Bio-bank protocols and regulations: state-of-the-art review

Project Number: FP6-2005-IST-026996
Deliverable id: D12.2
Deliverable name: Bio-bank protocols and regulations
Submission Date: 10/2/2008

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	ACGT
Project Full Name:	Advancing Clinico-Genomic Clinical Trials on Cancer: Open Grid Services for improving Medical Knowledge Discovery
Document id:	D 12.2
Document name:	Bio-bank protocols and regulations
Document type (PU, INT, RE)	PU
Version:	0.4
Submission date:	10/2/2008
Editor: Organisation: Email:	Dimitris Kafetzopoulos FORTH-IMBB kafetzo@imbb.forth.gr

Document type PU = public, INT = internal, RE = restricted

ABSTRACT:

Human specimens like blood, DNA, tissues and other biological materials are increasingly valuable resources in medical research. Bio-banks or bio-repositories refer to organized collections of biological samples and associated data. Through their strict organization they provide high-quality bio-specimens, and support discovery using advanced technology, like bioinformatics, nanotechnology, genomics, proteomics, metabolomics and molecular imaging. In recent years, many countries and private sectors have invested in bio-banks for many different reasons.

Several countries are establishing bio-banks associated with electronic health records in the hope that population genetics can speed the identification of disease susceptibility genes or

prognostic biomarkers. Concurrently, biotech companies are independently amassing private collections of DNA and tissues or seeking to collaborate with public population databases and bio-banks. The collection and storage of biological samples raises several ethical and policy issues concerning access to and use of these samples, particularly around issues related to informed consent. However no binding international regulatory framework addresses these issues. Instead, a patchwork of national laws, regulations and ethics advisory body guidelines govern the collection, storage and research use of biological samples.

This report is a state-of-the-art review on bio-banks in Europe and in the other countries. Based on our experience with ACGT (and the ACGT clinical trials) we will describe current experience in bio-banking, current regulations of ethical and legal issues, best practices in bio-specimen collection, processing, storage and distribution and use of bioinformatics systems and data management.

KEYWORD LIST: Bio-banks, biorepository, tissue bank, network, database.

MODIFICATION CONTROL			
Version	Date	Status	Author
0.3	23/11/2007	Draft	Dimitris Kafetzopoulos
0.4	4/2/2008	Draft	Dimitris Kafetzopoulos
0.5	15/3/2008	Final	Dimitris Kafetzopoulos

List of Contributors

- Giorgos Zacharioudakis (FORTH-ICS)
- Vasiliki Danilatou (FORTH-IMBB)

Contents

1 Introduction

1.1 Executive Summary

Bio-banking is an old concept as human tissues have been collected and stored at various institutions around the world for more than 100 years. Bio-banks provide the samples and information that are necessary to move toward the future of targeted therapies and a world of more personalized medicine. Research using bio-repository samples and data aims to discover the genetic basis of diseases and provide novel diagnostics, predictive biomarkers, and therapeutics that are safe, and more efficacious by understanding for which patients those products are appropriate. The promise of bio-banks is the advancement of science and medicine by providing essential high-quality information to researchers and clinicians.

The list of anticipated benefits derived from bio-banks generally include:

- new knowledge of:
  - □ disease aetiology, and natural history;
  - □ genomic contributors to health;
  - □ pathogenic and environmental contributors to disease; and
  - □ the genomic-organism-environment interaction;
- new treatments in the areas of:
  - □ pharmacology; and
  - □ genetic therapies;
- new tests to:
  - □ reduce harm from pharmacological treatments with genetic risks;
  - □ detect pathologies earlier;
  - □ personalize risk assessment and preventive strategies; and
  - □ support population-based risk assessment and preventive strategies;
- new preventive strategies to:
  - □ personalize risk assessments and dietary/environmental advice;
  - □ identify high risk populations most likely to benefit from closer follow-up;
  - □ develop medications or other treatments to supplement missing genetic functions associated with increased risk; and
  - □ stronger arguments for environmental policies.

It should be noted that bio-banks do not constitute the research that might produce these benefits; banks are an intermediate step in a process that produces a collection of data and materials that facilitate further research [1].

However, many issues concerning bio-banking have arisen around the world. The majority of issues fall under one of three category headings: scientific, clinical and laboratory requirements;

ethical, legal, and social policy (ELSP) requirements; and information technology (IT) requirements [2].

Scientific issues identified include:

- What procedures represent best practices for sample collection, processing, annotation, storage and distribution, and how can those practices be shared, standardized, and widely adopted?
- Maintaining quality controls on samples and data is a critical component of extracting real scientific and clinical benefit from these samples now and in decades to come. How can bio-banks work together to achieve high levels of quality control?
- How can researchers and clinicians find and gain access to samples and data needed to advance scientific and medical knowledge? How can bio-banks manage access while protecting privacy?
- Bio-banks have no small task in selecting appropriate terminology and taxonomy for sample classification that can serve as a common reference for researchers from a variety of scientific domains. Complicating the semantic issues are multi-lingual access and communication issues.
- Finally, vocal support and behavioral support of bio-banks from both the clinicians and scientists are critical to the long-term success of bio-banking.

Ethical, legal, and social issues include:

- Donation of a sample should not impose an increased risk for the donor and should be done in the context of patient diagnosis.
- Protection of the privacy is very important. Various national governments have already implemented laws or formal regulations to form a basis regarding the practice of bio-banking while excluding any possible misuse of the information.
- Effective informed patient consent and possibility for later withdrawal.
- Donor's access to research results.
- Returning to patient for later data.
- Ethics committee approval of studies.
- Ownership of bio-specimens.
- Intellectual property rights.
- Conflict of interests.
- Legal issues involved in international tissue exchange.
- Funding models can vary from the commercial pharmaceutical companies, to the government initiatives to the non-profit organizations.
- Concerns about commercialism.
- Ethical practices and regulations vary from country to country. The challenge of harmonization and networking is huge and a number of recent initiatives address it.

Information technology issues include:

- The need for consistent and individually unique sample identification pattern and supporting tracking systems.

- The need for clearly defined and agreed upon data standards was articulated as the IT counterpart to the semantic issue of ontologies for classification.
- The implementation of sound data quality assurance and control processes is critical to the long-term success of bio-banks because of the impact they have on accurate analysis and conclusions.
- The design and deployment of appropriate database architectures are vital to the ease with which bio-banks can scale up to manage larger collections and adapt to changing conditions.
- Data integration is a critical issue because the vision of widely agreed upon ontologies and well-defined data standards is still just an ideal. Data integration continues to be an important task both within individual bio-banks and across bio-banks in support of collaboration.
- As more and more researchers derive or create knowledge from the samples and associated data provided by bio-banks, the policies of bio-bank in managing that knowledge and the role that IT plays in implementing those policies will be important.
- One of the most critical aspects of the IT supporting bio-banking is maintaining privacy through sound IT security processes and supporting technologies. Bio-banks must be able to provide a strong assurance of confidentiality and privacy to donors, otherwise, buy-in from both the public and the scientific community will not be attained and the enterprise of bio-banking will be starved at its source.

1.2 Purpose of this document

The purpose of this document is to review the present situation on bio-banking in Europe and in the rest of the world and to describe some examples of a variety of bio-banks. We will also mention the current regulations and guidelines in human tissue research given at national or international level, guidelines and procedures in specimen collection, processing, storage and distribution, information technology tools and data processing, as well as management of ethical and legal issues.

1.3 Structure of this document

In Section 2 we will describe the current experience and the different types of bio-banks around the world which fall under four major categories: government-funded, private industry, academic and other non-profit organizations and international initiatives. In Section 3 we will also mention the current European, national and international regulations for human tissue research and we will give special emphasize at the current bioethical legislation in Germany, Belgium, UK and Greece. In Section 4 we will discuss some ethical and legal issues such as informed consent, risks, privacy, access and sharing, ownership of banked samples as well as special issues in research without adequate consent and in forensic settings. In Section 5 we will provide some recommendations concerning bio-specimen collection and processing, measures concerning facilities, quality assurance and quality control standards, data collection, annotation, storage conditions, and specimen distribution. In Section 6, we will discuss the use of bioinformatics systems in bio-banking, the types of data contained, data integration challenges, data accessibility, security and quality control of bioinformatics systems. The report is concluded in Section 7.

2 The impact of bio-specimen banking

The healthcare and life science communities have been buzzing with the promise of personalized medicine since the inception of the Human Genome Project and other large-scale scientific studies. The combination of bio-banking, high throughput research together with bioinformatics and the unraveling of the human genome and transcriptome have proven to be a powerful approach in biomedical research. Bio-banks supply the research community with biological samples and molecular, tissue, and clinical data. These samples, along with associated patient data, can be used to discover the genetic basis of the diseases as well as determine the protein biomarkers that may predict disease onset, prognosis, progression or therapy response. These indicators of disease or biomarkers can then be developed into clinically administered diagnostic tools to better predict and monitor disease. Research efforts utilizing clinical specimens hope to elucidate the information found in the human genome and enable personalized medicine by generating the information needed to get the right medicine to the right patient at the right time [3].

To fully realize the vision of personalized medicine, researchers and clinicians need access to two critical types of information. The first is **molecular information**, including genomics, proteomics, and other high-throughput molecular data. The second is **clinical information**, including data contained in medical records or in clinical trial records. The creation and analysis of both types of information have exploded in recent years — but only by seamlessly integrating these data types to reveal the complex underlying causes and outcomes of diseases can effective, personalized treatments be realized [2].

The promise of post-genomic biology is a detailed, molecular-level understanding of normal and disease biology. While some progress has been made in this direction, it is also a fact that fundamental discoveries have not been translated into effective therapeutic interventions yet. It is becoming increasingly clear that to translate advances made in understanding the molecular underpinnings of diseases into effective therapeutic interventions, there has to be a seamless integration between "bench and bedside." **Translational medicine** aims to bridge this gap by providing effective biomarkers of safety and efficacy and also to identify subsets of populations that would respond to a drug therapy. In order to put translational medicine into practice and to test the findings of discovery research, we need a well-annotated repository of disease-associated and normal tissue samples.

Samples from clinical trials are an integral part of translational medicine and are invaluable for the clinical development process. By helping to provide disease-relevant genetic markers and understand the expression and/or activation of the drug target, patient samples may aid in the selection of a population that will most likely benefit from a given treatment. In addition, samples taken throughout a clinical trial provide critical information relating to drug action, safety, and efficacy. Of importance here is the identification of biomarkers that correlate with the specific mechanism of drug action. In the ideal case, these may also have predictive value regarding disease prevention or therapeutic response. Clinical tissue samples would also be useful in choosing from a plurality of clinical indications and designing proof-of-concept clinical trials. Thus, clinical tissue samples form an integral part of the effort in translating research findings to effective clinical treatments.

Research on disease-associated samples is being used to elucidate the relationship between genotype and phenotype and help identify the genetic cause of diseases. Two new methods of conducting research — translational research and molecular profiling — rely heavily on

biological samples with associated clinical information. In particular, bio-banking is vital to closing the loop in this model of research and development (R&D), feeding back into research relevant biological samples, associated pathology reports, molecular characterization, and the associated treatments and clinical outcomes.

- **Translational research** is science that can be translated from the laboratory bench to the patient's bedside; that is, turning laboratory data into novel diagnostics and therapeutics.
- **Molecular profiling** flows in the opposite direction — from the patient's bedside to the scientist's bench — thus closing the loop between scientific research and clinical medicine. Molecular profiling interrogates clinical samples for a variety of biomolecules, including DNA, RNA, proteins, and metabolites (*Figure 1*).

Bio-banks — sometimes called **bio-repositories** or **tissue banks** — provide a natural focal point for both streams of information, thereby serving as a critical bridge to enable translational research. Bio-banks have the robust capacity to house large collections of well-defined and processed samples and data sets. This capability is a key for longitudinal clinical studies that are tracked over extended periods. Longitudinal data are critical for understanding the long-term benefits and/or side effects of drugs and various treatment options as well as the influence of various environmental factors. The data housed inside bio-banks can potentially reduce the time needed to develop a drug or diagnostic marker by providing the critical number of samples and information needed during R&D. Bio-banks can accelerate development in two main ways: by providing better logistical or operational support to technical users and by providing scientific and business users with a more seamless understanding of data, from molecule to clinic. For example, pharmacogenomic and toxicogenomic studies utilizing these samples may lead to the creation of safer and more efficacious therapeutics. Large-scale genotyping experiments may uncover the reason why some patients respond, and some do not, to a particular medicine. Moreover, these samples may help in the drug approval process by identifying which patient population will respond positively or negatively to an experimental drug. For these reasons and many more, bio-banking is an important bridge between basic and applied research that is currently attracting much attention and many resources. Collaboration among scientists, clinicians, governments, and the public will be required for these endeavors to succeed. At the same time, the advanced technical capabilities that enable the linkage of genetic data with clinical information have raised ethical, legal, and social concerns. Individual donors of bio-specimens and data for clinical research must feel confident that the privacy of their medical information will be honored by the research community.

The increasing demand for biological samples with associated clinical data is providing an impetus for the establishment of bio-banks. Hence, bio-banking has emerged as a dynamic new discipline conducted by a variety of institutions, including hospitals performing clinical research, large pharmaceutical companies engaged in clinical trials, and government-funded programs supporting healthcare-based initiatives. A growing number of bio-banks have established networks of collection sites throughout research hospitals and clinics and obtain a variety of samples linked to a broad range of diseases.

The collection of various biological samples, including tissue, cells, and bodily fluids such as blood, is central to bio-banking (*Figure 2*). Biological samples collected in a clinical setting are usually obtained under signed consent from an individual. Recorded with each biological sample are clinical and phenotypic data that often include general patient information, medical history, family history, lifestyle data, treatment history, geographic information, physician, date of

collection, collection address, and, if appropriate, diagnostic testing, pathology results, and molecular profiling data.

Scientific research using clinical samples (DNA, RNA, proteins, cells, tissue, as well as blood and other fluids) has dramatically increased over the past five years and will continue to increase for decades to come. The demand for clinical samples stems from numerous scientific advancements, including the ongoing elucidation of the human genome, the implementation of high-throughput "-omics" technologies, and the further development of more intelligent bioinformatics tools and algorithms. By providing researchers with adequate number of high-quality biological samples, processed with standardized procedures and associated with clinical data, bio-banks can accelerate medical discovery and improve health.

In particular researchers in the cancer field have called repeatedly for bio-specimens to be collected using standardized protocols, so that results can be reproducible and comparable. The heterogeneity in the sample processing affects the histological and molecular analysis and results in variable results that do not necessarily reflect biology. They also seek greater research data accessibility through an open, Web-based platform, while remaining committed to the proposition that the collection and use of bio-specimens and associated data must meet the highest possible ethical standards for protecting the privacy and confidentiality of the donor. They also recognize that the usefulness of bio-specimens is maximized if accompanied by relevant demographic, social history, clinical, pathology, and longitudinal data, as well as genomic and/or proteomic data. A searchable, Web-based bioinformatics system therefore is seen as crucial for facilitating scientific discovery. Many investigators also have expressed a desire for the services that accompany tissue sample analysis, such as tissue microarrays and DNA or RNA assays [2].

The potential sources for bio-specimens are expected to be derived primarily from academic medical centers and community hospitals. The potential users would be primarily scientists and researchers at academic institutions, government agencies, and biotech and pharmaceutical companies. The potential uses of bio-specimens and their associated data are many, including the following purposes:

- Target- and validation-discovery of molecular biomarkers*:* Primarily using RNA or protein analysis methods (large and small scale)
- Genomic analysis
    - DNA sequencing
    - SNP mapping and association studies
    - Mutation screening
    - Loss of heterozygosity and amplification studies
    - Methylation and other epigenetic studies
- Validation of diagnostic or therapeutic antibodies, or nucleic acid probes
- Drug discovery
- Pharmacogenomic analysis
- Comparative analysis of normal and disease-associated samples
- Tissue histology

- Generation of cell lines and *in vitro* studies
- Laser-capture microdissection (LCMD) [4].

Bio-banks have varied goals, objectives, and statements of purpose. For instance, governments and healthcare institutions are often interested in bio-banking as a way to manage their future healthcare costs; informatics companies are interested because of the data management challenges; and pharmaceutical and biotechnology companies are interested in understanding what drugs to develop and who will respond to those drugs. Some bio-banks exist for diagnostic purposes (e.g., pathology), some for therapeutic treatment (e.g., blood banks), and some purely for research of specific diseases or specific populations. Moreover bio-banks use different approaches to exploit tissue samples. Some collect material from a specific population whereas others study a specific disease. Most of them are project-driven and collect tissues from specific diseases or organs. Less frequently banks systematically collect tissues from a variety of diseases or organs. A bio-bank's mandate determines the types of material collected and stored and the scope of research that is performed on the samples. Because of the associated ethical and regulatory issues surrounding the storage of an individual's genetic and medical information, a well-articulated and well-protected mandate is critical to build public trust and avoid abuse. In addition, a well-articulated mandate sets the expectations for its contributors, end users (whether research or clinical), and the public [2].

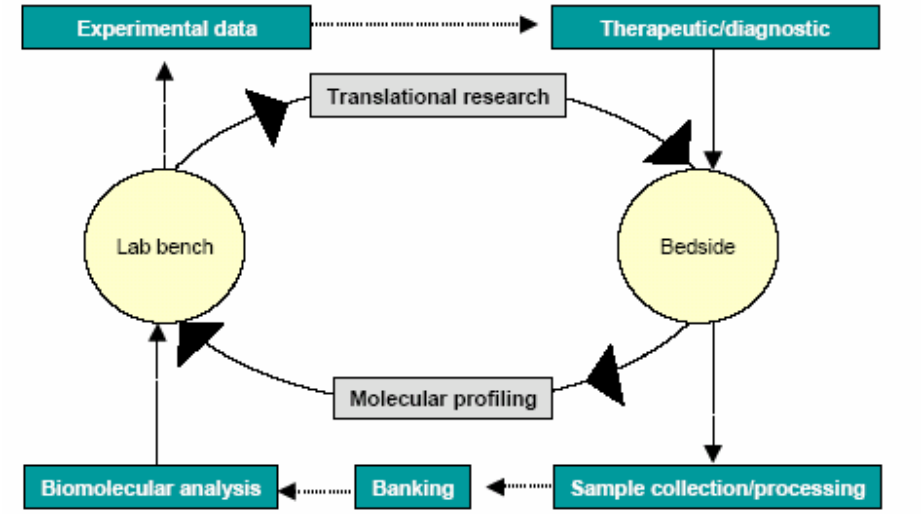

Figure 1. Illustration of the circular flow connecting the lab bench to the bedside [2].

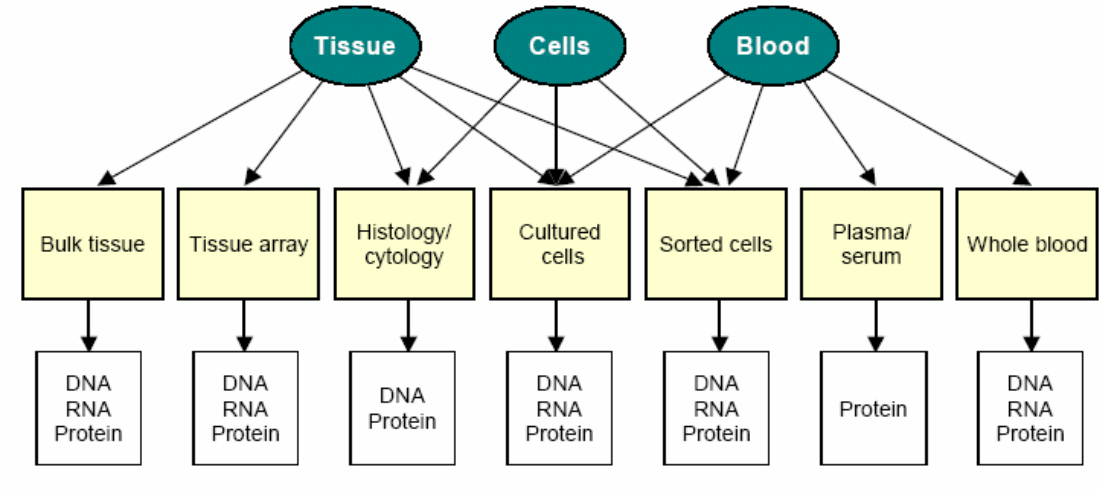

Figure 2. Illustratrion of the types of samples collected by repositories [2].

According the funding and the interests of bio-banks four major categories are recognized: Government-funded, private industrial, academic and other non-for-profit organizations and international initiatives.

2.1 Government-funded Bio-bank Initiatives

In recent years there is an increasing interest from various governments in sponsoring bio-banks, particularly to help understand and/or manage the health of their populations. Bio-banks are increasingly considered an important component of a national health strategy because understanding the genetic and environmental bases of disease and therapy response is critical to managing the healthcare of a population. The focus and mandates of national bio-banks vary based on a country's priorities. Some countries are primarily interested in controlling their long-term healthcare costs. Some countries with homogeneous populations with well-understood genealogies are turning that information base into a powerful asset. Some countries have a long history of individual banks, but now they are trying to manage or control those banks in a more collective or central way so that they can leverage new genetic and information technologies as well as apply ethical and social policies in a more consistent manner [2].

2.1.0 Estonian Genome Project

In Estonia, the **Estonian Genome Project** was founded as an agreement between the Estonian government and the Estonian Genome Centre Foundation, a non-profit organization, founded in January 1999, by Estonian scientists, doctors, and politicians to support genetic research in Estonia. Initially it was supported by private money from Egeen (Mountain View, CA, USA). Since January 2004, the project has received additional funding from the Estonian Government. Initiated in October 2002, the project plans to collect the next 10 years DNA from 1 million Estonian adults and children together with health and genealogical data. This population represents 70% of Estonian population. In addition to identifying the genetic causes of the diseases, the republic recognized several social objectives that go beyond healthcare itself. Within the framework of the project, a foundation was established to "collect health and genetic data of the Estonian population" and "use the results of genetic research to improve public health." In order to collect data, participants are required to complete a consent form, a

questionnaire and to donate a blood sample. Personal information on the questionnaires is replaced with a 16-digit code and stored in the coding centre. Health data are separated from personal information and stored in the database of the Gene Bank. This bank may be used only for scientific research and treatment purposes and was set up specifically to be of interest to research institutions as well as bioinformatics, biotechnology, and pharmaceutical companies "because it enables the analysis of genotype and phenotype relationships based on an entire population" [5, 6].

2.1.1 UK Bio-bank

UK Bio-bank has resulted from the collaboration between the Welcome Trust, the Medical Research Council and the Department of Health in June 1999. The UK Bio-bank has begun to collect genetic samples from half a million volunteers aged 45-69, which, when combined with their medical information, will help explain the effects of genetics and environmental factors in causing common adult diseases. In forming this national undertaking, the government was sensitive to public views of the collection of biological samples. Several objections have been stated, among them the inefficiency of large population- based cohorts versus specific disease case control studies, the narrow age range of the study which does not represents the general population, the long projected time-lines, and the limited phenotyping [2, 5, 7, 8].

2.1.2 Latvian Genome Project

The Latvian Genome Project was launced in January 2001 and funded by the Latvian Council of Science. The aim of the project is to establish a genome database of the Latvian population over a period of ten years. Samples as well as medical, genealogical and lifestyle information will be collected from the inhabitants of Latvia and coded. While this project has numerous objects, generally its main purpose is to study DNA polymorphisms in the Latvian population and their ethnogenesis in order to improve knowledge of monogenetic and multi-factorial diseases and ultimately to generate personalized prevention, diagnosis, and treatment tools. Scientists from various medical and educational institutions (the Latvian Medical Academy, the State Centre of Medical Genetics, the University of Latvia and the Biomedical Research and Study Centre) are involved in the study [5].

2.1.3 CARTaGene

In Canada, the **CARTaGENE** project was initiated in 1999 by a multidisciplinary team of the Quebec Network of Applied Genetic Medicine (RMGA). CARTaGENE ‘s aim is to create a database of genetic, physiological, medical, social and environmental data and a bio-bank for blood samples in order to draw a first map of the genetic diversity of the population of Quebec, Canada. CARTaGENE project has been supported by public money from Genome Canada and will collect DNA from more than 50,000 adults in Quebec between 25 and 69 years (representing approximately 1% of Quebec population) [5, 9].

2.1.4 Genome Austria Genome Bank (GATiB)

In Austria the **Genome Austria Tissue Bank (GATiB)** has collected samples from 700,000 patients, from 1983 to 2002, totaling 2.7 million diseased and normal tissue specimens. Major emphasis has been put on annotation of archival tissue with comprehensive clinical data, including follow-up data. GATiB has focused on both archival and prospectively collected material. One important aspect of the program is that all samples have been collected and processed in one institute, so that the same tissue-processing protocol was applied over the years. An assay for RNA quality control in formaldehyde-fixed, paraffin –embedded (FFPE) tissues has been developed, which is based on the quantification of the fragmented size of

preserved mRNA. A specific IT infrastructure was developed that supports sample annotation, tracking of sample usage as well as sample and data storage. Innovative data protection tools were developed which prevent sample re-identification. GATiB does not act as a tissue distributor but supports specific research projects [10].

2.1.5 National Bio-bank Program of Sweden

In Sweden, the **National Bio-bank Program** was established as a joint national program of Swegene and Wallenberg Consortium North and supported by the Wallenberg Foundation and by the major universities of Sweden. The project aims to a network of Swedish bio-banks with the following objectives:

- To increase the knowledge, quality, usefulness, efficiency and accessibility of the Swedish bio-banking system
- To increase the safety of individual donors and promote ethical awareness [11]

The major Swedish Bio-banks that participate in the program are four population-based research bio-banks (containing>185,000 samples), pathology bio-banks (containing more than 10 million specimens) and microbiological serology bio-banks (containing more that 1.7 million serum and plasma samples and expanding by 120,000 samples per year) [12].

2.1.6 HUNT database and CONOR-The Norway experience

In Norway, the Nord-Trøndelag Health Study (HUNT) is a result of several joint efforts. The main objectives have been both epidemiological and clinical research and preventive medicine. National institutions took interest in HUNT due to the need for population-based studies on a national level. In 1992, the Research Council of Norway initiated discussions concerning the need for population-based studies in Norway. Representatives from the Ministry of Health, the four Norwegian Universities, the National Health Screening Service, the National Institute of Public Health and the Cancer Registry participated in this debate. The conclusion was that Norway was particularly suitable for population-based studies, and that Norway should be obligated to carry out such studies, to generate new medical knowledge.

The HUNT database is a result of tight collaboration and joint actions between local, regional, national and international partners through the last 20 years:

- General practitioners and district nurses in the whole county

- The municipal (24 municipalities) and county authorities

- The two local hospitals, i.e. Levanger Hospital and Namsos Hospital

- A number of private regional and national organisations

- The Norwegian Institute of Public Health, The National Health Screening Service (from 2002 are both part of The Norwegian Institute of Public Health), and a number of other national institutions and universities

- The Norwegian Research Council

- The Ministry of Health

HUNT is also part of the nation-wide CONOR (Cohort Norway) collaboration, constituting a network of national health databases and bio-banks, in which HUNT is the largest single unit. HUNT has through several years initiated collaboration with various research groups, and has been invited to collaborate with research groups in other European countries and in the USA

[13].

A number of large population-based surveys have been conducted in Norway since the beginning of the 1970s. In the late 1980s the Research Council of Norway established a programme in epidemiology. This also gave stimulus to the idea of establishing a cohort including both core survey data and stored blood samples. In the early 1990s, all universities, the National Health Screening Service, The National Institute of Public Health and the Cancer Registry discussed the possibility of a national representative cohort. The issue of storing blood samples for future analyses raised some concern and it was discussed in the parliament. In 1994, the Ministry of Health appointed the Steering Committee for the CONOR collaboration. In 1994–95, the fourth round of the Tromsø Study was conducted, and became the first survey to provide data and blood samples for CONOR. During the years 1994–2003, a number of health surveys that were carried out in other counties and cities also provided similar data for the network. So far, 10 different surveys have provided data and blood samples for CONOR. The administrative responsibility for CONOR was given to the Norwegian Institute of Public Health (NIPH) in 2002. CONOR is currently a collaboration research between the NIPH and the Universities of Bergen, Oslo, Tromsø and Trondheim. In 2002, CONOR and the Norwegian Mother and Child study (MoBa), received a 5-year grant from the Norwegian Research Council to build a technology platform under the Functional Genomics programme (FUGE), called the Bio-banks for Health in Norway (Biohealth) platform. The overall aim was to investigate separate and combined effects of genes and environment on the risk of disease [14].

2.1.7 Japan Bio-bank

The “Personalized medicine” project started in June 2003 by the support of the Japanese government. The aims of this project are 1) discovery of genes susceptible to diseases, or those related to effectiveness or adverse reactions of various drugs, 2) identification of molecular targets for evidence-based development of drugs or diagnostic tools, 3) identification of the important genetic information that can be applied for establishment of “Personalized Medicine” and 4) studies on gene-environment interaction for prevention of diseases. To achieve these goals DNAs, sera and clinical information from 300,000 patients who have either of 47 common diseases were collected by March 2007. As the research resource bank, “Bio-bank Japan” was constructed that consisted of facilities that are able to store DNAs (the maximum capacity of 1,000,000 tubes; fully-automated sample handling system) and sera (the maximum capacity of 3,000,000 tubes in liquid nitrogen; semi-automated handling system). Clinical information database was also established as a part of “Bio-bank Japan” activity. For protection of individual privacy two measures have been taken 1) a two-step anonymization of individual identification code by two-dimension code system and 2) avoidance of placing individual identification information with genotyping information together. By January 15, 2006, a written informed consent from more than 130,000 patients was obtained (a total number of disease cases is more than 185,000 because of multiple diseases in one patient) from 66 hospitals participating in this project. Genome-wide association studies using 250,000 SNPs have been performed to identify genes of medical importance. These SNPs can cover 98.5% of the human genome [15].

2.1.8 Genome EUtwin project

The genome EUtwin project is coordinated by the National Public Health Institute and the University of Helsinki and received funding under the European Commission’s “Quality of Life and Management of Living Resources” of the 5th Framework Programme. This program aims to apply and develop new molecular and statistical strategies to analyze unique European twin and other population cohorts to define and characterize the genetic, environmental and lifestyle

components in the background of health problems. The population cohorts of the project are derived from a variety of sources including the Danish, Dutch, Finnish, Italian, Norwegian and Swedish twin cohorts. The core facilities are DNA isolation and genotyping (Helsinki, Uppsala), epidemiological expertise (Odense), database expertise (Stockholm), and bio-computing expertise (Leiden) [5].

2.1.9 Singapore Tissue Network

The Singapore Tissue Network (STN) is a national, non-profit research tissue and DNA bank funded since March 2002 by Singapore Biomedical Research Council, Agency for Science, Technology and Research, Ministry of Health and the Genome Institute of Singapore. Genomic information from Singapore population will be collected via a network of collaborating organizations and hospitals [16].

2.1.10 US Bio-banks Initiatives

The **Armed Force Institute of Pathology (AFIP)** National Pathology Repository, the single largest repository in the world, stores more than 92 million specimens [17].

The **Marshfield Medical Clinic** in Wisconsin is one of the largest population-based bio-banks in USA, which contains information from 40,000 participants living in northern and central Wisconsin. Some consider this effort to be a small step away from being characterized as a statewide population-based bio-bank. New Mexico recently proposed legislation to study the feasibility of collecting DNA samples from all residents in the state. This initiative is not designed for genetic research, but rather for forensic purposes such as crime-solving, clearing the wrongly accused and convicted, and identifying unknown persons.

Many other research institutions in USA have developed population-based bio-banks, drawing volunteers from neighbouring areas. The **NUgene Project**, which is developed at Northwestern University, is currently collecting and storing DNA linked with medical records of volunteer patients from neighbouring hospitals and clinics. Hospitals and clinics associated with Duke University participate in a DNA bio-bank, as do the University of Alabama and Mayo Clinics. The proposed African American Population Bio-bank, under the auspices of Howard University, plans to enlist 25,000 volunteers over 5 years. The main objective is to develop a population-based genetic epidemiology resource for the study of common complex diseases among members of the African Diaspora. The United States Department of Veterans Affairs has proposed a comprehensive bio-bank of veterans and their family members.

A number of other large-scale tissue-banking projects, primarily building on existing epidemiologic studies, have been proposed or are in various stages of development using a variety of bio-samples. These include: Physicians Health Study, Nurses Health Study, Multi-Ethnic Cohort, Women's Health Initiative, Cancer Prevention Study II of the American Cancer Society, and the National Health and Nutrition Examination Survey III (**NHANES III**) of the Centers for Disease Control and Prevention (CDC). The **Framingham Heart Study**, a longitudinal study started in 1948, recently included a genetics component involving 9000 volunteers from within the study. Recent efforts at the NIH have broadened the infrastructure to include multiple participating research institutions. The Environmental Genome Project at the National Institute of Environmental Health Sciences is a population-based effort to characterize a number of important genes and to relate variations in genotype with susceptibility to effects of chemical and physical agents. Started in 1998, this project is assessing approximately 200 genes based on analyses of 1000 samples of human DNA. Also located at the National Institutes of Health (NIH) is the recently launched **Genetic Association Information Network (GAIN)**, which is a public-private partner- ship in conjunction with Pfizer Global Research and

Affymetrix formed to study the genetic determinants of common diseases. GAIN is structured to capitalize on the existence of enormous numbers of well-characterized bio-banks housed in hospitals and academic centers across the United States. Via a competitive application process, the project is intended to augment existing, well-characterized case-control studies by providing free genotyping on existing DNA samples from these studies. All results will be made available in the public domain [18].

The **National Cancer Institute (**NCI) at the National Institutes of Health (NIH) supports numerous tissue resources, including the NCI Cooperative Group Human Tissue Resources CHTN), the Early Detection Research Network (EDRN), the Specialized Programs of Research Excellence (SPOREs), and NCI intramural collections. The evaluation of NCI-supported tissue resources entailed gathering background information about all the tissue resources using the Internet and available literature, performing site visits and interviews at representative repositories (e.g., CHTN, EDRN, and the Tissue Array Research Program [TARP]), and interviewing key personnel involved in coordinating NCI's cancer specimen resources [19].

2.2 Private Industry Bio-banks

In recent years there is an emergence of private companies that collect and analyze samples and personal information. Some private companies have arisen to conduct research using data from national sub-populations. Examples are DeCode Genetics (see description below), Oxagen (Abingdon, Oxfordshire, United Kingdom) and Galileo Genomics (St. Laurent, Quebec, Canada, see description below). Other bio-banks operate solely as tissue and sample banks. Many of these bio-banks are relatively new companies. Some United States-based examples include Ardais (see description below), Genomics Collaborative (recently acquired by SeraCare), Asterand, ILSbio, and others. These banks vary by size of repository, source and type of samples, target market, amount of associated clinical data, services offered, and so forth, but most are focused on selling their samples to the pharmaceutical and biotech research community [2].

In addition to buying from commercial bio-banks, the pharmaceutical and biotech companies build up their own repositories as they collect samples and medical data from enrollees in clinical trials. Biotech and pharmaceutical firms are interested in knowing the optimal candidates for their products so that they can develop appropriate products, conduct efficient clinical trials, and target products in the market effectively. Once the drug has been designed and developed, knowing responders and non-responders can be critical to the drug's success in the market. An example is the recent withdrawal of the nonsteroidal anti-inflammatory drug Vioxx by Merck and Co. An application of the information concerning response or sensitivity to a drug is the Food and Drug Administration (FDA) approval of label changes for two anticancer drugs, 6-mercaptopurine (6-MP) and irinotecan, to include pharmacogenetic testing as a potential means to reduce the rate of severe toxic events.

2.2.1 DeCode Genetics

One of the most well-known private industry endeavors is led by deCode Genetics (Reykjavik, Iceland), which has created a bank of genetic samples from 100,000 volunteers in Iceland. Iceland has 270,000 citizens and has the unique characteristic of being isolated for more that 1000 years. Moreover the country has extensive computerized genealogical records going back 10 centuries, population-wide clinical records beginning with the National Health Service in 1915, a large human tissue bank from the 40s and an educated cooperative population. The goals of the project are the study of the genetic cause of common diseases and the indexing of

the heredity of the entire nation of Iceland. deCODE has already discovered variations in the Icelander's genome that may indicate susceptibility to multiple sclerosis, hereditary hand tremors, and osteoarthritis. Many objections have been stated against this program such as the monopoly and the commercial aspects of deCODE, the potential violation of personal privacy, the presumed consent, the potential future racial discrimination and exclusion of Icelanders from insurance companies [5, 17, 20, 21].

2.2.2 Genomics Collaborative, Inc (GCI)

GCI was established in 1998 primarily as a for-profit private biotechnology research company designed to participate in and facilitate the application of genetic research to drug and diagnostic discovery decisions. Specimen collection began in earnest in March 2000 to collect multiple types of specimens (e.g., tissue, serum, and DNA) on the same patient, along with detailed medical and demographic information. GCI's approach links human genes, proteins, and clinical outcomes through proprietary technology platforms. GCI offers human DNA, RNA, sera, and snap-frozen tissue specimens linked to detailed medical information collected from patient populations worldwide. GCI personnel provide expertise in designing and conducting human genetic studies geared toward the development of therapeutics and diagnostics. They also offer high-throughput analysis tools, including single nucleotide polymorphism (SNP) genotyping, DNA sequencing, reverse transcriptase polymerase chain reaction (RT-PCR), and gene expression analyses. GCI uses a dual business model. It has a fee-for-service side that works primarily with the pharmaceutical industry to design and collect specimens for drug development. The other side of the business participates in larger collaborative programs with pharmaceutical companies, biotech companies, and academic and government institutions [17].

2.2.3 IMPATH Inc

IMPATH Inc., a private company formed in 1988 to improve outcomes for cancer patients by providing cancer information and analyses, was identified because it has a database of over one million patient profiles and outcomes data on over 2.3 million individuals and because it represents a for-profit repository model. The Vice President and Scientific Director at IMPATH was initially eager to participate in the study; however, once the interview request was referred to the legal department, the process was held up over concerns about proprietary issues [17].

2.2.4 Ardais Corporation

Ardais Corporation is a private clinical genomics company that grew out of a mutual interest among researchers at Duke University Medical Center and Ardais founders to develop a tissue/data banking centre that could simultaneously support the needs of Duke internal researchers and the broader research community. Ardais, in collaboration with its network of partner medical institutions, launched the National Clinical Genomics Initiative in September 2000, after obtaining initial funding in December 1999, to facilitate genomics-based biomedical research among academic and industrial researchers. The goals of the initiative are to develop systematic, large scale procedures to comprehensively collect, process, and store research quality clinical materials and associated information; to provide these resources in optimized formats for biomedical research; and to support the research and clinical programs at participating medical institutions. Ardais has established best practices working groups to provide advice, continually review, and ensure that operations are ethically appropriate, technically excellent, and practical [17].

2.2.5 Galileo Genomics (St. Laurent, Quebec, Canada)

Galileo is a genomics company focused on understanding the genes and biomarkers associated with the root cause of selected diseases and drug responses that it is leveraging from the genetic information of the Quebec population. Since the ancestry of almost 70% of the current Quebec population can be traced back (thanks to comprehensive genealogical databases) to some 2,600 founders with low levels of intermarriage, a high level of genetic sharing in the population is critical to speeding such research.

2.3 Academic and Non-Profit Organization Bio-banks

Many non-profit organizations exist around the world, and some focus on specific populations, such as women. In the United States, for example, the Women's Health Initiative was funded by the National Institutes of Health (NIH) to collect the biological specimens and health records of some 170,000 women aged 50-79. The initial focus was on understanding the effects of hormone replacement therapy and diet modification, but the biological specimens, data, and medical information have now been extended to include other diseases such as heart disease and osteoporosis to understand who develops the disease and why. Other bio-banks focus on certain diseases. For example, the National Marrow Donor Program and the Breast Cancer Family Registries collect biological and medical data on populations with specific cancers. They then share this information with researchers on a limited basis to help understand the causes and cures for the disease of interest.

Some of these projects are international in reach, driven by the need to share research and clinical data. For example, the dbMHC database is a cooperative effort of the National Center for Biotechnology Information (NCBI), the NIH, and the International Histocompatibility Working Group (IHWG) focused on providing publicly accessible databases for DNA and clinical data related to the human major histocompatibility complex (MHC). This database has two major components: reagents to trace DNA typing and anonymous clinical data from individuals engaged in MHC research projects.

Various universities and research hospitals are involved as well. In the United States, for example, the University of North Carolina is working with the National Institute of Environmental Health Sciences (NIEHS) to build the Environmental Polymorphism Registry, which will house DNA samples from some 20,000 patients, in order to study links between genotypes, environmental factors, and disease status. In Sweden, the Karolinska Institutet opened a bio-bank as "a national, non-commercial resource for collection, handling, and storage of human biological material aiming at promoting scientific excellence within molecular and genetic research." Typically, such a bio-bank limits use to researchers who must submit a specific research plan to a scientific advisory board.

Historically, many of these disease-based organizations and institutions have tended to operate independently. Now, cross-organization activity is increasing, in large part due to the amount, types, and complexity of information involved. For example, in North America, the Multiple Myeloma Foundation recently spearheaded the formation of the Multiple Myeloma Research Consortium (MMRC). The consortium comprises the Dana-Farber Cancer Institute (Boston, MA), the H. Lee Moffitt Cancer Center (Tampa, FL), the Mayo Clinic (Rochester, MN), and the University Health Network (Princess Margaret Hospital, Toronto). The goal is to advance the drug development process and, ultimately, find a cure for multiple myeloma. A unique element of the MMRC is the presence of the MMRC Tissue Bank, a resource of tissue and corresponding genomic and clinical data, and the MMRC Data Bank, a new information system that integrates

laboratory and clinical trial data. "By having one source of genomic data and tissue samples that have been uniformly standardized, as well as an integrated validation and multi-center clinical trials system, this will reduce the drug development process by months and possibly years," according to Kenneth Anderson, M.D., chairman of the MMRC.

This type of collaboration is developing on the international level as well. The work of the MMRC will also complement the work of the International Myeloma Foundation (IMF), which started the "Bank On A Cure" DNA bank in 2002 with the participation of some 20 myeloma institutions that are part of the IMF's International Working Group. The bank contains several thousand DNA samples with associated clinical data. Collecting and analyzing the DNA of myeloma patients are critical elements of the bank's fulfilling its purpose of helping to understand the genetic factors related to the disease, understanding individuals' response to treatment, and enabling the development of new therapies [2].

The **Specialized Programs of Research Excellence (SPOREs)** are funded through specialized center grants that promote interdisciplinary research and move basic research findings from the laboratory to clinical settings, involving both cancer patients and populations at risk of cancer. The outcome of interdisciplinary research is a bidirectional approach to translational research, moving laboratory discoveries to clinical settings or clinical observations to the laboratory environment. Laboratory and clinical scientists share the common goal of bringing novel ideas to clinical care settings that have the potential to reduce cancer incidence and mortality as well as improve survival and the quality of life. In order to achieve these goals, SPORE investigators work collaboratively to plan, design and implement research programs that may impact cancer prevention, detection, diagnosis, and treatment. Additionally, SPOREs approach these goals through collaborative efforts within the individual multidisciplinary SPORE teams, inter-SPORE collaborations, partnerships with other NCI/NIH programs, and public-private partnerships with industry and non-profit organizations. Key qualities of the program feature the inclusion of patient advocates in SPORE activities and international cooperation with investigators in Europe, Canada, Asia, and Mexico [22]. The Prostate SPORE National Bio-specimen Network (NBN) Pilot is a study designed to pilot key aspects of the NBN concept as described in the NBN Blueprint. The development of the NBN Blueprint was spearheaded by members of the National Dialogue on Cancer (NDC), which, in addition to researchers, includes representation from the National Cancer Institute (NCI), patient advocacy groups, and the pharmaceutical and medical diagnostics industries. The NBN concept calls for a national, “best practices”-based tissue resource to manage the standardized collection, processing, storage and distribution of high-quality bio-specimens and linked data to support and reduce variability in translational research. The Prostate SPORE NBN pilot will support the development of a common bio-specimen coordination system and informatics infrastructure to support collaborative projects related to prostate cancer research currently underway at participating SPOREs across the nation.

Duke University Breast SPORE. Duke University initially received funding for a Breast SPORE in 1995. It then lost funding for the SPORE in 2001 but was recently re-funded, in July 2003. The Tissue Resource Core is an integrated part of a pre-existing repository established by a molecular biology researcher and a surgeon in 1987 [17].

Mayo Clinic Prostate SPORE. Mayo Clinic Prostate SPORE was established in 2001. The goal of the program is to identify genetic susceptibility factors for prostate cancer that can improve the understanding of the aetiology of the disease and potentially identify men at increased risk of developing prostate cancer for whom prevention strategies might be targeted. The Prostate

Cancer Tissue Procurement Core of Mayo Clinic Prostate SPORE is an integrated part of an ongoing program at the Mayo Clinic to collect, process, and store tissue from prostate cancer patients. The goal of the Tissue Procurement Core is to procure prostate tissue from every prostate cancer patient undergoing radical prostatectomy at the Mayo Clinic. The Tissue Procurement Core is also electronically integrated with the Prostate Cancer Patient Registry and the Biostatistics Core to provide investigators with clinically annotated specimens [17].

University of Alabama at Birmingham Breast and Ovarian SPOREs.

The UAB Breast SPORE, established in 2001, is focused on the areas of breast cancer prevention, including genetics, chemoprevention, and therapy. The UAB Ovarian SPORE, established in 1999, is focused on areas of gene therapy, targeted immunotherapy, and chemoprevention for ovarian cancer. Both SPOREs compliment ongoing programs at the UAB Comprehensive Cancer Center in the areas of breast and ovarian cancer. UAB also has a newly funded Pancreatic SPORE. A Brain SPORE with an associated tissue resource is not associated with the breast, ovarian, and pancreatic tissue resources. The UAB Breast and Ovarian SPOREs Tissue Resource Cores are integrated with pre-existing shared facilities at the UAB Comprehensive Cancer Center. The goal of the UAB Breast and Ovarian SPOREs Tissue Resource Cores is to collect well-characterized breast and ovarian tumor specimens and matching adjacent specimens, along with clinical and demographic information, for use by SPORE members and by selected extramural users for special research purposes [17].

University of Pittsburgh Health Sciences Tissue Bank (HSTB)

University of Pittsburgh HSTB was established in 1991. It performs both prospective tissue collection and banking of resected tumours from the prostate, gastrointestinal tract, lung, liver, breast, head and neck, gynaecological sites, muscle, and skin. Its organ specific database (OSD) integration engine automatically integrates the tissue bank inventory system, the pathology report, and the longitudinal and outcome data from the cancer registry system. It also includes results from DNA microarray experiments and is both minable and Web-based. The repository is primarily publicly funded. The University of Pittsburgh HSTB OSD integration engine is the basis for the Pennsylvania Cancer Alliance Bioinformatics Consortium, a partnership comprising six institutes in Pennsylvania that are sharing tissue resources and data to enhance translational and clinical cancer research. This consortium has been funded to develop “a statewide serum and tissue repository, a data model for biomarker data storage, a statewide model for bioinformatics, a public access website for disseminating research results, and a strategic plan to support aggressive collaboration between industry and academia” [17, 23].

The HSTB initially started banking tissues as a resource for researchers at the university, but the program has grown to include NCI-funded resources (the Cooperative Prostate Cancer Tissue Resource [CPCTR], an EDRN Gastrointestinal grant, and a Lung SPORE), and institute-funded programs (a melanoma banking program and a cancer biomarkers laboratory).

2.4 International initiatives

To exploit the full potential of bio-banks, networking between individual bio-banks is indispensable. International tissue-banking initiatives have been financed temporarily in diverse areas, some more successful than others. In the future, there is a need of larger infrastructure to even enable studies of multi-factorial diseases in an efficient way. Such studies are in need of many more than a few hundred well-defined samples and can easily increase to thousands. In parallel, there will also be an increasing demand for more accurate and standardized sample

annotation. It would therefore be mutually beneficial to unite the successful initiatives again into cooperative and communicative networks opening up new challenges for future medical research [3].

2.4.1 European Bio-banking and Biomolecular Resources Research Infrastructure (BBMRI)

BBMRI is pan-European and broadly accessible network of existing and de novo bio-banks and biomolecular resources. The infrastructure will include samples from patients and healthy persons (with links to epidemiological and health care information), molecular genomic resources and bio-computational tools to optimally exploit this resource for global biomedical research.

The Preparatory Phase of BBMRI (2007) will focus on technical, legal, governance, and financial issues:

- to prepare to construct BBMRI, building on existing bio-banks, resources and technologies, specifically complemented with innovative components and properly embedded into European scientific, ethical, legal and societal frameworks,
- to provide the concept for a key resource to increase excellence and efficacy in biomedical sciences, drug development and public health,
- to expand and secure competitiveness of European research and industry in a global context and
- to develop a sustainable financial framework.

Biomedical quality-assessed samples and data as well as bio-molecular resources and molecular analysis tools are essential for academic and industry driven research to treat and prevent human diseases. Although currently established national bio-banks and bio-molecular resources are a unique European strength, valuable collections typically suffer from fragmentation of the European bio-banking-related research community. This hampers the collation of biological samples and data from different bio-banks required to achieve sufficient statistical power. Moreover, it results in duplication of effort and jeopardises sustainability due to the lack of long-term funding. BBMRI will comprise:

- bio-banks of different formats (collections of blood, DNA, tissue, etc., together with medical, environmental, life-style and follow-up data),
- biomolecular resources (antibody and affinity binder collections, ORF clone collections, siRNA libraries, proteins, cellular resources etc.),
- enabling technologies and high-throughput analysis platforms and molecular tools to decipher gene, protein and metabolite functions and their interactions,
- harmonized standards for sample collection, storage, pre-analytics and analysis,
- harmonized databases and bio-computing infrastructure and
- ethical, legal and societal guidance platform [24].

2.4.2 EuroBio-bank

The EuroBio-bank network was established in 2001 by two patient organisations: The French Muscular Dystrophy Association (AFM) and the European Organisation for Rare Diseases (Eurordis), who has since coordinated the network. The EuroBio-bank network is the first operating network of bio-banks in Europe providing human DNA, cell and tissue samples as a service to the scientific community conducting research on rare diseases. It is the only network dedicated to rare disease research in Europe.

- A total of approximately 170,000 samples are available via the online catalogue - 145 cell

collections, 544 DNA collections and 282 tissue collections.

- The network is currently composed of 15 members from 7 European countries (France, Germany, Hungary, Italy, Malta, Slovenia and Spain): 11 academic or private bio-banks, 2 IT services companies, 1 expert in bio-banking management and Eurordis (European Organisation for Rare Diseases) who has coordinated the network since its creation.
- Originally funded by the European Commission between 2003-2006, EuroBio-bank has received funding through grants from the AFM (French Association against Myopathies) and DGM (German Association for Patients with Muscle Disorders) and through membership fees. Since January 2007, the network has participated in the European Network of Excellence TREAT-NMD (FP6) with Eurordis taking the leadership of WP04.1:"Develop and Manage Supranational Bio-banks ". As such, it receives EC funding for several of its activities, including coordination of the network and hosting of the website, while each bio-bank of the network is financed by its own national institution or charitable organisation [25].

2.4.3 US Cooperative Human Tissue Network (CHTN)

The CHTN was formed in 1987 by three organizations along with a subcontract with the Children's Cancer Group (CCG). In 1991 the Pediatric Division became an independent group and an additional adult collection center was added to the CHTN. The University of Virginia and Vanderbilt University joined the network in 2001. All six organizations that currently comprise the CHTN have extensive experience in providing human tissues for research and include: the Ohio State University (OSU), the Pediatric Division, located at Children's Hospital of Columbus, the University of Alabama at Birmingham's (UAB), the University of Pennsylvania (UPENN), the University of Virginia and the Vanderbilt University. The number of samples distributed by CHTN since 1987 is 720,527.

The CHTN has contributed to a wide range of scientific advances in cancer research and research on other diseases. During the last 10 years, more than 1500 publications have resulted from studies using CHTN specimens. CHTN specimens have contributed to discoveries of the role of genetic alterations in cancer initiation, progression and metastasis and studies to improve diagnostic accuracy and classification of tumours. The CHTN has provided specimens needed for the development of emerging technologies and the application of these technologies to study problems in cancer biology, and to develop markers for diagnosis, prognosis and prediction of response to therapy. In recent years, the CHTN has supported a wide spectrum of onco-genomic studies including:

- Characterization of epigenetic lesions as non-random and tumour specific type events.
- Studies of gene expression profiles and studies to determine the clinical usefulness of such profiles.
- Identification of novel diagnostic molecular markers by exploiting genome wide expression tools.
- Methylation microarray analysis of cancers [26].

2.4.4 The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) pilot project is a joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), both part of NIH. The TCGA pilot was announced in December 2005 to test the feasibility of a large-scale, systematic

approach to identifying the changes that occur in the genomes of cancer cells. The goal is to generate genomic information that the research community can use to develop new and improved strategies for detecting, treating and, ultimately, preventing cancer. The TCGA has two key components, the Cancer Biomedical Informatics Grid (**caBIG™**), a network connecting individuals and institutions to enable the sharing of data and tools, and the Bio-specimen Core Resource (BCR), a centralized collection center to review and process bio-specimens and associated data [27].

2.4.5 Human Genome Epidemiology Network (HuGENet™)

The Human Genome Epidemiology Network or HuGENet™ is a global collaboration of individuals and organizations committed to the assessment of the impact of human genome variation on population health and how genetic information can be used to improve health and prevent disease. Since 2001, HuGENet™ has maintained a database of published, population-based epidemiologic studies of human genes extracted and curated from PubMed [28].

2.4.6 COGENE

With the objective to co-ordinate and further genomic research related to human health in Europe, the European Commission created, in November 2000, the "Forum of Genome Program Managers" with representatives from 25 countries. Funded by the European Union under the Quality of Life Program of the 5th Framework Program (FP5) the associated "Co-ordination of genome research across Europe" (COGENE) aimed to implement the Forum's objectives through a web service that listed the European human genetic research databases and offered data organized by country, on genome research in the European Union. Administered by the Academy of Finland, COGENE ended in 2004 [5].

2.4.7 Public Population Project in Genomics (P^{3}G)

The Public Population Project in Genomics (P^{3}G) is a non-for-profit international consortium to promote collaboration between researchers in the field of population genomics. P³G aims to facilitate international harmonization between major population bio-banks and longitudinal studies, as well as, to offer a focal point for knowledge sharing and collaboration. P^{3}G is primarily funded by its partners (up to 75% of the costs) and by Genome Canada. Thus far, P^{3}G has formed three main international working groups: i) social, environmental, biochemical and investment, ii) knowledge, curation and IT and iii) ethics, governance and public engagement [5, 29].

2.4.8 Shared Pathology Informatics Network (SPIN)

The Shared Pathology Informatics Network (SPIN) is a 5 year project (2001-2006) initiated by NCI. The SPIN will use state-of-the-art informatics techniques to establish an Internet-based virtual database that will allow investigators to locate appropriate human tissue specimens for their research. The SPIN will develop software to give researchers limited access to de-identified patient data. Accessible data will be associated with archived tissue specimens. Institutions will have full control over any data leaving their sites, and there will be no centralized database that stores SPIN data.

The need for this system has been fuelled by the growing use of tissues, diagnostic specimens, and their related clinical data in modern biomedical research. There is a wealth of archived tissue, and most pathology laboratories have at least ten years of pathology reports stored electronically. Searchable databases with patient data exist at hospitals and medical institutions.

Increasingly sophisticated software tools are now available that can facilitate communications among disparate computer systems, even among those that employ different architectures and search strategies. This combination of the need for tissue and the opportunities provided by new internet technologies led the NCI to initiate SPIN.

The systems being developed:

- will allow searches of the archived information in multiple institutions
- strip patient records of all identifiers
- encrypt patient information for security during transmission

Challenges in addition to the technical hurdle of secure transmission of information stored in different institutions include:

- the ability to accurately identify required comparable data that has been stored in multiple formats, including free text
- developing common data elements to map to the information
- designing the structure of the report that responds to the query [30].

2.4.9 TubaFrost Consortium

The TubaFrost consortium is a European Project which aims to create a European Virtual Tumour Tissue Bank and was funded by the European Commission within the 5th Framework of the division “Quality of life and living resources”. The central database is the European Organization for Research and Treatment of Cancer (EORTC). The storage of the samples is decentralized at the institute of the original collection. The aims formulated in the mission statement were: Create an innovating virtual European human frozen tumour tissue bank for the whole scientific community composed of high quality frozen tumour tissue sample collections with a corresponding accurate diagnosis stored in major European cancer centres and universities, searchable through an uncomplicated query system on the Internet provided with rules for access and use of the tissues complete with a European code of conduct to comply with the various legal and ethical regulations in the different European countries [31].

In our view, the “home country control” concept that underlies the Code of Conduct generates scepticism. This principle respects the legal and ethical regulations of the country where the tissue was collected, regardless of where it is eventually used with the minimal requirement of an opt-out system as a consent procedure: if tissue may be legitimately used for certain research purposes in the country where it was collected and under whose jurisdiction the patient falls, it may also be used for such research in the country where it is sent to in the context of a scientific program, even if that other country has more strict regulations enforced for residual tissue research collected from patients under their jurisdiction. This principle does not respect the laws of the country where the research will take place and under the cloak of a scientific collaboration and coordination covers legally the use of the tissue even if the country has strict and opposite regulations.

2.4.7 US Tissue Array Research Program (TARP)

In 1999, Richard Klausner, M.D., then Director of NCI, started to develop a research program that would utilize tissue microarray technology developed by Olli Kallioniemi and colleagues at the National Human Genome Research Institute (NHGRI). TARP was conceived as a joint program (i.e., a collaborative effort) by NCI and NHGRI. NCI produced the tissue microarrays

that were to be distributed extramurally by the Eastern Division of CHTN (see above for a description of CHTN), while NHGRI was to be responsible for development of the tissue microarray technology. Since that time, the NHGRI laboratory involved in developing tissue microarray technology has ceased to exist, and the TARP laboratory has expanded to produce and develop the technology. TARP was originally in the Office of the Director, NCI, but moved to the Centers for Cancer Research. The primary objective of TARP is to develop and disseminate tissue microarrays containing samples of multiple tumors (300 to 500 tissues per array) to cancer researchers to expedite the discovery of novel cancer targets for the detection, treatment, and prevention of cancer. These microarrays provide a tool for high-throughput screening of multiple tumor tissues with immunohistochemistry, in-situ hybridization, and fluorescence in-situ hybridization (FISH) [17].

2.4.8 US Early Detection Research Network (EDRN)

EDRN was initiated by NCI in 1998 to improve methods for detecting the signatures of cancer cells. EDRN is funded through peer-reviewed Cooperative Agreements. It is a consortium for collaborative research to link the discovery of biomarkers directly to the next steps in the process of developing early detection tests. This network brings together researchers across disciplines and institutions to identify, develop, and validate biomarkers. Network participants act as a team in a streamlined process through a distributed physical network of geographically dispersed repositories with a centralized bioinformatics and data management system. The network primarily consists of approximately 40 research universities but also includes more than a dozen private companies as industry partners [17].

3 Regulations and guidelines for Human Tissue Research

3.1 Introduction

The increasing emergence of various bio-banks has raised important ethical and legal issues, for example concerning consent and confidentiality. Regulations are at an early stage in most European countries and the rules for exchanging and sharing of information and material are not clear. Some countries have their own laws and federal legislations. At the international level some organizations like WHO, UNESCO and the Council of Europe have published recommendations regarding biomedical research and bio-banks. Below we will present some of these regulations and guidelines.

3.2 The Council of Europe

Founded in 1949, the Council of Europe (COE) seeks to develop throughout Europe common and democratic principles based on the European Convention on Human Rights and other reference texts on the protection of individuals. It now has 47 member states that accept the principles of democracy, human rights and the rule of law. Public debate is of special importance in the field of bioethics and the importance of such debate on the European level cannot be underestimated. The Council of Europe provides a forum where 47 countries, their governments, parliamentarians, local authorities and civil societies can discuss this issue and find common solutions. In the most recent recommendation of the COE in March 2006, biological materials

are defined and are distinguished from ‘‘collections of biological materials’’. A definition of bio-banks is also given but restricted to ‘‘population bio-bank’’ with the following features: ‘‘1) the collection has a population basis; 2) it is established, or has been converted to supply biological materials or data for multiple future research projects; 3) it contains biological materials and associated personal data, which may include or be linked to genealogical, medical and lifestyle data and which may be regularly updated; and 4) it receives and supplies materials in a organised manner’’ [32].

3.3 The Health Council of the Netherlands

The Health Council of the Netherlands is a scientific advisory body of the government in the field of public health. In 1994, in response to the Health Secretary of Welfare, Health and Cultural Affairs' request concerning the uses of human tissues in the Netherlands, the Health Council has published a report "Proper use of human tissue". As stated in the executive summary, the principles are as follows:

● Intended use must be morally acceptable in so far as its purpose is to promote human health

● Human tissue should always be used in the greatest of care

● Patient's needs come first and the doctor should exercise openness regarding the storage and use of human tissue and must duly inform the patient thereof

● People cannot be forced to cooperate with the use of material obtained from them, even if this it is in a good cause

● Privacy of those whose material is put to further use must be respected and protected

● Collection of human tissue should not be of commercial character and should not be transferred to a third party with a view to making profit [20].

3.4 The Nuffield Council on bioethics

The Nuffield Foundation is a private foundation in England that promotes science education. In 1991, it has established the Council on Bioethics, a multidisciplinary panel of experts which would become a bioethics committee with national standing, as the British government had not established a similar committee. One of the first projects of the Nuffield Council was the analysis of the ethical and legal issues concerning the stored tissue research, and for that reason the Working Party on Human Tissue has been appointed.

In contrast to the Health Council of the Netherlands, the Nuffield Committee does not think that consent is the primary concern, and they certainly do not agree with the view regarding the ownership of the tissue. The first ethical test is the avoidance and limitation of any injury. Secondly, the working party thinks that informed consent is important but relatively limited in actual practice. They summarize their general ethical guidelines with three questions:

a) Is the removal of tissue governed by intentions which respect human beings and their bodies in that gratuitous injury is avoided?

b) If the removal of tissue was in the course of a medical treatment, was consent given by the patient?

c) If the tissue from a cadaver was donated either by a previously healthy volunteer or port-mortem, was the appropriate consent procedure followed?

The Working Party has pointed out the following conclusions and recommendations:

In terms of ethical guidelines they conclude that:

1) The use of human tissue for medical treatment and research is ethically acceptable;

2) Using human tissue in harmful ways is unacceptable because it does not respect human beings and their bodies;

3) Uses of human tissues in medical treatment and research require informed consent of the donor;

4) There are strong arguments against the commercial acquisition and supply of the human tissue;

As to desirable legal standards they recommend that:

5) Tissues removed as in part of a medical treatment they are not any longer ownership of the person from whom they have removed;

6) Medical intermediaries should supply users of human tissue on a non-profit making basis;

7) The government should join with other member states in the European Patent Convention in adopting an immorality exclusion to patents in the area of human and animal tissue.

The council's positions look simplistic and they do not cover all the aspects of ethical and legal issues [20].

3.5 The Common Rule

In USA, the Federal Policy for the Protection of Human Subjects, is known as the "Common Rule". The federal policy sets forth the most significant and comprehensive legal requirements for the conduct of research involving human participants. This policy applies only when there is federal funding or other support for research; it does not govern privately or state-funded research. The final modification of the Privacy Rule is accessible through http://privacyruleandresearch.nih.gov/.

3.6 CIOMS International Ethical Guidelines

The Council for International Organizations of Medical Sciences (CIOMS) is an international non-governmental organization in official relations with the World Health Organization (WHO). It was founded under the auspices of WHO and the United Nations Educational, Scientific and Cultural and Organization (UNESCO) in 1949 with among its mandates that of maintaining collaborative relations with the United Nations and its specialized agencies, particularly with UNESCO and WHO. In the late 1970s, CIOMS, in association with WHO, undertook its work on ethics in relation to biomedical research. It was thus that CIOMS set out, in cooperation with WHO, to prepare guidelines " to indicate how the ethical principles that should guide the conduct of biomedical research involving human subjects, as set forth in the Declaration of Helsinki, could be effectively applied, particularly in developing countries, given their socioeconomic circumstances, laws and regulations, and executive and administrative arrangements". The World Medical Association had issued the original Declaration of Helsinki in 1964 and an amended version in 1975. The outcome of the CIOMS/WHO undertaking was, in 1982, Proposed International Ethical Guidelines for Biomedical Research Involving Human Subjects.

Guideline 1: Ethical justification and scientific validity of biomedical research involving

human beings

The ethical justification of biomedical research involving human subjects is the prospect of discovering new ways of benefiting people's health. Such research can be ethically justifiable only if it is carried out in ways that respect and protect, and are fair to, the subjects of that research and are morally acceptable within the communities in which the research is carried out. Moreover, because scientifically invalid research is unethical in that it exposes research subjects to risks without possible benefit, investigators and sponsors must ensure that proposed studies involving human subjects conform to generally accepted scientific principles and are based on adequate knowledge of the pertinent scientific literature.

Guideline 2: Ethical review committees

All proposals to conduct research involving human subjects must be submitted for review of their scientific merit and ethical acceptability to one or more scientific review and ethical review committees. The review committees must be independent of the research team, and any direct financial or other material benefit they may derive from the research should not be contingent on the outcome of their review. The investigator must obtain their approval or clearance before undertaking the research. The ethical review committee should conduct further reviews as necessary in the course of the research, including monitoring of the progress of the study.

Guideline 3: Ethical review of externally sponsored research

An external sponsoring organization and individual investigators should submit the research protocol for ethical and scientific review in the country of the sponsoring organization, and the ethical standards applied should be no less stringent than they would be for research carried out in that country. The health authorities of the host country, as well as a national or local ethical review committee, should ensure that the proposed research is responsive to the health needs and priorities of the host country and meets the requisite ethical standards.

Guideline 4: Individual informed consent

For all biomedical research involving humans the investigator must obtain the voluntary informed consent of the prospective subject or, in the case of an individual who is not capable of giving informed consent, the permission of a legally authorized representative in accordance with applicable law. Waiver of informed consent is to be regarded as uncommon and exceptional, and must in all cases be approved by an ethical review committee.

Guideline 5: Obtaining informed consent: Essential information for prospective research subjects

Before requesting an individual's consent to participate in research, the investigator must provide the following information, in language or another form of communication that the individual can understand:

- that the individual is invited to participate in research, the reasons for considering the individual suitable for the research, and that participation is voluntary;
- that the individual is free to refuse to participate and will be free to withdraw from the research at any time without penalty or loss of benefits to which he or she would otherwise be entitled;
- the purpose of the research, the procedures to be carried out by the investigator and the subject, and an explanation of how the research differs from routine medical care;
- for controlled trials, an explanation of features of the research design (e.g., randomization,

double-blinding), and that the subject will not be told of the assigned treatment until the study has been completed and the blind has been broken;

- the expected duration of the individual's participation (including number and duration of visits to the research centre and the total time involved) and the possibility of early termination of the trial or of the individual's participation in it;
- whether money or other forms of material goods will be provided in return for the individual's participation and, if so, the kind and amount;
- that, after the completion of the study, subjects will be informed of the findings of the research in general, and individual subjects will be informed of any finding that relates to their particular health status;
- that subjects have the right of access to their data on demand, even if these data lack immediate clinical utility (unless the ethical review committee has approved temporary or permanent non-disclosure of data, in which case the subject should be informed of, and given, the reasons for such non-disclosure);
- any foreseeable risks, pain or discomfort, or inconvenience to the individual (or others) associated with participation in the research, including risks to the health or well-being of a subject's spouse or partner;
- the direct benefits, if any, expected to result to subjects from participating in the research
- the expected benefits of the research to the community or to society at large, or contributions to scientific knowledge;
- whether, when and how any products or interventions proven by the research to be safe and effective will be made available to subjects after they have completed their participation in the research, and whether they will be expected to pay for them;
- any currently available alternative interventions or courses of treatment;
- the provisions that will be made to ensure respect for the privacy of subjects and for the confidentiality of records in which subjects are identified;
- the limits, legal or other, to the investigators' ability to safeguard confidentiality, and the possible consequences of breaches of confidentiality;
- policy with regard to the use of results of genetic tests and familial genetic information, and the precautions in place to prevent disclosure of the results of a subject's genetic tests to immediate family relatives or to others (e.g., insurance companies or employers) without the consent of the subject;
- the sponsors of the research, the institutional affiliation of the investigators, and the nature and sources of funding for the research;
- the possible research uses, direct or secondary, of the subject`s medical records and of biological specimens taken in the course of clinical care;
- whether it is planned that biological specimens collected in the research will be destroyed at its conclusion, and, if not, details about their storage (where, how, for how long, and final disposition) and possible future use, and that subjects have the right to decide about such future use, to refuse storage, and to have the material destroyed;
- whether commercial products may be developed from biological specimens, and whether the participant will receive monetary or other benefits from the development of such products;

- whether the investigator is serving only as an investigator or as both investigator and the subject`s physician;
- the extent of the investigator's responsibility to provide medical services to the participant;
- that treatment will be provided free of charge for specified types of research-related injury or for complications associated with the research, the nature and duration of such care, the name of the organization or individual that will provide the treatment, and whether there is any uncertainty regarding funding of such treatment;
- in what way, and by what organization, the subject or the subject`s family or dependants will be compensated for disability or death resulting from such injury (or, when indicated, that there are no plans to provide such compensation);
- whether or not, in the country in which the prospective subject is invited to participate in research, the right to compensation is legally guaranteed;
- that an ethical review committee has approved or cleared the research protocol.

Guideline 6: Obtaining informed consent: Obligations of sponsors and investigators

Sponsors and investigators have a duty to:

- refrain from unjustified deception, undue influence, or intimidation;
- seek consent only after ascertaining that the prospective subject has adequate understanding of the relevant facts and of the consequences of participation and has had sufficient opportunity to consider whether to participate;
- as a general rule, obtain from each prospective subject a signed form as evidence of informed consent – investigators should justify any exceptions to this general rule and obtain the approval of the ethical review committee;
- renew the informed consent of each subject if there are significant changes in the conditions or procedures of the research or if new information becomes available that could affect the willingness of subjects to continue to participate; and,
- renew the informed consent of each subject in long-term studies at pre-determined intervals, even if there are no changes in the design or objectives of the research.

Guideline 7: Inducement to participate

Subjects may be reimbursed for lost earnings, travel costs and other expenses incurred in taking part in a study; they may also receive free medical services. Subjects, particularly those who receive no direct benefit from research, may also be paid or otherwise compensated for inconvenience and time spent. The payments should not be so large, however, or the medical services so extensive as to induce prospective subjects to consent to participate in the research against their better judgment ("undue inducement"). All payments, reimbursements and medical services provided to research subjects must have been approved by an ethical review committee.

Guideline 8: Benefits and risks of study participation

For all biomedical research involving human subjects, the investigator must ensure that potential benefits and risks are reasonably balanced and risks are minimized.

Interventions or procedures that hold out the prospect of direct diagnostic, therapeutic or preventive benefit for the individual subject must be justified by the expectation that they will be at least as advantageous to the individual subject, in the light of foreseeable risks and benefits,

as any available alternative. Risks of such 'beneficial' interventions or procedures must be justified in relation to expected benefits to the individual subject.

Risks of interventions that do not hold out the prospect of direct diagnostic, therapeutic or preventive benefit for the individual must be justified in relation to the expected benefits to society (generalizable knowledge). The risks presented by such interventions must be reasonable in relation to the importance of the knowledge to be gained.

Guideline 9: Special limitations on risk when research involves individuals who are not capable of giving informed consent

When there is ethical and scientific justification to conduct research with individuals incapable of giving informed consent, the risk from research interventions that do not hold out the prospect of direct benefit for the individual subject should be no more likely and not greater than the risk attached to routine medical or psychological examination of such persons. Slight or minor increases above such risk may be permitted when there is an overriding scientific or medical rationale for such increases and when an ethical review committee has approved them.

Guideline 10: Research in populations and communities with limited resources

- Before undertaking research in a population or community with limited resources, the sponsor and the investigator must make every effort to ensure that: the research is responsive to the health needs and the priorities of the population or community in which it is to be carried out; and
- any intervention or product developed, or knowledge generated, will be made reasonably available for the benefit of that population or community.

Guideline 11: Choice of control in clinical trials

As a general rule, research subjects in the control group of a trial of a diagnostic, therapeutic, or preventive intervention should receive an established effective intervention. In some circumstances it may be ethically acceptable to use an alternative comparator, such as placebo or "no treatment".

Placebo may be used: when there is no established effective intervention; when withholding an established effective intervention would expose subjects to, at most, temporary discomfort or delay in relief of symptoms; when use of an established effective intervention as comparator would not yield scientifically reliable results and use of placebo would not add any risk of serious or irreversible harm to the subjects.

Guideline 12: Equitable distribution of burdens and benefits in the selection of groups of subjects in research

Groups or communities to be invited to be subjects of research should be selected in such a way that the burdens and benefits of the research will be equitably distributed. The exclusion of groups or communities that might benefit from study participation must be justified.

Guideline 13: Research involving vulnerable persons

Special justification is required for inviting vulnerable individuals to serve as research subjects and, if they are selected, the means of protecting their rights and welfare must be strictly applied.

Guideline 14: Research involving children

Before undertaking research involving children, the investigator must ensure that:

- the research might not equally well be carried out with adults;

- the purpose of the research is to obtain knowledge relevant to the health needs of children;
- a parent or legal representative of each child has given permission;
- the agreement (assent) of each child has been obtained to the extent of the child`s capabilities; and,
- a child`s refusal to participate or continue in the research will be respected.

Guideline 15: Research involving individuals who by reason of mental or behavioural disorders are not capable of giving adequately informed consent

Before undertaking research involving individuals who by reason of mental or behavioural disorders are not capable of giving adequately informed consent, the investigator must ensure that:

- such persons will not be subjects of research that might equally well be carried out on persons whose capacity to give adequately informed consent is not impaired;
- the purpose of the research is to obtain knowledge relevant to the particular health needs of persons with mental or behavioural disorders;
- the consent of each subject has been obtained to the extent of that person's capabilities, and a prospective subject's refusal to participate in research is always respected, unless, in exceptional circumstances, there is no reasonable medical alternative and local law permits overriding the objection; and,
- in cases where prospective subjects lack capacity to consent, permission is obtained from a responsible family member or a legally authorized representative in accordance with applicable law.

Guideline 16: Women as research subjects

Guideline 17: Pregnant women as research participants.

Guideline 18: Safeguarding confidentiality

The investigator must establish secure safeguards of the confidentiality of subjects' research data. Subjects should be told the limits, legal or other, to the investigators' ability to safeguard confidentiality and the possible consequences of breaches of confidentiality.

Guideline 19: Right of injured subjects to treatment and compensation

Investigators should ensure that research subjects who suffer injury as a result of their participation are entitled to free medical treatment for such injury and to such financial or other assistance as would compensate them equitably for any resultant impairment, disability or handicap. In the case of death as a result of their participation, their dependants are entitled to compensation. Subjects must not be asked to waive the right to compensation.

Guideline 20: Strengthening capacity for ethical and scientific review and biomedical research

Many countries lack the capacity to assess or ensure the scientific quality or ethical acceptability of biomedical research proposed or carried out in their jurisdictions. In externally sponsored collaborative research, sponsors and investigators have an ethical obligation to ensure that biomedical research projects for which they are responsible in such countries contribute effectively to national or local capacity to design and conduct biomedical research, and to provide scientific and ethical review and monitoring of such research.

Capacity-building may include, but is not limited to, the following activities:

- establishing and strengthening independent and competent ethical review processes/ committees
- strengthening research capacity
- developing technologies appropriate to health-care and biomedical research
- training of research and health-care staff
- educating the community from which research subjects will be drawn

Guideline 21: Ethical obligation of external sponsors to provide health-care services

The complete report of International Ethical Guidelines for Biomedical Research Involving Human Subjects from CIOMS is available online http://www.cioms.ch/frame_guidelines_nov_2002.htm [33].

3.7 The Canadian Tri-Council Policy Statement

In Canada, the Parliament has funded three research councils, the Canadian Institute of Health Research, the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council, to promote and regulate all federally funded research done with human participants. In 1998, this group has published the tri-Council Policy Statement: Ethical conduct for research involving humans (with 2000, 2002 and 2005 amendments). This document has binding authority with federally funded researchers in Canada. The ethical principles and federal regulations that govern research with humans in Canada are characterized by two conflicting values: the ongoing need for research and the moral imperative of respecting human dignity. In their report they give recommendations concerning research ethics boards, prospective and retrospective research, clinical trials, free and informed consent, conflict of interests, human genetic research. The minimum information that researchers should provide to the participants should include the purpose of the study, the type and the amount of the tissue, the manner in which tissue will be taken and preserved, the potential uses, the protection measures of privacy and confidentiality, the identifying information that will be attached to the sample and possible ways in which research could cause harm. The complete report of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans is accessed through http://www.ncehr-cnerh.org/english/code_2/ [34].

3.8 The European Society of Human Genetics (ESHG)

The ESHG was established on March 15, 1967 by a small group of outstanding European Geneticists [European Journal of Human Genetics: 3, 63-64, 1995]. In 1996 the ESHG together with the American Society of Human Genetics and the Human Genetics Society of Australasia founded the International Federation of Human Genetic societies, which represents a growing number of national societies. In 2003 ESHG has published a policy about data storage and DNA banking for biomedical research.

The statements and recommendations given included the following:

1) Reasons for DNA banking, potential benefits and misuses;

2) Status of collection, anonymized or identifiable collections;

3) Consent requirements, review of ethical committee, protection of vulnerable subjects, free consent, right to withdrawal, consent requirements for existing collections, population studies;

4) Management, quality control and security issues;

5) Access, sharing and ownership of the data [35, 36].

3.9 The World Medical Association (WMA)

The World Medical Association (WMA), an international organization of physicians, was formally established in 1947. In 2007, the WMA had a membership of 84 national medical associations and represents some 9 million physicians. The World Medical Association has developed the Declaration of Helsinki as a statement of ethical principles to provide guidance to physicians and other participants in medical research involving human subjects. Medical research involving human subjects includes research on identifiable human material or identifiable data. In 2002, the General Assembly has published a policy on ethical considerations regarding health databases. The principles of the policy involve access to information by patients, confidentiality, informed consent, data integrity, documentation and management. The full WMA declaration on ethical considerations on health databases is accessed through http://www.wma.net/e/policy/d1.htm [37].

3.10 The Human Genome Organization (HUGO)

HUGO was established in 1988 as an international organization of genome scientists. Its funding derives from private foundations in the United States, Britain, France, and Japan and its central offices for genome research coordination are located in Bethesda, London and Osaka.

In 1998, the HUGO Ethics Committee produced a document specifically regarding genetics research with banked DNA samples. Entitled a “Statement on DNA Sampling: Control and Access”, this second document continues the committee’s emphasis on the international ethical and legal standards that are needed for human tissue research in the era of genomic medicine. They made several recommendations:

1) Policies and practices regarding genetics research with banked DNA samples should distinguish between tissue samples obtained during “routine medical care” and during “specific research protocol”.

2) Research with “routine samples” should be done only after individual patients have been given “general notification” of an institution’s policies regarding future use of diagnostic samples, the patients have been given an opportunity to object to the research option and the samples have either been coded or anonymized.

3) The practice of secondary research with “research samples” should be done only after research participants have been given notification for this possibility, individual participants have not objected, and the samples have been coded or anonymized.

4) Policies regarding third-party access to stored DNA samples and the personal and familial genetic information contained in the samples should limit such access to “immediate relatives”.

5) Individual’s requests for the destruction of their DNA samples in genetics labs should be carried our only if immediate relatives have no need for learning their own genetic status and if the samples have not already been entered into a research protocol or provided to other scientific investigators.

6) These ethical requirements regarding “the control and access of DNA samples and information” should be standardized world-wide [20, 38].

3.11 The World Health Organisation (WHO)

The European Partenrship on Patients' Rights and Citizens' Empowerment, a network of the WHO has published a report on genetic databases [39]. The report examines the ethical, social and legal issues that surround the creation and operation of genetic databases.

Key recommendations:

There follows a brief account of the central recommendations made in the report.

1) Individuals are entitled to control the use of their samples and information, in a manner akin to a **property right**.

2) Collection of genetic data should normally be allowed only for the purpose of promoting **public health**.

3) The onus will be on those who seek to create the database to justify its nature, purposes, content and uses. The creators should ensure adequate security measures for the data and for **privacy protection**, and **respect human rights**.

4) An appropriate **ethical approval mechanism** should be established to oversee the creation and maintenance of genetic databases.

5) Public debate should precede the establishment of new genetic databases. No database should be established if public trust is seriously in doubt.

6) When obtaining **consent** to the provision of a sample or information for a genetic database, participants should be informed to the following extent:

i) Participants should be given sufficient information to make a meaningful choice about participation, including information about the purposes of the database and its commercial potential;

ii) Sufficient information should be provided to ensure that participants comprehend the nature of the enterprise to their own satisfaction;

iii) Participants should be given the opportunity to ask questions and have these answered;

iv) Participants should be informed of the risks of participating, where these exist;

v) Participants should be informed of the security provisions that exist to protect their personal data;

vi) Participants should be informed of the alternatives to participating, and in particular, should receive assurances that no adverse consequences will follow if they choose not to participate;

vii) Participants should be informed of the uses to which data might be put;

viii) Participants should be informed of the possibility of future uses of data, beyond the limits of the present consent, and should be provided with an opportunity to withhold consent to such uses.

7) **Blanket future consent** is permissible where anonymity can be guaranteed, and there is no risk that unexpected results will filter back to the subjects concerned. If this guarantee is not possible, or if linking of data is necessary for the research, then specific consent to the research must be obtained.

8) For **vulnerable adults** all current international guidelines should be complied with. Children are also considered as a separate category. In both cases, contributions to a database are

possible within stringent safeguards.

9) Research using **archival material** – for which no specific consent has been obtained – is permissible if the material and information derived from it is anonymized, and there is no prospect that research results will be used to identify the sample sources at any future time.

10) The **anonymization** process should be overseen by an independent body with attendant duties explained in the report.

11) **Feedback** to participants should not normally be provided from genetic research data. This principle should be departed from only with ethical approval and if:

i) the data have been instrumental in identifying a clear clinical benefit to identifiable individuals;

ii) the disclosure of the data to the relevant individuals will avert or minimise significant harm to those individuals;

iii) there is no indication that the individuals in question would prefer not to know.

12) Adequate account must be taken of the privacy interest that individuals have in not knowing information about themselves.

13) Those who would seek to depart from the practice of requiring active informed consent prior to participation in the creation of a genetic database must justify this position in strong ethical terms.

14) The gathering and storage of genetic samples and information must be subject to rigorous privacy protection measures in conformity with international and national data protection laws. These privacy measures must be transparent and subject to ethical approval by a suitable body.

15) It should be the role of an independent body to oversee and regulate access to genetic databases. This same body should hold the key(s) to any anonymization methods. The body should receive applications for access for consideration in light of the nature and purposes of the database. The body must be satisfied that the party seeking access is able to make responsible use of the data and to respect their status. The use of finite resources such as genetic samples must equally be regulated by ethically appropriate means.

16) Every participant has the absolute right to withdraw at any time and without reason. All information and samples should be destroyed on request unless it can be shown that it is not reasonably practicable to do so.

17) Serious thought should be given to the role of property rights for individuals in their own samples and benefit sharing to participant communities as a means to foster and maintain public trust and confidence in genetic databases.

18) The establishment, maintenance and operation of genetic databases should be carried out in an atmosphere of openness, transparency and appropriate ethical scrutiny. Accountability of database creators, managers and users should be a given. Consideration should be made of the ways in which this can be achieved, including legal measures to regulate and control the creation and management of databases, and duties to report publicly on activities in respect of the database [39-41].

3.12 United Nations Educational, Scientific and Cultural Organisation (UNESCO)-International declaration on human genetic data

The aims of this Declaration are:

(a) to provide a universal framework of principles and procedures to guide States in the formulation of their legislation, policies or other instruments in the field of bioethics;
(b) to guide the actions of individuals, groups, communities, institutions and corporations, public and private;
(c) to promote respect for human dignity and protect human rights, by ensuring respect for the life of human beings, and fundamental freedoms, consistent with international human rights law;
(d) to recognize the importance of freedom of scientific research and the benefits derived from scientific and technological developments, while stressing the need for such research and developments to occur within the framework of ethical principles set out in this Declaration and to respect human dignity, human rights and fundamental freedoms; (e) to foster multidisciplinary and pluralistic dialogue about bioethical issues between all stakeholders and within society as a whole;
(f) to promote equitable access to medical, scientific and technological developments as well as the greatest possible flow and the rapid sharing of knowledge concerning those developments and the sharing of benefits, with particular attention to the needs of developing countries;
(g) to safeguard and promote the interests of the present and future generations;
(h) to safeguard and promote the interests of the present and future generations [42].

3.13 The German, Belgian, UK & Greek Bioethical legislation and recommendations

3.13.1 Germany

There is no federal German law concerning human tissues or bio-banks. The question of tissues and cadaver samples are regulated by common law. Nationwide rules can only be enforced by criminal law. According to the patrimony law (“Erbrecht”), the human body is not considered an object during life, so that it is not part of the heritage. As a consequence, a dead body has no owner in Germany. The legal condition of parts and fluids separated from the body has to be examined carefully: during life, the integrity of one’s own body is a fundamental right of the person. Once parts are separated from the body, they are regulated by both property and personality right law (overlap thesis). Interestingly the property rights and the right to privacy can belong to two different persons, and they differ in their scope of protection: property of samples can be completely transferred, but the right to privacy cannot! In practice, if a patient gave no particular directive for the use of his tissues/body fluids, the physician (*e.g.* the pathologist) is free to trash them or to use them for quality control purposes in the laboratory or for research purposes if anonymization is warranted, if no immortalization is undertaken, and if the research project is not ethically debated. However, if the samples are not only used for research, but also for therapeutic purposes, the need for a specific informed consent given by the patient is increasing, in particular with large samples, or in the presence of personal characteristics (*e.g.* genetic material). For the researcher, the duty remains to protect the right to privacy of the donor/or the patient [43].

- If the banking of human genetic material includes the collection and storage of personal data associated with an identifiable person, the Data Protection Law has to be taken into account. The German Data Protection Law is very complex, since the applicable law depends on the status of the data collecting and storing institution (public, private, federal, state), it contains different permissions for collecting, storing, using, transferring for own or for other purposes and there are numerous exceptions in other laws. The Data Protection Laws are subsidiary towards these special provisions in certain areas.

- *Federal Data Protection Law (BDSG) 1990.* Non public DNA banks, which process personal data, have to take into account the Federal Data Protection Law if data are processed into or out of file commercially or for professional or commercial purposes. The processing of personal data and their use is only allowed if there is a legal permission or consent of the affected person. The transmission of personal data is permitted if it is necessary for scientific research. This permission does not apply for medical data.
- *Data Protection Law of the states.* The application field of the data protection laws of the states include all public authorities of the respective states. They have to be applied in public hospitals or the state or the municipalities. The University hospitals are either non-self-maintained public state institution or state companies with restricted independence. If these institutions bank identified or identifiable human genetic material, they have to take the state data protection acts into account. These laws are subsidiary towards special data regulation in law, specific for certain areas, for example the hospital laws.
- *Law of 20 June 1990 to regulate matters relating to gene technology.* The aims of this law are: (1) to protect life and health of human beings, animals and plants against possible threats of gene technology, and (2) to give a legal framework for research, development and support of scientific, technical and economic possibilities of gene technology.
- *The German Bundestag, Chancen und Risken der Gentechnologie Enquete-Commission, 1987.*
- *Resolution of the Conference on the Privacy of Information of Rheinland–Pfalz on the Subject of Gene Analysis and Informational Self-Determination (26–27 October 1989).*
- *The Federal and State Governments, Final Report of the Bund–La¨nder–Arbeitsgruppe Genomanalyse, 1990.*
- *Minister of Research and Technology, Arbeitskreis Genforschung Report, Frankfurt, 1991.*
- *German Parliament, Periodic Report of the Bu¨ro fu¨r Technikfolgen, 1992, 2000 [44].*

In 2004, the Bio-bank Working Group of the German Telematics Platform for Medical Research Networks (TMF) initiated a project to construct a generalized legal basis for the establishment and operation of bio-banks in Germany. Most bio-banks in Germany are currently operated by public institutions such as university clinics, institutes or departments. However, a substantial proportion of these bio-banks are considered “going private”. This project involved the planning, writing and evaluation of an expert report that addresses in great detail the legal issues concerning property rights, medical professional regulations, general liability insurance, resource continuity and research secrecy [45].

3.13.2 Belgium

In September 2002 Belgium adopted a general law on patient rights. This law regulates topics as informed consent, the medical file, representation of incompetent patients, etc. Consent has in principle to be explicit but no written consent is needed except when the patient wishes so. The law does not relate to 'further use' of tissue.

The law on organ donation after death dates from 1986 and is based on an opt-out or no objection system. Donation from living donors (kidney, bone marrow) is regulated as well, but 'further use' of tissue cannot be headed under it. There is a law on the processing of tissue for therapeutic purposes, which again does not relate to 'further use'. No advisory body has issued a statement on 'further use' yet.

In 1992 the law on Privacy Protection in relation to the Processing of Personal Data was accepted. This law modified in 1998 following the European Directive. In 2001 secondary legislation in the form of a Royal decree was adopted which especially relates to the use of data for research. The Belgium law considers two types of data: personal data (directly, indirectly and

coded data) and anonymous data. This means that coded data remain personal data according to the Belgian law, because the possibility remains to reverse a code, even if not by the processor of the data. The law contains the usual exception to the principle of explicit consent (which even has to be written) to the use of sensitive data when this is necessary for – in short - the treatment of the data subject and the processing takes place under the supervision of a health care practitioner.

Data for research coded data play an important role in the regulation of the processing of data for research. Before they can be transmitted to the researcher these data should in principle be coded. The data subject must be informed about this procedure and can object to it. The Royal Degree contains an exemption if the data subject cannot be reached or this would be unreasonably difficult. In that case an application to use these data can be made to the Data Protection Commission which can give a permit, if necessary subject to terms and conditions. The Royal Degree also contains provisions on the use of non coded data. The Belgian legislation does not contain specific provisions on the use of genetic data for research [46].

3.13.3 UK

UK is an interesting case as much debate has taken place on 'further use'. The Nuffield Council was probably the first advisory body to issue a report on 'further use' of tissue for research. In a much later stage the debate was heavily influenced by a scandal around the retention of organs and tissue of deceased infants without the consent of the parents. This reaction has lead to an increased awareness of the public and authorities on the issue of research on left over tissue as well. The UK Medical Research Council made some sensible recommendations regarding feedback to individual patients. It proposed a kind of 'layered consent' where a choice is given to the patient for consent to various forms of research. The Department of Health published in April 2003 an interim statement regarding the taking, storage and use of human tissue from the living and the dead pending new legislation.

The statement proposes written consent for use of left-over tissue for research. It is unclear how specific this consent needs to be probably somewhere between blanket consent and consent specific for one research protocol. About the use of existing stored tissue for clinical research the following was put forward in the interim statement. A decision will need to be taken as to whether consent or further consent need to be sought. Here, three situations can occur:

- Valid consent may previously have been given to a particular use or uses, in which case it is usually lawful to use the organs or tissue as already authorized. But consideration is needed whether the form of consent is sufficient on its own to be regarded as valid today.
- The donor is identifiable and unambiguous consent has not already been obtained. In that case consent should be sought from the person concerned or if the person is not alive or cannot be traced, from someone close to that person.
- The identity of the donor is unknown or the donor cannot be traced. In this case the tissue may in principle be used for research after careful consideration of the possible effects of the research.

In general Data protection is most of all regulated by the Data Protection Act of 1998 (hereinafter the DPA) and following statutory instruments. However the Health and Social Care Act 200160 also contains provisions on the processing of medical data and professional secrecy. Data which are coded but are anonymous to the controller are not considered to be personal data in the sense of the DPA.

In December 2003 a Bill was proposed to parliament called the Human Tissue Bill encompassing all aspects of the handling of human tissue. With regard to 'further use' of tissue for research and tissue banking the key elements are:

- Consent is needed for the use of tissue, whether from living patients or from the deceased, in research. The Bill does not, however, specify what form the consent should take or how and when it should be taken or recorded. The new Human Tissue Authority will give guidance on this, taking account of the practical aspects of seeking consent in the course of medical, diagnostic and post mortem procedures.
- The Bill does not specify that consent, when given for research, should be specific to any particular research project. Indeed, it is envisaged that a general consent to research use would be the norm. This would then be subject to the guidance from the HTA and, as at present, to the general standards required by Local Research Ethics Committees in approving research proposals.
- The Bill provides that the use of tissue from living patients for the particular purposes of clinical audit, quality control, 'on the job' education and training, and public health monitoring, will not need consent over and above the consent required for the taking of the tissue from the patient [46]

3.13.4 Greece

Until now Greece did not have a bio-bank. Recently, the Biomedical Research Foundation of the Academy of Athens (IBEAA) has been funded and approved from the European Union to establish a Bio-bank that will join the European Bio-banking and Biomolecular Resources Infrastructure Network. The existing cord blood bank that already is active in IBEAA will be incorporated in the new bio-bank.

The National Bioethics Commission has provided recommendations on the collection and use of genetic data (2002). The general principle is the respect of the value of human beings that requires the free and informed consent of the person whose biological sample is collected for the purpose of genetic testing. i) A bio-bank is legally set up by a license issued by the responsible authority, which may be a university.
ii) In case the bio-bank will collect and process named personal data it must also obtain a license from the Authority for the Protection of Personal Data (art. 7(2) and 8(3), Act 2472/1997 respectively).

The collection and processing of information that may be of interest to a bio-bank is governed by the special provisions on sensitive data in the sense of Act 2472/1997 (art. 7) in all cases where the identity of the person the data pertains to is or may become known. In principle, collection and processing are allowed by authorization of the Data Protection Authority.

The issue of the origin of the information collected by a bio-bank must be treated separately. If this information is obtained from (genetic or other) tests conducted in the bio-bank itself (or by a third party on its behalf), the donors must be informed accordingly and give their consent for harvesting of related biological specimens pursuant to art. 5 of the Oviedo Convention, in addition to the consent required for the specific processing (collection of information) pursuant to art. 7 of Act 2472/1997.

When this information originated from already existing collections (hospitals, social security funds, private practices, diagnostic labs, etc.) its transmission and processing for different purposes in a bio-bank requires specific informed re-consent. By way of exception, re-consent is not required only when the identity of the donor is not disclosed to the bio-bank or is safely

encrypted by the latter. Prior authorization of the Authority is always required. Physicians may disclose information protected by the duty of confidentiality to bio-banks if the donor specifically agrees to that (art. 13 CME).

The processing of information stored in a bio-bank is governed by the general rule of art. 10 Act 2472/1997 which lays down strict conditions to ensure the confidentiality and security of any such processing. The key person in this respect is the one responsible for processing who must fulfil specific duties-also vis-à-vis the Data Protection Authority. The law does not expressly require that all relevant research projects go through ethics control but the Authority may impose such requirements.

Freedom of research for bio-banks is basically determined by the above framework. Bio-banks are also obliged not to hinder access to the results of research to third parties outside the framework of protection of sensitive personal data. Thus, third parties may be allowed access only if informed consent was obtained from the donors. Naturally, bio-banks are free to communicate any research findings not disclosing the identity of donors to third parties.

In the context of economic freedom, a private bio-bank would normally be entitled to find commercial uses for the results of its research by applying for patents (if they are “inventions” in the sense of p.d. 321/2001) or by negotiating deals (e.g. with pharmaceutical companies, etc). It must be noted that, in principle, the State does not hold any rights on the personal data stored in collections of public agencies which may, therefore, be used by bio-banks, neither on the results of the bio-bank’s research. Any entitlements to the data are owned only by the donors who are free to use them even for profit. Indeed, greek legislation does not regulate the commercialisation of such data in so far as it does not encroach on the rights of third parties or violate the Constitution (particularly, the value of human beings) or morality (art. 5(1) of the Constitution, arts 178, 179 CC). As long as this framework remains in place, donors may exchange consent for specific economic advantages which are ultimately borne by the bio-bank [47].

4 Management of Ethical and Legal Issues

Excitement over the value that bio-banks can bring must be tempered with concerns about security, privacy, and the appropriate use of the information they contain. Not only are genes heritable — and thus have predictive value that could be used inappropriately — but the genes of individuals contain information about their families as well. So when organizations collect this data prospectively and individuals do not know the exact purpose of the use of their information, concerns about the potential use, or mis-use, of this data is high. Such concerns become amplified when people worry about the potential motivation of private organizations for commercial gain. As a result, not only are private companies putting in place ethics policies, but international organizations, such as UNESCO, WHO, and the OECD, as well as governments have developed, or are developing, policies regarding the regulation of bio-bank research and management. Given the worldwide nature of scientific research, harmonizing such policies and establishing standards at the international level are critical [2].

Governance

Because of the complexity of the healthcare, quality, security, and privacy issues surrounding bio-banks, governance is a key issue. Recognizing this fact, many governments have established regulations concerning the operations of bio-banks at both local and national levels. For example, in January 2003, The Swedish Ministry of Health and Social Affairs enacted the Bio-banks in Medical Care Act, which "regulates how human biological material is to be collected, stored, and used for certain purposes with respect for the personal integrity of the individual" and to which all bio-banks in Sweden must adhere. Of course, each of these banks in turn has its own governance policies.

Also at a national level, the Estonian Genome Project was established under the guidance of the Human Genes Research Act, which in turn was guided by international documents dealing with genetic research from UNESCO and the Council of Europe. The act focuses on ethics and security. It established a unique public/private governance structure with the Estonian Genome Project Foundation, which owns the medical and genetic database and acts as a privacy guarantor, and EGeen, the exclusive commercial licensee of the database. Key governing groups include an Ethics Committee, a Management Board, and a Supervisory Board. The latter board consists of members of the government, Parliament, and the Estonian Academy of Sciences.

Within the United States, recognition of the need for consistency of governance policies and an interest in establishing a National Bio-specimen Network led the National Dialogue on Cancer (NDC) to commission a study last year by the RAND Corporation on best practices in United States-based bio-banks.

Control of funding sources can lead to policy standards. For example, in the United States, all federally funded institutions are required to have institutional review boards (IRBs) to establish and oversee the operation of repositories and their data management centers (Office of Human Research Protection, 1997). This includes oversight of all aspects from policies on ethics and patient privacy to review of research proposals, grants, clinical trials, and publications. All bio-banks typically follow a similar procedure. Research bio-banks often have management groups that report into a steering committee that has overall responsibility for the bio-banks. That committee, or a subset, also has responsibility for allowing access to the bio-banks and approving research

proposals. For example, bio-banks with tissue samples may establish a tissue utilization committee to prioritize sample distribution based on review of research proposals according to standardized criteria. In addition to IRB oversight, review, and approval, all bio-banks have established specific guidelines, policies, and procedures to protect patient privacy and confidentiality. These policies are extended to the informatics area with specific rules on data standardization, sample anonymization, infrastructure and database security, password protection, encryption, firewalls, and so forth [2].

Furthermore, bio-banks governance is also increasingly characterized by its deterritorialized, transnational/global character, which neither excludes the national character of projects, nor a national rhetoric.

The emerging topography of bio-bank governance displays a complicated picture. Ultimately, the success and failure of bio-banks projects will depend on their integration into the political and socio-cultural fabric. Thought not always viewed in that way, answers to the problems of science tend to be answers to sociopolitical problems, needs and demands. Therefore, bio-bank governance should not be misunderstood as a mainly technical challenge. Bio-bank governance is always a response to socio-cultural challenges and requires the building of trust, acceptance, and careful political negotiation. Bio-bank development refers to the emergence of a multilayered biomedical apparatus that introduces new medical knowledge, rationalities and principles of action, a set of novel strategies of practicing medicine, medical research, interacting with patients, and locating patients within the system of health management. In such networks of governance the state is one actor next to others, and the ordering and structuring of the interaction between bio-banks, society, and politics operates through a variety of actors, on different levels and along particular rationalities. Such networks of governance reflect, to some extent, a post-regulatory state in which governance has become a complicated architecture and field of action involving a multitude of forces and rationalities. Bio-bank governance is still a relatively new field of political-legal-intervention and it will be crucial for the future of bio-banks to establish governance regimes that appropriately link research with society and politics [48].

Cross-Border Issues

A large variety of practices and procedures is associated with the regulation and management of bio-banks; however, these institutions also need to share information and learning.

The March Bio-bank Summit led by IBM was one attempt to highlight these issues. Other groups have also started to address some of them. In January 2003, the "Bio-banks for Health" workshop was held in Oslo as "a strategic initiative to bridge and standardize bio-banks, health registries, and population surveys at a European level in order to improve the research potential and find better ways to tackle diseases." The need to exchange material and data brings a need to standardize processes and legal, ethical, and regulatory guidelines. Currently, practices vary widely and coordination is highly variable. Those who promote a pan-European, if not global, approach note, "because of the enormous variability in environment, lifestyle, and genetic dispositions at different locations and in different ethnic groups, it is necessary to have a globally oriented approach." (EU workshop "Optimizing the Use of European Bio-banks and Health Registries for Research Relevant to Public Health and Combating Disease," Oslo, January 2003). Some of these differences arise out of simple history — for example, some of the Nordic repositories have been in existence for a long time and thus cannot take advantage of some of the new storage and analysis technologies

available to the more recent banks. Levels and types of privacy and security concerns can vary by country as well. This particular workshop identified a range of key problems that arise in connecting bio-banks across borders:

- Bio-banks are a mixture of disease-specific banks and general population bio-banks, and most often are not linked.
- Some registries and bio-banks are effectively inaccessible because they are institution specific.
- Storage systems for material vary widely and are of inconsistent quality.
- Consents to use the biological material vary and thus raise ethical questions about their potential use.
- Health registries have been compiled for different purposes at different times.
- The extent to which registries in one or more countries can be cross-correlated varies from region to region and from country to country.
- The ways in which researchers are allowed to access and use the registries and bio-banks vary with different access methods, screening and evaluation systems, and policies.

Even within national borders, policies and technologies can vary widely across institutions. As noted above, in the arena of cancer research, this variation has prompted the United States to start to examine best practices across organizations. As the RAND study noted, "Tissue collection, processing, and storage techniques vary depending on the purpose of the repository, as do the quality and extent of information collected with the bio-specimens" [2].

4.1 Informed Consent in the Era of Genomic Medicinal Research

Informed consent is a fundamental principle that has marked the emergence of modern medical ethics based on personal autonomy. The need for informed consent in biomedical research was emphasized by the Nuremberg trials that revealed in human experimentation on prisoners in concentration camps. In the clinical context, the importance of informed consent has been recognized as a consequence of the rising patients' rights movement and emerging biomedical technologies that emphasized the necessity to decide about the complex health-care choices to be made by the patient him/herself [49].

Obtaining bio-specimens from individuals who are fully informed about and have consented to the collection of their tissue by the repository and its use for research purposes is a best practice. Using a tiered consent process that allows individuals to choose the type of specimen(s), if any, they want to donate (e.g., tissue, blood, or urine), the type of research the specimen can be used for (e.g., a specific research project, general research, or genetic research), and whether their medical records and outcomes data can be accessed is also a best practice. Ideally, the consent process should occur separately from the surgical consent. However, since this is not always possible, at a minimum the informed consent for the collection and research use of specimens should be a separate section of the surgical consent form that requires a separate signature.

Free and informed consent is a guiding principle of the *Ethical Conduct for Research Involving Humans*. Described as being "at the heart of ethical research involving human

subjects" (Tri-Council Policy Statement 2003Article 2.1), informed consent—along with Research Ethics Boards—is intended to act as a barrier that protects research subjects from abuse. Under the notion of informed consent that has developed in Canada even competent and un-influenced research participants cannot authorize a particular use of data or material if it is not described to them. The value of free and informed consent has also been recognized by the Public and Professional Policy Committee of the European Society of Human Genetics (PPPC), which insists on the need for informed consent as well as oversight by an ethical review committee in the case of bio-banks. Research participants must voluntarily authorize participation based on an informed understanding of the risks and benefits of the research, including options to participating [1].

Two models of informed consent exist: the **opt-in** and the **opt-out** model. Under opt-in model, potential donors give specific consent to participate in the research described in the informed consent document. Opt-out models presume consent, unless a person specifically elects not to participate. Although opt-out models facilitate the creation of a national tissue repository, depending the laws and regulations of each country, this model would only be permitted under specific constraints, e.g. informed consent waived by an institutional review board [4].

Recommended elements to be disclosed to potential study subjects include not only the risks and benefits of the research, as well as the limitations and outcomes. They also should cover how results will be communicated and how confidentiality will be protected. According to the PPPC, consent must be freely given, provided in writing, and based on information regarding the use of the samples, including access to and sharing of the samples, and the timeline for storage of the samples.

Informed consent is often taken as general consent for research applications well beyond those envisaged when the research subject signed the form. Insofar as informed consent can be reduced to a simple signature on a consent form, monitoring consent becomes rather simple. The general reliance on consent forms and the lack of follow-up to ensure compliance can lead to the conclusion that the ethics review process is primarily intended to protect research institutions, researchers and sponsors and not the research subject. This apparent emphasis on accountability for, rather than to subjects is particularly concerning in the case of bio-banks, due to the sensitivity of the information they hold and the potentially lengthy time that they might hold it.

This time component is, in fact, a crucial part of any discussion concerning informed consent and bio-banks. Since the objective of bio-banks is to establish a database for future research questions that are either unanticipated or questions that cannot be framed in any detail without first conducting research based on the data held in the bio-bank, consent to participate in research conducted with data from a bio-bank cannot be simply "informed." This is further complicated by the fact that the social risks of particular research results and any potential duty to disclose to family and blood relatives cannot be described.

Typically, there are two responses to this problem: "re-consenting" and "blanket consent."

• **Re-consenting** explicitly rejects the presumption that it is possible to authorize future research, stating that informed consent is only valid if participants understand the specific nature, risks and benefits of particular research. This necessitates each research project using bio-banks to seek and receive newly informed consent specific to each project.

A compromise is sometimes struck where consent to be included in the bio-bank is sufficient for the use of data or materials in any research where identifying characteristics can be separated from the data through aggregation or other processes. In this case only identifiable or additional participant contacts require additional consent.

• **Blanket consent** suggests that the challenge presented by future research to informed consent can be described to research participants who can then decide whether or not to authorize that research. Variations include options to authorize only certain kinds of research—such as research on a familial disease—or only under certain conditions—such as when data is anonymous. This permits people who require more specific information before giving consent to decline participation, while enabling others to authorize future research. Adopting blanket consent addresses the concern of some research groups that re-consent could "bring to a halt all research on existing, archived material".

One persistent ethical issue with this approach is the presumption that future research will not be objectionable to those who provide blanket consent. However, possibly more pressing a challenge is the fact that there is good reason to think that blanket consent is not consistent with accepted ethical guidelines in some countries.

A possible third response is to inform people that their tissues will be used in research subject to certain constraints. This is "**informing without consent**." Even using anonymized or exclusively aggregated data and materials can create problems. Not only is there debate over whether research using aggregate data that does not identify individuals can be conducted without informed consent specific to the research, but participants in bio-banks often want to be contacted if research reveals even potentially relevant information about them. And although many genetic tests lack definitive clinical relevance, which may be a good reason for refusing to fund access, when data exists as a result of research there is probably a fair presumption in favour of disclosure with careful explanation.

A fourth possible form of consent, described, is "**dynamic informed consent**." Appearing to be a form of re-consent, dynamic informed consent allows study participants to be included or excluded in research as they feel appropriate in response to updated information on the goals, risks, and benefits of ongoing studies, follow-up research, or wholly new research projects. First Genetic Trust, an American company that provides a web-based infrastructure for dynamic informed consent, explains on its website that blanket consent:

…for unspecified future use of biological samples and data generated from clinical trials, is no longer adequate for genetic research… Research subjects cannot provide truly informed consent for unspecified future research that he/she does not and will not know about.

However, according to the PPPC:

As it is difficult to foresee all the potential research applications that a collection may be used for, individuals may be asked to consent for a broader use. In that case, there is no need to recontact individuals although the subjects should be able to communicate should they wish to withdraw. [1]

The committee also allows for the use of pre-existing collections of anonymous and anonymized samples [44], but recommends re-contact of subjects when feasible, or review by an ethics oversight committee.

It should be noted that informed consent generally does not require that the reason someone refuses to participate in certain research be disclosed, or rational according to some standard. Since there is no obligation for someone to participate in any particular research project, refusing to participate is usually accepted unconditionally and efforts to clarify reasons are sometimes considered a form of harassment. Foregoing the requirement to seek specific informed consent to genetic research precludes individuals' ability to refuse to have their data or materials used in genetic research.

Finally, it should be pointed out that much of the literature on informed consent and bio-banking focuses on protecting privacy through safeguarding access to identifying information. For some people, certain tissues have symbolic meaning and are regarded as sacred, some even requiring ritual and special storage. Indeed one of the objections to retaining the Nuu-chah-nulth samples has been the violation of cultural norms for the disposal of tissue [1].

International Bioethics Committee of UNESCO has also published a Report on Consent which is accessible through:
http://unesdoc.unesco.org/images/0015/001505/150520e.pdf.

4.2 The Relevance of Ethical Principles to the Bio-banking

Legal Status

The mandate of a bio-bank — biomedical research, drug or diagnostic development, preventive or curative medicine — and its scale (e.g., national or disease specific) will be primary factors in its legal status. Currently, bio-banks are found in multiple forms as public foundations, government institutions, or commercial operations. The March Bio-bank Summit posed the question of whether there is an ideal legal status. The notion of establishing bio-banks as a "Public Utility Idealistic Foundation" is starting to be discussed, particularly in the European Union, where some common structures and monetary systems exist across countries. The concept of a hybrid public/private model presents itself because:

- Tying research to specific diseases and therapeutics depends on the work going on within pharmaceutical and biotechnology companies as well as disease foundations and government-sponsored research.
- The need for genealogical data and clinical data usually extends beyond the boundaries of any one organization into public health data.
- Bio-banks cannot afford to be entirely at the whims of government budgets or of commercial profit motives if they are going to achieve some of their broader goals.

It will probably take some time to develop the framework and operating principles for the form and structure for such an organization in a way that will work smoothly for all parties concerned. It will also be important to take into account cultural, political, intellectual, and socioeconomic differences as more and more of this work occurs across countries.

Each organization needs to address legal issues, in addition to the legal form and status of the bio-bank organization. Chief among these — once policies, procedures, and general risk management practices are addressed — is intellectual property policy. Many bio-banks act solely as "banks" and do not retain downstream rights to any

intellectual property developed using their samples and data. On the other hand, some bio-banks exist to have such rights and, in some cases, are developing physical products — diagnostics and therapeutics — based on those rights. There will not be one correct approach, but each organization, consortium, or network will need to publish a specific policy on intellectual property rights in order to avoid confusion and conflicts [2].

Ethics Boards

Many countries do not have legislation or regulation in place regarding ethics policies and practices, but some have taken an active stance. The 2002 Swedish Act on bio-banks regulates the storage and use of samples and also defines quality and security requirements for all bio-banks within the country. In 2000, the government of Singapore established the Bioethics Advisory Committee (BAC) with a broader agenda to "address the potential ethical, legal, and social issues arising from biomedical sciences research in Singapore." BAC includes international expert advisors and offers many resources to the public, such as "classrooms" with information on various topics and the *BIOEthics* newsletter. Iceland has likewise formed a National Bioethics Committee to review research concerning genetics and the use of biological samples. As noted above, the Estonian Genome Project has an Ethics Committee that aims "to assist in ensuring the protection of the health, human dignity, identity, security of person, privacy, and other fundamental rights and freedoms of gene donors and resolution of general ethical problems related to human gene research."

Some cross-border organizations have stepped forward as well. For example, in late 2002, the European Society for Human Genetics (ESHG) published a paper on European storage practices for both DNA and data regarding issues of consent, control, quality, ownership, and benefits. Earlier this year, the Research Directorate General of the European Commission published a report on the ethical, legal, and social aspects of genetic testing, with a section addressing some of the specific ethical issues confronting bio-banks in particular, such as data protection and donor consent.

At the individual organizational level, many institutions have established bioethics review boards or committees to supplement IRB and/or steering committee oversight. These boards have responsibility for ensuring that any specimen use is in accordance with the mission and policies of the organization, including being in accordance with the consented use, and also for ensuring that privacy and confidentiality procedures are in place and are followed. The RAND study on best practices identified some ethics boards as having established tiered consent processes in which individuals may choose the type of specimen(s) they want to donate, the type of research for which the specimen(s) may be used, and whether their clinical records may be accessed in relation to the research. This consent should be separate from any consent required for delivery of healthcare, such as an examination or surgery. These policies should also allow individuals to withdraw their consent in the future and have corresponding procedures in place to account for removal of samples and deletion of associated data [2].

4.3 Risks associated with research using human specimens

Potential risks to subjects whose specimens and associated health data are used in research may include **physical risks**, particularly if the specimens are taken specifically for research purposes. Physical risks can include those involved with medical procedures, such as blood draws or extra biopsies taken for research purposes. Often, however, residual specimens taken during the course of routine medical care (e.g.,

diagnostic specimens) are used for research, which means that additional physical risk beyond that involved in the diagnostic procedure generally is not incurred. In some cases researchers might request permission from patients to take extra tissue for research during the course of clinical care, for example, an extra 5 cc of blood. In these situations, there is usually minimal additional physical risk.

The primary risk in research using human specimens is likely to be **loss of privacy and confidentiality,** which can occur anywhere along the continuum of obtaining, storing and distributing materials for research use. Concerns about breeches of confidentiality have increased as a result of advances in genetic and other molecular technologies, as well as the broad sharing of data made possible by sophisticated information technologies. Today, research involving specimens has the potential to identify genetic or other molecular alterations that may have implications for an individual's current or future health, or that of their immediate family, such as the presence of disease or other unsuspected risks. In addition, the improper use or disclosure of such information could result in psychosocial harms (such as anxiety) or the loss of employment or insurability.

An additional risk to subjects involves the **improper use of unvalidated research results for clinical decision-making**. This includes the use of the results of tests that have not been approved by FDA or performed in Clinical Laboratory Improvements Amendments (CLIA)-approved laboratories for patient treatment and care.

In addition, in some cases, research on human specimens and associated health data also could raise risks to groups of individuals (**racial or ethnic discrimination**). For example, research using specimens may determine that a particular group of individuals (e.g., a specific racial or ethnic group) have an increased risk of developing disease. Disclosure of such information could have implications for insurability and/or employment and the potential for stigmatization. Fathoming the desires of the subject is often difficult when there is the potential for group harm from possible stigmatization of the group. IRBs must be sensitive to this concern when reviewing the requests for subsequent research activity and examine the sampling schemes both for the bank and for the tissues or data to be examined.

A risk that is less quantifiable to the subject is that associated with **future, unspecified, research purposes**, i.e., the uses are unspecified at the time of the initial specimen banking. These are the risks that often give IRBs and ethicists the most pause. Here, the subject does not have the opportunity to speak for himself or herself given a specific research protocol; the IRB has the responsibility to assess risks and make decisions that do not abrogate the rights and welfare of the individual. The IRB must try to understand the subjects' interests and decide under what circumstances consent can be waived and when a new consent might be warranted. Most of the risks associated with future, unspecified, research are non-physical harms; such as loss of confidentiality. The risks associated with loss of confidentiality can be mitigated in a number of ways, including anonymization, good security and privacy practices, and certificates of confidentiality.

Meeting the promise of research using human specimens requires that custodians and users of that material ensure that the proper safeguards are in place to meet ethical, and legal/ regulatory requirements [4, 20].

4.4 Institutional Review Boards (IRBs)

IRBs are responsible for the oversight and review of research that involves human participants to ensure that their privacy is protected and confidentiality of data is maintained. In the United States, Food and Drug Administration (FDA) and the Department of Health and Human Services (HHS) regulations have empowered IRBs to approve, require modifications in (to secure approval), or disapprove research. Therefore, the collection, storage and distribution practices of federally funded repositories, as well as the research for which the investigator is requesting tissue samples, should be subject to IRB review and approval.

By design, IRBs are multidisciplinary bodies with a mandated minimum membership of five persons, including at least one person whose primary interests are in a scientific area, one whose primary interests are nonscientific, and one who is neither associated with the institution nor immediate family of anyone at the institution. IRB composition should reflect appropriate race, gender, and cultural diversity, and the board cannot be comprised exclusively of men or of women. Additional relevant expertise is to be included when the research under review involves especially vulnerable populations (e.g. children, disabled people). Many IRBs are much larger than this requisite membership, perhaps having as many as 20 or 30 members. The regulations specifically prohibit conflicts of interest for IRB members (e.g. no member should be involved in the research study under review). In situations of multi-institutional collaborative research, each location is responsible for compliance with the federal policy.

IRBs evaluate research protocols based on a number of criteria that assess the risks and benefits of the proposed research, the nature of the informed consent process, and investigators' approach to ensuring privacy and confidentiality. More specifically criteria for review include:

1) Risks to participants are minimized
2) Risks are reasonable in relation to anticipated benefits
3) Selection of participants is equitable
4) Informed consent is sought and documented
5) Adequate provision is made for monitoring the data collected to ensure safety of participants
6) Adequate protections for privacy and confidentiality are in place
7) Additional protections are in place in cases of vulnerable populations [20].

4.5 Conflicts of interest

Institutions and IRBs increasingly are confronting conflicts of interest as more investigators and institutions enter into financial arrangements in which they stand to benefit from the results of their research. Conflicts of interest are not new, but they have changed and intensified as the research enterprise has evolved. Disclosure and management of investigator and institution conflicts still seem to be the best strategies. Conflicts affecting the IRB can be handled by increasing the percentage of members who do not have any direct interests in the institution or its research program (unaffiliated members). Increasing the percentage of non-scientists and members who represent the

views of participants can also reduce conflicts. Ultimately, clear policies and guidance will provide the best way for IRBs to fulfil their responsibilities and meet their obligations in educating their members, monitoring for compliance, and avoiding conflicts of interest [50].

4.6 Privacy of Genetic and Other Health Information

The expectation and assurance of privacy are critical to gaining the cooperation of patients/subjects to donate samples and data in the first place. Potential donors may be concerned that information about them derived from tissues and medical records may be released and harm them. One of the primary concerns is that employment and insurance discrimination might result from exposure of information about health history, genetic makeup, or familial predisposition to disease. Although there are anecdotal reports of employment termination and denial of insurance coverage based on genetic predisposition to disease, there is no definitive research on the extent of these risks; however the perception is real and the risks exists [4]. The Health Insurance Portability and Accountability Act (HIPAA) of 1996 in USA, holds the promise of more significant, uniform privacy protections. HIPAA is perhaps best known for its efforts to improve access to and protability of health insurance, but it also promulgates new privacy and security standards for use and access to individual health information [20, 51].

A few years ago, the American Society of Human Genetics described four types of identification of samples for research purposes:

a) **anonymous**: biological materials that were originally collected without identifiers and are impossible to link to their sources;

b) **anonymized**: biological materials that were originally identified, but have been irreversibly stripped of all identifiers and are impossible to link to their sources;

c) **identifiable** or **coded** or traceable: biological materials that are unidentified for research purposes, but can be linked to their sources through the use of a code. Decoding can only be done by the investigator or another member of the research team;
d) **identified**: biological materials to which identifiers, such as name, patient number, or clear pedigree location, are attached and made available to the researchers.

Recently, the European Agency for the Evaluation of Medicinal Products gave some guidance in dealing with coded samples, adding the category "**double**-**coded** samples and results", where a first code is assigned to the sample and a second code is provided to link it to the results. The key-code linking the double coded samples and the information is kept by a third party. The research subject can only be linked with the sample or data obtained from it by bringing the two code keys together [52].

The more identifiable the specimens are, the greater the risk to donors' privacy and confidentiality. Tissue samples are more useful to researchers when accompanied by demographic and clinical information, some of which may make them identifiable. In order to protect privacy, some repositories keep identifiable information on tissue sources at the collection sites rather than at the main repository site. All of the repositories that keep identifiable information on site generally limit access to identifiable information to select staff. A few repositories use the "honest broker" model, which uses a neutral intermediary, between the individual whose tissue and data are being studied and the researcher, to collect and collate pertinent information regarding the tissue source, replace identifiers with a code, and release only coded information to the

researcher. Limiting access to the codes that link patients' identifying information to their tissue specimens, through physical and/or cyber procedures, is a best practice to ensure that patient privacy and confidentiality are protected. NBN must be able to assure its partners in the system that they will not be at risk for privacy violation through their relationship with NBN.

IRBs are responsible for the oversight and review of research that involves human participants to ensure that their privacy is protected and data confidentiality is maintained. IRB review is critical to oversee the repository practices, and approve or disapprove specific research projects. Some of the repositories also rely on separate bioethics advisory boards or committees to oversee privacy and confidentiality procedures.

Requiring repositories to have IRB approval for the collection, storage, and distribution of bio-specimens and associated data, and requiring researchers requesting samples to have IRB approval of the research projects that will use the samples are best practices.

Having a bioethics advisory board or other governance and oversight board/committee to oversee privacy and confidentiality procedures is a best practice that provides another layer of review [17].

4.7 Non-physical risks

Non-physical risk or "information-related" risk is related to the information collected during the storage of a biological specimen, the research or produced by genetic analysis. This type of risk involves:

• psychological damage to the subject through awareness of individual results

• damage to the individual and breach of confidentiality caused by disclosure of individual results to third parties

• damage caused by stigmatization of the subject's ethnic group, if the research involves or focuses on a specific ethnic group [35].

4.8 Genetic Privacy and Genetic Discrimination

The development of genetic research has been accompanied by concerns of stigmatization and discrimination of groups of individuals. This threat, which is certainly not peculiar to genetics, has often been perceived in the past – for instance, in attitudes to disability, rare diseases, HIV infections and many other conditions. In genetics, stigmatization and discrimination could occur on the basis of the genotype or ethnic group. Population groups who are "genetically disadvantaged", for example because their genotype exposes them to the risk of developing a specific pathology or resistance to a certain medicine, could be stigmatized or discriminated against (with regard to such issues as access to health insurance, mortgages and employment). Since many points are still relatively unclear in this field and there is a lack of clear regulatory provisions, the actual risks are at present difficult to determine. The following considerations should afford some guidance.

Every individual has genes creating susceptibility to important diseases, and in the vast majority of cases genotyping does not make it possible to predict with certainty that the disease will actually occur. For instance, a person who is APOE•4-positive has an

increased risk of developing Alzheimer's disease, but an APOE•4-negative subject's genotype may indicate susceptibility to stroke, cancer or even Alzheimer's as a result of alleles other than APOE•4 or of environmental factors. There are thus no genetic grounds for considering the two subjects concerned at all different with regard to their risk of developing any disease. However, this does not mean that there is no risk of stigmatization and discrimination – an employer might decide not to hire the first person on the basis of the available information, but accept the possible risks of not having any information on the second person.

Another source of stigmatization can be the ethnic group. It is well known that specific mutations in BRCA1 and 2 genes responsible for hereditary breast cancer are more frequent in Ashkenazi Jewish women than in other ethnic groups, including Sephardic Jewish women. This means that if a woman is known to be of Ashkenazi Jewish descent, she is implicitly known to have an increased risk for early development of hereditary breast cancer, even if the result of her genetic test is not known. Similar considerations apply to the whole range of environmental risk factors, for example in the case of populations living in highly polluted areas as opposed to the countryside, or of a Sephardic Jewish woman who may be susceptible to other diseases or to the sporadic form of breast cancer, far more common than the hereditary type. The EPO (European Patent Office) nevertheless granted a patent to the University of Utah Research Foundation on a BRCA2 gene mutation analyzed for "in vitro diagnosis of predisposition to breast cancer in Ashkenazi Jewish women". The European Society of Human Genetics, which is opposed to the patent, regards it as discriminatory against the Ashkenazi Jewish ethnic group.

The possibility of discrimination against patients with a genotype predisposing to a reduced drug response is sometimes raised. This information cannot harm the patient – indeed, it is surely useful to know in advance if the patient will experience benefits or adverse effects from a medicine, rather than find this out only when treatment is in progress. It is sometimes thought that systematic use of pharmacogenomics and pharmacogenetics in drug research and development could channel research toward drugs for common responder genotypes, to the detriment of medicines for less frequent genotypes. However, even without pharmacogenetics, a pharmaceutical company developing a new drug will have access to the same sort of information on the proportion of responders and non-responders once early phase II studies have been carried out.

The risks of stigmatization and discrimination on genetic grounds thus seem to be based more on overestimation of the predictive power of genetic testing and genetic predisposition, rather than on valid scientific and medical assumptions [35].

4.9 Some Special Issues Involving Research Without Adequate Consent

There are a number of clinical and research settings in which research is often done on stored tissues without express, appropriately specific consent from the source of the tissues, their legal representatives, or their families. If consent is solicited for research the consent is frequently (1) some version of general consent for undisclosed research purposes or (2) a version of presumed consent with an opt-out procedure. Some examples that involve bio-banks are:

- *Neonatal blood spots in government-sponsored DNA banks:* In many countries millions of dried blood spots are stored each year in government-sponsored DNA

banks without appropriate consent. Some research studies use dried blood spots that are linked to specific children; other population-based studies use anonymized spots. At least some information should be disclosed to the parent(s) about post-screening research: plans for research with the dried blood spot; plans regarding the identity status; time limits if any; information about future access by third parties to the samples and/or data; and options parents have to communicate their consent (or refusal) to the post-screening research.

- *Secondary research*: The term secondary research refers to the use of a tissue sample or cell line by scientists other than the original investigator(s) or for scientific purposes unrelated to the original study. Secondary research poses complex ethical, political, technical and social challenges.
- *Research on tissues from the now dead.*
- *Research with cryo-preserved human embryos.*
- *Vulnerable research participants* [20].

4.10 Ownership of banked samples

Many of the agreements about ownership of banked samples and access to biological material and information are determined by multiparty contracts and are not regulated by legislation. The general practice is that information belongs to the researcher or team that creates it and the individual who may have been a subject of the research has no legal entitlements to that research. The claim that subjects should own their sample even while it is entered into research is the minority view. In 1988, the ASHG statement established that 'banked DNA is the property of the depositor unless otherwise stipulated'. On the other hand, the British MRC Working Group on Collections of Human Tissue and Biological Samples for Use in Research considered that the funding body retains ownership of the collection while the researcher is the custodian of the collection. The custodian has the responsibility of control over the access to the collection and ensuring that standards of confidentiality are maintained. Funding bodies need to determine the purpose of the collection and if it is available to both commercial ventures and academic researchers. It has been suggested that potential subjects should decide whether they are willing to participate only after they have been informed about who will own the sample (as it is for tissue) and whether or not there is a plan for subsequent use of sample. Subjects should have the right to decide to donate their tissue to the research team. As with other aspects of life, competent individuals have the right to make gifts and these gifts are owned by those who receive them.

The issue of ownership arises primarily because of the possibility that DNA samples could have some commercial value. This issue is most controversial in connection with the patentability of 'inventions' derived from the scientific analysis of human material. This patentability remains contested in various nations and within particular groups.

At the international level, HUGO has maintained a consistent position on the unpatentability of human DNA sequences. At the regional level, in 1994, the Council of Europe's recommendation on the Protection and Patentability of Material of Human Origin stated that 'human beings are subjects – not objects – of law'. After 3 years, in its Convention on Human Rights and Biomedicine, the Council of Europe insisted that 'the human body and its parts shall not, as such, give rise to financial gain'.

Finally, in 1998, the European Commission Directive concerning the Legal Protection of Biotechnology Inventions intended to complement international intellectual property provisions. By distinguishing between inventions and discoveries (the latter being unpatentable), the Directive responds to concerns that the human body or parts thereof could be patentable. According to the Article 3, 'an element of the human body in its natural environment or even a sequence or partial sequence is unpatentable', while Article 9 states that 'even a patentable invention – according to the European patent law – can be excluded if contrary to public order and morality'.

At the national level, the status of human genetic material is not so clear: some countries have not taken any position on the issue of status, or among the few that do, the position is often through an indirect reference. Whatever the country, any inventions even if patentable are still subject to the public order and morality filter of European patent law. Finally, while a property position may allow for actual or potential financial return, the personality approach avoids individual returns but not the possibility of commercialization by the researcher, through traditional intellectual property rules. Thus, irrespective of the qualification, ultimately, patenting is still possible but the locus of the financial benefits is different. There are important questions which remain unanswered about the social acceptability of the private ownership of gene patents, and the impact this might have on scientific research, innovation and the costs of new medical technologies [44].

4.11 Access and sharing

Worldwide research endeavors have raised the issue of access to and sharing of banked samples. While protecting confidentiality, the free circulation and the availability of genetic information and samples for research have been promoted by many instances. At the international level, HUGO (1996) and UNESCO (1997) mandated that DNA samples should be openly available to the scientific community: 'collaboration between individuals, populations, and researchers and between programs in the free flow, access, and exchange of information is essential not only to scientific progress but also for the present or future benefits of all participants. Cooperation and coordination between industrialized and developing countries should be facilitated'. WHO adopted the same position: 'qualified researchers should have access to samples, provided that strict confidentiality is observed or that identifying characteristics are removed'. Concerning access by relatives, HUGO stated that 'Where there is a high risk of having or transmitting a serious disorder and prevention or treatment is available, immediate relatives should have access to stored DNA for the purpose of learning their own status. These exceptional circumstances should be made generally known at both the institutional level and in the research relationship'. WHO went further and recommended that 'control of DNA should be familial, not individual. All blood relatives should have access to stored DNA for purposes of learning their own genetic status, but not for learning the donor's status'. At the regional level, in 1991, the Council of Europe proposed that 'standard research contracts have a clause to the effect that the sharing of all knowledge and distribution of materials will be obligatory. Data and materials as much as possible should be free of charge or available at a nominal cost or to cover distribution costs'. In 1997, the same instance adopted the Recommendation on the

Protection of Medical Data which provided that 'Medical data can be used by health professionals for their own medical research as long as the data subject has been informed of this possibility and has not objected'. At the national level, access to medical

records or to samples for genetic research is normally restricted to qualified investigators and subject to institutional oversight, be it legislative or via ethics committees.

Generally, consent of the patient is also required. Ethics committees or legislative provisions may also grant access to records or samples without consent for research purposes, mainly if the data are anonymized (French Law No. 94-548 1994). In most international and national legislation and protocols, individuals are considered to have an absolute right to give or to withhold information about their genetic status, and equally an absolute right to prevent their stored genetic data being transmitted to a third party for whatever purpose. However, one of those purposes might be further research in which individuals' genetic data might assist in securing health benefits for a large number of other people. Some people wonder if in such a case should such individuals be able to exercise a right to withhold their genetic information, particularly if it is encrypted or anonymized, when it might play a part in establishing links and patterns with genetic defects in members of their families or some wider social grouping and thus contribute significantly to their well being. It has been argued that the principle of inalienable individual rights that lies at the heart of much of the legislation about data protection may not be in the best interests of the population's health [44].

4.12 Special Issues in Forensic Settings

Some tissues are used for non-biomedical purposes, for example for identification of human remains from destructive scenes of death. In USA, the Department of Defense's DNA Repository was established in 1991 and contains more than 2 million DNA specimens along with a database that matches each DNA sample with the Social Security Number and abbreviated last name of the individual from whom it came. A more recent example of human remains identification after tragic deaths is the work of numerous pathologists in New York and other states as they have tried to identify the thousands of persons killed by terrorists on September 11, 2001.

A more publicized and more controversial use of stored tissues for non-biomedical purposes is the practice of using DNA evidence in court to convict (or sometimes exonerate) individuals facing criminal charges. This widely used practice of DNA forensics has been carried out in a number of countries since it began in England in 1985, with the British legal system providing substantial support over the years for the ever-widening use of this law enforcement technology, including the controversial practice of "sweeps" (the collection of large numbers of DNA samples from persons living in an area where a terrible crime occurred). Now, laws for forensic DNA databanks have been enacted in Australia, Canada, China, France, Germany and several other countries.

The use of DNA evidence for forensic purposes is likely to increase in the future, especially as the field of molecular genetics changes with the development of new analytic technologies that can be used in the pursuit of justice. Equally predictable, the forensic DNA database laws in the Europe and other countries are likely to undergo many changes in the future as crime rates, imprisonment numbers, government budgets, public opinion polls, and civil libertarian concerns rise or fall. Some issues that seem imperative for legislators to consider regarding existing DNA database laws and use of DNA databases for forensic purposes are the following:

Crimes targeted by law: Only sexual felonies? Also property crimes? Various non-violent crimes? “White-collar” crimes? All felonies?

Inclusion of juvenile offenders.

Timing of collection of DNA samples. After conviction? After arrest? Before parole from prison?

Length of sample and data storage.

Procedures for destruction/deletion.

Use(s) of stored samples and data. Restricted to “law enforcement”? To “identification purposes”? Or also to be used for research on forensic techniques? What about use in non-forensic research?

Security provisions and third party access.

Provisions for unauthorized disclosures. Financial or criminal penalties? If any.

Population-wide DNA profiling [20].

5 Bio-specimen Collection, Processing, Annotation, Storage and Distribution

Best practices for bio-specimen collection that will increase the quantity and variety of high-quality samples available to researchers, while maintaining appropriate normal controls, include establishing a network of collection sites at academic medical centers and community hospitals, and collecting tissues from a broad range of diseases, non-diseased matching adjacent tissue, normal tissue, and other biological specimens (e.g., whole blood, serum, and plasma). It is also important that tissue be collected from ethnically diverse populations of all ages to ensure that the tissue available for research purposes is diverse and demographically representative of the population and to expand biomedical research to include understudied/underrepresented populations and the study of health disparities.

The prioritization of patient diagnosis over collection of specimens for research purposes is key to ensuring that patient care is not compromised and that patients continue to donate bio-specimens. Pathologists at the collection site play an important role in the initial procurement of the specimen for the repository, and repository pathologists are central to the quality control procedures for verification and evaluation of the specimen. In addition, repository-trained personnel using standard operating procedures and standard collection and processing equipment are important to promoting standardized tissue collection and processing.

Best practices for data collection depend on the mission of the repository. However, no matter what the requirements for the amount of associated data are, certain best practices are applicable. It is important to collect consistent and high-quality data associated with bio-specimens and to employ a standardized set of common data elements that are collected with every bio-specimen. It is also important to define the data set that is optimal for fulfilling the mission of the repository and the needs of its customers, and to collect the data (such as demographic and pathologic data, family history, medical history, lifestyle and diet history, treatment history, and clinical outcomes) required to meet those needs.

Once data are collected, they must be entered into the repository's bioinformatics system. The use of common data elements and standardized terminology for data collection procedures allows the use of standardized data-entry forms with features that minimize the errors introduced while typing information into forms. In addition, scannable bar codes are used to track bio-specimens and associated information throughout their lifetime at the repository. Parsing techniques are used to flag discrepancies and to record errors and their reconciliation. The use of standardized terminology and computer data entry forms, scannable bar codes, and data reconciliation techniques are best practices that ensure data accuracy.

Standards for storage depend on tissue type and preservation condition (e.g. snap-frozen, paraffin embedded, tissue microarray). Snap-frozen specimens are commonly stored at -80°C in mechanical freezers or in liquid nitrogen. Paraffin-embedded tissue and tissue microarrays are stored at room temperature or in a climate-controlled environment to protect them from melting or other damage. However, there is no consensus on the optimum storage conditions for specimens.

Once specimens are placed in storage, it is necessary to monitor storage conditions and

maintain equipment. Standard operating procedures for freezer maintenance, adequate backup equipment, and redundancy in storage location are best practices for ensuring that specimens are stored and maintained at the necessary temperature and condition and that specimen integrity is not compromised. Periodic auditing, inventories, and certification of the location, identity, and quality of specimens ensure the quality and integrity of samples sent to researchers. Bar coded inventory systems are used to track specimen storage location.

Standardized and carefully monitored shipping procedures track all shipments in and out of a repository. Bio-specimens sent to a repository from remote/satellite collection sites and samples sent from that repository to researchers are tracked using electronic technologies, such as bar coded inventory systems or smart chips and radio-frequency identification tags.

Specimen distribution practices clearly depend on the mission of the repository. If the mission is to provide tissue samples to as broad a base of researchers as possible based on the quality of the proposed research, then bio-specimen distribution policies should be established to fulfill this mission. If the mission is clearly defined, and if the repository evaluates its ability to meet its goals and changes its policies, procedures, and practices when not meeting those goals, then this is a best practice.

Quality assurance is fundamental to the successful operation of any bio-specimen repository. The use of standardized protocols for collection, storage, processing, and distribution of specimens, and the use of common data elements for the annotation of specimens at each of the individual network participant locations make comparative research across participating institutions possible. To ensure that the collection, processing, annotation, storage, and distribution of bio-specimens occur at a consistently high level of quality, it is necessary to have a multitiered, fully integrated quality assurance system and standard operating procedures. Quality assurance starts with the training of personnel before bio-specimens are ever collected and includes everything up through considering researcher feedback on sample quality.

5.1 Organization of a Tissue Collection Network

5.1.1 Collection locations

NBN infrastructure could be organized as: a) a centralized virtual repository with data networked across the nation or b) a decentralized network, possibly of non-profit, tissue repository organizations located near academic medical centres [4].

There are three main models of collection and storage operations: decentralized collection with centralized storage, centralized collection and storage, and decentralized collection and storage. Many of the repositories are decentralized in the collection of specimens—meaning there are multiple collection sites that are geographically dispersed, usually involving some combination of academic medical centers and community hospitals—but store their specimens in a centralized facility. Some repositories have both centralized collection and storage. Whatever the infrastructure is, repositories should have clearly formulated and documented mission. It should be defined as to whether it is a freestanding entity, virtual, or part of an institution [53].

5.1.2 Staff

Bio-banks should be adequately staffed and the personnel selected for these tasks must

have an appropriate level of specialized training. Running a bio-bank requires dedicated staff for specimen processing, storage and for data management. The job description, tasks and reporting system of all supervisory and technical staff involved in the bio-bank must be documented. This is of particular importance in instances where the staff involved performs other tasks within the institution. Staff must have adequate educational background, experience, and training to ensure that assigned tasks are performed in accordance with the bio-bank established procedures [54].

5.1.3 Records management

Each repository should develop a records management system that permits detailed records to be made concurrently with the performance of each step in the collection, processing and distribution of specimens. Records maintained may include but are not limited to: protocols, standard operating procedures (SOPs), informed consent documentation, procurement documentation, processing records, testing, equipment maintenance, audit/review documents, specimen storage location information, sample distribution, and quality control activities. Records should be created and maintained in a manner that allows steps to be clearly traced. Security systems should be adequate to ensure the confidentiality and safety of all stored records. Access to records should also be on a "need to know" basis.

In some cases it may be necessary to either destroy or remove specimens at the request of study participants. Under these circumstances, records should be appropriately modified to indicate that the specimen is no longer part of the collection and the information management system should be adequately modified to reflect this event.

Forms should be numbered, have a distinct title, and include the date that the version of the form was created.

Corrections or changes in a hard copy record should be made in ink with a single line drawn through the altered text. Corrections should be initialed and dated by the individual making the correction or change. Changes in electronic records should be noted and tracked. Changes tracked should include the name of the individual making the change, and the time and date at which the change was made.

Unless otherwise specified by contract or other agreement, each repository should establish a period of time during which all records are maintained. A policy should be in place for the destruction or return of records that no longer need to be maintained.

Electronic records should be backed up daily on a network or remote server and weekly on a CD or diskette or other appropriate media. Computers operated by repository staff should be password protected and should make use of automatic screensavers. Permission levels should be created for staff at different operational levels as well as for users who are not repository staff, where this is allowed.

Records should be readily accessible for inspection by authorized personnel from regulatory agencies (these may vary for each state or country, depending on the regulatory agencies with jurisdiction over those activities) and Quality Assurance (QA) personnel [53].

5.2 Bio-specimen Collection

5.2.1 Standard operating procedures (SOP)

Bio-banks should develop, document and regularly update policies and procedures in a standardized written format incorporated into a Standard Operating Procedures Manual that is readily available to all laboratory personnel. The SOP Manual should specifically include:

● Specimen handling, processing and labelling.

● Procedures including supplies, methods and equipment.

● Policies and procedures for shipping and receiving specimens.

● Records management policies.

● Quality assurance and quality control policies and procedures for supplies, equipment, instruments, reagents, labels, and processes employed in sample retrieval and processing.

● Emergency and safety policies and procedures, including reporting of staff injuries and exposure to potential pathogens.

● Policies and procedures for the investigation, documentation and reporting of accidents, errors, complaints and adverse outcomes.

● Policies and procedures and schedules for equipment inspection, maintenance, repair and calibration.

● Procedures for disposal of medical and other hazardous waste.

● Policies and procedures describing requirements of training programs for staff [54].

5.2.2 Specimen Types

A variety of specimens may be collected for storage:

• Blood and blood fractions (plasma, serum, buffy coat, red blood cells)

• Urine

• Buccal cells/saliva

• Hair

• Nail clippings

• Breast milk

• Feces

• Exhaled air

• Tissues—surgical, autopsy, frozen, paraffin-embedded

• Cell lines

• Products of conception

• Cord blood

• Bone marrow

• Fluids from cytology (ascites, pleural fluid, synovial fluid, etc.) [53].

5.2.3 Collection and labelling procedures

***Special considerations for specimen collection protocols*:** Various protocols exist for the collection of different specimens.

***Timing of specimen collection*:** Biological marker levels may vary according to the time of day; however, a much greater effect may be the time and conditions associated with specimen processing. In general, specimens should be processed as rapidly as possible and it is important to document the collection, processing, and storage times.

Temperature: The temperatures at which specimens are collected, processed and stored are very important and depend on the specimen type and intended analysis.

***Specimen biological stability*:** The stability of specimens may be affected by any of the following:

• Anticoagulants used in blood collection.

• Stabilizing agents (*e.g.*, EDTA, ascorbate necessary to preserve folates and vitamins) should be included in the collection device or added as soon as possible after collection to assure stability of analytes.

• The time elapsed between specimen collection, or removal from a storage unit, and subsequent processing. Cell viability may be affected negatively as this period of time is extended.

• The temperatures at which specimens are collected and subsequently processed and stored are critical and must be carefully considered depending on the type of specimen and intended analyses. Warmer storage environment may be permissive for macromolecular degradation (RNA is particularly susceptible).

• The sterility of the instruments, surfaces, and equipment is an important consideration to prevent subsequent contamination. Sterility is particularly important if the intention is to collect RNA or to culture cells from the sample.

• Degradation—Enzymatic degradation affects many biochemical markers. RNA and proteins are particularly susceptible to enzymatic degradation and require special procedures to maintain their integrity during collection and processing.

• Containers/equipment—Collection containers and freezers or other equipment vary according to specimen types being collected [53].

Labelling of Bio-specimens

- Each specimen should be labeled in such a manner that the labeling will survive all potential storage conditions, in particular dry ice and liquid nitrogen. Ink used on the label should be resistant to all common laboratory solvents.
- Labels should be printed with a linear barcode if possible, thus providing a direct link to database software. However, it is also important to include human-readable indications of contents.
- All specimens should ideally be labeled with at least 3 human-readable forms of identification on them, for example patient number, specimen number, patient date of birth.

- Suggested information for the label is the tissue bank's unique identifier number, sample type, and date of collection / banking, plus a barcode if available.

Collection of Blood

- All blood should be treated as potentially infectious.
- Staff handling blood should check their Hepatitis B titer every 2 years.
- It is 'best practice' to take tissue bank blood samples concurrently with routine clinical blood samples, so as to minimize additional inconvenience and discomfort to patients.
- In the case of lymphoma, leukaemia, and myeloma donors, blood should be collected by fresh venipuncture rather than from an existing line.
- Blood may be collected into EDTA, ACD, Lithium Heparin, or into a clotted tube containing separating gel. If DNA is to be extracted, either EDTA and ACD tubes can be used. However, ACD is more appropriate if there is to be an extended time lapse between blood collection and processing. Lithium Heparin is generally only used if cytology studies are being performed. Lithium Heparin interferes with DNA yields and PCR applications. ACD tubes can be used for blood from remote locations, as viable white blood cells can be recovered up to 10 days after collection.
- The amount of blood usually collected varies for different diseases. In most cases, 2 tubes (18-20 ml) blood is an ideal collection amount. This volume collected is guided by ethics clearance. However, sometimes there will be less than the expected amount, and in this case priorities need to be set as to the final blood products made. Reduced volume of blood in a tube containing additives should be noted so as to avoid confounding of results by variation in additive concentration.
- Ensure tubes are clearly labeled.
- Time of collection and time of freezing should always be recorded, as well as any variations to the processing protocol.
- Blood should be transported at room temperature, although some proteomic applications require transport on dry ice.
- All blood should be processed within 48 hours of collection. Cell viability decreases rapidly after 48 hours, resulting in poor cell structure in slide preparations, or degradation of proteins and nucleic acids.
- Serum and plasma should be stored within 2 hours.

Collection of Solid Tissues

- Treat all tissue as potentially infectious.
- The collection process should be carried out in the most aseptic conditions possible.
- The intact operative specimen should be sent as soon as possible to pathology.
- It is considered to be 'best practice' to collect and process specimens within one hour of excision. Transfer of specimens must be carried out as quickly as possible in order to minimize the effect of hypoxia upon genetic expression, and degradation of RNA and other tissue constituents.
- For transport from surgery to pathology, or to the repository, specimens should remain fresh (not fixed) and be placed in a closed, sterile container on wet ice. Do not immerse them in liquid of any kind.
- The collection of samples for research should never compromise the diagnostic integrity of a specimen. Ensuring that only tissue exceeding diagnostic needs is banked means that patient care is never compromised, and helps retain public confidence in donating.
- It is important to have a pathologist supervise the procurement of the tissue (for quality assurance purposes).

- The pathologists will examine the sample, and, allowing adequate tissue for histological diagnosis and assessment of margins, will remove a portion of the tumour and adjacent normal tissue, if appropriate for the tissue bank.
- When selecting specimens, those areas with massive ischaemia and / or necrosis should be avoided.
- The anatomical site from which the tissue is taken must be recorded.
- Tissue bank staff will be present in pathology to freeze and fix the tissue as quickly as possible to maximize the RNA preservation.
- Samples requiring snap freezing can be frozen in a dewar of liquid nitrogen or on dry ice at the time of collection. Do not allow direct contact of the tissue with the liquid nitrogen.
- The approximate time that elapses before freezing and fixation should be noted.
- Tissue may be embargoed for an agreed period with the donating pathologist (commonly 2 weeks) before homogenizing or cutting in case of diagnostic recall.
- If the tissue is being delivered from a remote site, ensure good communication between the sender and the receiver.
- Tissue should be placed in labelled histo-cassettes on biopsy pads if being sent from remote locations. They should be sent in liquid nitrogen or dry ice. For remote sites where dry ice / liquid nitrogen isn't readily available, tissue collections into RNA*later* act as a good alternative.

Collection of Urine

- When in transit, urine collections should be maintained on ice or refrigerated.
- Plastic or glass containers should be clean and dry, and have a 50-3000ml capacity, wide mouth, and leak-proof cap.
- Depending on the analyte to be measured, a preservative may be added.
- Urine should be processed and stored within 48 hours.

Collection of Buccal Cells

The collection of buccal cells is logistically not difficult and does not require highly trained staff; thus, it may be more feasible than blood collection when staff are not on site, or the population chosen for a study is geographically diverse. Buccal cell collection should therefore be considered when non-invasive, self-administered, or mailed collection protocols are required. Donors who do not give blood should also be asked to donate a buccal cell specimen; however, buccal cell collection will yield only limited amounts of DNA in comparison to blood.

A collection kit (containing mouthwash, 50 ml plastic tube, plastic biohazard bottle, and courier packaging) is mailed or given to the participant, along with an instruction sheet.

- The participant is to brush their teeth as usual, rinse their mouth well twice with water, and then wait 2 hours. They should not eat or drink anything other than water during this time.
- After 2 hours, 10 ml commercial mouthwash should be poured into the tube, and then 10 ml tap water added. This diluted mouthwash should be placed into the mouth (without swallowing) and swished around vigorously for 30 seconds.
- The mouthwash should then be spat back into the plastic tube and the tube should be sealed tightly.

- The sample should be sent back to the tissue bank immediately for processing, or stored at 4°C until sent (it should be sent within 24 hours).

Collection of Bone Marrow

- Treat all bone marrow aspirated as potentially infectious
- Staff should check their Hepatitis B titer every two years.
- Samples should be collected in an EDTA tube.
- For leukaemia and myeloma donors, a 2 ml aspirate should be collected in a 4 ml Lithium Heparin or EDTA tube.
- Transport bone marrow aspirates at room temperature.
- Process bone marrow aspirates within 24 hours of collection [55].

5.2.4 Bio-specimen Processing

Bio-specimen processing depends on the tissue type and purpose of the intended study and differs among different institutes and repositories. The purpose of this document is not to provide complete protocols but to present some examples of standard processing procedures.

5.2.4.1 Processing of Blood specimens

Guthrie Cards

Guthrie Cards are made from pure cotton and can be used for the extraction of DNA. Blood spot collection should be considered as alternative to whole blood when protocols call for easier collection and cheap room-temperature storage. Always handle Guthrie Cards wearing gloves and only by the upper corner, marked out for labelling. Do not allow the card to come into contact with any unclean surface e.g. bench, base of hood.

(use EDTA / ACD tubes to make Guthrie Cards)

1. Mix blood thoroughly by inversion before starting.

2. Wipe top of vacutainers with ethanol before opening.

3. Use the fullest vacutainers to make 2 Guthrie cards by placing 40 µl of blood in the circle using a P200 pipette. Wipe the top of the vacutainer with 70% alcohol before removing the lid.

4. Air dry thoroughly in the back of the Class II Biological Safety Cabinet.

5. Store in a paper envelope (not plastic) at room temperature. Protect from moisture and rodents.

Blood Pellets (White Cells)

Blood Pellets can be used for the isolation of DNA (from EDTA / ACD tubes)

1. Transfer blood from the original tube to a labelled 50ml tube.

2. Fill tube with Tris-EDTA buffer (formula) and mix vigorously. Place on ice for 5 to 10 minutes.

3. Spin as soon as possible at 1200g for 10 minutes.

4. Carefully pour off supernatant into a beaker containing chlorine bleach. Briefly vortex the pellet and add 50 ml Tris-EDTA buffer. Shake vigorously.

5. If division of the sample is necessary, at this point pour 25 ml of sample into another

falcon tube.

6. Spin both tubes at 1200g for 10 minutes.

7. Repeat washing if red cells persist.

8. Carefully pour off supernatant.

9. Using a swirling motion, remove the pellets (and a small volume of supernatant) with a P1000 pipette and transfer to 2 labelled cryovials.

10. Store in -80°C until DNA required.

Plasma

Plasma can be used for bioassays, plasma DNA isolation, proteomic analysis, and biomarker discovery (from EDTA / ACD tubes)

1. Spin one vacutainer (about 9 ml) at 815g for 10 minutes at 4°C to separate plasma.

2. After wiping each tube with alcohol, remove about 3 ml plasma (but not the white cells in the buffy coat). Tube can be retained for white blood cell extraction.

3. Transfer to a clean, labelled 15 ml tube.

4. Centrifuge at 3200g for 10 minutes at 4°C.

5. Aliquot plasma into 1 ml labelled cryovials (3 to 4 aliquots).

6. Place in liquid nitrogen Dewar to snap freeze.

7. Store at -80°C.

The purpose of double spinning the plasma is to remove all cellular contaminants so that the plasma is suitable for plasma DNA analysis. It is extremely important, therefore, not to disturb the buffy coat after the first spin, and any pellet after the second spin.

Serum

1. Spin blood at 1200g for 10 minutes.

2. Aliquot 1 ml into labelled cryovials.

3. Place into liquid nitrogen Dewar or dry ice to snap freeze.

4. Transfer to -80°C freezer.

White Blood Cells

White blood cells can be used for DNA extraction and the creation of cell lines. (From EDTA / ACD tubes)

1. Transfer the remaining blood from the plasma spin to a labelled 50 ml 15 tube containing 10 ml RPMI-1640.

2. After alcohol swabbing the lid of this tube, aliquot 3 ml Ficoll into each of 2 clearly labelled 15 ml tubes.

3. Carefully layer 9 ml diluted blood onto each tube of Ficoll. Treat gently, do not mix, but spin as soon as possible.

4. Spin at 450g for 30 minutes.

5. Remove most of the top layer (RPMI-1640) using a 1 ml Eppendorf tip and discard (3-

4 ml) into waste container containing chlorine bleach.

6. Collect white blood cells with the same Eppendorf tip using a swirling motion to 'vacuum up' white blood cells. Do not take too much Ficoll (third layer), as it is toxic to the cells. Place the white blood cells in a labelled 15ml tube containing 10ml RPMI.

7. Spin at 450g for 10 minutes.

8. Pour off the supernatant into a waste container containing chlorine bleach. Add 3 ml of cold freezing mix (10% DMSO, 20% FCS, RPMI-1640) and resuspend.

9. Dispense the white blood cells into 3 x 1 ml labelled cryovials which have been sitting on ice.

10. Place on ice. Place vials in a rate-limiting freezer as to cryopreserve cells in conditions that maintain cell viability. This should be done as soon as possible as DMSO is toxic at room temperature.

11. Transfer on a weekly basis to liquid nitrogen tanks.

Buffy Coat

The buffy coat is a thin, greyish-white layer of white blood cells (leukocytes) and platelets covering the top of the packed red blood cells after 450g centrifugation (from EDTA / ACD tubes).

1. After having spun the blood, take buffy coat off with about 100 µl of plasma using a disposable sterile Pasteur pipette. NOTE: be careful not to lift red cells (if possible).

2. Aliquot as appropriate into labelled cryovials.

3. Place in liquid nitrogen Dewar to snap freeze.

4. Transfer to liquid nitrogen freezer.

Whole Blood

(To be prepared from EDTA tubes).

1. Dispense 50 µl DMSO into two 1 ml sterile cryovials.

2. Invert EDTA tube twice then add 450 µl of blood to each cryovial.

3. Invert cryovial to mix the whole blood with the DMSO. Note: DMSO is cytotoxic at room temperature, therefore as soon as it is mixed with blood it should be placed in a rate limiting freezer filled up with 200 ml of isopropanol.

4. Transfer to -80°C after at least 4 hours [55].

Serum and Plasma for proteomics

As proteomics is an emerging field, there is a lack of concrete evidence as to the best way to collect and process serum and plasma. There are also many different factors that can affect proteomic profiles, and so researchers are often reluctant to focus on just one factor (e.g. storage). However, the conclusions from several recent studies are presented here:

► Collection Tube Type

- The choice of tube in which to collect blood depends on the expected downstream uses of the serum / plasma.

- Platelets are most stable in Sodium Citrate tubes, and therefore these tubes may be suitable if there will be a long elapse of time between collection and processing (i.e. greater than 2 hrs). However, as these tubes contain liquid, they dilute the blood sample, which can act to lower immunoassay measurements.
- Samples in Lithium Heparin are also reasonably stable, but Heparin binds with a significant number of proteins. This can result in interference with some affinity processes, as Heparin can compete for or prevent binding of molecules to charged surfaces.
- Blood samples in EDTA tubes are slightly less stable than those in other tube types, and so must be processed quicker, however there is evidence that EDTA may act to inhibit the breakdown of proteins by proteases. EDTA tubes are not recommended for assays where metal ions are necessary as it binds with them [56].

► Processing Considerations

- Serum that has been collected in clotting activator tubes should be processed 30-60 minutes after venipuncture to minimise the effect of coagulation events.
- Plasma and serum in other tubes should be processed within two hours of the blood draw to allow analysis of the maximum profile of proteins in the blood sample.
- To ensure that platelet contamination is minimized, be careful to not centrifuge blood at too low a speed, and also when removing platelets from the blood sample. Ensure platelets are removed before freezing.
- Serum and plasma should be snap-frozen and stored at -80°C or below. As with any bio-specimen, it is best to minimize the number of freeze-thaw cycles.
- Make sure all variables are tracked in your database, e.g. tube type, storage temperature, time to freezing [55].

5.2.4.2 Processing of solid tissue specimens

Snap-freezing

Tissues must be snap-frozen within 30 minutes after removal from the subject. If this is not possible, the specimen should be put on ice until dissection. The delay time should be recorded on the tissue bank database. Freezing by direct immersion in liquid nitrogen is potentially damaging to the tissue and can hinder subsequent microscopic examination, especially when micro dissection techniques are required. For liquid nitrogen snap freezing, the protocol is:

1. Place the tissue sample into a clearly labeled plastic disposable tube or histology cassette (note: glass or pop-top vials should not be used as they may break or pop open).

2. Optimal size for snap freezing permeability is 0.5 cm^3, although smaller fragments should still be frozen. If there is sufficient material, freeze multiple samples.

3. When using a cassette, the cassette is wrapped in foil and a label is placed on the outside of the foil parcel. This parcel is then put into a rapid freezing medium (e.g. isopentane) in a liquid nitrogen Dewar. Slow freezing by placing tissue in a refrigerator, chest freezer, or cryostat must be avoided, as it results in the formation of ice crystals

4. The tubes or cassettes are then transferred into liquid nitrogen tanks for long-term storage. When appropriate, specimens may be kept for two weeks in a -80°C freezer in case of diagnostic recall.

Optimum Cutting Temperature (OCT) Embedding

Embedding specimens in OCT medium before freezing is useful when good preservation of histological detail is required. OCT samples can also be used for DNA and RNA extraction.

1. Fill Dewar with liquid nitrogen.

2. Pour approximately 200-300 ml isopropanol into a plastic beaker.

3. Place the beaker into liquid nitrogen. Note: do not put beaker into liquid nitrogen before pouring in the isopropanol or the beaker will crack.

4. While isopropanol is chilling, prepare tissue.

5. Form a foil 'boat' approximately 12 mm in diameter. Place a label on the side of this 'boat'. Commercial plastic moulds can also be used.

6. If tissue is too big for the 'boat', divide as appropriate in a biohazard hood using a scalpel blade and a Petri dish.

7. Put two drops of OCT into the 'boat' and place tissue on top in correct orientation for cutting.

8. Carefully pour OCT on top of tissue so that it is well covered. Do not allow any air bubbles to form.

9. Label a 20 ml tube and also place a label on the inside of the tube.

10. When isopropanol is very cold, remove the beaker from the liquid nitrogen and hold the foil 'boat' in the isopropanol using forceps. Do not submerge.

11. The tissue block should freeze slowly from the outside in, obvious by the whitening of the OCT.

12. When completely white, place the OCT block inside the labelled tube and drop into the liquid nitrogen canister. Note: rather that storing the OCT block inside a tube, a snap-locked bag can be used instead as long as it is tightly sealed to prevent desiccation.

13. Store OCT blocks in a -80°C freezer.

14. Allow the isopropanol to warm up before pouring back into the recycling bottle.

Formalin fixation

Formalin fixation is standard practice in most routine histopathology laboratories. The following guidelines address specific issues related to preservation of formalin-fixed specimens in bio-banks

1. Tissue specimens should not be bigger than 1.5 x 1 x 0.5 cm.

2. Specimens will be fixed in fresh 10% neutral buffered formalin (NBF) for a minimum of 4 hours and a maximum of 48 hours, after which time they will be embedded in paraffin following conventional techniques.

3. All reagents should be DNAse, RNAse free (e.g. prepared using DEPC water)

4. Fixation media containing Picric acid (e.g. Bouin's) should be avoided as this compound interferes with subsequent PCR analysis of extracted nucleic acids.

5. Alcohol fixation may be used as an alternative to formalin. Tissue is placed into 70%

alcohol made with DEPC water for a minimum of 4 hours.

RNA later

1. Cut tissue to be less than 0.5 cm in at least one dimension

2. Submerge tissue in 5 volumes of RNAlater (e.g. a 0.5 g sample requires about 2.5 ml of RNAlater).

3. Resuspend pelleted cells in a small amount of PBS before adding 5-10 volumes of RNAlater.

4. Samples can be stored at 4°C for one month, at 25°C for one week or at -20°C indefinitely. Archive RNAlater-treated tissues at -20°C.

5. For RNA isolation, simply remove the tissue from RNAlater and treat it as though it was just harvested [55].

5.3 Facilities

An efficient repository has taken adequate measures to ensure safe storage of the materials, support and maintenance of the equipment required, and safety for the staff. The design of the facility should include sufficient space to accommodate not only the initial material but also future storage [53].

5.3.1 Heating, ventilation and air conditioning

For optimal life of the mechanical refrigeration equipment, repository temperatures should be maintained between 15-22 ºC. Sufficient heating capacity is critical to prevent the freezing of water in drain lines. Moreover, sufficient air conditioning is necessary to prevent excess load on the compressor systems of mechanical freezers and refrigerators that may result in excess wear and early failure. Excess humidity can lead to fungal growth, which can possible affect specimen integrity and may cause health problems for the staff. Compressor function can be affected if air conditioning does not provide enough space for air circulation. In spaces where liquid nitrogen and dry ice are used, adequate ventilation is also critical to ensure maintenance of sufficient oxygen levels. Similarly, when services are performed that generate potentially harmful vapors (e.g. formaldehyde), the ventilation system should ensure that personnel are protected [53].

5.3.2 Lighting

Some specimens are photosensitive and require special protection from the light. The general lighting of the facility should provide a safe working environment and allow staff to accurately identify and retrieve materials. In situations where task lighting is employed, care must be taken that the lighting method does not adversely influence the sample integrity and the storage conditions. Fluorescent lighting or another type of lighting that does not create a source of heat is recommended for use in task lighting near frozen materials.

Emergency lighting in cases of power loss is also very important for the safety of the staff [53].

5.3.3 Backup power

Given that all commercial power will be interrupted at some time, a back up power system is required to maintain specimens in constant temperatures. Computer systems and electronic systems should be protected by a UPS (uninterruptible power supply). A generator should have a fuel supply to run continuously for a minimum of 48 hours and preferably a minimum of 72 hours, with an ability to refill fuel storage supplies. An emergency plan for sources to replenish fuel supplies is necessary and should include lists of suppliers and back-up suppliers. The power generator system should be tested for automatic starting and power generation weekly and load tested monthly unless sensitive laboratory equipment is also supported by this generator. In this case load testing may place the laboratory equipment at risk and should be tested less frequently [53].

5.3.4 Access

Access should be restricted to authorized personnel only. Doors should be locked. Mechanical keys employed should be ones that cannot be readily duplicated. Magnetic locks which control and record entry should be placed at entry points to the repository. Written records of repository visitors should be maintained and archived, and visitors should wear badges and accompanied by staff during their visit [53].

5.3.5 Security systems

Repositories should establish an automated building access detection system that simultaneously governs facility access and records who enters and exits the facility. The system should accommodate changes to security codes and keys when individuals leave the organization.

A fire prevention system is also required. The most common type of fire suppression is a sprinkler system that sprays water upon activation. In cases that water is unsuitable for certain equipment and stored materials, other chemicals may be employed for fire suppression. Personnel must be trained to evacuate facility immediately upon activation to prevent asphyxiation [53].

5.3.6 Emergency preparedness

Emergency contact numbers should be posted in prominent locations in the repository and should be carried by staff at all times that are “on call”. “On call” staff will be able to respond to an emergency at the repository and follow a check list of activities. Where possible, emergencies should be simulated to ensure proper follow-through for the established emergency plan [53].

5.4 Quality assurance and Quality Control

Quality Assurance (QA) is an integrated system of management activities involving planning, implementation, documentation, assessment, and improvement to ensure that a process, or item, is of the type and quality needed for the project. Quality Control (QC) is the system of technical activities that measures the attributes and performance of a process, or item, against defined standards, to verify that the stated requirements are fully met.

Each repository should have a Quality Assurance Program/Quality Management System (QA/QMS) or adhere to the QA program of the organization with which the repository is associated. The program should describe the repository's commitment to its QA and QC programs and describe approaches for ensuring that the requirements of the QA and QC programs are met. Should it not be possible to have a formal Quality Assurance Program with dedicated staff, a program should be in place to review procedures and records to assess the efficacy and quality of repository operations. This review should be conducted at least on an annual basis [53].

5.5 Quality standards

A variety of systems have been devised to allow for confidence and reproducibility in repository practices. While each of the standards described below are resources for repositories, there are costs involved in the attainment of each and may not be appropriate for every repository.

5.5.1 Current Good Practices

Current Good Practices (cGP) are regulatory guidelines that should be interpreted by the repository to fit its particular circumstances. cGP may be preclinical (Good Laboratory Practice, or GLP), clinical (Good Clinical Practice or GCP) or manufacture (Good Manufacture Practice, or GMP). cGP may be more relevant to large corporate repositories, but academic and other small repositories may wish to aim toward cGMP guidelines to instill confidence in the implementation of its SOPs. Generally, these standards are interpreted as follows:

- The facility is in a secure, locked area with limited access.
- Personnel must be trained in all procedures and such training is documented.
- The facility is subject to internal QA audits and/site visits by external clients and agencies as appropriate. The agencies that would audit depend on the agencies having jurisdiction for the repository activities included in their oversight. These may vary by local, state, national or international regulations.
- Policies and procedures are documented in SOPs that are approved by appropriate personnel and changed or updated only under strict document control rules.
- Records are maintained with respect to the purchase of new equipment, maintenance and repair activities, as well as equipment disposal. Examples of information tracked may include but are not limited to the name and model number for the equipment, manufacturer name and contact information, dates of maintenance and repair, etc.
- Records should also be maintained for critical materials and reagents used by the repository. Examples of information tracked may include but not be limited to the item name, company from which the item was purchased, date of purchase, and expiration date.
- Deviation reports are produced for all events that fall outside SOPs.

5.5.2 ISO

ISO9001 was created through the International Organization for Standardization (ISO). ISO is a worldwide federation of national standards bodies with headquarters in Geneva, Switzerland. The organization was founded in 1946 to develop a common set of standards for manufacturing, trade and communications organizations [53].

5.5.3 ISO9001:2000 Requirements of Quality Management Systems

ISO9001:2000 is a system standard, not a product standard. Its primary purpose is to provide organizations with useful internationally recognized models for operating a quality management system. ISO9001:2000 specified requirements for a quality management system where an organization needs to demonstrate its ability to consistently provide products that meet customers' and applicable regulatory requirements. It aims to enhance customer satisfaction through the effective application of the system. ISO9001:2000 is the benchmark of all standards. It is a level of quality standardization that some repositories are working to implement. ISO is similar to cGMP but is more recognizable in international settings [53].

5.5.4 ISO/IEC 17025: General Requirements Testing & Calibration Laboratories

ISO/IEC 17025 provides general requirements for producers of reference materials including tests and/or calibrations, and sampling. ISO/IEC 17025 covers the use of standard methods, non-standard methods, and laboratory-developed methods. ISO/IEC 17025 incorporates key requirements of ISO9001:2000.

The ISO Guide 34 provides the general requirements that a reference material producer must demonstrate if they are to be recognized as competent to carry out the production of reference materials. ISO guide 34 references ISO/IEC 17025 as a normative document [53].

5.6 Data Collection and Specimen Annotation of Bio-specimens

Definitions and Ontologies

Another important issue is the semantics that forms the basis of a bio-bank's systems and databases: What are the accepted definitions and ontologies? Scientific terms and ontologies are often not universally accepted. What definitions and ontologies should be used? Issues of consistent semantics exist on several levels, whether among countries, institutions, individual banks within an organization, or even individual researchers. The semantics issue even extends into the protocols used during sample collection, processing, and storage.

In addition to less-than-complete agreement on ontologies within both medicine and the supporting biological scientific domains, we must consider the element of time. As scientific knowledge increases, ontologies and definitions also evolve. This suggests that bio-banks need a periodic process for reviewing and updating the ontologies on which classification systems rest.

Related to, and yet distinct from, issues of consistency around definitions and ontologies is the maintenance of semantic consistency across multiple languages. While this may be a minor issue with regard to bio-banks established within a particular country with a focused mandate to support researchers and clinicians from that same nation, it becomes more complicated when we contemplate inter-bio-bank cooperation and collaboration. In all likelihood, those collaborations will cross national and linguistic barriers. Although simply defaulting to English as the scientific lingua franca may be pragmatic, it may not be a particularly satisfying resolution for local primary users whose first language is not English [2].

The Impact of Consistent Ontologies

In the section of this study addressing information technology issues, we articulate the impact of consistent ontologies on IT issues such as data integration, data querying, and data mining efforts. Some contributors to this document point out that without such standards and consistency, the sharing of data that is a pre-requisite for bio-banking to make its fullest contribution on a global scale is not possible [2].

Tissue and patient information database

Following are some general guidelines on basic bio-repository database requirements:

- Ideally, a bio-repository database should ensure that data is:

- Maintained in a standard format,
- Available over a long time,
- Able to be distributed to others as needed,
- Collected from satellite sites and combined.

- A bio-repository database must be able to:

- Provide all samples with a unique identifier generated by a standard coding mechanism. This identifier should be assigned at the time of collection, specific clinical and epidemiological data should be identified with the same number, and this number should be used to track a bio-specimen from collection through processing, storage, and distribution.
- Define a minimum datasets which an acceptable file must contain.
- Keep track of consent (and withdrawal of consent).
- Track original and current quantities of bio-specimens, as well as record when samples are exhausted.
- Record sample location and bio-specimen movements within or out of the bio-repository.
- Provide reports and audit information.
- Monitor quality control.
- Store data securely and authenticate users to protect against un-authorized access to the data.
- Have tiered levels of access so that users can only perform those operations for which they have permission.

Annotations on patients/subjects

Information that should be included is: Local patient case code, specimen topography and morphology according to the International Classification of Disease (ICD), TNM staging and tumor grade (in case of tumors), age at time of specimen collection, gender, place of residence, ethnicity, medical history, familial history, involvement into clinical trial, treatment and response to therapy, laboratory data.

Annotation on stored specimen

Information that should be included is: Local or (inter)national repository inventory code, tissue condition, preservation protocol, time elapsed between tissue removal and fixation/freezing, date of collection/storage and record of storage incidents, history of freezing/thawing, amount of tissue collected, and amount leftover in storage [54].

5.7 Bio-specimen Storage

5.7.1 Storage Techniques

The following general practices apply to all types of specimens, such as wet tissue, frozen tissue, paraffin-embedded tissue, glass slides, blood, serum, and urine. Individual types of specimens are handled according the type and the biomolecule analysis (RNA, DNA, protein).

Standardized protocols are applied consistently in preparing and storing specimens to ensure their quality and to avoid introducing variables into research studies.

Bio-specimens are stored in a stabilized state. Unnecessary thawing and refreezing of frozen samples are avoided, and appropriate size aliquots are used. A validated protocol of

thawing/refreezing is followed in cases that thawing/refreezing is necessary. Methods, such as inventory tracking, are established to minimize disruption of sample retrieval.

Storage temperature depends on specimen type, anticipated length of storage, bio-molecules of interest, and requirements of preserving viability. Paraffin blocks are stored at below 27 ºC in an area controlled for pests and humidity. In the case of liquids, components should be separated before storage, to preserve each constituent. Whole blood cryo-preservation is considered efficient and cost-effective for processing viable cells in large-scale studies. For possible future uses, tissues are stored in the vapour phase of liquid nitrogen freezers to ensure long-term viability. Lower storage temperatures and cryo-preservants (e.g. DMSO) are efficient in maintain cell viability for extended periods of time. The temperature at the top of the liquid nitrogen is consistently below -140 ºC.

Storage vessels are stable under planned storage conditions. Vial size and number are suitable for typical aliquots, anticipated investigators uses, and number of investigators. Volume and type of containers prevent sample loss. Screw-cap cryo-vials are used for long-term, low-temperature storage. Glass vials or vials with popup tops are unsuitable for long-term storage. Snap-frozen specimens are wrapped in aluminium foil or placed in commercial storage containers [57].

5.7.2 Storage equipment

The variety of environmental storage systems available for specimen collections continues to expand as technologies advance. Environmental storage equipment selections should be based on the type of specimens to be stored, the anticipated length of time the specimens will be stored, and on the intended use for the specimens. Also important are the size and physical design of the repository and the number of specimens stored. Some freezers and refrigerators now provide automated sample entry and retrieval components which may reduce long term costs for the repository. Often these larger systems are accompanied by increased initial costs which may be more than smaller repositories are capable of supporting. Equipment selections must take into consideration staffing requirements, quality issues and equipment support and maintenance [53].

Liquid Nitrogen Freezers

The use of liquid nitrogen (LN_2) freezers for long-term specimen preservation is optimal for the storage of some types of biological material, provided that the critical temperature for storage of those materials is not exceeded. The critical temperature for the storage of cells is called the glass transition temperature (T*g*) and is generally considered to be -140°C or below. This is the temperature at which all molecular activity is thought to cease. Some newer LN_2 freezers are able to maintain temperatures of <-186°C for extended periods of time, even with the lid of the freezer open. Many older models, however, cannot consistently maintain -140°C at the top of the tank and opening the freezer may result in temperature increases, especially in the upper region of the freezer. Care must be taken to minimize the number of times a freezer is opened within a given time frame.

Vapor or Liquid Storage

In general, vapor phase storage is preferred over storage in the liquid phase of nitrogen because the vapor phase provides sufficiently low temperatures to maintain temperatures below the T*g.* Storage in the vapor phase also avoids the safety hazards inherent in liquid phase storage. Most commercially available vials are penetrable by liquid nitrogen so vials selected for storage should be tested before they are used.

Note that storage in either vapor or liquid nitrogen carries specific requirements for freezer design that must be considered when the decision for vapor *vs.* liquid nitrogen is made.

Storage Containers

Liquid nitrogen expands to 700 to 800 times its original volume when brought to a gaseous phase at room temperature. This situation may produce an explosion hazard. Plastic and glass containers can easily explode if liquid nitrogen is trapped inside the container when it is removed from the freezer. Any container that has been stored in the liquid phase should be allowed to equilibrate in the gaseous phase of the freezer prior to removing it from the freezer.

Liquid Nitrogen Supply

Where liquid nitrogen (LN_2) refrigeration is employed, an adequate supply of refrigerant must be maintained. For freezers filled from Dewars or supply tanks, a minimum three-day supply of LN_2 at normal usage and replenishment intervals should be maintained, with the assumption that a re-supply is readily available. Bulk supply systems should maintain a minimum supply of 20% of the bulk tank capacity, or greater than three days working capacity. Bulk supplies should be checked for re-supply at least once a week.

Bulk storage and piping systems require relief valves to prevent rupturing of the pipe and bulk tanks in the event of over-pressure. If relief valves trip unexpectedly, a person near a valve can be sprayed with either the cold gas or the liquid. More likely, in the event of a blockage or excessive pressure, a number of relief valves may vent nearly simultaneously. This can cause a “white-out” condition in a matter of a few seconds. Visibility can drop to near zero and oxygen levels in the area may become less than that necessary to sustain life. Under these circumstances personnel should evacuate immediately. This unlikely event, which is usually caused by an error during the filling of the bulk tank, can be mitigated by well-designed procedures and practices.

Daily liquid nitrogen usage should be recorded either by monitoring the display levels or by manual means. Excessive liquid nitrogen usage can indicate problems with the vacuum component of the freezer. Self Contained Breathing Apparatus (SCBAs or “air packs”) should be available for use in the event of a “white out” condition. If a SCBA or other respiratory protection gear is used, compliance with regulatory standards may be required.

Oxygen Sensors

Because nitrogen displaces oxygen, care must be taken when LN_2 freezers are employed. The risk is inversely correlated with the size of the room. Oxygen level sensors should always be employed when LN_2 freezers are used in a repository. Both installed and mobile/personal monitors may be appropriate depending on the size of the facility. Even when installed units indicate an alarm condition, it may be useful to employ a personal monitor to enter the room carefully to validate the alarm condition if the area is not visible from the outside. Mobile oxygen monitors may be the best to use in a secure area where liquid nitrogen freezers operate because the sensors in installed units will degrade over time and sound false alarms.

Mechanical Freezers

Mechanical freezers are employed in a variety of storage temperature ranges, including -20, -40, -70 to -80°C, and occasionally as low as -140°C. Mechanical freezers come in a wide variety of sizes, configurations, and electric voltages.

The length of time that results in the significant warming of the stored material will vary by the properties of the stored material, the temperature of the material stored in the freezer (thermal loading) the ambient conditions and the design and maintenance of the unit. It is incumbent on the facility operator to establish the critical temperatures and response times to alarms.

Some mechanical freezers are equipped with the capacity to automatically cool their contents with either liquid nitrogen or liquid carbon dioxide as emergency back-up systems in the event of power loss for an extended period of time. Any freezer implementing this type of emergency back-up cooling system must be specifically designed to accommodate whatever coolants are utilized.

Refrigerators

Refrigerators are commonly employed where the life of the material being stored is enhanced by storage below ambient temperature. This is the preferred storage medium when the material must be kept cool, but does not require freezing. It is important to ensure that the temperature is maintained within the specified operating range, not just below a maximum temperature. Some high value materials, vaccines for example, must be maintained precisely between 2°C and 8°C. The facility operator must insure that high and low set points are monitored, and that alarm response time is adequate to prevent excessive temperature fluctuation

Walk-in Environmental Storage Systems

- **Compressors**

For the storage of valuable materials, walk-in refrigerators and freezers should be equipped with dual compressors that operate under an electrical alternating control system.

- **Door Release**

In most countries building codes require that walk-in units have internal safety releases to prevent a person from being trapped within a unit by the accidental closing of doors (*i.e.,* interior door release mechanism).

- **Floor Covering**

Refrigerators can generate slipping and falling hazards if water condenses on the floor. Freezers can occasionally create ice on the floor. Both types of units should have some type of mat or grate to prevent slipping.

- **Dry Ice**

Walk-in freezers should be kept free of dry ice *(i.e.,* the solid phase of CO_2). Carbon dioxide can rapidly build-up, displace the oxygen in the room, and cause personnel working in the units to lose consciousness. In confined areas the carbon dioxide can displace oxygen, presenting an asphyxiation hazard. Where dry ice is employed, engineering controls are required to insure sufficient air or oxygen level monitoring.

- **Motion Detection Devices**

Because of the special hazards involved in personnel working in a -20°C environment, it is desirable that some form of monitoring system be employed. This is especially applicable if only one person is working in the freezers. Systems which detect and alarm when motion does not occur are readily available (such systems are commonly employed by fire-fighters and other emergency personnel) [53].

5.7.3 Freezer Maintenance and Backup

A system for maintenance and repair of storage equipment, supporting systems, and facilities should be in place. System maintenance should be performed at regular, established intervals per manufacturer's recommendation.

Calibration

A system for the calibration of all instruments should be in place. Calibration should be done annually or per manufacturer's recommendation.

Calibration records should include the appropriate standard readings taken both before and after calibration.

A log of calibration records should be kept that includes the date of the calibration, the name of the individual performing the calibration, and a reference to the Standard Operating Procedure used to perform the calibration.

Equipment Preventative Maintenance and Repair

Essentially all equipment comprised of multiple components wears out with time and exposure to various environmental conditions. The duration of the lifetime for equipment used in the repository may be extended by performing routine assessments and modifications to the equipment according to the manufacturer's specifications. For mechanical freezers this may include a periodic changing out of fluids, cleaning of filters, calibration of probes, or manually removing ice from the tops and sides of the interior chamber of the freezer. Routine maintenance recommendations should be determined before a piece of equipment is put into service.

Resources for equipment repair should be identified when the repository is being established before an emergency is experienced. These resources should be reviewed on an annual basis.

Repositories should maintain spare parts for critical equipment (*e.g.,* spare compressors for mechanical freezers or refrigerators) especially for aging equipment that may not be readily available.

Repair *vs.* Replacement

Repositories should plan for the orderly replacement of equipment. If multiple pieces of the same equipment need replacement at one time, it might be best to use interim equipment or back-up equipment while introducing the new equipment in over time.

Back up

Bio-banks should have back up freezers available at all times. The total amount of back-up storage required for large repositories must be determined empirically, but will typically be 1.5% to 3% of the total freezer capacity for liquid nitrogen storage and will be 10% for mechanical freezer storage.

Repositories should have a written procedure for transferring samples from a failed unit (one that has exceeded or is on the verge of exceeding its acceptable operating temperature range) and for the return of the samples to their original location once it is considered safe to do so [53].

5.8 Specimen Distribution

Sample access and distribution are important topics for many bio-banks. Issues include: Who has access to the samples in a bio-bank? Is access limited to particular sets of samples? How are the samples distributed? How are the distributed samples tracked, and how is feedback obtained? How is the customer relationship managed? Several issues around sample access and distribution are critical to the overall security and safety of the samples, data, and personal privacy.

5.8.1 Review and Prioritization of Requests for Tissue

Access to tissue specimens should be regulated by a Bio-specimen Utilization Review Committee [4]. Members of this committee would be recruited from the research community, academia, advocacy groups, industry and possibly government. A peer review process is necessary to guarantee fair prioritization and equitable access. The investigators should justify the number and the type of specimens required for the specific proposed study.

5.8.2 Packaging and shipping

Packaging and shipping should conform to all governing regulations. Air shipments should conform to International Air Transport Association (IATA) standards. Ground shipments should conform to applicable federal standards. All personnel involved shipping dangerous goods (including infectious materials) should be trained properly for both air and ground shipments.

The first step in the preparation of a shipment for transport is the determination of the specifications for the specimens that are travelling. The shipper should determine what regulatory requirements are to be met as well as the physical requirements necessary to ensure proper shipping conditions.

Temperature Requirements

Specimens may be exposed to temperature fluctuations during transit. The following are typical temperature conditions required for transport of specimens and the insulation/refrigerant helpful to maintain that temperature:

- Ambient (20˚C to 30˚C) - insulated packaging to protect from extreme heat/cold ambient conditions.
- Refrigerated (2˚C to 8˚C) - wet ice, or gel packs designed for refrigerated temperatures, conditioned at –15˚C or phase change material rated for refrigerated transport.
- Frozen (-20˚C) - gel packs designed for frozen temperatures, conditioned at or below –20˚C.
- Frozen (-70˚C) - dry ice pellets, dry ice blocks, or dry ice sheets. Note that dry ice (solid carbon dioxide) employed for frozen shipments is considered a hazardous materials and appropriate labelling must be included.
- Frozen (at or below –150˚C) - liquid nitrogen dry shipper. Dry nitrogen shippers are insulated containers that contain refrigerated liquid nitrogen that is fully absorbed in a porous material and is therefore considered a non-dangerous product and is not subject to IATA regulations as a dangerous good.

Shipments of material that are subject to cold chain management should be shipped with sufficient refrigerant to maintain temperature throughout the shipping cycle, with allowance for at least a 24-hour delay in arrival time.

Humidity Requirements

Specimens sensitive to humid conditions may need to be shipped in sealed bags with desiccant to prevent exposure to moisture during transit.

Light Sensitivity Requirements

Light sensitive material should be sent in packaging that does not allow penetration of light such as amber vials or amber coated bags.

Arrival Time Requirements

Time sensitive specimens such as fresh whole blood should be consigned to couriers with a proven reputation of successful on-time delivery. Time required for shipment processing should be considered as well. Shipments should be initiated when there are at least two working days left in the week, in case it does not arrive on the day it is scheduled for delivery.

Sample Quantities

The quantity of specimens to be transported will affect the type of packaging and amount of refrigerant required to maintain appropriate temperatures for all specimens in the shipment. The container size should be appropriate for the amount of refrigerant needed and for the number of specimens that will be included in the container. Avoid sending an excessive amount of specimens in a single container.

Other Packaging Considerations

- Specimens should be positioned between the refrigerants used, rather than to be placed on top of or underneath the refrigerant.
- Empty spaces in the container present after the specimens and the refrigerant have been loaded should be filled with waded paper to prevent movement of the specimens during shipment.
- Remove or mark through any labels remaining from a previous shipment.
- Airbills should not be reused.

Regulations

Infectious substances fall into two categories: *Category A* comprises substances which are capable of posing permanent disability, life-threatening, or fatal disease to humans or animals, when exposed to them. Examples of these are highly pathogenic viruses (Ebola, Marburg, Lassa etc). The proper shipping name for such substances is UN2814 (infectious substances affecting humans) or UN2900 (infectious substances affecting animals). *Category B* comprises substances which do not meet the above criteria. Most human specimens such as blood, tissues etc fall into this category. The proper shipping name for such substances is UN3373 (Diagnostic specimens or clinical specimens) [54].

Special permits or other requirements may be unique to certain countries and regions. Some countries have regulations related to ethical issues which prohibit the import/export of certain types of human specimens or have specific requirements concerning the import/export of such specimens. If collecting non-human biological samples that are endangered or protected, special permits such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) permit, as well as additional paperwork may be required.

IATA Dangerous goods regulations (DGR) (2007 edition) and Infectious substances and Shipping Guidelines (www.iata.org) give instructions to ship infectious substances, biological specimens and patient specimens by any mode of transport, anywhere in the world. The

Technical Instructions for the Safe Transport of Dangerous Goods by Air published by the International Civil Aviation Organization (ICAO) are the legally binding international regulations.

The following links refer to these regulations:

- UNECE (United Nations Economic Commission for Europe) *UN Recommendations on the Transport of Dangerous Goods. Model Regulations.*

http://www.unece.org/trans/danger/publi/unrec/rev13/13files_e.html

- IATA (International Air Transport Association)

http://www.iata.org/ps/publications/9065.htm

- ICAO (International Civil Aviation Organization)

http://www.icao.int/icao/en/m_publications.html

- WHO (World Health Organization)

Transport of infectious substances 2005

http://www.who.int/csr/resources/publications/biosafety/WHO_CDS_CSR_LYO_2005_22r%20.pdf

Identify all requirements for shipping to a designated country prior to the initiation of the shipment. Customs clearance note must be displayed.

Due to possible delays in completing Customs requirements, temperature sensitive material may be consigned with a courier capable of replenishing refrigerant in the event of a delay [53].

Paperwork

Before sending bio-specimens, the recipient needs to be notified and a material transfer agreement received. Both the sender and recipient need a log so that they can track packages [55].

5.8.3 Bio-specimen Tracking

Both the sender and recipient should track all packages while in transit.

The recipient should confirm that they are able to receive the package and that they have the proper facilities for storage before the sender releases the shipment. The sender should send a shipping manifest (electronic) to the recipient prior to the release of the shipment. Confirmation of receipt and the condition upon arrival should be obtained for every shipment [53].

5.8.4 Ensuring Responsible Use of Resources

In general, whenever a researcher obtains an identifiable specimen from a tissue bank, this is human subjects' research and requires an IRB review, an informed consent or a waiver of informed consent.

There are four general approaches to review and prioritization systems for tissue requests at bio-banks: 1) first come, first served; 2) priority to members of the network, collaborators, and/or contributors to the repository; 3) prioritization based on merit review of research proposals; and 4) prioritization based on a set policy of the repository. The use of a tissue utilization committee to prioritize tissue distribution based on merit review of all research

proposals using standardized criteria and to ensure equitable distribution is a best practice [17]. Committees are usually made up of physicians, administrators and researchers.

In order to create a framework for sharing and comparing research results, it is recommended that validated, investigator-derived data be returned to the NBN and linked back to original tissue samples. Encouraging the entry of validated data in a standardized format back into the NBN system would create a rich, valuable and unique resource available to all investigators, providing new possibilities for the corroboration of research findings across methodologies; the emergence of successful new therapeutic/preventive targets; and the acceleration of important scientific breakthroughs. Researchers should be encouraged and invited to submit data and report all publications resulting from the use of NBN samples, as well as reference the source of the samples in their report. If DNA, RNA and proteomic data are published, the raw data must be available for public use, similar to publication standards set by major scientific journals [4].

6 Bioinformatics and Data Management

The backbone of any repository is a standardized, scalable, and secure bioinformatics system that is appropriate for repository management, tissue acquisition and management, and data aggregation and analysis. Bioinformatics systems are used for repository management, clinical and pathological data management, collection and analysis of research results, and data mining and advanced statistical analysis to identify patterns and establish relationships. A bioinformatics system that is searchable and minable via varying levels of Web-based access for different individuals—including repository personnel, researchers, patients, and the public—is a best practice. Robust network security systems and access control are crucial to ensure that the privacy of the tissue source is protected and that the bioinformatics system is secure.

Bioinformatics systems can range from simple databases to proprietary systems developed in house. Close ties between bioinformatics system developers, researchers, data managers, and repository management allow the bioinformatics system to be designed so that it is responsive to the needs of multiple user types.

The use of a standardized language to categorize and describe bio-specimens and enter data into the bioinformatics system is essential for comparison of bio-specimen characteristics among collection sites. In addition to using a standardized language, it is also important to use either a system that can automatically extract data from medical records or multiple checks of data entry to ensure the accuracy of the data in the bioinformatics system.

Efficient and accurate sample distribution is important to the bio-bank and its end users. Time is an often critical component of research; successful bio-banks must be able to distribute samples accurately and in a timely manner. A security component incorporated into the distribution system ensures that sample and data access will be granted to only approved researchers. The distribution system may implement a sophisticated sample query and data mining tool that allows end users to search the repository for specific specimens.

Information technology is as critical to the operations of a bio-bank as are the core tasks of collecting, storing, and distributing biological samples. Information technology is not an option and cannot be separated from these core tasks if a bio-bank is to be successful and secure. Information technology is at the juncture where a sample is given context. IT enables researchers to efficiently select research subjects from a bio-bank; it is vital to the subsequent measurement and testing that those researchers conduct on biological samples and ultimately makes possible statistical analysis of control variables (demographics and clinical variables), predictor variables (typically biological measurements carried out on the samples), and dependent variables (typically clinical outcomes).

We review a number of issues and open questions on IT from the following perspectives:

- Sample identification and tracking methods
- Data standards
- Data quality — QA/QC
- Database requirements and architecture
- Data integration
- Knowledge management
- Privacy and security
- Balancing compatibility and standardization against adaptability and growth [2].

The advantages of integrating databases in different aspects of bio-banks for health are manifold. The first requirement that has to be fulfilled to enable bio-bank communication is a unique identity for each bio-bank. Karolinska Institutet Bio-bank has proposed an identification method in Sweden, and it is now being implemented as a national standard. The same method can be used internationally, and the reactions from bio-banks outside Sweden have been very promising. The second requirement is a common nomenclature to enable communication between bio-banks.

The potential impact of integrating will be:

- Promote communication within and between major bio-banking initiatives, thereby helping to overcome existing fragmentation of population genomic research.
- Enhance the effective sharing and synthesis of information, thereby addressing the need for very large sample sizes and helping to promote collaborative international genetic, epidemiological, and clinical research.
- Avoid the expensive mistakes and inefficiencies that can arise when individual initiatives repeatedly "re-invent the wheel," thereby saving funders and researchers a lot of time and money.

The key purpose of the population-based cohorts is to gather an extremely large quantity of information from a large sample of the population. Longitudinal research over a long period of time, hopefully generations, demands completely new methods and systems to handle the gathering of information and bio-bank information. These bio-banks bring to the fore the problems concerning the need for a standardization of research data and a safe computer and storage strategy.

The bio-bank will use freezers and liquid nitrogen tanks to store serum, tissues, or extracted DNA. A laboratory information management system (LIMS) will handle the management of the storage issues. However, one of the key goals of the Bio-banks is to create a link between the actual samples and large databases containing phenotype and genotype information of the individual sample donors. The middleware system responsible for this functionality has been attributed as a bio-bank information management system (BIMS) [58]. The project contains much work concerning defining, structuring, and standardizing the information that has been gathered within the clinical and epidemiological fields.

The project will significantly improve the possibility of using collected information in a cross-disciplinary fashion and in completely different contexts than the original.

The BIMS will be a middleware system that will handle communication between several other systems. Security will be a prime issue when designing BIMS, and careful thought will be given to issues such as data encryption and access control. Personal integrity for sample donors will be of greatest importance. The system will be responsible for a range of tasks, including:

Middleware functionality. BIMS will provide functionality for connecting a range of software systems. In particular, it will function as a link between the LIMS used for sample handling and large database systems used for storing epidemiological and clinical data. The communication between the different systems will mainly consist of eXtensible Markup Language (XML) streams. Depending on access privileges, BIMS will allow for powerful queries combining phenotype characteristics from large databases with laboratory and storage information from the LIMS.

Result storage. BIMS will allow researchers to store results in a controlled and secure environment, ensuring the quality and safety of the data. BIMS will provide functionality for storing both genotype and phenotype information collected during studies performed in cooperation with the bio-bank.

Donor consent handling. BIMS will be responsible for handling donor consent information. All donors have the right to change their minds regarding the use of their donations. In

addition, BIMS will provide functionality for complete removal of samples from the bio-bank upon requests from donors

Sample and information management. As described below, it will be possible to order biological samples as well as collected data from the bio-bank (or other bio-banks). However, a researcher or institution will own each sample and its corresponding data. Enabling an external user to access this information will involve BIMS functionality that provides secure routines for applying for and granting access.

BIMS will be accessed by an intuitive, secure Web interface, which will simplify the use of the system. Researchers who will mainly use the system fall into two categories, differing in their use of the software. The first class of users will be involved in collecting samples and data that will be stored within the bio-bank. These users will have full access (on the individual sample level) to their own data within BIMS, as well as the actual samples stored in the bio-bank freezers. This user group will use the system for a variety of tasks, including data manipulation, consent management, sample removal approvals, and data import/export. The second user class consists of researchers who want to utilize the information collected by other bio-bank users. This class will have significantly lower access privileges. However, some depersonalized search capabilities will be provided to allow the users to search for attractive samples.

Sample Identification and Tracking Methods

To manage sample identification and tracking, the first issue is deciding on an identification system. It would be advantageous to have all bio-banks worldwide ascribe to a single identification system. This would prevent overlap or confusion at research sites that may be receiving samples from multiple bio-banks. Today, each bio-bank decides on its own identification system for samples. Currently, it appears that bio-banks may agree to tie into a broader standard such as the new Generation 2 code supported by EPCglobal or wait for a new ISO-backed standard to emerge. Alternatively, bio-banks could agree to adopt a standard ID system from among their peers, such as the one developed at the Karolinska Institutet Bio-bank.

The second issue is deciding what information must travel physically with samples. At a minimum, it should contain the name or code identifying the bio-bank distributing the sample, an identification code that is unique within that center (if not universally unique), and some standard designator of the type of sample (e.g., blood). Additional pieces of information that might be included are date of collection, date of distribution from the center, details on sample type, and warnings or conditions of use.

The technological issue is how that tracking information should be encoded. Should the information be human-readable as text on a label? Presumably yes, likely in both the native language of the source bio-bank and in English, which is currently the lingua franca of science. In addition, it makes sense to decide on a machine-readable format to support automated inventory management and tracking systems. Barcodes are a mature, inexpensive technology, but they are somewhat limited in the volume of information that can be contained in the code. A newer technology, Radio Frequency Identification (RFID), is more expensive than barcode systems, but costs for tags and readers are dropping rapidly. RFID delivers some benefits over barcode systems, such as proximity reading rather than line-of-sight reading. However, RFID still has some readability issues when dealing with liquids and metals. In general, it makes some sense not only to start with data standards for labels that can be set and implemented in barcode technology but also to think ahead to ease a transition to RFID when that technology matures.

The other main issue in identification and tracking is arranging a look-up database for machine-readable codes. Each bio-bank should provide a look-up server for managing its own samples. If multi-bio-bank collaborations become the norm, and a single encompassing identification system is adopted, we would expect that replicate lookup servers would be set

up so that any sample ID could be entered and the look-up server could provide sample source, type, and dates as well as other common status information on the sample.

6.1 Use of Bioinformatics Systems

6.1.1 Clinical Trials

Clinical research enables doctors and researchers to find new and better ways to understand, detect, control and treat illness. A clinical research study is a way to find answers to difficult scientific or health questions.

The most commonly performed clinical trials evaluate new drugs, medical devices, biologics, or other interventions to patients in strictly scientifically controlled settings, and are required for regulatory authority approval of new therapies. Trials may be designed to assess the safety and efficacy of an experimental therapy, to assess whether the new intervention is better than standard therapy, or to compare the efficacy of two standard or marketed interventions. The trial objectives and design are usually documented in a Clinical trial protocol.

Bioinformatics systems are of high importance to Clinical Trials because:

- They facilitate the automation of many of the above procedures; enhance the efficiency and the optimal design of the study.
- They assist in the monitoring of the clinical trial over time and ensure that the required quality standards are met in all steps of a trial.
- They provide the technological mechanisms to enforce the legal and ethical policies, such as consent procedures and patient privacy.
- They provide tools to aggregate data from different fields or previous studies, correlate the findings, visualize or analyze the results.

6.1.2 Data Mining & Knowledge Discovery

A key-technology for addressing some of the challenges in biomedical informatics relates to *data mining* systems, methods and tools. Data mining is a step in the process of generating knowledge in databases. It includes techniques for query databases, on-line analytical processing, and machine-learning algorithms. In the medical area, many applications have been created for decision support to address issues such as image and signal analysis and outlining clinical prognoses for patient conditions. In biology, efforts have been targeted on research issues such as the prediction of protein structures and drug studies.
Both types of predictive exercises present considerable challenges for future research. Text mining is a discipline that aims to extract data, information or knowledge from texts. Finding information in biomedical databases using text mining and information-retrieval techniques is expected to leverage a substantial amount of biomedical information that has escaped analysis until now.

An indicative list with some of the key biomedical tasks to be tackled with data-mining is:

- Genome Database Mining
- Computational/Mining for Gene Discovery
- Sequence Similarity Searching
- Gene Expression Mining
- Proteomics and Data Mining
- Metabolomics and Data Mining

6.1.3 Semantic Information Integration

Database integration requires bridging the syntactic and semantic gaps existing across data sources. Given the suitability of ontologies to provide a semantic layer to applications, database integration is moving towards an ontology-based approach. This approach seems to be very promising, although there are still several issues that must be addressed.
In the biomedical domain, it has been demonstrated that ontologies can aid Bio-Medical integration, since they are mainly used to facilitate knowledge distribution, sharing and reuse. In order to apply ontologies to database integration, several systems use ontology-based views to facilitate the mapping from objects of specific databases to shared vocabularies. There are some less common approaches, such as the use of ontologies for automatic mediator generation. Database integration is evolving towards ontology-based approaches, where ontologies are used to support mapping between equivalent concepts for integration and query formulation, given that ontologies provide a common and shared vocabulary that can be used to facilitate the communication and information transportation between users, systems and databases.

6.1.4 In-Silico Modelling and Simulation

In the last decades considerable efforts have been made in order to mathematically simulate tumour growth and tumour and normal tissue response to various therapeutic schemes. Mathematical analysis and discrete mathematics (theory of algorithms, graph theory, cellular automata, finite state machines, etc.) along with probability theory have played central roles in this process. The ultimate goal of tumour and normal tissue simulation is to contribute to the optimisation of cancer treatment by fully exploiting the individual data of the patient. The vision is that by utilizing an “oncosimulator”, the medical doctor will be able to perform in silico (on the computer) experiments corresponding to different candidate therapeutic scenarios for any cancer patient in order to facilitate and better substantiate his or her treatment decisions. Therapeutic scenarios may refer to differing radiation fractionations, differing drug administration schedules etc.

It is clear from the above that in order to build such complex computational models and algorithms, sufficient data must be collected, analyzed, processed and evaluated. Also, enough input data from clinical trials and genetic researches must be accumulated for the validation of these models, so it is bio-banks and bio-repositories which can promote the research in this field. Furthermore, bio-banks store the information regarding the discrete characteristics of the patients, their clinical history and their genetic profile and assist the medical doctors into the individualized application of such models to the patients.

6.2 Types of Data Contained in Bioinformatics Systems

Database Requirements and Architecture

Databases for bio-banks must be able to effectively manage the following data types: numerical data; date and time; text data; images, ranging from microarray images to spectroscopy; and digitized forms of analog signals.

In addition to the basics of storing a variety of data types, ideal database tools and architecture will enable numerous types of analysis, ranging from simple queries and sorts, to multi-variable statistical analysis, to mining of text fields or automated analysis of images. It is not necessary for all those analytical tools to reside as functions in the database itself, but the database architecture and tools should enable easy extraction of data from those fields in forms that work with leading analytical tools. Further, complicating database requirements is the need to be able to coordinate textual analysis across multiple languages. Ideally, tools would allow a query or analysis to be executed in any of the major languages of the world, and the tool would automatically construct and execute a parallel query in the various languages used in bio-banks around the world.

Database management tools should control access and manage rights to various activities in the data set. It is also necessary for database tools to maintain logs that track changes to records and restrict the ability to change a record once it is confirmed or accepted.

From an architectural perspective, the traditional relational database management system (RDBMS) has a long history with a base of deep experience and knowledge among programmers and IT administrators. However, most current trends show specialized industries moving toward XML data structures and specialized XML vocabularies. RDBMSs and XML standards are not mutually exclusive. In fact, institutions maintaining databases that are open to public researchers often have a traditional RDBMS-type architecture for storage and back-end processing, but they use an XML-enabled front-end interface that enables automatic downloads and querying by researchers.

Data Integration

Data integration is important on multiple levels. At the intra-organizational level, integration is important for individual bio-banks to aggregate data from multiple collection sites. Across organizations lies an idealistic vision of integrating data from multiple bio-banks in different countries for a single study. Where there is strong agreement on ontologies, field naming conventions, data formats and definitions, as well as consistent adherence to SOPs, data integration becomes relatively straightforward. However, that description of consistency is still an idealistic vision of the future. As the situation exists today, data integration tools and appropriate methodologies are important to maximizing scientific value from the multiple data sets at bio-banks. In particular, newer-generation text mining tools may assist in mapping disparate ontologies. Moreover, participants in the World Wide Bio-bank Summit suggested, with regard to phenotype classification, that bio-banks should agree on a "minimum data set to be interchangeable between bio-banks and identify the complete ontology and multilingual definition for this data set."

Typical types of data stored or communicated over bioinformatics systems include **demographic data** of patients, **clinical data** of the individual persons and **genomic information** regarding the stored specimens. The information can be encoded using specific vocabularies, but data can also be stored in free text form, imaging data, collections of spreadsheet files, etc.

6.2.1 Demographic Data

In most bioinformatics databases, demographic data about the involved patients must be stored. These data can be used for the identification of patients, the grouping of patients to population groups with similar characteristics, the evaluation of clinical trials against specific population groups, etc. The identification of patients usually is highly connected with the security and authorization policies that are used to ensure the legal and ethical requirements for the usage of these types of systems. Typical types of demographic data include

- Patient identification data, social security numbers or pseudonymized tracking IDs, which provide the means to uniquely refer to a specific individual person. These often include name, ethnicity, ancestry information, etc.
- Gender
- Age or birth date
- Height and weight

6.2.2 Clinical Data

Information about the clinical history and treatment of each patient is of paramount importance for all kinds of biomedical studies and research activities. These data are necessary for the realization of clinical trials, the decision for suggested treatment and the evaluation of the therapies and the outcomes of each study.

Clinical data contained in bio-repository databases usually include information such as:

- Signs and symptoms
- Special diet and habits
- Clinical diagnosis
- Histologic diagnosis
- Pathologic status
- Tissue anatomic site
- Surgery data
- Medical history
- Family history
- Treatment history.

6.2.3 Genomic Data

The development of new technologies like DNA sequencing, gene expression analysis, genotyping, SNP mapping and pharmacogenomics have changed the concept of diagnostics and therapeutics. Thus, this genetic information will gradually included in databases and in the future it will probably contribute to the patient care. Examples of this information include:

- Gene expression signatures
- SNPs and risk associations
- Mutation screening
- Proteomic profiling
- Epigenetic markers
- Pharmacogenomic analysis

6.3 Data Integration Challenges

The amount of information generated by biomedical research has increased almost exponentially due to the use of modern information technology tools. Particularly in the field of epidemiology, a discipline tightly associated with the use of bio-banks, there is a growing need to manage the huge quantities of data collected from the study subjects themselves, their medical records, and their biological samples.

In describing the challenges of dealing with large amounts of data, we focus on the following aspects of data integration because these play a vital role in the management of medical data.

6.3.1 Quality and comparability of data

When dealing with data from several sources, it is important to first determine if it is possible to compare them. Data may have been collected by methods that are not consistent. It is possible that data variables which at first seem identical may not convey the same information at all. In addition, the quality of data varies as very few scientific data sources are regulated by a quality control system. Mistakes are often made in the data collection process. Data quality and comparability are not primarily technical issues, but IT can provide the mechanisms to detect errors and enforce quality control.

6.3.2 Differing data models

In different data sources, data may be structured according to different data models. These data models usually range from completely normalized relational data models to very simple spreadsheet-based models. The same type of information can be stored under different models, and when data is integrated, it is important to be able to handle these differences.

6.3.3 Differing ontologies

Data sources often make use of differing ontologies and taxonomies to represent the same information. This lack of standardization is a general problem that undermines the widespread use of bio-specimens. Mapping ontology A to ontology B is a challenging research problem. However, once the mapping exists, the ability to integrate data based on different ontologies has important practical uses.

6.3.4 De-identification

For a variety of reasons data integration often requires the de-identification of data. The process of de-identification (sometimes known as anonymization) generally involves the removal of information that can identify the person associated with a medical record. The de-identification process is sometimes reversible, in which case we most commonly refer to it as pseudonymization.

6.3.5 Differing data formats

Medical data is stored in a variety of formats, including spreadsheets, text files, data sets produced by statistical applications in proprietary formats, relational databases, images, various binary streams (such as instrument data), and Internet-based formats. All these data formats have to be accommodated by the information management system.

6.3.6 Ownership of data

When integrating data from sources under different ownership, it is important to be able to control access to data and track ownership. This is especially important for gaining the trust of data source owners and for promoting a spirit of collaboration. When integrating data from two data sources, the combined data set often contains more information than the two sources viewed separately from one another (a fact that is the primary driving force behind the integration of information). Thus, it is important to be able to control authorization (i.e., access to data and services) at a very detailed level.

6.3.7 Legal and ethical aspects

Integration of data from multiple sources requires that we comply with the legal and ethical requirements, such as complying with the wishes of the study participants as expressed in the consent agreements. The main challenge is to comply with these requirements and, at the same time, make the data as useful as possible to the research community.

6.4 Data Accessibility

The primary reason for the creation of bio-banks is to make the information of the biological material more readily available to scientists for their research projects. Therefore, in order for the researchers to gain access to this network, facilities must be maintained to allow them to interrogate the entire central database or information system.

6.4.1 Online searches

Most existing databases usually offer access through web interfaces to authorized users. Often, the information provided through web is publicly available as long as it doesn't reveal sensitive data. Information can be presented in an aggregated form, such as graphs or other visualization techniques, or in a fine-grained level.

A user who searches online is usually given two primary options to access the information, either by using text matching queries or traversing through predefined hierarchical categories.

- **Text matching queries** - The queries are formed using string matching expressions combined together with simple Boolean functions. With this kind of search, the user must be well accustomed with the specific domain in order to use appropriate searching terms and keywords and to form search expressions that will ultimately return the aggregated information which he was aiming at. It is a powerful technique that gives access to the information in great detail combined with the easiness of the web-access. Nevertheless, the returned data usually is unorganized free text and, since it cannot be easily combined with other tools or visualization techniques, it is difficult to extract useful information out of the abundance of "noise".
- **Hierarchical categories traversal** – The user usually locates information by selecting specific sub-categories of more general groups of data, or by selecting categories which correspond to predefined queries. This type of search can be used easily by none experts, as the user doesn't have to form the query expressions himself, but lacks flexibility as the user cannot easily append other types of categories.

6.4.2 Direct Database Access

A usual way of accessing the stored data of a database is by allowing direct connection to users and, most commonly, to applications and tools. Users or programs provide some type of credentials to prove their identity and claim authorization of accessing the data. This kind of connection is the most powerful type of access since the user has the ability to query directly to the source of data and combine information in whatever way he/she likes. Most importantly, direct access to a database permits the implementation of software tools that can handle great amounts of data, transform it to other formats thus enabling "chains" of tools to act on data, and visualize it in a flexible manner.

This power of the direct access to database is also a drawback, since unauthorized, malicious or naive users can harm the system or be permitted access to data which they shouldn't be. For this reason, direct access to databases must be accompanied by strict security protocols and mechanisms, to guarantee the integrity and safety of the stored information.

6.4.3 Data Wrappers

Many times, due to various reasons such as legacy or proprietary systems usage, data are provided in a specific format but cannot be used directly without transformation to another format. In such situations, data wrappers can be used to apply the needed transformations. In addition, data wrappers can be used to combine data from several sources, or to handle the user queries and transform them to another form without the need to alter data, or without revealing hidden information. Also, this mediation layer can be utilized to encapsulate knowledge of the specific domain of expertise (business logic) and decouple the data sources from the end users, visualization tools or further analysis.

6.4.4 Federated Databases

A federated database is a logical association of independent databases that provides a single, integrated, coherent view of all resources in the federation [66]. The federation architecture makes several distinct physical databases appear as one logical database to end-users. Federations also provide a cohesive, unified view of data derived from multiple sources. The data sources for federated systems can include databases (relational or object-based), flat files, text documents, spreadsheets and various other forms of structured and

unstructured data. Moreover, the federation can be largely vendor neutral since databases from almost any vendor can be supported.

Remote databases (federates) in the system expose a subset of their resources to the federation. These resources can include metadata (database schemas), raw data, summary data and utilities for management of remote data and API's (application programming interfaces) for direct manipulation of remote data. The union of these exposed resources, in conjunction with a central integration database, constitutes the federation infrastructure. And federates respond to queries from other federation members while remaining largely autonomous from them.

Some key features of federated systems include:

- Autonomous data sources
- Heterogeneity of data sources
- Data sources often geographically distributed
- Data sources controlled by independent administrative domains
- Logical integration of distributed datasets
- Coherent, unified, and integrated view of data from multiple resources
- Largely vendor neutral

The caBIG project [65] is an example of a bioinformatics system designed around federation architectures to make all resources readily available to anyone plugged into the caBIG informatics grid. For example, bioinformatics tools and databases are currently available for clinical trial management systems, integrated cancer research, tissue banks and pathology tools, controlled vocabularies, genomic and proteomic databases, biochemical pathway analysis, and image repositories. By joining the grid, researchers can decide which applications, tools and databases fit their needs, and they can access these resources from anywhere. The underlying federation architecture enables a high level of integration among these resources.

6.5 Bioinformatics System Security

Data integration is important on multiple levels. At the intra-organizational level, integration is important for individual bio-banks to aggregate data from multiple collection sites. Across organizations lies an idealistic vision of integrating data from multiple bio-banks in different countries for a single study. Where there is strong agreement on ontologies, field naming conventions, data formats and definitions, as well as consistent adherence to SOPs, data integration becomes relatively straightforward. However, that description of consistency is still an idealistic vision of the future. As the situation exists today, data integration tools and appropriate methodologies are important to maximizing scientific value from the multiple data sets at bio-banks. In particular, newer-generation text mining tools may assist in mapping disparate ontologies. Moreover, participants in the World Wide Bio-bank Summit suggested, with regard to phenotype classification, that bio-banks should agree on a "minimum data set to be interchangeable between bio-banks and identify the complete ontology and multilingual definition for this data set."

The expectation and assurance of privacy are critical to gaining the cooperation of patients/subjects to donate samples and data in the first place. Security breaches and compromised patient privacy could significantly hinder the ability of the offending bio-bank to continue collecting new samples and data. In addition, such an incident could also harm, at least for a time, the effectiveness of all bio-banks. Because most, if not all, critical data on the patient contributors is stored in electronic form, the role of IT is particularly critical to securing patient privacy.

Electronic security starts with well-understood processes that articulate who needs access to what information at which points in the process. Those processes and the accompanying policies must then be incorporated in electronic access controls. The tightest controls (i.e., most stringent authentication process) must be for those persons who would have access to database records that allow them to tie together personal identification information (i.e., name, address, phone number) with specific biological samples and associated clinical data. Less restrictive access can be deployed once personal identification information is effectively removed.

Key elements of security include:

- Unique user identifiers
- Passwords
- Digital signatures
- Encryption
- Physical security for server rooms/datacenters
- Network firewalls
- Network intrusion detection
- Staying current (what is secure today may be relatively easy to hack in six months)
- Periodic security training for key users/administrators who have the broadest access

One of the challenges faced by every organization that makes electronic security a priority is how to implement security systems and processes without making the system onerous for administrators and users; ultimately, one still wants the data and tools to be relatively "easy to use."

Depending on the various mechanisms and the procedures in use of each bioinformatics system, different techniques apply to ensure the safety, integrity, authorized access and any other security concern over the stored information. Security must not be hindering the research purposes, as much as possible, but be intrinsically applied by the used protocols and procedures.

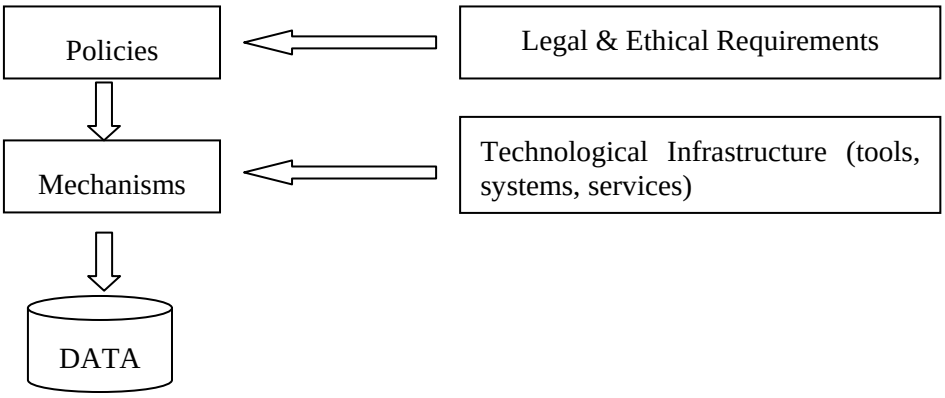

Figure 3 : Security Architecture

Based on the legal and ethical requirements, policies and protocols are developed and, depending on the available technological tools, various mechanisms may be utilized to implement, monitor, log and assess the conformance with these policies.

Below we quote some use cases drawn from large bioinformatics systems which have developed such policies, such as the UK Bio-bank [7] and The Cancer Genome ATLAS (TCGA) project [27]. The specific mechanisms vary and change to each individual case, as each project has different requirements.

6.5.1 The UK Bio-bank case

The processing and storage components of UK Bio-bank systems will be hosted in dedicated facilities. Strict controls over physical and logical access will be implemented which permit access only to authorised individuals.

Consideration will be given to resilience issues such as off-site backup and escrow facilities to facilitate the resumption of operations in the event of system failure or disaster. UK Bio-bank is currently developing a detailed Information Security Management System with external experts, working towards ISO 27001 compliance. This will put in place a set of controls, consisting primarily of policies and procedures, to manage.

6.5.1.1 Overall security

An information security governance structure that provides strategic direction and implements the high level processes for monitoring the success or failure of the underlying security processes. This is comparable to the high level PDCA (Plan, Do, Check, Act) processes implemented in a Quality Management System.

6.5.1.2 Organisational assets

Understanding what information assets are held, and managing their security appropriately. Policies and procedures will cover the classification of information, and its appropriate handling by UK Bio-bank, to ensure that sensitive data are not compromised.

6.5.1.3 Communications and operations

Security controls for systems and network management will ensure that IT systems are configured and used in a secure manner, mitigating against intrusion and failure. Control of logical access to IT systems, networks and data will prevent unauthorized use.

6.5.1.4 Human resources security

Access rights for staff, including acceptable usage policies and suitable security awareness and training activities.

6.5.1.5 Physical and environmental security

Protection of valuable IT systems against malicious or accidental damage, or loss through overheating or mains power failure. Use of equipment will need to be controlled and monitored in order to ensure that the data collected by, or stored on, this equipment are accurate and not compromised .

6.5.1.6 Systems development and maintenance

Taking information security into account in the processes for specifying, building/acquiring, testing and implementing IT systems.

6.5.1.7 Security incidents

Prompt reporting and proper management of information security events, incidents and weaknesses (including near misses) provides a key feedback mechanism for the monitoring and improvement of information security systems.

All policy and procedure documents will be integrated with the Quality Management System being developed by UK Bio-bank laboratory operations.

6.5.2 The Cancer Genome ATLAS case

The TCGA Pilot Project produces volumes of genomic information derived from human tumour specimens collected from patient populations, and also grants access to clinical information associated with these specimens. The aggregated data generated is unique to each individual and, despite the lack of any direct identifying information within the data, there is a risk of individual re-identification by bioinformatics methods or third-party databases. Due to patient privacy protection concerns, data access policies are being implemented to minimize the risk that the privacy of the donors and the confidentiality of their data will be compromised. As part of this effort, data generated from TCGA are available in two tiers.

6.5.2.1 Open-Access Data tier

A publicly accessible tier of data that cannot be aggregated to generate a dataset unique to an individual. The open-access data tier does not require user certification for data access.

The data types in this layer may include:

- Tissue pathology data,
- De-identified clinical data,
- Gene expression data,
- Copy-number alterations for non-genetic platforms,
- Epigenetic data,
- Data summaries, such as genotype frequencies, and
- DNA sequence data of single amplicons.

6.5.2.2 Controlled-Access Data tier

A controlled-access tier with clinical data and individually unique information. This tier requires user certification for data access and is available to researchers who agree to the terms of usage and conform to the applied restrictions.

Data within this tier include:

- Demographic and clinical data up to the level of detail permitted,
- Genome-wide genotypes, and

- Information linking all sequence traces to a single donor, whose associated data has been stripped of direct identifiers.

6.6 Quality Control, Auditing and Standardization for Bioinformatics Systems

The other main issue in identification and tracking is arranging a look-up database for machine-readable codes. Each bio-bank should provide a look-up server for managing its own samples. If multi-bio-bank collaborations become the norm, and a single encompassing identification system is adopted, we would expect that replicate lookup servers would be set up so that any sample ID could be entered and the look-up server could provide sample source, type, and dates as well as other common status information on the sample.

Data standards have numerous levels on which to achieve consistency; for example, agreeing on how height will be recorded (e.g., centimeters not inches, with or without shoes, to the nearest tenth of a centimeter) — and this is a simple example. A more complex example is deciding on a single naming convention for gene identification. Alternatively, what ontology will be used for classifying phenotypes? The 2004 World Wide Bio-bank Summit participants suggested that SNOMED and HL7 may be candidates for phenotype classifications. One quickly realizes that the heart of the matter is not really information technology; rather, it is the level of consensus in each medical and scientific domain that determines whether there is a generally accepted ontology on which to base data standards.

Each bio-bank could make its own evaluations and choose a preferred ontology and related standards without any coordination with other centers. However, active coordination with other bio-banks would increase the overall level of consistency on data standards within bio-banking. In addition, at a minimum, it would be best practice for each public or non-profit bio-bank to publish and maintain a publicly available detailed data map. This would make possible the creation of third-party integration utilities and tools. Further, bio-banks should set organizational goals that they will periodically join with other bio-banks to agree to adopt existing data standards or collectively create data standards that are specific to bio-banks. Wherever possible, bio-banks should elect to adopt existing data standards rather than create new ones.

Data Quality, QA, and QC

Good quality data starts with well-defined standard operating procedures for data generation, data capture from instruments, as well as consistent approaches to annotation, provenance, and metadata. Standardization of data management SOPs across multiple bio-banks will greatly enhance researcher confidence and ease of use. In addition to SOPs for data collection and entry, it is important to include periodic auditing of workflows and data quality to improve existing processes, remove systemic errors, and reassure investigators of the quality of the data with which they work.

Examples of common data quality techniques include well-thought-out structuring of data fields and response categories. Whenever possible, electronic entry should be constrained to menu choices from pre-structured response categories. Additionally, out-of-range rules should be implemented to constrain open entry fields to physically possible ranges. Data entry forms should be thoroughly tested and validated at the time of creation to confirm that they accurately record entered data and store it in the designated cells in the database structure. Manual processes that involve cutting and pasting among spreadsheets should be avoided. Processing steps that involve manual manipulation of spreadsheets introduce errors related to improper placement of data in columns or sometimes even transposing columns and rows when pasting.

7 Conclusions

In this document we have carried out a state-of-the-art review on current bio-banking protocols and regulations and we have discussed some ethical and legal issues that arise during exploitation of biological samples. Health care depends on research and modern research requires access to biological samples, including DNA. The potential benefits justify the establishment of DNA banks but the possibility of misuses imposes a responsibility of proper management and protection of the subjects' interests.

Collecting biological specimens and samples and storing them into bio-repositories has played an important role in the advancement of medical science. At the close of the 20th century, medical research was increasingly concerned with the genetic components of disease and their applicability to personalized medicine, which led to a heightened interest in the bio-banking of human DNA. Providing a large, well-annotated repository of biological samples for scientific research is the main idea of bio-banking. Although the storing of the actual human tissue is currently a prerequisite for preserving the information for future scientific studies, it is the information itself, contained within the tissue, which is the key to understanding the genetic background of human traits. Of equal or higher importance is the information about the tissue donors themselves, that is, their phenotypes.

The potential impact for integrating databases in health will be:

- Promote communication within and between major bio-banking initiatives, thereby helping to overcome existing fragmentation of population genomic research.
- Enhance the effective sharing of information, thereby addressing the need for very large sample sizes and helping to promote collaborative international genetic epidemiological and clinical research.
- Avoid the expensive mistakes and inefficiencies that can arise when individual initiatives repeatedly 're-invent the wheel", thereby saving funders and researchers a lot of time and money.

In the context of ACGT-clinical trials and the emerging needs of post-genomic era we integrated the necessary information that should be included in the future research databases and we have reported current efforts of bio-bank networking. In the future, there is a need of larger infrastructure to even enable studies of multi-factorial diseases in an efficient way. Such studies are in need of many more than a few hundred well-defined samples and can easily increase to thousands. In parallel, there will also be an increasing demand for more accurate and standardized sample annotation. It would therefore be mutually beneficial to unite the successful initiatives again into cooperative and communicative networks opening up new challenges for future medical research. In the context of oncology, a global cancer tissue database will improve the quality, robustness and speed of research.

Information technology is a necessary enabler, by improving existing pharmaceutical and medical practices with knowledge generated from the integration of diverse clinical and biomedical data. Bioinformatics could help to coordinate information from different institutes and integrate that data. In that way, information taken from samples can be maximised. For example, researchers could take the same tissue from the same sample and conduct parallel analyses of the genome, transcriptome and proteome. Clinical measurements of the nutritional, metabolic and immune status of patients might also be collected and stored and correlations between environmental exposures and clinical changes noted.

Appendix 1: Abbreviations and acronyms

ACD: Acid-Citrate-Dextrose

AFM: French Muscular Dystrophy Association

APOE•4: Apolipoprotein E4

ASHG: American Society of Human Genetics

BAC: Bioethics Advisory Committee

BBMRI: European Biobanking and Biomolecular Resources Research Infrastructure

BCR: Biospecimen Core Resource

BIMS: bio-bank information management system

BRCA2: Breast related cancer antigen 2

caBIG™: Cancer Biomedical Informatics Grid

CCG: Children's Cancer Group

cGP: current good practices

CHTN: US Cooperative Human Tissue Network

CIOMS: Council for International Organizations of Medical Sciences

CITES: Convention on International Trade in Endangered Species of Wild Fauna and Flora

COE: Council of Europe

COGENE: Co-ordination of genome research across Europe

CONOR: Cohort Norway

DEPC: Diethyl-pyrocarbonate

DGM: German Association for Patients with Muscle Disorders

DGR: Dangerous goods regulations

DMSO: Dimethyl-sulfoxide

DNA: Deoxyribonucleic acid

EDRN: US Early Detection Research Network

EDTA: diaminoethanetetraacetic acid,

ELPS: Ethical, legal and social policy

EORTC: European Organization for Research and Treatment of Cancer

EPO: European Patent Office

ESHG: European Society of Human Genetics

Eurordis: European Organization of Rare Diseases

FCS: Fetal calf serum

FDA: Food and Drug Administration

FFPE: formalin-fixed paraffin embedded

FISH: Fluorescence in-situ hybridization

GAIN: Genetic Association Information Network

GATiB: Genome Austria Tissue Bank

GCI: Genomics Collaborative

GCP: Good Clinical Practice

GMP: Good Manufacture Practice

HHS: Health and Human Services

HIPAA: Health Insurance Portability and Accountability Act

HSTB: University of Pittsburg Health Sciences Tissue Bank

HuGENet™: Human Genome Epidemiology Network

HUGO: Human Genome Association

HUNT: Nord-Trøndelag Health Study

IATA: International Air Transport Association

ICAO: International Civil Aviation Organization

ICD: International Classification of Disease

IHWG: International Histocompatibility Working Group

IRB: Institutional Review Board

ISO: International Organization of Standardization

IT: Information technology

LCMD: Laser capture microdissection

LIMS: laboratory information management system

LN_2: iquid nitrogen

MHC: Major histocompatibility complex

MMRC: Multiple Myeloma Research Consortium

NBF: Neutral buffered formalin

NBN: National Biospecimen Network

NCBI: National Center for Biotechnology Information

NCI: National Cancer Institute

NDC: National Dialogue on Cancer

NHGRI: National Human Genome Research Institute

NIEHS: National Institute of Environmental Health Services

NIH: National Institute of Health

NIPH: Norwegian Institute of Public Health

OCT: Optimum Cutting Temperature

OECD: Organization for Economic Co-operation and Development

OSD: Organ specific database

PBS: Phosphate buffer solution

P^3G : Public Population Project in Genomics

PCR: Polymerase Chain Reaction

PDCA: Plan, Do, Check, Act

PPPC: Public and Professional Policy Committee of the European Society of Human Genetics

QA: Quality Assurance

QC: Quality Control

R&D: Research and development

RDBMS: relational database management system

RFID: Radio Frequency Identification

RNA: Ribonucleic acid

RT-PCR: Reverse transcription polymerase chain reaction

SCBA: Self contained breathing apparatus

siRNA: small interfering RNA

SNOMED: Systemized nomenclature of medicine

SNP: Single nucleotide polymorphism

SOP: Standard Operating Procedures

SPIN: Shared Pathology Informatics Network

SPORE: Specialized Programs of Research Excellence

STN: Singapore Tissue Network

TARP: US Tissue Array Research Program

TCGA: The Cancer Genome Atlas

T*g*: glass transition temperature

UAB: University of Alabama

UNECE: United Nations Economic Commission for Europe

UNESCO: United Nations Educational, Scientific and Cultural and Organization

UPS: uninterruptible power supply

WHO: World Health Organization

WMA: World Medical Association

XML: eXtensible Markup Language

Appendix 2: A minimum (MDS) and recommended (RDS) data set (OECD Best Practice Guidelines for BRCs-2007)

Data set for DNA

MDS (data without asterisks should be provided to users)	**RDS**
Identification of depositor*	Consent
Identification number of the donor*	Family tree
Indentification number of the family*	
Identification number of the biological material	Samples from relatives available
Consent/approval by ethical committee	Form of supply
Gender and age of the donor	Maximum delay for delivery
Pathology of family with OMIM number	Karyotype
Status of the biological material (e.g. affected, non-affected, indication of suspected diagnosis)	Quantity of families and subjects available for the specific disease
Date, year and month of collection	Detail information for treatment/medications
Nature of human biological material where DNA was extracted from (e.g. affected, non-affected)	Information on disease outcome
Preservation or storage conditions	Associated clinical data
Quantity of biological material: µg/µl and number of µl	Information on life style
	Information on family history
	DNA finger printing of another method of authentication
	Hazard status

Data set for tissues and isolated cells

MDS (data without asterisks should be provided to users)	**RDS**
Identification of depositor*	Consent
Identification number for the donor*	Details of diagnosis
Identification number of the biological material	Related biological material (DNA, biopsy)
Consent/approval by ethical committee	Quantity or concentration available
Gender and age of the donor	Quaracteristics of the sample (e.g. sample composition, content tumour cells)
Disease diagnosis	Delay of freezing
Status of the biological material (e.g. affected, non-affected, indication of	Form of supply
	Maximum delay for delivery

suspected diagnosis, indication of grade tumor)	
Origin of the biological material	
Date, year and month of collection	Detail information for treatment/medications
Nature of human biological material where DNA was extracted from (e.g. tissue, slides, cell, pellets)	Information on disease outcome
Preservation or storage conditions	Associated clinical data
Documentation on processing method (e.g. chemical preservation)	Information on life style
Hazard status	Information on family history
	DNA finger printing of another method of authentication

Data set for cell cultures

MDS (data without asterisks should be provided to users)	**RDS**
Identification of depositor	Consent
Identification number of the biological material	Details of diagnosis and outcome of disease
Disease diagnosis	Related biological material (tissue, DNA)
Type of cell line (cell line, primary cultured cells, transformed cells)	Quantity or concentration available
Gender and age of the donor	Characterization of cells (doubling time, tumorigenicity, karyotype etc)
Disease diagnosis	Delay of freezing
Origin of the biological material	Form of supply
	Maximum delay for delivery
Date, year and month of collection	Number of passage
Nature of the cells (e.g. epithelia, fibroblast, lymphocyte)	Information on treatment
Preservation or storage conditions	Morphology and growth characteristics
Culture condition (medium and subculture routine)	Information on life style
Hazard status	Reference paper for cell lines
	DNA fingerprinting of another method of authentication

8 References

1. http://www.ethics.ubc.ca/workingpapers/deg/deg005.pdf.
2. http://www-03.ibm.com/industries/healthcare/doc/content/bin/Bio-banks_Accelerating_Molecular_Medicine.pdf
3. Riegman PH, Dinjens WN, Oosterhuis JW. Bio-banking for interdisciplinary clinical research. Pathobiology. 2007;74(4):239-44.
4. National Bio-specimen Network Blueprint. http://prostatenbnpilot.nci.nih.gov/FINAL_NBN_Blueprint.pdf
5. Creation and governance of human genetic research databases. http://213.253.134.43/oecd/pdfs/browseit/9306091E.PDF
6. Lähteenmaki R. Estonian parliament considers genome law. Nat Biotechnol. 2000; 18(11):1135.
7. UK Bio-bank. http://www.ukbio-bank.ac.uk/
8. Palmer LJ. UK Bio-bank: bank on it. Lancet. 2007;369(9578):1980-2.
9. CARTaGENE project. http://www.cartagene.qc.ca/accueil/index.asp.
10. Asslaber M, Abuja PM, Stark K, Eder J, Gottweis H, Trauner M, Samonigg H, Mischinger HJ, Schippinger W, Berghold A, Denk H, Zatloukal K. The Genome Austria Tissue Bank (GATiB). Pathobiology. 2007; 74(4):251-8.
11. The Swedish National Biobank Program, www.bio-banks.se
12. http://www.bio-banks.se/documents/RapportBio-banksprogrammet.pdf.
13. http://www.hunt.ntnu.no/index.php?side=english/background.
14. Næss O, Søgaard AJ, Arnesen E, Beckstrøm AC, Bjertness E, Engeland A, Hjort PF, Holmen J, Magnus P, Njølstad I, Tell GS, Vatten L, Vollset SE, Aamodt G. Cohort Profile: Cohort of Norway (CONOR). Int J Epidemiol. 2007 Nov 4 [Epub ahead of print].
15. "Bio-bank Japan" toward the personalized medicine. Abstract for presentation at 11th International Congress of human genetics. http://www.ichg2006.com/abstract/497.htm.
16. Singapore Tissue Network. http://www.stn.org.sg/.
17. Case studies of existing human tissue repositories. Best practices for a bio-specimen resource for the genomic and proteomic era. Eiseman E, Bloom G, Brower J, Clancy N, Olmsted SS. NCI, National Dialogue on Cancer. RAND Science and Technology.
18. Swede H, Stone CL, Norwood AR. National population-based bio-banks for genetic research. Genet Med. 2007;9(3):141-9.
19. NCI. www.cancer.gov.
20. The stored tissue issue. Biomedical research, ethics and law in the era of genomic medicine. Weir R and Olick RS. Oxford University Press 2004.
21. www.decode.com
22. http://spores.nci.nih.gov/
23. http://pcabc.upmc.edu/
24. http://www.bio-banks.eu/documents/INFO_BBMRI.doc

25. EuroBio-bank
26. http://chtn.nci.nih.gov/
27. The Cancer Genome ATLAS, http://cancergenome.nih.gov/
28. http://www.cdc.gov/genomics/hugenet/default.htm
29. http://www.p3gconsortium.org/
30. www.cancerdiagnosis.nci.nih.gov/spin/#description
31. Riegman PH, Dinjens WN, Oomen MH, Spatz A, Ratcliffe C, Knox K, Mager R et al. TuBaFrost 1: Uniting local frozen tumour banks into a European network: an overview. Eur J Cancer. 2006;42(16):2678-83.
32. Cambon-Thomsen A, Rial-Sebbag E, Knoppers BM. Trends in ethical and legal frameworks for the use of human bio-banks. Eur Respir J. 2007; 30(2):373-82.
33. CIOMS International Ethical Guidelines for Biomedical Research involving human subjects. Revised draft, January 2002.
34. Tri-Council Policy Statement. Ethical Conduct for Research Involving Humans.
35. Guidelines for clinical protocols in genetic research. Anneca T, Bompiani A, Buttura R, Calori G. et al. http://www.fsk.it/eventi/guidelines.pdf.
36. Data storage and DNA banking for biomedical research: technical, social and ethical issues. Recommendations of the European Society of Human Genetics. Eur. J. Hum. Genetics 2003; 11(suppl 2) : S2-S10.
37. World Medical Association: http://www.wma.net/e/about/index.htm
38. Human Genome Organization: http://www.hugo-international.org/mission_history.htm.
39. Laurie G. Genetic databases: assessing the benefits and the impact on human and patient rights--a WHO report. Eur J Health Law. 2004;11(1):87-92.
40. WHO. Genetic databases: Assessing the benefits and the impact on human and patient rights [http://www.law.ed.ac.uk/ahrb/publications/online/whofinalreport.pdf].
41. http://whqlibdoc.who.int/hq/2002/a74580.pdf.
42. UNESCO. Universal Declaration on Bioethics and Human Rights http://portal.unesco.org/en/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html.
43. Reymond MA, Steinert R, Eder F, Lippert H. Ethical and regulatory issues arising from proteomic research and technology. Proteomics. 2003;3(8):1387-96.
44. Godard B, Schmidtke J, Cassiman JJ, Aymé S. Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. Eur J Hum Genet. 2003;11 Suppl 2:S88-122.
45. Simon J, Paslack R, Robienski J, Cooper DN, Goebel JW, Krawczak M. A legal framework for bio-banking: the German experience. Eur J Hum Genet. 2007;15(5):528-32.
46. http://www.tubafrost.org/research/moreinfo/deliverables/TUBAFROST%20Deliverable%207.1.pdf.
47. Vidalis T, Manolakou K. Report on banks of biological material of human origin (bio-banks) in biomedical research. [http://www.bioethics.gr/media/pdf/reports/bio-banks_rep_eng.pdf].

48. Gottweis H, Zatloukal K. Bio-bank governance: trends and perspectives. Pathobiology. 2007;74(4):206-11.

49. International Bioethics Committee. Draft report on consent.

50. Ethical and policy issues in Research involving human participants. Report and recommendations of the National Bioethics Advisory Commission. [http://bioethics.georgetown.edu/nbac/human/overvol1.pdf].

51. Annas GJ. HIPAA regulations - a new era of medical-record privacy? N Engl J Med. 2003;348(15):1486-90.

52. Auray-Blais C, Patenaude J. A bio-bank management model applicable to biomedical research. BMC Med Ethics. 2006;7:E4.

53. ISBER. Best practices for repositories: Collection, Storage, and Retrieval of Biological Materials for Research. 2008.

54. Working Party 1, Draft. International Network ofBbiological Resource Centres for Cancer Research: Recommendations on common minimal technical standards. 2006. http://www.iarc.fr/News/RecommendationsBRC.pdf

55. Australasian Bio-specimen Network. Bio-repository Protocols. 2006.

56. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD et al. HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics. 2005;5:3262-77.

57. NCI Best Practices for Bio-specimen Resources 2007.

58. Olund, Lindqvist, and Litton, “BIMS: An information management system for biobanking in the 21st century”, IBM SYSTEMS JOURNAL, VOL 46, NO 1, 2007

59. Lickerman, Erik, Bry L., Herring N., Pandelidis Y., Afsheen, Nighat, Sharif A., “Grid-Enabled Bio-Sample Management Application for Data-Intensive Biomarker Analysis”, Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)

60. M. Isabelle, I. Teodorovic, M.M. Morente, D. Jamine, A. Passioukov, S. Lejeune, P. Therasse, W.N.M. Dinjens, J.W. Oosterhuis, K.H. Lam, M.H.A. Oomen, A. Spatz, C. Ratcliffe, K. Knox, R. Mager, D. Kerr, F. Pezzella, M. van de Vijver, H. van Boven, S. Alonso, D. Kerjaschki, J. Pammer, J.A. Lopez-Guerrero, A. Llombart Bosch, A. Carbone, A. Gloghini, E.-B. van Veen, B. van Damme, P.H.J. Riegman, “TuBaFrost 5: Multifunctional central database application for a European tumor bank”, EUROPEAN JOURNAL OF CANCER 42 (2006) 3103 –3109

61. Väliaho, J., Riikonen, P., and Vihinen, M. “Biomedical data description with XML. Distribution of immunodeficiency fact files - from Web to WAP”, BMC Med Inform Decis Mak. 2005 Jun 26;5:21.

62. GENATLAS, http://www.dsi.univ-paris5.fr/genatlas/

63. Stenson et al (2003), The Human Gene Mutation Database, Hum Mutat (2003) 21:577-581

64. Swiss-Prot Protein Knowledgebase, http://www.ebi.ac.uk/swissprot/

65. The caBIG project, https://cabig.nci.nih.gov/

66. Richard Casey, “How Federated Databases Benefit Bioinformatics Research”, http://www.b-eye-network.com/view/2164