRD. Health-care professionals responsible for the recruitment of participants and collection of the data (*i.e.* interviews, questionnaires, medical examination, drawing of blood, transfer of the information collected) may not be familiar with genomic research. Thus, a policy for training such health-care professionals may be important. For example, such a policy may include a protocol explaining their role to general practitioners, outlining the information they may divulge to the participants, and guiding them in the appropriate handling of contentious situations.

---

#### *Database management and governance*

---

The governance of databases involves numerous operational, technical and legal issues, including consideration of applicable legislation and regulation, the role of ethics and oversight committees, the powers, compliance and enforcement attributes granted to the HGRD, the security features of the database, access to the database and the demise of an HGRD.

In the establishment of an HGRD the question arises of whether the database should result from a statute or be independent of an act of parliament. For example, the Estonian Genomic Database and the Icelandic Health Sector Database would be the creation of an act of parliament. Conversely, the UK Biobank and the Canadian CARTaGENE initiative were established independent of specific legislation but are subject to a number of existing legislation. There are advantages and disadvantages in establishing a database via an act of parliament versus a softlaw instrument, such as a memorandum of understanding.

A review of most HGRD initiatives reveals that they require some form of oversight committee, but the composition and formation vary across projects. The role, function and nature of the oversight committee in the governance of an HGRD should be addressed. The composition of the oversight committee, especially whether it should be multi-disciplinary, the term for an individual's appointment, a policy or approach for determining which issues should be submitted for the consideration of the oversight committee are all important issues. For example, the Ethics Committee of the Estonian Genomics Database ensures compliance with ethical guidelines, and any requests on behalf of projects seeking access to the Database may be submitted for its consideration.

An important set of issues for HGRDs concerns the powers, the ability to ensure compliance and the mechanisms for enforcement of decisions. Such powers, or the lack thereof, may have implications for ensuring that privacy and security policies are respected, and for ensuring that a private entity to which commercialisation rights are granted respects these rights and does not act in an abusive manner. For example, since the Icelandic database would operate by licence, the power to revoke that licence is a means of effecting compliance. If any of the provisions of the operating licence or the enabling legislation are violated, the Minister can issue a written warning, and set a deadline by which action must be taken. Inaction or intentional and gross negligence may result in revocation of the licence.

Given the potential for misuse of data and samples collected in HGRDs, the security of the database is primordial. This is both a legal and a technical question. Given the objective of such databases, considerations of the best methods for ensuring security, ensuring that access occurs only in the permitted manner, and ensuring that access to the data and samples is not stifled are significant. One method proposed is through custody of a code registry. This method would be most valuable for protecting the confidentiality of data in databases wherein links are maintained between data and personal identifying information. In this approach, the custodian could be a person who is under the duty of non-divulgence of confidential information, such as a medical doctor. An additional security feature of this approach may be the use of stand-alone computers for handling personal identifiers and other personal information, including health data, so as to reduce the risk of unauthorised access of networked computers.

Another approach to ensuring security and protecting privacy is to limit the amount or type of data released or accessible to researchers using the HGRD. This may involve a combination of legislation and technical solutions. For example, it could include limiting the bin size or ensuring that data and information are not made available to researchers if the data sample does not involve more than a certain number of individuals. Similarly, privacy and confidentiality may be protected by limiting or monitoring access to the data. A simplified way of doing this is to allow only researchers with approved passwords to access a database. A more sophisticated version of this strategy is to use rule-based control of access to data, instead of, or in addition to, human intervention. This method would allow different users to access different categories of data and information according to their roles. An alternative is to allow a very limited set of analysts to query the primary data directly. In this scenario, “outside” researchers would be allowed to query the data only indirectly, through these analysts, and would only receive summarised answers to their queries (*e.g.* means, P-values, etc.).

A common method for enhancing data privacy is encryption, most often either one-way or public-key. This method may also be used in conjunction with the methods described above. Since encryption can be fairly easily implemented, it is assumed that data transferred to and from research databases will be encrypted in some way. However, it must be recognised that encrypted data can also be decrypted, so encryption should not be relied upon as the only means of privacy protection.

Given that the fundamental purpose of HGRDs is to foster research, access to the database raises crucial issues. Key questions for consideration include who should have access to the database (*e.g.* only researchers, public- and/or private-sector researchers), the manner in which access should be given (directly versus via an internal researcher), whether access should be free or for a fee (who should pay the fee and what should it be), and to what should access be given (*e.g.* to the whole database, to parts and which ones,

only to certain data and then only in an anonymised manner). Another key issue is the purpose for which access should be granted.

While there are few examples of HGRDs that failed or were terminated (for example, the Tongan database that was to be established by Autogen Limited), the possible demise of an HGRD should be considered as early as possible, especially when establishing its governance structure. Consideration should be given to clearly determining the consequences of its demise. This could include, for example, whether all of the data and the samples are to be maintained or destroyed and whether participants should be notified of the demise of the database. If the database is operated by a private undertaking, consideration of whether provision should be made for a government to retain the right to have the database handed over to them or whether a government should reserve at least a right of first refusal. Such considerations will be influenced by the applicable legislation. For instance, many countries have enacted legislation prohibiting the sale of human tissue or materials. The consequence of the application of such statutes would need to be taken into account.

---

### *Commercialisation considerations*

---

HGRDs raise issues with respect to commercialisation, including intellectual property, the actual commercialisation of the database and/or its components and benefit-sharing.

Intellectual property broadly refers to the legal rights which result from intellectual activity in the industrial, scientific, literary and artistic fields. One set of issues raised by HGRDs pertains to the intellectual property rights (IPRs) that may arise as a result of research employing data or samples accessed from a database. In such circumstances, questions arise of who owns the invention and who is under an obligation to ensure that the relevant IPRs are protected. Another important consideration is access to a follow-on innovation developed using data and samples from the database. A policy that would balance follow-on access while permitting a return on investment would also be an important consideration. Intellectual property issues may also arise with respect to the database itself, including database rights, where they exist, copyright protection for the software and other rights that allow the database to operate effectively.

With respect to commercialisation, the first consideration is whether or not it is desirable to commercialise the database and/or whether commercialisation is in line with the participants' expectations. If commercial exploitation of the database is undertaken, consideration should be given to the manner in which this should be carried out. For instance, commercialisation could be on an exclusive or non-exclusive basis. If the commercial exploitation were to be undertaken on an exclusive basis, it would be essential to consider how to ensure fair access to the database and how to ensure compliance with competition law. The question of whether or not the database could be sold or transferred for consideration would also be important.

The issue of benefit sharing is a complex one with many aspects that vary depending on the structure of the database. For instance, in the context of an HGRD established as a public-private partnership or as a private undertaking, consideration should be given to whether the government should be entitled to some compensation from the private entity and the form of such compensation. In the case of the Icelandic Health Sector Database, for example, the Licence Agreement would have required payment to the government of

*i*) an annual fixed fee, earmarked for the promotion of health care and R&D; and *ii*) 6% of the profits of the Licensee, capped at ISK 70 million a year. If monetary compensation is the option favoured, the use made of these sums would be an important consideration. Moreover, compensation may also take a non-monetary form, such as technical or scientific support. Benefit-sharing also raises the issue of whether or not the participant should be entitled to individual benefits arising from the database. For example, whether participants would be entitled to share in the profits of a successful invention developed using samples, data and information contained in the database. A further issue is whether participants should be given the right to access other, non-monetary benefits, such as the products developed as a result of outside research but involving data and samples from the HGRD.