Semantic Interoperability and Biobanking – The Politics of Setting Technical Standards in Tissue Economies

Aaro Tupasela

Abstract

During the past decade a great deal of effort and funding has been given to the development of national or regional biobanking initiatives. The goal of many such initiatives is to study disease and thus provide the basis for developing treatments and medicines. More recently, however, local biobanking initiatives and the data that can be derived from them are increasingly being embedded into a global network of biobanks. This process requires the development of technical standards to facilitate the inclusion of various actors and the exchange of data. This enrolment of actors and data on the human body is inherently a political one in that it relates to forms of control and power that researchers and national governments seek to exert over medical and lifestyle information on the human body. In this chapter, I will look at the way the notion of semantic interoperability allows for such data exchanges while exhibiting the operation of power relation between actors participating in the exchange. From this I will derive some conclusions as to the direction and form that current and future biobanking initiatives will take in terms of practices in the biomedical sciences.

Introduction

> The challenge lies not only in data harmonization and constant update of clinical details in various locations, but also in the heterogeneity of data storage and confidentiality of sensitive health-related and genetic data. Solid infrastructure must be built to provide secure, but easily accessible and standardized, data exchange also facilitating statistical analyses of the stored data (Muilu et al. 2007, 2).

During the last ten years there has been a surge in the amount of national and regional biobanking efforts that have been launched around the world. These initiatives include DeCode Genetics in Iceland, the UK Biobank,

Generation Scotland, CARTaGENE in Quebec, Egeen in Estonia, the NIH initiative in the USA and UmanGenomics in Sweden, just to name a few.[1] Some of these ventures are public, others a mixture of public and private, while some are strictly private. All share, however, a common goal of collecting new samples and analyzing existing tissue samples along with healthcare and lifestyle information to ascertain the molecular and environmental causes behind many of today's common diseases, such as heart disease and diabetes.

The use of large human tissue sample collections has become an important tool in biomedical research because of the versatility in providing new information not only about the human body and disease, but also about populations (Collins et al. 2003). Genetic databases or population databases are electronic databases containing information and data that has been gathered or extracted from physical tissue samples using various methods. These databases are derived from collections of human tissue samples or biobanks. Austin et al. (2003, 37) have defined biobanks as 'a stored collection of genetic samples in the form of blood or tissue, that can be linked with medical and genealogical or lifestyle information from a specific population, gathered using a process of generalized consent'. There is, therefore, an important link between the physical material collections of tissues and the data that is derived from them.

Although the size of the collections vary from one cohort to another, it is becoming increasingly clear that regional and national collections offer only a limited 'snapshot' of local or national communities and that in the future it will become necessary to combine and compare such collections with one another. Indeed, a scientific debate has even emerged as to what should be the size and composition of the ideal population, where some argue that supposedly homogenous populations, such as in Iceland or the population isolates in Finland provide the best scientific material (see Palotie & Peltonen-Palotie 2004). Others, however, such as in Estonia, argue that heterogeneous populations will provide the best (representative) genetic material for analyses of complex disease traits (see Petrone 2003). Kere (2007, 864), however, has argued that current research using genetic association studies are difficult to publish unless several different sample collections, preferably from different countries, have

been combined. This change in the perceptions concerning the validity of national collections indicates the need for researchers to begin developing networks of research cooperation, which invariably integrate national collections into this network. This integration of collections and actors serves as the basis through which samples and their related information become enrolled in an ongoing project of scientific knowledge production and ultimately the production of various forms of biovalue (Sunder Rajan 2006; Waldby 2002).

In the study of simple or monogenic disease small collections have served an important purpose and led to the identification of numerous genetic causes behind such diseases. This has also given rise to the development of new industries around genetic diagnostics. The study of complex or multivariate diseases, however, has invariably brought forth the need to combine smaller sample sizes for comparative studies. Given the fact that many diseases may be the result of numerous factors, not just genetic, but environmental and lifestyle as well, researchers are finding it necessary to conglomerate smaller samples to form larger ones in order to gain a better level of statistical accuracy and reliability. This necessity is driving biobanking towards increased cooperation between research groups and initiatives which have collected large and small sample collections to combine their data for statistical analysis. Given the increasingly international scope of biobanking, it has become necessary to develop common standards for biobanking practices (ECVAM 2002). Standards play an important role in facilitating transfer of materials, as well as harmonizing common practices from one context to another (Bowker & Star 2000).

The processes of standardization of practices, however, are taking place at multiple levels. One of the main areas where standardization has been taking place has been in the development of common global social, ethical and legal (ELSA) frameworks around which biobanking activities can be organized. Examples of such standards in regulation include the Council of Europe's Convention for the Protection of Human Rights and Dignity of the Human Being (ETS 164).[2] Standardization practices also extend to the everyday activities associated with the technical side of collecting, handling and storage of physical tissue samples. These standards

include the temperature of storage, the content of the solution in which blood samples, for example, are stored as well as the types of meta-data (health information) that are collected and in what formats they are available. Although such details might seem trivial, they are crucial in facilitating the exchange, comparison and transfer of data between various actors. At the same time, however, the setting of technical standards represents the development of political interests in the way information on the human body is made available to the research community and their commercial partners. Standardisation facilitates the introduction of human tissue samples and related health information into global regimes of scientific knowledge production and value creation. In this chapter, I would like to focus on a less obvious area of development in developing standards, namely software development and argue that the development of software standards to allow for data inclusion also represents a political perspective on the ways in which data is made available and the conditions under which various actors may or may not participate in specific scientific collaborative undertakings and thus in the production of scientific knowledge.

Given the fact that the need to compare data from different biobanks and genetic databases is steadily increasing, there has also emerged a need to develop common standards for the way information on health and tissue samples is transferred and compared across physically disparate databases. Such processes require the development of both common software programs which can be used to facilitate either the transfer or data analysis of databases that can be physically located on the other side of the world, as well as accepted practices between different actors. Biobanks and the data that can be mined from them are a scarce and valuable commodity in that the acquisition of large collections is expensive and time consuming. It is, therefore, understandable that those actors who have access and control over such resources want to maintain tight controls over how such data and resources are maintained, shared and eventually introduced into commercial systems of value production. The setting of standards, therefore, also represents a certain political ideology in relation to the way scientific knowledge ought to be produced in relation to the data that is available around the world.

This presentation is an extension of material and interviews that have been conducted on biobanking in Finland since 2004. The primary research has focused particularly on the ways in which data from large collections of human tissue samples, or biobanks, are being activated. From this research I have also become interested in the ways in which such collections become the sites of leverage within broader networks of population genetics, which rely on regional or national biobanks. In doing this research I have also been able to meet and interview those people who are charged with developing the information technology infrastructures within and between research groups to facilitate cooperation and date exchange. These actors, although seen as technicians, play an important role in developing and implementing various standards related to the ways in which information of human tissue samples and health are organized and analyzed across broad networks.

The work presented here draws on Waldby's (2002) notion of tissue economies as systems of circulation which are formed through the acquisition, storage, handling and distribution of tissue samples and the information that can be produced from them (see also Waldby & Mitchell 2006). In this sense, the notion of tissue economies also relates to what has been termed the informational turn (Beaulieu 2004), where the materiality of objects is increasingly embedded within information economies (see also Waldby 2000). In addition, I try to connect the development of tissue economies to work done on processes of standardization in technological development (Bowker & Star 2000; Werle & Iversen 2006). The connection between the development of tissue economies and standards is important in that it helps to identify the practices through which socially accepted practices emerge and come about, as well the process by which actors and their data become enrolled in an international system of knowledge production and value creation. Such practices are by no means straightforward, but reflect important underlying politics, policies and strategies for the acquisition, management and control of information on the human body. This information, in turn, has significant leverage in the biomedical research community, but more importantly in the commercial pharmaceutical industry as well.

Setting standards in data sharing

> The move towards a universal information infrastructure for biobanking in Europe is directly connected to the issues of semantic interoperability through standardized message formats and controlled terminologies. The information infrastructure has become a critical component in life sciences research (BBMRI 2007, 29).

With the completion of the map of the human genome, there emerged new possibilities for the identification of the genetic causes of diseases. Already in the production of the map of the human genome it has been evident that networks of cooperation have played a major role in such undertakings. These networks operate at several different levels: data sharing, publication, research cooperation, as well as program and software development to better manage and analyze the data that is being produced. Although there are many commercially available software programs and hardware platforms for such tasks many research groups also actively develop their own software (Kleinman 2003).

One key element within this global development has been the emphasis that large publicly funded research programs also promote the sharing of data and information that is produced through these programs (Arzberger et al. 2004). The promotion of data sharing has become a major policy issue relating to scientific research (OECD 2004). For example, Teri A. Manolio (2005), director of the Epidemiology and Biometry Program at the National Human Genome Research Institute (USA), has noted that one of the major priorities in large cohort studies of genes and environment is the promotion of data sharing and protocols. Implied in this statement is the assertion that phenotypic and exposure information is collected in a standardized and exchangeable format. To achieve these ends the development of standards has become an important element in both the collection of samples and the development of the tools used in managing the data.

Bowker and Star (2000, 13–14) have noted that standards are any set of agreed-upon rules which span more than one community of practice and are deployed in making things work together across long distances and heterogeneous settings. In addition, standards can be enforced by legal bodies, but there is no way of predicting which standards will prevail in the end. Finally, standards can gain significant inertia and can be difficult to change once

they have become accepted and commonplace. Standards also impose a system of classification in that they are used to denote and distinguish good and bad ways of organizing objects and practices. Such activities are therefore central in relation to the way information on the human body is organized.

The sharing of data, however, is also seen as problematic from certain perspectives in that the sharing of raw, un-analyzed data without any type of reciprocation from users is seen as a waste of national resources. Biomedical data in genetic databases and health registries have come to be seen and compared to any other type of resource. From an industrial production perspective, this also means that it is better for each country or organization to try and process the raw material as far as possible domestically before 'shipping' or exporting it abroad. The process of adding value through local analysis is taking on the same significance as it has in many other industrial sectors, such as steel and the forest industry, and genetic databases are very quickly becoming highly commercialized objects (cf. NORDICUM 2007, 27).

It is within this context of needing to develop and promote the development of tools and methods for sharing data, on the one hand, and maintaining strict control over the data, on the other, that many data sharing techniques are being developed today. The sharing of data is a highly competitive field where participation and access is based on one's input (*quid pro quo*). Without having something to give one cannot gain access to the raw data of others or participate (Mayrhofer & Prainsack 2008). Although such practices (guarding resources) take place within public research organizations, there has emerged a highly enclosed or private sphere within public research organizations. Such practices can be seen to promote both the sharing of data, as well as setting forth new forms of protectionist practices in relation to genetic data. Before we look at this more closely, however, it is important to examine the underlying framework that is facilitating this development in data sharing.

Within biomedical research it is becoming possible to combine and compare a multitude of different types of information resources. These include genotype and phenotype information from disease and population databases, and information on lifestyle and environmental conditions. In doing so, however, it is necessary to develop ways in which these different

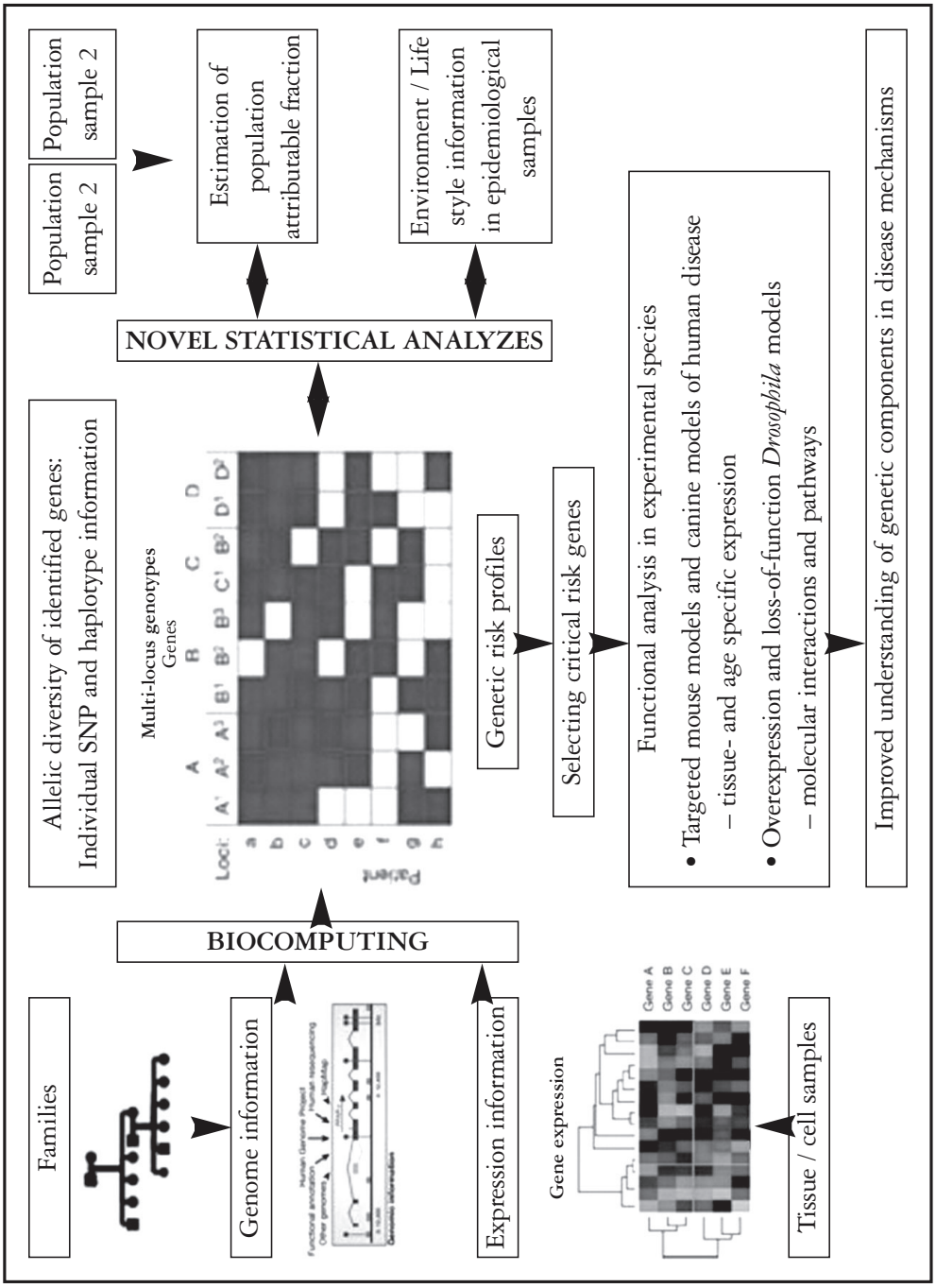
types of information and data can be compared and analyzed together. Figure 1 shows an example of the way one Finnish research group has characterized their activities, as well as the different data resources that they use for their analyses. In order for these different statistical analyses and bio-computing processes to take place, information and data need to be in a standard format.

The processes of standardization are by no means straightforward in that they require negotiation and indeed sometimes competition between different actors. Werle and Iversen (2006) make a distinction between coordinative and market based forms of technical standardization where, according to them, voluntary (coordinative) forms of standardization require some type of democratic legitimation as well. The coordinative approach aims at interoperability and compatibility through negotiation among actors. Such activities are voluntary and their goal is to reduce transaction costs and produce positive externalities. Within the context of tissue economies the reduction of transaction costs has become an important feature through which the production of scientific knowledge is facilitated.

One example of the ways in which voluntary standards are developed in biobanking comes from the setting up of various networks between researchers, such as the European Network for Research Tissue Banks (ENRTB). In late 2000 researchers from several European countries attended the UK Human Tissue Bank Workshop which was the first research tissue bank meeting that was organized in Europe. A total of twenty-two participants from six countries attended the workshop. The purpose of the meeting was to set up a network of research tissue banks in Europe that would actively support the acquisition of tissue sample collections and develop collectively acceptable standards and procedures within Europe for the acquisition and use of tissues. The network would also help develop quality standards and organize training among its members. According to its mission statement the goal of the group was

> To establish a sustainable network for sharing information to guide in the establishment and running of human tissue banks with the ultimate goal of sharing human tissue / information derived from use of these donations across this network under harmonized guidelines and agreed best practices to promote the use of human tissue (Orr et al. 2002, 136).

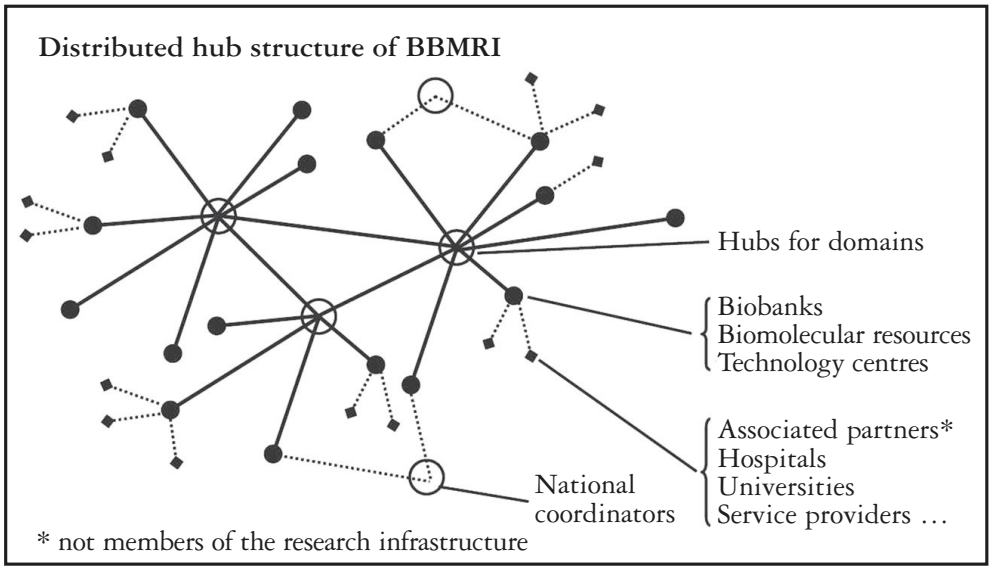
Figure 1. A schematic of various data resources used by a Finnish research group



The mission statement highlights the development and codification into practice of the way in which *donations are shared*. At the same time, however, the network seeks to implement harmonized guidelines according to which the practices of various actors will be governed. These practices are voluntary, but are clearly meant to facilitate the exchange of biological material across the network. At the same time, however, the activities of this network can serve as a standard according to which other actors, operating outside the network can organize their activities. This process represents the foundations on which the enrolment of tissue samples is made possible into global regimes of knowledge production. It remains, however, more programmatic than technical in nature.

A more recent venture is the pan-European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) initiative which is a preparatory effort under the EU's 7th Framework Programme to coordinate the activities of 50 participants involved in biobanking from around Europe. One of its goals is to harmonize and standardize collection, storage and analysis techniques associated with biobanking in Europe (BBMRI 2007).

Figure 2. Networks of biological resource centres according to BBMRI plan



Source: BBMRI 2007

According to its organizers, BBMRI will give Europeans a distinct advantage by developing a 'broad and unified access to the catalogued information on biological samples and collected data which is presently difficult due to different data structures and incompatible regulations for their access and exchange in different countries' (BBMRI 2007, 13). The BBMRI effort includes a total of 104 biobanks that comprise more than 12.5 million samples. As such, these types of efforts play a crucial role in the way information about our health is collected, stored and analysed, and has an impact on the way biomedical knowledge is produced.[3] The main idea behind the BBMRI infrastructure, however, remains the connection of a vast network of biological resource centres around Europe, as is shown in Figure 2, through which actors, samples and data become enrolled and included into networks of sharing.

In terms of genetic databases and other databases containing health information, a great deal of standardization of software applications is needed in order to support such broad initiatives. Within such programs, new ways of managing data are also being developed in order to meet the needs of the research community, as well as to assure that the sensitive data that is being managed and analyzed remains safe and secure. In the next section I will look at the development of the concept of database federation and semantic interoperability as one solution to data management in large projects.

Database federation and semantic interoperability

> The concept of a database federation is not new. It has been available on relational database management systems for over two decades. In a federated system, remote data tables or data objects in general are made available through an integrating database using special database views, which are like local view tables that can be joined in SQL queries with other tables and views (Muilu et al. 2007, 2).

As the size and scope of genetic databases has increased so too has the need to develop better tools for their management. Since the collection, storage and analysis of various biobanks have taken place within particular historical and socio-technical conditions, there is considerable variation among the databases that exist. This poses several challenges in terms of

data integration. The biggest challenge relates to the way in which data from these separate databases can be integrated without each actor losing its autonomy over the data that they have collected. This process of communication also necessitates the requirement for all databases to be in similar formats, which allows for semantic interoperability. Most databases rely on some type of central authority (database administrator) who is charged with designing and maintaining the database architecture. This authority also necessarily imposes constraints on those who use the database by, for example, setting the input formats and parameters. A traditional drawback in combining separate databases thus meant that flexibility and control had to be ceded to some central authority. It became increasingly important during the 1980s for data management specialists to develop a solution to this problem.

One of these solutions was the development of federated database management systems (FDBMS). Originally developed in the US through funding by the Air Force, the National Science Foundation and the Defence Advanced Research Projects Agency (DARPA), the goal of the federated system was to interconnect databases (and thus facilitate sharing), but at the same time minimizing central authority. According to Heimbigner and McLeod (1985, 254) the '*federated database architecture* allows a collection of database systems (*components*) to unite into a loosely coupled federation in order to share and exchange information. The term *federation* refers to the collection of constituent databases participating in a federated database' (italics in original). The main benefit of the federated system is that each participant in the system retains control over their own data and resources. Participation in the system, however, requires that some of the data is made available to others as well.

Although relational databases are not new, the use of federated database management systems in the analysis of geographically dispersed genetic databases is more recent. Such database management systems address two of the important criteria that I have discussed above. *First*, they provide an infrastructure which allows and promotes the sharing of data among participating actors. This has been a key policy goal within the biomedical research field. *Second*, the federated system allows each participant to maintain control over their own data, which might otherwise be compromised

under other data management architectures. At the same time, however, data sharing requires a level of semantic interoperability which can only be achieved through the setting of standards. What we are now witnessing within the development of tissue economies are therefore the processes through which exchange standards are being developed transnationally.

The ability to share data among the components is based on cooperative activity and 'freedom of association' between each component. The implementation of such components between various actors thus requires a high degree of voluntary (coordinative) forms of standardization (cf. Werle & Iversen 2006). In addition, however, such practices also have fundamental implications for the way in which future data management architectures will be implemented. Given that the role of biobanks and genetic databases in biomedical research is increasing, such practices may have an important impact on the tools and protocols that are used to analyze and produce knowledge on the human body.

At the same time, however, the federated database system represents a type of political approach to the way in which national resources are used and shared across national borders. Essentially, this political perspective outlines the contours of the way in which the economy of tissue use and sharing will be implemented and the ways in which information on the body becomes introduced in the first place into systems of scientific knowledge production, and subsequently into systems of commercial exploitation.

In the following I will describe the way in which a federated database architecture was implemented in a large international research consortium which integrated phenome and genome data from over 600,000 twin pairs. The case highlights the way in which cooperative engagements between actors enrols data into coherent systems through which it is analysed and subsequently deployed.

GenomEUtwin – implementing a federated database

The GenomEUtwin research project was a six year research project funded by the European Commission and included twin registries from the Netherlands, Denmark, Norway, Sweden, Finland, Italy, UK and Australia. The goal of the project was to pool the epidemiological and phenotype data

from over 600,000 twin pairs and genotype data from a smaller fraction of the whole cohort. The aim of the project was to identify genetic variants of most common diseases, such as diabetes and heart disease. The twin research project was also able in many cases to connect study samples to national health care registries, (e.g. discharge registries, cancer registries and cause of death registries), which enabled even further data mining of the unique data sets.

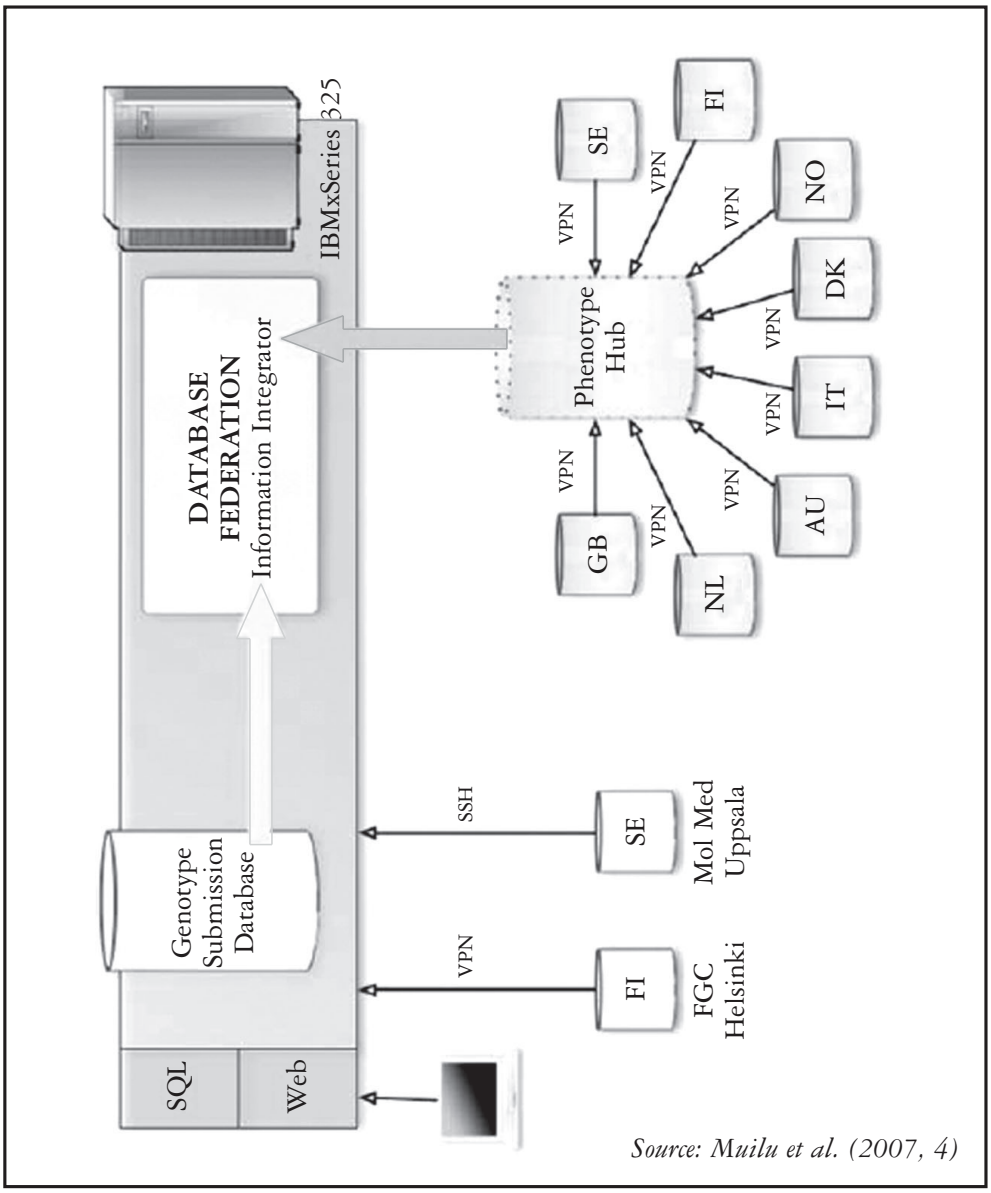
The GenomEUtwin project is important since it is one of the first (if not the first) effort to combine biomedical information from twins using the database federation model. In this sense it has served as a development platform for more recent developments in database federation. I am using the notion of *platform* (Keating & Cambrosio 2003, 27) here to denote both the technical aspects of biobanking in standardization and development, as well as the political dimension of organizing vast amounts of resources in both a discursive and material sense. The idea of a platform in biobanking is important from an information architecture perspective, in that such infrastructures will play an important role in the way health information will be analyzed and managed in the future.

The model used in the analysis of the twin data is based on the database federation model. The main advantage of this is that each participant retains their control over the samples and data that they have collected. The data of all participants is at the same time available to everybody through the composite of all constituent databases in the system. Figure 3 provides a schematic overview of the way the database federation system operated in the GenomEUtwin project.

As one can see, there is both a genotype and phenotype hub to which various participants contribute their data. This data is then integrated into the database federation or information integrator. The area within the IT architecture that has been developed to house the information composite from all the participants, or TwinMart as it is called, is also known as a *demilitarized zone*. The idea of a demilitarized zone within information architectures highlights the sensitive nature of the information that is being integrated, as well as the highly competitive nature of biomedical research. Demilitarization refers to the fact that no participant can claim a stake to the data that is integrated into the hub, but rather it is available to

all those who participate and contribute to the project. Such zones of demilitarization also attest to the highly competitive nature of biomedical research, as well as the commercial expectations that are increasingly being attached to biobanking ventures (Tupasela 2006; 2007).

Figure 3. The database federation model of the GenomEUtwin project



Source: Muilu et al. (2007, 4)

Such platforms for the analysis of both genotype and phenotype information bring forth several important features that are emerging in relation to biomedical research and tissue sample collections. *First*, physical tissue samples and the tissue economies that they make up are becoming increasingly connected to global information economies and are related closely to what Beaulieu (2004, 367) has described as the 'informational turn', which is characterized by the production of large amounts of data based on physical specimens (see also Thurtle & Mitchell 2004). The informational turn is placing specific data management needs on the ways in which information architectures are being designed and implemented. Issues such as privacy and security are at the top of the list of these requirements. *Second*, local or national collections are increasingly becoming points of leverage and access to trans-national cooperative efforts in bioinformatics (cf. Mayrhofer & Prainsack 2008). Although national projects emphasise their local (national) significance, such collections are also becoming entry points to an international market for health and genetic data on populations. This places further demands on the information that is passed on to the people and patients from whom information is collected and stored. The conglomeration of data from across the globe also challenges many of the arguments that have been set forth on the validity of certain types of populations as ideal populations for the study of the genetic causes of disease in that it is becoming increasingly evident that large, multi-national samples are being needed for this task. *Third*, the development of networks of data exchange also necessitates coordinated efforts between actors to standardize data and the way it is managed.

As mentioned above, standardization relating to biobanks takes place at many levels, from the ethical, legal and social aspects through to the technical specifications relating to the storage of samples. Standardization processes also relate to what types of information are collected and analyzed, both for phenotype and genotype data. In the GenomEUtwin project, for example, participants had to agree on the types of clinical phenotype information that would be collected and analyzed. This included serum lipid values, insulin and glucose content, as well as other metabolic traits. In addition, the database also integrated data on weight, height and BMI, as well as data on questionnaires relating to migraines. In

terms of the genotype data, the participants also had to agree on quality control, as well as the genotypes and alleles that would be used at two different sites (Uppsala and Helsinki). Currently the database includes over 20 million accessible genotypes (Muilu et al. 2007, 5).

In order for the data in these databases to be accessible across different platforms, bioinformaticians are also developing common programming languages that can be used across the internet. In the case of GenomEUtwin, for example, programmers and developers are using and developing what is called Polymorphism Markup Language (Open PML). PML is a variation of the common internet programming language XML (Extensible Markup Language). PML 'defines exchange format of genomic sequence variation data, which enables us to easily and effectively assemble and analyze genomic sequence variation data from distributed heterogeneous database in XML documents described in PML. PML was adopted as the OMG standard specification in June 2005.'[4] Although such groups do not maintain or have any type of legal mandate or power to set standards, they nonetheless have considerable power as voluntary organizations to set and implement standards across broad networks of actors.

The uptake and acceptance of these standards, however, is dependent on many factors. At the development stage, the group of people involved in the development of standards and programming remains quite small. Usually it consists of the core set of data management specialists who are directly involved in the project that is being funded. As projects develop, however, there emerges a need to enlist larger actors to accept the standards that have been developed among this small group of people. As one program and standards developer noted in an interview:

> The challenge for us is to get the big players interested in what we are doing. At the beginning it's just a small group of us trying to develop a good tool for our needs, but obviously we want others to accept our work and this is where we see if we have succeeded in that. And it also indicates to us whether we've developed something useful. (Interview with data management specialist, 20.9.2007)

The interview excerpt highlights the challenge that confronts the developers of new standards for the management of genotype and phenotype data; although there is a relatively small group of actors who par-

ticipate in the development of these standards through a particular research program, there is no guarantee that it will be accepted by the scientific community at large. Being among the first to develop such standards, however, and the application of these standards to the management and analysis of such large collections of data makes the uptake of the standards much more likely. The data that is made available through database federation, for example, therefore has a further function besides enabling each participant to gain access to the data of other participants, it also serves as further leverage to indirectly impose technical standards and solutions developed along the way on others who wish to make us of this data sharing scheme in due course. Following Bowker and Star (2000, 13–14), however, the standards that are developed within the more limited circle of initial developers are not standards until they are adopted outside of the group and span broader networks than simply the initial group.

It is for this reason that the development of standards in conjunction with the analysis of genetic, phenotypic, lifestyle and environmental data may have such an important impact on the way we study diseases and the corresponding tools in that if these initial specifications become accepted more broadly, they will then also have great potential for the way research tools come into broader operation.

Discussion

The development of standards for semantic interoperability in the analysis of data collected from biobanks is taking place at multiple levels and numerous different sites. Biobanking activities and its related bioinformatics infrastructure are, however, beginning to develop into various platforms for the analysis of the data, which inevitably require the development and uptake of standards. This process is also political in that the choice of a particular standard and approach to the sharing of data reflects the premises upon which access and use of other data is allowed. In this sense, the tissue economy of European biobanks reflects an approach where access is premised on contribution, forcing would-be participants to collect

their own samples and data according to the standards that have been set, and then make it available to other participants in return for access.

In using the notion of *platform* (Keating & Kambrosio 2003, 27) to denote both the technical aspects and the political dimension of organizing resources in both the discursive and material sense, I have tried to analyze the means through which such information architectures are being realized in biomedical research. This information architecture is closely aligned to, and indeed reliant on, the material tissue economies which are being amassed in a number of different countries (cf. Waldby & Mitchell 2006). These tissue economies, in turn, activate and encompass a whole host of other national database resources related to human health and activities (Tupasela 2006). They also play a crucial part in the way biomedical knowledge is produced and made available to the European healthcare sector, as well as private businesses. The analysis of these platforms and their related tissue economies is therefore by no means a trivial venture, given the scope and direction in which biomedical research is heading, but rather extends the scope from tissue economies ever further into new domains.

What begins to emerge, however, is also the global development and establishment of standards for the exchange of tissues and information. The development of such tissue economies is by no means mundane since it can have far reaching repercussions as to the material and informational conditions under which scientific knowledge is produced, exchanged and enrolled into global systems of commercial profit. At the centre of these developments one can locate the often mundane activities associated with the processes of standardization, which are in turn taking place at multiple levels. Standardization is the process by which disparate information architectures and resources are made interoperable and interconnected. Standards allow for data to be exchanged and shared, but they also create the possibility for access and leverage through which research groups enter transnational flows of data on the human body. In this sense, the development and uptake of standards also has political and social consequences in relation to the ways in which information on the human body becomes integrated into increasingly larger systems of biomedical research.

I have looked at the development of database federation and semantic interoperability as a possible model on which future analysis and data sharing will be modelled. Database federation as such represents only one possible solution to the problems posed by sharing and the interoperability of data that has been collected in different locations and different times using variable criteria and standards. The standards that are being developed around database federation are still very much in the development phase. Database federation, nonetheless, purports to solve many of the difficult issues associated with sharing sensitive data and thus allowing the custodians of databases to maintain a high level of control over what they have collected over many years. At the same time, the processes associated with standard setting in small, close knit groups also helps to create a favourable working environment for developing standards (i.e. not too large a development group). This small size, however, poses challenges as to whether these standards will be adopted by the larger scientific community as the communication and exchange tools of the future.

Notes

1 For more information see http://www.decode.com/, http://www.ukbiobank.ac.uk/, http://129.215.140.49/gs/, http://www.cartagene.qc.ca/accueil/index.asp, http://www.egeen.ee/.

2 Other important international documents include the *International Ethical Guidelines for Biomedical Research Involving Human Subjects* (Council for International Organizations of Medical Science, CIOMS), *Universal Declaration on the Human Genome and Human Rights* (UNESCO) and the *Helsinki Declaration* (World Medical Association).

3 The BBMRI initiative to harmonize activities and develop standards builds on work already done, for example by the Organization for Economic Co-operation and Development (OECD) best practice guidelines for the operation of Biological Resource Centres, and the International Society for Biological and Environmental Repositories (ISBER) Guidelines on Biobanking.

4 Available at http://www.openpml.org/. Object management Group (OMG). The OMG meets regularly at the International Bio-Data Interoperability Conference (IBIC) to discuss and agree upon the standardization of genome sequence variation data description form and propose these forms as official standard specifications.

References

Arzberger, P. et al. (2004), 'An international framework to promote access to data', *Science* 303 (19 March): 1777–1778.

Austin, M. A., Harding, S., and McElroy, C. (2003), 'Genebanks: A comparison of eight proposed international genetic databases', *Community Genetics* 6: 37–45.

Beaulieu, A. (2004), 'From brainbank to database: The informational turn in the study of the brain', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 35 (2): 367–390, special issue on 'the Brain in the Vat'.

BBMRI (2007), 'Biobanking and biomolecular resources research infrastructure. Construction of new infrastructures – preparatory phase', FP7-INFRASTRUCTURES-2007-1 research proposal.

Bowker, J. and Star, S. L. (2000), *Sorting Things Out – Classification and its Consequences*, Cambridge, MA: The MIT Press.

Collins, F. S., Green, E. D., Guttmacher, A. E., and Guyer, M. S. (2003), 'A vision for the future of genomics research – A blueprint for the genomic era', *Nature* 422: 835–847.

ECVAM (2002), 'The establishment of human research tissue banking in the UK and several Western European countries', the Report and Recommendations of ECVAM Workshop 44, reprinted with minor amendments in *ATLA* 29: 125–134.

Heimbigner, D. and McLeod, D. (1985), 'A federated architecture for information management', *ACM Transactions on Office Information Systems* 3 (3): 253–278.

Keating, P. and Cambrosio, A. (2003), *Biomedical Platforms – Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*, Cambridge, MA: The MIT Press.

Kere, J. (2007), 'Miten Suomessa kerättyjä DNA – ja kudosnäytteitä voidaan hyödyntää?' [How can DNA and tissue samples collected in Finland be utilized], *Duodecim* 123: 864–866.

Manolio, T. A. (2005), 'The NIH initiatives and plans for biobank studies', presentation at the 'International Biobank and Cohort Studies: Developing a Harmonious Approach' workshop, 7–8.2.2005, Atlanta, GA.

Mayrhofer, M. and Prainsack, B. (2008), 'Being a member of the club: The transnational (self-)governance of biobank networks', *International Journal on Risk Assessment and Management* 12: 64–81.

Muilu, J., Peltonen, L., and Litton, J.-E. (2007), 'The Federated Database – A basis for biobank-based post-genome studies, integrating phenome and genome data from 600,000 twin pairs in Europe', *European Journal of Human Genetics*: 1–6.

NORDICUM (2007), 'The secrets of the genome', *Scandinavian Business Journal* 3: 26–28.

OECD (2004), 'Declaration on access to research data from public finding', available at www.oecd.org.

Orr, S. et al. (2002), 'The establishment of a network of European human research tissue banks', *Cell and Tissue Banking* 3: 133–137.

Palotie, A. and Peltonen-Palotie, L. (2004), 'Pitäisikö perustaa suomalainen biopankki? Epidemiologisten aineistojen hyödyntäminen on kaikkien etu', [Should a Finnish biobank be founded. Utilizing our epidemiological data is in everyone's interest], *Duodecim* 120: 1710–1712 .

Petrone, J. (2003), 'Scientists debate drug research at gene forum', *Baltic Times,* 18.9. 2003.

Sunder Rajan, K. (2006), *BioCapital – The Constitution of Postgenomic Life*, Durham: Duke University Press.

Tupasela, A. (2006), 'Kudostalous ja kaupalliset mallit – Biolääketieteellisen tutkimuksen muuttuvat ehdot', [Tissue economies and commercial models – The changing terms of biomedical research], *Tiede&edistys* 31 (1): 105–118.

Tupasela, A. (2007), 'Re-examining medical modernization – Framing the public in Finnish biomedical research policy', *Public Understanding of Science* 16 (1): 63–78, special issue 'Publics and Science: A New Understanding'.

Waldby, C. (2000), *The Visible Human Project – Informatic Bodies and Posthuman Medicine*, London: Routledge.

Waldby, C. (2002), 'Stem cells, tissue cultures and the production of biovalue', *Health: An Interdisciplinary Journal for the Study of Health, Illness and Medicine* 6 (3): 305–323.

Waldby, C. and Mitchell, R. (2006), *Tissue Economies: Gifts, Commodities, and Bio-Value in Late Stage Capitalism*, Durham: Duke University Press.

Werle, R. and Iversen, E. J. (2006), 'Promoting legitimacy in technical standardization', *Science, Technology & Innovation Studies* 2 (March), available at www.sti-studies.de.