

GLOBAL GOVERNANCE IN GENETIC, GENOMIC, AND PROTEOMIC DATABASES - A RIGHT TO PRIVACY AGAINST A SOCIETAL GOOD

HEDLEY CHRIST¹

ARTICLE INFO

ABSTRACT

Ever since the mapping of the human genome, the amount of human genetic data now being collected has been called a tidal wave of data. Such data is being stored with the principal aim of using such data for research, primarily in genetic diseases, but not exclusively. Such databases become more powerful when they are linked, because of the increased number of DNA sequences that can be searched. However, this provides for significant problems of management and governance of these databases, not least because they hold genetic information on identifiable individuals and therefore, there has to be control over access to these databases. But, beyond such issues lie a number of legal problems which relate to patients' rights and patients' duties to society and medical research; questions of ownership, not only of the databases themselves, but also the genetic information stored in such databases, particularly with the issues related to intellectual property rights. This brief paper examines the need for governance of such databases, principally through soft law techniques of international regime analysis.

KEYWORDS

Genetic, genomic, and proteomic databases. Global governance. Right to privacy. Societal good.

The concept of human rights in patient care has wide application. It includes bioethics, patients' rights, right to health, and patient safety. But beyond this there is a societal good in that, information gathered from patients may help in providing cures for specific diseases or conditions. Human rights in patient care therefore, addresses wider rights, including the benefits to other patients than the one undergoing investigation and treatment. However, this inevitably encompasses a conflict between the right of the patient and the information that they provide, particularly when the information they provide is their own genetic information². The European Convention on Human Rights and Biomedicine provides a number of rights and protections to individuals who provide genetic information useful for research, however, this does not address the issues of storage and the provision of access to this information for research purposes. Furthermore, this form of information as data is significantly different from that conceived of in the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. We therefore, need to consider how genetic information, provided by patients, can be governed, such that, both protection is given to the patient and researchers have the opportunity to use such information in medical research.

¹ University of Brighton.

² The genetic information provided is specific to the individual, and thereby may readily identify them and their health issues.

THE NATURE OF GENETIC INFORMATION

Since the mapping of the human genome the amount of genetic sequence information now being generated has been called a tidal wave of data. These data are being collated and stored in genetic, genomic, and proteomic databases³ within a computer-assisted data management system known as bioinformatics. As more and more genetic and protein information is generated, the intention is to link these databases, both nationally and internationally, to other genetic, genomic, and proteomic databases and/or public health information systems, leading finally perhaps to a few international mega-database systems. This network of database systems will then be able to provide vast genetic and protein data and information important in the development of medicine, both in research (pure and applied) and clinical practice. However, genetic, genomic and proteomic databases contain information about individuals' and their families' health which has long been considered as sensitive information with strict confidentiality rules. Fundamental to the development of these databases will be the balance to be drawn between an individual's privacy rights and the social welfare in health that these bioinformatic systems will be able to provide. The notion of this balance is compounded by the development of intellectual property rights within these genetic, genomic, and proteomic databases. This network of national and international databases will, therefore, need to develop management and governance within an international framework. Such management and governance will need to develop within an international regime to enable the efficient, effective, ethical, and best use of this network system of genetic, genomic, and proteomic databases. However, there is no clear understanding of who the actors are in producing and controlling the development of bioinformatic systems, what type of databases are being produced and whether and to what extent these databases are being linked. It is important, therefore, to address these issues and consider what will constitute the agenda-setting issues of the developing international regime in genetic, genomic, and proteomic databases, particularly in relation to individual privacy, intellectual property, and social health welfare.

³ The range and type of these databases are numerous. International collaborations have provided three primary sequence databases.

- GenBank, maintained by the National Center for Biotechnology Information (NCBI).
- Nucleotide Sequence Databases, maintained by the European Molecular Biology Laboratory (EMBL).
- DNA Databank of Japan.

These primary sequence databases have subsidiary databases for the storage of particular types of sequence data. These include:

- dbEST for expressed sequence tags,
- dbGSS for single pass genomic survey,
- dbSTS for sequence tagged sites, and
- HTG for high-throughput genomic division used to store unfinished genomic sequences data.

Two important protein sequence databases are:

- SWISS-POT, and
- TrEMBL.

There are a range of specific databases including, rDNA, tRNA, Promoter Sequences, Regulatory Elements, and Inbase, a database of inteins which are small peptides that are spliced out of some microbial protein. OMIM (Online Mendelian Inheritance in Man) is a database of human genes and genetic disorders maintained by NCBI. Incyte is a commercial database containing DNA sequences, transcripts, extensive annotations, expression data and access to cDNA for experimental studies. There are, therefore, a vast number of databases widely distributed. The UK Biobank created between 2006-2010 has collected data from 500,000 people aged between 40-69 across the United Kingdom. Within the EU the EGA provides a service for the permanent archiving and distribution of personally identifiable genetic and phenotypic data resulting from biomedical research.

WHAT ARE GENETIC, GENOMIC AND PROTEOMIC DATABASES AND WHY IS IT IMPORTANT?

Genetic, genomic, and proteomic databases are a computer-assisted data management system that gathers, and analyzes genetic, genomic, and proteomic data. It stores and searches⁴ data on gene and protein sequences⁵, structures⁶, expression profiles, and biochemical pathways. The importance of these databases is that it analyzes genetic and proteomic sequences and structures, enabling an understanding of how genes and proteins influence the working of the body at a molecular level. It enables biological data to be used to understand the higher-level functions of the cell, such as biochemical pathways, regulatory networks, signal transduction pathways, and what influences the behaviour of cell, organ and organisms. It enables the modelling of target proteins' interactions with drug molecules and the way individuals are affected by the interaction between their genetic make-up and their environment or life circumstances (UK HOUSE OF LORDS, 2001). In 2001 the UK Parliament, House of Lords Select Committee on Science and Technology on "Human Genetic Databases" recognized that understanding the genome will bring substantial improvements in medicine, particularly through the use of genetic data⁷ and medical histories.

Over the last twenty years the development of genetic, genomic, and proteomic databases, along with combinational chemistry, has made a significant contribution to the understanding of disease and drug development. Both genomics and proteomics have revolutionized the way target molecules are identified and validated for drug targeting. Thousands of potential new targets can now be identified by sequences, structures, and functions. The contribution of these databases, particularly in medicine, is their ability to analyze sequences and structures, enable an understanding of how genes and proteins influence the working of the body at a molecular level, and model target protein interactions with drug molecules *inter alia* and is being practised worldwide by academics' groups, companies, and national and international research consortia.

WHY DO WE NEED GLOBAL GOVERNANCE - THE PROBLEM

Genetic, genomic, and proteomic databases and their database management systems are resources that must be managed. Data ownership⁸ and the legal regimes which regulate their control are vital in determining the availability and control of the information which is stored, manipulated, and disseminated. This is particularly important for these databases where "Information about individuals and their health can be very sensitive and genetic data are particularly so... Information will say something about family members as well as the individual directly concerned... The challenge is to

⁴ There are a range of retrieval tools which allow a text-based search within a number of linked databases. Most widely used are Entrez, DBGET, and SRS. Sequence searches can be done with BLAST or FASTA.

⁵ Sequences are derived from DNA, RNA, and Proteins. Genomic DNA is taken directly from the genome, e.g. its natural state and therefore contains introns, regulatory elements. cDNA is generated by reverse transcribing of mRNA. rDNA includes the sequences of vectors (e.g. Plasmids), modified viruses and other genetic elements.

⁶ Structures related to proteins and nucleic acids.

⁷ The UK Medical Research Council and The Wellcome Trust have launched the UK BioBank. The aim of this BioBank is to collect genetic data and health statistics from 500,000 volunteers.

⁸ In database systems the data site is often termed master site (or primary site) which makes data available to slave sites (or subscribers). A master site may own the data; however, there may also be multiple sites in which ownership is invested in distinct fragments.

find a way of protecting the interests of individuals while at the same time making essential information available to medical research” (UK PARLIAMENT HOUSE OF LORDS, 2001). A principal concern is, therefore, that these data systems must be managed in such an efficient way as to bring social welfare benefits to the international community while at the same time respecting the informational privacy of the individual. However, set against this is the demand for open science. Organizations campaigning against any notion of ownership of biological information are working to develop a public or open licensing plan for information⁹. In contrast to open science is the issue of intellectual property (IP) in these databases and how such IP should be managed.

During February and March of 2004 the OECD held a Workshop on Human Genetic Research Databases (HGRDs) - Issues of Privacy and Security (Tokyo Workshop). Recognizing that HGRDs are invaluable tools which will have immense possibilities for medicine, the Workshop felt that clear procedures must be in place for informing patients about the way that data based on their genetics might be used, whether current informed consent is sufficient to assure patients’ privacy and achieve an appropriate balance with research and access. Whether or not such a balance is achieved in public policy will affect how successful genetic science is as a driver for innovation products and processes and delivery of better health. The Workshop also concluded that the OECD should develop principles of best practice for the management and governance of HGRDs, but as yet has been unable to provide such principles of best practice.

The reason for this lack of management and governance guidance of HGRDs lies in the complexity of the task-in-hand. In part the OECD acknowledges: “The protection of individuals’ genetic information relies on a combination of health-related, confidentiality, and/or personal data protection laws¹⁰. Many countries also provide constitutional protection or human rights legislation. In general, however, no specific laws distinguish the processing of genetic data from other personal or sensitive information. Yet genetic data is perceived as being special because it can reveal important information about both an individual and his/her family and can have a significant impact on an individual’s life, including his or her reproductive choices”.

Genetic, genomic, and proteomic databases may also play a significant role in public health management. As GOSTIN (1996) has pointed out, the development of public health information has produced a vast reservoir of information on health status giving states a means of health surveillance. Public health investigation through epidemiological research, testing and screening for disease, and now genetic and proteomic databases “enable the public health systems to identify health problems, inform the public, intervene, and influence funding decisions”. However, like the OECD Gostin also recognizes: “Patients, often physically and mentally vulnerable, divulge intimate details of their lives to their physician, medicine’s paternalistic

⁹ Bioinformatics.org and the Open Lab offer web hosting and project support relating to bioinformatics. The projects within the Open Lab are primarily end-user software tools for scientists looking to solve particular biological and bioinformatics problems. The Distributed Sequence Annotation systems (DAS) developed at the Cold Spring Harbor Laboratory and Ensembl are both projects intended to bring the human genome into the public domain. Cold Spring Harbor Laboratory is looking at using MP3 players as a means to sharing genetic information, peer-to-peer.

¹⁰ In the UK, the construction of a genetic, genomic, or proteomic database has to receive ethical committee approval before it can be developed. Such approval will only allow the database to be constructed on the basis of the application for ethic approval. Any alteration, development, or linkage not in the initial application requires further ethical committee approval. These ethical committees are particularly concerned with patient consent which cannot be ‘blanket’ consent for any use of the database.

traditions have long-recognized that the patient's weakened position compels strict confidentiality assurances even in the face of government demands... Law and ethics in the late twentieth century emphasize autonomy as a theoretical justification for privacy, patient autonomy encompasses the right to control the dissemination of person health information... Confidentiality is central to a trusting physician/patient relationship - this promotes patient's candour about health and disease risks".

This conflict between an individual's informational privacy and the disclosure and access to genetic and proteomic data is compounded by the introduction of intellectual property and the possibility of a commercial application to these databases. In the final report from the OECD Working Group on Biological Information it stated: "The actual holding of biological materials clearly leads to the control of the access to related information and, in practice, also to the control of any possible invention that could derive from those materials. In this perspective, the problem of intellectual property of biological information is also connected to this problem of ownership of media where that information is stored".

In the European Union, many EU countries have implemented the 1996 Database Directive (96/9/EC) which gives an intellectual property right in databases. As BOVENBERG (2000) states: "The database right in a DNA database would almost undoubtedly represent valuable intellectual property, especially with the uncertainty that still surrounds DNA sequence patenting... The database right allows its holder to prevent extraction or reutilization of substantial parts of the database for fifteen years after completion of a database. In fact, the right may pertain in perpetuity given that it can be rolled over following any substantial modification to the database that requires a substantial investment. Such modifications include the routine business of DNA databases: extension, deletions, and amendments". The UK Chorley Report identified a wide range of public datasets which could have commercial application and where possible should be commercialized to recoup the development costs. Many states now recognize the potential for the commercialization and intellectual property exploitation of public research.

As more and more genetic and protein information is generated with the purpose of linking this information either to other genetic, genomic, and proteomic databases or other types of databases such as clinical records, the need to manage these resources efficiently and within strict legal boundaries has become paramount. Governments and international organizations are driving greater and deeper linkages within these databases at the same time as demanding tight controls over informational privacy and intellectual property rights. Central to this debate is the issue of ownership. The contribution that these databases can make to medicine, both in research and in clinical practice will depend on the network system that is created for it. This network system will be of an international character bringing about both economies of scale and economies of scope. Good governance should bring about economies of scale through efficiency gains in coordinating rulemaking, enforcement activities, and the acquisition of specialized skills and organizations, while reducing unnecessary regulatory disharmony. Economies of scope will bring about reductions in costs resulting from centralized access points and wider benefits to the international community.

Within this framework are national legal systems (developed in a pre-digital age), which were not designed for genetic, genomic and proteomic databases. In fact, these legal systems may positively inhibit the efficient management of these database systems. Furthermore, this still emerging technology requires international cooperation and an understanding between a wide range of actors (epistemic community) as to the

nature of efficient database regulation. If an international regime is to be developed then it is necessary to identify the actors concerned (the epistemic community), the range of and type of genetic, genomic, and proteomic databases being produced, and identify the legal issues that will constitute the agenda setting.

REGIMES AS A MEANS OF GOVERNING GGP DATABASES

The concept of regimes was conceived as an issue area in which relevant actors share the principles, norms, rules and determine how decision-making processes are made and implemented¹¹. However, little attention has been paid to actors, other than government or governmental actors, despite considerable evidence that in many regimes other types of actors play a crucial role (ROSENAU, 1999). RUGGIE (1975) argued that the existing literature on technology change and international cooperation was inadequate because of its restricted focus on law and organization. “What was required was a wider view encompassing implicit understanding between a whole range of actors”. More recently VOGLER (2000) stated “A form of regime analysis that has been relatively neglected is the fundamental one of how agendas are set and issues arise, altered and are aggregated together. Who defines what this social construct - the issue area - will be? Who are the actors?”

Such views are particularly important for genetic, genomic, and proteomic databases. Throughout the international molecular biology community many genetic, genomic, and proteomic databases are being produced with an agenda to integrate these databases. The result could be a few international databases accessible by national public health systems and an international research community. However, it is not clear how many research groups are producing genetic, genomic, and proteomic databases? How many of these databases are already linked, either to other genetic, genomic, or proteomic databases or public health records such as clinical records? How many of these databases are linked nationally and/or internationally? What type of genetic, genomic, and proteomic databases are being produced, i.e. specific disease types, protein interaction drug targeting, and diagnostic testing, and for what purpose? Who has access to these databases, and who has control over the development, use, and access to these databases?

Just as important as understanding who the actors are in this developing technology are the issues that will need to be addressed by these actors as these databases become more and more linked, particularly on an international level. It is already apparent that, apart from the issues of standardization, these issues will principally surround the balance to be drawn over individual privacy rights, intellectual property rights, and social welfare. However, it is not apparent how these agenda-setting issues affect the epistemic community. What are the principal concerns the actors have in developing genetic, genomic, and proteomic databases? How do these concerns manifest themselves when these databases are linked? What do the actors believe will be the important issues that will determine whether an efficient, effective international database system will be developed, and how will an international system be managed, monitored, controlled, developed, and accessed?

The policy choices open to the international community wishing or needing to take action to manage genetic, genomic, and proteomic databases fall into a number of

¹¹ International regimes were first defined by KRASNER (1982) as a “set of implicit or explicit principles, norms, rules, and decision-making procedures around which actors’ expectations converge in a given area of international relations”.

categories. These include regulation, persuasion (often considered as soft law), the use of property rights, the use of targeted economic instruments, and adjusting direct or indirect economic policies that do not have as their goal the efficient regulation of these databases, such as the overriding aim of commercial exploitation of these databases. Alongside these policy choices is the risk of fragmentation, that is, where policies develop on individual lines lacking the coordination that efficient regulation will require. It will be necessary, therefore, to consider which policy variables are most likely to affect the efficient regulation of these databases set against the desired goals and concerns over fragmentation.

Set within the overarching aims and objectives: what is the typology and linkage of the genetic, genomic, and proteomic databases, who forms the epistemic community and what is the nature of their capacity, and what are the agenda setting issues for an international regime, will be the policy options. Principal among these, are the policy choices which can be taken within the framework of international regime analysis will be:

1. Epistemic community. The policy choice of who constitutes the epistemic community will have a strong influence on the nature, form, structure, and management of the genetic, genomic, and proteomic databases, i.e. certain interest groups may impose externalities on others, how interests are balanced, and how a wide epistemic community may not understand the uncertainties that actors commonly face regarding the nature of the particular issues in everyday management.
2. Network Externalities¹². Linkage of genetic, genomic, and proteomic databases can be enhanced by the network effect. However, an open lab approach would severely restrict the network effect. Which policy option is taken will have a major impact on a developing international regime.
3. Form of Approaches Taken. Three approaches may be taken to form an international regime: Framework Convention Pathway¹³, Plurilateral Pathway¹⁴, and Soft Law Pathway¹⁵. Of these policy options, which is taken will determine the form, structure, and type of agreement within the agenda setting issues.
4. Fragmentation. Resulting from “market failure”, i.e. conflict of interest among states¹⁶ and conflict of interest across states¹⁷, the approach of agenda setting issues must develop policy choices of cooperation, particularly within the distinct areas of database usage, privacy, and intellectual property, which will coordinate rather than contradict one another.
5. Content. In developing an agenda setting programme for an international regime what should constitute the substantive content governing the area of international genetic, genomic, and proteomic database regulation. This reflects not only the

¹² These arise when a good or service is more valuable to a user the more users adopt the same good or service.

¹³ Begin with an agreement that has broad participation and is at least moderately legalized but includes only shallow substantive commitments, and deepen the substantive content over time.

¹⁴ Begin with an agreement that includes deep substantive commitments and is highly legalized, but has limited membership, and expand participation in the agreement over time.

¹⁵ Begin with an agreement that contains significant substantive commitments and has wide participation, but is not highly legalized, and strengthen legalization over time.

¹⁶ Conflicts that preclude states from agreeing on certain practices as legally binding.

¹⁷ These conflicts have been particularly noted in international environmental law in which certain domestic interest groups impose externalities on neighbouring countries.

policy choices of form of approach taken, and elimination of fragmentation, as above, but also on the nature of the obligation it provides and on the delegation to third parties (WHO, OECD, or newly created institution).

6. Constraint. Certain policy options have already been undertaken by Governments covering the subject matter of database usage, privacy, and intellectual property (note fragmentation above). These will constrain the type of options available and it may be unrealistic to expect the desired policy changes to happen at once. It might be appropriate to take a longer-term view entering into “policy dialogue” with the epistemic community perhaps involving the exchange of information at a professional and technical level.

These policy options and choices will fundamentally determine the nature, form, structure, and overall efficient regulation of an international regime in genetic, genomic, and proteomic databases and how these databases will be used for research, clinical practice, and public health welfare regulation. Furthermore, they will determine the balance of public and private involvement of this international regime.

REFERENCES

BOVENBERG, J. A. (2000). “Should Companies Set-up Databases in Europe?” *Nature Biotechnology*, 18, 907-909.

GOSTIN, L. O. *et al.* (1996). “The Public Health Information Infrastructure: A National Review of the Law on Health Information Privacy”. *Journal of the American Medical Association*, 275, 1921-1927.

KRASNER, S. D. (ed.) (1983). *International Regimes*. Ithaca and London: Cornell University Press.

ROSENAU, J. N. (1999). “Towards an Ontology for Global Governance”. In M. HENSON & T. J. SINCLAIR. *Approaches to Global Governance Theory*. New York: State University of New York Press.

RUGGIE, J. G. (1975). “International Response to Technology: Concepts and Trends”. *International Organizations*, 29, 557-583.

UK PARLIAMENT (1987). *Select Committee Report into the Handling of Geographic Information*. The Chorley Report, London: HMSO.

UK PARLIAMENT (2001). *House of Lords Select Committee on Science and Technology. Human Genetic Databases: Challenges and Opportunities*. London: HMSO.

VOGEL, J. (2000). *The Global Commons. Environment and Technology Governance*. Chichester: John Wiley & Sons.