biopool

SERVICES ASSOCIATED TO DIGITALISED CONTENTS OF TISSUES IN BIOBANKS ACROSS EUROPE

Review of Impact of Data Pooling in Biobanks

Deliverable 7.1.2

Project acronym:	BIOPOOL
Grant Agreement:	296162
Version:	v1.0
Due date:	Month 3
Submission date:	18/12/12
Dissemination level:	Internal
Author:	Roberto Bilbao (BIOEF)

SEVENTH FRAMEWORK PROGRAMME

Part of the Seventh Framework Programme
Funded by the EC - DG INFSO

Table of Contents

1 Document History

Version	Status	Date
V1.0	Delivered to EU	18/12/2012
V0.1	Francesco Moscone, Elisa Tosetti	08/12/2012
V0.2	Review by Irune Vegas, Oihana Belar, Roberto Bilbao	12/12/2012
V0.3	Review by Aranzazu Bereciartua	14/12/2012

Approval		
	Name	Date
Prepared	Oihana Belar, Francesco Moscone, Elisa Tosetti	08/12/2012
Reviewed	Aranzazu Bereciartua	16/12/2012
Authorized	Roberto Bilbao	18/12/2012
Circulation		
Recipient	Date of submission	
Project partners	08/10/2012	
Project partners (2nd review)	16/12/2012	
European Commission	18/12/2012	

2 Introduction

The past few decades have witnessed rapid growth of aging population. in the last 50 years, in the OECD countries, life expectancy for males has increased markedly, rising from an average of 12.7 years in 1960 to 16.8 in 2007, suggesting not only that greater numbers of individuals are reaching old age but also that elderly people are living longer (Jagger et al., 2008). All this has caused a rapid growth in health expenditure, putting pressure on policy makers about the economic and financial sustainability of many healthcare systems. That technological progress has an important impact both on health outcomes and spending is well known. Novel technologies indeed can improve the diagnosis, prevention, treatment and management of disease, and accelerate advances in personalized medicine. All this is possible if we can access to comprehensive and well-organized collections of human biological samples and associated clinical and research data. These collections, referred to as biobanks, are crucial infrastructures for scientific and medical research, ultimately impacting on both public health and individual patient care.

The aim of this document is to review the benefits of biobanking, and to assess the economic costs associated to biobank's activities. In the following, we first review what biobanks are, and their use. We then focus on the benefits of using biobanks for improving health and healthcare and for carrying research. Finally, we provide an overview of the economic costs associated to maintaining a biobank.

3 Biobanks

Biobanks are organised collections of biological specimens and associated data on their donors, for medical-scientific research and diagnostic purposes. Examples of biological specimens are cells, tissue and blood, as well as DNA. Depending on the purpose of the biobank, genetic information on the donors and/or health- and lifestyle-related information may be associated with the samples.

Biobanking is not a new concept, as the collection of samples and data for research purposes has a long history in the arena of medicine. Biospecimens have been collected and stored in conjunction with clinical and epidemiological studies for several decades. In the past, biorepositories were relatively uncontroversial, and were mostly stored in pathology institutes. In recent years, the scale of biobanking activities, in terms of quantity of samples and data as well as the number of institutions involved have increased considerably. Further, these collections are being configured so that they can be used as a resource for the whole scientific community. Such expansion of biobanking activity has raised ethical, regulatory, technical, and managerial questions that still need to be addressed (European Commission, 2012).

Biobanks vary considerably in size, scope and focus. Samples can be collected during routine clinical care for therapeutic or diagnostic purposes, gathered during clinical trials as part of specific research projects, or could be collected as part of population-based biobanks. Roger et al. (2008) distinguish between three major types of human biobanks:

1. **Population banks**. Their primary goal is to obtain *biomarkers of susceptibility* and *identity*, and their operational substrate is germinal-line DNA from a huge number of healthy donors, representative of a concrete country/region or ethnic cohort.
2. **Disease-oriented banks for epidemiology**. Their activity is focused *on biomarkers of exposure*, using a huge number of samples, usually following a healthy exposed cohort/ case-control design, and studying germinal-line DNA or serum markers and a great amount of specifically designed and collected data.
3. **Disease-oriented general biobanks**. Their primary objective is to find *biomarkers of disease* through prospective and/or retrospective collections of tumour and no-tumour samples and their derivates (DNA/RNA/proteins), usually associated to clinical data and sometimes associated to clinical trials. Those data are usually not collected for a concrete research project, except in case of clinical trials, but from the healthcare clinical records.

Recent years have witnessed the rising need for collaboration between research groups, and combined efforts to obtain high-quality large biobanks as infrastructure for future studies. This is because sufficient numbers of samples must be available in order for the results to be valid, and often single-centre studies cannot achieve them. Even when large-scale biobanks are used, they may not contain enough samples to study rare diseases. Several recent initiatives, both at national and international level, have emerged throughout Europe and worldwide. One notable

example is the Biobanking and Biomedical Resources Infrastructure (BBMRI),[1] a network of over 280 biobanks throughout Europe, which recently became a legal entity. The aim of BBMRI is to create a research infrastructure within Europe by linking existing clinical collections and harmonizing and standardizing them, as well providing an ethical, legal and societal guidance platform.

The role of biobanks and repositories in these networks is to act as "nodes" on the flow of information between researchers and institutions enabling data and samples to be stored, organised and reconfigured for the use in different projects (Kaye, 2011).

3.1 Virtual biobanks

In the last few years, the increase in computational power and digital image acquisition has allowed the development of new approaches in the representation and storage of specimens' information. One important technology that has completely revolutionized traditional pathology practice is the virtual microscopy. **Virtual microscopy** (virtual slide system or virtual pathology slide) is the technique of digitalizing an entire glass microscope slide at the highest resolution, to produce a "digital virtual slide" with diagnostic image quality (Teodorovic et al., 2006). Such "digital virtual slide" can be used in conjunction with image processing software tools to view, manipulate, position and specify the magnification of the image on screen, in the same way one would use regular microscope to view the original glass slide. As the glass slide is now available in a digital or virtual format, it is possible to use the image for archival, replication, transferring over networks, remote consultation, as well as integration with other media types for educational use on the web (Teodorovic et al., 2006).

Virtual microscope can be used to generate tissue histological images, to study the composition of cells, glands, tissues and organs and the diseases that affect them. Histological tissue images are used by pathologists to confirm presence or absence of diseases, disease grading or measurement of disease progression, or in some cases, best type of treatment for the patient, by researchers to discover and validate biomarkers, and by lecturers to create comprehensive resources to train students. The use of scanned images reduces errors in tissue selection thus diminishing unnecessary retrieval, shipping and the associated potential degradation of valuable samples. We refer to Dee (2009) for an overview of virtual microscopy technology.

Virtual biobanks are large databases and can provide high-resolution images of samples as well as other data. There are several advantages in using the virtual microscope with online databases. First, it is less expensive than replicating, creating and storing glass slide sets. Secondly, the digitalized slides have a high resolution, while glass slides are fragile, can be easily broken or damage, and can fade. Finally, its use does not depend on availability of space, equipment or specimens. Hence, digital histological images are potential sources of information and knowledge, and making them available to the scientific and medical community is desirable.

[1] See http://www.bbmri.eu.

One important example of virtual biobank is the OECI-TuBaFrost[2] project. This is a consortium of 11 biobanks established to create a virtual tissue banking where tissue samples remain stored at the collector's institute, and the clinical, sample inventory as well as digital images are stored in a central databased. See Riegman et al. (2006) for further details.

[2] www.tubafrost.org.

4 Benefits of biobanking

Data pooling coming from biobanks or networks of biobanks has great potential to the understanding of diseases and their treatments. However, we observe that there is still little empirical evidence to support the hypothesis that the use, application, and dissemination of biobank's resources may generate positive outcomes. In this section, we review the benefits of biobanking on health and health care, and how these may translate into economics benefits. We focus on large biobanks, such as a national biorepository, or multi-country network of biobanks. In our review, we adopt a classification of benefits similar to that proposed by Rogers et al. (2011).

4.1 Improvements in patient diagnosis and therapeutic care

There is ample evidence that the preservation of the integrity of biospecimen molecular and architectural components has a direct impact on the reliability of the analyses performed on these specimens. High-quality biospecimens with appropriate clinical annotation are critical to the advancing of basic and translational cancer research, particularly with the development of genomic research, in which researchers are searching for the specific genes and proteins responsible for cancers' start and spread (Rogers et al., 2011). The current system for collecting and maintaining biospecimens, however, is often described as decentralized and *ad hoc*, resulting in poor quality, inefficient processes, and limited utility. A survey in 2011 on more than 700 cancer researchers found that 47 per cent of them had trouble finding samples of sufficient quality. Low-quality biospecimens resulted in 60 per cent questioning their findings and 81 per cent limiting the scope of their work (Massett et al., 2011).

Performing standardised, evidence-based procedures and quality controls in tissue preservation process in biobanks may play an important role in improving biospecimen integrity, and reducing DNA, RNA and protein degradation, ultimately increasing the accuracy of genomic and proteomic analyses for better diagnosis (Baker, 2012).

Using biobanks' resources has the advantage of exploiting a larger set of information on patient's individual characteristics, by combining, or pooling, data of different types. Data pooling allows the adoption of a "disease network" approach that integrates genetic, genomic, biochemical, cellular, physiological, and clinical data to create a network that can be used to model predictively disease expression and response to therapy. Loscalzo et al (2007) discuss the beneficial consequences of using such disease network analysis, identifying the following major key benefits:

1. Identify the determinants or combinations of determinants that strongly influence network behaviour and disease expression
2. Give insights into disease mechanism and potential therapeutic targets
3. Help studying diseases as the interaction between genomic and environmental factors

4. Help in defining phenotypic differences among individuals with the same disease through consideration of unique genetic and environmental factors that govern intermediate phenotypes contributing to disease expression
5. Offer a method for identifying therapeutic targets or combinations of targets that can alter disease expression.

Disease network analysis offers a novel approach to human disease classification, since it defines disease expression on the basis of its molecular and environmental elements in a holistic and fully deterministic way. Although the application of these principles to specific diseases is still in its infancy, the integration of genome-based knowledge into epidemiological and public health research, policies and health services for the benefit of all can be considered as one of the most important future challenges that our health care systems will face. We refer to Brand et al. (2012) for further discussion.

4.2 Impacts on epidemiological research

Biobanks can be used not only for basic research aimed at understanding fundamental biological principles such as molecular mechanisms, but also for clinical and epidemiological research. Biobanks in which human bodily substances are collected and linked to information on donors' states of health or lifestyle, as well as on their working and environmental conditions, may be expected to contribute to identifying the causes and mechanisms of diseases. By investigating the frequency and distribution of diseases among the population, epidemiologists have in many cases established correlations between environmental factors and the incidence of disease. For example, epidemiological methods have established links between a number of cancers and chemical substances to which sufferers were exposed at work. The high prevalence of limb malformations in newborns in the late 1950s was ultimately found to be due to the mothers' ingestion of thalidomide during pregnancy. With today's ever greater understanding of the human genome, the methods of epidemiological research can increasingly identify not only "external" but also "internal" pathogenic factors. This applies in particular to the correlation between diseases and genetic predisposition (genetic epidemiology).

Large-scale prospective studies have several advantages for a reliable quantification of the combined effects of lifestyle, environment, genotype and other determinants of disease. In particular, they allow the study of a large range of conditions, avoid recall bias as exposures can be assessed prior to disease development, and allow investigation of factors that may be affected by disease processes and consequent treatment.

From the above discussion it becomes clear that biobanks form a necessary infrastructure for high-quality epidemiological research, as they allow disentangling the association between genetic background, lifestyle and environmental determinants on the incidence, natural course and treatment response for various complex diseases and health traits. Also in this case, the

scale and standardization of data collection are crucial, suggesting the establishment of prospective collaborative biobanks for patients with specific disease entities.

4.3 Reduced clinical trials evaluation costs

Biobanks are largely used in the pharmaceutical industry for research purposes mostly in connection with clinical drug trials. The aim may be, for example, to identify molecular drug targets in cells, or to discover genetic factors responsible for the various effects and side-effects of drugs observed in patients (pharmacogenetics). Such biobank research may help improving the efficacy of drugs, while at the same time reducing their side-effects.

One important barrier to clinical research is the availability of a sufficient number of clinical biospecimens, permitting the validation and verification of new theories. Biospecimens residing in various repositories may differ in quality because there are no standardized procedures for collection and storage across tissue collection sites. In addition, biospecimens may vary in actual availability due to differing repository access privileges and specimen-associated patient consent policies. In addition, researchers acquiring tissue samples and associated data for their studies often need to exclude samples because biospecimens and associated data are not in alignment with the specific research requirement. This results in the need to replace excluded tissue samples to complete the laboratory analysis, or to repeat the entire research experiments, with an increase in expenses associated with the unbudgeted, incremental project time. A survey in 2011 by Massett et al. (2011) found that 39 per cent of interviewed cancer research had difficulty in obtaining biospecimens of adequate numbers. Nine in every 10 respondents reacted positively to the idea of a national biospecimen resource, with 62 per cent reporting that they would obtain biospecimens from it and 53 per cent reporting that they would be willing to contribute biospecimens to it.

A business case project[3] produced by Gartner Inc. reports that that rigorous use of data standards can significantly improve processes in a single clinical study, and enhance re-usability in “knowledge” thus saving time and cost. According to this study, it would reduce the average clinical study time taken by the industry for a trial from around 14 months to less than 6 months.

It becomes clear that a nation-wide biorepository may positively impact on clinical research and reduce the costs associated to clinical trials by disseminating standardized human tissue and associated best practices for use in clinical trials.

[3] See http://www.cdisc.org/business-case.

4.4 Benefits from the implementation of best practices

The specialized nature of biobanking requires that a standard operating procedures manual describes the following policies and procedures (see Vaugh and Lockhart, 2012):

- Biospecimen handling
- Laboratory processing
- Shipping and receiving protocols and material transfer agreements
- A record management system
- Building, personnel and biospecimen security
- Safety and waste disposal
- Procedures to investigate, document and report on staff injuries and dangerous exposures
- Equipment maintenance, repair and calibration records.

Implementing best practices and setting appropriate standard operating procedure could potentially result in time savings for activities related to preparing and processing biospecimens. As a centre of excellence, a large biobank would promote best practices as well as provide valuable information to other commercial and university biorepositories to help them avoid contaminating or compromising proprietary specimen collections (Rogers, et al., 2011).

5 Economic costs of biobanks

In the following, we review the costs of developing and maintaining biobanking operations. In doing this, we mainly follow the work by Vaugh et al. (2011b), and the classification therein.

- **Case collection**. This category mainly includes costs for: tissue acquisition, pathology/histology review, informed consent documentation, case data collection, Health Insurance Portability and Accountability Act adherence, preparation of collection kits, and the shipping of those kits back to the biobank
- **Tissue processing**. This category includes costs for: data annotation, confirmation of diagnosis, detailed pathology work-up including digital imaging and image analysis, molecular analysis and genetic marker tests, establishment of a comprehensive case profile, and any bioinformatics system data entry that is necessary, quality review to test the integrity of collected biospecimens
- **Storage management**. Storage costs are specific to the type of biospecimen collected (e.g, blood, tissue, DNA, etc.), and will vary with the volume of each item. Bard and Frome (2005) provide a detailed description of the steps needed to implement a biologic repository and the associated costs. According to their study, storage infrastructure costs include: liquid nitrogen freezers, mechanical freezers, and room temperature cabinet and rack systems. Freezer storage areas need to be constantly monitored by video and alarm systems, requiring dedicated staff or use of a contracted service firm. Other costs include freezer maintenance and periodic recalibration as well as backup power systems. Last, storage management will also include costs associated with following standard protocols, such as periodic auditing, inventory control and reconciliation, and certification of biospecimen identity to ensure the accuracy of data maintained in the biobank's biomedical informatics, laboratory information management, or other enterprise resource planning systems
- **Sample distribution.** This category of costs is associated with the retrieval and distribution of biospecimens to customers in the research community. Following an order of a particular sample made by the customer, the sample must be retrieved from inventory, packaged and shipped. Other costs entail verification of order receipts and validation that samples received met quality and fit-for-use standards. System integration costs must be considered as well to ensure that order entry systems are integrated with the biobank's financial accounting system to record transactions, affix the appropriate pricing, and generate customer invoices
- **Infrastructure and administration**. This category mainly includes costs for administrative personnel required to carry out enterprise operations. Major capital investments are required to purchase or build out the site; design and outfit laboratory facilities; install fire prevention, building access, security and surveillance systems; furnish offices; and to install backup power generator systems to maintain freezer storage stability and ensure business continuity. In addition, comprehensive biomedical informatics systems are necessary, as they are critical to optimizing the value of biospecimen collections. Other information

technology costs that must be considered are the data centre, storage, local networks, as well as any web-based portals and applications that are necessary for customers to access the biobank for research or business purposes.

Given the above described costs, it is clear that a nation-wide biorepository resource could potentially offer smaller organizations that are contemplating development of their own biobank, or struggling with budget limitations, an alternative to such capital investment by leveraging existing infrastructure for either storage management or bioinformatics systems and applications.

6 Concluding remarks

Biobanks can have a crucial role in revealing disease etiology, and advancing public health. However, from the above discussion it emerges that one necessary condition for biobanking to be a powerful platform for health innovation and knowledge generation is the strong collaboration and networking between different research groups and institutions. This is because often single-centre studies cannot achieve sufficient numbers of samples needed for research results to be valid. Our discussion also points at standardisation of biobanks as of fundamental importance, in order to reduce the clinical study time and make the analysis, like a medical trial, more cost-effective. Finally, the understanding of biobanking by the media and the general public as well as policy makers need to be enhanced, for example through the development of formal tools for the evaluation of the impact of biobanks (Harris et al., 2012).

7 References

[1] Baird P.M., Frome R.J. (2005), Large-scale repository design. Cell Preservation Technology, 3, 256-266.

[2] Baker, M. (2012), Building better biobanks, Nature, 486, 143-146.

[3] Brand A., Schulte in den Bäumen T., Probst-Henschet N.M. (2012), Biobanking for Public Health, in Dabrock P. et al. (eds.), Trust in Biobanking Dealing with Ethical, Legal and Social Issues in an Emerging Field of Biotechnology, Springer.

[4] Dee F. (2009), Virtual microscopy in pathology education, Human Pathology, 40, 1112-1121.

[5] European Commission (2012), Biobanks for Europe: a challenge for governance, Report of the Expert Group on Dealing with Ethical and Regulatory Challenges of International Biobank Research.

[6] Harris J.R., et al. (2012), Toward a roadmap in global biobanking for health, European Journal of Human Genetics, 20, 1105-1111.

[7] Jagger, C., Gillies, C., Moscone, F. and EHLEIS Team. (2008), Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis, The Lancet 372, 2114- 2114.

[8] Kaye J. (2011), From single biobanks to international networks: developing e-governance, Human Genetics, 130, 377-382.

[9] Riegman P.H.J., Morente M.M., Betsou F., de Blasio P., Gearye P., and the Marble Arch International Working Group on Biobanking for Biomedical Research (2008), Biobanking for better healthcare, Molecular Oncology, 2, 213-222.

[10] Riegman P.H.J. et al. (2006) TuBaFrost 1: Uniting local Frozen Tumour Banks into a European Network: an overview, European Journal of Cancer, 42, 2678-2683.

[11] Rogers J., Carolin T., Vaught J., Compton C. (2011), Biobankonomics: A Taxonomy for Evaluating the Economic Benefits of Standardized Centralized Human Biobanking for Translational Research", Journal of the National Cancer Institute Monographs, 42, 32-38.

[12] Teodorovic I. et al. (2006), TuBaFrost 6: Virtual microscopy in virtual tumour banking, European Journal of Cancer, 42, 3110-3116.

[13] Vaught J., Rogers J., Carolin T., Compton C. (2011), Biobankonomics: Developing a Sustainable Business Model Approach for the Formation of a Human Tissue Biobank, Journal of the National Cancer Institute Monographs, 42, 24-31.

[14] Vaugh J., Lockhart N.C. (2012), The evolution of biobanking best practices, Clinica Chimica Acta, forthcoming.

[15] van 't Veer L.J., Dai H., van de Vijver M.J. et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer, Nature, 415, 530-536.

[16] Zielhuis G.A. (2012), Biobanking for epidemiology, Publich Health, 126, 214-216.