Inter- and Intra-Biobank Networks: Classification of Biobanks

Darren Shickle  Marcus Griffin  Karen El-Arifi

Academic Unit of Public Health, Institute of Health Sciences, University of Leeds, Leeds, UK

Abstract

Background: Many biobanks have struggled to deliver on the high expectations and claims made for them because of insufficient samples, inadequate infrastructure, cost of establishing and maintaining a large enough resource over the long term, and satisfying legal, ethics and governance requirements. Increasingly, networks have formed to help with the collection, processing, storage, advertising, and distribution of samples. However, there are also challenges to establishing and maintaining biobank networks. Aim: To classify biobanks in order to better understand the problems faced by biobank networking. Methods: Interviews were conducted with principal investigators and/or managers responsible for 33 biobanks in 9 countries. Results: Biobanks were classified into the following categories: ‘storage’, ‘bring-and-share’, ‘catalogue’, ‘partnership’, ‘contribution’ and ‘expertise’. Conclusion: It was possible to allocate all of the biobanks visited to one of the network categories although some fitted better than others. Thus, the classification may not be mutually exclusive nor encompass all types of biobanks. Many of the governance and operational problems associated with the biobanks visited were due to networking functions: either intra- or inter-biobank networks. Thus, this proposed classification system should help better understand these issues and identify solutions.

What Is a Biobank?

According to the Organisation for Economic Co-operation and Development (OECD), 'research involving human genetic or genomic information analyzed in conjunction with other personal or health data has become increasingly important for understanding risk factors underlying complex (multi-factorial) diseases' [1].

The OECD defined human biobanks and genetic research databases as 'structured resources that can be used for the purpose of genetic research, which include: (a) human biological materials and/or information generated from the analysis of the same; and (b) extensive associated information' [1].

However, there is no consensus on a definition for biobanks. Some, for example, the Council of Europe, only use the term biobank population collections, and have a separate definition for other collections of biological material biobanks [2]. Swedish legislation defined a biobank as 'biological material from one or sev-
eral human beings collected and stored indefinitely or for a specified time and whose origin can be traced to the human or humans from whom it originates’ [3]. The size of the sample collection has no significance within the Swedish Act: a single human sample is sufficient. The Swedish definition does not apply to unidentifiable human biological material where it is not possible to identify the person from whom the tissue originates. Portuguese law defined biobank as ‘any depository of biological samples and related derivatives, with or without a pre-defined period of storage, based on prospective collection or made up of previously collected material, obtained for health care purposes, public health monitoring programs, or for research, and that includes identified, identifiable, anonymized or anonymous samples’ [4]. The approach taken within the Swedish and Portuguese legislation differs from the OECD definition because they could apply to biobanks which do not have a research element.

The term ‘biobank’ also need not be restricted to human biological material, so for example, there are biobanks containing samples from other animal species [5], plants [6], or microbes [7].

**The Difficulties with Biobanks**

Despite the OECD claiming that biobanks have already contributed ‘significantly to our understanding of genetic and environmental factors that influence disease risk and treatment’ [1], many biobanking initiatives have failed to deliver on the high expectations and claims made for them [8]. The main areas of difficulty for biobanks are:

- insufficient samples to conduct research on rarer diseases;
- inadequate infrastructure to process, store and retrieve samples to meet the necessary quality standards for research;
- cost of establishing and maintaining a large enough resource over the long term, and
- satisfying legal, ethics and governance requirements.

As there are many parameters that influence and characterize disease (e.g. genetic polymorphism, comorbidity, lifestyle, treatment), thousands of samples must be studied in order to assess the clinical usefulness of a marker or drug [9, 10]. This could entail the follow-up of a large number of cases with homogeneously treated tissue samples in the context of multicentre and multinational projects. However, as Morente and Alonso [10] point out, the problem is that there is no institution in the world, and almost no country on the planet, that is capable of collecting a sufficiently large number of cases for studying. For this reason, networking appears the best solution for Tumour Bank to expand’.

**The Move to Networking**

Morente and Alonso [10] suggested that networking has the following advantages:

- access to a larger series of common and uncommon cases;
- access to cases previously included in multicentre and international clinical trials;
- access to cases included in multicentre co-operative projects;
- improvement of effective multicentre connections;
- implication of national and international scientific societies in research projects, and
- avoidance of the intrinsic bias of multicentre studies.

However, Morente and Alonso [10] also recognized that effective networking required:

- standardized technical procedures;
- strict and common quality control programmes;
- homogeneous ethics requirements, and
- open mentality for sharing.

While networks seek to address the inadequacies of biobanks, they are also faced with similar challenges. For example, the cost issues are magnified, not just because the number of samples increased. Storing more samples can result in relative cost savings through economies of scale, but can also introduce more complexity if samples are being stored and retrieved from multiple locations. More organizations will be involved in the network, each with their own overheads and aspirations that must be satisfied. There may also be a network of funders, each with their own remit and political agenda [11]. More stakeholders, spread over more jurisdictions further complicates governance and ethics challenges.

**The GeneBanC Project**

The research described in this paper was conducted as part of a work package of the GeneBanC project (www.genebanc.eu) funded by the European Commission’s 6th Framework Programme. The aim of the work package was to understand the practical, legal and ethical considerations in establishing and maintaining a biobank. However, as a result of the work of this work package, it became apparent that a new way of categorizing biobanks, and in particular biobank networks, would be helpful in order to better understand the problems facing biobanking and for the development of solutions.
**Description of Biobanks**

Biobanks can be very different from one another, depending on:
- how long they have been in existence;
- whether data and sample collection is prospective or retrospective;
- whether they are focused on population exposures or specific disease outcomes, and indeed whether the diseases are very rare or very common;
- whether samples were collected for clinical purposes (secondary use) or specifically for research (primary use);
- the size of the biobank, in terms of number of subjects recruited;
- the type of samples being collected, e.g. tissue, blood, urine, DNA;
- the extent to which samples can or cannot be linked to identifiable individuals;
- the ownership of samples;
- the way samples are stored;
- storage location, in particular whether there are a number of locations where samples/data will be stored;
- the degree to which access is permitted by third parties, if at all;
- whether they are a member of a network;
- whether they are public sector (not for profit) or commercial (for profit);
- the way that they are funded;
- the number of organizations involved in funding or running the biobank;
- the political and regulatory environment in which they operate, in particular if samples and/or data are to be transferred between countries with very different legislative frameworks, and
- the social environment, in particular attitudes of the media and general public.

Many authors have attempted to describe biobanks based on various combinations of these characteristics [12–15]. However, such lists only serve to demonstrate the heterogeneity of biobanks. Indeed, the initial intention for the GeneBanC work package was to construct a sampling frame to select biobanks to be interviewed according to whether they were prospective or retrospective; size of biobanks, and to cover a range of diseases and exposures. However, following an initial trawl of the literature and the internet, it was apparent that it is difficult to categorize biobanks in this way. Indeed, it was also difficult to identify biobanks because very few biobanks are actually called biobanks.

Part of the problem is that biobanks appear too diverse, and it is too difficult to devise governance and regulatory frameworks to cover all eventualities, resulting in a form of ‘biobank exceptionalism’. ‘Genetic exceptionalism’ [16] is the claim that genetic information is unique and deserves special consideration. But the confusion is actually due to the fact that the term ‘genetics’ covers such very different concepts and technology, and a separate regulatory framework is required for each. What is required is a Linnaean taxonomy to distinguish between the different species of biobank. This paper proposes a new way of thinking about biobanks, recognizing that all to a greater or lesser extent depend on networking, even if it is only intrabiobank networks for collection and storage of samples, or sharing the expertise to do so.

**Methodology**

Biobanks were identified by searching the internet and Medline using various synonyms of ‘biobank’. The principal investigator or manager for the biobank was contacted and an interview requested. Previous studies of this sort have tended to only survey larger or better publicized biobanks. In order to minimize such a selection bias, a ‘snow-balling’ recruitment strategy was used with interviewees asked to suggest contacts for other biobanks in their locality.

For the purposes of this paper, the OECD definition [1] was used for a ‘biobank’ and a modified version of a definition used by Oxford English Dictionary [17] was used for ‘network’: ‘A network is an interconnected group of people or organisation having certain connections which may be exploited to gain preferment, information, etc., especially for professional advantage.’

Interviews lasted between 1 and 2 h. Interviewees were asked about co-investigators and organizations responsible for the biobank, funding arrangements, governance arrangements, ethics scrutiny, public/patient consultation, recruitment strategies, response rates, informed-consent procedures, samples and data collected, data protection, intellectual property right and access to the biobank. Interviews were recorded and transcribed. Framework analysis [18] was performed to identify practical, legal and ethical issues faced by biobanks and how these have been addressed.

**Results**

Between 2007 and 2009, 33 biobanks were visited in 9 countries (Australia, Canada, Denmark, Italy, Latvia, The Netherlands, Singapore, Sweden, and the United Kingdom). The biobanks were at differing stages of development from some still being constructed to others that had been in existence for over 30 years. The smallest biobank visited intended to collect samples from 200 families while the biggest aimed to recruit 500,000 people.
Different Forms of Network

Some biobanks were overtly members of a network. But even those that did not seem to be a network, on closer inspection had internal networking functions. The following forms of networks were observed during the interviews (table 1) although many networks are a mixture of these forms:

- storage networks;
- bring-and-share storage networks;
- catalogue networks;
- partnership networks;
- contribution networks, and
- expertise networks.

Storage Networks (table 2)

Some networks have developed mainly as storage solutions. The cost and speed of genetic procedures have improved significantly over the last decade as evidenced by the Human Genome Project coming in ahead of schedule and below budget. Nevertheless, the costs associated with establishing and maintaining a biobank are significant. Storing samples in a freezer in a corner of a laboratory is no longer sufficient.

Storage networks are mainly collaborations having a shared facility to reduce cost and raise quality. Smaller collections cooperate for economies of scale and to utilize cutting-edge technology for processing, storage and retrieval of samples. Typically, the network is preceded by one or two collections, and the storage facilities are developed with spare capacity that can be made available for other biobanks to use.

Any other connections between the studies in terms of research may be limited or non-existent. Typically, they may have a focus on a relatively small geographical area,
Inter- and Intrabiobank Networks: Classification of Biobanks

There is no externally viewable catalogue describing the contents of the network although individual biobanks within the network may advertise their samples for others to access. The samples within the storage facility remain the property of the originating biobank which controls access to the samples for further research.

As a proviso to use the storage facility, some storage networks require evidence of a minimum standard of informed consent, so there may be some convergence in ethics procedures and legal requirements.

**Bring-and-Share Storage Networks (table 3)**

Other forms of storage network, in addition to charging their usual fee for their storage and processing services, will have a lower fee structure for researchers willing to allow access to their samples. Thus it will cost less for those who bring their samples for storage but are willing for them to be shared, compared to those researchers who want to use the facility for storage until they want to access the sample. For an intermediate price, there may be a time-limited embargo on the release of material to permit researchers time to complete research studies arising from their cohort recruitment before the samples become generally available to the research community.

**Catalogue Networks (table 4)**

The simpler form of storage networks must maintain a database of samples for retrieval and governance purposes but this is not accessible outside the storage facility. As with bring-and-share networks, catalogue networks maintain a database that is searchable by external researchers seeking samples for their research. They bring together researchers to increase statistical power to study rare diseases or exposures. They are useful at the international level, where it would be inefficient or inappropriate to send samples to a single storage facility. The problem of transferring samples between countries with incompatible ethics and legislative requirements remains. Thus, these issues typically consume a lot of attention when the network is established.

Catalogue networks are built around the information technology (IT) needed for uploading and searching data. A significant obstacle to communication between biobanks is the different information technology platforms, message formats and terminologies that they use [11]. Without IT compatibility, it is difficult to mine and

<table>
<thead>
<tr>
<th>Table 3. Bring-and-share networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main coverage</td>
</tr>
<tr>
<td>Comments and further information</td>
</tr>
<tr>
<td>Genetic Repositories Australia</td>
</tr>
<tr>
<td>Istituto Nazionale per la Ricerca sul Cancro, Biological Bank and Cell Factory</td>
</tr>
</tbody>
</table>

1 Location of coordinating centre.
BBMRI = Biobanking and Biomolecular Resources Research Infrastructure.

<table>
<thead>
<tr>
<th>Table 4. Catalogue networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main coverage</td>
</tr>
<tr>
<td>Comments and further informaion</td>
</tr>
<tr>
<td>TuBaFrost</td>
</tr>
</tbody>
</table>

1 Location of coordinating centre.
BBMRI = Biobanking and Biomolecular Resources Research Infrastructure.
interface datasets, which hinders the ability to publish a catalogue of samples available within the network. Catalogue networks are also only as effective as the effort invested in uploading and maintaining information about the content of collections. For example, although there has been cooperation and exchange of samples, TuBaFrost has not yet been used for exchange of samples as it should. Whilst TuBaFrost was seen as an excellent platform, researchers did not feel very confident using the central database and were in need of support uploading all sample data. “This shows there is a serious bottleneck between the local database and the central database. It is far too much work and more importantly also too difficult to efficiently do a time consuming translation of the sample description for a batch upload of sample data also without really knowing the benefits of the system for the local biobank’ [19].

**Partnership Networks (table 5)**

Partnership networks attempt to share costs and effort in recruitment. The material and data may be stored in a single location, or in multiple storage sites at the location where they were collected. But there is always a central coordinating hub for network administration. There is a formal contract or partnership agreement that specifies the rules of the consortium. Externally, it may appear to be a single biobank. The samples collected as part of the collaboration would be considered to ‘belong’ to the partnership. Ownership of samples if the network were to be dissolved would need to be specified within the consortium agreement. However, if the network brings together biobanks that are already in existence then the historical collections may remain separate in some way. This could mean that there are different mechanisms for accessing these samples, even if they are declared within the partnership network catalogue. Often there are national initiatives to encourage clinicians and researchers to collect samples within a network rather than set up and/or maintain their own smaller collections with all of their problems of scale and limited expertise. The partnership network model was considered during the formation of TuBaFrost. However, it was felt that this model ‘may potentially deter collectors as they have no part to play in the decision-making process and may feel that their samples could be wasted in projects which they consider useless’ [20].

The process of working within a partnership encourages convergence between the consortium partners. Recruitment processes may vary to reflect operational differences within each organization. However, information sheets and consent forms may be very similar. Even if samples are stored at the separate sites, processing may follow similar protocols to allow consistency of samples that are sent outside the network. The partnership will need to maintain a single catalogue of samples owned by the network, and so there is usually harmonization of IT systems.

The lack of sustained funding can be a major problem for the long-term maintenance and operation of central resources in the life sciences such as biobanks [11]. There are economies of scale from research centres collaborating within all of the forms of network. By working to-

<table>
<thead>
<tr>
<th>Table 5. Partnership networks</th>
<th>Main coverage</th>
<th>Comments and further information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victorian Cancer Biobank</td>
<td>Victoria (Melbourne)</td>
<td><a href="http://www.viccancerbiobank.org.au">www.viccancerbiobank.org.au</a></td>
</tr>
<tr>
<td>onCore UK</td>
<td>United Kingdom (Borehamwood)</td>
<td>In 2009 it ceased receiving samples and now only operates as an expertise network Member of The National Cancer Research Institute’s Confederation of Cancer Biobanks BBMRI associate member <a href="http://www.oncoreuk.org">www.oncoreuk.org</a></td>
</tr>
<tr>
<td>Dental Deformities</td>
<td>United Kingdom (Leeds)</td>
<td></td>
</tr>
<tr>
<td>Wales Cancer Bank</td>
<td>Wales (Cardiff)</td>
<td><a href="http://www.walescancerbank.com/">www.walescancerbank.com/</a></td>
</tr>
</tbody>
</table>

1 Location of coordinating centre.
BBMRI = Biobanking and Biomolecular Resources Research Infrastructure.
Together, researchers can request money for centralized activities, e.g. cataloging, advertising and distributing samples, as well as facilitating sharing of expertise and development of operating procedures. Much of the money given by outside bodies to a network does not go on central administrative costs, but is distributed to the collaborating centres for maintenance and upgrading of the biobank. Thus scientists get access to more staff, newer and better equipment that they may not have had if they were operating their own smaller biobank.

**Contribution Networks** (table 6)

Contribution networks depend on much loser collaboration between the biobank and the people responsible for collecting samples/data. The biobank is not usually responsible for the collection of samples, and only stores and distributes tissues. Typically, clinicians recognize a case that may be of relevance to the biobank, e.g. because the disease is unusual or rare, and send in the sample to the biobank. Some clinicians may be regular contributors, others may only ever send one or two cases. The network may be established by or on behalf of a professional body or an association of clinicians and researchers with an interest in the disease. Contributors never have a formal contract to supply samples. Usually, contributors are aware of the biobank through conferences, publications and newsletters.

**Expertise Networks** (table 7)

In order to function properly, both catalogue and partnership networks will need to share expertise and to a greater or lesser extent harmonize procedures. However, there are also networks where the central focus is to share expertise rather than samples. They also usually have an...
objective to encourage harmonization although the nature of this harmonization does not need to be specified nor is it a requirement of membership. They may be bottom-up with biobanks spontaneously coming together to share good practice or top-down with funding provided by a national or international body to facilitate the joint working. They can share expertise on the processing and storage of data and samples or training of junior researchers. They may publicize procedures used by existing biobanks so that newer biobanks do not have to ‘re-invent the wheel’, thus facilitating sharing data and tissues as there will be more alignment of protocols.

Table 8. Population cohorts

<table>
<thead>
<tr>
<th>Main coverage</th>
<th>Further information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prospective birth cohorts</strong></td>
<td></td>
</tr>
<tr>
<td>Generation R</td>
<td>The Netherlands (Rotterdam)</td>
</tr>
<tr>
<td>Danish National Birth cohort</td>
<td>Denmark (Copenhagen)</td>
</tr>
<tr>
<td>Born in Bradford</td>
<td>United Kingdom (Bradford)</td>
</tr>
<tr>
<td><strong>Prospective adult cohort</strong></td>
<td></td>
</tr>
<tr>
<td>Aussie Normals</td>
<td>Australia (Canberra/Sydney)</td>
</tr>
<tr>
<td>UK Biobank</td>
<td>United Kingdom (Manchester)</td>
</tr>
<tr>
<td>Airwave Health Monitoring Study</td>
<td>United Kingdom (London)</td>
</tr>
<tr>
<td>The Genome Database of Latvian Population</td>
<td>Latvia (Riga)</td>
</tr>
<tr>
<td>CARTaGENE</td>
<td>Quebec, Canada (Montreal)</td>
</tr>
<tr>
<td>The Rotterdam study</td>
<td>The Netherlands (Rotterdam)</td>
</tr>
<tr>
<td>EPIC Denmark</td>
<td>Denmark (Copenhagen and Aarhus)</td>
</tr>
<tr>
<td>Copenhagen City Heart and General Population Studies</td>
<td>Denmark (Copenhagen)</td>
</tr>
<tr>
<td><strong>Retrospective adult cohorts</strong></td>
<td></td>
</tr>
<tr>
<td>Post-War cohort of obesity among Danish young men liable for military service</td>
<td>Denmark (Copenhagen)</td>
</tr>
<tr>
<td>ERF/GRIP Research in genetically isolated populations</td>
<td>Southwest Netherlands (Rotterdam)</td>
</tr>
</tbody>
</table>

BBMRI = Biobanking and Biomolecular Resources Research Infrastructure; EPIC = European Prospective Investigation of Cancer; ERF/GRIP = Erasmus Rucphen Familieonderzoek/Genetic Research in Isolated Populations; P3G =Public Population Project in Genomics.
Many expertise networks have a significant function around ethics and governance. Harmonization of ethics procedures such as informed consent will help sharing of samples between countries with different legislative and governance frameworks, but also allows pooling of expertise on ethics and governance issues for biobanks which may not have the time or expertise to develop high-quality effective ethics and governance procedures.

**Networking of Population Cohorts** *(table 8)*

Population cohorts are usually prospective, comprising adults or children as part of a birth cohort. However, retrospective cohorts also exist, exploiting historical data on exposures, and follow-up data collected later (often prospectively) on morbidity and mortality. Retrospective cohorts are heavily dependent on the quality of the exposure data, as it was not originally collected for research purposes. For example, one of the biobanks visited utilized the data collected between 1943 and 1974 to assess young men's fitness for military service. The Danish Central Person register allowed all the obese men and some of normal body mass index within the original cohort to permit follow-up, in some cases 40 years after the original data were collected [21, 22].

Population biobanks should normally be designed on the basis of sample size calculations, meaning that in principle they should already have sufficient power to answer research questions for common disease. Rarer diseases will occur within the cohort, and in theory these samples could be shared with other biobanks to generate sufficient sample size. Despite the overwhelming size of many prospective population biobanks, greater networking between such cohorts would be of benefit to further understanding of the most common causes of morbidity and mortality [23]. In order to understand the aetiology of common diseases, sample sizes may be required that were greater than any existing national prospective biobank. The European Science Foundation [11] urged more coordination and harmonization. Despite this, most population cohorts do not network other than within expertise networks. This may reflect the fact that most population biobanks are still in the recruitment and data/sample collection phases, and so have no samples to share at this moment. However, since many of the population cohorts are national initiatives and are high profile with media and political attention, there may be other factors at play that would inhibit the formation of a network to share samples.

**Barriers for Networking**

Ethics and legislation are perhaps the most commonly quoted excuses given by researchers for ineffective networking. Much has been written about ethical problems with biobanks, in particular in relation to consent [24]. However, these concerns largely relate to prospective biobanks where it is difficult to describe the sort of research that is likely to utilize the biobank. With the exception of a few examples, e.g. genomeEUtwin [25], most networks incorporate retrospective case-control biobanks, and issues relating to ‘blanket consent’ or ‘consent waivers’ are not as relevant for retrospective biobanks. Retrospective biobanks rarely collect any data after the initial recruitment period other than perhaps mortality data. Thus participants know (approximately) what sort of information is going to be recorded (usually basic diagnosis and demographic data). They also know that their samples/data are only likely to be used for research into their particular disease. The main uncertainty is around who will have access to the samples and where the samples will be sent. However, given the importance of the time and effort tied up in a biobank, it may be convenient for principal investigators to be able to blame external factors such as ethics committees as the main reason why they do not share their precious resource with others.

The benefits for the people offering to share their samples may not be immediately obvious. It could be considered counter-intuitive to give away samples that may have taken significant amounts of time and money to collect. While researchers are altruistic most of the time and keen to see research advance in their specialist area, ownership of samples, especially of rare conditions, is a means of entry into collaborations with other research teams. While this is not always made explicit, access to samples may be conditional on inclusion in research bids and publications. There may also be an expectation that transfers of samples will be reciprocal with a research team both giving and receiving samples, such that all benefit from free exchange. While scientists can do this outside a network, the prominence of a network can increase the probability that they will be asked for samples and hence collaboration. Given that funding can be precarious, funders will be looking for evidence of outputs from previous funding when considering requests for extensions. An important element of this evidence would be publications resulting from research on samples obtained from within the network or through collaboration with research groups secured by virtue of network membership. Morente and Alonso [10] thought that of their list of require-
ments for effective networking an open mentality for sharing was the most important. But they also noted that it was frequently the most difficult to achieve.

**Inter- and Intrabiobank Network Taxonomy**

It was possible to allocate all of the biobanks visited to one of the network categories. Some fitted better than others. Thus the classification may not be mutually exclusive nor encompass all types of biobanks. Indeed, other categories of network may exist or emerge over time. However, the starting point for this paper was not a desire to develop a new classification system, but to find a way of better understanding the problems faced by biobanks.

Many of the governance and operational problems associated with the biobanks visited were due to networking functions: either intra- or inter-biobank networks. Thus, this proposed classification system should help better understand these issues, and more importantly aid identification of solutions.

**Acknowledgements**

This paper is based on research conducted for the GeneBanC project funded by the European Commission’s 6th Framework Programme (SAS6-CT-2006-036751-GeneBanC). Thanks to the various biobanks for their help with the interviewees. Thanks to the reviewers of an earlier draft for their helpful suggestions.

**References**

2. Council of Europe, Committee of Ministers: Recommendation Rec(2006)4 of the Committee of Ministers to member states on research on biological materials of human origin (adopted by the Committee of Ministers on 15 March 2006 at the 958th meeting of the Ministers’ Deputies) https://wcd.coe.int/ViewDoc.jsp?id=977859.