Sample Identification and Tracking in Biobanks

A thesis submitted to
University of Dublin, Trinity College
for the degree of
Doctor of Philosophy

Atieh Zarabzadeh
Department of Computer Science
Trinity College
Dublin

October 2009

Declaration

I, the undersigned, declare that this work has not been previously submitted as an exercise for a degree at this or any other University, and that, unless otherwise stated, it is entirely my own work.

_____

Atieh Zarabzadeh

October 2009

Permission to lend or copy

I, the undersigned, agree that the Trinity College Library may lend or copy this thesis upon request.

_____

Atieh Zarabzadeh
October 2009

Dedication

To my parents...

Acknowledgment

I would like to express my appreciation to all those without whose help I would not have been able to complete this thesis.

First and foremost I offer my sincerest gratitude to my supervisor, Prof. Jane Grimson for her invaluable support, supervision and encouragement. I would like to express my profound thanks to Prof. William Watson, University College Dublin, for his patient support and valuable suggestions that enabled me observe the world from a different window. Completing this thesis without Jane and Bill's advice would not have been possible.

I am grateful to the Irish Prostate Cancer Research Consortium members for their support in the first year when I was learning about the biobanks, and for their kind cooperation during the evaluation phase. I am especially grateful to Ms. Marie Burke, Dr. Linda Sullivan, Dr. Ruth Foley and Dr. Antoinette Perry. I really appreciate the kindness of Dr. Darach Golden and Dr. Geoff Bradley from the Trinity Centre for High Performance Computing during all these years and especially during the last few months for setting up a test server and allowing me to work in their office.

I also wish to express my thanks to the Irish Research Council for Science, Engineering and Technology and Hewlett-Packard for providing funding for this project, without which life would have been difficult.

In my daily work I have been blessed with a friendly and cheerful group in the Center for Health Informatics (CHI) laboratory. Their support and encouragement made long working hours tolerable. I wish to express my appreciation to former and current staff and students in the School of Computer Science and Statistics.

I am thankful to all my friends for their supports, encouragements and advice. I am deeply indebted to Farzad Hayati for his support, encouragement and help in the implementation phase of the project and beyond. I am grateful to Karen Tierney and Sinead Kelleher for editing and proofreading my thesis and to Ronan McDonnell and

Dr. Bridget Kane for their advice throughout all these years. I am also indebted to Maureen D'Arcy for her kindness during the times I was under pressure. I would also like to express my thanks to Dr. Mohammad Jamshidi.

I would like to express my love and gratitude to my beloved family; for their understanding and endless love not only throughout the duration of my studies but also throughout all my entire life. My parents, Nasrin and Mohsen; and my sister, Talieh are the best I could ever have.

Abstract

Biobanks or bio-repositories facilitate the storage and maintenance of biological samples and data to support discovery of biomarkers, therapeutic targets, and the underlying causes of diseases. Such discoveries require high quality samples that are supported by their associated data. While samples and their data are of crucial importance, they are expensive to collect and maintain. Samples and data are subject to various procedures from data collection from the sample donors, to clinical laboratory analyses, to processing in the research laboratory (omic procedures) and ultimately to knowledge discovery. Samples are transferred between different locations and moved in and out of freezers at several different points in this workflow, increasing the potential of error in sample identification and incorrect linkage with the corresponding data. Implementing a robust and reliable Sample Identification and Tracking System (SITS) which can track samples through all the various phases in different locations and at the same time protecting the confidentiality of the sample donors represents a major challenge.

In this thesis a novel method has been developed based on Radio Frequency Identification (RFID) technology to support reliable tracking of samples from collection to knowledge discovery. In this system RFID tags are attached to the tubes containing the samples and send and receive data to and from the users through a web based interface. These RFID tags are passive, allowing longer life span than active RFID tags, rewriteable and survive extreme temperatures. They have small physical size and store dynamic data to support biobank activities. In addition to sample tracking, this system improves quality control of samples by recording details of collection, procedure and storage data such as Standard Operating Procedures (SOPs) under which the sample has been collected and processed. The SITS prototype has been evaluated by the Irish Prostate Cancer Research Consortium (PCRC) biobank, a federated biobank for prostate cancer.

Publications Related to This PhD

1. Zarabzadeh, A., Hayati, F., Watson, R. W. G., Bradley, G., Grimson, J. (2009) *An Overview of Sample Identification and Tracking System for Biobanks*: Abstract for the 14th Annual Conference and Scientific Symposium of the Healthcare Informatics Society of Ireland (HISI). Dublin, Ireland.
2. Zarabzadeh, A., Hayati, F., Watson, R. W. G., Bradley, G., Grimson, J. (2009) *A Sample Identification and Tracking System for Biobanks*: Proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems (CBMS 2009). New Mexico, USA.
3. Zarabzadeh, A., Watson, R. W. G., Bradley, G., Grimson, J. (2008) *Participants Confidentiality in Prostate Cancer Research Consortium Bio-repository*: Abstract for the 13th Annual Conference and Scientific Symposium of the Healthcare Informatics Society of Ireland (HISI). Dublin, Ireland.
4. Zarabzadeh, A., R. W. G. Watson, et al. (2009). *Ensuring participant privacy in networked biobanks*. Principles and Practice in Biobank Governance. J. Kaye and M. Stranger, Ashgate.
5. Zarabzadeh, A., Watson, R. W. G., Bradley, G., Grimson, J. (2008) *Ensuring Participant Confidentiality in Bio-repositories*: Abstract for Conference on Governing Biobanks – What are the challenges? Oxford, UK.
6. Zarabzadeh, A., Watson, R. W. G., Grimson, J. (2008) *The use of Radio Frequency Identification to Track Samples in Bio-repositories*: Proceedings of the IEEE 1st International Conference on Information Technology (IT2008). Gdansk, Poland. pp.351-354.
7. Zarabzadeh, A., Watson, R. W. G., Bradley, G., Grimson, J. (2007) *Sample Tracking and Identification in Biobanks*: Abstract for the 12th Annual Conference and Scientific Symposium of the Healthcare Informatics Society of Ireland (HISI). Dublin, Ireland.

Table of Contents

List of Figures

List of Tables

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
ABN	Australian Biospecimen Network
AF	Accessibility Format
AIAG	Automotive Industry Action
AIDS	Acquired Immune Deficiency Syndrome/Acquired Immunodeficiency Syndrome
AJAX	Asynchronous, JavaScript, And XML
ANSI	American National Standard Institute
API	Application Programming Interface
ASHG	American Society of Human Genetics
BBMRI	Biobanking and Biomolecular Resources Research Infrastructure
BGA	Blood Group Analyser
BIMS	Biobank Information Management System
BIRF	Brain Injury Resource Foundation
BMA	British Medical Association
CASPIAN	Consumers Against Supermarket Privacy Invasion and Numbering
CAVD	Collaboration for AIDS Vaccine Discovery
CD	Compact Disk
CC	Composite Component
CFR	Code of Federal Regulations
CHI	Center for Health Informatics
CMOS	Complementary Metal Oxide Semiconductor
CPCTR	Cooperative Prostate Cancer Tissue Resource
CRC	Clinical Resource Centre
CSO	Central Selling Organisation
CSV	Comma Separated Value
DDE	Dynamic Data Exchange
DNA	DeoxyriboNucleic Acid
DoS	Denial of Service

DTC	De Beers' Diamond Trading Company
EAN	European Article Number
ECTMS	Electronic Clinical Transfusion Management System
EGP	Estonian Genome Project
ELN	Electronic Laboratory Notebook
EMEA	European Agency for the Evaluation of Medicinal Products
EPC	Electronic Product Code
EPCglobal	Electronic Product Code global
EPIC	Electronic Privacy Information Center
ETSI	European Telecommunications Standards Institute
EU	European Union
FDA	Food and Drug Administration
FCC	Federal Communications Commission
GCP	Good Clinical Practice
GHAVE	Global HIV/AIDS Vaccine Enterprise
GHRC	Global HIV Vaccine Research Cryorepository
GLP	Good Laboratory Practice
GPS	Global Positioning System
GS1	Global Standards One
GUI	Graphical User Interface
GWU	George Washington University Medical Center
HGRA	Human Genes Research Act
HIPAA	Health Insurance Portability and Accountability Act
HIV	Human Immunodeficiency Virus
HP	Hewlett-Packard
HTML	HyperText Markup Language
http	Hyper-Text Transport Protocol
IBM	International Business Machines
ID	Identifier
IE	Internet Explorer
IEEE	Institute of Electrical and Electronics Engineers
IET	Institution of Engineering and Technology

IFF	Identity Friend or Foe
IMM	Institute of Molecular Medicine
IP	Internet Protocol
IPAS	Integrated Patient Administration System
IRB	Institutional Review Board
ISBER	International Society for Biological and Environmental Repositories
ISM	industrial, scientific and medical
ISN	Institute Study Number
ISO	International Organisation of Standards
IT	Information Technology
ITF	Interleaved Two of Five
KCN	Kimberley Certificate Number
KI	Karolinska Institutet
KPCS	Kimberley Process Certification Scheme
LAN	Local Area Network
LIMS	Laboratory Information Management System
LIS	Laboratory Information System
LN	Liquid Nitrogen
MCW	Medical College of Wisconsin
MEC	Medical Ethics Committee
MHRA	Medicines and Healthcare products Regulatory Agency
MIT	Massachusetts Institute of Technology
NBAC	National Bioethics Advisory Commission
NBN	National Biospecimen Network
NBTC	National Blood Transfusion Committee
NHS	National Health Services
NHS CFH	NHS Connecting for Health
NPSA	National Patient Safety Agency
NY	New York
NYU	New York University School of Medicine
OCR	Optical Character Reading
OLE	Object Linking and Embedding

PAS	Patient Administration System
PBMCs	Peripheral Blood Mononuclear Cells
PCRC	Prostate Cancer Research Consortium
PDF417	Portable Data File 417
PITT	The University of Pittsburgh
PJM	Phase Jitter Technology
PPI	Positive Patient Identification
PPID	Positive Patient Identification
QA	Quality Assurance
QC	Quality Control
REST/RESTful	REpresentational State Transfer
RF	Radio Frequency
RFID	Radio Frequency Identification [Devices]
RNA	Ribonucleic acid
RS232	Recommended Standard 232
SDMS	Scientific Data Management System
SHOT	Serious Hazards of Transfusion
SITS	Sample Identification and Tracking System
SMS	Sample Management System
SOA	Service-Oriented Architecture
SOP	Standard Operaing Procedure
SRS	System Requirements Specification
TAM	Technology Acceptance Model
TMA	Tissue Microarray
TTP	Third Trusted Party
UCC	Uniform Code Council
UK	United Kingdom
UML	Unified Modeling Language
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UPC	Universal Product Code
UPI	Unique Patient Identification
URL	Uniform Resource Locator

USB	Universal Serial Bus
UUID	Universal Unique Identifier
VPN	Virtual Private Network
WAN	Wide Area Network
WfMS	Workflow Management System
Win32	Windows 32
Win32 Controller	Windows 32bit Compatible Controller
WMA	World Medical Association
XML	Extensible Markup Language

Chapter 1. Introduction

1.1 Motivation

Recent advances in the fields of genetics and proteomics offer the realistic prospect of personalised medicine – of a healthcare system which is truly targeted at the needs of the individual (2005). In order to understand the underlying genetic characteristics of a human being two types of data are required; namely molecular data that are derived from biological specimens, and phenotypic data collected about the individuals themselves (Betsou et al., 2004, Schmitz et al., 2005, Compton et al., 2005, Ölund et al., 2007). By analysing these two types of data together, the underlying genetic causes of diseases and protein biomarkers can be discovered which may lead, for example, to the ability to predict how a particular disease will progress in a particular individual based on their genetic and phenotypic make-up (Betsou et al., 2004). The collection of molecular and phenotypic data has been made feasible by the development of biobanks which provide long-term storage of biological samples (Schmitz et al., 2005, Compton et al., 2005). Biobanks allow optimised storage of samples and data such that their quality is preserved. Biological samples are stored using a variety of methods depending on their type. For example, tissue samples can be stored in the form of paraffin blocks; blood samples can be stored as whole blood or centrifuged, processed and stored as Peripheral Blood Mononuclear Cells (PBMCs) and serum aliquots at -20°C or -80°C freezers or in Liquid Nitrogen (LN) tanks (Betsou et al., 2004).

An essential prerequisite for successful biobanking is ensuring the security of donors' data and protecting their privacy in order to gain their trust and maintain their confidence (Schmitz et al., 2005). The trust of donors can be facilitated by the open, transparent and ethical management of consents (Schmitz et al., 2005). Various types of consent can be used depending on the purposes of the study and the nature of the particular biobank. For example, under informed consent the donor is given detailed information about the purpose of the study, the data to be collected, who will have access to it, what happens if they change their mind and wish to withdraw consent, and so on. Whereas under broad consent, the donor is simply given a general indication of

the purpose of the study (Hansson, 2005, Godard et al., 2003). Security of data records regardless of their storage medium should be assured (Dodek and Dodek, 1997).

Modern biobanks rely on Information Technology (IT) (Betsou et al., 2004). IT plays an important role, for example, in the integration of molecular and phenotypic data; statistical analysis of the data; and in understanding the complex relationship between them (Compton et al., 2005). IT tools are available to provide such functionality but typically only in a fragmented fashion. The software tools need to be standardised and integrated in order to fulfil the requirements of biobanks and to allow effective and efficient sharing and integration of data (Schmitz et al., 2005).

A major challenge faced by researchers in many clinical domains is the lack of sufficiently large numbers of donors to allow statistically valid analyses of the data. Equally, the absence of a system to support long-term follow-up of individuals imposes significant limitations (Melamed et al., 2004). Furthermore, confining the study to a single institution not only results in small study cohorts but also can be source of bias in the study which further undermines the validity of the research (Melamed et al., 2004). To overcome these limitations, multi-institutional biobanks are being developed to facilitate larger study cohort size and avoid bias. Examples of multi-institutional biobanks include the Irish Prostate Cancer Research Consortium (PCRC) biobank (PCRC, 2009) and the Cooperative Prostate Cancer Tissue Resource (CPCTR) in the United States (CPCTR, 2007). The employment of a central database with a web-based query tool plays a vital role in the success of multi-institutional or distributed biobanks (Patel et al., 2006). Issues of comparability of quality of samples collected and processed in various institutions can be eliminated by enforcing the application of Standard Operating Procedures (SOPs). SOPs are written policies and procedures that all staff and personnel working within its scope are obliged to adhere to (Pitt et al., 2005). SOPs should be in place for every procedure and action, including sample and data collection, storage, distribution and ethical issues such as consenting donors (Schmitz et al., 2005, Compton et al., 2005). Errors occurring during a procedure can be determined based on the divergence in carrying it out from the SOP of that procedure (Pitt et al., 2005). Staff and personnel who carry out the procedures and actions, should

be involved in developing and approving SOPs, and these should be updated as and when necessary (Schmitz et al., 2005).

The situation is further complicated by the fact that each individual sample is typically sub-divided into a number of aliquots to maximise the utility of the sample and to facilitate different types of processing on a single sample (for example proteomics, metablomics, and so on). Furthermore, each individual aliquot must be linked not only to its parent sample but also to the data generated from it.

Thus a vital and crucial responsibility of biobanks is maintaining a standardised link between the physical sample itself, its associated data records on the database (Ölund et al., 2007) and its aliquots. Biobank Information Management System (BIMS) has been developed to enable such a link and effectively act as middleware supporting communication between other systems in the biobank (Ölund et al., 2007, Litton, 2004).

The need to have a robust and flexible sample tracking system in biobanks has been emphasised for consistency in quality control measures, adherence to SOPs (Schmitz et al., 2005), establishing and maintaining a link between the physical sample, its associated data records and aliquots. A unique identifier is required to support linking the physical sample with its associated data records as well as integrating various information related to that sample (Schmitz et al., 2005). Electronic sample tracking reduces the instances of data-transcription error and improves efficiency and confidence (Martin et al., 2007, McGiven et al., 2007). Such an electronic system for identifying and tracking samples should ideally have a web-based interface and be supported by a database while the infrastructure may be based on, say, barcoding, or Radio Frequency Identification (RFID) (Schmitz et al., 2005).

A Laboratory Information Management System (LIMS) handles the management and storage of samples, while a BIMS handles communication in a biobank, supports integration of data and links samples and data (Litton, 2004). Many biobanks are implemented using LIMS software. Barcodes have been successfully tested and used in LIMS setting. They are found to have reduced data-transcription errors while increasing throughput, confidence, speed and efficiency (McGiven et al., 2007, Martin et al., 2007). The UKBiobank (UKBiobank, 2008), a major biobank collecting blood samples

in UK, has deployed a configured version of a LIMS product along with barcodes to manage samples and data and track aliquots (Elliott and Peakman, 2008, Schreier, 2008). In addition to barcodes, RFID technology has also been tested in biobanks. For example, the Paoli Calmettes (Paoli-Calmettes, 2009) biobank in France has used RFID to track samples and store data (Bettendorf et al., 2005). In addition, intelligent freezers and storage cabinets have been designed that can automatically locate samples within themselves, such as the system by Magellan Technology (MagellanTechnology, 2009) in Australia (Bacheldor, 2008).

RFID was first proposed by Stockman (1948), however it was not deployed until 1970s (Adams, 2007, Roberts, 2006). Transfer of data is based on radio wave transmissions in RFID technology (Adams, 2007, Piramuthu, 2007). RFID consists of three main components: tags or transponders, reader/writer device, and an antenna that allows transferring data between the tags and the reader/writer device (Domdouzis et al., 2007). Communication between the three components must be based on an agreed standard in respect, for example, of the frequency and protocol to be used (Adams, 2007). There are a variety of different types of tags depending on their functionalities and capabilities. Active, passive and semi-passive tags are available depending on whether they are powered by an onboard battery, receive their energy from the radio waves, or both (Roberts, 2006).

While RFID is based on penetration of radio waves it has the limitation of behaving strangely at certain frequencies when near liquid or metal (Domdouzis et al., 2007). There are also other challenges with regard to the security of RFID tags such as sniffing, tracking individuals without their consent, spoofing, replaying attacks and Denial of Service (DoS) (Rieback et al., 2006, Knight, 2006). However, RFID is already widely used in a wide variety of environments such as hospitals, supply chain, retail and logistics (Ngai et al., 2008).

The existence of a Sample Identification and Tracking System (SITS) is of crucial importance for biobanks where samples and data are valuable resources for future data and knowledge discovery and must therefore be reliably and securely managed. Sample and data loss and mix-up are one of the major sources of error in multi-institutional

biobanks. Thus a fundamental requirement for the Irish multi-institutional PCRC biobank is an effective, integrated SITS.

This research proposes a novel technology-based Sample Identification and Tracking System (SITS) that tracks samples across various locations in a multi-institutional biobank and utilises a unique identifier for each sample. Since ensuring the confidentiality of donors is of such of crucial importance in biobanks, this area will be the subject of detailed scrutiny in order to ensure that the solution proposed guarantees confidentiality. The proposed SITS supports longitudinal tracking of samples at all stages and throughout all processes. The system will support the collection and storage of both static and dynamic data. It will operate reliably at all times and in particular when samples are subjected to processes such as centrifugation, storage in extremely cold freezers and Liquid Nitrogen (LN) tanks. SITS will reduce the possibility of sample mix-up and data loss. Potential technologies such as barcodes, Radio Frequency Identification (RFID), databases and web will be investigated and the best technology or combination of technologies will be adopted to develop a prototype SITS.

The key motivation of this research is to develop a SITS that while maintaining the confidentiality of sample donors allows:

(i) Tracking each sample through the workflows of the procedures it goes through regardless of the environmental (for example freezing, thawing) and procedural (for example centrifugation) conditions, and tracking each sample to its parent sample from which it was initially collected,

(ii) Tracking each sample back to its parent and vice versa,

(iii) Accompanying samples with their associated data throughout the procedures in order to identify them accurately,

(iv) Tracking samples collected from the same participant longitudinally over time and

(v) Facilitating users with a web-based query tool to search for samples of their interest.

1.2 Research Question

The research question posed in this thesis is *What is a more reliable and effective way of identifying and tracking biological samples in a multi-institutional biobank while (i) maintaining confidentiality of donors, and (ii) supporting samples collected longitudinally over time?*

In addition to an investigation of confidentiality maintenance approaches, the focus of this thesis will be on designing and implementing a SITS that is based on technologies that best satisfy the needs and requirements of biobanks. The two main candidate technologies to emerge and which are analysed in detail are barcodes and RFID, supported by database and web. The designed and implemented system will be based on RFID technology that is the subject of thorough investigation.

The proposed SITS will allow identification of samples through a unique identifier. The system supports tracking samples throughout all the procedures to which they are subjected thereby allowing a full history of samples to be maintained. A web-based interface allows users to pose queries about the samples, their location, procedures undergone and results obtained, and so on. Evaluation of this system will also be carried out in the context of the Prostate Cancer Research Consortium (PCRC) multi-institutional biobank (PCRC, 2009).

1.3 Research Goals and Objectives

In response to the research question of this thesis, the objectives of the research are the investigation of approaches to and the development of a solution to:

(i) Mapping the sample journey and associated processes,

(ii) Providing a system of unique identification for samples,

(iii) Providing a mechanism for linking derived samples (aliquots) to their parent sample and vice versa,

(iv) Tracking sample history,

(v) Linking the sample to its associated data,

(vi) Ensuring confidentiality of donors including consenting, consent management, concealing of data to prevent identification of donors,

(vii) Providing a role based system of granting access privileges to users,

(viii) Validating the resulting SITS in the context of the Irish PCRC biobank.

The quality and validity of the research, such as the discovery of biomarkers, supported by biobanks is dependent on a number of factors; in particular the preservation of sample quality and correct linkage to its associated data. Thus the motivation behind this research is to reduce sample loss and mix-up, improve the quality of samples by observing every action carried out on them and consequently improving the data that will be used for knowledge discovery.

The outcome of the investigation on confidentiality maintenance is to document best practice in relation to:

- Approval of studies that samples will be used for,
- Consent types, procedures and management,
- Methods of preventing identification of donors by unauthorised individuals while allowing removal of samples in case of withdrawal from consent, and
- Security means that should be put in place to ensure individuals have correct access to right information.

The main goal of this thesis is to develop a SITS that allows identification and tracking of biological samples through sets of complex procedures while ensuring that all confidentiality requirements are met, using RFID technology. In order to validate the priori choice of RFID, this technology will be compared with other candidate technologies such as barcodes with respect to the needs of biobanks at theoretical level. Database and web are also utilised as the backbone and interface of the system, respectively.

1.4 Research Contribution

This research makes a number of contributions to the development and implementation of biobanks and their associated information systems. Specifically, the research has provided novel approaches and insights into:

(i) Best practice guidelines for the management of confidentiality in multi-institutional biobanks

(ii) A robust, reliable, and secure method of sample identification and tracking in multi-institutional biobanks

(iii) Mapping of Standard Operating Procedures (SOPs) and their representation in the standard Unified Modelling Language (UML).

Firstly, best practice guidelines for maintaining confidentiality in multi-institutional biobanks are proposed and validated in the context of the PCRC biobank. This best practice is based on an in-depth analysis of a variety of types of consents and data concealing approaches. It is a comprehensive guideline for maintaining the studies and activities of biobanks within the approved scope and ensuring that samples and their associated data are stored safely and securely. Ensuring confidentiality is essential both ethically and legally. It is also essential to ensure the trust of donors and hence to facilitate the participation of significant numbers in the cohort to support knowledge discovery.

Secondly, the proposed SITS tracks samples through the range of complex omic procedures. Data needed for each stage of the procedure is identified and a database to house tracking and identification data is designed. This SITS is based on RFID technology, supported by the database and a web application. The prototype supports various conditions that samples go through such as extreme temperatures and different processes such as centrifugation. A unique identifier keeps track of all of the samples and aliquots.

Finally, the thesis makes an important contribution to the documentation of Standard Operating Procedures (SOPs). SOPs are essentially workflow specification, which detail

what needs to be done at each stage in the collection, processing and analysis of samples. These are essential in the context of biobanking where the robustness of research results depends critically on the consistency of sample collection and processing. The Unified Modelling Language (UML) offers a graphical tool for representing these workflows clearly and unambiguously.

1.5 Thesis Overview

An overview of the various types of biobanks and Biobank Information Management Systems (BIMSs), how they are used, is presented in Chapter 2 together with a detailed account of the Irish Prostate Cancer Research Consortium (PCRC) biobank which has been used as the validation domain for this research. The term "donor" and "participant" will be used interchangeably in this thesis and both refer to the study subjects.

Confidentiality issues are investigated in Chapter 3. This Chapter provides the reader with background and definitions and then develops a set of guidelines based on best practice for ensuring confidentiality. It discusses how confidentiality is maintained in the PCRC biobank as a multi-institutional biobank where samples are collected and processed in different locations and by different individuals. The approach taken by the PCRC biobank to confidentiality and how it adheres to best practice is provided in this Chapter.

The candidate technologies available for identification and tracking will be discussed in Chapter 4. These include linear barcodes and 2-dimentional (2D) matrices, RFID and a number of other emerging technologies. Barcodes and RFID emerge as the best candidates and hence are the subject of a detailed comparison. Their pros and cons are investigated and the technology that best fits the biobank needs, namely RFID, is chosen.

Applications of barcodes and RFID in different environments are surveyed and lessons are learned from each application in Chapter 5. The limitations and advantages that these technologies offer are studied in pilots and projects. To further investigate problems that may be encountered when implementing SITS, different scenarios that

have common characteristics with sample identification and tracking and biobanks are also discussed in Chapter 5.

Chapter 6 of this thesis develops the system requirements and validation for SITS. Based on the Institute of Electrical and Electronics Engineers (IEEE) recommended Software Requirements Specifications (SRS) a set of requirements are formulated and defined for SITS. High level requirements are given in Chapter 2 and these are developed further in Chapter 6. However, system requirements are vital for developing a system. The second part of this Chapter will focus on evaluating the feasibility of delivering the requirements described earlier in the Chapter, in the context of the Irish PCRC biobank.

Chapter 7 covers the detailed design of SITS based on the requirements specified in Chapter 6. The development of a SITS prototype for the PCRC biobank, PCRC-SITS, will be discussed in this Chapter.

Chapter 8 provides an evaluation of PCRC-SITS. While end-to-end tracking of a statistically significant number of samples is not feasible within the timeframe of this project, the evaluation strategy involved sub-dividing the process into various parts that could be individually evaluated in different institutes that are members of the PCRC biobank and by different individuals. Coverage of workflows is evaluated in Chapter 8 followed by an evaluation of integrating PCRC-SITS with BIMS. The results of a survey of PCRC biobank members on identifying and tracking samples and their perceptions of electronic tagging are also presented.

Finally, the thesis concludes by assessing the extent to which the thesis answers the initial research question and meets the objectives set out in this Chapter. The limitations of the current prototype and its evaluation are presented. The Chapter concludes with a discussion of future work.

Chapter 2. Sample Identification and Tracking Systems

2.1 Introduction

Biobanks are repositories of biological samples and data that are collected over time from patients with a certain disease or from the general population to facilitate researchers and scientists and to support data analysis and knowledge discovery. There are various types of biobanks based on a variety of parameters including scope, activities, and goals. Many successful biobanks have been established such as the UKBiobank in the UK (UKBiobank, 2008), the Cooperative Prostate Cancer Tissue Resource (CPCTR) in the US (CPCTR, 2007), the Karolinska Institutet (KI) biobank in Sweden (KI, 2007), the Estonian Genome Project in Estonia (EGP), and the Global HIV Vaccine Research Cryorepository (GHRC) (IBMT, 2009). Samples collected in biobanks will go through various procedures for analysis and knowledge discovery. In a typical biobank, these include sample collection, preparation, processing and, finally, long-term storage and retrieval. Samples and their associated data, including their identifiers, molecular and phenotypic data, must be kept linked throughout these procedures.

Regardless of the type of biobank, the quality of samples and their associated data play a vital role in biobanks and hence the existence of a management system to handle the activities carried out is crucial. It is the responsibility of the Biobank Information Management System (BIMS) to take care of data communication and exchange within a biobank. The BIMS is in charge of consent handling and sample and information management. Controlling the quality of samples is also part of the BIMS responsibilities. Standard Operating Procedures (SOPs) along with Good Laboratory Practice (GLP) and Good Clinical Practice (GCP) have been developed to ensure that samples collected and processed by different individuals are of comparable and appropriate quality. Also, each individual biobank typically develops its own guidelines and requirements.

The Irish Prostate Cancer Research Consortium (PCRC) (PCRC, 2009) has established a multi-institutional biobank collecting blood, urine and tissue samples from patients with prostate cancer. Samples in this biobank are collected in hospitals and after

preparation they are sent to the research institutes for omic procedures. Three main phases of a sample's lifecycle can be identified: initial processing phase, secondary processing phase, and long-term storage and retrieval phase. The initial processing phase is carried out in hospitals and includes sample preparation and initial aliquotting. At the end of this phase, samples are stored in -20°C freezers. Depending on the sample type, each sample collected in the initial processing phase is sub-divided into three aliquots. Aliquots are then transferred to the research institutes where researchers and scientists process them through omic procedures. These omic procedures typically involve the sub-division of the sample into a large number of very small sized samples. The need to have a Sample Identification and Tracking System (SITS) in place is emphasised by the complex workflows that samples go through. Therefore, approaches that successful biobanks have taken towards identifying and tracking samples will be discussed, and these will form the basis for a set of generic high-level requirements for SITS.

This Chapter is organised as follows: definitions of a biobank, its various types, and examples of biobanks are provided in Section 2.2. Sample and data journeys, the abstract workflow that samples go through are discussed in Section 2.3. BIMS, standards of sample collection and quality control, are given in Section 2.4. Details of the PCRC biobank infrastructure, sample and data flow are provided in Section 2.5. The development of a SITS in different biobanks is examined in Section 2.6. Finally, the requirements of SITS at a high level are drawn in Section 2.7. The Chapter finishes with a conclusion in Section 2.8.

2.2 Biobanking

2.2.1 Definitions

Discovering the underlying causes and the prognosis of diseases requires two types of data: molecular data and phenotypic data, that are retrievable from clinical analysis on biological samples and from individuals' personal information, respectively (Betsou et al., 2004, Schmitz et al., 2005, Compton et al., 2005, Ölund et al., 2007). Both types of data are obtained from the analysis of samples gathered from generally large cohorts of individuals and from their medical records longitudinally over time and, in some cases,

from family members of these individuals (Compton et al., 2005). A “biobank”, “bio-repository” or “bio-resource”, interchangeable terms in the literature, stores and maintains these data and samples from large populations for several years. Biobanks are resources for storing biological samples and data in an optimised fashion to support the discovery of biomarkers, therapeutic targets, and the underlying causes of diseases that have been made feasible by advancements in molecular science and technology.

A biobank has been defined by Winn et al. (1990), and Holland et al. (2005) as:

> “a system which will store one or many types of biological specimens for later analysis from single or multiple studies under conditions which permit efficient retrieval and optimum stability of the sample”.

Another definition of a biobank from a different angle is provided in the International Evaluation of Swedish Biobanks, March 2005 (Sorensen et al., 2005). It defines a biobank as:

> “A long-term depository of biological samples from an identifiable human population. The content of the depository must be of such quantity and quality that it is suitable for later biomedical analysis in epidemiological and clinical research for individual clinical purposes”.

A number of characteristics for describing a biobank can be drawn from the above definitions. A biobank is an infrastructure for storing single or multiple types of samples along with their associated data from a cohort of individuals and/or their family members over time, to populate study samples for later analysis by researchers and scientists which may lead to discoveries about understanding or diagnosing the disease.

Samples are collected, processed and stored in a biobank for several years. Processes that samples go through may include omic procedures such as proteomics and metablomics, and may also include centrifugation or similar processing. With the proteomics procedure, proteins and their changes after administration of a drug, or their changes by disease are observed (Anderson and Anderson, 1998). In the metablomics procedure, the chemical fingerprints that a given cellular process creates are studied (Daviss, 2005). Each omic procedure analyses samples from a different perspective.

Omic procedures lead to the production of a massive amount of information. The reason behind the creation of biobanks is to supply large and "well-annotated" samples for scientific analysis (Ölund et al., 2007). Samples are required to be annotated by donors' associated data, as well as the disease and laboratory data of the sample, but no donors' personal identifiable information (Compton et al., 2005). Data stored in biobanks range from clinical patient information to sample, molecular, annotated, analysed and interpreted information (Schmitz et al., 2005). Clinical, molecular and data about the participant him/herself are of central importance to knowledge discovery in biobanks (Compton et al., 2005). Samples without their associated data are not scientifically valuable for scientific analysis (Ölund et al., 2007).

Figure 2-1 depicts the key components of a typical biobank. These components include donors who may be either of control (healthy) or patient populations, samples and data collected from donors, analysis tools applied to samples and data, the infrastructure needed for storage of samples and data and, finally, research and applications carried out on samples and data (BBMRI). Samples collected from donors may be serum, Peripheral Blood Mononuclear Cell (PBMC), Deoxyribonucleic Acid (DNA), Formalin-Fixed Paraffin-Embedded (FFPE) tissues, frozen tissues and cells.

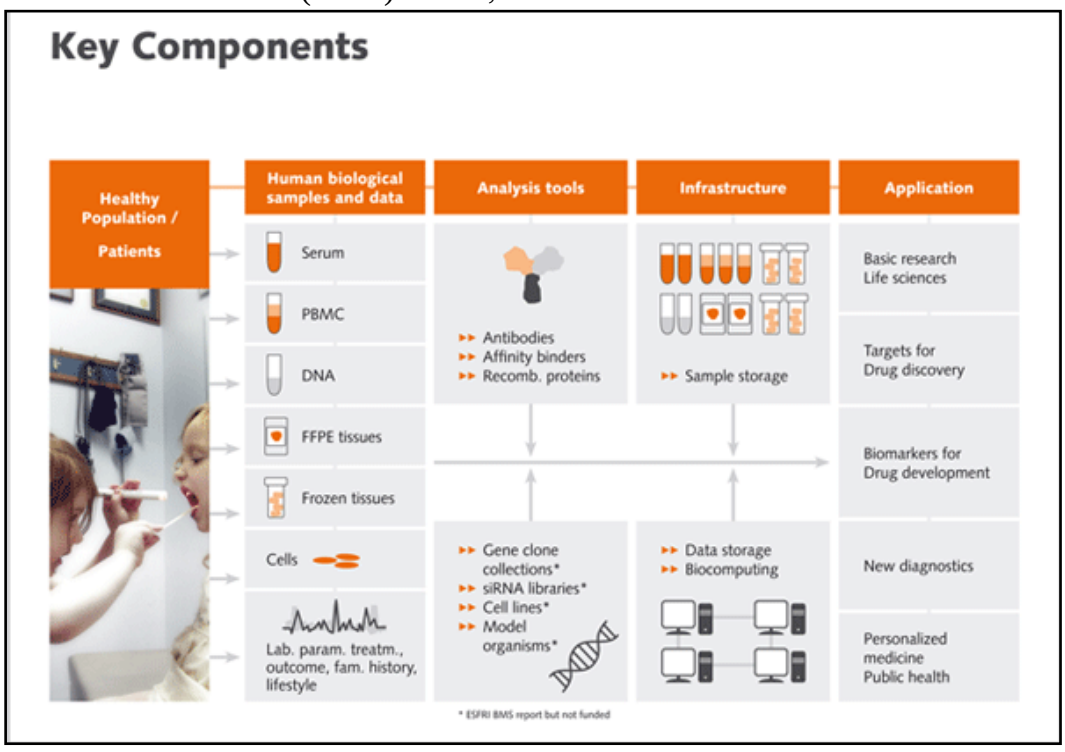


Figure 2-1: Key Components of a Typical Biobank (BBMRI)

2.2.2 Types of Biobanks

There are various different types of biobanks depending on their scope, goals and purpose (Betsou et al., 2004). Although a single biobank might belong to more than one type, they can be classified as disease, type, location, or population-specific, and multi-purpose or multi-institutional.

Disease-specific: In this type of biobank, samples are collected, maintained and stored for investigation of a particular disease and its causes (Betsou et al., 2004). For example, the Cooperative Prostate Cancer Tissue Resource (CPCTR) in the US only collects samples from prostate cancer patients (CPCTR, 2007).

Type-specific: Single or multiple types of samples are collected in type-specific biobanks. For example the UK Human Tissue Bank (UKHTB) (UK Human Tissue Bank, 2006) only facilitates collection of tissue samples.

Location-specific: The scope of a location-specific biobank's activities is limited to a particular location, i.e., they only maintain and store samples collected from a particular geographic region. The outcome of research on such a cohort might vary for another region. An example of this type of biobank is the UKBiobank that supports the collection and maintenance of samples from the UK only (UKBiobank, 2008).

Population-specific (Betsou et al., 2004): This type of biobank provides resources for studies on populations with particular characteristics such as gender, age, background or race. An example of this type of biobanks is the GenomEUtwin study whose population cohorts consists of Danish, Finnish, Italian, Dutch, English, Australian and Swedish twins (GenomEUtwin, 2009).

Multi-purpose: These are biobanks in which samples that are collected and maintained can be used for a number of studies. The KI biobank is an example of this type of biobank (KI, 2007). Depending on the consent approach employed in this type of biobank, studies may have to be defined in advance of any sample being collected, or studies may be defined as samples are being collected. Different types of consents will be discussed in Chapter 3 of this thesis.

Multi-institutional: When two or more institutes collaborate on the collection and maintenance of samples and data to populate larger studies they form a multi-institutional biobank. Such collaboration improves the limitations of the sample collection process in each individual institute (Melamed et al., 2004). It also allows the inclusion of samples from various cohorts in the study. One of the challenges that these biobanks face is the issue of standardisation and unification of samples and data so that they are of comparable quality (Betsou et al., 2004). The CPCTR (CPCTR, 2007) is an example of a multi-institutional biobank that is comprised of four academic institutions (Patel et al., 2006).

2.2.3 Examples of Biobanks

A number of successful biobanks have been developed and are currently in operation. Examples include the UKBiobank in the UK (UKBiobank, 2008), the CPCTR in the US (CPCTR, 2007), the KI biobank in Sweden (KI, 2007), the Estonian Genome Project in Estonia (EGP), and the Global HIV Vaccine Research Cryorepository (GHRC) (IBMT, 2009).

The UKBiobank: This biobank was launched in April 2003 (Watson and Cyranoski, 2005) with the aim of collecting samples and data from over 500,000 participants aged from 40 to 69 years (Johnston and Kaye, 2004, Ölund et al., 2007, UKBiobank, 2008) focusing on cancer and heart disease since these are found to be the main causes of death in this age group (Johnston and Kaye, 2004). In this biobank longitudinal research is carried out to understand the relationship between disease, lifestyle and genes that are of interest to the scientists (Johnston and Kaye, 2004). Samples and data are collected from various locations and therefore, the UKBiobank (UKBiobank, 2008) has developed protocols in the form of SOPs. Details of samples to be collected from each participant and the procedures of collection and preliminary processing, transportation as well as details of temperature and aliquotting are provided in these SOPs (Elliott and Peakman, 2008). The UKBiobank (UKBiobank, 2008) is a population-specific, multi-purpose, multi-institutional and location-specific biobank.

The CPCTR: This resource (CPCTR, 2007) was established in April 2000 in the US, with the aim of providing a large collection of samples from prostate cancer patients

that are supported by quality-controlled and standardised pathological review which will lead to quality-controlled and detailed outcome data. This data will in turn be used for biomarker validation studies by the researchers (Melamed et al., 2004). Four sites are participating in the CPCTR:

1. George Washington University Medical Center (GWU), Washington, DC,
2. Medical College of Wisconsin (MCW), Milwaukee, WI,
3. New York University School of Medicine (NYU), New York, NY, and
4. The University of Pittsburgh (PITT), PA (Patel et al., 2006).

A central database houses data from these four sites, while physical samples are maintained at each site (Melamed et al., 2004). The CPCTR (CPCTR, 2007) is a disease-specific, multi-institutional and location-specific resource for studies on prostate cancer. The aim of this biobank is the collection of tissue and fluid samples from a large population of patients with prostate cancer disease, with standardised details of their associated data and pathological review (Patel et al., 2005). These data and samples are then made available to the scientists for biomarker discoveries.

The KI biobank: The reasons for the success of Swedish biobanks are identified as being a long history of collecting samples and data from a large cohort of the population, the legal framework to support biobank activity, and an infrastructure for linking samples and data together (IA Sorensen et al., 2005). One of these successful biobanks in Sweden is the Karolinska Institutet (KI) biobank. The Karolinska Institutet in Sweden established its biobank with the aim of enhancing research and science (Betsou et al., 2004) in 2004 (KI, 2007). The KI biobank (KI, 2007) is a well facilitated biobank and is accredited to the ISO 17025:2005 standard (Beskow, 2009). The idea of a Biobank Information Management System (BIMS) was first introduced in the context of this biobank. The KI biobank BIMS is shown in Figure 2-2. Its BIMS is responsible for integrating sample, phenotype, genotype and medical record data (see Section 2.4). The integration of data facilitates complex searches being carried out on the biobank samples supporting knowledge discovery (Ölund et al., 2007, Karolinska Institutet biobank). The query searches are possible via the biobank website and its portal (Karolinska Institutet biobank). The four categories of data: sample, phenotype, genotype and medical record data, are stored in the database of the BIMS. The

integrated information is then made available to the researchers. The KI biobank (KI, 2007) is a location-specific multi-purpose biobank.

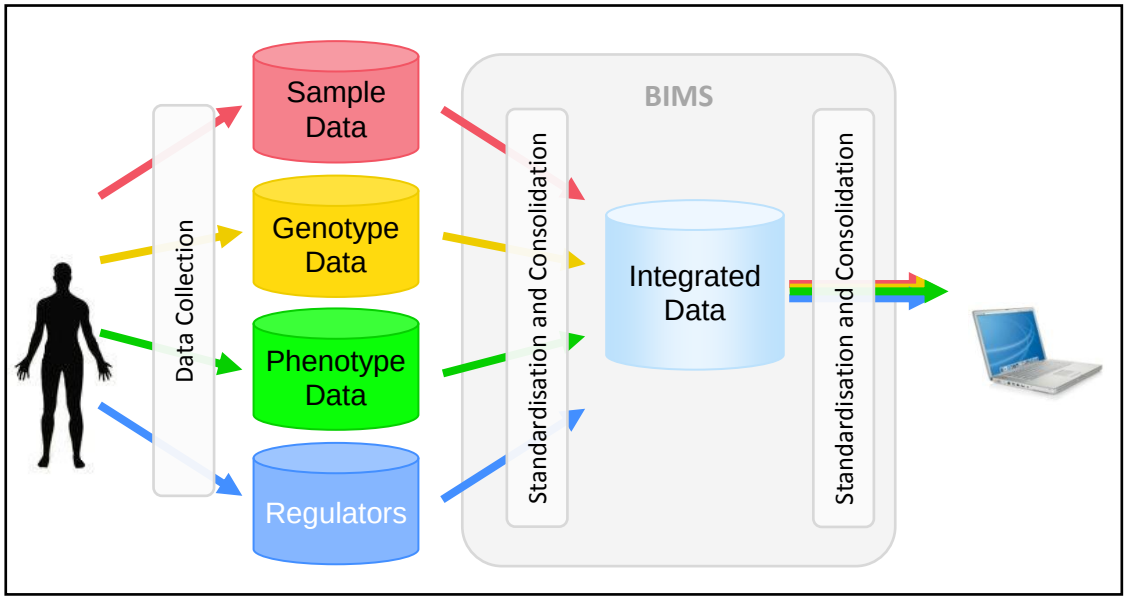


Figure 2-2: KI BIMS Redrawn (Karolinska Institutet biobank)

The Estonian Genome Project (EGP): This project was presented to the Estonian government by the Estonian Genome Foundation in June 2000 (Metspalu, 2004). Later that year, the Human Genes Research Act (HGRA) was passed by the Estonian parliament (Fletcher, 2004, Metspalu et al., 2004). Then the EGP began as a country wide project investigating the health of the public by establishing a database of genetic and medical records (Betsou et al., 2004). The EGP attempts to include about 1 million DNA samples in one database (Sutrop and Simm, 2004). The goal of establishing such a project is to bring together phenotype and genotype data from a large portion of the country's population and to support scientific and public health research (Metspalu et al., 2004). The population of Estonia is estimated to be about 1.3 million in 2009 (Statistics Estonia, 2009). Donors who take part in this project are randomly chosen to avoid study biases (Metspalu et al., 2004). The EGP biobank is a multi-purpose, multi-institutional, and location-specific biobank.

The Global HIV Vaccine Research Cryorepository (GHRC) (IBMT, 2009): This project provides central storage service facilities to the Collaboration for AIDS Vaccine Discovery (CAVD) with the aim of discovering a vaccine for AIDS (Ihmig et al., 2009, Shirley et al., 2009). CAVD supports Global HIV/AIDS Vaccine Enterprise (GHAVE)

(Global HIV Vaccine Enterprise, 2009) through its thirteen HIV vaccine discovery consortia and five central storage service facilities including the Fraunhofer-IBMT (Durst et al., 2007, Ihmig et al., 2009, Shirley et al., 2009). These discovery consortia and facilities are spread across 89 institutes in 22 countries worldwide (Ihmig et al., 2009, Shirley et al., 2009). The GHRC maintains a collection of preserved specimens collected from patients who are at the early stage of infection from the "Primary Sites" such as Brazil, Russia and South Africa (Ihmig et al., 2009). This biobank is a multi-institutional and disease-specific biobank.

The Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) (BBMRI): BBMRI has gathered more than 200 organisations from 24 European countries to improve the quality of samples and data, and to minimise fragmentation of the activities carried out across European countries. Currently, more than 10 million samples are managed under BBMRI (BBMRI).

Table 2-1 shows more examples of biobanks with their locations, scope and type.

Biobank	Location	Scope	Type
UKBiobank (UKBiobank, 2008)	UK	Country wide 40 to 69 years age group	Multi-purpose Multi-institutional Location-specific Population-specific
CPCTR (CPCTR, 2007)	US	Country wide Male participants	Multi-institutional Disease-specific Location-specific
KI biobank (KI, 2007)	Sweden	Country wide	Multi-purpose Location-specific
Estonian Genome Project (EGP)	Estonia	Country wide	Multi-purpose Multi-institutional Location-specific
DeCode Genetics (deCODE)	Iceland	Country wide	Multi-purpose Location-specific
GenomEUtwin (GenomEUtwin, 2009)	Multi-country	Danish, Finnish, Italian, Dutch, English, Australian and Swedish twins	Multi-institutional Multi-purpose Population-specific
GHRC (IBMT, 2009)	Multi-country	Worldwide	Multi-institutional Disease-specific

Table 2-1: Examples of different types of biobanks

2.3 Sample and Data Journey

2.3.1 Integrating Samples and Data

Figure 2-3 depicts a generic diagram of sample and data flow in biobanks. Three major phases of sample and data journeys are clearly delineated, namely:

1. The sample collection and preparation phase, where samples are collected from donors and are prepared by, for example, aliquotting or centrifuging,
2. The sample processing phase which is mainly the omic procedures carried out on samples and which leads to the generation of data used for knowledge discovery and data mining,
3. The long-term storage and retrieval of samples and aliquots for future analysis, which may involve storage of samples at extremely cold temperatures, typically a temperature of -90°C and -190°C, in freezers or Liquid Nitrogen (LN) tanks.

Samples must be supported by their associated data during each of these three phases. For example, date of collection of sample and sample identifiers must accompany sample throughout the three phases. Depending on the type of the biobank and its governing strategy, each phase might occur in an independent location and/or by various individuals. This is mostly due to the facilities at each location and the expertise of each individual. Each biobank may have a different approach to carrying out operations.

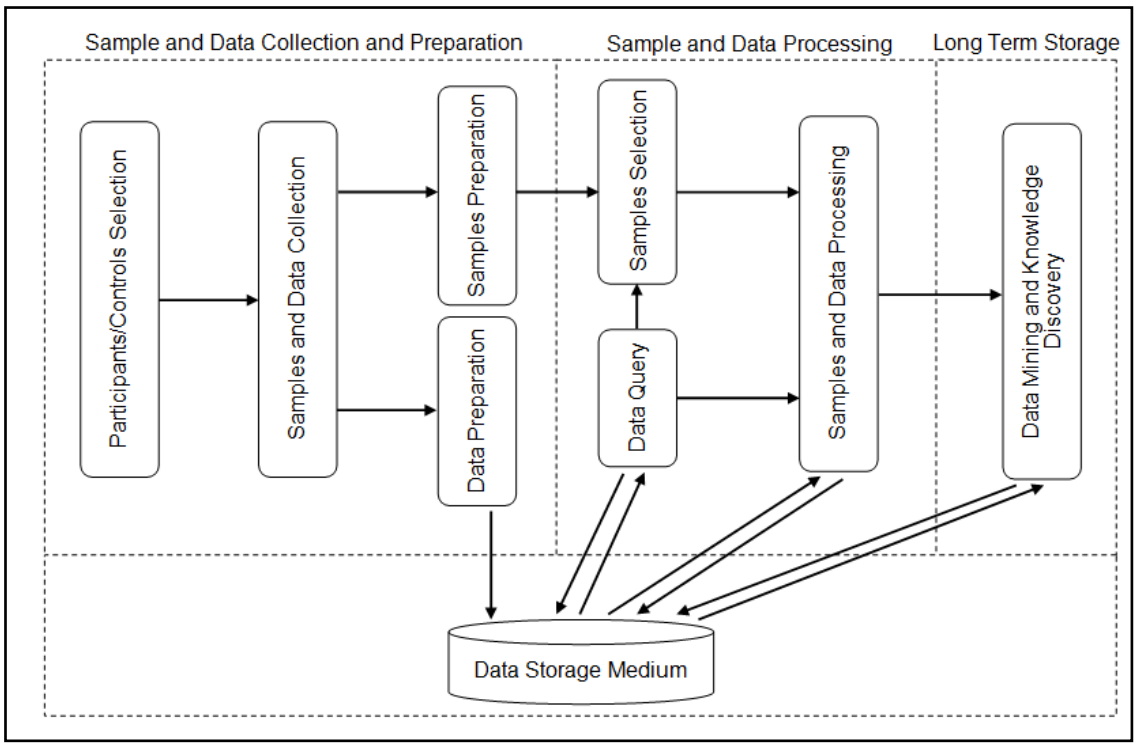


Figure 2-3: Sample and Data Flow in Biobanks

Data being collected from different locations and procedures are integrated through a five step model defined by Ölund et al. (2007). The procedures that the data will go through in this model are *extraction*, *deidentification*, *consolidation*, *abstraction* and *querying* as depicted in Figure 2-4. Data *extraction* is undertaken on databases or other types of data resources that include the larger population on which the study is going to be carried out. During *deidentification*, identifiers of the extracted data are removed, replaced with a code, or pseudonymised for the purposes of maintaining confidentiality of donors. In Ölund et al. (2007) model, *deidentification* is achieved either by *alteration* or by *exclusion*. Methods of *deidentification* will be discussed in Chapter 3 where confidentiality issues in biobanks are described. While different types of data may have been collected from different resources or by different methods, they need to be merged together after *deidentification*. This is known as *consolidation* in the Ölund et al. (2007) model. Having prepared and standardised the data, it undergoes *abstraction* to be made ready for *querying* by scientists.

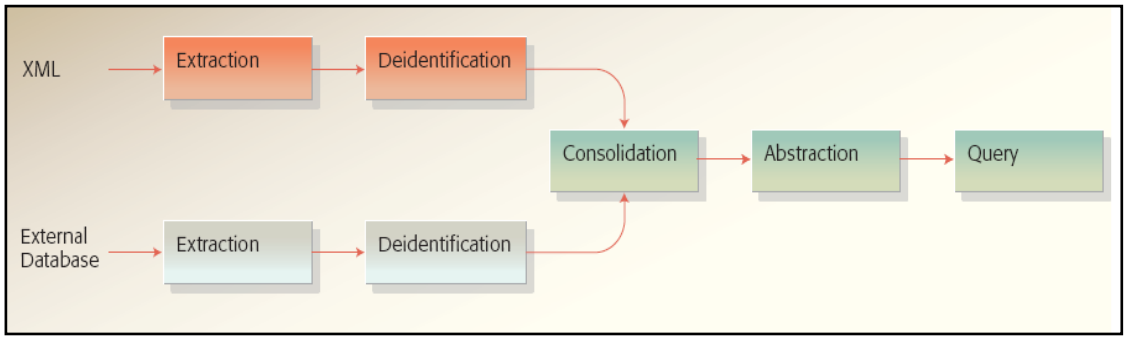


Figure 2-4: Data Integration Model (Ölund et al., 2007)

Incorporating the data integration process with samples and data flow to allow data availability when needed, is of crucial requirements for supporting biobanks. Each step of the data integration model can be mapped to the activities carried out in the three phases of sample processing as shown in Figure 2-5. The various phases involved in Figure 2-5 will be further explained in the following subsections.

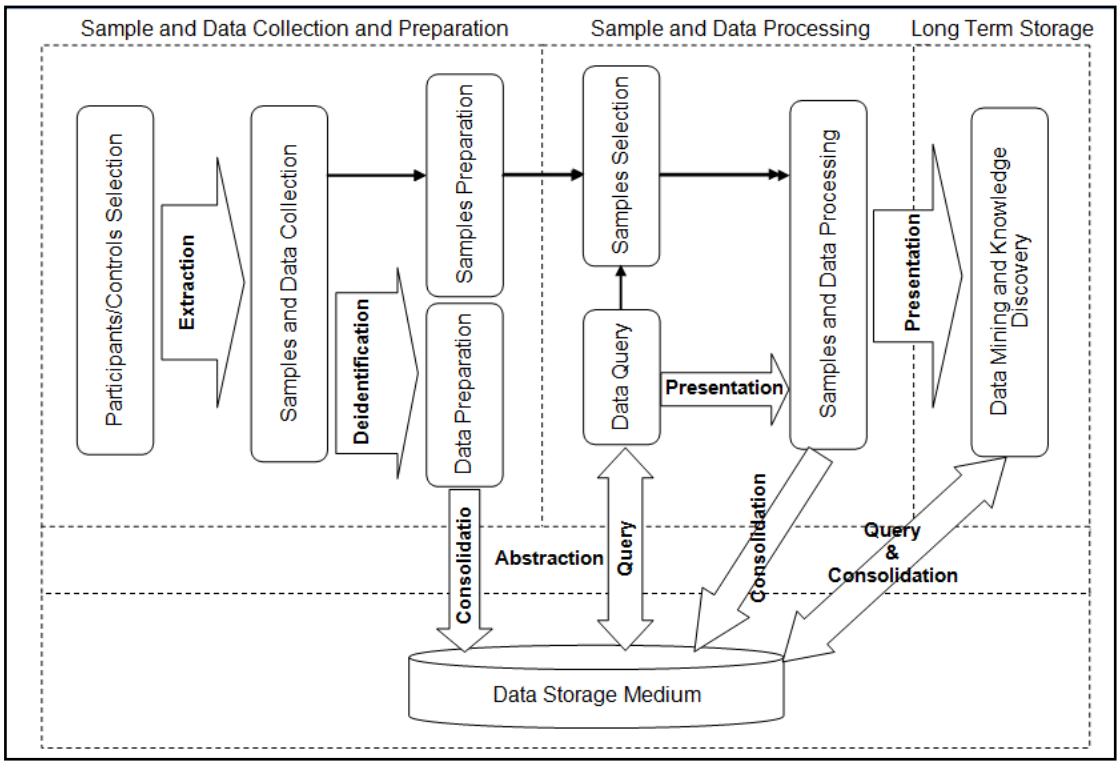


Figure 2-5: Data Integration Process Incorporated with Sample Flow

2.3.2 Sample and Data Collection

The initial step in sample collection is the identification of donors who are suitable for the purposes of study to be carried out. The disease under study, age group, gender and other criteria may influence population selection. Some clinical trials and scientific experiments carried out in biobanks require comparison between the subjects under study and a control group to recognise differences or to identify the effect of treatments. A control group does not receive the new treatment that is being studied (NCI, 2009). Generally, a user who already has access to the population records browses the database or other form of records and approaches individuals or patients who fit the study criteria. Another method for selecting donors is that the consultant who is caring for the patient decides if a patient is suitable for the study and then approaches them. The approach to consent depends on the rules and regulations of the biobank and the country. Then sample and data will be gathered from each identified and consented individual. Certain types of data, such as history of medication, that are already available in the hospital or on another database will be *extracted*; for example, using Extensible Markup Language (XML) models (Ölund et al., 2007). Longitudinal studies allow monitoring of the progression or regression of the disease and the influence of any treatment given. Hence, they are most desirably collected at specified time intervals. Samples then will have to be prepared by adjusting their volumes and other characteristics so that they match the SOPs employed in the biobank. At this stage the *deidentification* of data should be carried out as part of the data preparation. Data records of samples collected from different participants are then merged in a data storage medium, typically a database. This is referred to as *consolidation* in Ölund et al.'s (2007) integration model and in Figure 2-5. *Merging* data will require data being in comparable and identical format. These standardised samples and data will be made available to internal or external researchers depending on the design of the biobank, hence it is vital that a predefined model is in place for consolidating data.

2.3.3 Sample Processing

After completion of the collection phase, researchers select samples and data from the biobank resource based on their study parameters. They run *queries* with their study parameters on the database through an *abstraction* layer that displays data in an

appropriate format and through predefined access controls. The *query* results will also be made available to researchers through the *abstraction* layer. Then they carry out omic procedures on the provided samples and *presented* data.

The next major step is sample and data processing. This is an important step as in this step omic procedures and analyses will be undertaken and are the basis on which future research studies will be formed. The outcomes of these procedures will be used for analysis that may lead to drug and knowledge discovery and the invention of new therapy methods. Omic procedures are complicated in most cases and generate a large amount of data and aliquots. Data gathered during omic procedures will be stored in the database. Again, *consolidation* operations are applied to data for storing procedure data in the database. This database includes data collected previously, so in addition to *merging* different omic procedural data, procedures data is *merged* with collection data. Physical samples will be stored in low temperature freezers or Liquid Nitrogen (LN) for future use.

2.3.4 Long-term Storage and Retrieval

Investigations on longitudinal data require storage and maintenance of samples and data for several years and this will be the final phase of operation on samples. The outcomes of previous omic procedures stored over time are used for knowledge discovery and data mining during this phase, and hence it is vital to maintain samples that are stable and of appropriate quality. Samples of acceptable quality have their characteristics preserved as they were when collected from the donor. In order to monitor the progression or regression of the disease, statistical methods as well as other strategies are applied to the data that are *queried* by the investigators based on their study parameters. Data and samples are then retrieved for further analysis and to support knowledge discovery from the clinical data and the results of previous omic procedures on the sample. During this phase, data stored in a database or any other medium will be accessed and new data or research outcomes will be added to the records available. Thus, more data will be *consolidated* into the database. Generally, to achieve more accurate results, a larger cohort of population is investigated, requiring longer periods of sample collection and storage.

2.4 Biobank Information Management Systems

2.4.1 Definition

A management system capable of handling massive amounts of samples and updatable data, integrating information gathered from multiple sources and allowing queries to be carried out on the data, as well as having control over access to the samples and data in the most secure and reliable manner, is needed to facilitate suitable data and sample storage, maintenance and retrieval (Ölund et al., 2007). Hence the concept of a Biobank Information Management System (BIMS) has been developed. This system is in charge of supporting samples and data throughout the three phases illustrated in Figure 2-5.

Litton (2004) has described a BIMS as:

> "... a middleware system that will handle communication between several other systems. ... The system will be responsible for a range of tasks, including: Middleware functionality. ... Result storage. ... Participant consent handling. ... Sample and information management. ... ."

And Eiseman et al. (2003) has described it as:

> "The backbone of any repository is a standardised, scalable, and secure bioinformatics system that is appropriate for repository management, tissue acquisition and management, and data aggregation and analysis"

A BIMS design needs to include defining, structuring and standardising the information collected, such that they are comparable and of a standard quality (Ölund et al., 2007).

From a data management point of view, a BIMS must be able to take care of constantly updated data, control access to data with a user-friendly interface and, finally, support query searches on the data (Zarabzadeh et al., 2008). Integrating a BIMS with Information Technology (IT) or building an informatics structure for biobanks facilitates this. Biobank informatics is required to support collection, management, distribution and any other activities carried out in a biobank through standardised software (Schmitz et al., 2005). At the European Bio-Banking and Biomolecular Resource meeting held in Vienna, 2007, Professor Litton from the KI biobank described biobank informatics as an

arrangement of six elements: sample management, data integration, data collection, data query, security and administration management, and data analysis. Merriam-Webster (2002) has defined informatics as:

> "the collection, classification, storage, retrieval, and dissemination of recorded knowledge".

Applying this definition of informatics to biobanks for storing and organising data and samples leads to taking advantage of databases for safe data storage and management. A BIMS in general employs a database for data storage and retrieval. However, databases by themselves are not capable of managing and standardising samples and data.

Biobanks may deploy a BIMS that is currently available in the market by customising it or they can develop their own BIMS tailored to their particular requirements. Choosing to use an off-the-shelf BIMS or developing a new one depends highly on the strategies that each biobank might employ. Examples of off-the-shelf BIMS are Distiller by Slidepath (SlidePath) and Azura Biobank by AZURA (AZURA).

2.4.2 Sample Collection Standards and Quality Control

A vital requirement of any sample analysis to be carried out is having samples and data of appropriate quality, that is, sample quality being preserved by freezing them at collection time (Betsou et al., 2004) and for data to be accurate and reliable. High quality outcomes are obtained from analysis of high quality input. Scientific quality assurance is one of the six major issues discussed in the Second World Wide Biobank Summit, Collaborating for Cures, that was held in New York in November 2004 (Schmitz et al., 2005). According to the document compiled from this summit, two approaches towards ensuring quality of data and samples are found to be developing biobanking Good Laboratory Practice (GLP) and Good Clinical Practice (GCP), and requiring the biobanks to put them into practice (Schmitz et al., 2005).

In addition to GLP and GCP, Standard Operation Procedures (SOPs) are also used to standardise operations carried out in biobanks. Examples of these operations are obtaining donor consent, and collecting and processing samples. A SOP is defined as:

> "Detailed written instructions to achieve uniformity of the performance of a specific function across studies and patients at an individual site" (U.S. National Institutes of Health, 2008).

SOPs are used to standardise procedures, policies and processes within a single biobank or across multiple biobanks (Zarabzadeh et al., 2008). They are plain text file documents explaining the details of every operation performed in the biobank, from data collection through to data analysis, in an unambiguous language that is understandable to the person performing the action (Pitt et al., 2005).

Unified sample collection, and management and handling issues become more complicated with multi-institutional biobanks, and even more so in collaboration between different biobanks (Schmitz et al., 2005). Multi-institutional or collaborative biobanks are intended to overcome the problem of small numbers of samples for trials (Melamed et al., 2004). The requirement for such collaboration is standardising operations carried out in each location, such that samples and data collected from different locations are of comparable quality.

The Australian Biospecimen Network (ABN), established as an organisation to discuss the technical, legal/ethical and managerial matters regarding samples collected for medical research, has developed "Biorepository Protocols" to be used for standardising samples collected in different biobanks (Catchpoole et al., 2007). In addition to ethical issues, guidelines on the collection and storage of blood, tissue and other types of specimens are provided. In this guideline, sample handling, details of storage, transportation and quality control have been explained in detail (Catchpoole et al., 2007).

The National Biospecimen Network (NBN) in the US has developed a report entitled "Case Studies of Existing Human Tissue Repositories "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era" (Eiseman et al., 2003). This report offers best practice drawn from five repositories studied and can be used by the future NBNs as well as other biobanks.

In addition to the guidelines, protocols and SOPs mentioned above, the International Society for Biological and Environmental Repositories (ISBER) has developed its first

version of best practice for repositories (Pitt et al., 2005). This document provides details of the procedures that samples will go through, i.e. collection, processing, storage and retrieval as well as best practice to be considered for other activities in biobanks.

2.4.3 Sample and Data Management

Another issue that was raised in the Second World Wide Biobank Summit was having quality control measures in place (Schmitz et al., 2005). Sample tracking and an infrastructure for sample analysis should be employed in order to ensure adherence to the SOPs developed for each biobank. ISBER considers deciding on a unique labelling system and applying it across biobanks as the initial step in standardising sample identification and tracking (Pitt et al., 2005). Biobanks need to have an inventory system in place in order to locate samples and report on their status (Pitt et al., 2005). ISBER requires adherence to the Title 21 Code of Federal Regulations requirements of Part 11 (21 CFR Part 11) (FDA, 2000), for which any modification made to the database should be logged; the labels should be capable of providing complete information about the labelled sample; and connections to other associated databases should be facilitated (Pitt et al., 2005). Furthermore, according to ABN, labels and the ink used for printing should survive all potential processing and storage conditions. Using barcodes is also recommended so a direct automated link between the sample and the database exists (Catchpoole et al., 2007).

Sample loss and mix-up are major sources of medical error. Since samples in biobanks go through similar conditions at the time of sample collection and labelling as in normal clinical laboratories, it is expected to have similar rates of identification errors. A multi-institutional study focusing on five sources of error mislabelled, unlabeled, partially labelled, incompletely labelled, and illegible label on more than 3.3 million samples has revealed that a rate of 0.92 per 1,000 labels are identified inaccurately (Wagar et al., 2008). Another study (Martin et al., 2007) carried out on 21,351 surgical specimens showed that there were 91 identification errors from which 18 specimens were not labelled, 16 containers were empty, 16 stored with incorrect laterality, that is the label on the specimens were not consistent, 14 with tissue site reported incorrectly, 11 with incorrect patient reported, 9 with no patient name, and 7 with no tissue site.

2.5 The Irish Prostate Cancer Research Consortium (PCRC) Biobank

2.5.1 Infrastructure of PCRC Biobank and its BIMS

Figure 2-6 shows the infrastructure of the Irish PCRC biobank's BIMS. This multi-institutional, disease- and location-specific biobank, consists of four major components: the participating hospitals with their Clinical Resource Centres (CRC), research institutes, an identification and tracking system, and the central database. There are currently four Irish hospitals: Mater Misericordiae, St. James's, St. Vincents, and Beaumont and two research institutes: the University College Dublin (UCD) Conway Institute and the Trinity College Dublin Institute of Molecular Medicine (IMM) participating in this biobank (PCRC, 2009).

Data and samples are collected from participants in hospitals. Participants' identifiable data is gathered and stored in the local database of the hospital. Identifiable data and biological samples then undergo a de-identification process in the same location and by the same individual who collected samples at that site. De-identified data is then stored in the central database and the de-identified samples go for the initial processing phase including aliquotting. De-identified data from each collection site is accommodated in the central database so that it is accessible by users from other sites. This process is carried out in each hospital's corresponding Clinical Resource Centre (CRC) or Genome Resource Unit (GRU) by a research nurse. The research nurse is in charge of collecting samples from donors and preparing samples in the collaborating hospital. Samples are prepared and aliquotted into smaller tubes according to SOPs for that particular type of sample. These SOPs unify and standardise the process across all sites. Collection and preparation data is then stored in the central database for quality assurance purposes and may also be used by the research scientists subsequently. After the initial processing phase, de-identified samples are stored in designated freezers at the same site. These freezers are maintained at -20°C or -80°C depending on the SOP for the sample type.

Samples are then made available to the researchers and scientists in the research institutes to carry out omic analysis. Also, researchers and scientists can search samples from the central database and issue queries for the ones that are of interest to them. Omic analysis can be thought of as the secondary processing phase, that follows the

initial processing phase, in research institutes. Large number of small size samples will be derived from each aliquot of sample and are usually stored at -80°C at this stage. The omic data that is generated from the omic analysis will then be stored in the central database. Shared data and information on the central database include participants' clinical information, biopsy, radical prostatectomy, biochemistry, participant outcome and information on the collection of the three main types of biological samples: tissue, urine and blood. It should be emphasised that no participants' identifiable information is stored on the central database.

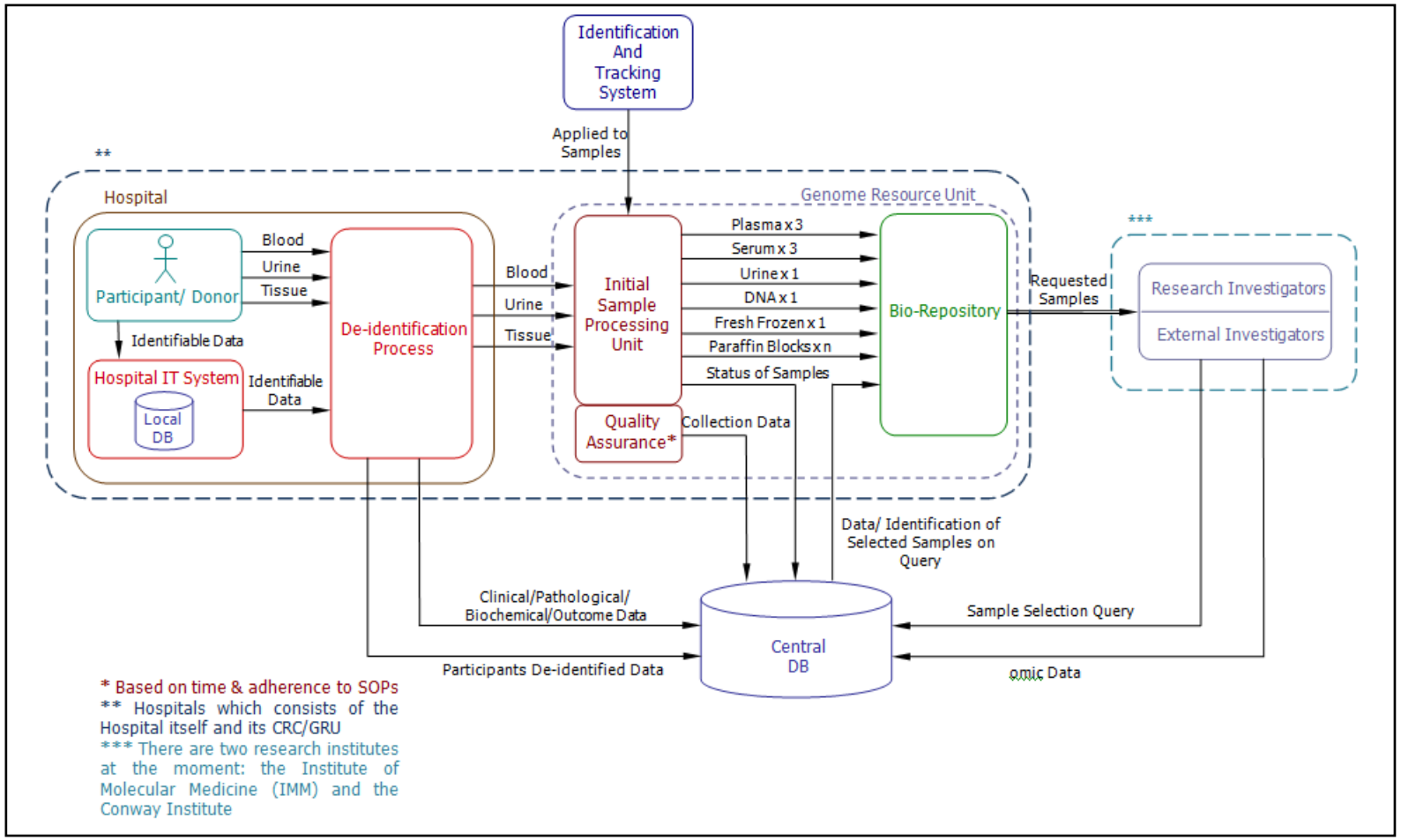


Figure 2-6: The Irish Prostate Cancer Research Consortium BIMS

2.5.2 Sample and Data Flow in the PCRC Biobank

As mentioned earlier, sample processing can be categorised into an initial processing phase, a secondary processing phase, and a long-term storage and retrieval phase.

Three types of samples are collected in the PCRC biobank namely, blood, urine and tissue. Blood is collected in both forms: plasma and serum, from which DNA samples are retrieved. Serum and plasma are aliquotted to three tubes. Tissue samples are collected during surgery and are prepared in two forms: paraffin blocks and fresh frozen tissue. Paraffin blocks are sent to the diagnosis laboratory or pathology for treatment purposes and are not part of the biobank's activities. Fresh frozen tissues are stored in RNA*later*™, from which normal and tumour tissues are dissected. RNA*later*™ is a storage reagent that is used for preserving RNA (Sigma-Aldrich, 2009). Figure 2-7 shows an overall view of sample types and aliquots from collection to storage. Details of procedures are omitted to simplify the diagram.

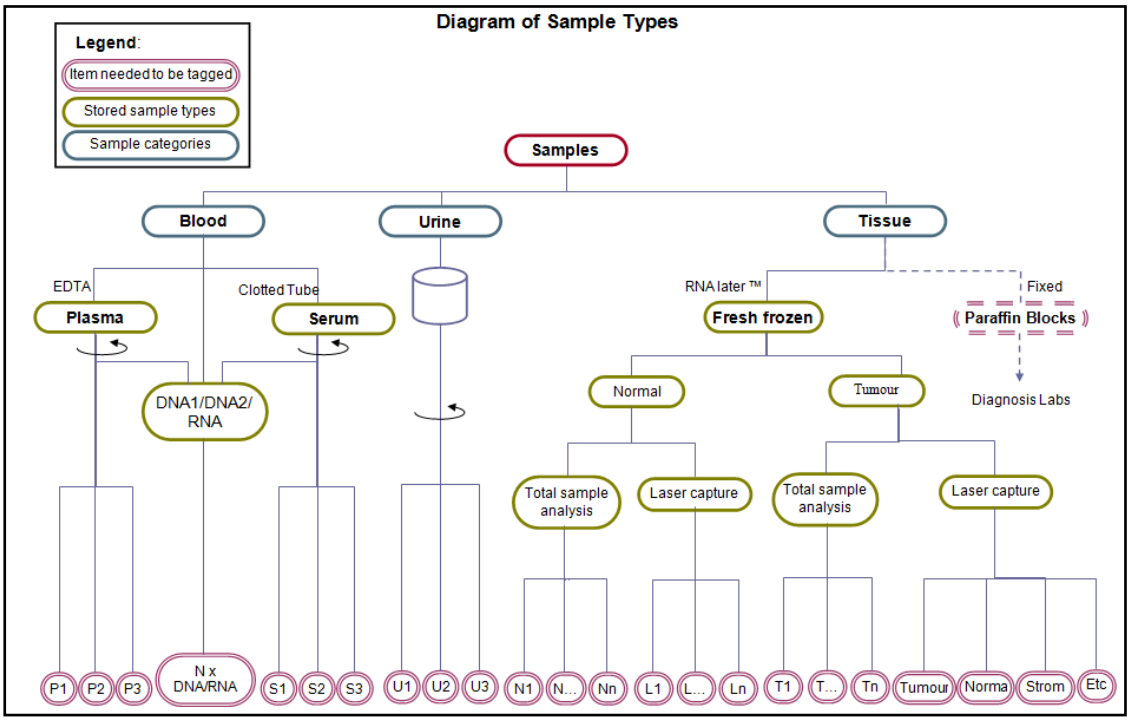


Figure 2-7: Sample Types

The initial processing phase begins in the collection sites and, as shown in Figure 2-8, DNA samples are sent to the IMM, plasma, serum and urine aliquots are sent to the

Conway Institute, and normal and tumour frozen tissues are sent to either, for the secondary processing phase.

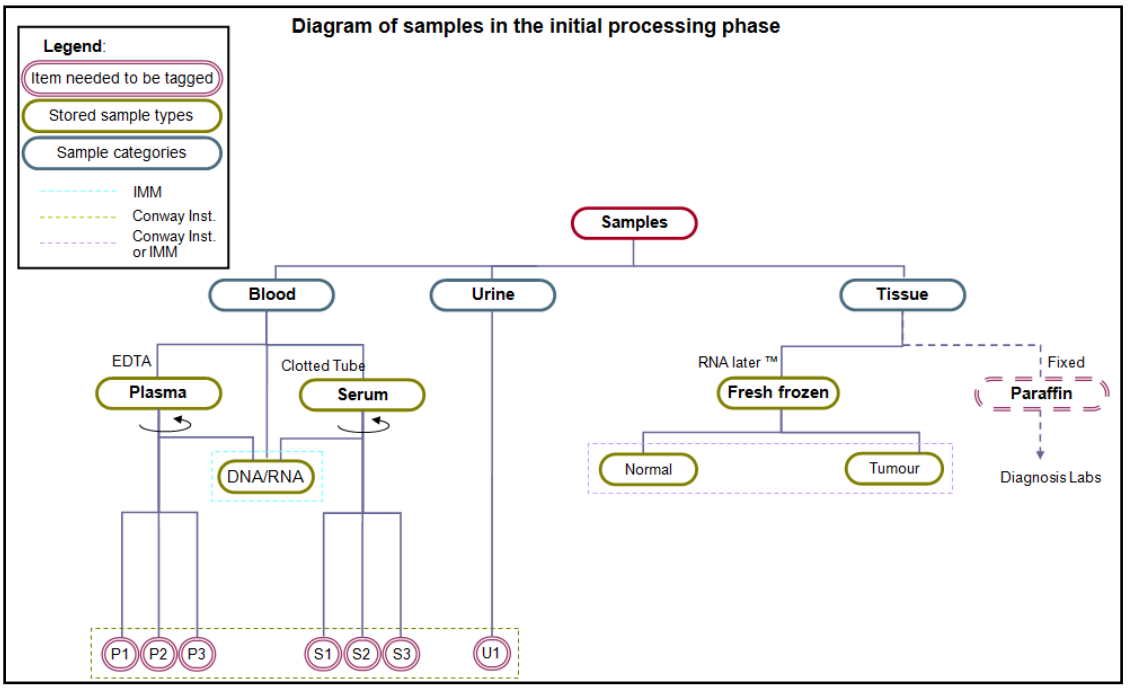


Figure 2-8: Samples in the Initial Processing Phase

The initial and secondary processing phases themselves include a number of operations applied to samples. These operations should be clarified in order to understand the conditions that samples will go through. A system for tracking samples can only be developed after understanding and structuring these techniques and procedures. It is of crucial importance to understand the flow of a sample, how it is transferred from one tube to another, and what each procedure involves in order to map the workflow documents and SOPs into UML sequence diagrams that clarify in a structured way what is involved in each procedure.

The initial processing phase for each sample type involves a number of operations carried out on the samples. The sequence of these operations can be mapped to UML sequence diagrams. For example, the UML sequence diagram for the urine collection procedure is shown in Figure 2-9. This UML diagram shows every step of the procedures that the sample will go through in order to be made ready for use in the secondary processing phase. It is clear from the diagram that the number of steps involved is large and the operations vary from centrifuging the sample to aliquotting it.

Blood samples pass through a more complicated procedure and since a blood sample is aliquotted to three the risk of mixing-up aliquots is higher compared to urine where the sample is not aliquotted until after the first freeze-thaw cycle as shown in Figure 2-9. The situation is even more complicated with tissue samples as they have to go pathology and from there frozen tissues that are preserved in RNA*later*™ are sent to the PCRC biobank. Normal and tumour tissues are then dissected from the preserved samples and each may undergo a different procedure (Figure 2-7). UML diagrams of blood collection in clotting and EDTA tubes, and urine collection are provided in Appendix 1 on the supplied Compact Disk (CD). EDTA tubes are used for collecting blood and have EDTA(K2) and EDTA(K3) as anticoagulants (Frank Healthcare, 2008). As an example the SOP for collection of blood in EDTA tube is also provided in Appendix A.

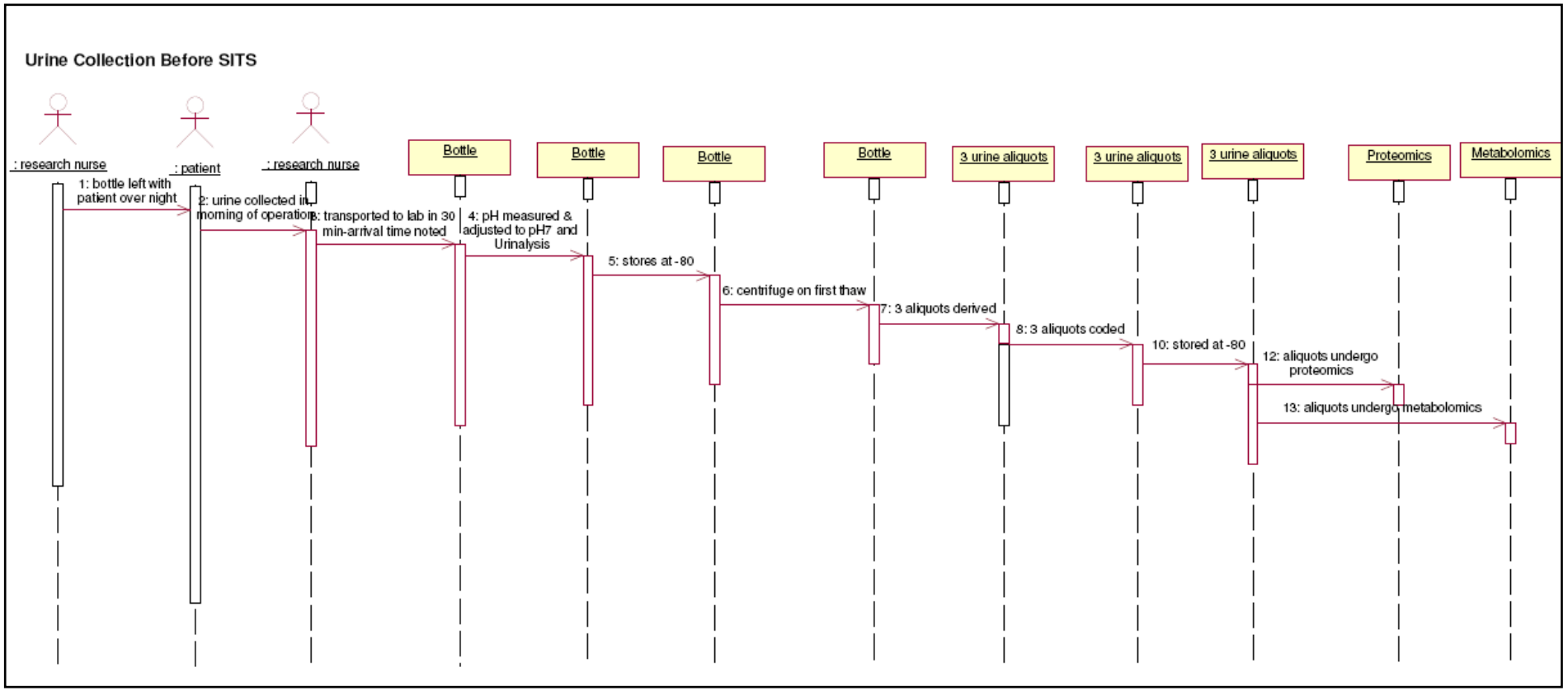


Figure 2-9: UML Sequence Diagram for Urine Collection in the Initial Processing Phase

After completion of the initial processing phase, samples are used in the research institutes for the secondary processing phase. The procedures that samples may undergo in the PCRC biobank include metablomics, proteomics and glycobiology and many more. Workflow documents or SOPs are developed for each of these procedures. Based on the workflow document for metablomics a UML sequence diagram has been drawn to visualise the procedure. The workflow document for the blood metablomics procedure is given in Figure 2-10 and its UML sequence diagram is provided in Figure 2-11. UML sequence diagrams for more complicated procedures such as blood and serum proteomics and urinary DiGE analysis are provided in Appendix 2 on the CD supplied. The workflow documents for these three procedures are included in Appendix B.

Metabolomics Workflow

1. Serum or Urine
Collected from patients using SOP and stored at –80°C

2. Patient groups identified and samples requested from Bio-Resource
We currently request a 500ul aliquot from one of the three 1500ul aliquots stored

3. Sample preparation
Serum (250ul) or urine (250ul) is diluted with D_2O, buffer and standard (TSP) to a total volume of 500ul.

4. Data Acquisition
Sample is run and acquisition of NMR spectra is carried out under optimised conditions. Once run the processed sample is stored at –80°C for additional analysis as required.

5. Data Analysis
NMR data is downloaded from the analyser into an Excel data file for importing into SIMCA P for multivariate data analysis.

Figure 2-10: Workflow Document for Metablomics Procedure

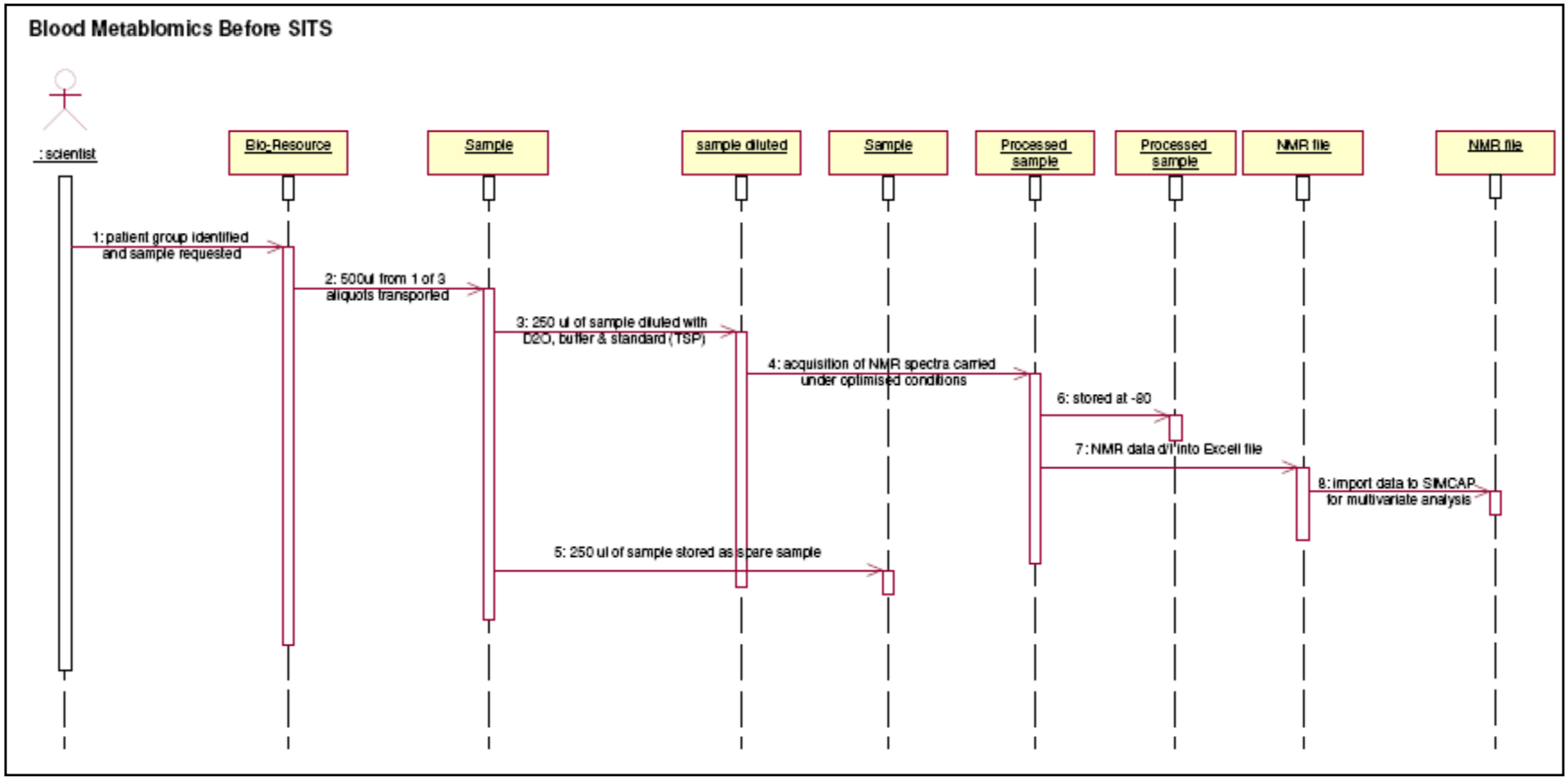


Figure 2-11: UML Sequence Diagram for Blood Metablomics in the Secondary Processing Phase

To reduce the complication of the procedures and provide an overall workflow diagram that covers the initial and secondary processing phases, details of procedures are omitted and workflow diagrams for each sample type have been developed. For instance, the workflow for blood samples is provided in Figure 2-12. As illustrated in this Figure, the initial processing phase involves the collection and aliquotting of serum and plasma samples. DNA samples are labelled as DNA1 and DNA2; although there are no differences, they have been labelled with numbers only to differentiate their sources, i.e., DNA1 and DNA2 are extracted from clotting and EDTA tubes, respectively. This figure also shows that serum and plasma aliquots will be further processed by metablomics, proteomics and glycobiology techniques. DNA samples will undergo methylation and genotyping techniques. The outcomes of each procedure will be a large number of small samples and what is left from the processed aliquot. The small samples and remainder will be stored in freezers for later retrieval and use.

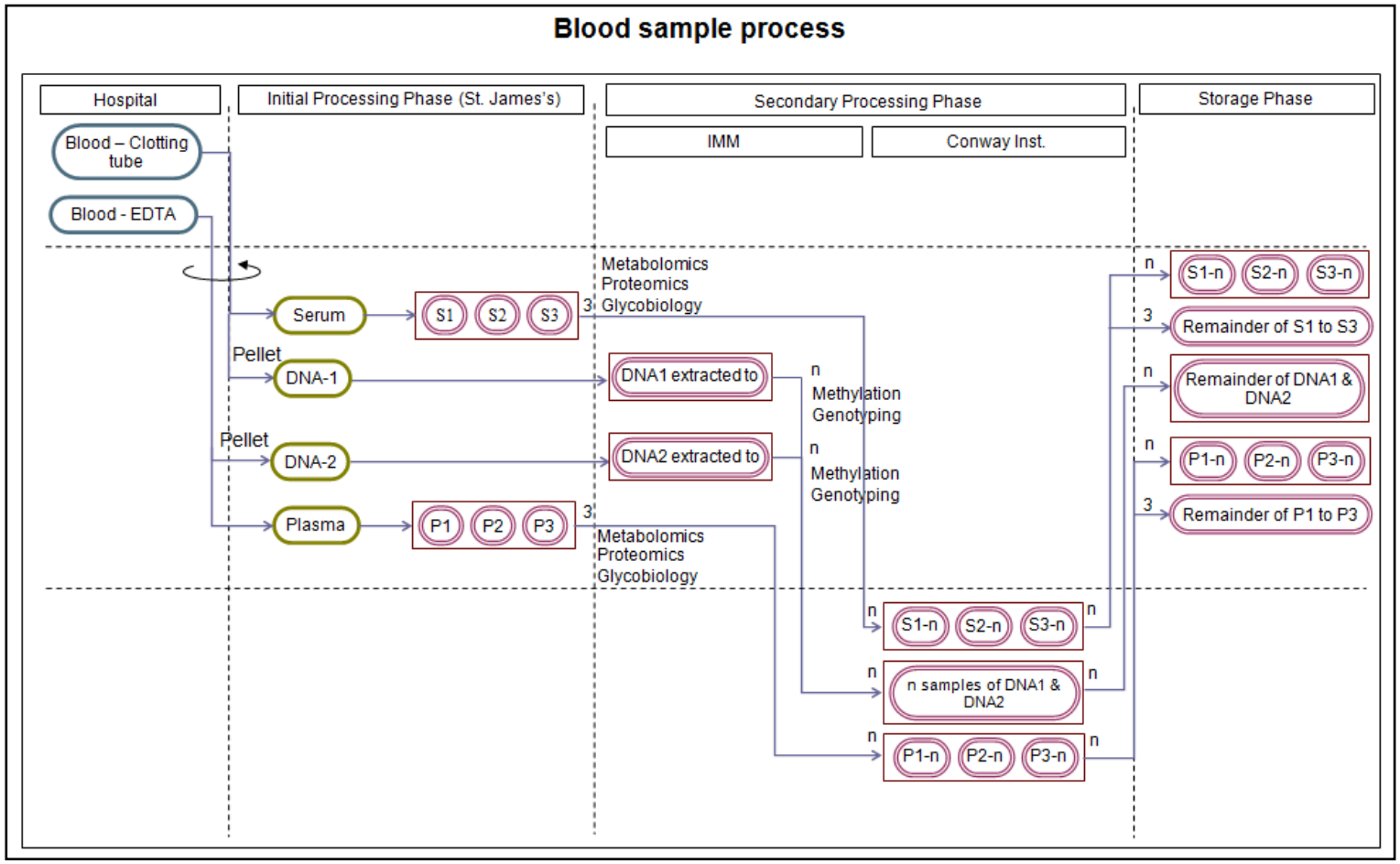


Figure 2-12: Blood Sample Process

Workflow diagrams for tissue and urine samples are provided in Appendix C.

Having described the procedures that samples go through and the complex workflows involved, the importance of having a Sample Identification and Tracking System (SITS) becomes clear. The PCRC biobank uses printed labels that are attached to tubes to identify aliquots. Tracking information, including the name of the institute, the freezer and shelf that the aliquots are stored in within the PCRC biobank, is stored in the BIMS central database.

2.6 Currently Available SITS

SITS is responsible for ensuring that samples are accompanied by related data and information through the workflows described earlier. It enables individuals to locate samples anywhere within the system. The identification and tracking system is applicable both to samples and data, such that data flow is also observed.

A SITS is as an open-loop supply chain (Gou et al., 2008)[1] system that deals with samples that increase in number and decrease in size. In this inventory system, no predetermined path can be identified for samples. It needs to be able to monitor samples across multiple locations and through procedures that vary from one sample to another depending on the type of sample and the desire of the scientists who carry out the procedures. The environment that an inventory system takes place within also varies with the procedure.

Approaches currently available for SITS will be described in this Section.

2.6.1 "System for Tracking Biological Samples"

A patent entitled a "System for tracking biological samples" has been registered (Torre-Bueno, 2007) in which a single or network of laboratories is supported by various data input equipments such as barcode readers, magnetic strip readers, keyboards or any other equipment that facilitates the input of data to a system. This may be located in any

[1] In open-loop supply chain, as opposed to closed-loop supply chain, the users of the end product are not the producers.

area in the laboratory or network of laboratories. Each sample in this system is allocated a Universal Unique Identifier (UUID) that is unique over time and space. The UUID is generated by known methods and algorithms. Samples that are derived from another sample are also allocated UUIDs. UUIDs may or may not carry information about the sample, for example, the location in which the sample was originally collected. The entire system is supported by a central database that houses all data about samples, the tests to be carried out on them, test results, information about how samples were processed and more information that may also be imported from laboratory equipment, devices and UUID readers. This database can be made available to authorised users through access from the laboratory, web, Local Area Network (LAN), Wide Area Network (WAN), Virtual Private Network (VPN) or other methods. This system for tracking biological samples provides features for selecting a therapy for a patient based on the samples and data available for that patient.

When a sample is received at a laboratory or laboratory station, its barcode or other storage medium of its UUID will be scanned by an input device and the UUID used to link the sample with its associated data on the central database. This operation is carried out in the grossing workstation (Thermo Scientific, 2009)[2], where the time and date will be stored on the database to represent the location of samples at that particular time and date. Samples are then sent to the microtome (Microtome, 2009)[3] where the UUID is scanned, and a time and date is stored on the central database. The central database at this stage may assist in identifying the procedures that should be carried out on the samples. Slides of samples may be created and their UUIDs may be assigned and affixed to them beforehand, or when samples are stored on them. For tracking samples, the UUIDs of slides or samples derived from an original sample will be stored on the original sample's record on the database, in such a fashion that only minimum data entry by a user is needed. Sample fixers, reagents and analysers can be scanned to ensure the samples are going through the correct procedure. Also, in case of sample misplacement, an alert or error can be displayed to warn the user. This scenario is

[2] A grossing workstation is a location in a laboratory where procedures are carried out.
[3] Microtome is a device for cutting tissue samples into slides.

depicted in Figure 2-13. Thus, the UUID and the central database allow the maintenance of a complete sample history.

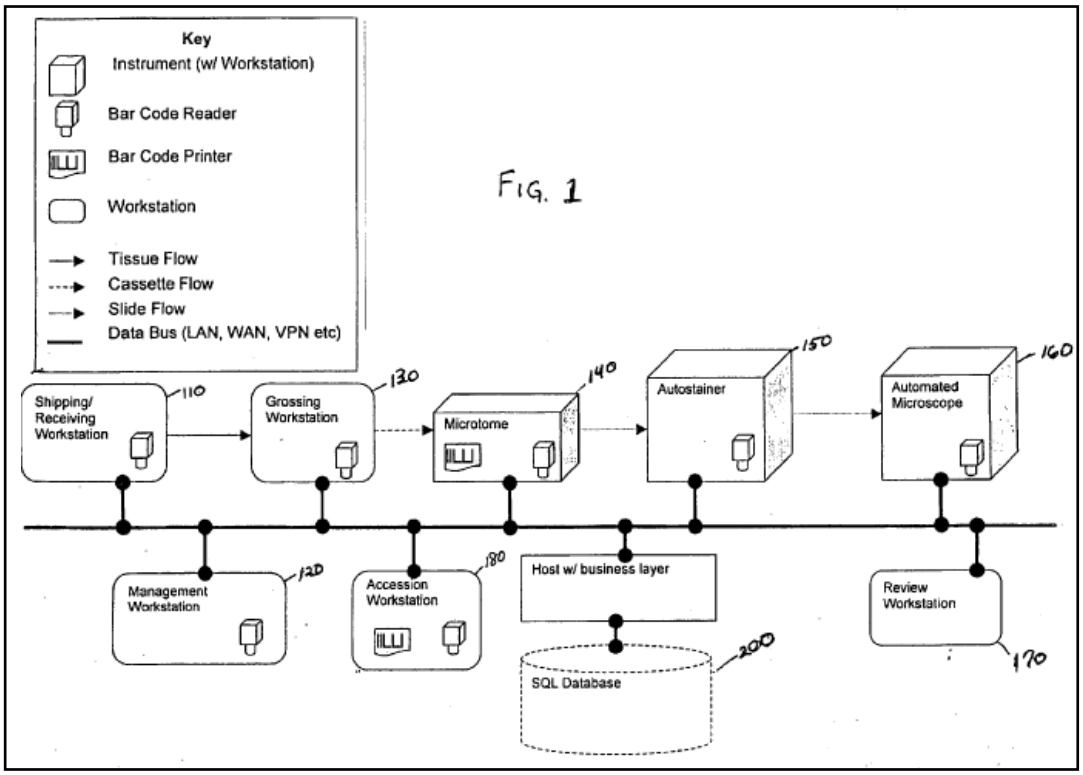


Figure 2-13: Scenario in “System for tracking biological samples” (Torre-Bueno, 2007)

2.6.2 The GHRC Project Approach to Sample Identification and Tracking

In a sophisticated system, flash-memory chips have been used for sample identification, documentation and hierarchical tracking in the GHRC project (Shirley et al., 2009). As part of the GHRC project, this system has been developed in Fraunhofer-IBMT that covers laboratory working areas as part of the hierarchy and provides a well-defined interface between its various levels (Shirley et al., 2009). In this system, samples are stored in containers that are equipped with flash-memory chips that include data about the sample itself, its container and procedures (Shirley et al., 2009). Items at each level of the hierarchy are equipped with flash-memory chips that facilitate identification of the item itself, its associated data, its handling procedures and location, resulting in a polymorphic structure (Ihmig et al., 2009, Shirley et al., 2009). In this way, introducing or adding new items or containers to the system is feasible.

2ml cryovials used for storage of samples are equipped with low-temperature RFID tags and barcodes in addition to the flash-memory chips at the time of manufacturing to prevent loss (Shirley et al., 2009). An image of this cryovial is presented in Figure 2-14.

Figure 2-14: 2ml Cryovial Equipped with Barcode, RFID and Flash-memory Chip (Shirley et al., 2009)

The schematic diagram of the 2ml cryovial is shown in Figure 2-15 (Ihmig et al., 2009). An electronic cryovial consists of five major parts, as labelled in the Figure:

1. Its coloured screw lid that is connected to the body of the cryovial by a ring
2. The body of the cryovial that has room for barcode labels to be stuck on it and the capability of storing samples with volume of up to 2ml
3. Connector of the cryovial to the base containing RFID tag and flash-memory chip
4. Slot for placing the RFID tag
5. Location of the flash-memory chip that can be plugged to the higher level of the system hierarchy, i.e., storage tanks or carrier plates (Ihmig et al., 2009).

Figure 2-15: Schematic Diagram of Flash-memory Chip Equipped Cryovial (Ihmig et al., 2009).

Information about the sample along with its identifier, legal documentation, data format and data about the processing and operating procedures for the laboratory workflow can be stored on the flash-memory chip (Shirley et al., 2009). The electronic cryovials, i.e., flash-memory chip equipped cryovials, are designed such that they can be plugged into the electronic carrier plates or directly into the storage tanks, as the higher level in the system hierarchy. The electronic carrier plates are compatible with standard 96-well-plates (Shirley et al., 2009). At a higher level in the system hierarchy, samples are stored in LN in a rack infrastructure that is designed in four quadrant wings with an aluminium frame (Ihmig et al., 2009, Shirley et al., 2009). The wings of the rack are connected to the backplane in its spine along with a controller that audits flash-memory chips and monitors temperature by its sensors (Ihmig et al., 2009, Shirley et al., 2009). Therefore, the rack is capable of maintaining a history of temperature changes, checking the availability of a sample inside it, and updating data on the flash-memory chip without removing the sample from the rack (Ihmig et al., 2006, Ihmig et al., 2009, Shirley et al., 2009).

There are four vital requirements for the application of electronic devices for identifying and tracking samples (Ihmig et al., 2006):

1. Low power consumption
2. Operational in extreme cold temperatures, LN and surviving multiple freeze-thaw cycles
3. Long life span to cover long-term preservation of data
4. Sufficient memory for storing required data

A two year long study carried out on the performance of flash-memory chips in cold temperatures revealed no data loss or degradation in functionality of the chips (Ihmig et al., 2006). Later on, the operation of 1,000 flash-memory chips has been tested in LN with temperatures ranging between -175°C and -195°C and on more than 50,000 write cycles, i.e. number of times data has been written, which is far beyond the expected 100 cycles (Shirley et al., 2009). The results of the test shows that 90% of the flash-memory chips had less than 200 bit errors (Shirley et al., 2009).

The GHRC system is supported by a Workflow Management System (WfMS) named ChameleonLab (Shirley et al., 2009) based on Extensible Markup Language (XML) (Durst et al., 2005). ChameleonLab is designed such that new processes can be added to it, and to support data and knowledge exchange (Durst et al., 2006). Running queries on the stored samples is facilitated by taking advantage of web services (Ihmig et al., 2009). ChameleonLab facilitates storage of preparation protocols and preparation documentation on the flash-memory chip attached to each sample while it is in long-term storage or in transit (Ihmig et al., 2009) in the form of XML documents (Durst et al., 2005). Definitions of the workflow that each sample should go through, how it should be carried out and its documentation are stored on the flash-memory chip attached to the sample (Durst et al., 2005). This system has a Graphical User Interface (GUI) to support users at each stage of the preparation and processing to ensure they adhere to the protocols (Durst et al., 2006, Ihmig et al., 2009). An evaluation carried out on a ChameleonLab prototype in 2006 revealed that the GUI and the workflow engine need to be customised to match the needs of the users, which in turn will improve system usability and operating time (Durst et al., 2006).

The data structure of the system that facilitates the ease of data exchange between its various levels and modules consists of three major parts:

1. Protected data: for storage of sample identifier and sample-associated sensitive data,
2. User data: for storage of process and workflow data,
3. Log data: for storage of data related to the long-term storage of samples, such as a log of temperature changes (Ihmig et al., 2009, Shirley et al., 2009).

Although the system developed in Fraunhofer-IBMT is a very comprehensive system it has a few limitations. Firstly, to gain full advantage of this system it has to be fully deployed, and deploying such a system will require rebuilding or restructuring major parts of the infrastructure of a biobank. Therefore, although it supports the introduction of new devices and tools, it would be difficult to adapt it for use in an existing biobank. Secondly, the system is designed for storing samples in LN which does not involve frost crystals; no other storage media are supported, for example, conventional freezers. Thirdly, in this system, flash-memory chip equipped cryovials, containers and well-plates are customised and produced for this specific application which increases the cost

hugely compared to the cryovials, containers and well-plates that are widely used elsewhere. The typical biobanks would not have the resources to deploy the solution adopted by the Fraunhofer-IBMT.

2.6.3 Other Biobanks Approaches to Sample Identification and Tracking

Other approaches for sample identification and tracking have also been developed. For example, in the KI biobank (KI, 2007), in addition to the database records being linked with the sample through a unique identifier, a barcode system is also deployed to minimise sample and data mix-up. Approaches taken by the UKBiobank (UKBiobank, 2008), the CPCTR (CPCTR, 2007) and the Paoli Calmettes (Paoli-Calmettes, 2009) are as described below:

The UKBiobank: This biobank has taken advantage of an automated blood aliquotting system from RTS Life Science, to allow reliable tracking of samples when aliquotted to different tubes. This biobank adopted and configured a Laboratory Information Management System (LIMS), Nautilus LIMS^{TM} (Thermo Fisher Scientific Inc, 2009), for storing and managing sample data, and integrating with other data (Schreier, 2008). Linear and 2D barcodes, depending on the type of tubes, are used to label and track samples (Elliott and Peakman, 2008). Nautilus LIMS^{TM} is a product from Thermo Scientific Incorporated that facilitates users with automated plate handling functionalities, and a Graphical User Interface (GUI) for data management that is brought to users through the web (Thermo Fisher Scientific Inc, 2009). This product allows users to define their customised and dynamic workflows, and to create sample lifecycle, and adheres to GLP and 21 CFR Part 11 (Thermo Fisher Scientific Inc, 2009). Title 21 of the Code of Federal Regulations (CFR) in Part 11 of the Food and Drug Administration (FDA) deals with "electronic records; electronic signatures" (FDA, 2000). Since Nautilus provides automated operations with reduced manual human interaction, sample throughput and accuracy as well as productivity of the system are improved (Thermo Fisher Scientific Inc, 2009). Nautilus is an automation product that carries out the procedures in a standardised method. It is not suitable for multi-location, multi-institutional biobanks unless each location and institution uses Nautilus.

The CPCTR: Data gathered in each site of the CPCTR is shared via a comprehensive database, a "virtual tissue bank", including all data related to samples while samples themselves are stored at each individual site (Melamed et al., 2004). Tissue Microarray (TMA) blocks in the CPCTR are barcoded (Datta, 2003). Hence, the CPCTR biobank uses barcode labels to identify samples and link them to their associated data on the "virtual tissue bank" database, on which no identifiable data is stored for confidentiality purposes. With the "virtual tissue bank" most recent data is concurrently available to users. To maintain the standardised operations, the CPCTR pathologists meet every six months to discuss quality assurance controls, and independent quality assurance audits of clinical data are performed at regular intervals (Melamed et al., 2004, Patel et al., 2006).

The Paoli Calmettes biobank: Although the KI biobank (KI, 2007), the UKBiobank (UKBiobank, 2008) and the CPCTR have based their sample management on central databases to house and collate data, and barcodes to identify samples and link them to their associated data record on the database, the Paoli Calmettes (Paoli-Calmettes, 2009) biobank in France has used Radio Frequency Identification (RFID) to store certain data on the RFID tags attached to tubes (Bettendorf et al., 2005). Data stored on RFID tags in this biobank include unique sample identifier, freezer number, date of preserving, and partial information about the sample itself, such as its type, whether it is tumour or normal tissue, viable cells, DNA, or Ribonucleic Acid (RNA). More detailed information about the sample is stored on a database. Data on the RFID tags is protected by unique identification numbers and a password (Bettendorf et al., 2005). Data stored on the tag can be used for sample identification purposes; however, tracking information about a sample is not available. Sample identification by itself is not sufficient for biobanks where longitudinally collected samples and data assist in understanding the progression or regression of a disease. It is important to track samples through the workflow from the collection point to processing and during procedures where there is a high risk of sample mix-up. Keeping a record of the sample flow through its journey would assist in improving the rate of sample loss and mix-up.

2.7 SITS Requirements

Based on the lessons learned from the approaches that other biobanks have taken (Ihmig et al., 2006) and knowledge of the biobanking operations, a number of generic high-level requirements have been identified for SITS, namely:

- As a vital prerequisite, the system must maintain confidentiality of donors which requires that data and samples be kept secure and confidential. For example participants' identifiers must not be disclosed or be made available to any individual without prior permission. Also, samples should only be made available to authorised individuals. Methods of confidentiality maintenance and consent forms will be discussed in detail in Chapter 3. This requirement must be met at both process and software levels in order to incorporate concealing processes and maintain the data securely.
- Fast, reliable and error free item level identification, where items are small biological samples that need special care when handled, is also essential. Biological samples must be reliably matched to their associated data, for instance identifiers, in a timely manner. Identification errors should not occur. Identification and tracking should be carried out very quickly to prevent sample degradation when exposed to room temperature. Identifying a group of samples is desirable in biobanks, where large numbers of samples are dealt with. The goal of SITS is reducing sample and data mix-up and loss, hence no error must occur in the process and the system has to be reliable. SITS activities should be integrated with the workflows. This requirement partially applies to the equipment and hardware and partially to the software that should take reliability into account.
- Data about a sample must accompany the sample at all times. For instance samples must have their identifiers attached to them so that the sample is identifiable at any stage of the process. This in turn requires sufficient space or memory to be dedicated to sample identification data. For example, in the case of barcodes, there must be enough space on the container for attaching the label and in the case of applying electronic tagging, there should be sufficient memory available for storage of data. Also, in the case of applying electronic tags or memories, the sample identifier must be static and cannot be updated or deleted. Prevention strategies should be in place to

ensure that samples have their identifiers attached to them at all times. This requirement is mostly related to the hardware and equipment and the methods by which they are implemented.

- SITS must support sample identification and tracking throughout their journey from collection to processing to storage and retrieval in complex workflows and in various conditions including storage in LN, extreme cold temperature, and number of freeze-thaw cycles. Depending on the protocols and workflows samples undergo various procedures and are stored under various conditions. Temperatures may be as low as -190°C. This requirement must be met by the hardware and the equipment that are going to be used.
- SITS must support long-term storage of samples. For example data must be maintained on the sample for several years. The technology used for identification of the sample must have a long life span to survive long-term storage of samples. Several years are an anticipated storage period in biobanks. This requirement is in association with the storage medium that will be used. It may be in terms of database on the physical media.
- SITS must support storage and retrieval of dynamic data at each phase of the workflow. This should allow collection of standardised and unified data from different institutions and sharing between them. This requirement must be met by the storage medium that will be used as well as the processes.

These requirements are listed in Table 2-2 below.

Code	Requirement
R1	Security and confidentiality of data
R2	Fast, reliable and error free
R3	Item level identification of small biological samples
R4	Data accompanying the small samples tubes at all times
R5	Sufficient storage space available for storing sample data apart from the sample identifier
R6	Support complex workflows
R7	Support storage in LN, temperatures as low as -190°C and multiple freeze-thaw cycles
R8	Long life span sufficient to cover long-term storage for several years
R9	Storage of dynamic data at each phase of the workflow

Table 2-2: Requirements of a SITS

These requirements are expected to be satisfied by incorporating an appropriate technology in the inventory system of biobank. Specified and detailed software requirements will be further explored in Chapter 6. Holland et al. (2003) highlights the importance of electronic data management programs and has mentioned the use of barcodes. Approaches to confidentiality maintenance will be discussed in Chapter 3, and following that, technologies that are available for identification and tracking applications will be analysed against these requirements in Chapter 4.

2.8 Conclusion

The aim of biobanks is to support biomarker discovery and advancements in molecular science by methods of data mining and knowledge analysis. Biobanks maintain collections of different types of samples which, depending on their protocols and needs, may be blood, urine, tissue, DNA and other types of biological samples. Multi-institutional or collaborative biobanks collaborate in order to facilitate larger cohorts of donors and to avoid bias in studies carried out on their samples. In order to use samples

and data from more than one location, they must be of comparable quality and this is done through adherence to SOPs and protocols deployed in each individual biobank. A BIMS is developed to coordinate communication across different institutions and units involved in biobanking. A major component of a BIMS is the database system that houses data, and that supports data communication.

The idea behind biobanks is the long-term storage of samples and data for cohort studies on specific diseases. However, storage and retrieval of biological samples and donor data, such that no sample or data is lost or mixed up, are major challenges by themselves. Biological samples are collected, prepared, stored and made available to researchers and scientists. This can be considered as the initial processing phase. Samples are then further processed by researchers and stored in extreme temperature freezers. This processing is omic analysis or the secondary processing phase. From an aliquot of a sample many smaller size samples will be drawn during the omic analysis. Data generated at each step of the process needs to be maintained for all of the small samples.

Successful biobanks such as the UKBiobank (UKBiobank, 2008), the CPCTR (CPCTR, 2007), the KI biobank (KI, 2007), the Estonian Genome Project (EGP) and the GHRC (IBMT, 2009) were introduced in this Chapter. The structure of the Irish PCRC biobank as a multi-institutional, disease- and location-specific biobank has also been explained. The BIMS of this biobank has been discussed and a need for having a SITS in place has been clarified. Approaches taken by the UKBiobank (UKBiobank, 2008), the CPCTR (CPCTR, 2007), and the KI biobank (KI, 2007) for sample identification and tracking are based on barcodes and databases. Barcodes provide as the unique identifier of each sample that is linkable to a unique data record on the database. The Paoli Calmettes (Paoli-Calmettes, 2009) biobank has deployed a SITS based on RFID. Certain data, including the sample identifier, are stored on the RFID tags attached to the sample's containers. However a system that supports workflows is needed to fulfil the aim of SITS. The GHRC project, in a comprehensive approach, has taken advantage of flash-memory chips being built in to the sample containers in addition to barcode labels and RFID tags being attached to the container. This is a hierarchical system in that each level of the hierarchy is plugged to a higher level. Although the system is very

comprehensive, there are limitations to it, such as the difficulties in its adoption by biobanks that have already established their infrastructure. Then the requirements of SITS at a higher level were discussed. More specific software and technical requirements of SITS will be detailed in Chapter 6.

Chapter 3. Confidentiality

3.1 Introduction

Patient confidentiality and privacy are important issues in the past and present. Medical practitioners have to handle patient biological samples, data, and information learnt during treatment in a highly confidential manner and must not reveal it to a third party unless as part of direct clinical care. Traditionally, doctors took the Hippocratic Oath upon completion of their qualification (MedicineNet, 2009). Treatments patients receive based on these confidential data and samples are considered primary uses of those data and samples. Secondary uses are when these confidential data and samples are used for the purposes of research or clinical trials that may not have a direct effect on the treatment plan of the patient and their use differs from the primary reason of collection. Patients should provide some sort of consent for secondary uses of their data and samples. In the case of research, they are then considered donors or participants who voluntarily take part in a study.

Since biobanks are resources to facilitate clinical trials and scientific analysis, and part of the activities carried out in them are secondary uses of data and samples that are collected for treatment purposes, they have to be handled with special care. Biobanks are based on the voluntary contribution of donors who might not receive any direct benefit. Therefore, it is important to maintain their confidentiality and to make them aware of the process and the purposes of such activities.

Maintaining the confidentiality of donors is a vital requirement for biobanks and therefore SITS, as has been described in Chapter 2, must ensure that confidentiality, privacy and security of data and samples are maintained as prerequisites. SITS must ensure that, while supporting longitudinal tracking of samples and data, and recognising samples from a particular donor, confidentiality of donors is preserved and adhered to.

This Chapter will provide an overview of donors' confidentiality in Section 3.2 with details of confidentiality in healthcare in general, and in biobanks specifically. The three elements of confidentiality in biobanks - consent, anonymity and rules and regulations - are investigated. Consent types, with their advantages and disadvantages as well as

various approaches to anonymity are also discussed. Biobank informatics, how it improves confidentiality, other available biobanks and the approach each has taken towards confidentiality is then discussed. Best practice for ensuring confidentiality in biobanks is provided in Section 3.3. Section 3.4 presents details of how the PCRC biobank maintains confidentiality by discussing the approach it has taken for consenting donors, its data warehouse structure and databases. Finally, how the PCRC biobank meets best practice is provided in Section 3.5.

3.2 Confidentiality in Biobanks

3.2.1 Confidentiality in Healthcare

Medical ethics is attracting increasing attention with the advancement of technology and the discovery of diseases and their causes. While privacy rights must be preserved for every human being, confidential data plays an important role in medical records. Information such as the medical history of an individual might be of interest to his/her employer and it might potentially be a cause for indirect discrimination against the individual (Woodward, 1995). Also, insurance companies often find interesting data in medical records that could potentially be used against the interests of individuals.

If patients trust their physician, and the healthcare system in general, they will feel more confident about disclosing personal information, and this in turn improves the quality of care. Medical practitioners undertake to maintain patient confidentiality and respect their privacy. Traditionally, physicians have taken the Hippocratic Oath at the end of their qualification. By this oath the practitioner swears by all gods and goddesses that

> " ... What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things shameful to be spoken about. ..." (MedicineNet, 2009)

Modern translations of this oath are also available and have been adapted by regulations and laws of countries in terms of security, integrity, privacy and confidentiality. Security mainly concerns data handling methods, both in terms of disclosure to inappropriate individuals and loss of data by deleting what is needed. Security maintenance should be

taken care of whilst designing systems. Consistency of data and its usage is concerned with integrity. The new Swedish law on biobanks refers to integrity as:

> "an interest of individual donors of biological samples to be respected and protected." (Hansson, 2006)

Security allows an individual to access a certain scope of data, and it has been shown that donors are concerned about the security of their data and samples more than the studies carried out on them (Hoeyer et al., 2004, Hansson, 2006). *Integrity* concerns the protection of the donor, their samples and data (Hansson, 2006). *Privacy*, on the other hand, is about the information which may be disclosed to other individuals. *Confidentiality* refers to the disclosure of data and information about someone without that person being recognised from that data. To maintain confidentiality in a system, privacy, security and integrity must also be assumed. Legislation that is applied to the health data of individuals, in order to maintain confidentiality, is based on the country and law of the community. Examples of this legislation will be discussed in section 3.2.2.1.

In its glossary, United Nations Educational, Scientific and Cultural Organisation (UNESCO) has described confidentiality as:

> "the quality of protection against unauthorised accessed to private or secret information"(UNESCO, 1997).

and, in a medical context as defined by the Brain Injury Resource Foundation (BIRF), confidentiality refers to:

> "A principle which states that personal information about others, particularly patients, should not be revealed to persons not authorized to receive such information" (BIRF, 2009).

Although the physicians promise to maintain confidentiality and to respect the privacy of their patients, with the growth of computer-based data management systems in hospitals and healthcare settings, the question of confidentiality and privacy of patients arises (Dodek and Dodek, 1997). Although there are concerns about the possibility of

compromising the confidentiality of an entire set of records in the event that security is breached, electronic records (Dodek and Dodek, 1997), if properly implemented, should be safer and more secure than paper-based records providing much more sophisticated and powerful access control than in possible with paper records. On the other hand, poor implementation or inappropriate definition of access controls and rights might result in allowing one to access an enormous amount of data. Healthcare organisations with computerised records are expected to develop a plan for confidentiality maintenance (Woodward, 1995). A policy for information security in each organisation needs to be developed and implemented. All staff must be fully aware of patients' rights and how they should maintain confidentiality. Computer systems require regular security checks, hence risk assessments and audits should be carried out in a timely manner. Taking regular backups of the entire system is considered vital in any computerised record system so that, in case of interruption to the system, no data is lost or damaged. There are also concerns about ensuring that confidential data is properly destroyed when it is no longer required (Dodek and Dodek, 1997).

Another advantage that computer systems have brought to medical sciences is the ability to analyse massive amounts of data rapidly. It therefore becomes possible to take advantage of data that has been already gathered as part of the care delivery process. Primary uses of samples and data focus on patient treatment and thus are directly beneficial to the patient. Cohort studies, that are carried out on the sample and data collected with the intention of treating patient, represent secondary uses of data. Running queries on clinical data helps to improve cohort studies on diseases and their causes, which in turn might lead to the discovery of biomarkers and drugs that might not be directly beneficial to the patient providing data and samples, but perhaps to patients in the future (Porteri and Borry, 2008). However, there are challenges that should be met before making data available for secondary uses. One of the major challenges that should be overcome is the issue of ensuring donor confidentiality. It is important to know what information should be made available to whom and for what purpose and for how long when performing secondary uses of data (Wallace, 2005).

3.2.2 Confidentiality in Biobanks

Biological samples in conjunction with clinical and pathological data gathered from patients are used in biobanks to support cohort studies. Patients visit hospital or their physician with the assumption that their confidentiality and privacy will be maintained during their visit and that their medical history and data will not be disclosed to any others in the future unless as part of their direct care. In the case where samples are to be made available to researchers beyond what is required for the treatment of the patient, (s)he must be informed. In a phone survey of 504 people living in the US, 65.8% "would require their consent for research on clinically derived, personally identified samples" (Wendler and Emanuel, 2002). Patients' willingness to take part in a study entails preserving their confidentiality. Most biobanks have taken steps towards confidentiality maintenance of patients, who become "donors" or "participants" taking part in the activities of the biobank by providing samples and data.

While biological specimens and donor information are being stored in biobanks, biobanks are responsible for maintaining donor confidentiality such that no unauthorised individual, according to the definitions and terms of the biobanks, has access to these specimens and information. Donors should also be made aware of the methods used to prohibit unauthorised access to their samples and information.

Hansson (2005) describes four principles for establishing trust in biobanks, and hence enhance their value. These four principles are:

1. Samples and data should only be collected after informing the donor and receiving consent, according to an ethical review board;
2. Samples and data should be kept safe through coding whilst being able to link different data about an individual;
3. Rules should be in place to prevent third parties accessing or requesting samples and data; and
4. Any research carried out on samples and data should be approved by an ethical review board based on their requirements.

The first three principles, i.e. anonymity, rules and regulations, and consent, described by Hansson (2005) will be further discussed in this Chapter. Biobank management and governance need to be approved by the relevant healthcare body. In addition, the prerequisite for any study to be carried out in the biobank is obtaining approval from the ethics committee of the research body, sometimes called the Institutional Review Board (IRB) (Godard et al., 2003).

3.2.2.1 Rules and Regulations of Countries

Biobanks are expected to fully adhere to the legislation and law of their community. Since biobanks are considered part of the healthcare sector of a country, rules that are applicable to health data are also applied to biobanks, in addition to potentially more complicated procedures of consenting and anonymity maintenance. Most countries establishing biobanks have developed legislation regarding confidential data management in biobank activities according to the legislation and law in each region.

The Data Protection Act, UK, has developed eight principles to ensure data safety in general healthcare settings. These eight principles are also applied by the Data Protection Commissioner in Ireland (Data Protection Commissioner Ireland, 2007) and are considered the baseline for any action carried out in the healthcare sector. According to this Act, data should only be used in accordance with the purpose of collection and within the same scope of location and time domain. Security and integrity are also taken into consideration by requiring the data to be accurate, adequate but not excessive, processed according to the donors' rights, and stored securely. These eight principles are as follows:

"1. fairly and lawfully processed;
2. processed for limited purposes;
3. adequate, relevant and not excessive;
4. accurate;
5. not kept for longer than is necessary;
6. processed in line with data subjects' rights;
7. secure; and

8. not transferred to countries without adequate protection." (British Medical Association, 2005)

In addition to the above eight principles, the Data Protection Commissioner in Ireland has published guidelines for research in the health sector (Hawkes, 2007). According to these guidelines, if pseudonymised data[4] is not sufficient for research purposes, consent should be sought from the patient. And if the patient has already consented to the use of his/her data for research purposes, then if that consent is still valid, research can proceed with that consent. Otherwise consent should be sought. Figure 3-1 shows the decision making process according to the guidelines. Types of consent and methods of concealing personal data will be discussed in Sections 3.2.2.2 and 3.2.2.3 respectively.

[4] Pseudonymisation involves the use of initials, codes or other replacements to conceal identifying data. More information about this method and other methods for concealing identifiable data will be described throughout this Section.

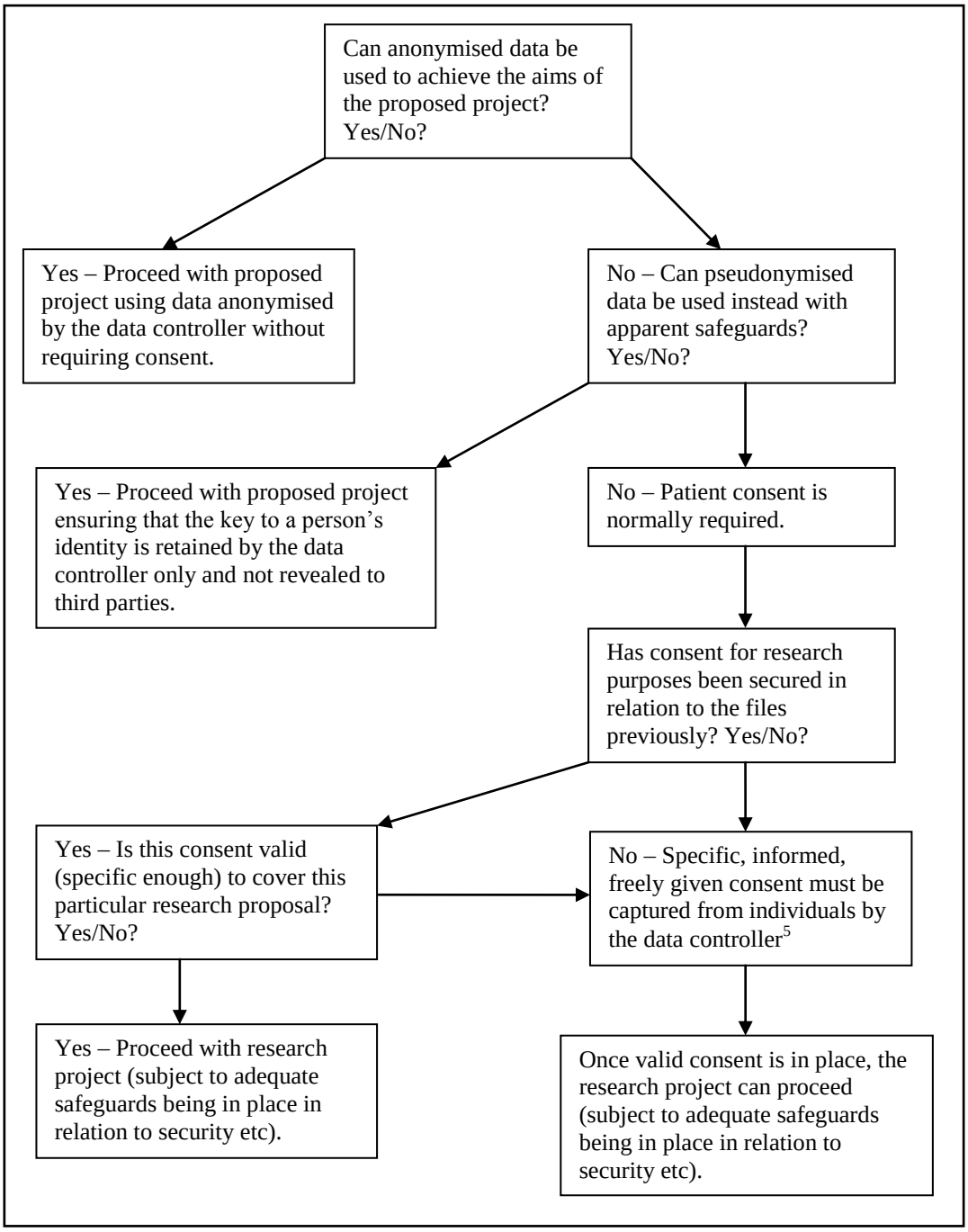


Figure 3-1: Research Projects Approach Involving Personal Data (Hawkes, 2007)

As mentioned earlier in this thesis, Sweden has long experience of collecting samples and therefore has regulations regarding biobanks in place. The Biobank Act in Sweden

[5] The term data controller is being used here in a generic sense; in practice consent would be usually obtained by the clinician or research nurse/nurse practitioner.

requires a specified informed consent to be maintained and well documented for each sample collected. This Act also grants the right of withdrawal to the donors at any time during the process. Any sample, whose donor has withdrawn his/her consent from, must be either destroyed or be fully anonymised depending on the wish of the donor. Swedish law allows samples with inadequate consent be stored for only two months (Hansson, 2006). According to the Biobanks in Medical Act published by the Swedish Ministry of Health and Social Care (Ministry of Health and Social Affairs, 2002), new consent has to be granted for samples that are already collected for other purposes. Donors, who have granted consent for the use of their samples in a particular study, need to be approached again. The Act also mentions that in case the new purpose is for research or clinical trials, the research ethics committee that has approved the study, may also decide on the requirement of information and consent regarding the samples already collected (Ministry of Health and Social Affairs, 2002).

According to Health Insurance Portability and Accountability Act (HIPAA) of 1996 that is in place in the US, all "individually identifiable health information" should be protected, regardless of its medium. However, de-identified data and information that do not reveal the individual's identity can be used or disclosed. Use and disclosure of data and information should be kept within the minimum requirement and no extra data is allowed to be provided (United States Department of Health and Human Services, 2003).

The British Medical Association (BMA) has published guidelines for ethics in general healthcare. According to the guidelines, any research has to be approved by the research ethics committee before being carried out. Safe storage of data and the anonymised or coded identification of personal data is regarded as good research practice. The BMA recommends that donor consent is obtained only after informing donors of the purpose of data collection. Donors must be made aware of their rights before consenting and have the right of refusing to consent. According to the BMA, confidentiality rules apply to all patient identifiable data irrespective of how they are stored and maintained. However, the BMA does not consider the disclosure of anonymous data as a violation of confidentiality (Nathanson et al., 1999).

Comparing rules in the UK, Ireland, Sweden and the US shows that although they all share the common aim of gaining patient trust, they adopt a variety of approaches regarding consent and data concealing.

3.2.2.2 Consent

The value of specimen data and donor information has been emphasised earlier, however, collecting and using this information without the willingness of donors is breaching their confidentiality and is morally unacceptable. Discussions on whether donor consents are needed to be taken when the study has been approved by the relevant ethics committee have led to different decisions being made by different countries and governing bodies. In general, donor consent is required before collecting any data and samples. Many consent types have been defined depending on the freedom that the donor is granted.

Assumed or *presumed consent* is a type of consent in which donors are by default considered to have consented to their data and samples being used for secondary uses. The deCODE Genetics project in Iceland has built a database from the population of the country on the basis of assumed consent (Kaiser, 2002). Donors who are not willing to take part need to opt-out of the database individually instead of being asked if they wish to participate beforehand. Assumed consent is not suitable for the purposes of biobanks as it is not morally appreciated by the patients who, in the first instance, seek diagnosis and treatment. Although given opt-out options, patients might feel disadvantaged in their treatment if they do so. In many cases the patient is unaware of the secondary uses that are carried out on his/her data and samples, so that (s)he cannot investigate if (s)he wishes to opt-out. So, in the case of assumed consent, one would have to make sure that individuals are aware of the existence of the studies being carried out on their samples and data, as well as being aware of their right to opt-out and have the opportunity to do so (Nathanson et al., 1999).

Donors consenting to future unspecified studies without further notification of the purpose and details, are said to have provided *blanket consent*. Blanket consent is a flexible type of consent from a biobank perspective. According to Godard (Godard et al., 2003), this type of consent is regarded as improper by the American Society of

Human Genetics (ASHG) where samples and data are identifiable. Blanket consent is not desirable for the purposes of biobanks as donors need to know what will be done with their samples and data in order to sign the consent form. Since biobanks are based on the voluntary contributions of donors, it is important to make them aware of what they are taking part in.

Hansson et al. (2006) describes broad consent as a consent that is not specific to a study nor as vague as a blanket consent. Hansson et al. ethically validates this type of consent if:

- Personal information of the donor is treated securely;
- Donors have the choice to withdraw their consent at any time during the study; and
- The ethics review board approves new studies to be carried out on the samples as well as any changes made to the biobank.

Implied, also called *implicit*, consent is a consent that is recognised from the donor's action or words. For this consent the donor does not sign a form nor is informed of the study (Shickle, 2006). Since biobanks supply samples and data for secondary uses, implied consent is inappropriate, where the patient is only acting or orally consenting for sample collection for the treatment purposes, i.e. the primary uses.

Under *explicit consent*, the donor is clearly made aware of the details of the study. The donor needs to sign for this type of consent or, if given verbally, an independent observer should witness the process (Shickle, 2006). Explicit consent is used in the UK Biobank (Shickle, 2006); however, other biobanks may find it necessary that the donors be made aware of the details of the study, which might not be done adequately orally. A more strict and documented consenting system is often desired by the ethic committees approving the biobank.

Finally, *informed consent* has been defined by CancerlineUK (2009a) as:

> "A process by which a subject voluntarily confirms his or her willingness to participate in a particular trial, after having been informed of all aspects of the trial that are relevant to the subject's decision to participate. Informed

consent is documented by means of a written, signed and dated informed consent form." (CancerlineUK, 2009b)

Informed consent is widely considered as the most suitable approach for maintaining donors' rights to confidentiality. In this type of consent, donors must be fully informed of the purpose of the study, and of how their data and samples will be used to contribute to current and possibly future studies. Informed consent has to be taken for each and every sample collected for each study. The fundamental idea in participating in research is the willingness and voluntary nature of involvement and, while the research will be dealing with confidential information and biological samples, it is crucial to disclose adequate information to the donors so that they are aware of the advantages and disadvantages of taking part and are therefore enabled to make the most informed decision (Shickle, 2006). The differences between informed and explicit consent, therefore, relate generally to the level of detail about the study provided to donors and the method of indicating assent. In the case of informed consent, donors are provided with detailed written information about all aspects of the study and their consent must be given in writing and formally recorded. With explicit consent, typically less detail is provided and may be given verbally. Explicit consent may also be given verbally, whereas informed consent must always be given in writing.

Porteri and Borry (2008) have proposed two models for informed consent for research purposes, one is suitable for one single research project and the other one is suitable for biobanks. According to their model for one single research project, sample and donor consent must be linked so that in case the donor is concerned about his/her samples (s)he can control or limit the use of them. In their comprehensive proposed model for biobanks, sixteen essential issues have been identified. In addition to pointing out the voluntary nature of the contribution by the donor, issues that they recommend to be included in the informed consent form include details of the biobank itself, e.g. the location; details of the research e.g. scope and duration; risks involved in taking part; confidentiality and privacy maintenance; donor's rights of withdrawal as well as rules and regulations for transferring samples to other biobanks and for case of biobank being unable to afford maintaining the samples or data (Porteri and Borry, 2008). It is important to note that while different consent types carry various level of freedom to the

donor and the researcher, they are not mutually exclusive. For instance, it is possible to have broad, informed consent, which would allow researchers to use the samples and data for multiple studies without having to re-consent the donor separately for each study but which also provides detailed written information about the research programme and formally records consents.

Different countries have adopted different approaches towards consenting depending on many factors. The British Medical Research Council Working Group supports blanket consent so that no further consent be taken from donors for the future research regardless of the study (Godard et al., 2003). Although broad consent has been used in many biobanks, many believe that donors should be contacted again for obtaining consent for each individual study that samples will be used for. Re-approaching individuals whose samples are already collected to obtain a new consent can be time-consuming and cause delays in the study (Godard et al., 2003), while the donor might have passed away or the contact details might be out of date. Many counties, like Sweden, require informed consent and give the donor the right of withdrawal at any point during the study.

3.2.2.3 Anonymity

To ensure donor privacy and confidentiality, data and sample collection procedures are required to be designed such that donor identification is prevented. However, many studies require the ability to collect samples from individual donors longitudinally over time, and these samples therefore need to be linked together with the clinical data of the donor (National Bioethics Advisory Commission, 1999). Furthermore, the scientific value of a specimen or a data set is mostly determined by the clinical, short and long-term outcome data that supports them (Isabelle et al., 2006).

Information such as date of birth of the donor, clinical history and long-term outcome require disclosure of identifiable attributes and are therefore confidential information. Donors need to know how their samples and data will be stored and managed confidentially when they sign the consent form. BMA regards personal health information as:

> "Any personal information relating to the physical or mental health of any person from which that person can be identified" (Nathanson et al., 1999)

and since individuals are identifiable from this information, they should be treated as confidential and some sort of barrier should be put in place to maintain privacy. There are a number of approaches defined for concealing confidential data.

The National Bioethics Advisory Commission (NBAC) has described four methods for concealing donor confidential information for use in biobanks in a guideline it published in 1999 (National Bioethics Advisory Commission, 1999). Also, the American Society of Human Genetics (ASHG) has defined similar methods, as mentioned by Godard et al. (Godard et al., 2003). NBAC and ASHG each label methods differently; however, they are basically the same idea. The four approaches extracted from both resources are as follows:

1. *Unidentified* samples, also called *anonymous*: are samples for which personal identifiable information has not been collected or if collected, is not maintained. Therefore, no confidential data can be retrieved from them.

2. *Unlinked* samples, also called *anonymised*: are stored without identifiers or codes that would link samples to other identified or identifiable specimens or individuals. These samples are supplied to researchers with no identifiers or confidential information. The biobank or "a disinterested party" maintains individual identifiable information but since no code is defined for unlinked samples, linking this category of samples to identifiable information would be a tedious if not impossible job.

3. *Coded* samples, also called *linked* or *identifiable*: do not disclose any individual identifiable information; however they are coded by the biobank or its agent before being made available to external scientists and researchers. The code can be used only by the coding body in order to associate individual identifiable data with the specimens. This approach is referred to as *pseudonymisation* by the Irish Data Protection Guidelines on Research in the Health Sector (Hawkes, 2007).

4. *Identified* samples: are those that are made available to researchers while accompanied by the individual identifiable information such that linking samples with the individuals from whom the specimens are collected can be done by researchers.

Nietfeld developed a terminology (see Figure 3-2) for confidential data concealing to be used in biobanks to clarify the distinction between different terms used in this context (Nietfeld, 2007). Comparing with NBAC (National Bioethics Advisory Commission, 1999) and ASHG (ASHG) terminologies, the *anonymised* category should be added to Nietfeld's classification.

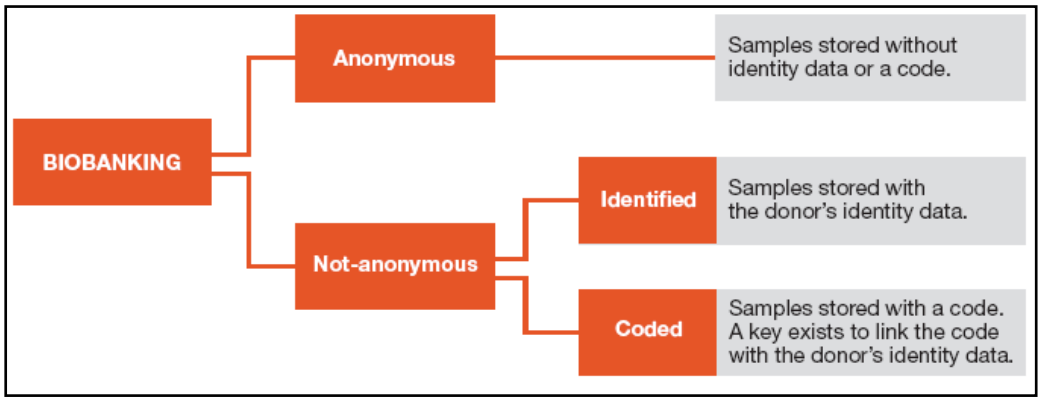


Figure 3-2: Nietfeld's Terminology for Biobanks (Nietfeld, 2007)

The classification presented in Figure 3-3 is more comprehensive than Nietfeld's classification as it covers more scenarios and is more structured than NBAC and ASHG categorisation whilst covering all three of them. There are two distinguished categories for concealing information in relation to *anonymity*: the information that should be fully anonymised and the information that should not be fully anonymised. In the first case, *anonymisation*, samples can be collected with no confidential or donor identifiable information. On the other hand, if the information is collected but is then discarded after collection, the sample is *anonymised* and so are the aliquots taken from it. For the second major category, samples are not anonymised. They are either *de-identified* through coding or are left *identifiable*, which endangers donor confidentiality. The first case is also known as *linked* since samples and data are linkable to the individual through the code. This code is often a single code that links the sample and data to the donor. However, the European Agency for the Evaluation of Medicinal Products (EMEA, 2009) has introduced *double-coded* samples and data to add additional security

to samples and data. It is proposed that these codes be maintained by a third party (EMEA, 2002). It is important to make sure that codes are managed and stored safely such that they are only accessible by authorised individuals. The Medical Care Act requires the Swedish Biobanks to use *depersonalised* or *decoded* samples from the first of January 2003. The key to decrypt the codes should be maintained securely by the collecting body (Ministry of Health and Social Affairs, 2002).

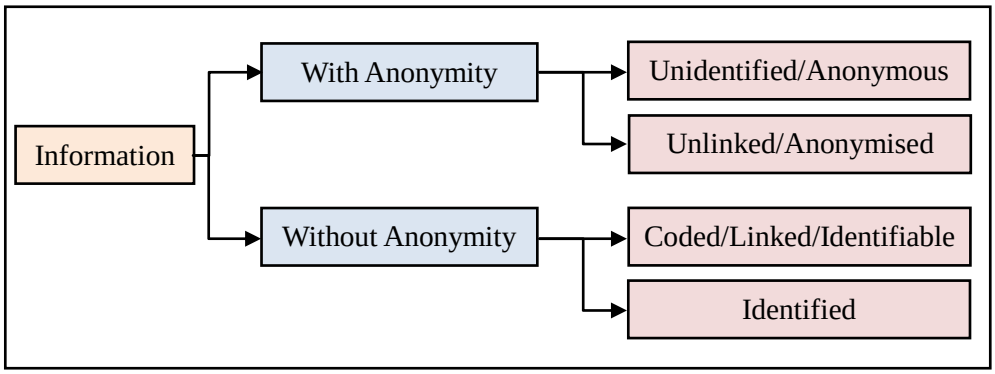


Figure 3-3: Combination of Three Definitions

In addition to the described methods, Ölund has illustrated two approaches towards concealing donor confidential information, namely *exclusion* and *alteration* (Ölund et al., 2007). *Exclusion* of data will lead to fully anonymised data by omitting donor identifiable data; however, *alteration* applies a one-way hashing algorithm to the personal data. This algorithm is not reversible and hence, as with fully anonymised data, it will be a tedious job to track donor samples over time. The code generated by this hashing algorithm will produce the same code for a given identifier, but it has the limitation (or maybe the advantage) of not being able to back-track samples to their donors without accessing the conversion table. Clearly, the conversion table must be kept secure and preferably be managed by a trusted third party. It is essential to be able to identify data and samples belonging to the same donor, as well as being able to select those with similar characteristics and features, hence *exclusion* and *alteration* of donor identifiers might lead to loss of the ability to track the data.

From the given definitions and terminologies, two approaches towards confidentiality maintenance in biobanks can be drawn, *anonymisation* and *de-identification*, also known as *coded*. Through *anonymisation*, all the identifiable information associated

with an individual is fully removed and while it is the easiest to apply, it significantly limits the usefulness of the data for research, specifically longitudinal follow-ups. It also prevents the identifying of samples if a particular donor withdraws his/her consent (if the regulation requires consent for fully *anonymised* samples for subsequent usage or collection). The process of *de-identification* involves removing or replacing identifiable information although the sample is still linkable to the donor through the code. In this process, samples are coded with arbitrary alphanumerical codes that reveal no personal information. Through *de-identification*, confidential information is either removed, or replaced, for example, date of birth is replaced with age (year only) at the time of collection. In contrast to fully *anonymised* data, *de-identification* is reversible by methods of re-identification through the code. It also supports longitudinal studies on samples and data. *De-identification* can be thought of as the most suitable approach for maintaining donor confidentiality in biobanks.

The World Medical Association (WMA, 2009) in its Declaration on Ethical Consideration Regarding Health Databases has defined *de-identification* as:

> "'De-identified data' are data in which the link between the patient and the information has been broken and cannot be recovered"(WMA, 2009).

However, throughout this thesis the term *de-identification* refers to what has been depicted as Coded/Linked/Identifiable in Figure 3-3. It is important to distinguish the difference between the terminology given for *de-identification* in this thesis and the definition given by WMA. *De-identification* here is used to refer to a method of concealing data by using a code to link de-identified data, that is, data that does not include any personal identifiable data, to the individual. So in the *de-identification* approach described here, unlike the definition given by WMA, the link is preserved.

3.2.3 Biobank Informatics

Although IT applications have been used in widely ranging environments, there are concerns about security and access controls of IT systems. IT systems can be very useful for biobanks by allowing data collection, storage, management, query and distribution as well as other supportive facilities they can provide (Schmitz et al.,

2005). These facilities could be delivered in the form of data storage on databases, remote access to the database, sharing data via a database, internet or other types of media, automatic data extraction and de-identification from available health records, cohort data analysis, image processing and many more applications on data and sample handling. However, it is vital to implement such systems carefully, and in accordance with rights and regulations.

Username and password protections, possibly using biometric-based authentication (Li et al., 2006), should be in place for any application that involves data that should not be open to the public community. Databases as the infrastructure of data management and storage for biobanks are of central importance. User accounts and strong passwords should be defined carefully for accessing the database (Dodek and Dodek, 1997, Graves, 2002). Roles and privileges need to be established and based on their necessities, access only to authorised and authenticated users should be guaranteed (Pitt et al., 2005). An issue with IT systems is that, if not properly insulated or roles and rights not carefully assinged, large amounts of data might be revealed to inappropriate individuals, for instance, granting the right to view a certain type of data to a particular role whereby a large amount of data is revealed to a number of individuals who are not supposed to view it (Dodek and Dodek, 1997, Graves, 2002).

An advantage that IT brings to biobanks is providing remote access to a shared database. This eliminates duplication of data and allows concurrent access to the database, which in turn facilitates up-to-date data being presented to the user. Remote access to the database can be made feasible through networks and/or the web (Schmitz et al., 2005). A means of network security should be in place, for instance, firewalls could be set up to enable external access, or a Virtual Private Network (VPN) could be deployed so that only the registered users of the system can access the biobank VPN.

Attempts to develop automatic data extraction from records and automatic de-identification of data are made as part of the IT systems. For example, Uzuner et al. (2007) and Szarvas et al. (2007) have developed such systems. An issue that is often experienced with automatic extraction of information is the content of open text fields, in which the text might not directly name confidential data; however, it might give descriptions that are specific to an individual.

Another application of IT in biobanks, or in laboratories in a more general setting, are sample management applications. These applications vary from barcoding systems and applications that track samples, ensure adherence to SOPs (Schmitz et al., 2005) and link samples to database records (Martin et al., 2007), to applications that locate samples in freezers. Although in most cases sample management applications deal with data records that are already de-identified and stored in databases, it is required that they supply accurate data, that is, no data and sample mix-up should occur. In addition, sample tracking is required for maintaining chain-of-custody of samples during sample processing and data analysis (Martin et al., 2007).

IT applications cover a wide range of operations in biobanks from automatic sample de-identification to data storage and sharing to sample management. However, central to all of these are security and accuracy maintenance, as data leaking will cause breaches of confidentiality, and data mix-up will lead to waste of resources.

3.2.4 Summary of Approaches to Consent and Confidentiality

Different biobanks use different types of consent and approaches for confidentiality maintenance. A list of selected biobanks and their details has been provided in Table 2-1 in Chapter 2. Table 3-1 displays the consent types that these biobanks employ and how donors are assured of their confidentiality. It is clear from the table that while these biobanks de-identify data, they also collect informed consents from donors, except for the UKBiobank that uses explicit consent and the DeCode Genetics biobank that is based on presumed consent and uses data records from the country population to populate its database (Kaiser, 2002). There have been objections towards presumed consent used for this database (Shickle, 2006). However, individuals have the right to opt-out of the study (Nathanson et al., 1999). This raises the question of awareness of individuals that such uses of their data are undertaken.

Biobank	Data Management	Consent Type	Confidentiality Maintenance
UKBiobank	Database	Explicit	De-deidentification
CPCTR	Data stored in Central Data Center Samples stored locally	Informed	De-identification by an “honest broker”
KI biobank	BIMS database	Informed	De-identification
Estonian Genome Project	Database	Informed	De-identification by encoding (Lenk et al., 2007)
DeCode Genetics	Sample Management System (SMS) database	Presumed	De-identification
GenomEUtwin	Database	Informed	De-identification

Table 3-1: Examples of Different Biobanks Consents and Confidentiality Maintenance

3.3 Best Practice for Ensuring Participant Confidentiality in Biobanks

Based on informed consent requirements, private data concealing methods, rules and regulations of different countries and locations, as well as the lessons learnt from the experience of successful biobanks in the US and Europe, key rules and regulation requirements and best practice for ensuring the confidentiality of donor data and samples are summarised below.

1. Each study must have approval from the individual ethics committees of the participating hospital collection sites or the related Institute Research Board (IRB). Methods of collecting and handling sensitive, de-identified data must adhere to national data protection regulations, and the biobank must be registered with the appropriate body.
2. A strategy should be in place to record and manage donor consents. No sample or data is allowed to be collected or stored without first obtaining donor consent.
3. No personally recognisable information, i.e. confidential information, should be revealed or go beyond the hospital system or be seen by unauthorised individuals.

Donor identities must be considered as highly confidential data and shall not be made available to unauthorised parties.

4. SOPs are required to be in place for every task being carried out in the biobank to make sure no additional data is gathered beyond that which is required for the study or consulted by the donor. SOPs for recruiting donors, consenting donors and sample collection, processing, storage and retrieval are required to be developed and adhered to. Samples and data quality controls are met through SOPs. Samples' processing duration and factors that influence their quality, are recorded according to SOPs.
5. Data should be stored in databases that provide appropriate levels of security. Electronic records are more secure in terms of preventing loss and unauthorised access, easier to query and search, and are much easier to manage than paper records. Personal data is stored on machines with limited access, in terms of physical security, i.e. locked offices with restricted access, and in terms of access control, through the use of a login username and password. These machines are either completely disconnected from a network, or are connected to the secure hospital network. In either case, clear procedures are required to ensure these measures are followed.
6. Authorisation and authentication of system users need to be carefully granted. Privileges must be defined carefully and access should be permitted on a strictly "need to know" basis. Guidance from the BMA's Medical Ethics Department defines the term "need to know" as:

   > "... justification applies to the sharing of information necessary to provide care or treatment for an individual patient." (Nathanson et al., 1999).

   Also, there are guidelines available for choosing strong passwords (Burnett and Kleiman, 2005).
7. Operations performed on donor records should be monitored and logged for possible future needs. An audit trail of all accesses should be maintained.
8. It should be possible to trace data and samples back to their donors but this functionality should be restricted only to authorised parties under strictly controlled circumstances.

It is important that biobanks carry out regular review of security and data protection procedures to ensure that best practice is being maintained throughout the lifetime of the biobank. Also the best practice will need to be updated as new technologies emerge.

3.4 Confidentiality Maintenance in the Irish PCRC Biobank

The PCRC biobank system has been reviewed and approved by the Irish Data Protection Commissioner with regard to the criteria defined for donor confidentiality.

3.4.1 Informed Consent

Before collecting any sample in the PCRC biobank, an information leaflet is provided to donors by the dedicated research nurse who is available to address all questions. Obtaining consent and collecting samples is done on an individual basis. In the information leaflet details of the study including purpose, information that will be collected, reasons for choosing this particular person and who is organising the study are provided. More information about the process, advantages and disadvantages, and how their confidentiality is maintained are also provided to the donor. They are also made aware that this is only a voluntary participation and they will not be treated differently if they do not wish to take part. Donors are assured of their anonymity and are given the right of withdrawal at any stage during the study. In case of withdrawal, all of their specimens are located and destroyed. An informed consent form is signed once they agree to participate. The information leaflet and the informed consent form are provided in Appendix D. In the event that a donor becomes incapable of making decision or dies, a close relative is in charge of making decisions regarding on-going consent.

3.4.2 The Irish PCRC Biobank Data Warehouse

The approach that the PCRC biobank has taken towards data de-identification and confidentiality maintenance is like the CPCTR approach (Patel et al., 2006), where personally identifiable data is stored only at the collection sites on the hospital IT system. However, the role of “honest broker” in the CPCTR is filled by the research nurses of each collection site, who unavoidably are aware of the donor’s identity at the time of sample collection.

3.4.2.1 Samples and Data Coding

In the PCRC biobank each collection site has introduced its own unique identifier to distinguish donors, samples and data at the local level. This identifier is called the Institute Study Number (ISN) consisting of an abbreviation of the name of the institute followed by an arbitrary number unique to the donor. The ISN is only suitable for identification within the domain of the corresponding institute. Since the ISN reveals the collection site and hence the hospital at which the donor was treated, a unique PCRC Study Number is used to prevent recognition of donors within BIMS. The ISN is regarded as confidential data as it could assist in identifying individual donors in the event that additional information, such as date of surgery, is known. Hence the ISN is replaced by a globally unique study number prior to transmission to the central repository. The PCRC Study Number is used as a global identifier of donors within the Irish PCRC biobank and is the key identifier which can be mapped back to the original donors' data via the local database of hospitals. This identifier is generated on a random basis and will be used by the identification and tracking system as the unique identifier.

3.4.2.2 Double Layer Database System

To maintain donor confidentiality in the PCRC biobank, the CRC at each collection site maintains a local database for storing identifiable donor data. Information, such as date of birth, name and medical record number of donors, is only stored in the local database of the collection site. These databases are standalone and are considered as part of the hospital's IT system with all the appropriate data protection and firewalls to protect the data. Local databases are only accessible by the research nurses at those sites who are already aware of donor's identity. As employees of the hospital, they are bound by all the hospital's confidentiality rules and regulations. Local research nurses at each of the sites maintain the mapping between local and global identifiers, thus enabling linkage of individual donor records over time. So, the identity of the donor is known only to the research nurse who is responsible for collecting the initial samples, clinical information and all subsequent follow-up data.

The central database receives data from the hospital or its local database via the manual data input done by the research nurse and allows sharing de-identified data with the

research institutes such that no donor identifiable data is stored in the central database. Figure 3-4 depicts the situation.

The data warehouse architecture of this biobank maintains a double layer security system to protect donor data. The results of omic techniques and other data from the research institutes are stored in the central database and are made accessible to the cooperating bodies in the biobank. Depending on predefined privileges, users have the right of editing the data. The PCRC biobank maintains a complete audit trail of all data accesses to the central database. These logs can be used to track back and identify unauthorised access to the database as well as to monitor operations performed by each account holder.

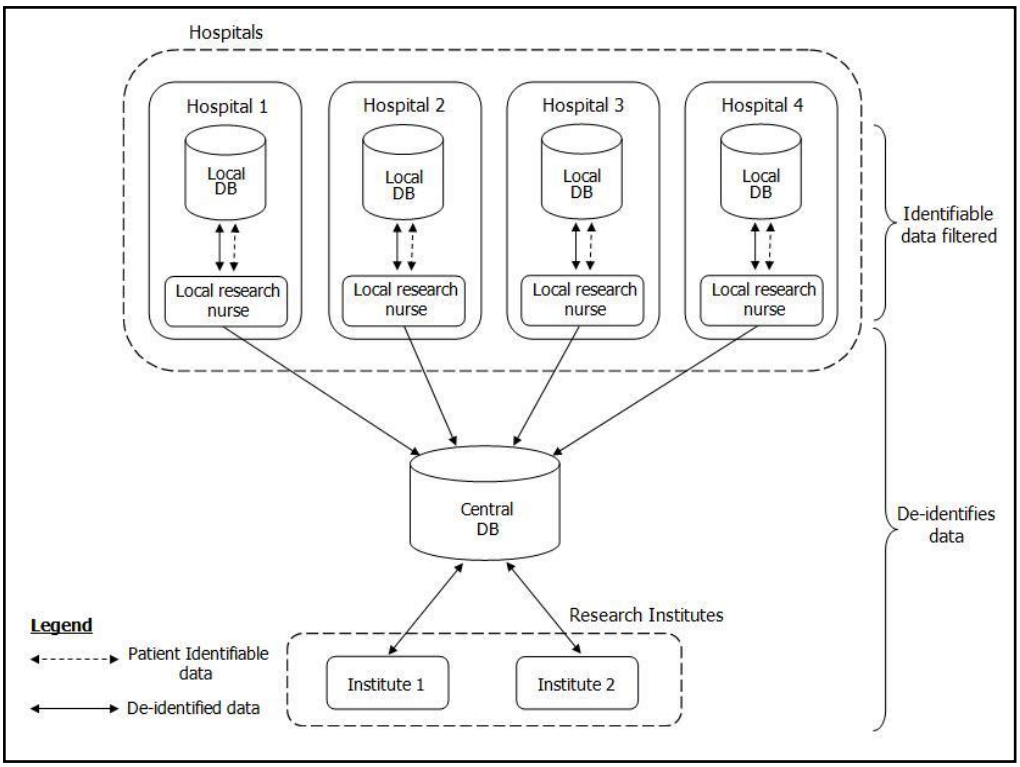


Figure 3-4: The Site Plan of Irish Prostate Cancer Research Consortium Biobank

Participant identifiable data are entered manually by the research nurses into the local database at each site. They then manual enter the de-identified data into the BIMS central database. While this approach offers the advantage of added security by providing no automatic linkage between the local database and the BIMS thereby preventing leakage of confidential information, it has the disadvantage of potentially

introducing transcription errors. When the PCRC biobank was established security at the hospital sites was considered of paramount importance and permission to link the PCRC BIMS to the hospital network would not have been given. It is possible that over time this may change and it may be possible to provide automatic secure transfer of de-identified data to the BIMS thus improving efficiency and eliminating the possibility of transcription errors, while ensuring confidentiality and security.

3.5 The Irish PCRC Biobank Adhering to the Best Practice

The implementation of the BIMS by the method described in this Chapter results in a secure and fully confidential online query system which can be made available across a shared network of cooperating biobanks. This is only possible when a globally standardised method of data collection and storage is made available and every biobank follows identical SOPs, so that data are of comparable quality and hence more reliable and valid biomarkers will be discovered.

In order to satisfy the confidentiality requirements established earlier in this Chapter the PCRC biobank has taken actions as below:

1. Studies that are carried out on the Irish PCRC biobank samples and data are approved by the individual hospital ethics committees and the Data Protection Commissioner, Ireland. Some of the criteria taken into consideration in the approval process are consent management, role based access to data, SOPs, audit trail facilities and separation of confidential data.
2. SOPs regarding consent and sample collection have been developed. Clear and unambiguous information leaflets and consent forms are made available to the donors before any sample is obtained. Donors are granted the right of withdrawal of their consents. The Irish PCRC biobank does not fully anonymise data and samples so they can be removed if required by the donor.
3. Donor confidential data is stored on a separate system at the collection site and is only available to research nurses who are already aware of the donor's identity and are bound by the confidentiality agreements made within the hospital work place.
4. Dedicated SOPs for all procedures from consent management to sample collection, storage are developed and adhered to.

5. Central and local databases have been developed and facilitate a double layer of data protection. Both layers are approved by the associated bodies. The local databases which contain personal information are on very limited swipe card access and only to biobank staff who are already aware of donor identities due to interactions at the time of collection. Also, the central database is designed to minimise accidental upload of data, e.g. the use of open text fields is reduced. A firewall is put in place to block access to the BIMS by unauthorised machines on the internet. The proxy server is only available to certain predefined IP addresses.
6. Roles and privileges for each user are defined for each database. The PCRC BIMS operates on Certificate Authority. After the user submits an application form to access the PCRC BIMS and gets approval from the biobank, a Client Certificate is sent to him/her by email. The password, which allows insertion of the Client Certificate to the browser, is then disclosed over the phone. The Client Certificate is only valid for one year and only works on the particular browser designated. Users are required to use desktop machines for browsing BIMS.
7. Logs of central database operations are maintained with details of all operations carried out by users.
8. Tracking back data to its donor is only possible via the local database which in turn is available only to authorised personnel.

Furthermore, the PCRC biobank is reviewed for accordance to best practice every six months.

3.6 Conclusion

Confidentiality maintenance, being a vital requirement for any healthcare system, is also of crucial importance in biobanks that are developed to facilitate secondary uses of samples and data.

Studies undertaken in biobanks must be approved by the associated ethics committee before commencement. Donors who wish to take part in the approved studies will need to provide some sort of consent for the secondary uses of their samples and data. Although a number of consent types exist, the one that best meets the requirements of confidentiality maintenance for the purposes of biobanks is informed consent. With

informed consent, donors taking part in the activities carried out in biobanks must be fully informed of the details of the study. During consenting donors, they need to be made aware of how their privacy and confidentiality will be satisfied.

A set of best practice guidelines has been developed for confidentiality maintenance in biobanks. This best practice is based on the literature review and the regulations in place.

The PCRC biobank has deployed a data warehouse structure for its BIMS database, where personal data is stored in a separate local hospital database that is linked to the central database by a study number. This database is maintained and accessed locally. This allows the individual hospitals to follow-up the donor's outcome and populate the central database with emerging data while the users of the database are not aware of the identity of the donors, i.e. only de-identified data is stored in the central repository. Thus the central repository and the local databases at the collection site comprise the biobank information system. In this way the PCRC biobank adopts a double layer approach which employs exclusion of personal data from the central database used by the researchers.

Meeting and adhering to the best practice is part of the requirement for any system that supports biobank activities, thus BIMS and therefore SITS, that deal with samples and data that are collected from donors, must meet these requirements.

Chapter 4. Technologies for Identification and Tracking

4.1 Introduction

In addition to confidentiality maintenance covered in Chapter 3, further requirements of SITS drawn in Chapter 2 must be accomplished by deploying a technological infrastructure that is supported by IT. Technologies that are available for identification and tracking of other objects in different environments can be applied to SITS for the identification and tracking of biological samples. One of these environments is supply chain management. Automated identification of each physical product in the domain of supply chain management improves tracking and tracing of products and inventory management (McFarlane and Sheffi, 2003). Examples of technologies allowing automated identification of products are barcode, RFID (McFarlane and Sheffi, 2003) and Zigbee. In addition to barcode and RFID there are other emerging technologies such as Hewlett-Packard (HP) Memory Spot (Alto, 2006) and Intel® Mote (Intel Corporation). More traditional approaches such as magnetic stripe and Optical Character Reading (OCR) are also available.

In order to apply the most suitable technology for the purposes of each individual application, it is important to draw a comparison between available technologies and to identify the one that most satisfies the system requirements, as for most applications there is no single technology that satisfies all of the requirements. Also the future developments of the system must be considered. The feasibility of each technology should be validated in the biobank environment. HP Memory Spot, Intel® Mote, magnetic strips, OCR, Zigbee, barcodes and RFID are potential candidates for the infrastructure of SITS.

Technologies available for identification and tracking in general will be discussed in Section 4.2. Based on these details, competing technologies that best satisfy the requirements outlined in Section 2.7 will be discussed in Sections 4.3 and 4.4. Then the most appropriate technologies will be compared in more detail in Section 4.5. Finally the Chapter is concluded in Section 4.6.

4.2 Technologies Available for General Identification and Tracking

In this section, technologies that are applicable to identification and tracking systems are introduced and their functionalities with regard to the requirements listed in Section 2.7 and Table 2-2 are evaluated. These technologies include HP Memory Spot, Intel® Mote, magnetic strips, OCR, Zigbee, barcodes and RFID technologies.

4.2.1 HP Memory Spot

HP has introduced a chip called "Memory Spot" due to its size being about 2 to 4 square millimetre with a memory capacity of 256 kilobits to 4 megabits (Genuth, 2006, Alto, 2006) in 2002 as the result of research carried out in HP research laboratory in Bristol UK (Genuth, 2006). The development of the Memory Spot was aimed at attaching sound to still pictures in a simple and cost effective way (Genuth, 2006). The data transmission rate for memory spot is about 10 megabits per second; however the reading range is very small about 1 millimetre (Genuth, 2006), making the technology particularly suitable for applications dealing with a large amount of data where the close reading range is not an issue. Furthermore, the short reading range improves the security (Genuth, 2006). The short reading range is due to the absence of a battery on the circuit, which requires the chip to receive power from its special reader device (Genuth, 2006). The six components of HP Memory Spot include its processor, its memory and memory driver, its modem, the capacitor array and the antenna as shown in Figure 4-1 (Genuth, 2006).

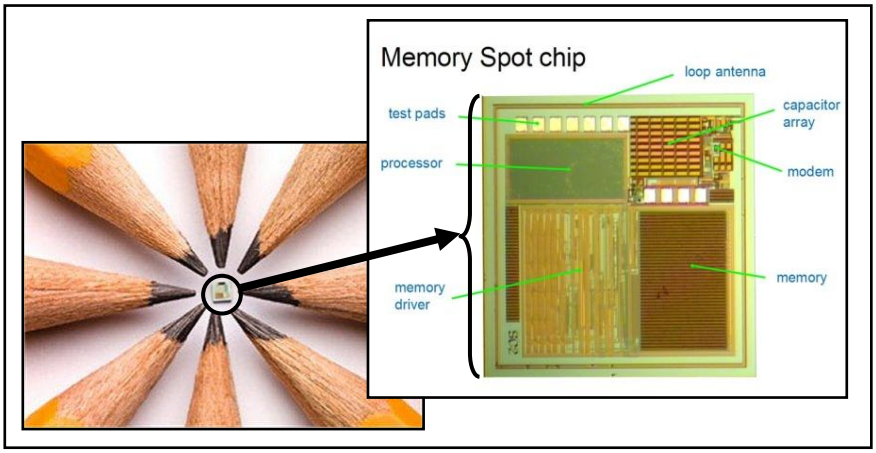


Figure 4-1: Components of HP Memory Spot (Genuth, 2006)

4.2.2 Intel® Mote

A technology introduced by Intel® is the Mote (Intel Corporation). The Intel® Mote is the result of a collaborative project between the University of California Berkeley, the Intel Research Berkeley Laboratory and Intel Research Seattle (Intel Corporation). These motes can be thought of as a sensor node platform (Kling et al., 2004) or as the basis of wireless sensor networks (Intel Corporation) that is made possible by adding network layers to the TinyOS to allow "multi-hop" functionality based on Bluetooth (Nachman et al., 2005). TinyOS and Linux are the operating systems running on the Intel® Mote (Adler et al., 2005, Nachman et al., 2005). This technology includes a battery onboard and is a self-contained computer that links with other Motes through radio communication (Intel Corporation). An Intel® Mote is about 3x3 cm and is shown in Figure 4-2 (Nachman et al., 2005, Intel Corporation). This mote is an integration of wireless microcontrollers such as ARM 7 core, and Bluetooth radio, RAM and Flash memory in addition to I/O options that can be stacked connected to the mote, such as UAER, 12C, SPI, USB (Kling et al., 2004, Adler et al., 2005). It is intended to develop the Intel® Mote in the form of a single microchip (Intel Corporation).

Motes are designed for creating ad hoc wireless sensor networks amongst themselves and are suitable for applications such as health monitoring, industrial control and military applications (Intel Corporation, Nachman et al., 2005).

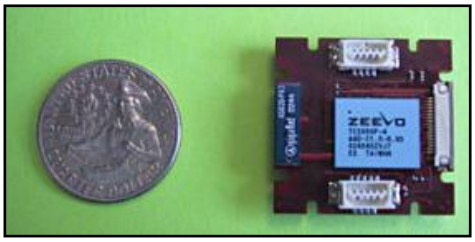


Figure 4-2: Intel® Mote Size Compared with a Coin (Intel Corporation)

4.2.3 Magnetic stripe

With magnetic stripe technology data are stored on a thin plastic film and the orientations of small magnetic particles on this film encode bits of data (Groover, 2007). Any damage to the film will cause loss of data. Films damage easily in laboratory environments when samples get wet and are processed in magnetic intense

environments. Furthermore, data can only be read when they are in direct contact with the scanning equipment (Groover, 2007). However, an advantage that this technology could offer to the SITS is that data stored on the film can be updated.

4.2.4 Optical Character Reading

Optical Character Reading (OCR) technology uses an optical reading device to read alphanumerical characters (Groover, 2007). The nature of OCR operation is based on line-of-sight. The error rate of data entry using OCR for a 12-character piece of data is 1 character error in 10,000 characters entered and it takes 4 seconds to read this 12-character piece of data (Muller, 2003).

4.2.5 Zigbee

Zigbee is a set of communication protocols for short-range wireless networks with low data rate (Farahani, 2008). There are three roles defined for Zigbee devices: *Zigbee Coordinator*, *Zigbee Router*, and *Zigbee End Device* (Farahani, 2008). Zigbee has many benefits such as while being battery-based it is low cost and has long battery life (Farahani, 2008).

4.2.6 Barcode

Barcodes are a series of printed bars with free spaces in between and are available in forms of linear and 2 Dimensional (2D) (Frazelle, 2001). Various types of barcodes use different patterns to represent data by dark bars and spaces. 2D barcodes include two barcodes, one horizontally and one vertically (Frazelle, 2001) and therefore are capable of storing more data in a smaller area compared to linear barcodes. Barcodes are readable by optical devices that read the barcodes based on the reflection of light from dark bars and spaces (Muller, 2003). Different coding styles are available for both linear and 2D barcodes such as code 128 for linear and PDF417 for 2D barcodes (Muller, 2003, Frazelle, 2001). Although barcodes do not support storage of dynamic data, they allow fast reading speed, 0.3 to 2 seconds for a 12-character long data, and the error rate is 1 character in 15,000 to 36 trillion characters entered (Muller, 2003). Also attaching the barcode labels to the tubes may be feasible depending on the coding style used.

4.2.7 RFID

RFID technology is based on the principles of transmission of data through radio waves and was first introduced by Stockman (1948). The three main components of RFID are the antenna, a transceiver also called reader/writer device, and transponder also called a tag (Domdouzis et al., 2007). To read a tag, data are transferred in the form of radio waves to the reader/writer device antenna and to the device. To write on a tag, data are sent to the antenna from the reader/writer device and to the tag through radio waves. All three components must comply with a certain technical specifications in order to make communication feasible. RFID tags can be passive, active or semi-passive depending on the power source they use (Roberts, 2006).

4.2.8 Analysis of Technologies Based on SITS Requirements

HP Memory Spot, Intel® Mote, magnetic strips, OCR, Zigbee, barcodes and RFID technologies should be analysed according to the requirements described earlier in Section 2.7. These requirements are:

R1. Security and confidentiality of data

R2. Fast, reliable and error free

R3. Item level identification of small biological samples

R4. Data accompanying the small samples tubes at all times

R5. Sufficient storage space available for storing sample data apart from the sample identifier

R6. Support complex workflows that do not follow a predefined path

R7. Support storage in LN, temperatures as low as -190°C and multiple freeze-thaw cycles

R8. Long life span sufficient to cover long-term storage for several years

R9. Storage of dynamic data at each phase of the workflow

HP Memory Spot:

Although the amount of the memory that HP Memory Spot provides make it suitable for applications involving large amount of data, it has a very short reading range that makes

it unsuitable for SITS. With a 1 millimetre reading range, the reader device should be almost kept at touch distance of the Memory Spot to read the data. Furthermore, this technology is intended to be attached to paper based documents and its functionality in temperatures as low as -190°C for long-term storage would need to be evaluated. In correspondence with a member of HP it was ascertained that the Memory Spot consists of a Complementary Metal Oxide Semiconductor (CMOS) chip and it is the chip which would determine the operational temperature range. A standard CMOS chip is operational between -55°C and +125°C (Edwards, 2008). While the Memory Spot is a research prototype, it has been tested for mechanical effects of multiple freeze-thaw cycles on the chip. The result of the test was that freeze-thaw cycles have no mechanical effect on the chip. However, the Memory Spot was only tested below -55°C, whereas the SITS must operate at temperatures as low as -190°C. Therefore due to the short reading range and the fact that it cannot operate at the temperature needed in biobanks, the HP Memory Spot will not be further investigated for SITS application.

Intel® Mote:

Although Intel® Mote has many advantages such as having sufficient memory to store large amounts of data, it is not suitable for SITS due to its physical size. Attaching an Intel® Mote to sample containers with its current size of 3x3 cm is not feasible. Also, its security, survival at extreme temperatures and various other parameters would need to be evaluated. However, this option for deploying in SITS is rejected merely because of its large size.

Magnetic strips:

Although magnetic strips support storage of dynamic data, they are not suitable for SITS. The issues of direct readability that requires the strip be “swiped” to the reader as well as vulnerability to physical damage make it unsuitable. So no further investigation will be carried out on this technology.

OCR:

With OCR, data must be printed on a label that is readable by the naked eye. This reduces the security and breaches confidentiality of data. Also SITS requires a large

amount of dynamic data to be stored on the sample tubes and containers. It is not feasible to machine print or hand write a large amount of data on a label attached to the tube. Also these data are static and cannot be updated. This technology is omitted from the options available for SITS.

Zigbee:

Long battery life has made Zigbee particularly interesting for SITS. However, the fact that Zigbee is intended for network applications, it will not be considered for further investigation.

Barcodes:

Depending on the coding style, large amount of static data can be stored on barcodes. Data stored on barcodes are not readable by the naked eye and therefore support security and confidentiality. Although barcode scanners need line-of-sight in order to read labels and only static data can be stored on them, they are a potential candidate to be used for SITS application due to their long life span, survival in various conditions and low error rate. Furthermore, they are already widely used in a number of existing biobanks.

RFID:

Storage of dynamic data on RFID tags, tags being of various sizes, shapes and types that may suit various containers, reduced error rate in reading tags, and tags being readable from a distance with no line-of-sight make RFID a candidate for SITS. However operation in extreme temperatures, in laboratory environment and near liquid samples should be further investigated for the particular RFID system specification that will be deployed.

Based on the description provided HP Memory Spot, Intel® Mote, magnetic strips, OCR, Zigbee, barcodes and RFID are evaluated against the requirements listed in Table 2-2 and the results are represented in Table 4-1. In this table, ✓ shows that the requirement listed in the column is met by the technology listed in the row, × shows that

the requirement is not met by this technology, and ? is used to indicate that there is currently insufficient information available to evaluate if it meets the requirement.

Technology	R1	R2	R3	R4	R5	R6	R7	R8	R9
HP memory spot	✓	✓	✓	✓	✓	✓	×	?	✓
Intel® Mote	?	?	×	✓	✓	✓	?	?	✓
Magnetic stripe	✓	×	×	✓	×	×	×	?	✓
OCR	×	×	✓	?	×	×	?	✓	×
Zigbee	✓	✓	×	✓	?	✓	?	?	?
Barcode	✓	✓	✓	✓	?	✓	✓	✓	×
RFID	✓	✓	✓	✓	✓	✓	?	?	✓
✓: The Requirement is Met by This Technology × : The Requirement is Not Met by This Technology ? : Insufficient Information Available to Evaluate if it Meets the Requirement									

Table 4-1: Technologies Evaluated against SITS Requirements in Table 2-2.

Based on the brief introduction given for each of the available technologies, barcodes and RFID are considered for further discussion and analysis in this Chapter. In the next section of this Chapter barcodes both in form of linear and 2D will be investigated. The technology underlying RFID, its components, different types, specifications and issues will be then discussed. Then barcodes and RFID will be compared in more detail and finally the one that best meets the requirements will be chosen to be deployed for SITS.

4.3 Barcodes

Barcodes refer to graphical representation of data printed on a surface, for example paper or body of a product, in the form of dark bars followed by free space (Frazelle, 2001). In some cases the barcode is printed in 3D form or bumpy form where the height of the lines represent data (Adams, 2008). Barcodes are readable either by an optical device that projects light on the barcode and depending on the reflected light the code is transferred (Muller, 2003), or by image analysis software which uses an image of the barcode to decode it. Barcodes have been in use in the supply chain for inventory maintenance, stock control and payment systems for several years (Muller, 2003). Depending on the graphical representation barcodes are classified as either linear or 2D

(Frazelle, 2001). These two types of barcodes will be described in this section. Each type of barcode supports a variety of coding style. Coding styles vary due to the application that the barcode is designed for, the amount of data it represents and the standard that it follows.

4.3.1 Linear Barcodes

Linear barcodes are created by a set of narrow and thick dark parallel vertical lines that are placed at different distances to each other and are read by standard optical scanners or other types of barcode readers (Frazelle, 2001).

A barcode is said to be *continuous* when dark bars and spaces both represent data, and *discrete* if only dark bars are used to represent the data (DirectBarCodes). Also the width of the bars and/or spaces encodes a different set of data. If the width is taken as narrow or wide, then only two widths are defined regardless of how wide or narrow they are as long as they are narrower or wider than a defined value. *Many-width* or *variable-width* barcodes are encoded such that the widths of the bars and/or the spaces in between them encode different data depending on how wide they are (Bylund and Jirhede, 2003). Different algorithms for encoding data are used based on these properties along with other properties such as the size of the barcode.

Examples of the GS1[6] (GS1, 2008b) approved methodologies are:

UPC (Universal Product Code) A or UPC 12: capable of encoding 12 digits. The first digit represents the content of the barcode, the next five digits are the identifier of the producer, the next seven to eleven digits are the identifier of the product defined by the producer and the last digit is the check digit (ActiveBarcode, 2008f). An example of label of this type is given in Figure 4-3.

[6] GS1 is a global organisation for designing and implementing standards to improve efficiency and visibility of supply and demand chains.

Figure 4-3: UPC A Label (ActiveBarcode, 2008f)

EAN 8 and EAN 13: EAN 13 is used by retailers at the point of sale. EAN 8 is a short version of EAN 13 labels. With EAN 13 the first two digits are the country code, the next five digits represent the producer, the next five are the item number provided by the producer and the last digit is the check digit (ActiveBarcode, 2008d). An example of this barcode is displayed in Figure 4-4.

Figure 4-4: EAN 13 Label (ActiveBarcode, 2008d)

EAN (European Article Number) 2 and EAN 5: add-on to the EAN 13 used for newspapers and magazines, and books, respectively. They do not have a check digit (ActiveBarcode, 2008c).

GS1-128: a standard from GS1 defining the kind of data and the data format to be encoded. A set of application identifiers has been listed under this standard. Four groups of data stored by this standard are: start character, function code, element string including application identifiers and user data, symbol check character and stop character. This standard is used for Code-128 and is displayed in Figure 4-5 (GS1-128, 2008).

Figure 4-5: Code 128 Label (ActiveBarcode, 2008a)

ITF-14: encodes 14 digits including one digit for “logic variant”, 12 for product number and one final digit for check (ActiveBarcode, 2008e). An example of this code is shown in Figure 4-6.

Figure 4-6: ITF-14 Label

4.3.2 2D barcodes

A 2D barcode is based on a similar idea to linear barcodes except that it is capable of encoding a larger amount of data vertically. 2D barcodes include multiple barcodes in rows (Frazelle, 2001). These barcodes hold data both in rows and columns allowing for more data to be stored in smaller area (Frazelle, 2001). Different categories of 2D barcodes are defined based on the data representation symbologies and encoding algorithms. Three major categories of 2D barcodes are: stacked code, matrix code and composite barcodes (Teklynx).

Stacked Code

A Stacked Code is made of multiple linear barcodes being added to each other. Types Code 39 and Code 128 of linear barcodes create Code 49 and Code 16K respectively. Most types of this code allow error detection and correction (AIMglobal, 2008). The density of data stored on a Stacked Code is about five to seven times the density of data on a Code 39 linear barcode and rows on this Code are read sequentially by a laser scanner (Groover, 2007). An example of this code is:

Code 49: encodes 49 alphanumerical or 81 numeric characters in two to eight rows of linear barcodes. This code has three error detection methods in place, specifically parity for each character, one check character for each row that comes as the last character in the row, and finally depending on the number of rows two or three characters are at the end on the barcode, two characters for six or less rows and three for seven or eight rows (Teklynx). Figure 4-7 shows a two row Code 49.

Figure 4-7: Code 49 Label (Teklynx)

Matrix Code

The matrix Code is similar to the stacked code except that the heights of rows are shorter, facilitating a more intense label encoding more data (Groover, 2007). Data cells in a matrix Code are usually square (Groover, 2007) and the matrix Codes themselves can be square, hexagonal or circular. This type of 2D barcodes allows error detection and correction for more reliable reads (AIMglobal, 2008). Matrix Codes allow storage of more data on them compared to Stacked Codes (Groover, 2007). A typical example of a matrix Code is "Data Matrix" shown in Figure 4-8. This particular type can store up to 2335 alphanumerical or 3116 numeric characters (TEC-IT).

Figure 4-8: Data Matrix Label (TEC-IT)

Composite barcode

Composite barcode consists of a linear and a 2D barcodes that are linked. The linear barcode can be EAN-13, UPC-A, EAN-8, GS1-128 or a Databar symbol, and the 2D barcode can be one of the types of Micro PDF417 or PDF417 depending on the type of the linear barcode used. The linear barcode can be scanned without reading the 2D barcodes, however there's a flag that prevents 2D barcode to be read without the linear barcode (dLSoft, 2009). An example of composite barcode is the composition of EAN-8 and CC-A as shown in Figure 4-9. This example can encode 56 digits of alphanumeric data (GS1, 2008a).

Figure 4-9: EAN-8 and CC-A Label (dLSoft, 2009)

More types of 2D barcodes from all three categories and their capacity are given in Table 4-2.

Category	Type	Example	Amount of data
Stacked Code	CODE 49 (Adams, 2008)		7.5 mils width by 8 rows with height of 0.5475 inches maximum: 170 alphanumeric characters per square inch
	CODE 16K (Adams, 2008)		8025 ASCII characters, or 16050 numeric digits
	Codablock F (TEC-IT)		Maximum 2725 ASCII or 5450 numeric characters
	PDF 417 (TEC-IT)		1108 ASCII/ Bytes or 1850 alphanumerical or 2725 numeric characters
Matrix Code	Data Matrix (ActiveBarcode, 2008b)		Maximum: 3116 numeric or 2335 character or 1556 bytes
	MaxiCode (UPSCode or Code 6) (Adams, 2008)		Approx. 100 ASCII characters per square inch
	QR Code (Adams, 2008)		7366 numeric characters, or 4464 alphanumeric characters
	Aztec Code (TEC-IT)		3067 alphanumeric or 3832 numeric or 1914 Bytes

Composite barcode (GS1, 2008a)	Composite Component A (CC-A)	 UPC-E with CC-A	56 digits of alphanumeric
	Composite Component B (CC-B)	 GS1 DataBar (RSS) with CC-B	338 digits of alphanumeric
	Composite Component C (CC-C)	 GS1-128 (SSCC-18) with CC-C	2361 digits of alphanumeric

Table 4-2: 2D Barcodes

4.4 Radio Frequency Identification

4.4.1 Definition and Components

Radio Frequency Identification (RFID) is based on the principles of data transmission over Radio Frequency (RF). The idea was first introduced by Stockman in a paper on "Communication by Means of Reflected Power" (Stockman, 1948).

Glover and Himanshu Bhatt (2006) have described an RFID system as:

> "any system of identification wherein an electronic device that uses radio frequency or magnetic field variations to communicate is attached to an item."

The RFID tag, the "electronic device" in the above description, consists of a memory chip onboard that holds the identification data. An RFID system consists of a reader which might also act as a writer, tags, and an antenna that allows data transmission between the tag and the reader/writer device. In most cases data received from or sent to the RFID system needs to be processed on a computer and depending on the system design there might be a need for a database for data storage and retrieval.

RFID tags contain a memory, antenna, processor, and depending on its type it might also include a battery onboard. Three major types of tags are available: active, passive and semi-passive tags (Roberts, 2006).

Active tags have a battery onboard to power the circuit. Active tags are always on and often have longer reading ranges compared to passive tags, but with shorter life span as the battery life is limited. They also need to be maintained under the conditions that guarantee the survival of battery (Roberts, 2006).

Passive tags power their circuit from the energy they receive from the antenna of the reader/writer. This type of tag has no battery onboard to facilitate communication, therefore has shorter reading ranges with longer life span. Passive tags often are smaller than active tags and can survive in harsher environments. These tags only turn on when they are within range of the antenna (Roberts, 2006).

Semi-passive tags have a battery onboard which is simply used to keep the circuit turned on. In order to communicate with the reader/writer they get their power from the reader/writer antenna (Glover and Bhatt, 2006).

The above mentioned categories of tags can be further grouped into read-only, write-once and rewriteable tags. Read-only tags only carry the serial number that they are given when being produced by the manufacturer and allow identification of tag producer. This serial number can be treated as a unique identifier of that tag. Write-once tags allow writing data on the tag to be done only once. Rewriteable tags include a memory onboard that handles updatable data and can be written many times (Adams, 2007). In order to write data on a tag, an RFID writer device should be employed whereas with read-only tags only RFID reader device is required.

One of the small size RFID tags is µ-Chip produced by Hitachi (Hitachi, 2009). The size of this tag is 0.4x0.4 mm^2, its reading range with a 300mW reader is about 300 mm. This tag operates on 2.45 GHz frequency (Usami and Ohki, 2003).

4.4.2 Specifications

For making communication feasible across different components of a RFID system, certain specifications should be met by all components. Components have to operate under the same frequency and protocol. Protocols that are developed internationally should be adhered to allow interoperability.

Frequency

The frequency used for RFID is dependent on the country and region in which it is deployed. Federal Communications Commission (FCC) in US and European Telecommunications Standard Institute (ETSI) (European Telecommunications Standard Institute, 2009) in Europe are examples of standard bodies. FCC and ETSI have allocated band 13.553 to 13.567 MHz for Industrial, Scientific and Medical (ISM) applications (Domdouzis et al., 2007).

Five major frequency bands allocated to RFID are shown in Table 4-3 (Domdouzis et al., 2007). Depending on the reading range, amount of communicated data and environmental conditions required for each application the operating frequency should

be chosen. In 125-134 kHz and 13.56 MHz frequency bands radio waves penetrate water, but not metal and in higher frequencies they do not penetrate water or metal. The data transfer rate increases as the frequency increases. The reading range although depending on the power of the reader device varies between 0.5 meter and 100 meter.

Frequency		**Reading range**	**Data Transfer Rate**	**Application**	**Comments**
125-134 kHz		0.5 m	1 kbit/sec	Animal identification	Penetrates water, not metal
13.56 MHz		1.5 m	25 kbit/sec	Access and security	Penetrates water, not metal
433-956 MHz	433-864 MHz	Up to 100 m	100 kbit/sec	Logistics	Does not penetrate water or metal
	865-956 MHz	0.5 to 5 m	100 kbit/sec	Logistics	Does not penetrate water or metal
2.45 GHz		10 m	100 kbit/sec	Mobile vehicle toll	Does not penetrate water or metal

Table 4-3: Frequency Ranges and their Properties (Domdouzis et al., 2007)

Protocols and standards

International Standards Organisation (ISO), EPCglobal, American National Standard Institute (ANSI) and Automotive Industry Action (AIAG) are the major organisations responsible developing the standard for RFID (Morgenroth and Fobes, 2004). The EPC system that was first established in the AutoID centre in MIT was converted to EPCglobal Inc in 2003 as a joint organisation of EAN International and Uniform Code Council (UCC) (Roberts, 2006).

The most widely used ISO protocols are:

- ISO 10536: used for close-coupled cards with a distance of less than 1 cm (Knospe and Pohl, 2004).
- ISO 14443: used for proximity cards in range of 10 cm. All four parts of ISO 14443 should be taken into account for interoperability. They are mainly describing physical characteristics, RF interface power and signal interface, initialisation and anti-collision, and transmission protocols. There are two types of this protocol, A and B,

depending on the method of power transmission from the card, that is the RFID tag, to the reader/writer device (Bashan, 2003).

- ISO 15693: used for vicinity cards with distances up to 1 meter. This standard is in three parts specifying physical characteristics, RF interface power and signal interface, and anti-collision and transmission protocol (ISO, 1999b, ISO, 1999a, ISO, 2000).
- ISO 11784, ISO 11785 and ISO 14223: used mainly for animal identification for frequency less than 135 kHz. ISO 11784/85 originally designed for 64 bit identifiers. ISO 14223 describes the read/write and write-protected data blocks and its communication protocol is related to the part 2 of ISO 18000 (Knospe and Pohl, 2004).
- ISO 18000: used for item level management in supply chain. In six part it specifies the air interface, collision detection mechanisms and the communication protocol for different frequency bands (Knospe and Pohl, 2004).

Designing RFID systems

At the time of designing an RFID system it is important to take the following factors into account:

- Reading range: the distance between the position of the reader/writer and tags. Orientation of tags and the antenna should also be taken into account when measuring the reading range.
- Environmental conditions: under what conditions the tags are read and written, what conditions they go through in their life and how the environment influences the reading range and functionality of the tags and reader/writer device. Also the size of the tagged object is an important factor since RFID tags vary in physical size and therefore their specifications differ.
- Amount of data to be stored: the memory of the tag has a limited capacity depending on the type. The maximum memory capacity of tags with suitable size for SITS is currently available from TAGSYS (TagSYS RFID, 2008) and TexasInstrument (TexasInstrument, 2008), the two major RFID suppliers, is 2 kbits (Hibccau, 2005, TI, 2008) with potential increase. Some RFID tags can be locked and/or password protected.

- Nature of data, i.e. static versus dynamic: to decide on read-only, write-once or re-writeable tags.

In addition to the above factors, the cost of the RFID system should also be taken into account as cost can be a major barrier of deploying RFID systems in many environments (Goodrum et al., 2006, Brown and Russell, 2007). However, by using rewriteable tags costs can be reduced (Zhang et al., 2006). In applications such as SITS in which the tagged items are precious and valuable, a balance between the investment in the RFID technology and the value of samples must be considered. However, conducting a detailed cost-benefit analysis was deemed to be beyond the scope of this thesis.

4.4.3 Comments on Maturity of RFID Technology

While RFID has been around for many years, its applications are still relatively limited, although they have started growing significantly recently. One of the preventing factors from widespread deployment of RFID applications has been the issue of maturity. Many industries and enterprises do not regard RFID as a mature technology.

The Gartner Hype Cycle (Gartner, 2009a) model is used to analyse technology lifecycles. This Hype Cycle initiates with the emergence of a new technology and reaches a peak of expectations when almost no adaptation is taking place. Expectations and enthusiasm then downgrade and finally after an increase in the slope of adaptation and expectation, the technology reaches its maturity and is understood properly (Linden and Fenn, 2003). There are five stages of maturity according to this model: *Technology Trigger*, *Peak of Inflated Expectations*, *Trough of Disillusionment*, *Slope of Enlightenment*, and finally *Plateau of Productivity* (Fenn and Linden, 2005, Fenn and Raskino, 2009). *Technology Trigger* and *Slope of Enlightenment* are both due to increase of visibility. After the *Peak of Inflated Expectations* from a technology, its visibility is lost to some degree, part of which is later recovered during *Trough of Disillusionment*. *Slope of Enlightenment* is when the technology is visible again, however, its limitations and capabilities are known and well understood. *Plateau of Productivity* is the final stage of maturity in this model and is when the technology is accepted and deployed according to its capabilities.

Figure 4-10 depicts the status of maturity of many new technologies including passive RFID in 2005. This figure is published in a Special Report by Gartner and is accurate as of August 2005 (Fenn and Linden, 2005). At the time, passive RFID technology was about to ascend through the *Slope of Enlightenment* and its *Plateau of Productivity* was expected to be reached between 5 to 10 years after the time of publication. Technologies like RFID with a number of different applications in industry and research are likely to follow different paths during their *Slope of Enlightenment* and hence *Plateau of Productivity* (Fenn and Raskino, 2009). Figure 4-11 is based on Gartner Hype Cycle as of July 2009 shows that the case/pallet level RFID (Bottani and Rizzi, 2008)[7] is in the *Trough of Disillusionment* expected to reach *Plateau of Productivity* in 5 to 10 years.

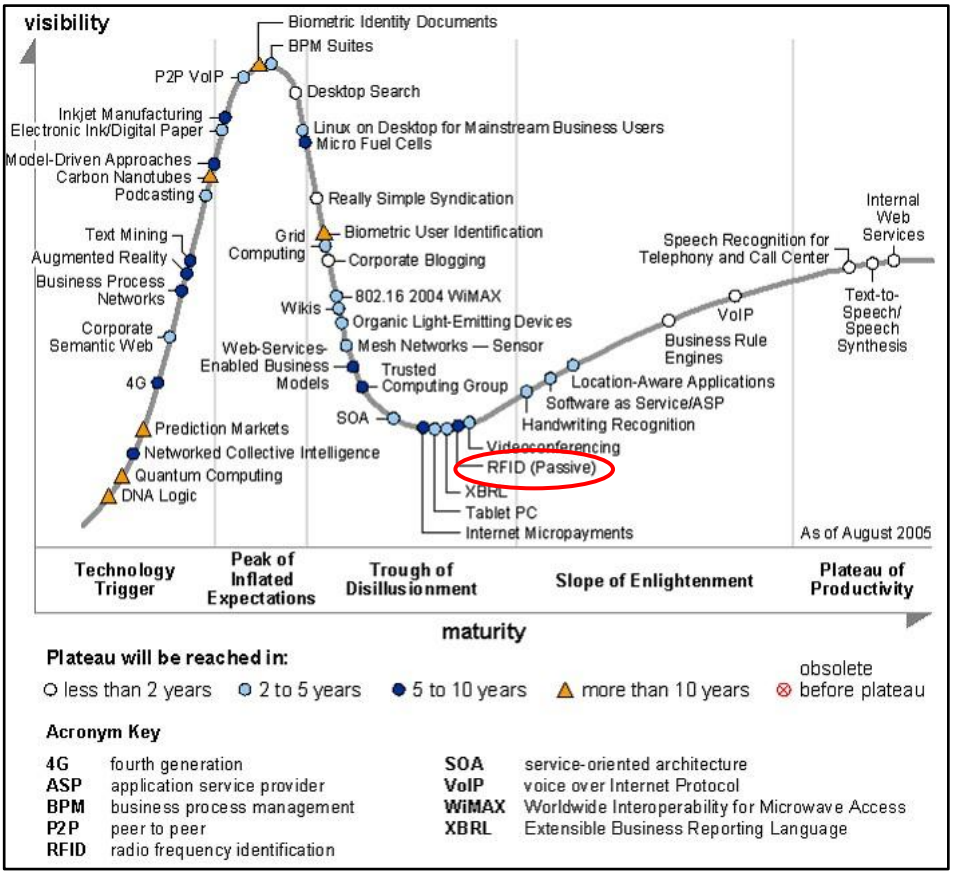


Figure 4-10: Gartner Hype Cycle as of August 2005 (Fenn and Linden, 2005)

[7] Pallet/case level RFID is about tagging different batches of product in the supply chain.

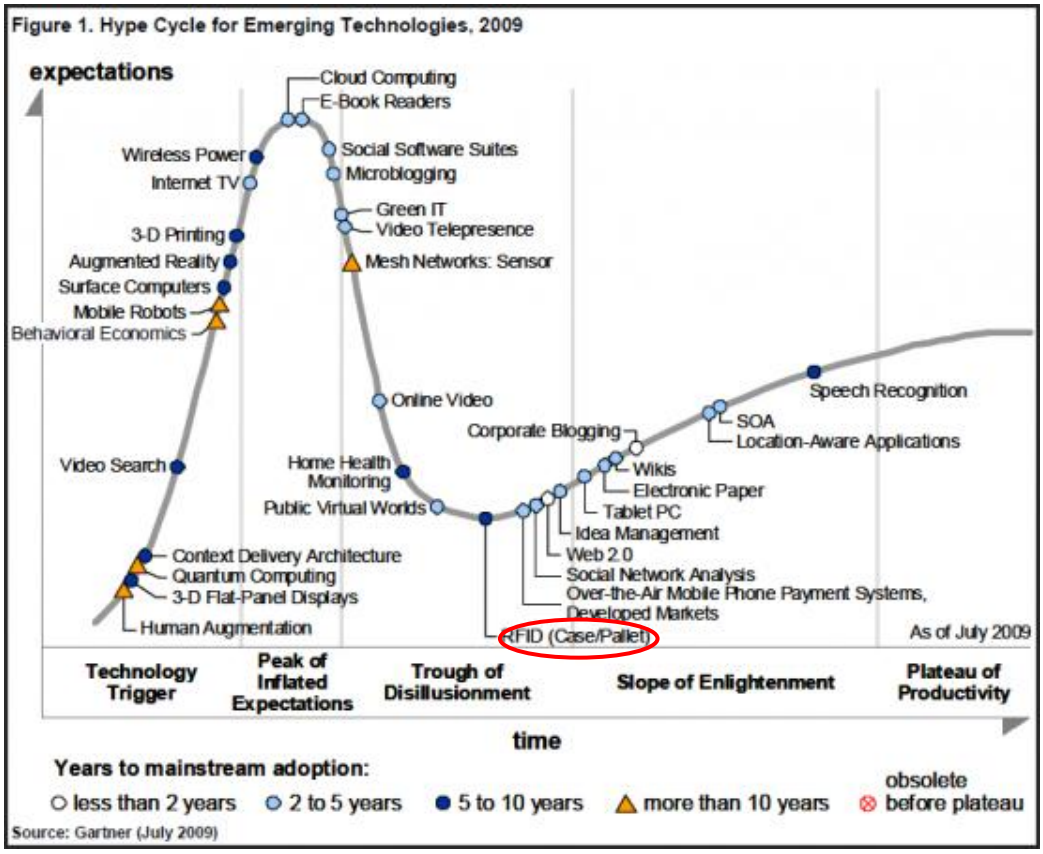


Figure 4-11: Gartner Hype Cycle as of July 2009 (Gartner, 2009b)

4.4.4 Privacy Concerns of RFID Applications

Privacy and security issues have been identified as a major cause of prevention or delay in wide adoption of RFID in a variety of sectors (Ayoade, 2007).

Major properties of RFID such as readability from distance where both the reader and the tag are hidden, data storage on the tag memory and built-in unique identifier, when implemented on item-level tracking, can be misused by endangering consumer privacy by tracking and profiling their activities (Spychips, 2003). Data mining and activities profiling, tracking and secretly reading tag memory with a unique identifier that can be linked to an individual are examples of privacy violation by RFID (Ayoade, 2007). The European Union (EU) Article 29 Data Protection Working Party published a "Working document on data protection issues related to RFID technology" in January 2005 and invited comments on the document (Electronic Privacy Information Center (EPIC)). This document emphasises the data protection concerns that are raised due to:

- RFID used to gather information that is linkable to personal data
- Ability to store personal data on RFID tags
- RFID used to track without notification.

The ability to "kill" an RFID tags has been proposed by Auto-ID Center and EPC global (Ayoade, 2007). Although it is possible to kill the RFID tag at the point of sale, or put them to sleep so that they could be awakened if needed (Ayoade, 2006, Roberts, 2006, Dobson and Todd, 2006, Jacoby, 2006), there still remain serious concerns about the existence of memory chips that can be detected without consumers' knowledge for example, Consumers Against Supermarket Privacy Invasion And Numbering (CASPIAN) (CASPIAN, 2004) and Electronic Privacy Information Center (EPIC) (EPIC, 2009) in US and stopRFID (stopRFID, 2005) campaign in Germany (Jacoby, 2006).

CASPIAN has published a position statement which sets out three principles in relation to RFID:

> "First, RFID must undergo a formal technology assessment, and RFID tags should not be affixed to individual consumer products until such assessment takes place. Second, RFID implementation must be guided by Principles of Fair Information Practice. Third, certain uses of RFID should be flatly prohibited" (Spychips, 2003).

Obviously, the formal technology assessment must be carried out by a neutral multi-disciplinary body. The position statement has listed five set of requirements based on the eight-part Privacy Guidelines of the Organisation for Economic Co-operation and Development (OECD). These requirements are:

- Openness or transparency: so that consumers are made aware of any RFID tag reading activities;
- Purpose specification: the reason behind the use of tags and readers should be clearly stated;
- Collection limitation: only necessary information should be collected and stored;

- Accountability: individuals who deploy RFID are in charge of its appropriate implementation and compliancy with principles; and
- Security safeguards: security and integrity of the RFID system with regard to transmission, database and access to the system should be verified by a third party (Spychips, 2003).

In addition, it is recommended that certain ways of deploying RFID tags should be forbidden (Spychips, 2003). These include, for example, leaving live tags on products after purchase; consumers being unable to locate tags and readers and therefore unable to kill them on detection; tracking individuals without their consent; and using the tag in a manner that compromises an individual's anonymity.

Similarly, Simson Garfinkel has developed "An RFID Bill of Rights" based on Principles of Fair Information Practice (Garfinkel et al., 2005, Garfinkel, 2002). Five basic rights are given to the users of RFID in this bill of rights as below:

- Knowledge of existence of RFID tag on a product;
- Removal, deactivation, or destroying RFID tag on the product once it is purchased;
- The right of not losing their other rights if they opt-out from having RFID tag on the product;
- Knowledge of what information is stored on the RFID tag; and
- Knowledge of when, where and why the RFID tag on their product is being read.

While privacy plays a major role in the adoption of RFID, there need to be rules governing its applications to ensure conformance with consumers' rights. There are also approaches proposed for maintaining the privacy of users. In addition to the kill or sleep commands that have their shortcomings, Faraday cage, active jamming, blocker tag and Authentication Processing Framework (APF) (Ayoade, 2006) approaches are also available.

Kill or sleep commands: these commands are used to put RFID tags in disable mode. A tag in sleep mode can be awakened with a predefined command. The kill command included in the Auto-ID Centre specifications, is a random kill code that can be retrieved from a secure database to disable RFID tags, for example at the point-of-sale

(Roberts, 2006). A kill code is generated for each RFID tag. The issue with the kill command is that consumers may decide that they need the tag to be operational after they have purchased it. It also limits applications that can be deployed beyond the point-of-sale (Dobson and Todd, 2006, Ayoade, 2006, Ayoade, 2007, Jacoby, 2006, Juels et al., 2003). Hence procedures for opt-out decisions by consumers need to be developed (Jacoby, 2006).

Faraday cage: Faraday cage prevents radio waves from reaching the RFID tag by surrounding the tag by metal or foil wraps. This approach has been used in burglaries (Ayoade, 2006, Juels et al., 2003).

Active jamming: in this approach the reader is blocked or overloaded by too many radio signals sent so that the RFID is hidden from view. The issue with this approach is that it is illegal to broadcast high power radio waves that cause problems for other readers in addition to problems that may arise in certain environments (Ayoade, 2006, Juels et al., 2003).

Blocker tags: also called selective blocking is introduced by Juels et al. (2003) and takes advantage of interfering with the tree-walking singulation protocol (Boukerche, 2005)[8] of RFID. By this approach the blocker tag simulates signals to the reader. There are legal implications involved in this approach.

Authentication Processing Framework (APF): based on registration of the tag and the reader on APF, this approach is proposed by Ayoade (Ayoade, 2006). The reader will receive the decryption key from the APF only if the reader is recognised. The issue with this approach is that it is operational when reading from the tag, and further work is needed to be undertaken for writing.

4.4.5 Security Threats of RFID

Five common security threats to RFID have been identified (Rieback et al., 2006, Knight, 2006):

[8] A method using which, the RFID reader handles multiple tags.

1. Sniffing: RFID tags are readable by any compliant reader device without the notification of the individual;
2. Tracking: Tracking individuals without their knowledge will lead to invasion of their privacy rights;
3. Spoofing: Spoofing is writing well-formatted cloned data of a tag on a blank or rewriteable tag;
4. Replay attacks: In this threat RFID queries are intercepted and retransmitted using a relay device; and
5. Denial of Service (DoS): This issue is caused when the system is not functioning as expected and might be due to Faraday cage or active jamming activities.

In addition to the above threats Rieback et al. (2006) have added RFID malware to the categories of threats. RFID malware consists of RFID exploits, worms and viruses (Rieback et al., 2006).

RFID exploits: in this attack the data on the tag are modified such that the back-end system is targeted. Buffer overflow, code insertion and SQL injection are approaches taken to hack system components, such as the database, web interface and APIs (Rieback et al., 2006).

RFID worms: In attacks based on worms, the malicious program propagates itself over the network and causes security flaws. An RFID worm is spread over the tags and might target the RFID online services.

RFID viruses: In this attack an infected tag is used to spread the virus upon the availability of a network connection.

4.5 Barcodes vs. RFID

Having defined linear barcodes, 2D barcodes and RFID, a comparison is made in this Section based on the following criteria derived from the biobank requirements:

- Reading/writing method
- Reading range
- Nature and size of data handled

- Data handling
- Life span
- Cost
- Extendibility and integration
- Survival and functional conditions
- Security

Reading/writing method: Barcodes can be designed by special software packages that create barcode labels, and can be printed on ordinary office printers. Linear and 2D barcodes are readable by optical devices (Muller, 2003). Barcodes need direct observation of the actual label, i.e. line-of-sight is required in order to be able to read data on barcodes. Although certain types of Stack Code 2D barcodes facilitate error detection and correction when scanning damaged labels, linear barcodes cannot be read without the availability of the full label. On the other hand, RFID tags are read by reader/writer device and no line-of-sight is required. This helps reading tags even if they are covered, provided that they are within the reading range of the antenna and there is no source of interference, such as metal or strong electromagnetic waves, in the environment that influences readability. Writing data is also facilitated by the same device if supported by the tags. Sample tubes stored in laboratory freezers are often covered by frost crystals. Hence, RFID is a better solution for reading the tags.

Reading range: The nature of barcode technology requires close proximity of the label and reader. The reading range is up to 12 meters for barcodes (Tzeng et al., 2008a) and about 100 metres for the frequency band 433-864 MHz for RFID. According to Flint (2006) an active RFID tag with a certain specification can be read from up to a 10 km distance. Active tags, that include their power source onboard, have longer reading range compared to passive tags that need to receive their power from the reader/writer antenna.

Nature and size of data handled: Barcodes are physically applied to paper or another surface by printer or similar device and therefore the data encoded in the label cannot be updated. The nature of the data that are represented by barcodes is then static, whereas with rewriteable RFID tags data can be updated many times and this allows dynamic data to be stored and updated on the tag. Although 2D barcodes can encode a larger

amount of data in a smaller area compare to linear barcodes, RFID tags are available in a range of sizes and shapes and provide various amounts of memory depending on the application and their cost. The amount of data stored on 2D barcodes is directly proportional to the area of the label, whereas with the RFID it is related to the Integrated Circuit (IC) of the tag. The amount of data that various 2D barcodes can hold are provided in Table 4-2. Data handled in biobanks is dynamic in nature. Although 2D barcodes provide storage of large amount of data, RFID tags are able to handle dynamic data.

Data handling: Barcode scanners can only read one single label at a time. With RFID technology multiple concurrent reads of tags are feasible. The number of tags that can be read depends on the tag type and the protocol. Reading multiple tags at once can be considered as a positive point in applications where reading speed is important. For example, for supply chain management where a large number of tagged items need to be processed this characteristic will speed up the process. In biobanks, although data entry has to be done individually for each sample, the ability to perform concurrent reading allows the potential of faster sample identification when processing groups of samples.

Life span: The life span of barcode labels highly depends on the quality of the medium they are applied to e.g. paper and the ink and the environment in which they are stored. The life of an RFID tags depends on the battery life which can be several years, and is also dependant on the environment and storage conditions. The battery life depends on the frequency of reads and duration of its "on" status. In the context of SITS where samples are stored for several years, this criterion should be taken into consideration with regard to the technology that is going to be deployed.

Cost: Typically barcodes are printed by ordinary printers on paper and read by inexpensive scanners, although they may be printed using special ink and printer when the application requires exposure to special environmental conditions. There are more expensive papers and specialised printers available. Although chipless or printable RFID tags are being developed (Webb, 2008) RFID tags are expensive because of the circuit they have onboard. In addition to the tags, RFID reader/writer devices and their

antenna are also expensive. To reduce the costs of RFID tags, re-useable RFID tags can be employed (Zhang et al., 2006).

Extendibility and integration: RFID tags can be integrated with sensors and Global Positioning System (GPS) receivers which is particularly useful in global supply chain, for example. Barcodes do not facilitate such developments. In the context of biobanks, the integration of RFID with temperature sensors offers the advantage of being able to maintain the temperature of the samples on an on-going basis. Maintaining samples at the correct temperature is valuable to maintaining the quality of the sample.

Survival and functional conditions: As long as barcode labels are not damaged or covered, i.e. the line-of-sight is available, they are readable. Active RFID tags survive temperatures down to a point that does not damage their onboard battery. This temperature depends on the type of chip and battery used in the tag. Studies have shown that temperature affects the reading range of tags (Goodrum et al., 2006). As mentioned earlier in this Chapter, RFID tags of certain frequency do not operate reliably in metal and/or liquid intense environments due to RF wave penetration issues. This problem does not exist for barcodes. In the case of developing a SITS based on RFID, the particular system and equipment that is going to be used for biobanks should be analysed for its reading range, operational and survival temperature ranges.

Security: Barcodes do not facilitate password protection of the label data; however, 2D barcodes can encode encrypted data. Data encoded in linear barcodes are often printed in human readable format on them. 2D barcodes display encrypted data that is not readable by human eyes. RFID tags allow data encryption, password protection and locking of the memory of the tag as well as maintaining a record of accesses to the memory. These features are available based on the specification of the tags that are used. Other security issues regarding RFID were discussed earlier in this Chapter, however they can be addressed. Cloning RFID tags is difficult, as tags are manufactured, whereas barcodes can be copied using copying devices and scanners. Thus, RFID provides better security functionalities which is an important consideration for biobank purposes. Although SITS includes no participant identifiable data, the data stored about the sample may reveal the participant identity in conjunction with other

information. Furthermore, RFID provides more security for SITS implemented in biobanks that do store participant data.

There are advantages and disadvantages that barcodes and RFID provide based on the above nine criteria. Each application should prioritise the facilities that are expected from the technology based on the application and the project definition and requirements.

For SITS, that requires storage of dynamic data, RFID is a better approach than barcodes. Handling multiple aliquots simultaneously would increase the speed of procedures. RFID facilitates reading data with no line-of-sight which is a major advantage in the context of biobanks, where tubes are often covered with layers of crystal. A major and important advantage than RFID provides over barcodes, is extendibility and integration with other sensors. Integrating RFID with temperature sensors provides a more comprehensive history of samples. Finally, RFID technology offers significant security advantages over barcodes. While, in the context of the PCRC biobank, where these data is de-identified, this may not be a major issue, for other biobanks is which identifiable information is stored, RFID is preferable. Thus, taking all the requirements for SITS into consideration RFID offers significant advantages over barcodes and will therefore be used as the basis for implementation.

4.6 Conclusion

Memory Spot, Intel® Mote, magnetic strips, OCR, Zigbee, barcodes and RFID are considered as technologies available for automatic identification of objects. Each of these technologies was briefly introduced. It was shown that barcodes and RFID are the two candidates that are suitable for SITS application. These two technologies then were further investigated and a comparison between them was carried out based on nine criteria: reading/writing method, reading range, nature and size of data handled, data handling, life span, cost, extendibility and integration, survival and functional conditions and security.

The comparison between barcodes and RFID shows that multiple RFID tags can be read simultaneously and require no line-of-sight compared to barcodes. Data stored on RFID

tags are updateable unlike barcode data. Although the reading range of RFID tags varies depending on the environment, it provides longer reading range than barcodes. Barcodes survive extreme temperatures while survival of RFID tags in such temperatures affects their reading range. RFID systems can be integrated with temperature sensors that will be particularly useful for SITS applications. Barcodes do not allow such integrateability. The advantages that RFID technology offers make it particularly suitable for SITS application. However, since details of RFID operation depends on the specific tag and reader/writer device these features should be further investigated for the particular equipment that are going to be used: survival in extreme temperatures and its effect on the reading range, its operation near liquid and metal and life span.

Although RFID is perceived as an expensive technology by many individuals and in many industries (Adams, 2007, Roberts, 2006), the cost of the tags and reader/writer is dropping rapidly. Analysis of cost is beyond the scope of this project, and will not be investigated as part of this thesis.

Chapter 5. Survey of Applications with Commonalities with SITS

5.1 Introduction

Barcodes and RFID are found to be the most applied technologies for supply chain management and inventory systems for tracking and managing goods. However, before deploying these technologies for sample identification and tracking, further investigation must be carried out to learn from the experience of prior applications. A large number of pilots and studies in various areas have been carried out and some of the major projects in a number of areas have been listed in Appendix E along with their scope, advantages and limitations. These areas include construction, retail, library, food, postal and cargo services environments as well as individuals' care, healthcare settings, and clinical and laboratory settings.

Objects, that is samples that are intended to be identified and tracked in SITS, are highly valuable and precious biological samples that need to be handled with special care, maintained under certain conditions, and are small in size and large in number. These samples change form and are moved from one container to another throughout the workflows. Also, the environment in which tracking takes place consists of multiple locations. Gemstones are also small in size, highly valuable, transferred across multiple locations and change their form from rocks to precious stones. Identification and tracking of gemstones will be described in Section 5.2.1.

Studying systems dealing with biological samples could reveal interesting results. The blood transfusion process involves blood samples that are collected from patients and are processed at a number of locations and requires a high rate of accuracy in order to be error-free. The blood transfusion workflow is complicated as are the workflows that samples in a biobank go through. Characteristics of this application make it particularly interesting for further analysis. The blood transfusion process will be discussed in Section 5.2.2.

Finally, biobanks can be thought of as a multi-site laboratory that collects, processes and stores samples for different purposes. Section 5.2.3 studies laboratory management

systems to provide ideas as to what should be anticipated in biobanks. In particular, the development of the Laboratory Information Management System (LIMS) would be an invaluable experience for developing BIMS and SITS, specifically from software development point of view.

5.2 Applications

5.2.1 Small Precious Stones: Diamond Tracking

Diamonds are small in size, not perishable, highly valuable compared to size, exchangeable for goods and cash, and impossible to track when polished (Tailby, 2002). These characteristics make diamonds the target of illegal activities and smuggling as well as increasing the potential of loss during processing and retail levels. It is estimated that approximately 20 per cent of the rough diamonds traded globally are illicit (Tailby, 2002).

From mining to consumer, diamonds go through a supply chain, also called the diamond pipeline. The initial stage of the pipeline is mining in the major mines worldwide (see Figure 5-1). The output is then sent to De Beers' Diamond Trading Company (DTC) as part of the Central Selling Organisation (CSO) in various locations. Rough stones are then sorted into 16,000 categories depending on their size, colour and quantity. This process may be automated or may be carried out manually. Each category will be boxed and sold to sightholders by DTC (number 3 in Figure 5-1). Sightholders are often diamond manufacturers, cutters, and retailers. Rough stones are then cut either in cutting houses at different locations worldwide (number 4 in Figure 5-1) or by the sightholders themselves (number 5 in Figure 5-1). Diamonds are then sold to wholesalers, jewellery retailers or jewellery manufacturers by the cutting house or by the sightholders (AllAboutGemstones, 2009). During this process, five steps of operation are identifiable. They are mining, rough stone trading, cutting, polishing and mounting on jewellery for the end consumer (Tailby, 2002). Transitions between locations are good opportunities for illicit traders.

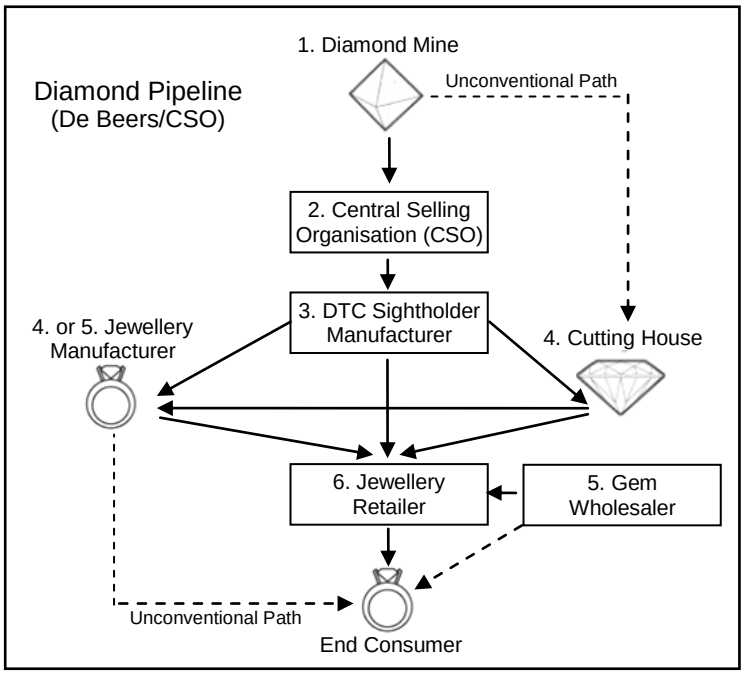


Figure 5-1: Diamond Pipeline Redrawn (AllAboutGemstones, 2009)

Another challenging issue that has been attracting attention in the diamond industry is mining "conflict diamonds". Conflict diamonds are defined as:

> "Conflict diamonds, also known as 'blood' diamonds, are rough diamonds used by rebel movements or their allies to finance armed conflicts aimed at undermining legitimate governments." (kimberleyprocess.com, 2009a)

To monitor the flow of diamonds and prevent conflict diamonds, the Kimberley Process Certification Scheme (KPCS) has been developed. The KPCS has been in place since November 2008 and, by this scheme, members have to validate the transport of rough diamonds so that conflict diamonds do not enter the pipeline (kimberleyprocess.com, 2009b).

Apart from the KPCS, there have been efforts by RFID technology companies to prevent loss and illicit activities in the diamond pipeline.

Spacecode-rfid (Spacecode-rfid, 2009) has made substantial efforts to develop a full tracking system for diamonds being monitored from mines to retail stores with full reliability, accuracy and security. They have mentioned the potential application of RFID at various points of the process and have anticipated the merging of these distinct

systems together in the near future. Figure 5-2 depicts the layout of their proposed system in schematic form. The Kimberley Certificate Number (KCN) is assigned to rough stones once they are mined to track their movements and validate their accordance to the KPCS. At the initial stage, numbered 1 in the Figure 5-2, mined rough stone packages are tagged. The IDs of the packages are read at the DTC for verification, and mining data as well as sorting data is tracked (point 2 on the Figure 5-2). At the third phase of the process rough stone packages are scanned and an inventory of their count and data is instantly maintained for the sightholder. Sightholders can also identify and verify receipt of any particular package. At the cutting house, or when rough stones are undergoing grading procedures, a full record of their origin, their KPCS certificate and image, and other related data can be retrieved simultaneously, improving efficiency and data availability. At the jewellery manufacturing phase, functions provided in previous phases are feasible. Once diamonds are made available to the retailers or end consumers, number 6 in the Figure, RFID improves stock protection, inventory control and management (Spacecode-rfid, 2009).

Spacecode-rfid has developed a system for identifying diamonds in their polished form (spacecode-rfid, 2007). This system is capable of identifying multiple diamonds simultaneously. Such a system has also been proposed in a white paper by Rasilant Technologies Pvt. Ltd. (RasilantTechnologies, 2009). This white paper proposes that diamonds should be stored in small envelopes with printed barcodes or handwritten labels. Barcodes are used to link diamond data to the actual diamond. The process of scanning data using the barcode is time-consuming when reading a box of diamonds. Also, the associated records are often outdated as they require data being entered manually into the database. Barcodes would also not support tracking the diamond when it is moved in the store or within locations along the pipeline. Rasilant Technologies, in this white paper, has claimed that an RFID system when deployed “can reduce shrinkage, as well as cut inventory by nearly 50 percent within a span of two years.” (RasilantTechnologies, 2009).

A patent for maintaining a gemstone inventory has been registered (Rubinstein, 2006). This system uses small RFID tags to monitor the flow of gemstones or diamonds using RFID tags’ unique identifiers. In addition to the mentioned systems, Favourite Diamond

Inc., a major diamond manufacturer in New York, has used RFID to monitor the inventory of its products (Gray, 2008). They have applied Magellan's Phase Jitter Technology (PJM) (MagellanTechnology, 2009). It has been found that reading multiple tags simultaneously and without the need to remove them from their container or envelope has saved the company labour costs and time.

In the past, attempts to store data on the table[9] of a diamond using a data matrix have been carried out (Roskin, 2001). With 3Beams Technologies of Hillsboro, Ore., data is stored on a 50 nanometres to 1 millimetre sized matrix and is readable by a laser beam.

[9] The surface on adiamond.

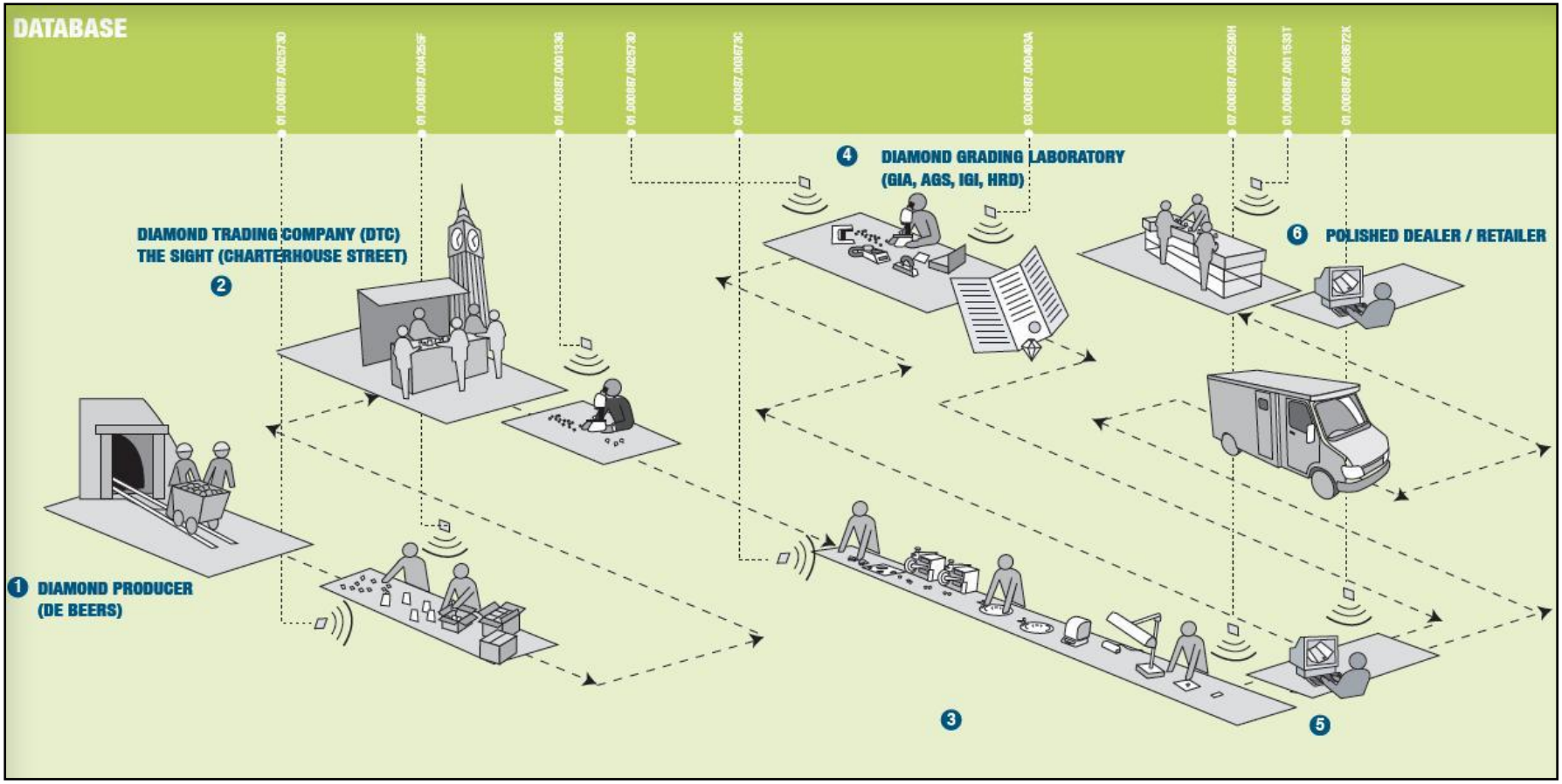


Figure 5-2: RFID Enabled Diamond Industry (Spacecode-rfid, 2009)

5.2.2 Biological Samples: Blood Transfusion Process

Blood transfusion refers to the process of transferring blood from a donor individual to a recipient individual (World Health Organization, 2009). The blood transfusion process involves a series of complicated procedures to be carried out safely and in accordance with the guidelines (National Patient Safety Agency, 2006). Expired, inadequately tested, and ABO incompatible blood being transferred to patients puts the lives of patients in danger (GS1 UK, 2009). There have been 115 cases of "Death in which transfusion reaction was causal or contributory" between 1996 and 2007 in the UK (Serious Hazards of Transfusion, 2007).

The two main points that are extracted from the Blood Safety and Quality Regulations (Office of Public Sector Information, 2005) as requirements for the transfusion process are: full tracking of blood samples from arrival at the hospital to administration to the patient should be maintained electronically, and the system should be based on good practice for the purpose of quality control. It is also vital that sufficient blood samples are maintained safely at all times and that they are appropriately used for the right recipient. In the UK, the National Patient Safety Agency (NPSA, 2008), Medicines and Healthcare products Regulatory Agency (MHRA, 2009), the Chief Medical Officer's National Blood Transfusion Committee (NBTC) (UK Blood Transfusion Services) and Serious Hazards of Transfusion (SHOT) organisation have made attempts to reduce ABO incompatible blood transfusions.

A document published by the National Health Services (NHS), November 2006, describes the specifications and IT requirements for an Electronic Clinical Transfusion Management System (ECTMS) that focuses on tracking blood products outside the laboratories (National Patient Safety Agency, 2006). This document, "Electronic Clinical Transfusion Management System", is used as the context of this Section. It was developed such that it could be applied when deploying either bar code or RFID technologies.

The clinical transfusion management system interacts with the local Patient Administration System (PAS), the NHS Spine and laboratory systems. This system provides services to a number of user roles, for example Blood Group Analyser

(BGA), clinician, patient and phlebotomist. Therefore, it is vital to have access rights and controls in place, in addition to maintaining a full audit trail of all access.

As part of blood transfusion activities, the system should be capable of taking care of a series of actions, including the processing of requests for blood, the storage of blood in stock and issue fridges, and the administration, prescription, collection and transfer of blood.

Based on the described specification, a pilot plan for ECTMS was deployed in Mayday Healthcare NHS Trust in collaboration with NHS Connecting for Health (NHS CFH) and NPSA. NHS CFH has funded the project. The Julibee and London wing wards are involved in this pilot (GS1 UK, 2009).

The aims of the pilot are to determine the possibility of adopting ECTMS in other hospitals, to feedback lessons learned to the NHS and to determine the feasibility of using the same technology in other areas of healthcare (GS1 UK, 2009). Deployment of the pilot started in June 2007 and went live early in 2009 (National Health Services, 2009a).

In this pilot, active and passive RFID tags have been used for tagging blood bags and patient wristbands, respectively. Active RFID tags are deployed for tracking blood bags in the hospital, and passive tags and RFID hand-held readers allow the carrying out of a final "Right patient, Right blood" check at the bedside before transfusion (National Health Services, 2009b). RFID tags have been chosen for this pilot because more data can be stored on them. The particular type of tag that is used complies with GS1 standards and hence supports the Department of Health's "Coding for Success" policy (GS1 UK, 2009).

The process begins with the admission of a patient to the treatment plan. Upon arrival a wristband with a passive RFID tag, which uses its unique GS1 identifier to link to the patient's record, is issued for each patient. The patient wears the wristband for the duration of their stay at the hospital. During the pilot, doctors use a hand-held computer device with an RFID reader and barcode scanner. This portable computer is linked to the hospital network and allows instant updating of records unlike the previous situation where the doctor had to update records from a different room,

increasing the potential for error. Doctors use the computer scanner to read the identifier of the patient on their wristband in order to retrieve the correct record from the database. If the doctor orders a blood sample to be taken from a patient, the order is placed electronically on the computer. A list of the patients who need a blood test is then updated. From this list, the phlebotomist visits the patients and, after checking the wristband for Positive Patient Identification (PPI), blood is taken and a label including all data is printed and attached to the blood bag instantly. This reduces the chances of the sample being mislabelled or illegible due to handwriting. The sample is then transferred to the laboratory where a compatibility label is attached to the blood bag. This label includes the patient's information in a 2D barcode format. The blood bag is then stored in the issue fridge. Before any blood is transferred, it has to be screened and cross-matched. When a doctor orders a blood transfusion for a patient and clinical staff carry out the checks successfully, an "authority to collect" slip is issued to the porter. Then the porter collects the blood bag, affixed with an active RFID tag, from the issue fridge and transfers it to the ward. Finally, the blood bag compatibility label and the patient's wristband are checked and if the matching process is successful, transfusion takes place. When the blood bag has been issued and tagged, its hospital journey until it reaches the patient can be monitored using the WI-FI that is already in use for asset tracking in Mayday Hospital (GS1 UK, 2009).

The pilot evaluated the effectiveness of ECTMS for blood management, decrease in transfusion errors and its potential application for other areas of healthcare delivery (National Health Services, 2008). The benefits of such a system include safer patient care with the reduction of adverse events, automatic checking of information, reduction in losing unused blood samples and improved traceability of blood products.

In addition to the ECTMS pilot in the UK, there have been similar pilots internationally. A hospital in Hong Kong has gained experience from 1999 to 2002. The Pamela Youde Nethersole Eastern Hospital, Hong Kong, deployed a barcode system for identifying patients for blood transfusion in all of its departments except for the psychiatric wards and accident and emergency departments (Chan et al., 2004). Patients admitted to a hospital receive a unique identifier, a "HN" prefix followed by eight numbers, on each visit. This identifier, along with other patient information,

extracted from Integrated Patient Administration System (IPAS) is then printed on a label. The identifier is in both human readable format and barcode. The system developed for the Unique Patient Identification (UPI) then extracts the patient record, that is, the same information being printed on the label, and prints it on a second label attached to the patient wristband. The difference between the two labels, in addition to their shape, is the prefix of the patient identifier. "WB" is used as the prefix in the UPI system. The device used to match the two labels is a hand held barcode reader that is operated by battery and includes a printer and a display screen. Upon blood collection from the patient, medical staff check the identity of the patient using the data on the wristband and the blood request form. If they match, the barcode label is attached to the blood request form and the second check is carried out using the barcode scanner. Then the UPI device prints the barcode with the blood collection data on a label to be attached to the tube. The sample is then sent to the hospital blood bank. The Laboratory Information System (LIS) scans the barcode on the blood tube with the barcode on the blood request form and if they match, tests will be carried out on the blood sample. Any new sample unit being generated from the original tube will have a barcode label with the same identifier and patient information attached to it. Finally, before blood administration, the patient's identity and blood group are checked with the label on the blood tube. Then a verification blood label is printed and attached to the patient's blood transfusion record.

There had been 13 instances of mislabelling of blood samples or request forms from May 1995 to April 1999, a period of three years prior to the introduction of the new barcode system. With the introduction of the new identification method, no instances of mislabelling of blood samples or request forms were reported from May 1999 to April 2002. During the three year use of the system, 12% of identifications had to be done in the traditional way, due to the device running short on battery or its failure. The majority of the recurring costs involved in this system were on device batteries and paper supplies. The next generation of the devices is expected to have improved batteries. Although other issues with the system were discovered during the three year interval, improvements were expected to be made in the newer versions.

5.2.3 Biological Samples: Laboratory Information Management System (LIMS)

A laboratory's success is measured by its error-free reports on processing a number of samples in a period (Stafford, 1998). Computer-based automation of this process assists in improving performance by analysing samples more accurately and in a shorter period of time (Stafford, 1998), and also by compiling standard reports.

A Laboratory Information Management System (LIMS) is a vital component of a laboratory whose outcome is information (Krasovec, 2007) and data reports (Stafford, 1998). A LIMS, being the backbone of a laboratory, is in charge of a variety of tasks including sample identification, work scheduling, data analysis, reporting and other administrative and management tasks (Krasovec, 2007).

A LIMS supports data manipulation to be carried out appropriately in order to provide information of suitable quality that meets the standards and regulatory systems in place. As has been described by Grauer (2003), archiving, retrieving and auditing data must comply with regulations, GLP and GMP (Stafford, 1998), and must allow traceability and accountability. In his article, Grauer (2003) broadens the scope of "traceability" to cover "recordkeeping, sample tracking, staff training, certification maintenance, and more" and explains how databases can successfully deploy traceability. He has also described the functionalities of a compliant LIMS as having to support various methods of uploading data from different formats, as well as having to support communication with analytical equipment and many other features that relate to laboratory maintenance. Furthermore, interoperability, flexibility and integration should also be supported by the LIMS (Krasovec, 2007).

A number of companies are working on the development of LIMS, for example, StarLIMS (STARLIMS, 2009), GraphLogic (GraphLogic, 2009), LabVantage (LabVantage) and Pardus (Pardus, 2009). StarLIMS, with a long history in LIMS development, will be discussed.

StarLIMS has been working in the area of LIMS since 1987 with PC-based LIMS that are now offered as web-based systems. StarLIMS' LIMS solution supports laboratory activities by a number of different modules that it provides. For example, its Data

Capture Utility (DCU) supports numerical data entered automatically by laboratory instruments, or its StarDoc archives "unstructured data" such as photos in a manner that complies with 21 CFR Part 11 (Grauer, 2003). StarLIMS takes advantage of other available tools to ensure data integrity, traceability and quality. Another module that StarLIMS has provided is an RFID module that handles multiple RFID readers from different sites. Tagged samples movements are monitored when they enter or exit a room site. Details of sample identification data along with movement information are logged (Venkatesan and Grauer, 2004). The Chain-of-Custody (COC) log in StarLIMS records the previous and current locations of the sample by logging where and when the sample was last detected before moving to its current location. If the tag is rewritable, then this data is written on the tag (Venkatesan and Grauer, 2004).

StarLIMS supports the managment of structured and unstructured data, and allows data and features integration from other systems such as the Scientific Data Management System (SDMS) and the Electronic Laboratory Notebook (ELN) (Wood, 2007).

The most recent version of StarLIMS, that is STARLIMSV10, is a web-based LIMS whose multi-tiered architecture allow for future developments (STARLIMS, 2008). This system supports the full integration of a SDMS to allow one unique platform to take care of all the operations carried out in laboratories including sample identification and tracking. It is designed to minimise the learning curve and its GUI is based on drag-and-drop methods. To ensure data integrity and validity, StarLIMS maintains full documentation relevant to the operation. It stores images downloaded from the instruments, text SOPs, reports and more (STARLIMS, 2008) to ensure samples are processed according to the requirements. The architecture of StarLIMS is depicted in Figure 5-3. A firewall is installed to ensure secure client access, which is an internet browser, to the server to ensure security of the system. Since STARLIMSV10 is web-based, no installation on the client machine is needed. This LIMS provides each user, group, or location staff with their own personalised and role-based reporting standard. The reporting tool used is Business Objects' Crystal Reports. A wizard, STARLIMS Role Design Wizard, is also provided to administrator accounts to assist them in defining and customising roles.

Modules provided in version 10 of STARLIMS include the Material Manager that is in charge of maintaining an inventory of materials used for sampling and processing, and the Instrument Management Module that deals with routine tasks and scheduling (STARLIMS, 2008).

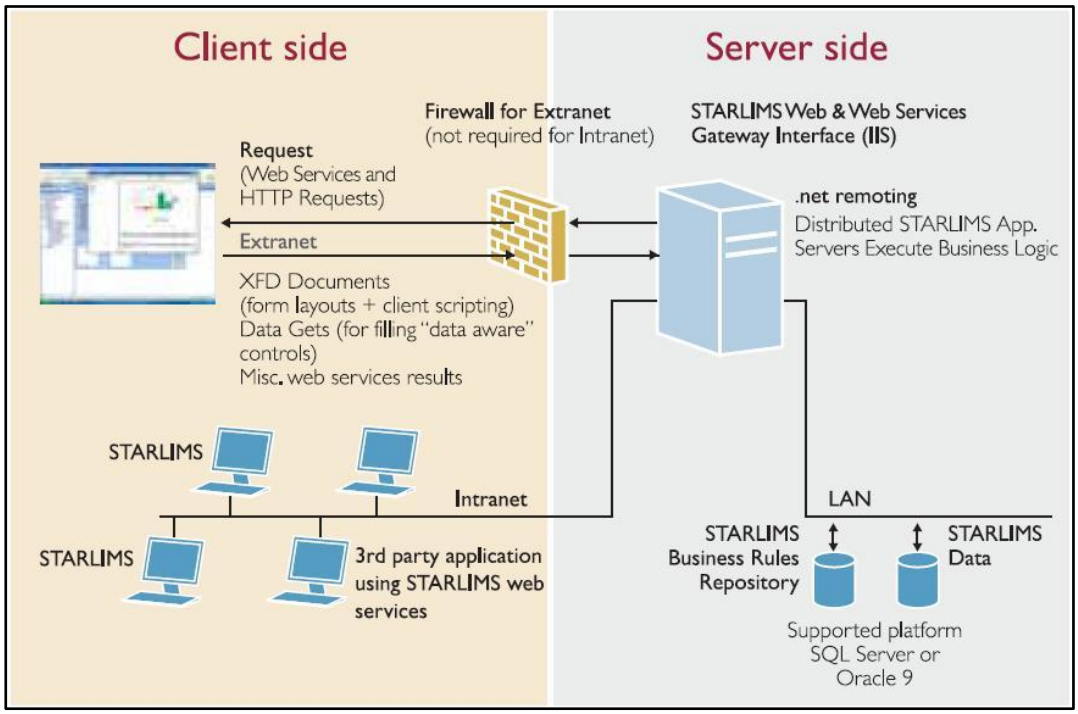


Figure 5-3: STARLIMSV10 Architecture (STARLIMS, 2008)

5.3 Summary

Three different scenarios have been investigated in this Chapter: tracking diamonds, the blood transfusion process and LIMS. Similarities between each of these scenarios with biobanks, workflows, and sample tracking and identification, have made them interesting resources to draw experience from before deploying a SITS.

Diamonds, being small in size and high in value, are particularly comparable to biological samples, although the interested group in the two may not be the same. Also, the distribution of the processing locations is a familiar scenario in the context of multi-institutional biobanks. Furthermore, the idea of storing diamonds in barcoded envelopes brings to mind the potential possibility of tagging samples in batches or boxes when they come to a stage where they have identical characteristics or when

they are stored in well-plates[10]. The tracking diamonds project has shown that RFID is a strong candidate for tracking small, precious items through various locations.

The blood transfusion process involves blood samples that are liquid in nature and require highly accurate identification of blood bags. It is crucial to ensure that the right patient receives the right blood. In this application, RFID has been used for matching the blood bag with the patient receiving the blood. Each blood bag is RFID-tagged and checks are carried out against the patients' wristband prior to the administration of the blood to the patient. This project has successfully proved that RFID tags can operate on blood bags that contain liquid substances. It also shows the feasibility of tracking samples through the transfusion process which is also a complicated scientific workflow.

Finally, the third application under study was a LIMS and, more specifically, StarLIMS as a major LIMS product. StarLIMS is web-based and has a multi-tiered architecture that supplies various modules for different purposes and supports both RFID and barcodes. StarLIMS allows the storage of SOPs, documentation and reports. It also takes advantage of a Material Manager to maintain an inventory of samples, like a SITS. Developing a web-based modular system to support a SITS would support integration with other systems and allow future developments.

In addition to the three scenarios studied, major pilots, studies and projects that are based on RFID and/or barcodes are provided in Appendix E. These studies cover applications in construction environments, supply chain management, library, food, postal and cargo services, individuals' care, healthcare settings, and clinical and laboratory settings. From these studies it can be concluded that:

- A wide range of applications are available for RFID and barcodes. For example, tracking components in construction sites (Furlani and Stone, 1999) and tracking cheese (Regattieri et al., 2007) in food industry.

[10] A well-plate is a flat surface with multiple wells on it that are used as small tubes.

- RFID can be used on meat and tissue products with no influence on them as shown in Kerry et al. (2006), hence this technology can be used for tissue samples without any effect.
- RFID, although expensive, can be integrated with other technologies, such as GPS and sensors to fulfil requirements. For example, RFID has been integrated in a system with barcodes, Personal Digital Assistant (PDA) and GPS to track components in construction environments (Furlani and Stone, 1999). RFID integrated with PDA have also been used for quality inspections and the management of concrete specimens (Min et al., 2007). Also, a prototype has been developed with an RFID reader with an onboard gas sensor for fruit quality monitoring (Vergara et al., 2007). RF technology, that is the underlying technology for RFID, along with an infrared motion detector, magnetic switch and pressure pad have been used in a study on home monitoring for individuals' care (Almudevar et al., 2008). RFID integrated with GPS have been deployed for tracking patients, infection control, outpatient and newborn management and many more applications (Tzeng et al., 2008b).
- Specific RFID equipment and tags are operational in metal-intense environments, although the reading range is affected. However, difficulties were experienced when reading low frequency tags surrounded by metal (Song et al., 2006). There have been cases where the tags acted unreliably near metal (Tzeng et al., 2008a, Umetani et al., 2006). A reduction in the reading range is also experienced when tags are placed inside concrete blocks (Min et al., 2007). Since a biobank's infrastructure consists of freezers and metal racks, it should be taken into account that they may reduce the reading range of the RFID reader or may affect its operation.
- The reading range of RFID tags is reduced in extreme temperatures as was discovered in a field trial on tracking construction tools (Bajic and Chaxel, 2002, Goodrum et al., 2006). This characteristic of RFID technology should be taken into consideration when designing and implementing a system.

Chapter 6. System Requirements

6.1 Introduction

High-level requirements for SITS have been already described in Section 2.7. Briefly reviewing those requirements, a SITS must provide:

R1. Security and confidentiality of data

R2. Fast, reliable and error free

R3. Item level identification of small biological samples

R4. Data accompanying the small samples tubes at all times

R5. Sufficient storage space available for storing sample data apart from the sample identifier

R6. Support for complex workflows

R7. Support for storage in LN, extreme cold and multiple freeze-thaw cycles

R8. Long life span sufficient to cover long-term storage for several years

R9. Storage of dynamic data at each phase of the workflow

The abovementioned requirements are considered vital and must be met by a technology based SITS. In order to fulfil these requirements, system requirements at a lower level should also be met. System requirements will be formulated in this Chapter by following the "IEEE Recommended Practice for Software Requirements Specifications" (IEEE, 1998) and customising the recommendations to fit the SITS system. These recommendations consist of two major parts: the overall description of the system and the specific requirements. The overall description of the system is intended to provide a background to understand factors that influence the system and therefore the requirements of the system. This part of the Software Requirements Specifications (SRS) can be broken down to six parts including system perspective, system functions, user characteristics, constraints, assumptions and dependencies, and apportioning of requirements. Detailed requirements are then covered in the specific requirements section on which the system design is based.

This Chapter will provide an overall description of the system in the next Section, with details of system perspective, system functions, user characteristics, constraints, assumptions and dependencies, and apportioning of requirements. In Section 6.3 a list of specific requirements is provided along with description of each element of the list. Two major categories of the specific requirements identified are the functional and non-functional requirements. In Section 6.4 the SITS requirements from Section 2.7 and the specific requirements are validated in the context of the PCRC biobank, where the SITS prototype is intended to be evaluated. Section 6.5 concludes the Chapter by providing a summary of requirements.

6.2 Overall Description

6.2.1 System Perspective

At an abstract level, SITS can be designed in two approaches. The first approach in which the SITS is fully integrated with the BIMS uses the central database to share its data and provides users with access through the same interface. The second approach is a stand-alone system which interfaces to the BIMS. This approach is shown in Figure 6-1. With this approach the requirements specification is based on a generic independent SITS which communicates with the BIMS as required.

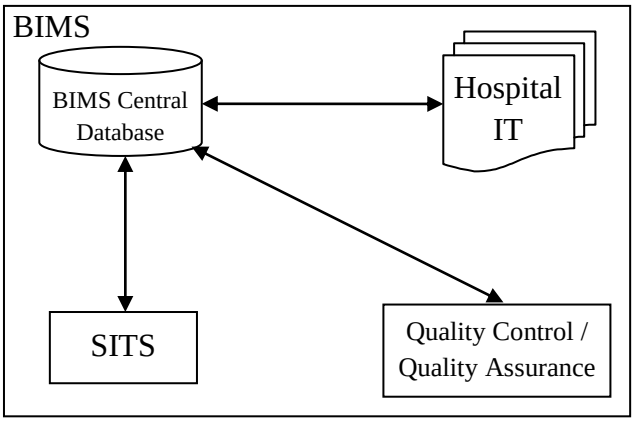


Figure 6-1: SITS Interaction with BIMS Central Database

System Interfaces:

With the first approach to SITS development where it is fully integrated with BIMS and thus the interface is within the system, SITS data can be stored on the BIMS

central database and thus no other interfaces between the two systems are needed. On the other hand, for the second approach where SITS is stand-alone, it is anticipated that SITS provide an interface to the BIMS central database. SITS data can be stored in a separate database, the SITS database, as a consequence requiring an interface between the two databases. Data from SITS should merge with the data on the BIMS central database. This database should be searchable from SITS. Operations such as adding, deleting and editing data must be facilitated through the SITS-BIMS central database interface.

User Interfaces:

The SITS user interface must be integrated with the BIMS user interface facilitating a single interface to the user. Users of SITS are geographically dispersed and hence it is vital that SITS is accessible from various locations and by various individuals accessing it concurrently. The interface must be easy to navigate and understand by non-technical individuals. Depending on the users' access rights and privileges they must be provided with access to different aspects of the system.

Hardware Interfaces:

SITS is going to be based on RFID technology and hence it must be able to control the RFID device through a physical port such as USB or RS232 that is used for data communication between the device and SITS or its user interface.

Software Interfaces:

SITS will need a database management system, which is the management system of the BIMS central database, with full integration of SITS into BIMS. With the stand-alone implementation, SITS will need a database management system that may be different from the one used for BIMS.

Communication Interfaces:

Communication with the RFID device shall follow its specific protocols as described in its supporting documentation.

Hardware:

Certain types of biological samples including plasma, serum and urine are liquid in nature. However, when frozen and after omic procedures, they may not be liquid any more, and this needs to be taken into consideration when deciding on the frequency to be used. Two available frequency bands that penetrate liquids are 125-134 kHz and 13.56 MHz from Table 4-3, Section 4.4. The reading range for the 125-134 kHz band is about 0.5 meter with 1 kbit/sec data transfer rate, and 1.5 meter with 25 kbit/sec for the 13.56 MHz band. It should be noted that samples are stored in freezers located in electromagnetically intense environments which reduces the reading range and data transfer rate. Hence 13.56 MHz is preferred. The protocol that matches this frequency is ISO 15693. Tags must be passive to allow for longer life span and survival in extreme temperatures. The lids of the tubes that will be tagged are 1 centimetre in diameter. Therefore tags must be less than 1 centimetre in diameter in order to fit on the lid.

Hence the RFID system must be of the following specifications:

- Tag type: passive and rewritable with maximum available memory
- Tag size: small enough to fit in a circle 1 centimetre in diameter
- Frequency: 13.56 MHz
- Protocol: ISO 15693

Operations:

Operations that must be supported by SITS include data operations on the database and the RFID tags. It must support initiating a sample in the process by adding its data to the database, and by adding information about it on its tag. SITS must support updating these data and information on the database and tag. SITS must maintain a log of accesses and operations carried out on it.

Site Adaptation Requirements:

RFID devices must be purchased and installed in the institutes collaborating with the biobank. Computers should be prepared and connected to the RFID devices. Pre-tagged tubes must be made available.

6.2.2 System Functions

Two major functions that SITS must perform are identification of samples and tracking them throughout the various procedures. As mentioned in Chapter 2, in PCRC biobank blood, urine and tissue samples are collected in the initial processing phase and depending on the type of sample they undergo one of the omic techniques, namely metablomics, proteomics, gelycobiology, methylation, epigenetic and genotyping. Upon collection in the initial processing phase, samples should be stored in tagged collection tubes, and after aliquotting, each aliquot should be stored in a new tagged tube. These tubes must be linked for tracking.

UML sequence diagrams provided in Chapter 2 and Appendix 1 on the supplied CD present the SOPs and workflows for collecting samples prior to the implementation of SITS in the PCRC biobank. These diagrams need to be modified in order to incorporate the modified SOPs and workflows required by introducing SITS. The UML sequence diagram for urine collection, given in Figure 2-9, has been modified to reflect incorporation of SITS and is provided in Figure 6-2. The modified UML sequence diagrams for blood collection in EDTA and clotting tubes are provided in Appendix 3 on the supplied CD.

The UML sequence diagram for the secondary processing phase should also be updated to take SITS on board. The blood metablomics that was given in Figure 2-11, is mapped to Figure 6-3 when incorporating SITS into the process. UML sequence diagrams for blood and serum proteomics and urinary DiGE analysis after incorporating SITS are provided in Appendix 4 on the supplied CD. In these diagrams the additional steps are circled in red.

To assimilate SITS into the procedures, the workflow documents and the SOPs for each procedure need to be modified. These documents should reflect SITS according to the following:

- Scanning and storing data on the tags upon arrival of tagged tubes at each location.
- Updating data already stored on the tags. For example temperature changes, storage location and volume should be constantly updated through the procedure.
- Adding new tagged tubes to the process when samples are transferred to new tubes.

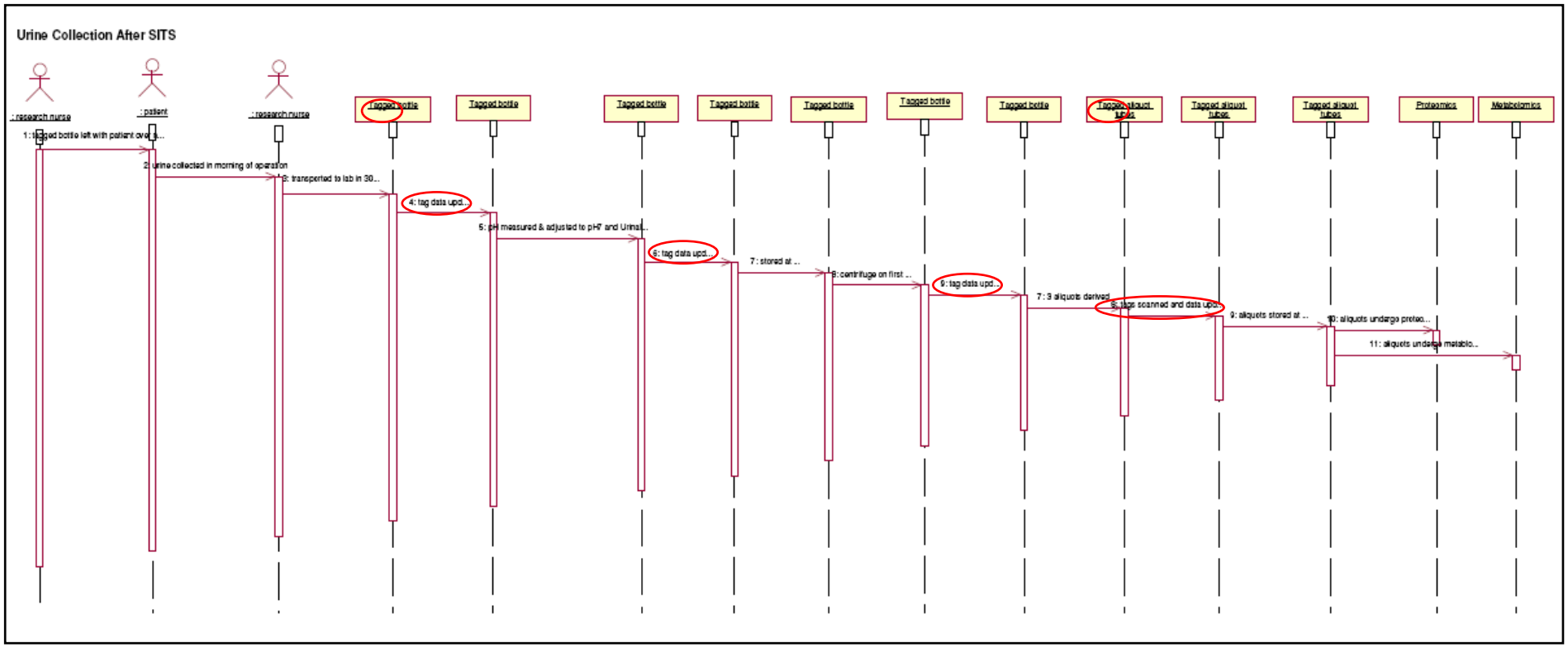


Figure 6-2 : UML Sequence Diagram Incorporating SITS for Urine Collection in the Initial Processing Phase

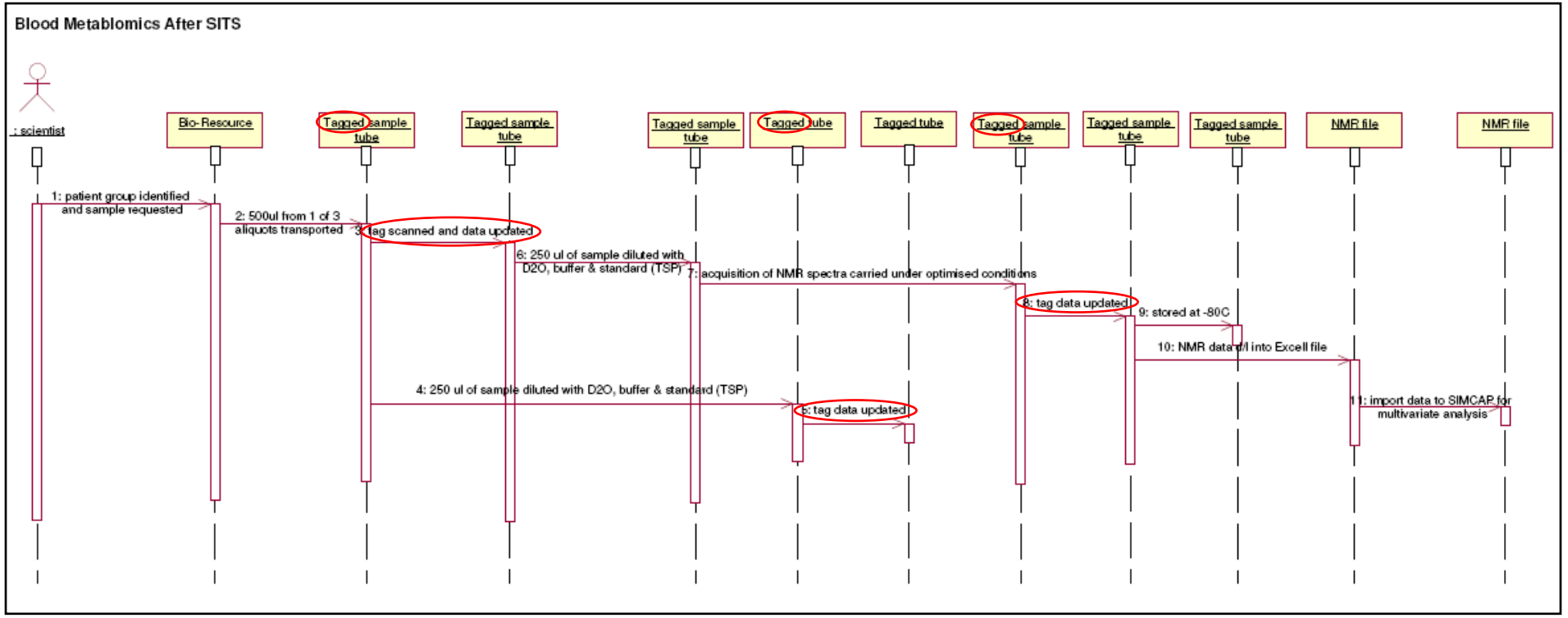


Figure 6-3: UML Sequence Diagram Incorporating SITS for Blood Metablomics in the Secondary Processing Phase

In the initial processing phase the preparation procedure depends on the sample type, and hence data elements for the initial processing phase are type-dependant. For example, all blood samples are processed through the same method, thus they all have the same data elements such as duration of centrifugation and time of processing. However, in the secondary processing phase, each aliquot may undergo a different procedure, so the tags and the SITS database should be designed so that omic data can be stored regardless of what procedure the sample undergoes. Data elements that are expected to be gathered through the process should be created according to the SOPs for each omic procedure. Some of the omic procedures include sub-processes. The omic procedures might be done in different locations and by different individuals. Data generated at each step of an omic procedure should be added to the tag and database to maintain a full record of the sample data and journey.

The UML sequence diagram for processing blood through metablomics is provided in Chapter 2, Figure 2-11. Figure 6-4 shows the proteomics procedure applied to serum samples. While this Figure exhibits the sample journey, it also includes a list of data generated at each step of the process. Data generated in the initial processing phase are labelled as “Tag data”. These data are the ones that should be stored with the sample. In the next phase “Added data to the tag” are listed. These data are generated during the secondary processing phase and like “Tag data” need to accompany samples for the next phase. It is important to note that values of these data items are similar across the small size samples as they originate from a unique aliquot. Furthermore, for the secondary processing phase data, different subcategories of data are included, that is 2D gel data, mass spec and significant differentially expressed proteins between donors. Pathology and biochemistry data are also expected to be added when they are made available.

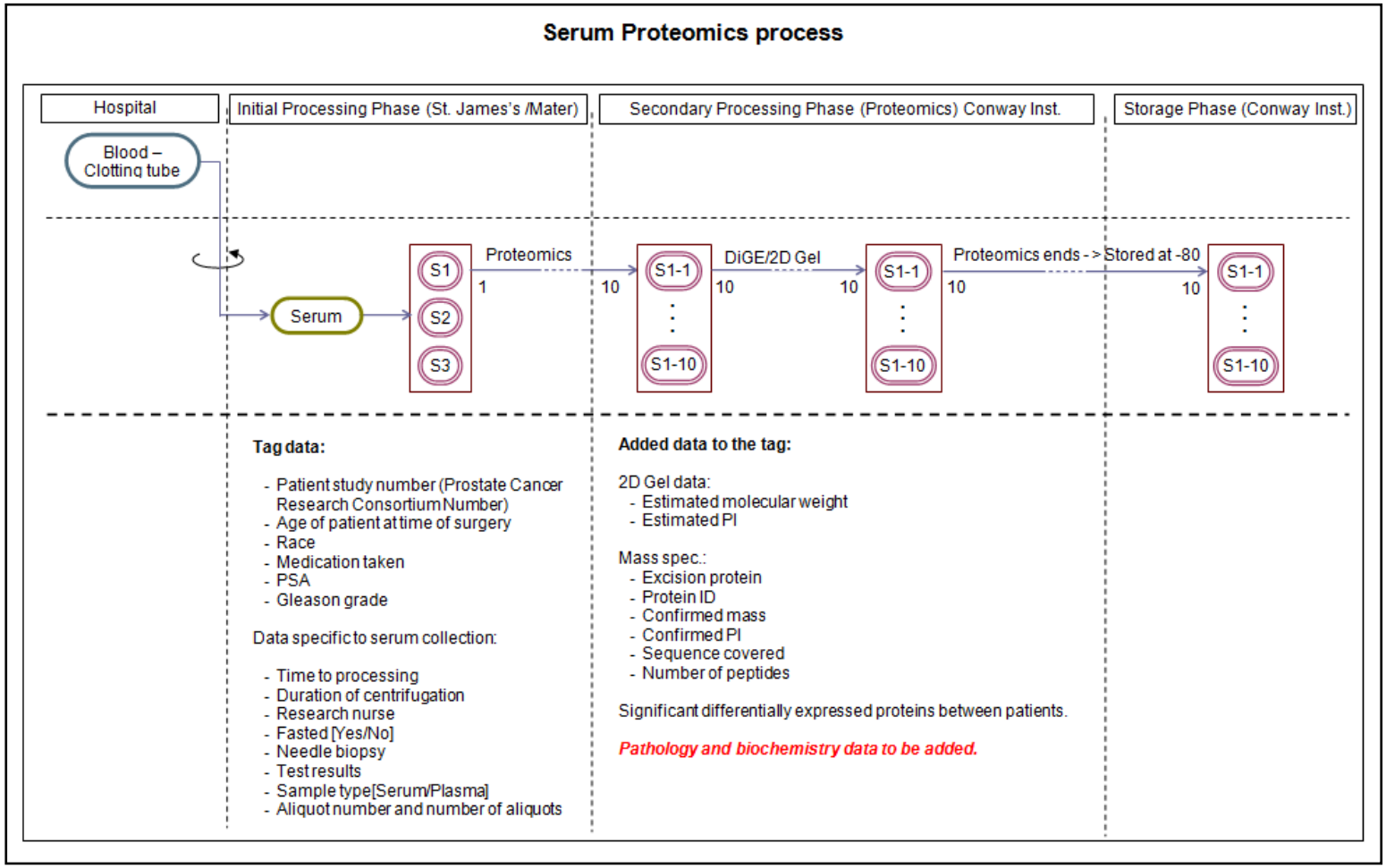


Figure 6-4: Serum Proteomics Process

Workflow diagrams of more techniques and omic procedures are attached in Appendix F. These workflows are drawn based on SOPs describing techniques and omic procedures. The data elements identified for initial and secondary processing phases based on each of their workflows are provided in Appendix G.

With the described processes that samples and data go through in the initial and secondary processing phases, a range of functions must be delivered in SITS. These functions are specifically:

- Identification of samples: Maintaining the link between the physical sample and its associated data is feasible through defining a unique identifier for each sample. There already exists a unique participant identifier.
- Storage of dynamic data on the physical sample: This functionality is required in order to improve sample management.
- Maintaining the records associated with samples up to date: In addition to the data stored on physical samples, SITS should store the most recent data on the BIMS database.
- Tracking samples: The journey of samples and aliquots through workflows must be monitored and recorded. Each aliquot must be linked to its parent sample or aliquot. Parent sample or aliquot is the one from which the aliquot is taken.
- Longitudinal identification and tracking of sample: Samples collected from donors over time should be linkable and identifiable.
- Attribute-based searches on samples and aliquots: Users must be able to query samples and aliquots for specific attributes of their interest. For example they must be able to query information about samples with a gleason grade of their interest.
- Multiple user roles access the system concurrently: Users with different roles must be defined and simultaneous access for each role must be made possible.
- Audit trail maintenance: Secure and automatically generated audit trails of accesses to the system, actions carried out on the records along with date and time stamp should be maintained.
- Administration functionalities: Operations such as adding new accounts, changing passwords and updating user access levels should be made available to appropriate roles.

In addition to above, data exchange with the BIMS central database should also be facilitated.

6.2.3 User Characteristics

The main users of BIMS are the members of the biobank who are mostly non-technical individuals. The three major groups of users are: Principal Investigators (PIs), internal or external researchers and scientists, and research nurses. PIs will fulfil the administrator role from SITS point of view by creating accounts and allocating access levels to users. Researchers and scientists are from a biomedical sciences background, and like PIs, are expected to be competent in the use of computers. Research nurses are expected to have sufficient skills for day-to-day use of computers.

6.2.4 Constraints

There exists a number of constraints imposed on the development of the SITS prototype and may limit its functionality. They are:

- SITS must be implemented with BIMS and its database schema must be the same as the BIMS central database management system. This limits the options available for implementation.
- SITS must be implemented on the same platform as BIMS or have communication links with BIMS. Thus the underlying technology used for SITS must be compatible with the BIMS.
- Metal-intense environment where there are metal freezers and centrifugation devices may limit penetration of radio waves.
- Information collected at each phase of processing, examples shown in Figure 6-4 and Appendix G, should ideally be stored on the tag attached to the container of the physical sample. The capacity of the tag memory may limit the quantity of the data that can be stored on it.

6.2.5 Assumptions and Dependencies

- SITS forms part of the overall BIMS and therefore the requirement for SITS must be compatible with the overall requirements for sample identification and tracking in the BIMS.
- The approach to system security maintenance of SITS depends in part on the functionality provided by the BIMS.
- It is assumed that BIMS supports concurrent access from various institutions as this is one of the SITS requirements.
- From an environmental point of view, it is assumed that samples tags are scanned and identified when out of metal freezers.
- It is assumed that the data required at each stage of the workflow remain as they are described initially
- It is assumed that no other user role is going to be added to the system apart from the three user categories defined earlier in Section 6.2.3.

6.2.6 Apportioning of Requirements

The future version of SITS should be implemented such that it allows flexibility in order to be able to track samples through various workflows in addition to the workflows currently developed.

The future version of SITS should also provide interfaces to other units in the BIMS, independent of the central database. For example, SITS should support facilities required by quality control units in BIMS.

6.3 Specific Requirements

An overall description of SITS has been provided with regard to its various interfaces, functions and users. Constraints that SITS will face, assumptions and dependencies as well as the requirements that will be covered in a future version of SITS are also explained. In this section the specific requirements of SITS will be listed and explained in two categories of functional and non-functional requirements.

6.3.1 Functional Requirements

SR1. Identification of samples

Each individual sample must be allocated a unique identifier. This identifier should be used for identifying the physical sample to which the tag is attached. Data gathered upon sample collection must be stored on the tag and subsequent scanning of the tag should reveal this data. Clinical data associated with sample should also be stored on the tag.

SR2. Maintaining the records associated with samples up to date

Samples pass through a number of procedures while being stored in one tube. Data about the location of the sample and temperature changes that the sample is exposed to in addition to the volume of sample are dynamic in nature and hence records on the tag and the database must be kept up to date at all times. The most recent data should be maintained on the database and tag. The user must be notified if either the tag or the database record is older than the records associated with the same sample stored on the other medium.

SR3. Tracking samples

A sample must be linkable to its parent sample or parent aliquot throughout the procedures, such that children and parent of any sample, whose identifier is given, are identifiable. Therefore the unique identifier of each sample should be stored as part of the data for its aliquots. Samples must be tracked through the initial processing phase, secondary processing phase and final storage and retrieval. It should also be possible locate it at any point in the procedure. The location data should be updated each time a sample is received at a location.

SR4. Longitudinal identification and tracking of sample

Samples collected from a particular donor over time must be identifiable. Participants in the PCRC biobank are identified uniquely by their PCRC NO. Therefore, the PCRC NO and date of collection should be stored as part of the record associated with each sample. Locating and identifying samples with a specific PCRC NO should be feasible.

SR5. Attribute-based queries on samples and aliquots

Researchers and scientists should be provided with a search tool that supports multiple fields and that can be carried out on specific types of samples and on samples at a given processing phase. User rights must be adhered to.

SR6. Audit trail of all accesses

An audit of accesses to SITS containing information about the user login session, the IP address from where SITS was accessed, date and time of access, operation carried out, and the identifier of the sample on which the operation was carried out should be available to administrator account holders.

SR7. Administration functionalities and access rights

Users must only be allowed to access pages permitted by their roles. The three main classes of users of the system must be carefully assigned privileges. While access to SITS must be only possible through username and password, each user must be able to change his/her own password and modify his/her own associated information such as location. Administrator account holders, who are the PIs in the biobank, should be given the option of modifying and deleting accounts and creating new accounts in addition to all actions granted to researchers and research nurses. Research nurses must have full access to the data about the collection of samples in the initial processing phase, while researchers can only view these data. Researchers have full access to the data about the aliquots at the secondary processing phase. However, the research nurses have no access rights to this.

6.3.2 Non-functional Requirements

SR8. SITS must be user-friendly to non-technical users

Users with little or no knowledge of the underlying technical details of databases, computer systems and RFID should be able to work with the system as well as being able to troubleshoot basic errors using a help system. Controlling the RFID device should be feasible through a user interface.

SITS must be developed such that it does not require any other configuration apart from first-time-use port and server configuration. Hence the installation process must be simple on the machines. Setting up SITS on a machine

should be straightforward, and once it is set up it should require no further attention.

SR9. The terminology and format used in SITS must be familiar to users.

The terminology used in SITS must be unified and must match what is used by the users in their activities.

SR10. Quick response time

The total time taken for scanning a tag, retrieving a record from database or tag, writing on the tag and other activities should be within 5 seconds. The user must be made aware at what stage of the process of reading, writing or updating the tag he/she is.

SR11. Context-sensitive messages must be provided by SITS

Tips on links and buttons will be useful to direct the users to the right page and appropriate operations. Users must be notified when they are overwriting data, deleting a record and cancelling a change they are at the process of making. Appropriate messages should be displayed upon the availability of RFID tag in the reading range, number of RFID tags if more than one found, and the status of tag (that is whether it is new or contains data) and if the tag found is the tag that was needed.

SR12. Multiple user roles access the system concurrently

The system should be accessible by multiple users concurrently with no interruption in their activities. Users should only be allowed to the sections of SITS they are entitled to access.

6.4 Requirements Feasibility

The SITS requirements are provided in Section 2.7 and its specific requirements have been formulated in Section 6.3. The SITS which is the subject of this thesis will be developed for use in the PCRC biobank. Hence SITS and the generic system requirements outlined above must be validated in the context of the PCRC biobank. In this Section, the constraints that the PCRC biobank and its BIMS enforce on the system will be discussed and based on these, the feasibility of delivering SITS requirements, functional and non-functional system requirements will be validated.

The SITS that will be developed for the PCRC biobank will be referred to as the "PCRC-SITS" from this point of this thesis onwards.

6.4.1 Constraints Enforced by the PCRC Biobank

The PCRC BIMS uses Distiller, a product of Slidepath, as the user interface to the BIMS database, and access to its code is not possible. Thus, integrating the PCRC-SITS interface with the BIMS interface is not currently possible. Furthermore, Distiller has not yet provided an Application Programming Interface (API) at the time of writing this thesis. Therefore, the PCRC-SITS prototype must be developed independently of the PCRC BIMS interface. It is vital to develop the prototype such that integration with BIMS is possible. Thus, the PCRC-SITS must develop its own approach to integration.

The BIMS database is implemented through an interface that Distiller provides to its administration account holders. Basic operations such as creating data groups as tables are done through the Distiller interface. This database management system does not fulfil the requirements for the PCRC-SITS; for example it cannot be queried by any other system, except by the Distiller search engine that is provided through its user interface. Accessibility to this interface is also restricted for security reasons, therefore the PCRC-SITS prototype must implement its own database and exchange data with the BIMS central database through the Distiller interface. This database will be referred to as the SITS database from this point of the thesis onwards. However, associating records on the PCRC-SITS and the BIMS databases must be feasible through the unique identifiers and common data elements. This will also ease potential integration of the two databases when an API is provided or access to the BIMS central database independent of its user interface is granted.

Thus the PCRC-SITS is effectively a generic self-contained SITS which can potentially be interfaced to any BIMS which provides a standard API. Integration of SITS and BIMS is needed to provide users with one unique system that they can query samples as well as browse data for statistical analysis or knowledge discovery.

6.4.2 Validation of SITS Requirements

R1. Security and confidentiality of data

Privacy and security of RFID have been described in Section 4.4.4 and Section 4.4.5 respectively and the confidentiality maintenance in the PCRC biobank has been described in details in Chapter 3. The security and confidentiality of the PCRC-SITS will be assured as follows:

- There will be no donor identifiable data stored on the tag or on the SITS database
- The short reading range of the RFID tags prevents tags from being read secretly
- Compatible RFID tags and readers are required in order to read or write the information on the tag. However data stored on the tags should be encrypted

R2. Fast, reliable and error free

The response time depends on the reading speed of the RFID devices and the time it takes for the SITS database to respond. The PCRC-SITS will guarantee response time of 5 seconds for scanning tags. The reliability of the PCRC-SITS database and interface will be enhanced by taking regular automatic backups of the data. The system will be thoroughly debugged and in so far as possible error-free.

R3. Item level identification of small biological samples

Item level identification of samples is provided by RFID tagging. The size of the tags that will be used will match the size of the containers.

R4. Data accompanying the small samples tubes at all times

RFID tags can store dynamic information about the samples themselves on their onboard memory.

R5. Sufficient storage space available for storing sample data apart from the sample identifier

The size of the RFID tag memory is expected to be large enough to store sufficient information about the sample. The PCRC-SITS database will support the storage of additional data.

R6. Support complex workflows

The PCRC-SITS supports the entire workflow from sample/data collection through processing to storage and retrieval. In order to reduce the complexity of the workflows, they are broken down into three main phases: initial processing phase, secondary processing phase, and long-term storage and retrieval phase.

R7. Support storage in LN, extreme cold and multiple freeze-thaw cycles

The RFID tags are the main component of the PCRC-SITS that must operate under these conditions. The datasheet and hardware specification of the specific tags should be checked for functionality, survival and reliability in these conditions.

R8. Long life span sufficient to cover long-term storage for several years

Testing storage of tags for several years in extreme temperatures and in LN is not feasible within the timeframe of this research project. However, supporting documents from the manufacturer can be taken into consideration when deciding on the hardware.

R9. Storage of dynamic data at each phase of the workflow

Certain RFID tags are rewriteable and databases support storage of dynamic data. Hence, both media of data storage should support dynamic data.

6.4.3 Feasibility of Functional Requirements

SR1. Identification of samples: This requirement is facilitated by taking advantage of the RFID tag identifiers. Tag identifiers can be used to identify each container uniquely.

SR2. Maintaining the records associated with samples up to date: This is deliverable by the design of SITS.

SR3. Tracking samples: Storing tag identifier of a tube on its subsequent and precedent tubes tags.

SR4. Longitudinal identification and tracking of sample: Using the PCRC identifiers of donors along with the date of collection of tubes can be used to identify samples by their date of collection.

SR5. Attribute-based searches on samples and aliquots: This requirement can be delivered by taking advantage of a comprehensive search engine to query the SITS database where SITS data are also stored.

SR6. Audit trail of all accesses: Maintaining this requirement is feasible and the audit trail data can be stored in a Comma Separated Value (CSV) file.

SR7. Administration functionalities and access rights: User information is stored on the database and their roles are also indicated as part of their data records. Based on their roles, access of each individual user to various parts of SITS is assessed and granted.

6.4.4 Feasibility of Non-functional Requirements

SR8. SITS must be user-friendly to non-technical users: the PCRC-SITS will provide a graphical user-friendly interface. Errors will be carefully handled.

SR9. SITS must be developed such that it does not require installation on individual machines: SITS requires communication with RFID hardware devices and therefore middleware will be needed to make this communication feasible. This middleware must be implemented such that it does not require any additional configuration of settings on the machines beyond the first time.

SR10. The terminology and format used in SITS must be familiar to users: This requirement is satisfied by deploying the system in a language that is understandable by users.

SR11. Quick response time: This requirement highly depends on the RFID device and the SITS database response time.

SR12. Context-sensitive messages must be provided by SITS: Tips on the buttons, links and graphical images are often helpful. This requirement is deliverable.

SR13. Multiple user roles access the system concurrently: Using web as the basis of the system allows such feasibility.

6.5 Summary

Based on the requirements discussed earlier a list of requirements is drawn. RS1 to RS13 are functional and non functional requirements that should be satisfied by the SITS.

SR1. Identification of samples

SR2. Maintaining the records associated with samples up to date

SR3. Tracking samples

SR4. Longitudinal identification and tracking of sample

SR5. Attribute-based searches on samples and aliquots

SR6. Audit trail of all accesses

SR7. Administration functionalities

SR8. The web application and the application program must be user-friendly to non-technical users

SR9. The application program must be developed such that it does not need to be installed on the machine

SR10. Use a terminology and format known to users

SR11. Quick response time

SR12. Having tips and messages appear to users when they need it

SR13. Multiple user roles access the system concurrently

Before implementing a full version of SITS it is appropriate to develop a prototype and evaluate its features. The prototype will focus on a limited number of sample types and procedures. While it only takes certain sample types and omic procedures into account, the principle can be extended for other sample types and procedures. Once a functional prototype is developed, based on the lessons learned a full implementation can be developed.

Chapter 7. Design and Implementation

7.1 Introduction

This Chapter discusses the design and implementation of a prototype for SITS, PCRC-SITS that meets the requirements detailed in Section 2.7 and Chapter 6.

The workflow of samples, data and procedures must be analysed in detail in order to identify the information that should be stored in the PCRC-SITS. A database should be designed based on these data. However while only identification and tracking data are of crucial importance to SITS and the memory available on the RFID tags is limited, only data that are used in PCRC-SITS, that is the identification and tracking data, will be stored in this prototype. The approach taken for storing these data can be applied to other data collected at various phases of the workflow.

The PCRC-SITS prototype will consist of a web application that is the user interface between the user and the PCRC-SITS database, and an application program that is the interface between the user interface and the RFID device. This application program also has an interface that is designed to be used only at the time of set up and troubleshooting. The prototype will facilitate identification and tracking of serum, plasma and urine samples from collection to omic procedures and storage and retrieval.

Section 7.2 will focus on designing the PCRC-SITS prototype. Details of the PCRC-SITS workflow, categories of data produced at each phase of the workflow, data management and database design, and the PCRC-SITS interface to the user will be explained. In this Section, the interface between the RFID device and the user interface will be designed and how the design meets the requirements will be explored. Section 7.3 will discuss implementation of the PCRC-SITS prototype by detailing the hardware used, interfaces to users, to the software and to the hardware. There will then be explanation of how functional and non-functional requirements are met.

7.2 Design

The PCRC-SITS prototype must cover all three processing phases: *initial processing phase*, *secondary processing phase*, and *the storage and retrieval phase*. In this Section, the workflow of samples and data will be discussed and, based on that, a database system will be designed. This database must be displayed to the user through a user interface whose design will also be discussed in this Section. In order to store data on the RFID tags attached to the physical samples, an application program is needed. This application program, or RFID device controller, makes data transmission between the PCRC-SITS user interface and the tag feasible.

The prototype should allow for integration with the BIMS. The PCRC biobank BIMS, Distiller, must run on Microsoft Internet Explorer and the machines that need to access the BIMS should be running on Windows as their operating system. Thus, the PCRC-SITS prototype must also operate on Microsoft Windows operating system and should be compatible with Microsoft Internet Explorer.

7.2.1 Overview of SITS Workflow Design

In the PCRC biobank, samples are processed in three major phases: *initial processing phase, secondary processing phase,* and *long-term storage and retrieval*. The *initial processing phase* is carried out in collaborating hospitals and by research nurses. Aliquots being taken from each sample are then stored in freezers and are transferred to the research institute for the *secondary processing phase*, or "Omic" procedures. The outcome of this phase is tens or hundreds of small size samples as well as their procedure data. Depending on the procedure and the sample characteristics, the number of samples varies. Data gathered during analysis on the small samples is used for data mining and knowledge discovery. The remainder of samples are then stored in freezers, where the *long-term storage and retrieval* phase is undertaken. Figure 7-1 depicts the sample flow from initial phase to secondary and where the long-term storage and retrieval phase commences. It is clear from Figure 7-1 that a large number of samples are produced by the omic procedures. These small samples must all be tracked and identified. Due to their size and number, it may not be feasible to tag every single sample. However, as they have identical characteristics and the same parent,

they can be batch tagged, that is, the box that they are stored in is tagged, with extra fields in place for the samples that might be removed.

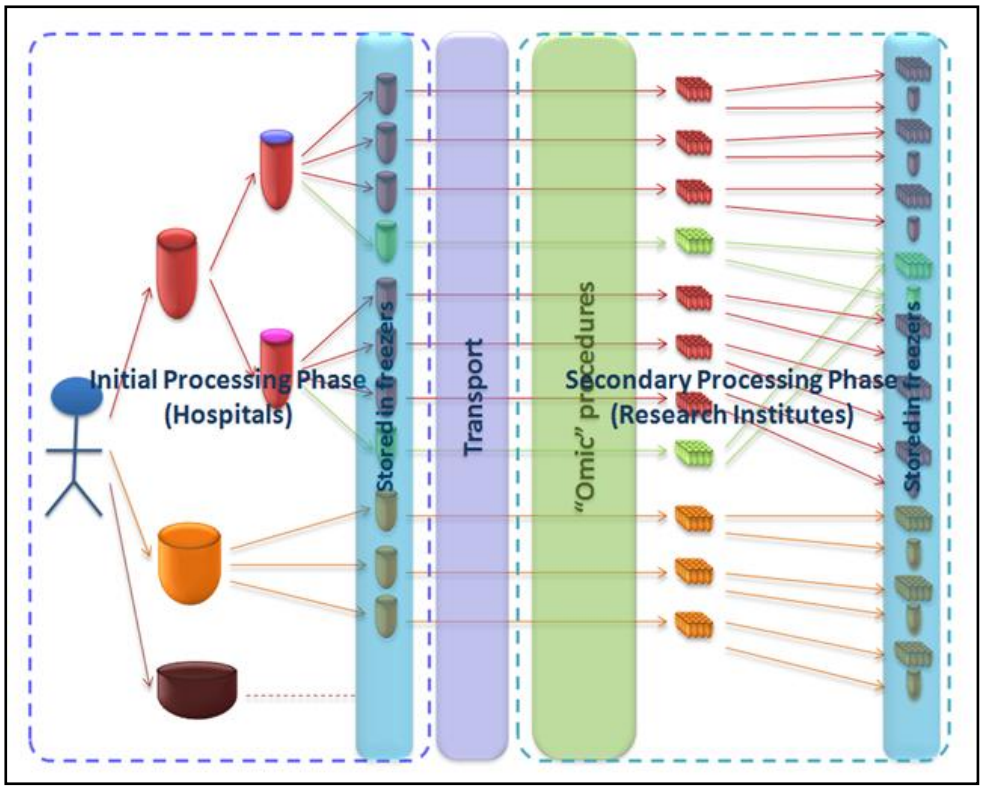


Figure 7-1: Sample Journey

The tubes used for storing aliquots in the PCRC biobank are shown in Figure 7-2. The RFID tags that will be used for tagging these tubes should be of a size that fits onto the lids of these tubes. These lids are round and have a diameter of about 1 centimetre. As described in Section 6.2 the RFID system that will be used in the PCRC-SITS must operate in 13.56 MHz frequency band and must be compliant with ISO 15693 standard. Attaching tags to the tubes must be done in a secure manner. Although the lids of this type of tubes currently in use in the PCRC biobank are not connected to the tube body, where the physical sample is stored, there are other types of tubes available where lids are connected to the body via a plastic loop. Separation of tagged lids risks data loss or mix-up, and as a consequence samples would become unidentifiable or supported with unrelated data. Thus, it is of crucial importance that the lids and corresponding tubes are always kept together.

Figure 7-2: Tubes Used in the PCRC Biobank

Consultations with researchers and scientists revealed four categories of data that should be stored on the sample to gain full benefits of the scientific data. These four categories of data are:

Type-specific data: This is collected in the initial processing phase and depends on the type of sample. For example, urinalysis is a data element or data field that is only meaningful in the context of urine samples. Therefore, urinalysis is considered as a type specific data for urine samples.

Procedure-specific data: This is collected in the secondary processing phase and is common across an omic procedure. For example, pattern of genes is a procedure specific data to the DNA epigenetic process.

General data: This category of data is very important for PCRC-SITS as it covers location data of samples, their identification, monitoring temperature data and tracking data. The PCRC identifier is an example of a general data element for a sample, regardless of its type and the procedure it has gone through.

SITS data: This category of data is collected by SITS as the sample travels through the procedure, and include data about identification of a sample, its location data, temperature changes and duration of exposure to the temperature, and tracking data.

A diagram showing these three categories of data, *sample data*, as well as the *SITS data* is show in Figure 7-3. *General data* about a sample in addition to *Type-* and *Procedure-Specific data* are labelled as *Sample data. Sample data* and *SITS data* are then stored in SITS. It is important to note that *SITS data* are collected as samples go through various phases and are considered the vital requirement for SITS. Data

collected at each phase of the workflow is depicted graphically in Appendix 3 on the supplied CD. *General data*, *Type-specific data* for blood and urine, and *procedure-specific data* for proteomics, DNA epigenetic and DNA genotyping are tabulated in Appendix G.

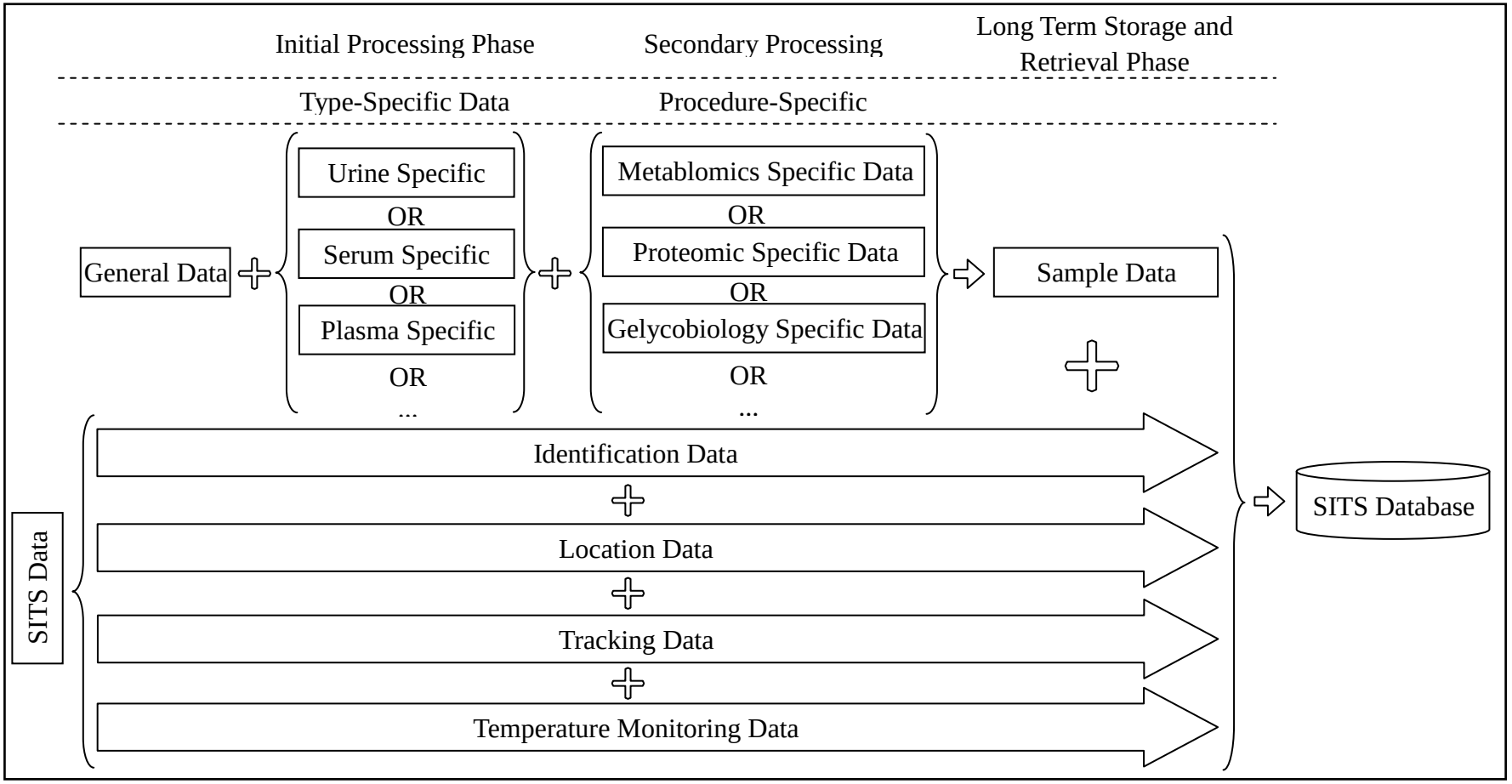


Figure 7-3: Data Collected at Various Stages of the Workflow

7.2.2 Management and Representation of Data in the SITS Prototype

As access to the PCRC BIMS central database is restricted, a database should be designed for PCRC-SITS. The SITS database will need to house the three categories of *sample data*: *type-specific data*, *procedure-specific data* and *general data* in addition to the *SITS Data*. Data stored in the SITS database and data stored on the RFID tag's memory should be identical. Thus the data elements are also the same. It is anticipated that there will be a limited amount of memory on the RFID tags so it is likely that only *SITS data* will be stored on the tags at the time of implementation. *SITS data* are the main data that are required for sample identification and tracking as shown in Figure 7-3. SITS data include *Identification Data*, *Location Data*, *Tracking Data* and *Temperature Monitoring Data*. Therefore, in this prototype only *SITS data* will be taken into consideration for storage on the SITS database and on the RFID tags memory. However, storage of more detailed data on the SITS database is feasible.

Database

The PCRC-SITS prototype database can be implemented in Microsoft Access or MySQL. Although Microsoft Access or MySQL are not the most robust database management systems, they are suitable for prototyping purposes. Schema of the PCRC-SITS database and tables relationships designed in Microsoft Access is shown in Figure 7-4. Details of all tables and their associated fields are given in Appendix H.

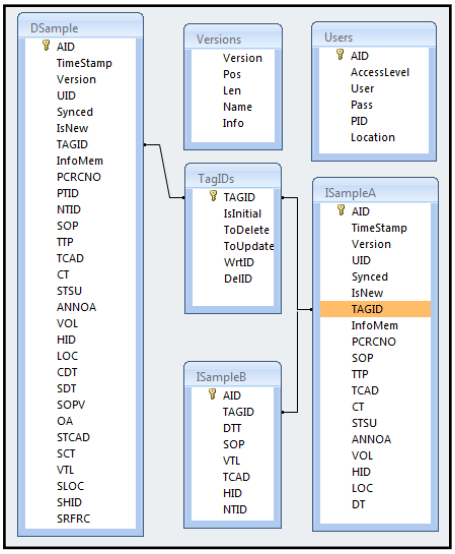


Figure 7-4: SITS Database Schema and Tables Relationships

The unique identifier of each sample is the RFID tag identifier, “tagID” field. This is stored on the tag memory at manufacture and is not erasable. This identifier is used for unique identification of samples, and the identifier of each donor is the PCRC number, “PCRCNO”, allocated to that donor.

When the user adds a new initial sample, that is, the sample collected from the donor in the initial processing phase, the initial data includes the date of collection, duration of centrifugation and many more data elements. Since three aliquots are derived from each initial sample, there is an extra table for storing the data of these three aliquots, before they are further processed. These are part of the *Tracking data*. A table, named “ISampleA”, will store the collection data of an initial sample that are gathered at the time of collection from the participant and another table, called “ISampleB”, will house data associated with the three aliquots derived in the initial processing phase. “ISampleA” and “ISampleB” tables should be written when the add-initial-sample operation is carried out. Updating records follows the same procedure, except that the tagID of the sample to be edited will be looked up and its associated data will be updated. The link between these two tables is the tagID of the initial sample and the previous tagID of the aliquots.

Similarly, when the user adds a derived sample and carries out the add-derived-sample operation, data is stored in a table. This table is called “DSample”. It assists users in learning about the origin of aliquots, as data related to the initial sample from which the aliquot is derived are retrieved from “ISampleA” and “ISampleB” tables. The data related to the initial collection of samples is retrieved from the “ISampleA” table and the basic data about the creation of aliquots is retrieved from “ISampleB”. The tagID of that record is used to update the record.

A table for storing user information is needed, so that when a new user account is created its data is stored in this table. Let the name of this table be “Users”. Update and deletion of the record is reflected in this table.

One of the approaches that the web application takes is storing the combined data that is sent to the application program on a data field in the initial sample or the derived sample tables depending on the sample type. The combination format is stored in a

table. This table is used to state for each of the initial and derived samples, how long each field of data is and from what position in an array they should be appended. This table is called “Version”.

It is important to note that in the current design changes made to records on the database, will over-write the existing data on the PCRC-SITS prototype database. Thus maintaining a full history of all records on the database is not supported in the prototype. However this functionality must be taken into consideration when deploying full implementation in order to conform to 21 CFR Part 11, GLP and GCP. Although the data are over-written on the database, the audit trail maintains old records that are updated so a full history of data can be re-created.

Interface

An interface to the PCRC-SITS database should be provided to users. This interface is designed as the main PCRC-SITS interface that users will interact with. The interface for SITS is available to geographically dispersed users and is delivered through the internet via a web application. The web application is implemented by PHP, ASP or other web programming languages. The web application is only available to registered users who are allocated usernames and passwords. Different user roles are defined and access levels on the pages available to each user or role are carefully implemented.

According to the requirements as detailed in Chapter 6, three user roles are taken into consideration when designing the system. The PIs, researchers and research nurses, who will have access to data collected at various phases, are the main users from biobank. PIs will hold “administrator” role. In addition, there need to be two extra roles, “guest” and “root”. The “guest” account is designed for providing temporary access to the database with no right to edit or delete data. Deletion of data is only allowed when the participant has retracted his/her consent. The “guest” role provides read-only access to the users. There is only one “root” account. “Root” account is designed for high level access and unlike all other types of accounts cannot be deleted. The holder of root account has all of the access rights that an “administrator” account has.

The web application has several sections including an administrative section and sections for managing data about initially collected sample (“initial samples”) and managing data about aliquot or derived samples (“derived samples”).

In the administrative section, the user can change their password, choose a RFID reader/writer device and edit, add or remove accounts depending on the permissions allocated to that user role. Only the “administrator” and “root” account holders are permitted to add or edit other user accounts.

Sections for managing data about initial samples and derived samples allow for:

Listing all samples: This refers to a page where all samples are listed. This list needs to include tag ID, PCRC NO of the participant, sample type and other optional fields. There is also a key for each row that leads to the full detailed record of each sample. The detailed record page should include edit, delete and cancel keys.

Adding a sample through a form: The fields are filled automatically where possible. There is an add key to submit the form, and a cancel key to suspend the action.

Editing or deleting data records associated with a sample: Editing is best done through a form similar or identical to the form used for adding samples. There is an “update” key for updating the record, a “delete” key to remove the record and a cancel key to abort action.

Searching for a specific sample: Users can query samples. This is expected to be done through a form similar to the add sample form except that certain unsearchable fields can be omitted to avoid complication. In cases where the results return more than one record, these should be displayed in a format where the user clicks on the record to see a detailed view. If only one record is returned as the results, the page should go the detailed view directly. In the detailed view users must be allowed to edit or delete the record, so there must be three keys: “edit”, “delete” and “cancel”.

7.2.3 RFID Device Controller

In addition to the web application that is the user interface to the PCRC-SITS database and is the main PCRC-SITS interface, an application program, called “SITS Win32

Controller", supports communication with the RFID device and hence with the RFID tags. SITS Win32 Controller needs minimum configuration by the non-technical user. Figure 7-5 represents a schematic diagram of the RFID device and the user interaction with the designed prototype. When the "User" decides to update data records associated with a sample, they browse the database and query the record through the "Internet Browser" that displays the Web pages from the "Web Server". This process is depicted by number 1, 2, 3 and 4 and the data is displayed to the "User" through the "Internet Browser", shown as number 5 and 11 on the Figure 7-5. The "User" then clicks on the "update" key on the Web page displayed, number 1. The TagID of the record to be updated is then sent to the "SITS Win32 Controller", number 6. This TagID is then sent to the "Reader/Writer" device, number 8, from where it will be sent to the "Antenna", number 9. If the "Tagged Tube" that is in the reading range has the same TagID, data transmission is carried out based on radio waves, labelled as number 10 on Figure 7-5. Once the data on the tag is updated, the acknowledgment of the update action is sent to the "SITS Win32 Controller" through the path of number 10, 9, 8 and 6 and is displayed to the "User" through number 7 and 2. Adding a record and reading tags is carried out similarly.

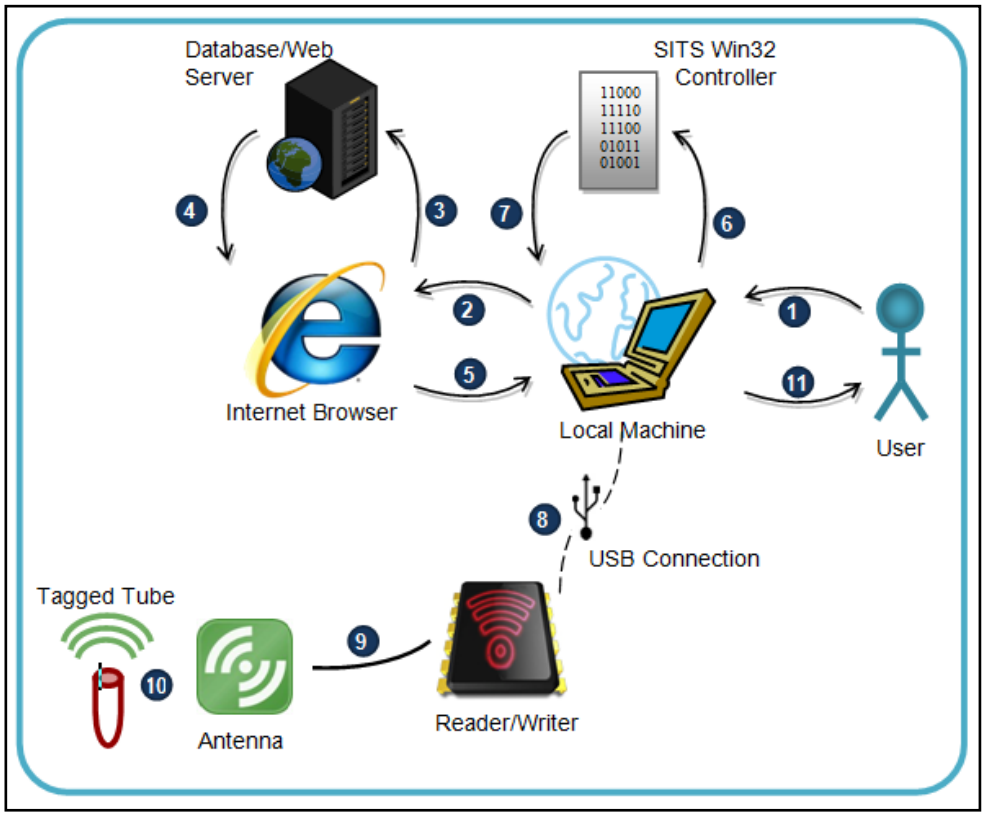


Figure 7-5: Schematic Diagram of RFID and User Interaction

Functionality

The SITS Win32 Controller is used to communicate with the RFID device and is developed based on the documentation supporting the particular RFID equipment. In order to establish and maintain the communication with the RFID device regardless of equipment specifications, SITS Win32 Controller should support the following operations:

- Turning the RFID device on/off
- Resetting the RFID device
- Checking the status of the connection constantly to report it to the web application
- Opening the port to establish the connection with the RFID device
- Setting the baud rate to what has been stated in the supporting documentation
- Counting the number of tags in the reading range
- Reading data from the tag. If there is only one tag present the entire memory of the tag that could be read. If more than one tag is present, the tag ID of all of the tags could be returned. Otherwise appropriate messages should be displayed.
- Writing data on the tag if there is only one tag in the reading range and the tag ID of the available tag matches that of the data record. Otherwise an appropriate message should be displayed.
- Displaying, hiding and clearing the buffer of the data and commands sent to the RFID device

User Interface

PCRC-SITS is designed so that the only interaction between the user and SITS Win32 Controller is when the system is being set up for the first time and when it needs troubleshooting. The following operations are manually possible through the application program interface by the user:

- Starting and killing the RFID device by turning it on and off through a key
- Learning about the status of the RFID device, if it is connected, disconnected or busy. A message should be displayed to the user and an icon on the tray bar of windows should be displayed.

- Resetting the RFID device using a key in case it did not act as expected
- Deciding on the port that the RFID device is connected to by selecting the port and clicking on a key. In case the RFID device is not connected to the given port a message is displayed and if it is connected a successful message is displayed. The message could be the specification of the RFID device connected.
- Clearing the buffer of the RFID device using a key. This is particularly useful in case an operation is not completed through the web application and the buffer is left full.
- Viewing the transmitted data between the RFID device and web application for troubleshooting.

Interface to SITS Database Interface

The SITS Win32 Controller must work closely with the web application, labelled "Internet Browser" in Figure 7-5. Communication between SITS Win32 Controller and the web application is based on command-response. The architecture that fits the needs for such communication is best delivered through REpresentational State Transfer (REST) software architecture or RESTful Web services (Fielding, 2000). RESTful Web services are based on four principles: resource identification through Uniform Resource Identifier (URI), "uniform interface", "self-descriptive messages" and "stateful interaction through hyperlinks" (Pautasso et al., 2008). Using RESTful, the client sends a request to the server and receives responses, otherwise the server and the client are at the rest status. In this architecture, the packets sent and received are small in size. Thus it is preferred over other architecture standards for applications with a high number of transactions. In the SITS scenario, the client is the application program and the server is the web application.

The SITS Win32 Controller launches a connection with the web application. This is stored on the server labelled "Database/Web Server" in Figure 7-5, and awaits commands from the web application. The web application should constantly check the availability of the RFID device. When the user submits a form, the data, in addition to being sent to the database, are combined using the "Version" table and sent to the SITS Win32 Controller. From there they are written on the RFID tag. Similarly, when the user decides to read data from a tag, a command will be sent to SITS Win32 Controller

in response to its request. SITS Win32 Controller then returns what it has received from the RFID device.

The response from SITS Win32 Controller should cover possible errors and confirmation messages. It could return the received command with the appropriate message.

The SITS Win32 Controller should log the sent and received data from the web application in a buffer. This will be particularly useful for troubleshooting purposes.

Integration with the PCRC BIMS

Three approaches are available for integrating SITS with BIMS, however depending on the implementation approaches taken for the PCRC BIMS and SITS, integration can be deployed. These three methods are:

1) Application Programming Interface (API): API has been defined as:

> "a set of standard program functions and commands that allow any programmer to interface a program with another application" (Collin, 2004)

Interaction between two system components based on an API depends on how the system is implemented. APIs can appear as functions provided by the Operating System (OS) to the developer or as a Web API interacting with other Web applications. The API is provided by the developer of a system and can be thought of as a standard that the developer of a system provides to the outside world.

The PCRC BIMS will need to provide an API that allows two-way data transmission between the two systems. The PCRC BIMS database will need to include tables for sample identification and tracking.

2) Service Oriented Architecture (SOA): SOA is an architectural approach to integration for services that is based on loose coupling between software components to allow reusability (Mahmoud, 2005). In this context a service is:

> "an implementation of a well-defined business functionality, and such services can then be consumed by clients in different applications or business processes." (Mahmoud, 2005)

By using web services to allow interaction between various machines on a multi-institutional biobank network, one can take advantage of SOA-based standards. Web services allowing interoperability and interaction over a network can be implemented by standards such as WSDL, SOAP and UDDI (Mahmoud, 2005). Biobanks can take advantage of SOA if interaction of their BIMS with a service is possible. In this case SITS must be developed as a service or module that will be used by BIMS and since the code for PCRC BIMS is not accessible this method cannot be used for integrating with the PCRC BIMS.

3) Dynamic Data Exchange (DDE) and Object Linking and Embedding (OLE) Automation: DDE protocol refers to methods provided by Microsoft to support the transfer of data between two applications (Microsoft, 2009a). Using DDE the memory is shared between the two applications and once data is updated by one side of the transaction, the other side receives the update. In DDE the basis of data exchange is shared memory.

API and SOA methods for integration are not feasible with the PCRC BIMS at its current status and thus DDE is the best option available. However, DDE will require the BIMS database to support data from SITS. DDE allows small flexibility such that any modification in BIMS and SITS should be reflected in the system.

OLE Automation, also termed Automation by Microsoft, allows access to objects in other applications (Microsoft, 2009b). The owner of the application that provides services is then the application accessing it. It can be thought of as a user of a service, where the user is the Automation Client and the service provider is the Automation Server. Remote Automation is carried out over a network of computers, where Automation with no prefix refers to the relationship between the two applications running locally. It is important to note that the Automation Server object must be created by the Automation Client to allow accessibility.

OLE Automation can be used when the only method for accessing BIMS is through its interface which is the case with the PCRC BIMS. With OLE Automation, like DDE, the BIMS database should support SITS so that the data transferred are storable in BIMS. Any modification to the interfaces of SITS and BIMS as well as their databases must be reflected in OLE Automation. This approach and DDE are not robust in terms of flexibility.

7.2.4 System Functions and Attributes

The PCRC-SITS prototype should be designed as below to meet the requirements.

Identification of samples (R1)

For initial samples, the user clicks on a link directing to "add initial sample" page, inputs the data and scans the tagID of the tag. Upon submitting the form, a copy of the record is stored on the database and the combined record will be sent to the SITS Win32 Controller, from where it will be stored on the tag. This data is used for identification of samples. An appropriate message is displayed to the user at each stage of this process as required by R12. Any data written on the tag, updates its tag header. The tag header is used for basic identification purposes such as checking if the sample belongs to the PCRC biobank, when and by whom it was last written and the type of sample it is.

Similarly, the user clicks on a link directing to "add new derived sample" page to add a derived sample. Part of the form in this page is filled in with the data that are retrieved from the initial sample data, when the user scans the tagID of the derived sample.

Subsequent scanning of the tag reveals the tagID and data, and based on the "Version" table they are separated and displayed to the user through a form on the web application. This form follows the same structure as the one used to input data in order to comply with R8.

Maintaining the records associated with samples up to date (R2)

When a user updates data about a sample on the web application, this modification is reflected on the database and the combination of the new data is sent to the application program. If the tagID of the only tag present in the reading range matches that of the updated record, it is updated right away. Otherwise, an appropriate message is displayed to the user stating that there was more than one tag in the reading range, there was no tag in the reading range, or the tagID of the only present tag does not match that of the record (R12). The record should then be flagged in the database, so that it is updated when it is found in the reading range. This could be done through a page that contains the tagID of all of the outdated tags.

Tracking samples (R3)

Data for each sample, regardless of its type, include the freezer in which the sample is stored and the institute where the freezer is located. This data should be updated every time a user receives a sample or moves it.

Also to track samples from their parents to their children, the tagID of the initial sample is stored as part of the data of the three aliquots taken from it. In addition to these tagIDs, PCRC number, sample type, collection SOP version, aliquot number and number of aliquots, original volume, time to initial processing, date of collection from participant, ID of the person who collected it originally from participant and location of the original collected sample should be looked up automatically from the collection sample data. These data are stored in relevant fields in the database as shown in Appendix H. These fields from Figure App. H-1 are PCRCNO, STSU, SOP, ANNOA, VOL, TTP, DT, UID and LOC, respectively. Thus, every initial sample should be linkable to derived samples that are in turn linkable to their parent and children. Therefore, each derived sample will have at least three fields for storing tagIDs: its parent tagID, its own tagID and the tagID of any other aliquots taken from it. These tagIDs are stored in “PID”, “TAGID” and “NID” fields from Figure App. H-3, respectively. This is achieved through the use of a double linked list.

On the web application, the user scans the tagID of the initial sample and enters its data along with the tagIDs of three aliquots taken from it that are also scanned. It is important to note that one tag can only be read at a time so if there are more than one

tag in the reading range, a message will be displayed to the user. There will be no data about the aliquots on them until they are added using the "Add derived sample" link on the web application. In fact, only the initial sample data includes these tagIDs and basic data such as date of collection. In the "Add derived sample" page, the user scans the tagID of the derived sample. This tagID is then found in the database, that is "ISampleB" table, and the tagID of the initial sample along with its data are looked up and part of the form is filled. Further, if another sample is derived from this derived tag, its tagID is stored as part of the data for this record. When the new derived sample is added, again, data from its parent and initial sample are looked up. This process may go for several generations.

Longitudinal tracking of samples (R4)

Samples are uniquely identified by their tagIDs and participants are uniquely identified by their PCRC NO. Any sample collected or derived should have the date of initial collection from participant as part of its data, therefore samples from a PCRC NO can be identified. The system should be such that by clicking or searching a PCRC NO all records associated with that PCRC NO are listed.

Attribute-based searches on samples and aliquots (R5)

A comprehensive searching tool is provided to users. It allows users to search for the attributes of interest and supports searching on more than one data field. Users may search on initial samples or derived samples. Then they can choose to view details of a record in the returned results and decide to update or delete the record.

Audit trail of all accesses (R6)

Actions carried out by the users on all the main pages of the web application are recorded and logged. The date, time, Internet Protocol (IP) address, username, action and tagID if applicable are recorded. Successful and unsuccessful logins to the system should be audited. These records are stored on a spreadsheet document such as a Comma Separated Value (CSV) file and are downloadable from the server to appropriate roles.

Administrative functionalities (R7)

The web application is available through username and passwords and each page is only accessible by roles that are permitted to see it. In addition to the three roles introduced in Chapter 6, that is, "administrator", "researcher" and "research nurse", "root" and "guest" accounts should also be defined. The "administrator" accounts are the accounts declared for PIs.

Each user may change his/her password and data by clicking on a page that leads to "change password" page. They must enter their existing password in addition to new password that must be entered twice. An appropriate message is displayed to confirm that the password has been successfully changed. Although an account password is not visible to other account holders, passwords are currently stored as plain text in the local database that is only accessible locally. The local path on which the database is stored is used by the web application. It would also be possible to store passwords as encrypted fields.

Also, in the administrative section, permitted user roles are able to add new user accounts by clicking on a page that leads to "add new user". They can enter username, password, name, role and location for each user. Drop down menu or radio buttons should be used where possible, for example, a list of omic procedures that the sample may undergo. Editing and deleting should also be possible on the user accounts. The user with relevant permission is able to log in to the system and on a page that allows editing or deleting accounts choose the account that needs modification. By clicking on a key he/she is able to view full data for that user with the option of editing role, name and location, and deleting the user account. A confirmation message is displayed.

Multiple user roles access the system concurrently (R13)

Each RFID device is allocated a unique identifier to allow concurrent availability online. A user can connect his/her RFID device on a local machine, run the application program and access the web application through his/her username and password. Users may browse their RFID device on the web application or on the application program. They can also choose a user-friendly name for their RFID devices. In this

way multiple accesses to the web application and application programs are feasible with no interruption. It is also possible to list the RFID devices and users that are online concurrently.

7.3 Implementation

A description of the requirements for SITS and its system were provided in Chapter 6 and the system design was supplied in the previous Section. These enable a PCRC-SITS prototype to be implemented. This system will be based on:

- VB for the application program
- ASP and AJAX for the web application
- Microsoft Access for SITS database management system
- RFID equipment purchased from TAGSYS RFID

Details of these technologies will be described in this Section.

7.3.1 Hardware

RFID equipment used for this prototype are from TAGSYS RFID. ARIO 370 SDM tags, shown in Figure 7-6, are attached to the lid of the tubes by SA 6101 adhesive from Flexcon as displayed in Figure 7-7.

Figure 7-6: ARIO 370 SDM Tags

Figure 7-7: Tagged Tube

An issue that had been taken into consideration is the approach taken to affix the tags to the tubes. RFID tags in case of being flexible can be twisted over the body of the tube. However this may affect their functionality and they may break in freezers when they are frozen and become fragile. ARIO 370 SDM tags are round and rigid and their diameter is 8.9 mm; they fit well on the lid of small tubes. However the lids of the tubes typically used in the PCRC biobank are not connected to their body and this may become another source of error. It is very important to have the tags attached to the tubes, attached to the lid that cannot be separated from the tube body or implanted in the tube. Attempts to implant tags on vials have already been carried out, for example Smart Test Tube™ shown in Figure 7-8 (Smart Medical Technologies, 2005) and BIOtrack® RFID as illustrated in Figure 7-9 (MAINtag, 2008). These tubes are not used for testing the SITS prototype since the PCRC biobank has deployed its standard tubes. However, tubes with implanted tags would be a better choice than attaching tags to tubes.

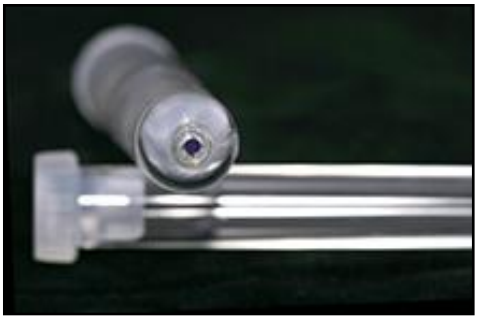


Figure 7-8: Smart Test Tube™ (Smart Medical Technologies, 2005)

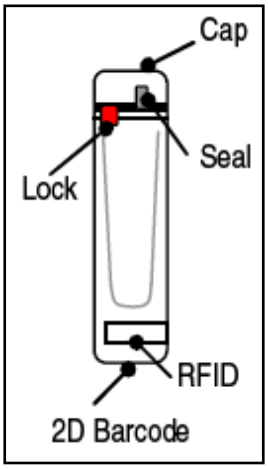


Figure 7-9: BIOtrack® RFID (MAINtag, 2008)

ARIO 370 SDM tags can store up to 1280 bits of user data, excluding the tag manufactured identifier. The tag and adhesive have been tested against extreme temperature in freezers, multiple freezes and thaws in conventional freezers, extreme temperature and multiple freezes and thaws in Liquid Nitrogen vapour. Details of these experiments are provided in 0. These experiments show that R15 and R16 are satisfied by this type of tag and adhesive. The entire data transfer from tag to the reader occurs on 13.56MHz frequency band and by the ISO15693 standard and is done by Medio P101 reader/writer from TAGSYS. The antenna that matches this device is a 5 turn spiral antenna also from TAGSYS.

7.3.2 User Interfaces

As shown in Figure 7-5, when a “Tagged Tube” is placed in close proximity to the antenna and its data is requested, data on the tag is read and transferred to the “Reader/Writer” device which in turn is connected to a “Local Machine” through a “USB Connection”. Two main components that make communication between the user and the tags feasible are:

- “SITS Win32 Controller” running on the local machine that the reader device is connected to. SITS Win32 Controller can be considered as a middleware preparing data for being sent to or received from the Reader/Writer.
- The web interface that is accessible by users via “Internet Browser” and by their predefined usernames and passwords. The PCRC-SITS database and web pages are stored on a server labelled “Database/Web Server” in Figure 7-5.

Web application

When the user logs into the system through the web interface, he/she will be directed to the main page that lists all the tasks that can be done by that user under his/her role. Figure 7-10 displays the first page for administrator accounts.

Links to other pages will be hidden for roles with limited permissions. The *User Management* section includes the links to administrative operations, allow for password changes and modifying other accounts depending on the account logged in.

Initial Samples and *Derived Samples* supply links to pages that are associated with actions that can be carried out on the initial samples collected in the initial processing phase, and the derived samples that are aliquots being taken in the secondary processing phase, respectively. These operations include listing and searching samples, reading tags, adding new samples, checking the update list and editing/deleting samples.

The *Security* section includes the link to the log file that can be downloaded by authorised roles.

Status Monitor shows the users who are online at the same time along with their roles as well as the RFID devices that are connected to the system and their status. Depending on the role access right, each section can be shown or hidden from the users with that role. Also by clicking on the blue tapes that separate sections, each section can be minimised. More images from each page are given in Appendix J.

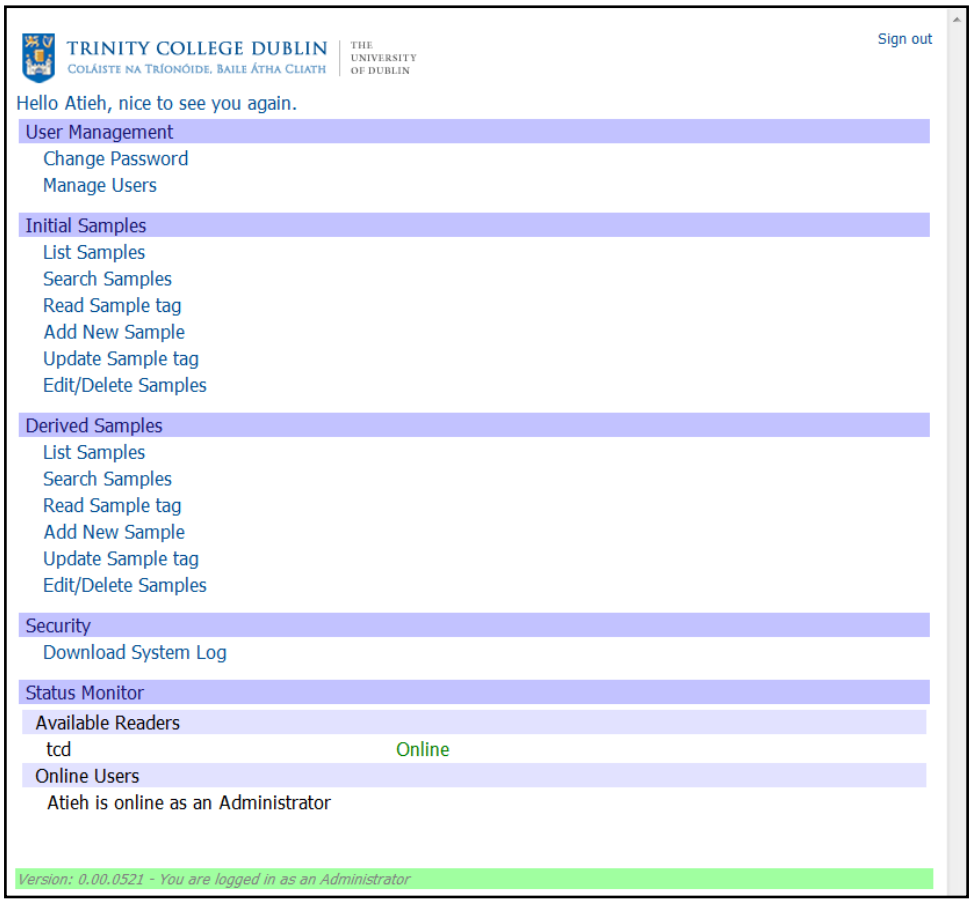


Figure 7-10: First Page of SITS

“List samples” displays all samples that are already in the database for each type of sample. The user can view details of each by clicking on the “View Details” option. The user can edit or delete samples in the same page, depending on privileges defined for his/her role. If editing the tag, it has to be placed in close proximity of the antenna so that data can be written on the tag and database simultaneously. If the tag is not in the reading range of the antenna or writing data is not successfully completed, then the tagID will be added to the to-be-updated list. Deleting can be done by requesting delete operation by non-administrator users or directly deleting by administrator accounts by clicking on the “Edit/Delete Sample” link. The “Reading sample tag” page allows the data that is stored on the tag to be read. A tag can only be read if it is written by the PCRC-SITS and according to its protocols.

A list of the tags to be updated is maintained so that if at the time of writing on the tag, the tag is moved out of range; its TagID is preserved. The tag can be updated to match the database at any time and under any login access right. Otherwise a field of the table that checks the sync status of tag data continues to display “No” and the TagID remains in the to-be-updated list. This list that is displayed in “update sample tag” page, only reveals the TagID, PCRC No and whether it is an initial or derived sample.

The date at which the aliquot is derived, SOP version, volume taken and left in the main tube, temperature changes and duration and the handler ID are the data that should be stored individually for each sample and aliquot. Their tagID should be scanned in by the user in the “Initial Sample List” page and “Add Derived Sample page”. For box objects, combinations of row and columns can be added to aliquot data to specify which individual sample is removed from the box.

The use case diagram of the system is shown in Figure 7-11. Also a detailed architecture diagram of the system is provided in Figure 7-12 and Figure 7-13 for initial and derived samples, respectively. The interaction between various components are also displayed in Figure 7-14 and Figure 7-15.

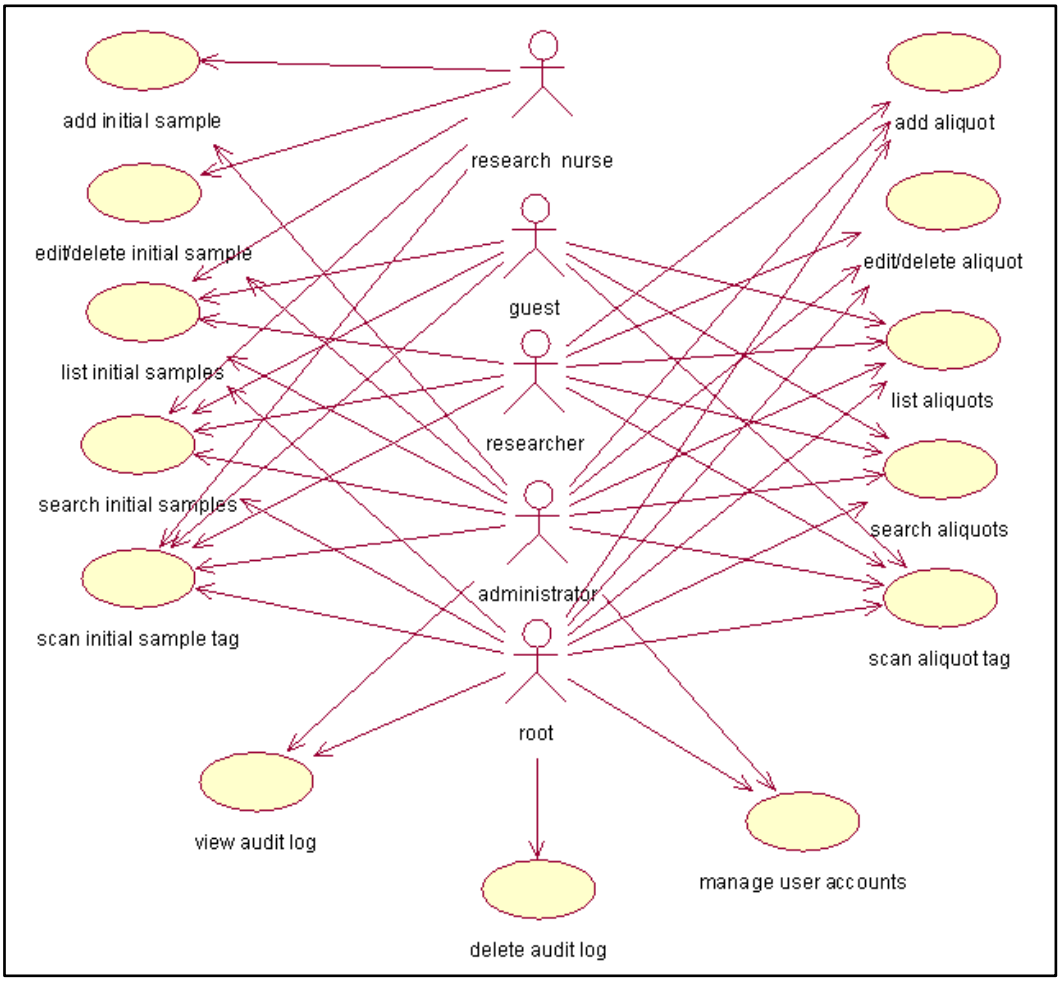


Figure 7-11: The Use Case Diagram of the PCRC-SITS Prototype

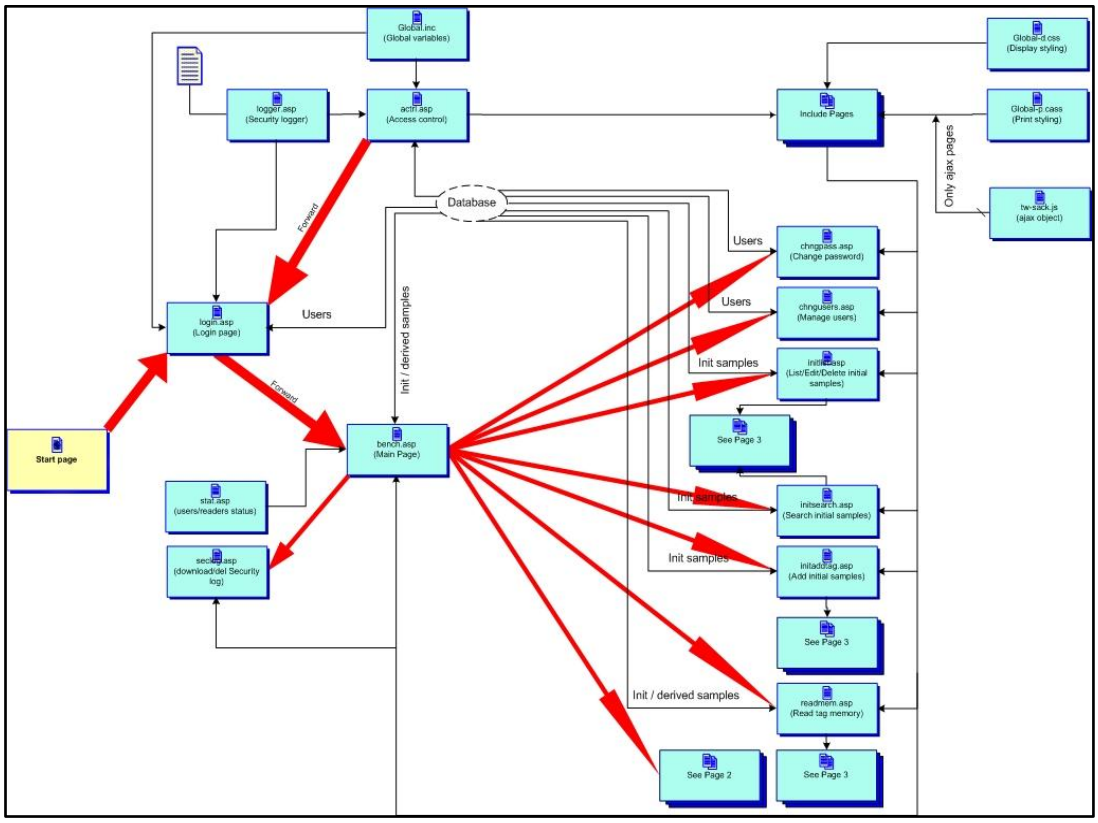


Figure 7-12: Architecture Diagram when Working with Initial Samples

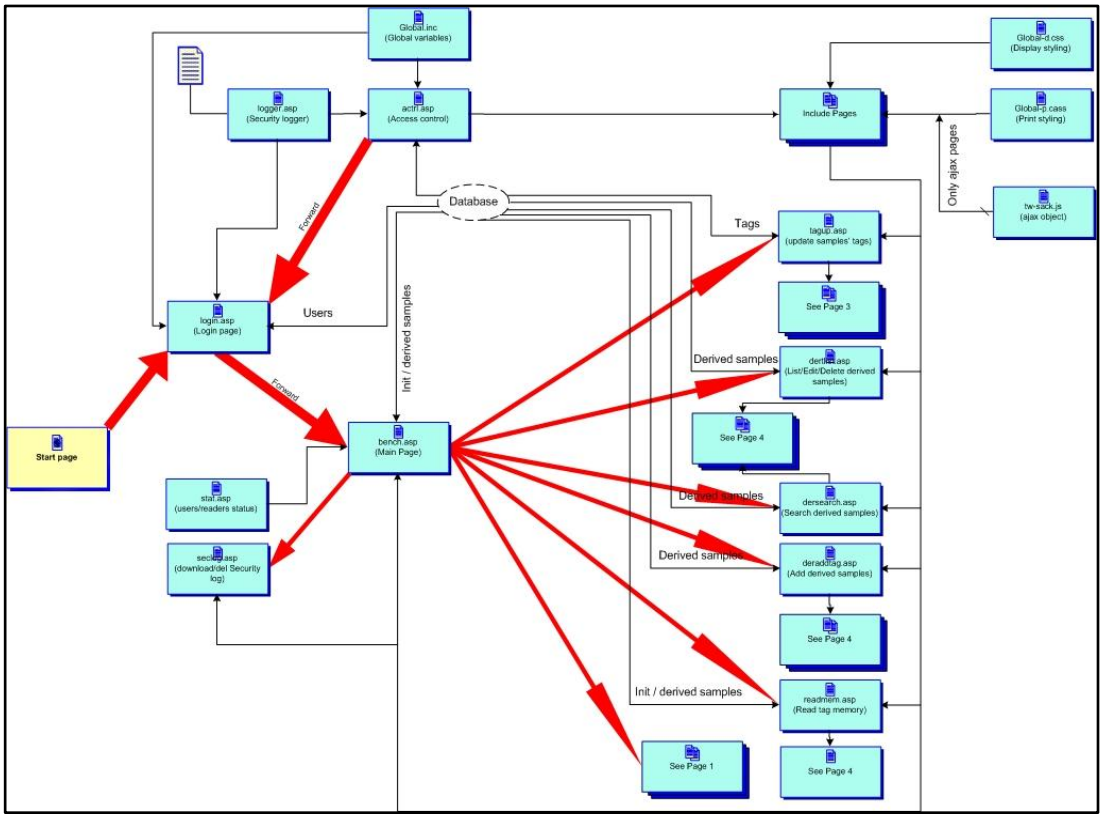


Figure 7-13: Architecture Diagram when Working with Derived Samples

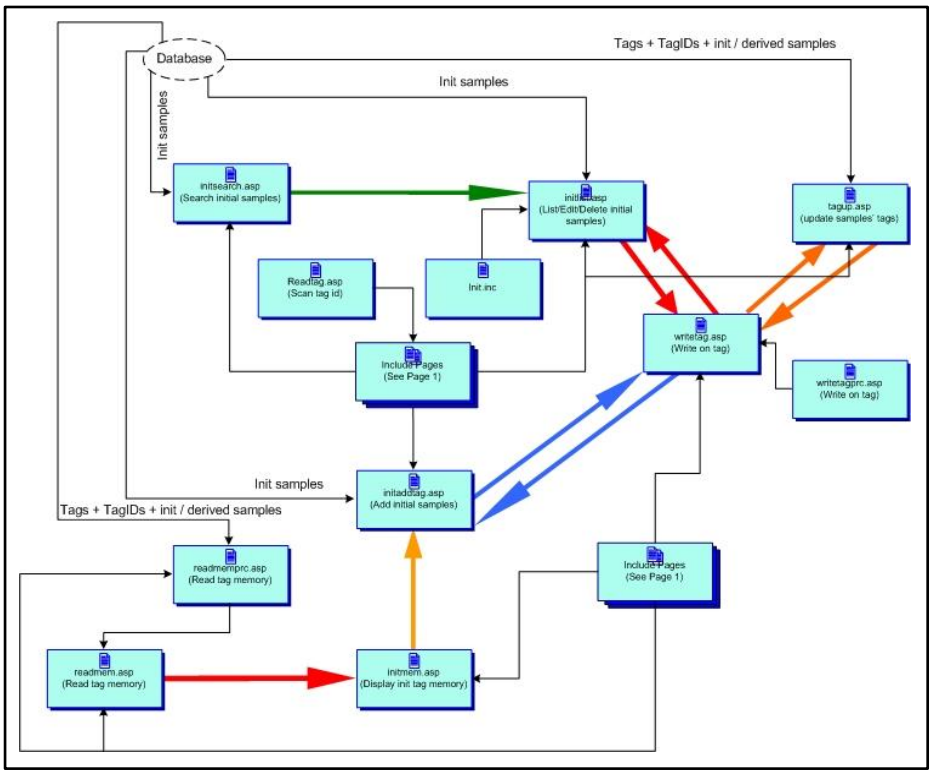


Figure 7-14: Interaction between Components for Initial Sample

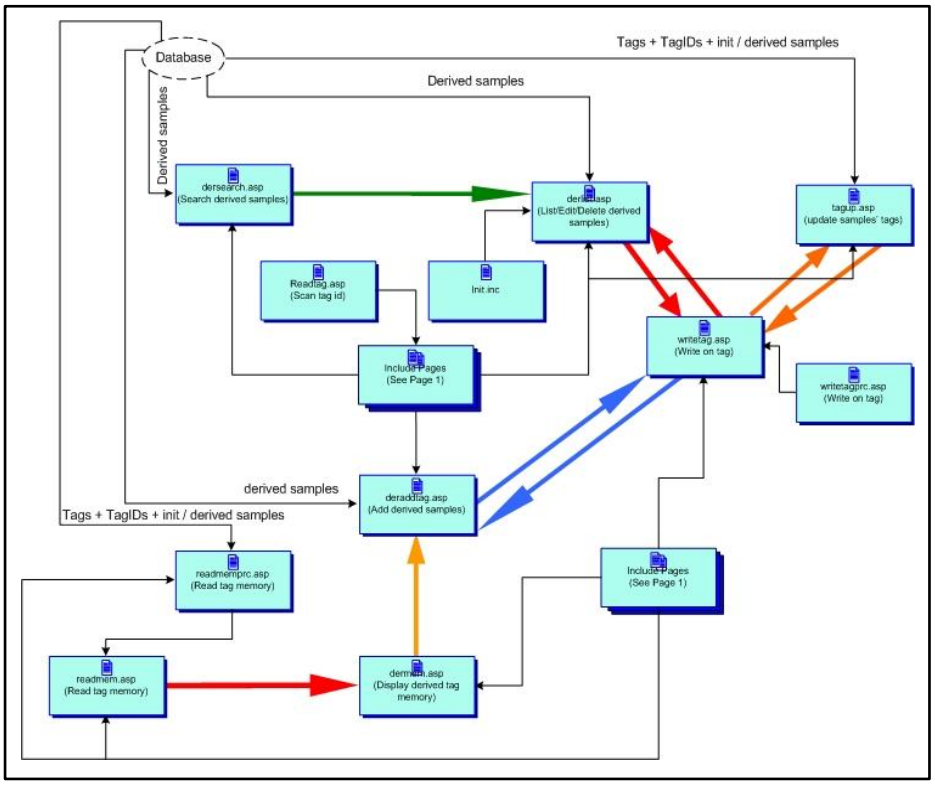


Figure 7-15: Interaction between Components for Derived Dample

Application program

The SITS Win32 Controller must run in order to control the RFID device. The user interface to this program is shown in Figure 7-16. Before connecting to the reader, the user must enter a name and an identifier for the reader connected to that machine. The host address must be entered and must match that of the web application. By clicking on the "Test" key the availability of the server is tested, shown in front of "Status" and an appropriate message is displayed. The user must then enter the port number to which the RFID device is connected. The "Test" button checks the availability of the port and displays a message. Details of the RFID device are shown in the message if the correct port is entered.

The "Save" button stores the specification of the connection in the windows registry for future use, if the user decides to store it. By clicking on "Save", previous specifications are overwritten. The "Connect" or "Disconnect" buttons establish or abort the connection to the RFID device. In Figure 7-16 the connection is already established. Finally, "Cancel" and "Exit" buttons cancel and exit the application respectively.

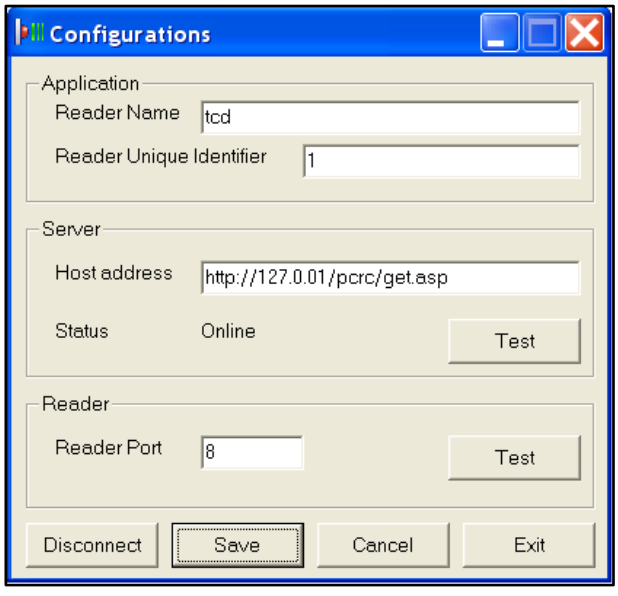


Figure 7-16: SITS Win32 Controller User Interface

In addition to the functions provided in the main user interface, more functionalities are available through the windows tray interface shown in Figure 7-17. When the

connection is established a blue antenna image is displayed, as in Figure 7-17, and when the connection is broken a red cross sign will appear on the blue antenna. The user interface can be hidden by clicking on “Hide” and if it is already hidden, it can be displayed. This item will change to “Display” if the interface is already hidden. “Disconnect” or “Connect” do exactly the same action that “Disconnect” or “Connect” buttons in the interface do. “Raw Data” is an important function when troubleshooting the application. This button displays all of the commands and data sent to and received from the RFID device. “Reset” clears the “Raw Data” buffer. “About” and “Exit” keys, display a window with information about the application and exit the application, respectively.

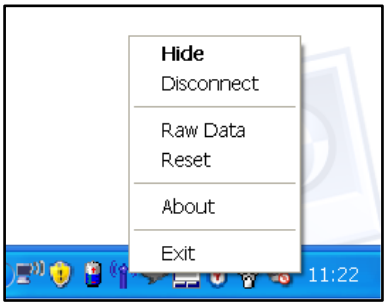


Figure 7-17: SITS Win32 Controller User Interface in Windows Tray

7.3.3 Software interfaces

Web application and SITS database

The database of the PCRC-SITS prototype consists of tables for sample data, user accounts and a table for expanding or combining data read from tags. Data submitted through web application forms are stored in these tables and tag data are also written accordingly.

Tables are created according to the designed database shown in Figure 7-4. Initial sample data, excluding the details of the samples that are aliquotted from them, are stored in a table called “ISampleA”. The aliquotted samples and their data that are excluded from the “ISampleA” table are stored in “ISampleB”. Further details of these aliquots are stored in a table called “DSample”. This way of data storage allows restrictions to be put in place for accessing initial and derived samples individually.

Sample data that are further derived from the aliquots are stored in the “DSample” table. So when the user submits the form for adding a new initial sample, the collection data will be stored in “ISampleA”, and the data of the derived samples will be stored in “ISampleB”. And when the derived samples are added, their data will be stored in “DSample” table. Editing a record follows the same procedure, except that when an initial sample record is updated, this data should also be updated on the derived sample tag data. So a list of the tags to be updated should be maintained. Finally, in case of deleting a record, upon confirming deletion, the record is removed from the relevant tables and the tag memory is cleared once it is in the reading range. A list of tagIDs of samples that are awaiting tag clearing is maintained.

A table called “TagIDs” is created to store all of the tagIDs existing in the database. This table includes their status as to whether tags are not updated according to the database record or they are being deleted. This table is also used to store the ID of the person who has requested deletion or has written the tag. This will be particularly useful when deletion by non-administrator accounts is not permitted.

Details of the users being added to the system through the web application are stored in a table named “Users”. Username, password, role and location of each user are stored in this table. It is used for authentication when signing in to the web interface.

The “Versions” table is used to split the data that is read from the tag memory. The position at which each field starts, its length and details of the field is given in this table. Two versions are currently defined for this prototype as tags are either on initial samples or on aliquots.

Application program and the web application

As depicted in Figure 7-18 when the connection between the reader device and the local machine in the laboratory is established, the following operations are carried out in order:

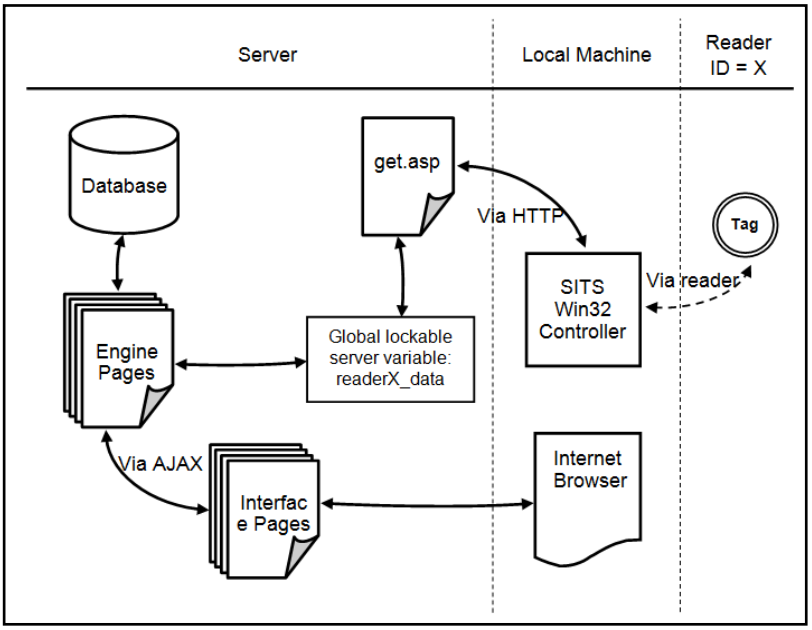


Figure 7-18: Data Exchange in the PCRC-SITS Prototype

1. The windows application that is running on the local machine, SITS Win32 Controller, tries to launch communication with the get.asp page. This happens through HTTP GET request.
2. SITS Win32 Controller keeps checking whether get.asp has any commands for the reader that it is connected to. User command and application response are stored in a variable named readerX_data, X being the ID of the reader that the command and response corresponds to. Reader IDs are defined to enable differentiation between different readers that might be online concurrently on different local machines. So, for each reader that is online a variable exists to act as a buffer for its related data transactions.
3. Users, depending on their defined privileges, send commands through a web interface. Commands are sent in four characters. Five commands are defined for this prototype: read tagID, read tag protocol, read tag header, read entire tag memory and write tag memory, and their matching commands are RTAG, RTGP, RDTH, RDTF and WRTM, respectively. RTAG is only used for scanning tags in range, so it does not take any parameter. In case of reading (except for scanning tags, i.e. except for RTAG), the command will be followed by the tagID of the tag to be read, and in case of writing, the command will be followed by the tagID of

the tag to be written followed by the data to be written on the tag. These data are stored in the readerX_data variable.

4. If there is a command for a particular reader that is connected to a local machine, the reader will carry out the requested command and will send its response back to the get.asp page via the readerX_data variable. Responses received from the reader are three characters, thus differentiating a response from a command. If the number of tags in the reading range is more than one, then the response will be "ANS:NO:More than one tag in range!" and if no tag is in range, then the response will be "ANS:NO:No tag in range!" and if there is only one tag in the reading range, depending on the command one of the following will happen (Table 7-1shows syntaxes of responses):
   - In case of RTAG, the response will be "ANS:OK:" followed by the tagID.
   - In case of RTGP, if the tag in range is written under the Accessibility Format (AF) protocol, then the response will be "ANS:NO:Tag is written before, Try to import it." and if not, then the response will be "ANS:OK:" followed by the tagID. AF protocol is specifically defined for the PCRC biobank. This command can be considered a safety check to ensure that the tag is not already used in the PCRC biobank, to prevent data loss.
   - In case of RDTH, if the header of the tag was read successfully, then the response will be "HDR:OK:" followed by the tagID, followed by ":" followed by the header data. If not, then the response will be "HDR:NO:Read Error!".
   - In case of RDTF, if the memory of the tag was read successfully, then the response will be "RDT:OK:" followed by the data of the memory, followed by ":" followed by the header data. If not, then the response will be "RDT:NO:Read Failure!".
   - In case of WRTM, if writing data on the memory of the tag is successful, then the response will be "WRT:OK:" followed by the data written, followed by ":Write Successfully!". If writing is not successful due to interruption while writing, then the response will be "WRT:NO:Write Failure!", and if the tagID of the tag in the reading range does not match that of the intended tag, then the response will be "WRT:NO:Wrong tag in range!".

Response	Syntax
Successful read tag ID, RTAG	ANS:<Status>:<ScannedTagId>
Successful read tag protocol, RTGP	ANS:<Status>:<ScannedTagId>
Successful read tag header, RDTH	HDR:<STATUS>:<TAGID>:<DATA>
Successful read full tag memory, RDTF	RDT:<STATUS>:<TAGID>:<Data>
Successful write tag memory, WRTM	WRT:<STATUS>:<TAGID>:<Message>
Unsuccessful operations	<Relevant response code>:<Error Status>:<Error Details/message>

Table 7-1: Responses Syntax

5. "Engine Pages" then read the readerX_data variable. These pages are pages that manipulate data and communicate with the database. This is the PCRC-SITS database that will be described in the next Section.
6. Data received from the database or the responses received from the windows application are then made available to the user, through the "Interface Pages" via internet browser. "Interface Pages" use AJAX to communicate and access to data from the "Engine Pages".

7.3.4 Hardware interface

The SITS Win32 Controller has been developed in Visual Basic (VB). The means used for the settings of the device are based on the documentation (Medio P101 Command set_V20.pdf) accompanying it from the manufacturer.

There are six files in total for the purpose of making communication with the device feasible, four forms and two main modules. The six forms are GUI.frm, GUI_About.frm, GUI_Data.frm and GUI_Settings.frm. The two modules are M_Kernel.bas and M_Tray.bas. Table 7-2 shows the forms and a brief description of their functionality.

Form	Functionality
GUI.frm	Handling and linking the three other forms and two modules
GUI_Settings.frm	The main interface for the user
GUI_About.frm	Working the about message
GUID_Data.frm	Displaying the raw data

Table 7-2: Forms and Their Functionality

GUI.frm handles and manages communication between the forms and modules. It also includes the common features accessible by other forms and modules. A "Tray" including Show, Connect, Raw Data, Reset, About and Exit is defined here. The Show command launches the Win32 application, Connect calls the `CMD_Connect_Click()` from GUI_Settings.frm to make connection to the reader, Raw Data calls the `rDataW` function from M_Kernel.bas module and Reset calls the `rBuffer`. "About" displays a message from the GUI_About.frm and Exit ends the application. Figure 7-19 shows how the Tray is displayed in the notification area of system tray in Windows XP.

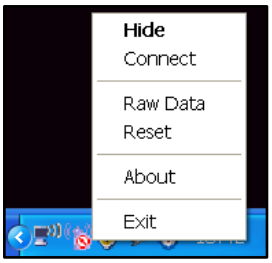


Figure 7-19: Application Tray in Windows XP Notification Area

The Reader Name and Reader Unique Identifier are set by the user in GUI_Settings.frm form. These appear on the web. The text entered to the "Host address" field, `TxT_Host`, is taken as the server where the data should be sent to and received from. Once the user presses the "Test" button a sub called `CMD_HCheck_Click()` is run to display the availability status of the server. Communication is only possible when the server is ready. The next field in the interface takes the port number which the device is connected to. Once the "Test" button is pressed the `CMD_RCheck_Click()` sub acts.

In case any problem occurs with the connection, an error message describing the issue is displayed. Not being able to communicate with the device or not being able to connect to the port (port being busy with a different device) are known and expected errors.

Once all of the fields are set and tested successfully, the user can save the current setting by pressing the "Save" button. The setting is stored in the machine registry using `getInfo` from M_Kernel.bas. This will be further explained below. By pressing the "Connect" button the `CMD_Connect_Click()` sub is run to establish connection with the device.

Connections to the server and the port are checked and "pTray", an image in GUI.frm, is made visible. pTray is the icon that is replaced with pTrayOFF and pTrayON when communication is disconnected and connected respectively. These icons are shown in the notification area of windows (disconnected case shown in Figure 7-19). The outline of this menu is implemented in the GUI.frm and the code for each key action is implemented in various locations.

`rMicro()` and `RF_ON_Reset()` are used to reset the reader device and the RF. These methods are implemented in M_Kernel.bas module and will be discussed later. `TMR_Cmd_Timer()` and `TMR_Stat_Timer()` are both defined in GUI.frm and from there they call `CMDPrc()` and `rPower()` functions from M_Kernel.bas module. `TMR_Cmd_Timer()` allows a delay for the command to be processed and `rPower()` checks the availability of the reader. A more detailed description will be given where the M_Kernel.bas module is explained.

Cancel and *Exit* buttons cancel and exit the application respectively.

There are two modules used in this application, M_Tray.bas and M_Kernel.bas. M_Tray.bas adds, updates and removes system tray icons as well as displaying notifications. This module is independent of this particular application and can be added to any VB program to handle notification area icons and tasks. This module uses shell32.dll and user32.dll from windows API. The `TrayAdd` function of this

module is called in GUI_Settings.frm and M_Kernel.bas, and `TrayRemove` and `TrayBalloon` are called in M_Kernel.bas.

M_Kernel.bas is where communication takes place.

`AHDR`, `ADR`, `CMD` and `ANS` are defined as `Enum` and they represent various attributes of the communication extracted from the Medio P101 Firmware Command Set document (Medio P101 Command set_V20.pdf).

`AHDR` is one byte and can take four different values for four different Protocol Transmission Modes. `0x02`, `0x06`, `0x0E` and `0x01E` for standard, not secured, not addressed and fast modes, respectively, are the values that `AHDR` can take.

`ADR` is also one byte and is the reader address. Its value is `0x00`.

`CMD` is one byte and is the command sent to the applicant by the reader. Various commands and their values are tabulated in Table 7-3.

Commands	**Value**	**Program Variable**	**Comments**
Detect Chip Type	`0x00`	`Detect_Chip`	Detect the type of any chip present in the RF field
Reset	`0x06`	`Reset_Micro`	Reset the microcontroller
Get/Send IO	`0x07`	`Get_Set_IO`	Returns the current status of the input and the output/Sets the status of the only output
RF On / Reset RF	`0x0C`	`RF_ON`	Turns on the RF amplifier or reset it
RF Off	`0x0D`	`RF_OFF`	Turns off the RF amplifier
Set Standalone Mode	`0x0E`	`Get_Set_Standalone`	Initialise standalone operating mode
ISO 15693 Inventory	`0x20`	`ISO_15693_Inventory`	Launches an inventory of the ISO15693 chips in the field, and manages the anti-collision sequence
ISO 15693 Raw Request	`0x21`	`ISO_15693_Raw`	Permits sending all commands respecting the ISO 15693 Standard
Identify Reader	`0x60`	`Get_Version`	Returns the firmware version and revision numbers and the product firmware reference
Monitor Power Supply	`0x76`	`Power_Level`	Provide the power supply level

Table 7-3: Commands, their Values, Program Variable and Comments

`ANS` is one byte and is the answer code received from the reader. Various commands and their values are tabulated in Table 7-4.

Commands	Value	Program Variable	Comments
Detect Chip Type	0x00	Info_Detect_Chip	Detects the type of any chip present in the RF field
Reset	0x06	Info_Reset_Micro	Resets the microcontroller
Get/Send IO	0x07	Info_Get_Set_IO	Returns the current status of the input and the output/Sets the status of the only output
RF On / Reset RF	0x0C	Info_RF_ON	Turns on or reset the RF amplifier
RF Off	0x0D	Info_RF_OFF	Turns off the RF amplifier
Set Standalone Mode	0x0E	Info_Get_Set_Standalone	Initialises standalone operating mode
ISO 15693 Inventory	0x20	Info_ISO_15693_Inventory	Launches an inventory of the ISO15693 chips in the field, and manage the anti-collision sequence
ISO 15693 Raw Request	0x21	Info_ISO_15693_Raw	Permits sending all commands respecting the ISO 15693 Standard
Identify Reader	0x60	Info_Get_Version	Returns the firmware version and revision numbers and the product firmware reference
Monitor Power Supply	0x76	Info_Power_Level	Provide the power supply level

Table 7-4: Answer Codes, their Values, Program Variable and Comments

Two packets are defined based on CMD and ANS Enums. These types are PktSnd and PktRec for a packet to be sent to the reader and for a packet to be sent to the application, respectively.

objport is a variable used to create a port object and is created by MSCommLib.MSComm. The baud rate for this device is 115200 Bps. objXML is also a variable that is used to communicate with the web interface MSXML.XMLHTTPRequest.

Other functions are used for various tasks as below:

ErrHndlr: This is in charge of handling possible errors.

Die: This is used whenever the user decides to terminate the application.

getInfo: This is used to either get the settings from the machine registry or set the current settings on the machine registry. This sub also handles errors that might occur.

`XMLRead`: This is used to communicate with the web. The `CRC_Calc` function is called in `XMLRead` function to calculate the checksum.

`OpenPort`: This opens a com port and uses `pkt2Str` function to communicate with the device. `pkt2Str` function converts a packet to a string. hex2Str is also a function to convert hex to string.

`BufferPrc`: This processes data received from the reader and returns a boolean variable depending on the success of procedure.

`CMDPrc`: This processes commands received from the web.

`Tag_Counter`: This is used to count the number of tags in the reading range. This function is used at the time of sending data to the reader for writing on the tag.

`ReadProtocol`: This is used to read the data from the tag memory.

`WriteData`: This is used to write the data on the tag memory.

`RF_ON_Reset`: This is used to turn on the RF or reset it.

`DataPrc`: This processes data using array. `DataPrc2` function processes data without using array.

`rPower`: This checks the power level of the reader and hence validates its availability.

`rMicro`: This resets the reader.

`rBuffer`: This clears the buffer by sending `RST` command to the `XMLRead` function.

`rDataW`: This inverts visibility of the raw data, that is, if the raw data is visible it makes in invisible and vice-versa. This is the function called in GUI.frm to display or hide raw data.

`aLog`: This logs the data along with date and time. This function is called in the `XMLRead` function.

7.3.5 Integration with the PCRC BIMS

The PCRC BIMS is only accessible through Microsoft Internet Explorer (IE). Using Automation the creator and owner of an instance of IE will be PCRC-SITS, or more specifically the Win32 Application written in VB. This application fills the Distiller web forms using data from the PCRC-SITS database. However the data fields on Distiller forms should be defined in advance so that the application can link the fields with data. Extracting every field of the sample identification and tracking tables is a tedious job.

Integration based on Automation is done as follows:

1. Tables, called groups in Distiller, are created according to the elements from the PCRC-SITS interface. One of the tables is shown as an example in Figure 7-20.

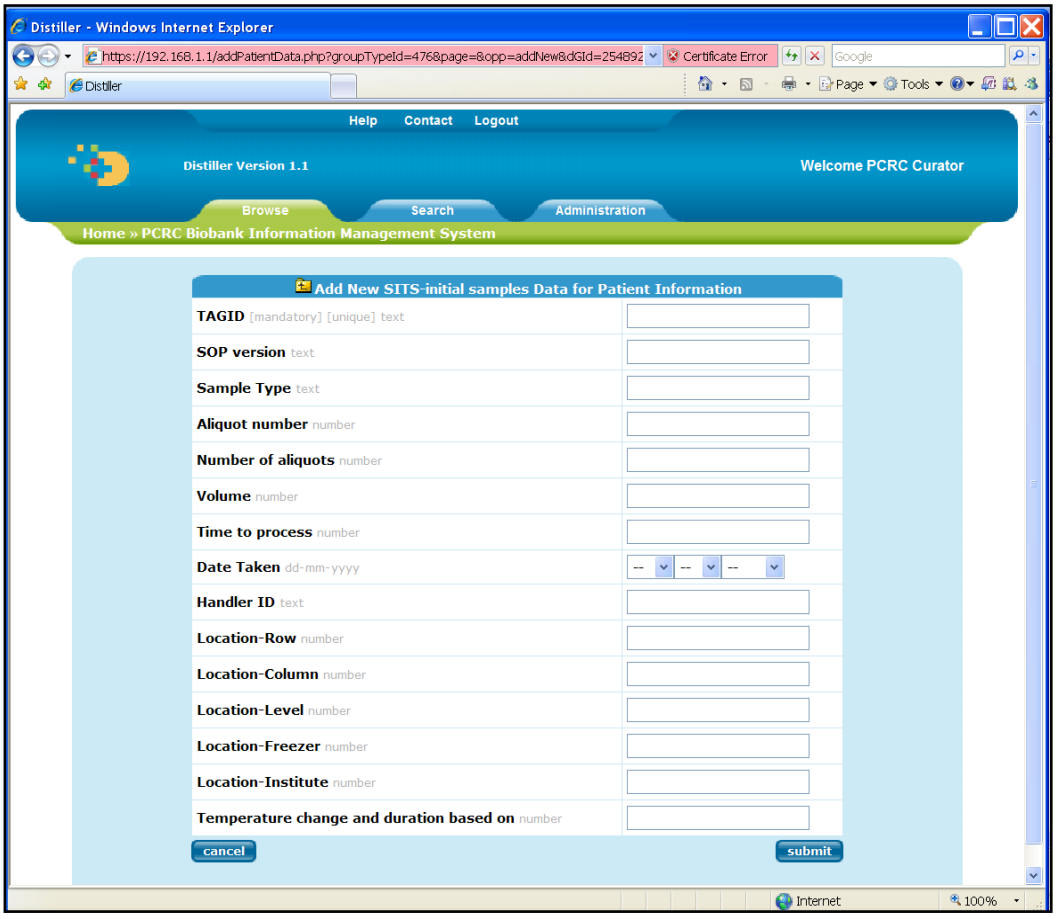


Figure 7-20: Data Groups in Distiller

2. Attributes of the fields are extracted from the HTML source of each page containing the created groups. This was done by viewing the source file of Distiller interface through the web browser.
3. SITS Win32 Controller code is modified to support creation of an IE object and also to handle commands and responses related to it in addition to previously defined commands and responses related to communication of the PCRC-SITS and RFID reader. Public oIE As cInternetExplorer and Set oIE = New cInternetExplorer are used to create the IE object. Then FillForm function handles the communication.
4. Based on the data fields shared between PCRC-SITS and Distiller in each page, SQL statements are created dynamically. This is done in xtobimsprc.asp and xtobims.asp pages when the user decided to export a record.
5. The results of SQL statements are then sent to Distiller via the Win32 Application and the form is filled.
6. When the user decides to export data from PCRC-SITS to Distiller, data is added to Distiller. TagID fields in each group are defined as the unique identifier of a record in that group of data. Therefore in the next attempt to export the same data, TagID is identified and the record is updated. It is very important to have TagID field set to be displayed in the first page (Figure 7-21) Since the only hyperlink that leads to the record is the "+" beside each row and only an identifier (that is not known to user) is used to retrieve the record, the HTML code of the page has to be searched for a pattern matching the TagID and record details.

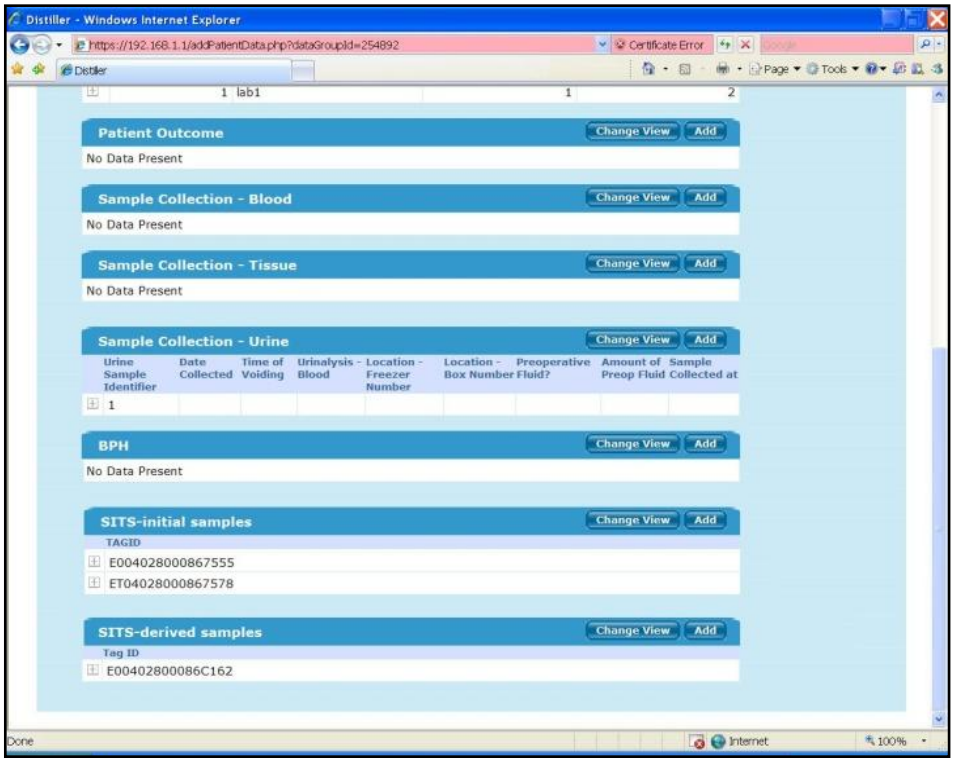


Figure 7-21: TAGID Fields Appearing on the First Page

7.3.6 System Functions and Attributes

Identification of samples (R1)

A RFID tag is manufactured with a unique identifier. This identifier is used as the sample identifier. Pages for adding, deleting, editing and updating tags are provided. The first ten bytes of the tag memory available for writing are allocated to the tag header that contains information about the tag and its data. Figure 7-22 depicts memory allocation of the tag header. The AF protocol, version number, sync time and date and user index are stored in the tag header.

The first byte is specified to the AF protocol that has been defined for identifying the PCRC biobank specimens. This byte is checked either by the SITS Win32 Controller or the SITS web interface, depending on the operation to be carried out. If its value is validated to be AF, further data of the tag will be read. Otherwise, if it is 0, the first 8 bytes will be checked. If these are all 0 then the tag is assumed to be new. If not, the tag will be considered unknown and no further reading of the data is done. This helps to improve reading speed and minimises data communication across the system.

The second byte of the header is the version number or version type of the tag data. It is denoted by "V" in Figure 7-22. This version number is used as a key to expand data to the fields. It can be regarded as an identifier that determines which tables of the database the tag is related to, i.e. whether the specimen is initial or derived. The value of this field retrieves a set of records in a table in the database which stores the length of each field and where they start, the "Version" table. This feature helps to safeguard the data stored on the tag. Even if the data are read by an unauthorised individual, they are meaningless without access to the table that contains the mapping key of the version number.

The next block is a two-byte block which stores the time when last the tag data and database record were synchronised for this tagID. The date is stored in the next four-byte block. Storing the full history of accesses to the tag requires massive amount of memory on the tag. Thus having an audit system in place eliminates the need for maintaining a history on the tag.

The last two-byte block contains the index number of the user who last updated the record, not necessarily written on the tag. It is important to note that this is only an index number of a record in the "Users" table and is only meaningful when reading from the database directly.

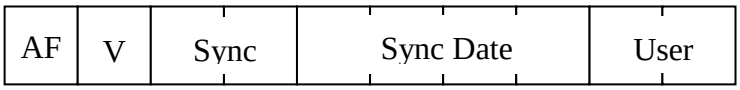


Figure 7-22: Tag Header

The remaining space on the 1280 bits of the tag is used as shown in Appendix K. The structure varies for initial and derived samples but for either case the tag has enough memory for storing identification and tracking data as required by R16.

Maintaining the records associated with samples up to date (R2)

RFID tags used for this prototype are rewritable so they can be updated many times. When updating records, the associated tag might not be in the reading range. The record is then updated on the relevant database table and the tagID of that sample is flagged in the "TagIDs" table. The "Update Sample Tag" page lists all of the tags that

are awaiting update. It is important to note that updating the tags that are listed in the waiting list does not reveal any database record or tag data to the user who is updating the tags. This list consists of tagIDs which do not expose any data. It is available to all users to facilitate faster update of data on the tags while they are in the reading range of the reader.

It is important to note that updating records will over-write the records on the database, hence history of data records are not maintained on the database. However, any update of the records is stored in the audit trail file from which a full history of the records can be re-created.

Tracking samples (R3)

Figure 7-23 displays the flow of samples, and describes pages, roles and locations for the corresponding action. After the initial processing phase, samples will be transferred to research institutes. Samples are individually tagged up to the point which a large number of them have identical characteristics, at which instant batch tagging will be used. The prototype records rows and columns from which the sample has been removed.

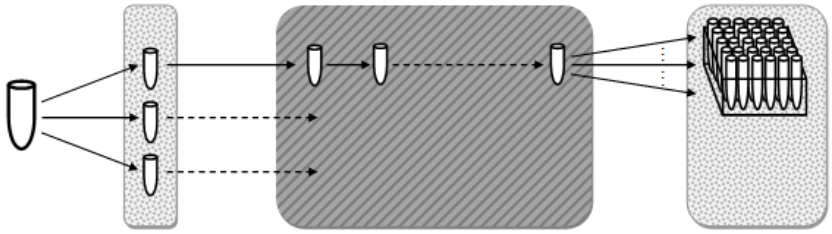
Phase	Initial Processing	Secondary Processing	Storage and Retrieval
			
Location	Collection Hospital	Research Institute	Research Institute
Actor	Research Nurse	Researcher	Researcher
Action	Collection – Aliquotting - Storage	Requesting – Omic Processing - Storage	Storage
Pages Involved	Add New Initial Sample (Tag ID Scanned) Update Initial Sample (With Derived Samples Tag ID)	Add New Derived Sample (Tag ID Scanned, Previous Tag ID and Collection Data Looked Up) Update Derived Sample (With Next Tag ID)	Add New Derived Sample (Tag ID Scanned, Previous Tag ID and Collection Data Looked Up) Update Derived Sample (Row and Columns of Removed Samples)

Figure 7-23: Pages Involved for Operations in Each Location and by each Role

Collection data are appended to the aliquots derived from each initial sample. These data include temperature changes and duration as well as details of the SOP under which collection has been made. In addition to collection data, the tagID of the initial sample should also be included in the header of the aliquots' tags. Once a sample is divided to aliquots, the tagID of those aliquots will be written on the initial sample tag and recorded on the database. This allows a two-way traversing of samples and aliquots. In case a change is made in the initial sample collection data the change is automatically propagated throughout the aliquots derived from that sample and the aliquots are listed in the update list. Tags attached to each aliquot have a field for storing the tagID of the next aliquots and once moved to the new tube, the tagID of the previous tube will be recorded on the new one. The situation corresponds to a double linked list. The first node is similar to a tree with three branches as shown in Figure 7-24. From Figure 7-24, the initial specimen includes ID, header, detail of the particular sample, temperature monitoring data, aliquot 1, aliquot 2 and aliquot 3 IDs. These three aliquots are branches of the tree. Details of the particular sample and temperature monitoring data are labelled as A and are passed to the aliquots. The aliquots in addition to the ID, header and collection data need to have aliquot data, IDs of the next and previous samples and the removed items in case of well-plates stored as well. The data shown in Figure 7-24 are provided in detail in Appendix K. Initial specimen and aliquot data from Figure 7-24 are detailed in Figure App. K-1 and Figure App. K-2 in Appendix K, respectively. Since there are three aliquots derived from each initial specimen, the data shown in Figure App. K-1 for "Derived Sample Data" are multiplied by three. In a double linked list, each item in the list has a link to the previous and next item allowing forward and backward transverse in the list (National Institute of Standards and Technology, 2008).

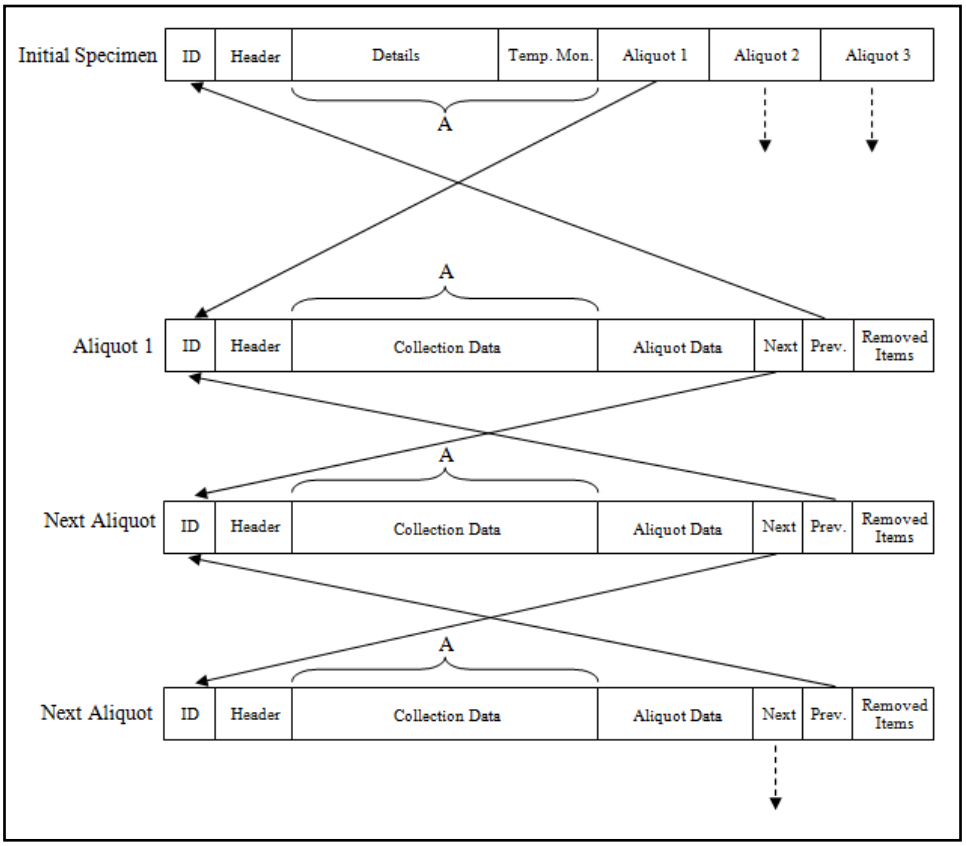


Figure 7-24: Tag Headers Being Linked

Longitudinal tracking of samples (R4)

Users can search for the PCRC NO they are interested in the "Search Sample" page. The results will appear on the top part of the page and by clicking on "View Details", the entire record for that sample can be observed. The date of collection for each sample is also stored on the database and the tag.

Attribute-based searches on samples and aliquots (R5)

A comprehensive searching tool is provided to users based on their rights and privileges. It allows users to search for attributes of interest and supports searching on more than one piece of data. Users can search on both initial samples and derived samples. Then they can choose to view details of a record in the returned results and decide to update or delete the record.

Audit trail of all accesses (R6)

An audit trail of major actions, along with users who carried out the action with their IP address and date is logged in a CSV file that is accessible by the root and administrator account holders and can only be reset by root account. The following actions are logged:

- The username and role for successful log in
- Incorrect username or password for unsuccessful attempts to log in
- Attempts to open pages beyond the access rights granted. The URL of the page is also logged. This case may occur for example when the user attempts to open a page from the browser history without logging on to the system.
- Adding records to the database, including the tag ID of the record and whether it is an initial or derived specimen
- Downloading and resetting the log file
- The missing tags and their IDs when trying to update the tags that are listed but are not in the reading range when attempting to read them
- If there is a mismatch between the update list and database the tag ID is logged. This is a rare situation that happens during prototype testing, when data are uploaded to the database and tag directly, without the use of the interface
- Requests placed for deleting a record in the database. This will store the tag ID of the record as well as whether it is initial or derived
- Changing a record from marked-to-delete to undelete status. This is only possible by root and administrator accounts. The tagID is also stored in the log
- Deleting records from the database or updating records on the database, including the tagID of the record and whether it is initial or derived specimen
- Modifications made to user accounts including their user identifiers (user IDs) only.
- Downgrading, upgrading and adding to database according to the tag's records. This will also record the tagID and whether it is initial or derived. Downgrading the database happens when the user chooses to overwrite the data on the database by the older data on the tag. This is an exceptional occasion that happens when inserting data directly to the database for testing purposes, so it has been taken into

consideration. Upgrading database is when the data on the database are written on the tag memory.

As discussed earlier, due to the massive amount of memory that the logs will take from the tag memory, only the last access will be stored on the tag header and details of all accesses to memory will be stored on the log file described above.

The web application displays a list of users who are online with their roles and the status of RFID devices connected to the system. The web application is accessible by multiple users concurrently as required by R13. An individual SITS Win32 Controller application must run for each local machine that the RFID device is connected to. In Figure 7-25 two users are online with no RFID devices connected.

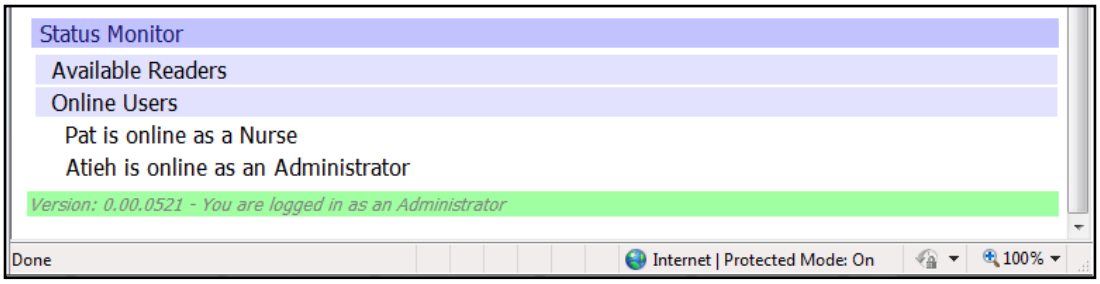


Figure 7-25: Online Users

Administrative functionalities (R7)

The first page that any user can access on SITS is the log in page where they are asked to enter their username and password. Then based on their roles they access pages which they are permitted to access. Access to pages that they do not have permission to view is prohibited both by direct links and by typing the address.

Excluding the root user, the five roles that are defined for this system are administrators, research nurses, researchers, guests and disabled accounts. Careful access rights are defined for each role and no access is given to disabled accounts; this role is needed when an account is suspended. Although predefining roles limits the functionality of the system for additional or customisable roles, for prototyping purposes these five roles are predefined based on requirements and discussions with users. In a fully operational system, it is likely that a more finely grained security system would be required. For example, it is probable that access rules could be

specified based on the institute in which the researcher is based and/or the particular study involved. It is not unusual for small groups of researchers to collaborate together on a particular piece of research and to share interim results amongst themselves. Only when the experiments have been completed and the results published, are the details made available to others. A comprehensive security system would have to be able to support such requirements. The roles rights are as in the table below:

Role	**Right**
Root	Full access to all user accounts, full access to collection specimen and also derived specimen tables and the ability to download or reset system log
Administrator	Full access to all users' accounts except for the root user, full access to collection specimen and also derived specimen tables and the ability to download system log
Research nurse	Change own password, full access to initial specimens table and also able to update the tags listed in the waiting list
Researcher	Change own password, full access to derived specimens table, browse and query the initial specimen table, read tags of the initial specimens and ability to update the tags listed in the waiting list
Guest	Browse and query both the initial specimen and derived specimens tables, read tags of both the initial specimens and derived specimens and ability to update the tags listed in the waiting list

Table 7-5: Roles and Rights

Figure 7-26 depicts accounts access privileges to each page and details of operations in each page are given in Table 7-6 based on this figure.

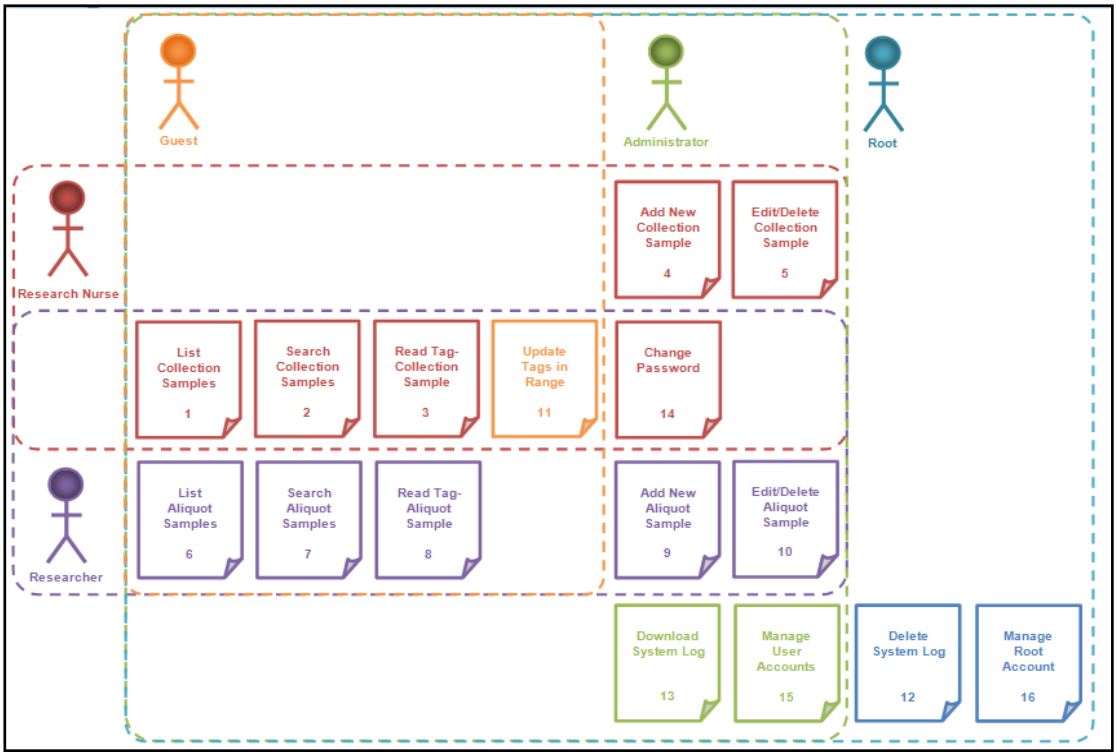


Figure 7-26: Pages Accessible by Roles

	Page		Description
1	Collection Sample	List	Lists the collection samples that are available in the database
			Scanning the aliquots tags taken from this collection tube
2		Search	Searching the database
			Scanning the tags
3		Read Tag	Reading the tags in range.
			Displaying the related record from the database
4		Add New	A form in which the tag ID should be scanned and the rest of data should be typed in
			This record will be written on the tag, if it is still in range
5		Edit/Delete	After viewing the details of the sample, user may choose to update, delete/request to delete, or set flag on a record
			In case of updating, if the tag is in range, it will update the tag data, otherwise it only updates the database and tag will be stored in the update list
6	Aliquot Sample	List	Lists the aliquot samples that are available in the database
			Tracing samples taken from this tube, or this sample is taken from
7		Search	Searching the database
			Scanning the tags
8		Read Tag	Reading the tags in range.
			Displaying the related record from the database
9		Add New	A form in which the tag ID should be scanned and the rest of data should be typed in
			This record will be written on the tag, if it is still in range
10		Edit/Delete	After viewing the details of the sample, user may choose to update, delete/request to delete, or set flag on a record
			In case of updating, if the tag is in range, it will update the tag data, otherwise it only updates the database and tag will be stored in the update list
11	Update Tags in Range		Lists tags whose database record is updated, but the tag is not
			Updates tags that are in the list and are in range
12	Delete System Log		Resetting the log
13	Download System Log		Downloads the log in CSV file
14	Change Password		Changing own password
15	Manage User Accounts		Add, modify and remove accounts except for root account
16	Manage Root Account		Add, modify and remove accounts root account

Table 7-6: Description of Pages

There is a section for user management in which the user can change his/her own password, once they know the password for the account he/she has logged in. Depending on their role they can manage other user accounts by modifying them, deleting them and adding new accounts.

7.4 Summary

The PCRC-SITS prototype has been designed and implemented such that requirements explained in Section 2.7 and Chapter 6 are satisfied.

The prototype consists of a user interface to the PCRC-SITS database and an interface that allows communication between the user interface and the RFID devices. The PCRC-SITS prototype has been developed for selected workflows that are more frequently carried out in the PCRC biobank, and for blood and urine samples. Blood and urine samples are chosen because they are liquid and this will allow testing the operation of RFID technology near liquid. Blood and urine samples are also more available in the PCRC biobank than other sample types such as tissue samples. SITS data, that is *Identification Data*, *Location Data*, *Tracking Data*, and *Temperature Monitoring Data*, are stored on the SITS database and the RFID tag memory attached to each sample.

The PCRC-SITS prototype is based on RFID and is web-based in order for it to be accessible by multiple users concurrently. Roles and privileges are defined carefully and a full audit trail is maintained. The PCRC-SITS prototype allows for reading data from tags, updating tags and deleting tags. Notifications are put in place to minimise accidentally overwriting data or deleting a record. Deletion by users who are not administrators or the root user must be confirmed by an administrator or root account holder. This system is designed in a way that minimises accidental upload of personal data. Open text fields are prevented, and where the data is expected to be of a fixed length and type, only that length of that type of data can be inserted. Locations and specimen handlers are coded so that they are only meaningful to the staff involved in the biobank. Nonetheless, to meet the requirements, access control barriers and monitoring, and also tag data mapping are put in place.

For integrating the PCRC-SITS prototype with the PCRC BIMS, OLE Automation was carried out in six major steps: creating tables and groups on Distiller to house the SITS data, extract the attributes of the fields from Distiller interface using its HTML source files, modifying the SITS Win32 Controller to facilitate data exchange, creating SQL statements to identify the data to be uploaded to BIMS, filling forms on Distiller

based on the results of the SQL statements, and finally programming the system to browse to that particular record and update it.

Tracking samples has been made possible by deploying a double link list in which the tagID of each tube is stored as part of the data for the next tube and the previous tube. PCRC NO and TagID are used to uniquely identify participants and samples, respectively. Identification, location, tracking, SOP, temperature changes and duration data are stored on the RFID tags which have 1280 bit of memory. The tag header is also defined to store basic identification data such as if the tag belongs to the PCRC biobank, if the sample is an initial or derived sample, when the tag was last written and by whom. Depending on the type of sample, data stored on the memory are read. The key for mapping data is stored in the "Version" table on the PCRC-SITS database.

The PCRC-SITS prototype conforms to part of the 21 CFR Part 11. This prototype does not maintain a history of the modifications made to the records on the database and data are updated on over-writing basis, although it does log any modification made to the sample data on the audit trail file and a full history of records can be re-created based on the audit trail file. This functionality should be taken into consideration when developing the full implementation of SITS. The labels used in this prototype, RFID tags basically, include complete information about a sample or an aliquot. It also allows connection to the associated database using the unique identifiers, the manufacture tagIDs. This prototype also conforms to the ABN requiting survival of labels, RFID tags in this prototype.

Chapter 8. Evaluation

8.1 Introduction

The PCRC-SITS prototype was designed and implemented as described in Chapter 7. However its functionality should be evaluated in real-life practice. The context for evaluating this prototype is the PCRC biobank in which samples are collected from donors in a number of hospitals, and processed in research institutes.

In this Chapter, three elements for evaluating the PCRC-SITS prototype are defined and carried out in order to respond to three evaluation questions:

1) What is the experience of the users of the system of the PCRC-SITS in a real-life setting?;

2) How is the PCRC-SITS integrateable with BIMS?; and

3) What is the general perception of PCRC biobank members of SITS and electronic tagging of sample?

The three elements are: using the system in the real-life setting in PCRC biobank to carry out procedures, evaluating integrateablity of the PCRC-SITS with BIMS and approaches for doing so, and collecting a survey of the PCRC biobank members to find out their opinions about sample tracking and identification and their perceptions of electronic tagging of samples. The users opinion about their experience of working with the PCRC-SITS, integrating the PCTC-SITS with BIMS, and analyses of the results of the survey will be used as measurements of evaluation.

In the next Section, the evaluation methodology will be described to familiarise the reader with the details of the evaluation strategy. Section 8.3 will then describe the approach carried out to evaluate the usability, and functionality. In Section 8.4 the integration of PCRC-SITS with BIMS in general is evaluated and with the PCRC BIMS specifically. Section 8.5 will cover the results of a questionnaire collected from the PCRC biobank members to elicit their opinions on current methods of identifying and tracking samples as well as their perceptions of electronic tagging samples. The

PCRC-SITS will then be compared with other available SITS in Section 8.6. Finally Section 8.7 will provide the summary of the Chapter and will discuss the evaluation method.

8.2 Evaluation Methodology

The evaluation of the PCRC-SITS prototype will be carried out in the context of the Irish PCRC biobank. The objective of PCRC-SITS is to identify and track samples as they go through workflows. Workflows have been segmented into three major phases as described earlier in this document. These three phases are:

1. The initial processing phase that is carried out in the collection hospitals by a research nurse at each hospital,
2. The secondary processing phase in which samples are processed in research institutes by researchers, and
3. Long-term storage and retrieval of the samples and data over several years.

In addition to identifying and tracking samples, PCRC-SITS is required to be integrateable with the BIMS. The feasibility of such capability and approaches available for it must be evaluated. Therefore as proof-of-concept the PCRC-SITS prototype should be integrated with the PCRC BIMS.

User experience and opinions about the current method of identifying and locating samples in the PCRC biobank needs to be compared against their experience with the PCRC-SITS prototype. A questionnaire was distributed to elicit their opinions on the current approach and their perceptions of the electronic tagging of samples. This will assist in determining the user acceptability of RFID in the biobank environment.

Therefore, the three elements for evaluating PCRC-SITS that will be carried out are testing the workflow, evaluating the feasibility of integration with the BIMS, and administering a questionnaire to learn about users opinions of current approach and perceptions of electronic tagging.

8.2.1 Workflow

The full benefits of SITS are realised in the secondary processing phase and long-term storage and retrieval phases, where samples and data are used for knowledge discovery and data analysis. End-to-end evaluation of PCRC-SITS requires data on a statistically significant number of samples collected from donors being processed through all three phases. The constraints that prohibit or limit the end-to-end evaluation are:

- Availability of a statistically significant number of samples is required in order to draw valid conclusions for evaluating the contribution of the PCRC-SITS to reduce sample mix-up rate, as it is not very common. This in turn requires a sufficient number of donors to be identified over the period of the evaluation. However, the participating hospitals typically recruit only 1 or 2 donors per month.
- Samples collected in the initial processing phase are queried based on their scientific characteristics that should match the requirements of the particular study being undertaken at that time. Samples may not be selected for the secondary processing phase for several months. In addition to the delay in the process that this will cause, it also increases the number of samples that should be collected in the initial processing phase.
- End-to-end evaluation of the PCRC-SITS also requires completion of long-term storage and retrieval phase. It may be several years before samples are used for knowledge discovery and data analysis.

Thus, end-to-end evaluation of the PCRC-SITS will require a long period of time for the workflow to be completed, which is beyond the scope of this project. For these reasons, an alternative methodology is required. With this method, PCRC-SITS is evaluated at each phase of processing separately and the initial processing phase and secondary processing phase of the workflow are analysed. However, identification and tracking samples for the long-term storage and retrieval phase as well as overcoming the limited number of referrals of donors to the hospitals are not feasible within the timeframe of this research.

The initial processing phase will be tested in the collection hospital to evaluate the functionality of RFID technology in electromagnetic intense environments and also to allow research nurses to experience using the system. This will help to determine how user-friendly the interface is and how it meets users needs. The secondary processing phase is mainly concerned with omic procedures and longitudinal tracking. This will be tested in a research institute in the PCRC biobank. The process has been carried out by a researcher. SR1-SR3, SR5-SR9, SR13-SR17 of the requirements identified in Chapter 6 will be tested in this element of the evaluation.

8.2.2 Feasibility of Integrating SITS with BIMS

A fundamental and crucial requirement of the PCRC-SITS is integration in the biobank environment and hence BIMS, to provide consistency and standardisation. Also to gain full benefits of the PCRC-SITS and improve user experience of the system, PCRC-SITS ought to be integrated with BIMS. Various approaches towards integration have been identified and obstacles to employing each will be discussed. Application Programming Interface (API), Service Oriented Architecture (SOA), Dynamic Data Exchange (DDE) and Object Linking and Embedding (OLE) Automation will be studied in order to select the most suitable option for integration. After validating these approaches for integrating PCRC-SITS prototype with the PCRC BIMS, the best approach is implemented to prove the feasibility of such integration. Issues, constraints and limitations that the PCRC BIMS, Distiller, imposes on the integration method will assist in assessing the overall feasibility of integration.

8.2.3 Questionnaire

A questionnaire was distributed among the PCRC biobank members, to learn about their opinions about the current approach used for identifying and locating samples in the PCRC biobank. The PCRC biobank members are the main users of the system. The PCRC-SITS prototype has been introduced to the PCRC biobank members through demonstrations and presentations at the semi-annual PCRC biobank meetings. However, they have not used the actual system in practice, and therefore, a comparison between the PCRC-SITS prototype and the approach taken by the PCRC biobank cannot be drawn. However, user perceptions of electronic tagging will be covered in

the questionnaire to learn about their attitudes towards the technology. While the number of members of the PCRC biobank is small for drawing any statistically significant conclusions, they are the main users of the system who can best respond to the questionnaire. It is very important to focus the domain of responses to the questionnaire on the actual users who are involved in the real-life practice of carrying out procedures and who are familiar with the type of issues and concerns that arise in multi-institutional biobanks.

The two parts of the questionnaire will be (A) the users' opinions about the current approach taken for locating and identifying samples in the PCRC biobank and (B) the perception of users of electronic tagging of samples. The respondents are provided with free text fields at the end of each part of the questionnaire to make comments and suggestions. The comments and suggestions assist in understanding their general experiences and attitudes.

It is important to note that the questionnaire was distributed among the users who had not used the PCRC-SITS in practice and the date of distribution was before the evaluation in the secondary processing phase to avoid bias.

8.3 Workflows

As part of the PCRC-SITS prototype evaluation it is vital to ensure its usability and functionality from the users' point of view. Real-time evaluation often reveals valuable comments on the system that lead to improvements.

Full evaluation of the PCRC-SITS prototype requires tracking of a statistically significant number of samples from collection to omic procedure to storage and retrieval which is not possible in the context of this thesis. Therefore, to evaluate the PCRC-SITS prototype, the workflow was broken into two phases - the initial processing phase and secondary processing phase as shown in Figure 8-1. The research nurse and the scientist who assisted in carrying out the evaluation in the initial processing phase and the secondary processing phase, respectively, were provided with user guides given in Appendix L.

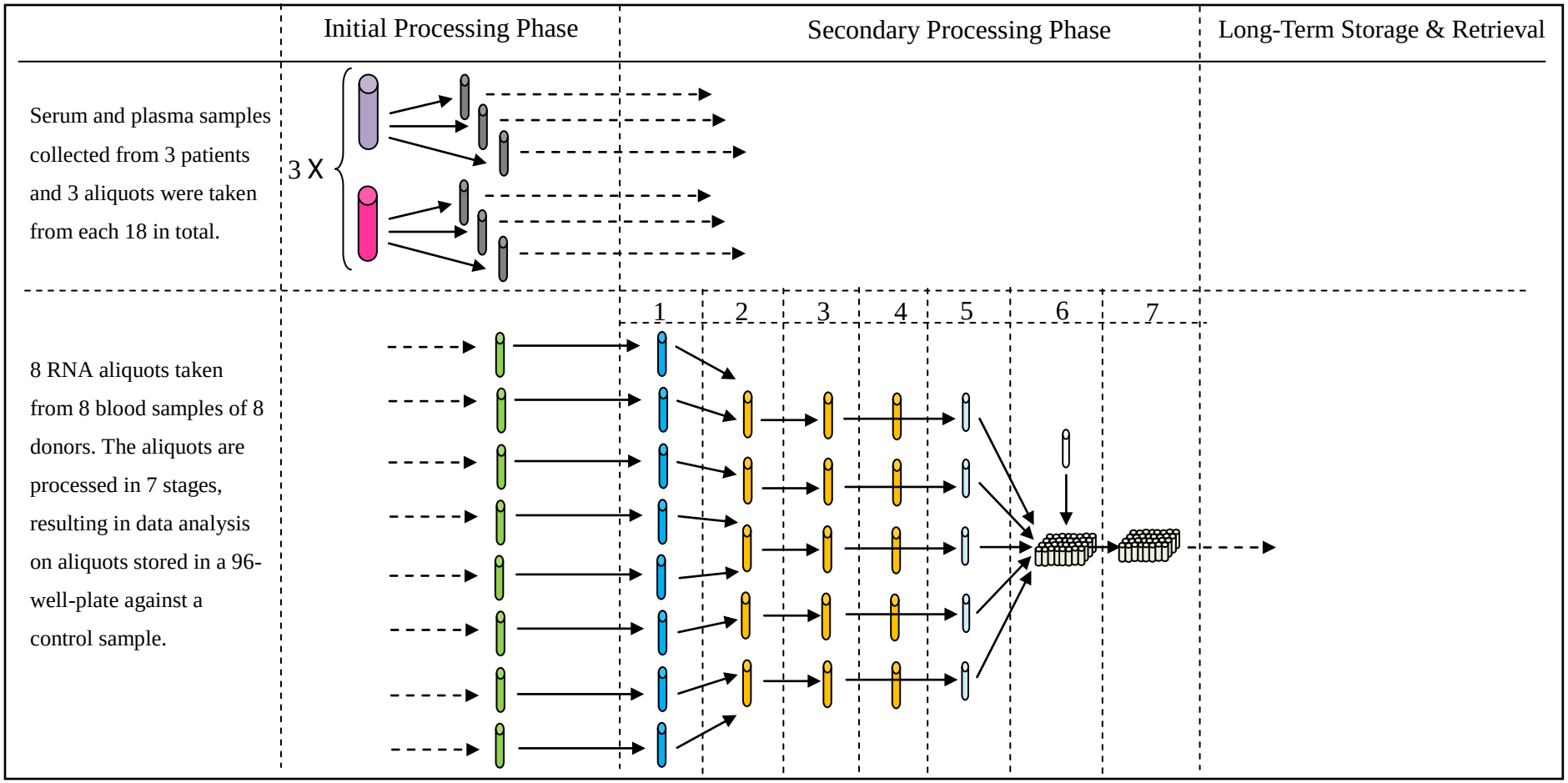


Figure 8-1: Evaluation of SITS on Workflows

8.3.1 Initial Processing Phase

To begin the procedures in the initial processing phase, three aliquots of each serum and plasma were collected from three participants, leading to a total of 18 samples. Since the samples of each donor were collected at the same time, tagged tubes were used for aliquots for the first time in the process. Prior to using the PCRC-SITS, the research nurse would collect samples in tubes that had typed labels on them, then later on she would aliquot samples in to pre-labelled tubes. She would record processing data in a book from collecting to aliquotting samples and would type data into the BIMS system when she would return to her desk. In Figure 8-1 plasma and serum samples are coloured in purple and pink, respectively, and the tagged tubes for aliquots are coloured gray. These aliquots are taken from the original tubes in which sample were stored at the time of collection. These samples were then stored in freezers at -20°C and transferred to research institutes where they were stored in freezers at -80°C. These 18 samples remained in the freezer of a research institute awaiting further analysis. Since using these aliquots for the secondary processing phase took longer than anticipated, given the timeframe of the project, evaluation of the secondary processing phase was carried out in a different research institute and on different samples. However, the principles and procedures involved are identical.

The research nurse who carried out the initial processing phase on the 18 samples was satisfied with the system based on the interface, ease of use, covering what is needed, meeting requirements, usability and functionality. She rated the system based on these criteria as shown in the form given in Appendix M. She rated 10 out of 10 for all of these criteria. The research nurse commented that the SITS interface is easy to follow and the layout follows the PCRC BIMS interface from her point of view. She also found storing data on the RFID tags easy and quick, specially the fact that data common across the three aliquots was retrieved automatically, so that the amount of typing required is minimised.

8.3.2 Secondary Processing Phase

The secondary processing phase was carried out on eight donor RNA samples extracted from previously collected blood samples stored in untagged tubes from

various PCRC biobank member hospitals. PCRC-SITS only allows collection data, that is initial sample data, to be entered and modified by a research nurse and each derived sample, that is the aliquot, must have its collection data. Therefore, dummy tags were used as collection samples, for which collection data were entered and from which aliquots taken. Dummy tagged samples are coloured in green in Figure 8-1. The aliquots went through RNA isolation and gene expression analysis. It is important to note that the tubes used for this procedure are different in shape from the ones used in other PCRC biobank institutes and hospitals. Two size of these tubes used for this procedure are shown in Figure 8-2. The lid of the tube is attached to the body with a loop and since the lid is a flat surface, the tag was glued onto it and survived the procedure. Figure 8-3 shows the tagged tubes stored in ice during the procedure.

Figure 8-2: Tubes Used for the Evaluation

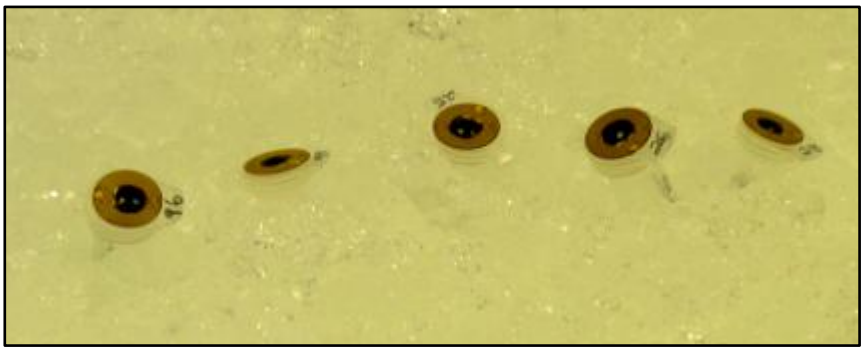


Figure 8-3: Tagged Tubes Stored in Ice

The SOP for the procedure, performing real-time Polymerase Chain Reaction (PCR) analysis, is provided in Appendix N. Based on this SOP a UML sequence diagram is drawn and is provided in Appendix 5 on the supplied CD. This diagram is then updated after incorporating the PCRC-SITS and is provided in Appendix 6 on the supplied CD. There are seven main stages of the process as labelled 1 to 7 in Figure 8-1. These seven stages are:

1. RNA isolation on small number of aliquots – eight samples from eight participants: 1 ml from each of the 8 aliquots was taken to adjust the amount of RNA in each. These tubes are tagged and their associated data is stored on PCRC-SITS.
2. Five of the eight RNA aliquots were of sufficient yield and purity. These were diluted to make a concentration. The remaining three aliquots were stored for future use and analysis. The five aliquots are shown in yellow in Figure 8-1.
3. Denature RNA by incubating at 70°C for ten minutes and let cool for five minutes.
4. RNA combined with reagents to make cDNA in tubes that are not tagged. This process is done instantly and no other procedure that involved relocating tubes was needed. Therefore, the tubes were not tagged.
5. cDNA aliquots then transferred to smaller tagged tubes, scanned and entered to the PCRC-SITS. These tubes are then processed in the PCR thermo cycler for two hours.
6. Finally three wells of the 96-well-plate were allocated for each of the five original aliquots and one control. The 96-well-plate is tagged and its data is stored on PCRC-SITS. However since PCRC-SITS is designed such that on each plate (or tray) only instances of one aliquot can be stored, only the first aliquot was linked to the 96-well-plate. Figure 8-4 shows the tagged 96-well-plate. To make the tag visible only part of the well-plate is shown in this Figure.
7. 96-well-plate is the final product of the process. Data analysis is carried out on the data associated with the samples stored in this well-plate.

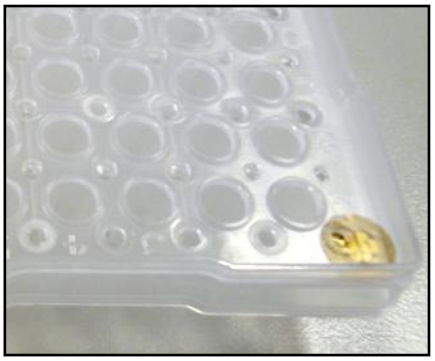


Figure 8-4: Part of the Tagged 96-well-plate

Part of the data stored in the PCRC-SITS database in the "DSample" table is shown in Figure 8-5. Each row of data represents a record associated with a derived sample. The field "InfoMem" combines all the data for each sample as they are written on the

tag memory. This field reduces the complication of storing each field of the table on the memory individually.

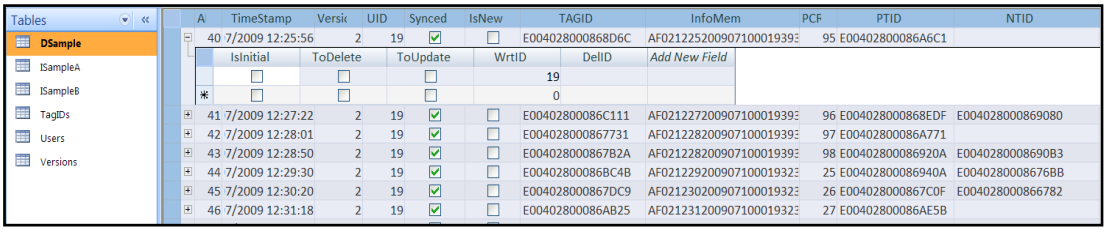


Figure 8-5: Data as They Appear in the SITS Database

The scientist found the system user-friendly and stated that its potential full deployment would be beneficial for controlling the quality and monitoring of sample availability. She found knowing what conditions the sample has gone through by storing temperature changes and duration of these changes on the sample particularly useful as it tracks the conditions the sample has gone through. This could assist her in selecting those samples which have undergone less variation for sensitive procedures. With the PCRC-SITS prototype the temperature data are manually entered by the user. She was interested in storing the identification and tracking data on the sample. She was satisfied with the facility that PCRC-SITS provides that would allow her to browse back and forth to the original samples that a given aliquot was taken from or to the aliquots that were taken from a particular sample. She also found the automatic upload of data already stored in the database to certain fields useful in reducing the typing errors. Automatic capture of sample identifier, TagID, from the reader was also appreciated. Her comments on the system emphasised the importance of keeping track of samples and aliquots that enter the biobank system. She was concerned about recording the aliquots that no longer exist and have been used up in another omic procedure. So that the data and results of analysis on that aliquot is maintained in the system, however it should be flagged as a used up aliquot. It was recommended that the record for such aliquots be maintained along with the outcomes of analysis; however it should be visible on the record that such sample does not exist anymore. Also she emphasised the importance of noting the volume of the aliquot left in the freezer that can be used by other researchers, and being able to access the results of analysis that have been carried out on the remainder of that aliquot.

Freeze-thaw cycles influence the quality and usability of samples significantly so automatic capture of major temperature changes would improve analysis. PCRC-SITS in its current status requires the users to manually enter the temperatures that the aliquot has been exposed to and the duration of this exposure. The scientist raised the issue of remembering the exact time that the sample has been at a particular temperature. Integrating PCRC-SITS with temperature sensors or temperature loggers is potentially feasible by allowing data entry to the PCRC-SITS database or a log file from sensors.

The PCRC-SITS needs constant update of data and multiple tags being scanned throughout the procedure. During the analysis it was noted that inserting data on the PCRC-SITS might be forgotten when processing aliquots, when the scientist is concentrating on the analysis. Incorporating PCRC-SITS operations in SOPs should help scientists to know where in the process they should be using the tagged tubes, what data to store and to integrate the use of the system with their analysis. Implementing PCRC-SITS operations in SOPs is a fundamental aspect of integrating PCRC-SITS into the biobank environment.

It was also noted that in addition to the User Guides that were provided to the research nurse and the scientist who tested the system, a set of training sessions need to be in place before full deployment of system. This would allow users to work with it to make sure they are confident with the technology. This should be included as part of the training for BIMS.

PCRC-SITS is designed so that each tray or box can only be linked back to one unique aliquot or tag. Although in this process five aliquots were transferred to one well-plate that had gone through the same procedure. It is important for future developments to allow such flexibility. Also, it is important to note that the RFID tags survived the process as expected.

8.4 Feasibility of Integrating SITS with BIMS

Since data management and communication in biobanks are carried out by BIMS, SITS operations should be undertaken as part of this system. Integration provides

users with a simple system for all biobank activities. It also reduces conflicts and complexities created by having multiple systems, and eliminates incompatibility issues that may arise between the two systems. It will also assist in maintaining the security and integrity of data.

Developing SITS concurrently with the development of BIMS allows full integration of the two systems. SITS in this case is part of the BIMS that is built in a compatible fashion with it. Standards for communication are set, and duplication of data is minimised. The three major components being considered for RFID- or barcode-based SITS, that is, database, application and interface, are then implemented in BIMS and hence maintained more securely while reducing duplication of records and integration effort. Approaches to implementation depend on the technology that is going to be used, hardware and devices involved, desired accessibility, data records available to biobank members and many other parameters depending on the rules and regulations of the biobank and its structure.

The biobanks that intend to develop SITS at later stages of their development have either purchased an off-the-shelf BIMS product or implemented their own BIMS. Biobanks with off-the-shelf BIMS purchased from a third party have little or no control over BIMS features. The most complicated scenario for integrating SITS is when the BIMS is a product from a third party where there is no access to the interface code or the database. In this case it is necessary to establish some form of interface between the two systems. It would be reasonable to expect such a BIMS to provide an Application Programming Interface (API) to make communication feasible between the BIMS and other systems.

Biobanks that have developed their own BIMS have control over the features of the BIMS and hence the BIMS, depending on the design, is fully or partially customisable. Biobanks with a long history of sample collection seeking implementation of SITS or improvements to their original SITS, that has been developed as part of their BIMS, have the option of modifying their entire BIMS or upgrading their current SITS, depending on the architecture of BIMS and/or SITS that is in place. In either case they would need certain tables being available in their BIMS database for storing sample identification and tracking data, in addition to the tables they have for storing clinical

and omic data. SITS also needs an interface which depends on the interface of BIMS. This might be a webpage or an application running locally. This interface should be such that communication with the barcode or RFID reader devices is feasible. Therefore, it is highly recommended that biobanks allow for expansion or further development when implementing their BIMS. It is also recommended that the new SITS be developed such that reusability of the interface and database is maximised. This will ease the process of upgrading the system to availability of new technologies in the future.

It is important to decide on the best approach based on the status of the biobank and its BIMS. In general, the more flexible the BIMS the easier integration with SITS will be.

8.4.1 Integrating PCRC-SITS and PCRC BIMS

The PCRC biobank has employed an off-the-shelf BIMS, developed by Slidepath (SlidePath). This system is provided as a product, Distiller. An image of a dummy participant record is shown in Figure 8-6. The database used for Distiller is MySQL and access to the tables and fields is made possible through an interface that is only available to Administrator account holders. Access to Distiller, or the BIMS database, is highly secured by the approaches described in Chapter 3. Therefore very limited room is left for the addition of new services or modules.

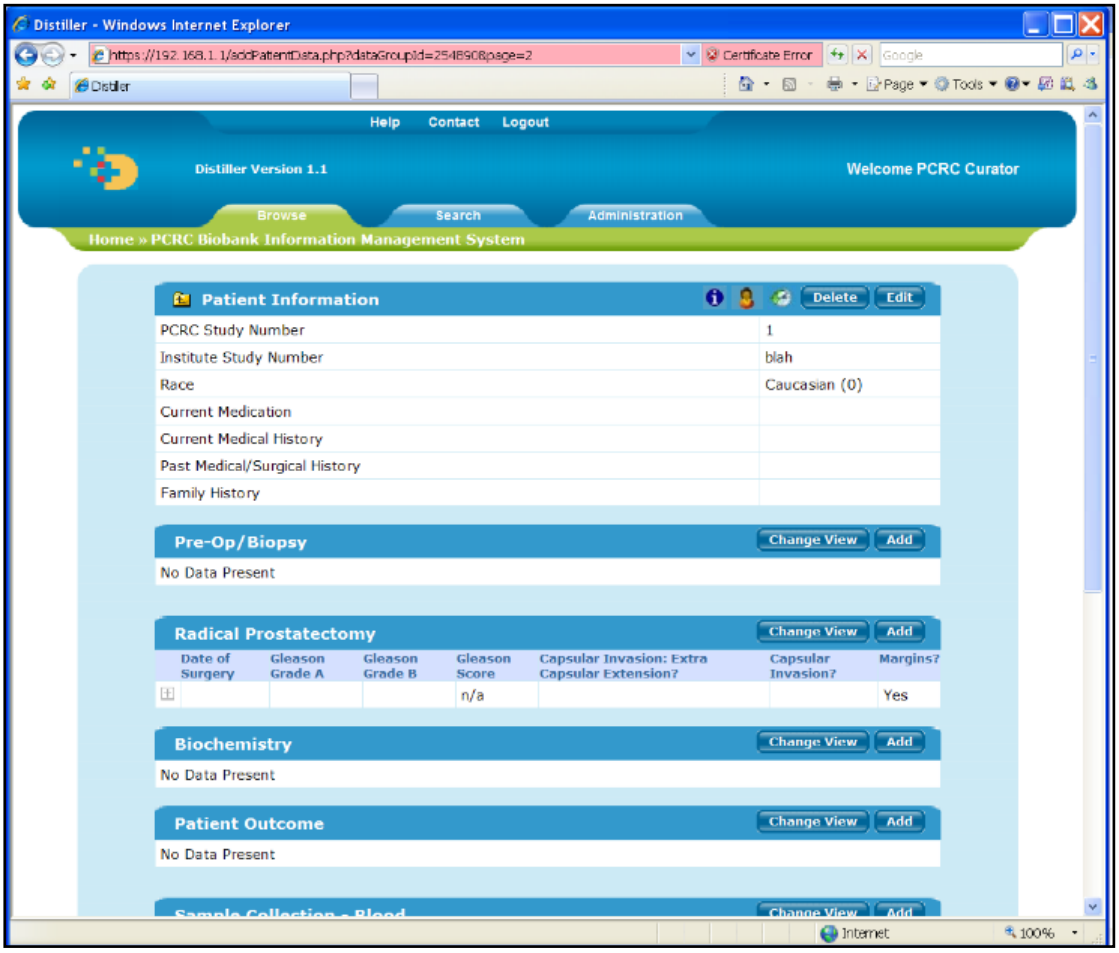


Figure 8-6: A Dummy Participant Record on Distiller

In its current status the BIMS central database includes three tables for each major type of the collected samples, i.e. Blood, Tissue and Urine. Figure 8-7 shows an extract from the PCRC biobank central database structure that depicts the current sample identification and tracking in BIMS central database. The full schema of the Distiller database for the PCRC biobank is provided in Appendix O. These three tables include details of collection data for each type of sample. Location information of aliquots is stored in tables linked to each major sample type. Hence the *Blood* table is further linked to four other tables that are *Plasma*, *Serum*, *DNA* and *RNA*, and the *Tissue* table is linked to the *Fresh-RNA Later* table. Since urine sample is not aliquotted and is stored as one unique sample in the initial phase, details of its location is stored in the same table as the collection data, the *Urine* table. It is important to note

that in the current status of the PCRC BIMS these tables are only used for locating samples. These tables might be considered as placeholders for future developments.

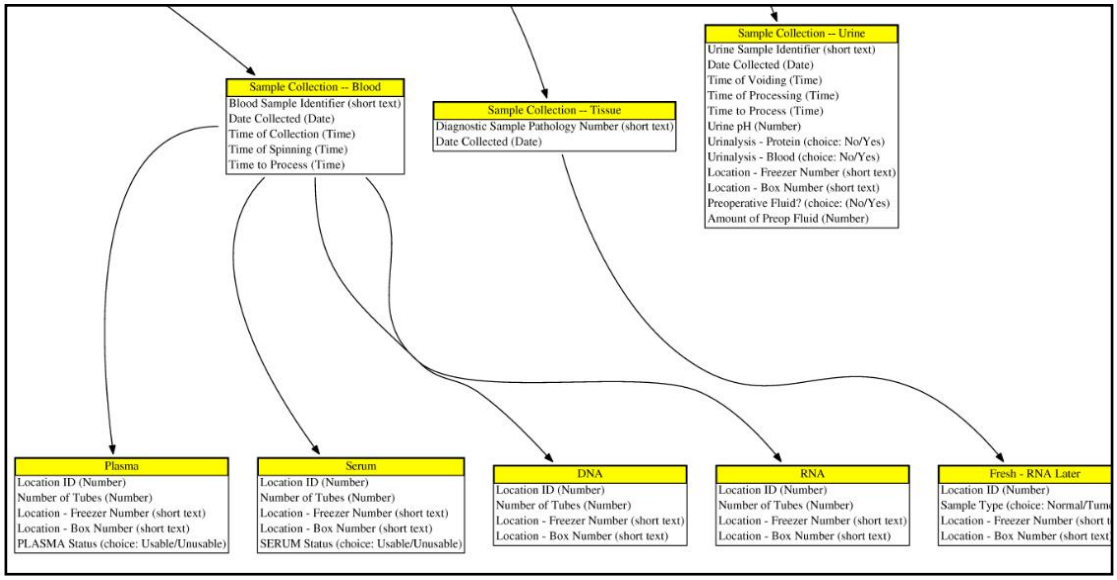


Figure 8-7: Sample Identification and Tracking in BIMS Central Database

The PCRC BIMS database can be edited by administrators and therefore the tables that the BIMS has in place for locating samples can be upgraded to include the tables PCRC-SITS has in its database. However, the interface Distiller provides does not support direct access to the PCRC-SITS application program. Hence, feeding data directly from the tag memory to the database and vice versa is not possible. Also its interface is not editable. At the moment, Distiller does not support URL as a data type. However it has a “bug” that allows html codes be inserted in data fields that a URL could be added. Linking the two systems requires certain data be sent to the PCRC-SITS by POST or GET methods, which is not feasible with the current version of Distiller. Furthermore, linking the two systems would not resolve the integration issues, as it is merely linking two pages. There are also other objections to linking systems using URL such as double authentication.

Application Programming Interface (API)

The API provided to the BIMS needs to facilitate export and import of data to and from other applications. This API can be considered as an add-on to the PCRC-SITS that links this system to BIMS and allows two way interactions. Some challenges with

regard to the technology used for the PCRC-SITS exist that need to be tackled at the time of programming the PCRC-SITS and depending on the API provided. However, in the most recent version of Distiller available at the time of compiling this thesis, no API has been made available.

An API that is compatible with the PCRC-SITS prototype would preferably have the following characteristics to make communication feasible:

- Act as an extension (add-on) that exports/imports the string message that get.asp works with. These messages are described as commands and responses in Chapter 7.
- Feed the PCRC-SITS data to the BIMS web interface automatically. The BIMS forms should be static, and in case of dynamic forms, a dynamic application should be developed.
- Setting the get.asp to send data to the destination of BIMS forms, i.e. the page that posts data to the underlying database. In this case data preparation should be done so that data match the database fields.

Service Oriented Architecture (SOA)

Deploying SOA for integration between the PCRC-SITS web and the BIMS web would need both of them to be considered as services. However, the PCRC BIMS web cannot be thought of as a service according to the definition given previously. BIMS, i.e. Distiller service at its current status is not consumable by clients or processes. Furthermore, the PCRC-SITS prototype will have to go through major adaptation to be used as a service in BIMS. However, for biobanks whose BIMS web meets the definition of a service, SOA is a feasible option for integrating SITS with BIMS.

Dynamic Data Exchange (DDE) and Object Linking and Embedding (OLE) Automation

Automation is a preferred approach for exchanging data across two applications, as DDE deals with a shared memory and Automation deals with processes. With the current status of BIMS, one would need transfer of data from PCRC-SITS to the Distiller web through the interface. This way data is uploaded to the central database.

Automation and DDE are not the best approaches to integration as they are incapable of handling variations that may occur in interfaces and databases. However, they are the only methods available in the worst case scenario when access to BIMS interface code is prohibited and this is the case with the PCRC BIMS. Although DDE approach is the only way that one could use to link to the BIMS interface, there are pros and cons to it. Through this approach data is being transferred from the database of PCRC-SITS, through its interface and SITS Win32 Application to the Distiller interface. The issues that may arise using this approach are:

1. Data fields in Distiller are named dynamically, so any changes in data fields will require the xtobimsprc.asp page to be modified accordingly.
2. Through this approach Distiller pages are browsed to get to a record. Any changes to the links of the middle pages should be updated on the xtobimsprc.asp page. This will require a lot of effort.
3. In case of interruption on the SITS Win32 Application no data can be transferred.
4. Implementation of this approach is highly dependent on the BIMS interface, hence it is not adoptable to other biobanks. However, the basic idea of adopting Automation is feasible.

8.5 Questionnaire

An anonymous questionnaire seeks to elicit the opinions of the members of the PCRC biobank on the sample identification and tracking method that BIMS provides as well as their perceptions of RFID and electronic tagging of samples in general. A blank version of the questionnaire is included in Appendix P. The aim of this questionnaire is to learn about the opinions of users and their perceptions of electronic tagging. Thus perceived usefulness and perceived ease of use, the basis of Technology Acceptance Model (TAM) (Davis, 1989), are not being considered.

The questionnaire was distributed in hard copy at a PCRC biobank semi-annual meeting and collected on the same day. In this questionnaire, 21 members of PCRC biobank were asked 12 questions in addition to a question about their position in the PCRC biobank. All 21 copies of the questionnaire were returned answered, so the return rate of the questionnaire was 100%. Part A, which is entitled "*Opinion of the*

users of the current BIMS identification and tracking method." consists of eight questions, including a few multiple part questions, and part B, which is "*Individuals perception of electronic tagging to identify and locate samples.*". This covers four questions with one multiple part question. The questionnaire included open text fields at the end of each section for their comments.

The first general question asks about the position of individuals by "*Which one of the following best describes your position?*" and the available options are "*Researcher/Scientist (R)*", "*Research Nurse (N)*", "*Principal Investigator (PI)*" and "*Other, Please specify....*". From the 21 responses received, the majority, 12, were "*Researcher/Scientist (R)*", none chose "*Research Nurse (N)*", two were "*Principal Investigator (PI)*", four chose "*Other, Please specify....*" and three left blank (Figure 8-8). Of the four people who chose "*Other, Please specify....*" one described him/herself as a computer scientist, one as an IT and one as biobank admin/research scientist. The size of the population under study is not sufficient for drawing statistically significant conclusions; however, the results provide insights which could be useful for further development of the system.

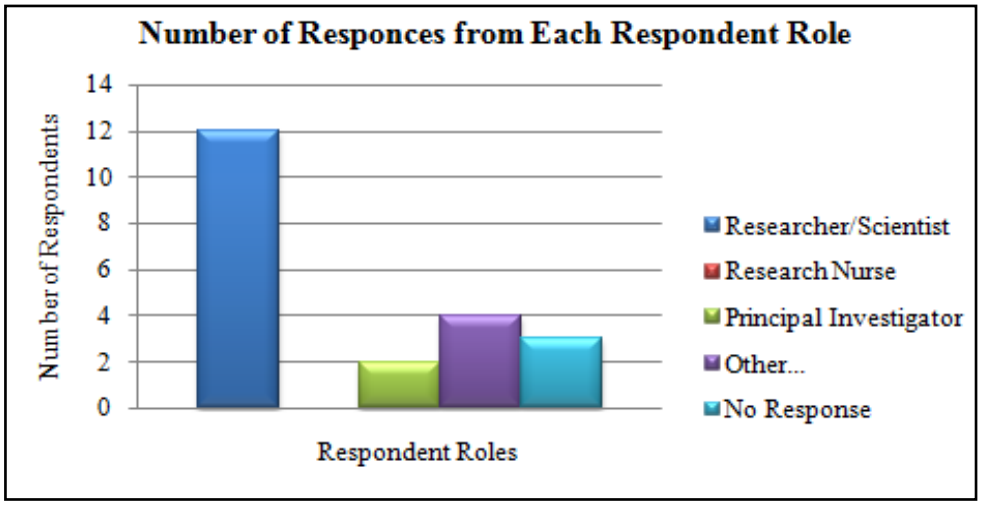


Figure 8-8: Number of Respondents from each Respondent Role

8.5.1 Part A: Opinion of BIMS Users

Of the 12 *Researcher/Scientists,* two *Principal Investigators (PI)* and three blank responses for position, four, one and two people, respectively, were engaged in activities that involved "*transferring a sample from a tube to another tube*". When

asked to "*rate the likelihood of mixing samples up, while aliquotting them*" one gave no answer, two people chose *very unlikely*, three people chose *unlikely*, and only one person chose *likely* (Figure 8-9). Excluding the no answer response, an average of the six responses is 1.83 (from (1+3+2+2+1+2)/6=1.83) which can be mapped to a point between *very unlikely* and *unlikely*, being close to *unlikely* when looking at the graded scale.

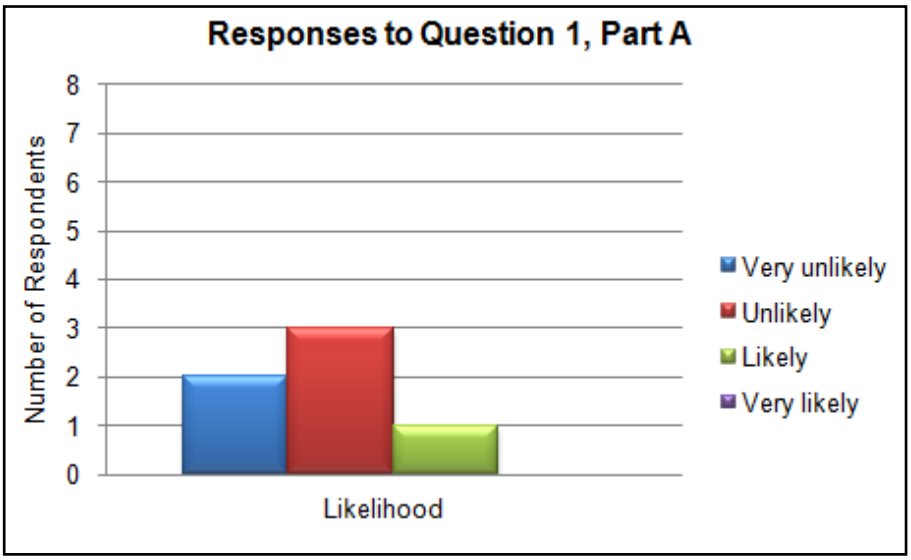


Figure 8-9: Responses to Question 1, Part A

Second question from Part A asks the users to rate the current tracking system by asking them "*How do you rate the current tracking system of samples using BIMS?*". They are given a scale with two points marked ranging from *not very useful* to *very useful*. The mid points are taken as *not useful* and *useful.* Out of the 21 respondents, 12 answered this question, from which six were *Researcher/Scientist*, one was *Principal Investigator*, one was an *IT person*, one was *biobank admin/research scientist*, and three were unanswered positions. Figure 8-10 shows the distribution of responses. The average of these responses is 2.42 (from (3+4+2+1+3+3+2+2+3+2+2+2)/12=2.42). On the scale, this is a point between *not useful* and *useful* and is closer to *not useful*. The roles of people who responded to this question vary and while the sample size is quiet small, the results can only be considered as a general opinion of the PCRC biobank members to the current tracking system.

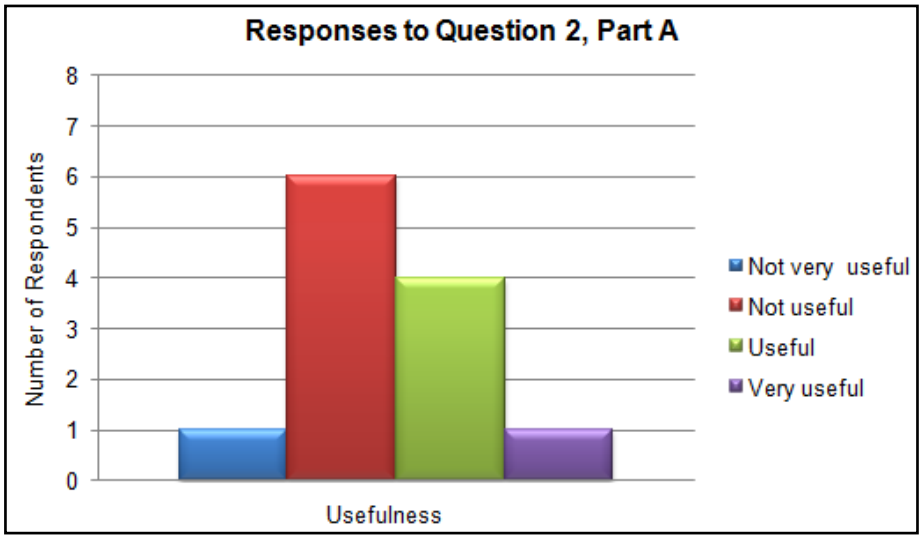


Figure 8-10: Responses to Question 2, Part A

Respondents were asked "*Are you involved in storing samples?*" and if they are involved asked "*how do you rate the procedure?*". A total of six of the 21 respondents' answers were positive, five of which are also involved in transferring a sample from a tube to another tube. They were asked to rate the procedure on two separate scales, one ranging from *very easy*, *easy*, *difficult* and *very difficult* and the other one ranging from *very quick*, *quick*, *time taking* and *very time taking*. Distributions of rates for both scales are depicted in Figure 8-11 and Figure 8-12. Average responses for the first scale, looking at the difficulty of the procedure is 2.17 (from (2+2+3+3+2+1)/6=2.17), a point between *easy* and *difficult*, being closer to being *easy*. For the second scale there is one missing answer (left blank). Taking the average of the rest five answers gives 2.6 (from(3+4+2+2+2)/5=2.6). According to the scale, 2.6 is between *quick* and *time taking*, closer to *time taking*.

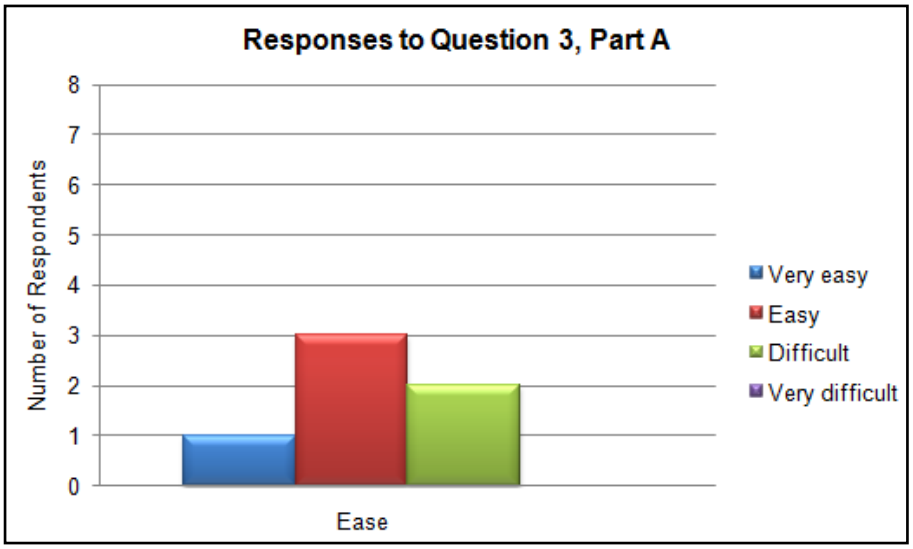


Figure 8-11: Responses to Question 3, Part A, Based on Ease

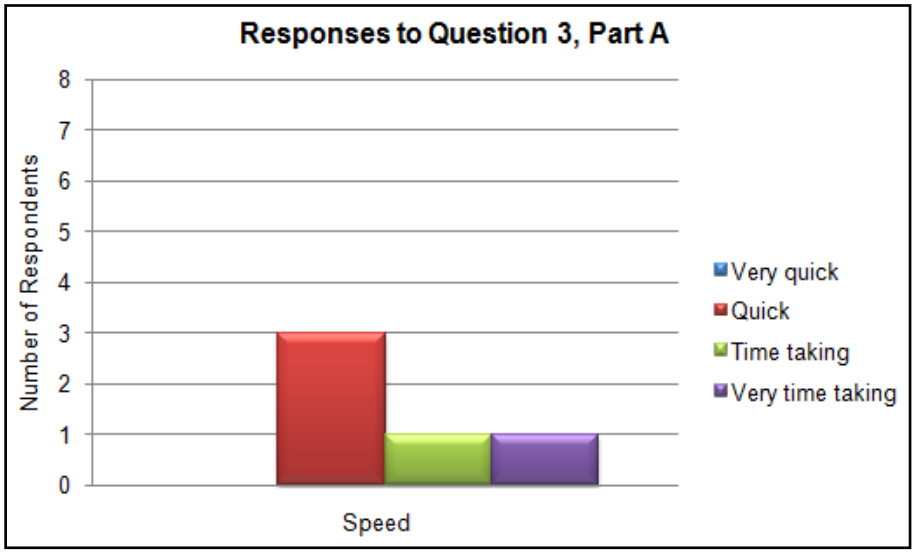


Figure 8-12: Responses to Question 3, Part A, Based on Speed

Then individuals who are "*involved in uploading the location data on the BIMS*" are asked how they rate the procedure with regard to difficulty and speed. They are given two scales identical to the previous question. Two people are involved in uploading data on the BIMS, one of them found the procedure *easy* and the other one found it *very easy*. It seems that both of them found uploading the location data on BIMS easy to some degree. Only the second person answered the question with regard to speed, and chose *time taking*. Responses to this question are so few that no significant analysis may be carried out on them.

Question five then asked the respondents to rate their "*confidence in the current BIMS system to prevent the following*" based on "*Sample mix up during aliquotting*", "*Sample degradation*" and "*Sample misplacement*". Five *Researcher/Scientists*, one *Principal Investigator*, one *biobank admin/research scientist*, one *IT person* and three unknown answered this question. From the 10 individuals who rated their confidence in BIMS on "*Sample mix up during aliquotting*" one chose *very unconfident*, three chose *unconfident*, four chose *confident* and two chose *very confident* as shown in Figure 8-13. The average is then 2.7 (from (2+3+3+3+1+4+3+2+2+4)/10=2.7) which on the scale is a point between *unconfident* and *confident* being closer to *confident*. Furthermore, from the same 10 people who rated their confidence in BIMS on "*Sample degradation*", only one person was *very unconfident*, five people were *unconfident and* four were *confident* as shown in Figure 8-14. From these 10 responses one may conclude the average of responses is 2.3 (from (2+2+2+3+3+3+3+2+2+1)/10=2.3), a point between *unconfident* and *confident* being closer to *unconfident*. It is worth noting that only one response was at the extreme of being *very unconfident*. From the 11 responses given to the third part of this question, "*Sample misplacement*", seven individuals chose *confident*, three chose *unconfident* and only one chose very *unconfident* as shown in Figure 8-15. The average based on these would be 2.55 (from (3+1+2+3+3+3+3+3+2+2+3)/11=2.55) which is almost a midpoint between unconfident and confident.

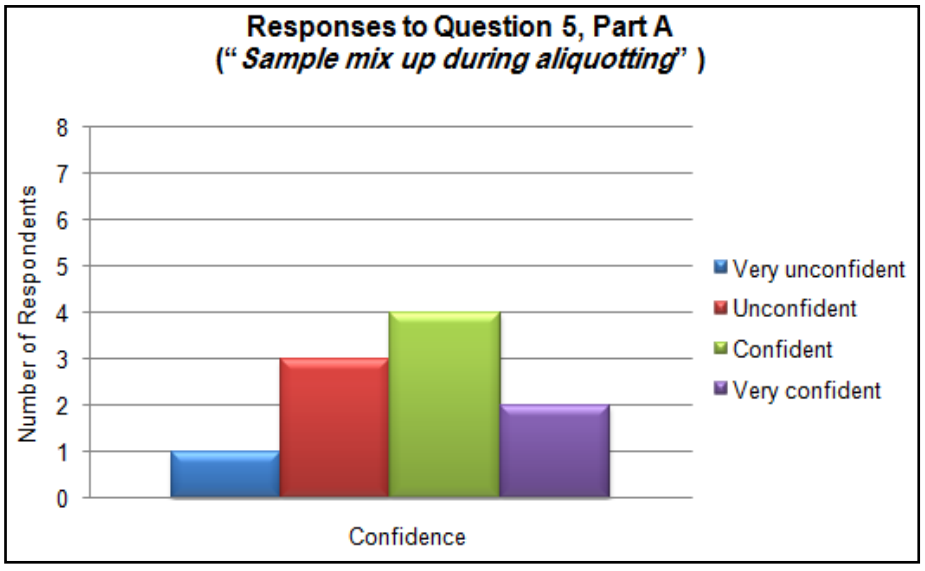


Figure 8-13: Responses to Question 5, Part A, "*Sample mix up during aliquotting*"

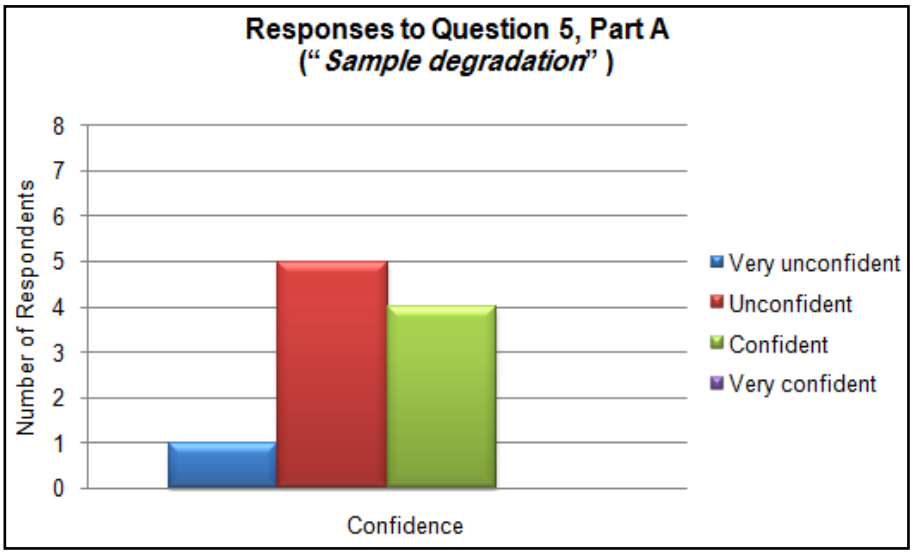


Figure 8-14: Responses to Question 5, Part A, "*Sample degradation*"

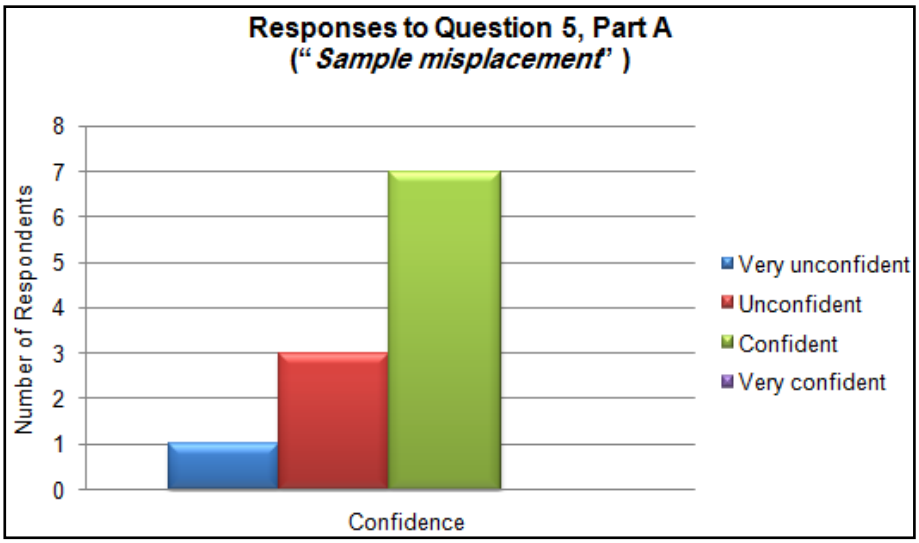


Figure 8-15: Responses to Question 5, Part A, "*Sample misplacement*"

Question six then asked if the respondents have ever come across a sample that they cannot identify. Six of the answers were *Yes* or *No* and the remaining were either left blank or marked as *Not applicable* (Figure 8-16). From the six, only one person answered *Yes*, and her/his role is not known. Hence, five people responded *No*, had not came across a sample that they could not identify.

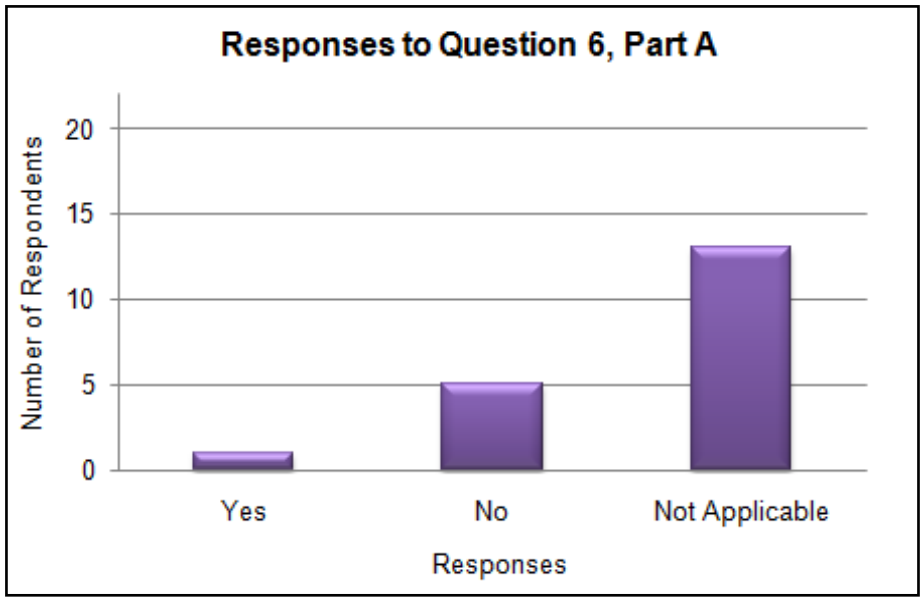


Figure 8-16: Number of *yes*, *no* and *not applicable* Responses to if Respondents have Came Across a Sample They Cannot Identify (Question 6, Part A)

From the population 16 people either marked *Not applicable* or left the answer blank to the seventh question asking "*Have you ever not been able to find a sample in its expected location indentified by BIMS?*". The remaining four people's answers were negative, showing that they have never come across a sample that they could not locate.

In answer to the final question of this part that asks "*Have you ever had difficulties reading the labels on the tubes for example due to being covered by frost layers?*", six peoples' responses were either *Yes* or *No*. The remaining answered either *Not applicable* or left the answer blank (Figure 8-17). Of the six, four individuals answered *Yes* and 2 chose *No*, showing that most of the small sample population had experienced difficulties reading the labels on the tubes at same time.

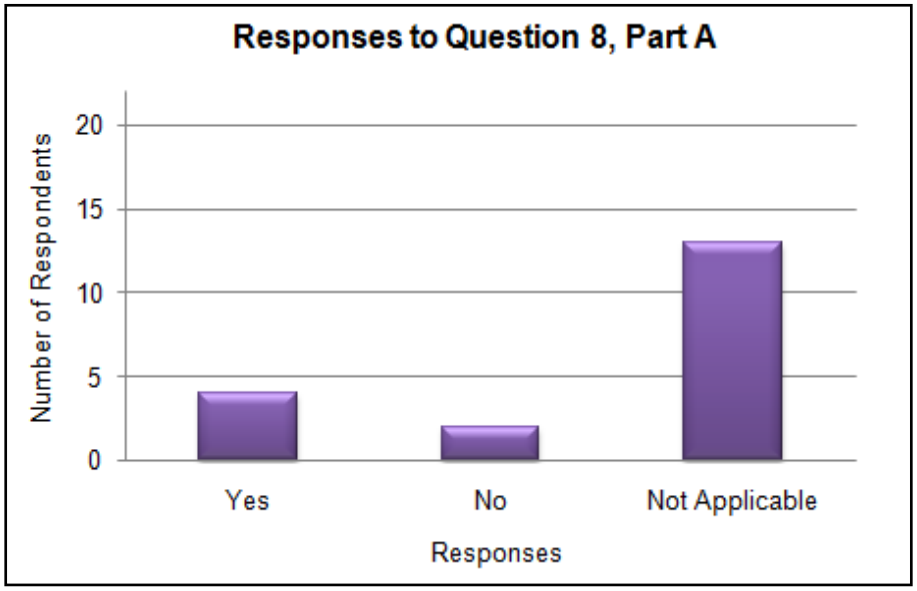


Figure 8-17: Number of *yes*, *no* and *not applicable* Responses to if Respondents have had Difficulties Reading the Labels on Tubes (Question 8, Part A)

8.5.2 Part B: Perception of BIMS Users of Electronic Tagging

Part B is about the perception of users of electronic tagging. There are four questions in this section.

The first question asks the PCRC biobank members "*Would a system for monitoring temperature changes during transfer of location of samples be an advantage to your research?*". Only one of the respondents left the question unanswered. 12 of the 20 answers were positive, stating that 60% of the sample population would find a system for monitoring temperature changes during transfer of location of samples be an advantage to their research. Five people answered *Maybe* to this question, while three others were negative answers showing that 15% of the sample population did not find such a system an advantage to their research. Results of this question are reflected in Figure 8-18.

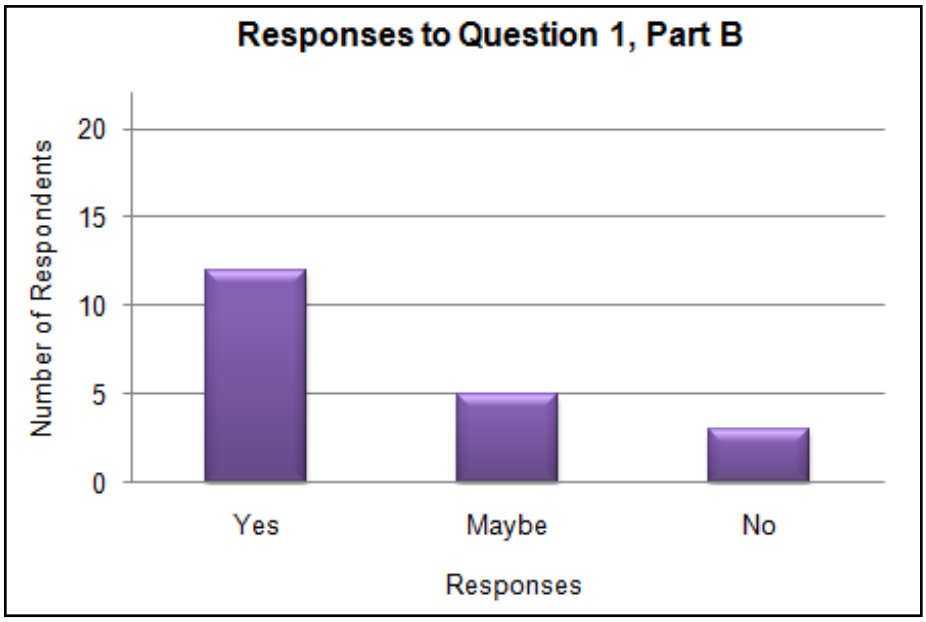


Figure 8-18: Number of *yes*, *maybe* and *no* Responses to if Monitoring of Temperature would an Advantage to Respondents' Research (Question 1, Part B)

The second question then asks "*Would you think it is useful for quality control purposes to know what processes the sample has gone through? e.g. Number of freeze-thaw cycles, centrifugation.*". Apart from one respondent whose answer was negative, all others answered the question by choosing *Yes* (Figure 8-19). It can be concluded that 95% of the sample population agree that it is useful for quality control purposes to know what processes the sample has gone through. In a note from one respondent it was stated that his/her response would be *Yes* for knowing number of freeze-thaw cycles. This answer is counted positive.

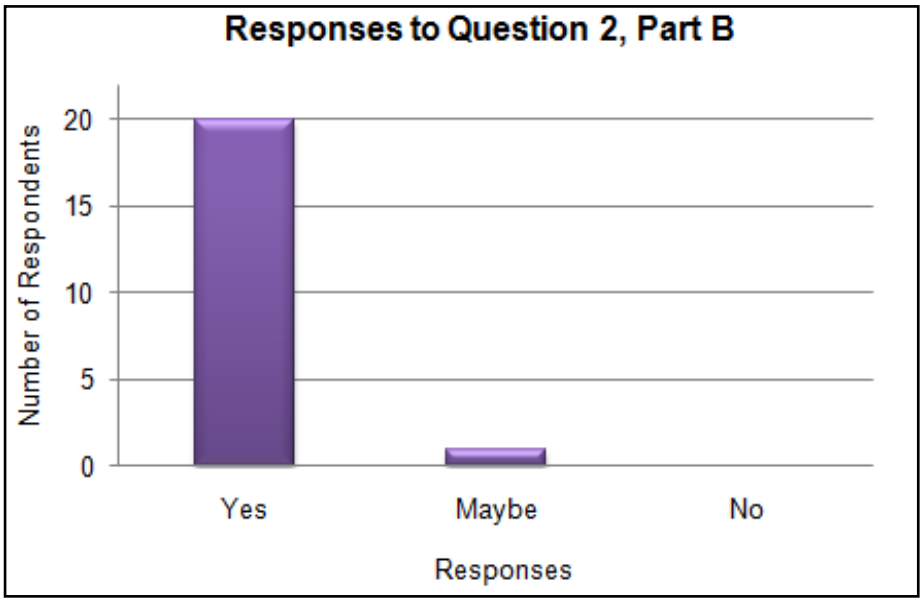


Figure 8-19: Number of *yes*, *maybe* and *no* Responses to if Respondents Think it is Useful for Quality Control Purposes to Know What Processes the Sample has Gone Through (Question 2, Part B)

The third question then asks if the respondents think "*storing sample information (including sample identifier, tracking information and temperature changes) on a tube would prevent incidents below?*" and then lists the three incidents that were also listed for question five part A, "*Sample mix up during aliquotting*", "*Sample degradation*" and "*Sample misplacement*". They are given three options to choose from, *Yes*, *No* and *Not sure*, unlike question five where they had to rate. 18 people responded *Yes* with regard to preventing *Sample mix up during aliquotting*, two chose *No* and one was *Not sure*. 16 also responded *Yes* to preventing *Sample degradation*, four chose *No* and one person was *Not sure*. Finally, from the 21 received responses for this question with regard to *Sample misplacement* only one person was *Not sure* and the remaining 20 were positive. Hence from the population 86%, 76% and 95% think storing sample information (including sample identifier, tracking information and temperature changes) on a tube would prevent *Sample mix up during aliquotting*, *Sample degradation* and *Sample misplacement*, respectively. Results of this question are depicted in Figure 8-20.

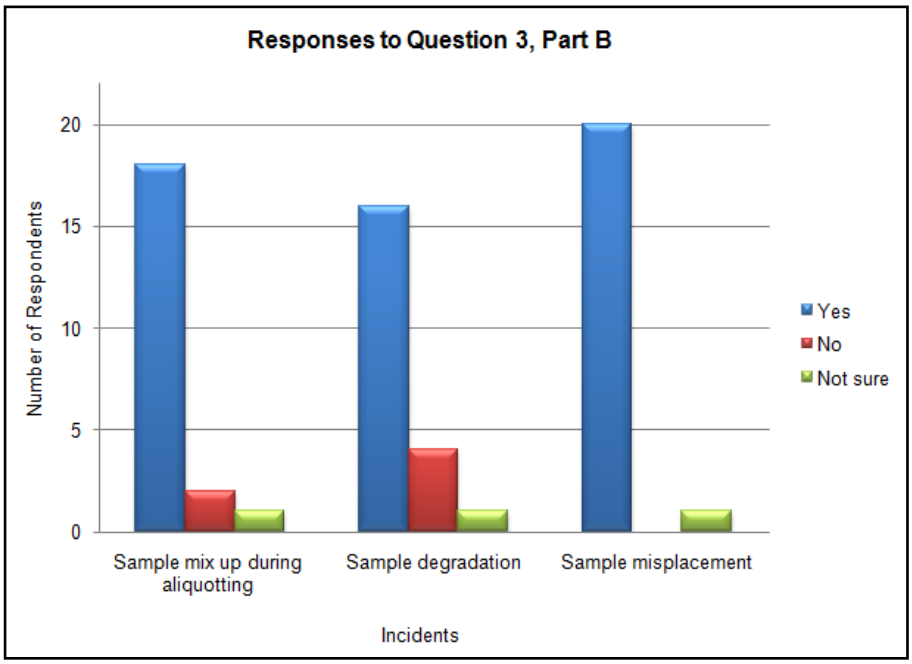


Figure 8-20: Number of *yes*, *maybe* and *no* Responses to if Respondents Think Storing Sample Information on a Tube Would Prevent Incidents the Three Incidents Shown in the Graph (Question 3, Part B)

Finally it is asked if they have "*ever heard about electronic or RFID (Radio Frequency Identification) tags?*" and were given *Yes* and *No* options. From 21 responses, 10 had not heard about RFID tags before and 11 had (Figure 8-21). Seven of the 10 people who had not heard about RFID tags before, categorised themselves as *Researcher/Scientist*, one as a *Principal Investigator*, one as a *biobank admin/research scientist* and one is not known. From this it can be drawn that about half of the people who took part in the study, had not heard about RFID tags before.

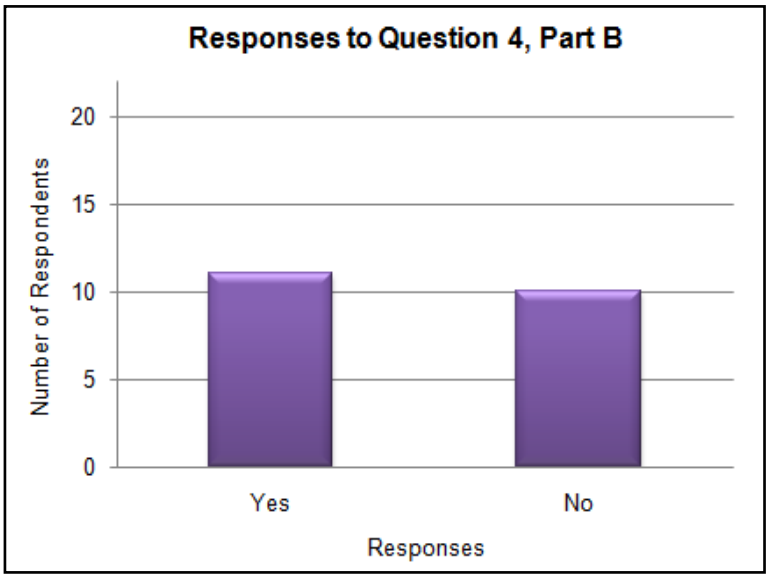


Figure 8-21: Number of Respondents Who have Heard and have not Heard About Electronic Tagging or RFID Before (Question 4, Part B)

8.5.3 Analysis

The size of the population who took part in the study is too small for any statistically significant analysis. However further analysis of the results of the questionnaire will be carried out to learn about the general opinion and attitude of the PCRC biobank members.

Two roles, *Researcher/Scientist* and *Research Nurse*, work closely with samples and are involved in procedures and hence are considered the potential users of PCRC-SITS. Therefore, responses from these roles in addition to the individual who described him/herself as a *biobank admin/research scientist* are extracted for further analysis.

Looking at responses to question five Part A, reveals that on average *Researcher/Scientists*, *Research Nurses* and *biobank admin/research scientist* are *confident* in the current BIMS system to prevent *Sample mix up during aliquotting*. On average they are more *unconfident* about prevention of *Sample degradation*, and are at a midpoint between *unconfident* and *confident* with regard to preventing *Sample misplacement*. This analysis is shown in Table 8-1 and is depicted in Figure 8-7. However, taking into account all responses to this question makes a slight change to

the averages of each. The average confidence for prevention of *Sample mix up during aliquotting* is reduced, while it has slightly increased for *Sample degradation* and *Sample misplacement* (Table 8-2 and Figure 8-23).

Respondent Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**16**	**18**	**19**	**21**	**Avg.**
Sample mix up during aliquotting	2	-	-	-	3	-	-	-	-	4	3	2	4	**3.0**
Sample degradation	2	-	-	-	2	-	-	-	-	3	3	2	1	**2.2**
Sample misplacement	3	-	-	-	1	-	-	-	-	3	3	2	3	**2.5**
Key: 1 = very unconfident, 2 = unconfident, 3 = confident, 4 = very confident														

Table 8-1: Responses from Three Important Categories of Users to "*How would you rate your confidence in the current BIMS system to prevent the following?*"

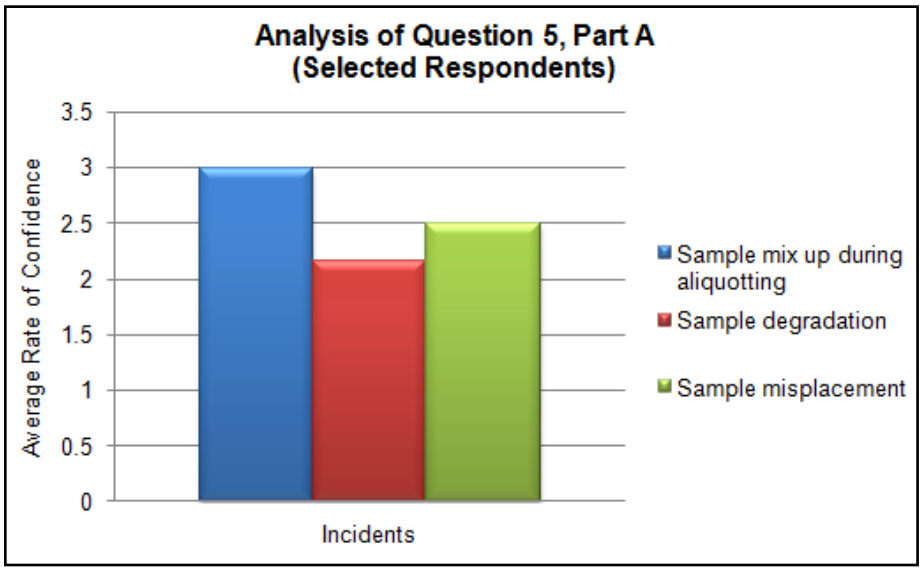


Figure 8-22: Graphical Representation of Table 8-1

Respondent Number	**1**	**5**	**11**	**12**	**14**	**15**	**16**	**18**	**19**	**20**	**21**	**Avg.**
Sample mix up during aliquotting	2	3		3	3	1	4	3	2	2	4	**2.7**
Sample degradation	2	2		2	3	3	3	3	2	2	1	**2.3**
Sample misplacement	3	1	2	3	3	3	3	3	2	2	3	**2.55**
Key: 1 = very unconfident, 2 = unconfident, 3 = confident, 4 = very confident												

Table 8-2: Responses from Respondents to "*How would you rate your confidence in the current BIMS system to prevent the following?*"

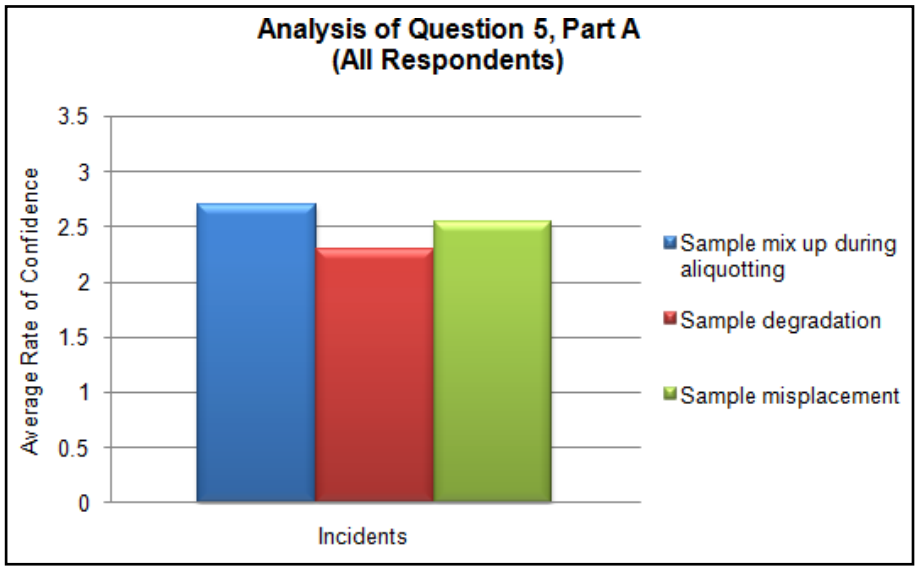


Figure 8-23: Graphical Representation of Table 8-2

Question three Part B, asks the users if they think "*storing sample information (including sample identifier, tracking information and temperature changes) on a tube would prevent*" each of the incidents rated in question five Part A, i.e. *Sample mix up during aliquotting*, *Sample degradation* and *Sample misplacement*. Analysing responses from the three potential users, *Researcher/Scientists*, *Research Nurses* and *biobank admin/research scientist* roles reveals that about 77% of them are positive with regard to *Sample mix up during aliquotting*, about 62% and 92% are also optimistic about prevention of *Sample degradation* and *Sample misplacement* if storing sample information on a tube. This analysis is reflected in Table 8-3 and Figure 8-24.

Respondent Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**16**	**18**	**19**	**21**
Sample mix up during aliquotting	3	1	1	1	1	2	1	1	1	1	1	2	1
Sample degradation	2	1	1	1	1	2	1	2	3	1	1	1	2
Sample misplacement	1	1	1	1	1	1	1	1	3	1	1	1	1
Key: 1 = yes, 2 = no, 3 = not sure													

Table 8-3: Percentage of Positive Responses to "*Do you think storing sample information (including sample identifier, tracking information and temperature changes) on a tube would prevent incidents below?*"

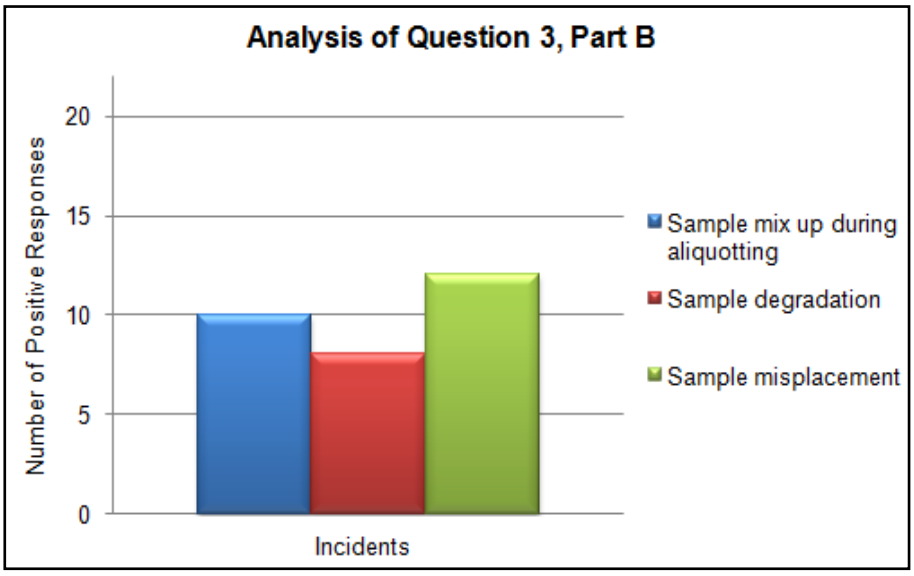


Figure 8-24: Graphical Representation of Table 8-3

8.5.4 Comments by Respondents

One of the respondents who described him/herself as a *biobank administrator/research scientist*, comments on the current method of sample identification and tracking using BIMS as *very simple, no information in relation to what happens after a sample leaves the biobank*. In line with this comment was another comment from a principal investigator, saying *look at potential for identifying and locating processed samples e.g. DNA/RNA*. This respondent also comments on electronic tagging emphasising what other potential information could be stored on the tags. Another comment also raises the idea of recording temperature information and freeze-thaw cycles to identify sample degradation.

8.6 Comparison with Other Similar Systems

The patented system for tracking biological samples (Torre-Bueno, 2007) was discussed in Section 2.6.1. The PCRC-SITS allows tracking aliquots taken from a sample and supports complicated workflows in biobanks. The system for tracking biological samples allows tracking of one unique sample using the UUID allocated to it. This system allows a complete history of samples to be maintained, however this is a limitation imposed by the PCRC-SITS prototype that the full history of records are not maintained in the database and have to be re-created from the audit trail. The

patented system does not support tracking samples through complicated workflows in multi-institutional biobanks.

The approach taken by the GHRC project (IBMT, 2009) as discussed in Section 2.6.2 is a very comprehensive hierarchical SITS in which every unit of the system is plugged in to the higher level layer in the hierarchy. The hierarchical approach would be an interesting method for identifying well-plates when implementing PCRC-SITS. Each well could be considered as an aliquot that forms the higher level of the hierarchy. However, the major limitation of the GHRC project method is adaptation by biobanks that already exist. This approach is an infrastructure that would best suit biobanks at start up stage. Another major limitation of this system is that it only operates in LN racks and freezers. This method is not applicable to biobanks using conventional freezers for storing samples due to the moisture in these freezers.

The PCRC-SITS has taken advantage of database and web technologies, similar to methods already used in many other biobanks as discussed in Section 2.6.3. However, the PCRC-SITS prototype takes advantage of RFID technology as a means of data storage on the sample. The advantages that RFID tags bring to the SITS are significant in terms of safe data storage. Biobanks such as the KI biobank (KI, 2007) and CPCTR (CPCTR, 2007) have deployed databases and barcodes to identify and track samples. The barcode is the unique identifier in these biobanks and databases are used as the storage media. The UKBiobank (UKBiobank, 2008) is using a customised LIMS and barcodes to identify samples. RFID tags are used in the Paoli Calmettes (Paoli-Calmettes, 2009) biobanks as the unique identifiers of the samples while storing partial information about the sample. The PCRC-SITS while allowing unique identification of samples, supports tracking them to their parents and vice versa by storing the identifiers of parent and child on the tag. This is a limitation imposed by the methods that other biobanks have deployed.

More details of comparison between the PCRC-SITS and other system will be provided in Section 9.3.

8.7 Summary and Discussion

8.7.1 Evaluation Summary

The PCRC-SITS prototype designed and developed as described in Chapter 7, must be evaluated in real-life practice to prove its functionality, integratablity and user acceptance. The evaluation of the PCRC-SITS prototype was carried out in the context of the PCRC biobank.

Due to the constraints imposed by various factors such as the amount of time required for a sample to be processed through all three major phases of the workflow, the delay in the commencement of secondary processing phase, and not having large numbers of donors who take part in the biobank activity for collecting samples, SITS was evaluated separately in two parts of the workflow - the initial processing phase and the secondary processing phase. The initial processing phase was carried out in a hospital and the research nurse was satisfied with the functionality and user friendliness of the SITS web application. The secondary processing phase was carried out in a research institution and by a researcher. The user who carried out the procedures found the system user-friendly and functional while satisfying their essential needs. Valuable comments on the system were made, such as the advantages that an automatic temperature monitoring system would bring to their research. It was also noticed that the fact that the well-plates can only be associated with one parent aliquot is a limitation imposed by the PCRC-SITS. Also it is necessary that space for the data for each well on the well-plate is allocated space on the tag memory.

As a fundamental requirement SITS integrateablity with the BIMS has been tested for feasibility. Data communication between the PCRC-SITS prototype and the PCRC BIMS, Distiller, has been made feasible by taking advantage of OLE Automation. With this approach data have been added to and updated on Distiller successfully. OLE Automation is the only operational approach for integration with the PCRC BIMS, although API, SOA and other methods are also available depending on the characteristics of the particular biobank BIMS. The problems that may arise with OLE Automation for PCRC BIMS include:

- Data fields in Distiller are dynamic and any changes in names or types of the fields must be reflected on the system,
- Distiller does not support direct access to a record page and therefore pages must be browsed to reach the page associated with a record. Therefore in case of any modifications to the intermediate pages, the SITS must be edited accordingly,
- the integration approach is dependent on the Win32 Application used to communicate with the RFID device. Thus in case of any interruption in that program no data will be exchanged, and
- the approach used for integration cannot be applied to any other BIMS.

Finally, to receive the opinion of users about the current approach for identifying and locating samples as well as to learn about their perceptions of electronic tagging of samples a questionnaire was distributed. The questionnaire was completed by the PCRC biobank members who had not used PCRC-SITS before and were introduced to the system through demonstrations and presentations at the semi-annual PCRC meetings. Although the size of the population is too small for drawing any statistically significant conclusions, it revealed the opinion of users of the current system and their perceptions of RFID and electronic tagging. The outcome of the questionnaire showed that most users did not find the current method of identifying and locating sample useful. However they found it easy. The outcomes also demonstrated a generally positive attitude of the PCRC biobank members towards the electronic tagging of samples. In general PCRC members are optimistic with regard to the potential of electronic tagging of samples. There have been comments about incorporating temperature monitoring sensors with electronic tagging of samples.

8.7.2 Assessment of Evaluation

The evaluation procedure consisted of three major parts, each evaluating the system from a different perspective: the experience of users using the PCRC-SITS in a real-life setting, integration of the prototype with the PCRC BIMS, and the users opinions and perceptions about the current status of BIMS, SITS and electronic tagging of samples.

Carrying out the workflows from the first processing phase to the end of the secondary processing phase in the real-life setting of the PCRC biobank and allowing the users to carry out the procedures revealed interesting results. The feedback received from the users will assist improving future versions of SITS. Being aware of the experience of users and their comments on the system, although necessary for such systems, was carried out with only two individuals. More comprehensive feedback could have been obtained if more than two individuals from each processing phases had attended the trials. However, this proved impossible to achieve in the context of the PCRC biobank. Also, evaluating the prototype in the third phase, long-term storage and retrieval, was not possible due to the lack of time. It would allow a more comprehensive evaluation if it was feasible to deploy the system for routine use.

Integration with BIMS while being a crucial feature of SITS, highly depends on the implementation of BIMS and the services it provides. However, methods for integration are discussed in the three most common scenarios. As noted above, the only choice available for PCRC integration was OLE Automation and of the three potential integration technologies it is probably the least desirable.

The analysis based on the questionnaire distributed among the PCRC biobank members was used to learn about the opinions and perception of users about BIMS, SITS and the electronic tagging of samples. Although this questionnaire revealed the opinions and perceptions of the PCRC biobank members about the subject under study, quantifying qualitative data using averages is a weak point of this analysis. Furthermore, the respondents were asked to choose their answers from a four point scale with the same semantic distances between descriptive values. It is also possible that respondents may have interpreted some of the questions differently thereby potentially affecting the results.

Chapter 9. Conclusions and Future Work

9.1 Introduction

This thesis presented a prototype SITS for biobanks, specifically the PCRC-SITS was developed for the PCRC biobank. This innovative approach proposes a system based on RFID technology that supports samples with their associated data throughout the entire workflow in multi-institutional biobanks where samples are collected, processed and stored in various locations and by different individuals. Throughout this thesis biobanks have been studied in depth along with the issues and challenges that they must overcome.

The purpose of this Chapter is to discuss the goals and objectives of this thesis and to illustrate how they have been achieved. It also identifies the contribution of this thesis to the state of the art of SITS. Finally, this Chapter concludes the thesis with a discussion of the possible future work which may be conducted to extend the research detailed in this thesis.

9.2 Objectives and Achievements

The research question posed in Chapter 1 of this thesis was *What is a more reliable and effective way of identifying and tracking biological samples in a multi-institutional biobank while (i) maintaining confidentiality of donors, and (ii) supporting samples collected longitudinally over time?*

In response to the research question, the objectives of the thesis were identified as approaches to and the development of a solution to:

(i) Mapping the sample journey and associated processes;

(ii) Providing a system of unique identification for samples;

(iii) Providing a mechanism for linking derived samples (aliquots) to their parent sample and vice versa;

(iv) Tracking sample history;

(v) Linking the sample to its associated data;

(vi) Ensuring confidentiality of donors including consenting, consent management, concealing of data to prevent identification of donors;

(vii) Providing a role based system of granting access privileges to users; and

(viii) Validating the resulting SITS in the context of the PCRC biobank.

This thesis has demonstrated that RFID currently offers a more reliable and effective way of identifying and tracking biological samples in a multi-institutional biobank while (i) maintaining confidentiality of donors, and (ii) supporting samples collected longitudinally over time. Moreover the implementation of PCRC-SITS based on RFID confirms the feasibility of this approach. It meets the requirements for robust, reliable secure sample identification and tracking. Specifically it satisfies nine identified requirements that SITS should meet. These requirements are maintenance of security and confidentiality of data, being fast, reliable and error free, facilitating item level identification of small samples and aliquots, having sufficient storage space for storing sample data in addition to the sample unique identifier, support complex workflows and procedures including omic procedures, support storage in LN, low temperatures and multiple freeze-thaw cycles, having sufficient life span to cover long-term storage for several years and facilitate storage of dynamic data at each phase of the workflow as samples go through them.

9.2.1 Mapping the sample journey

Mapping the sample journey and associated processes is of crucial importance for SITS deployment. Incorporating SITS into the procedures that samples go through requires a full understanding of the sample journey. This task is even more complex in the context of multi-institutional biobanks, where operations are spread across various locations and data must be integrated from different institutions.

Observing the sample flow through various locations, the sample journey was broken into three main phases: initial processing phase, secondary processing phase, and long-term storage and retrieval. The initial processing phase is mainly about sample

collection and preparation, while in the secondary processing phase omic procedures are carried out on samples. The SOPs and workflow documents that outline the precise operations that must be carried out during each phase of the sample journey were mapped to UML sequence diagrams and to workflow diagrams. UML sequence diagrams are detailed, while workflow diagrams are abstract and high-level. UML sequence diagrams are updated taking into account the effects of incorporating SITS. SOPs will also need to be updated to incorporate the modifications made to the process taking into account SITS activities.

9.2.2 Unique identification

In addressing the second objective of providing a system of unique identification for samples, the unique 16 alphanumerical character long RFID tag identifiers that are manufactured to the tag memory are used as the unique identifier of a tube containing samples. These unique identifiers are difficult to clone, erase from the memory or change. Samples are transferred from one tube to another during the procedures, and each tube is identified by a different tag identifier. Therefore, a tag identifier supports recognition of a particular sample at a particular point in the procedure. In the system developed for identifying samples, only tagged samples that belong to a particular biobank can be scanned, read, written and updated in that particular biobank.

9.2.3 Linking derived samples to their parents and vice versa

The third objective of providing a mechanism for linking derived samples (aliquots) to their parent sample is essential for tracking samples as they go along the procedures. An approach for linking samples to their parent and subsequent samples must be in place to allow full tracking of samples in SITS.

The mechanism for linking samples to their parent sample requires storing the identifier of each tube on the tags of its parent and subsequent tubes. In this approach the tagID that is the unique tag identifier, of the parent sample is automatically stored on its child tag, and then child tagID are stored on the parent tag. This allows a linkage between the sample aliquots at various stages of the procedure as well as between the samples that are taken from the same parent. Linking samples through their tag

identifiers enables the users to browse the aliquots or subsequent samples taken from a known sample.

9.2.4 Tracking sample history

Another objective of this thesis is to support tracking sample history and is required to provide the researchers with the information needed to assess the quality of samples on which they carry out procedures.

Tracking sample history is delivered by storing all data about the temperature changes and duration that the sample has gone through, data about the initial sample that was collected from donors and data about the intermediate aliquots in the secondary processing phase. Sample identification data, collection data, SOPs, aliquots, the procedure that sample has gone through, its parent sample, location, temperature changes and duration, are stored on the tag attached to the tube containing the sample. These data are collected for the parent of a sample as well as for the intermediate samples, and are accessible for any sample at any stage of the procedures.

9.2.5 Linking the sample to its associated data

Another functionality that SITS must deliver is linking the sample to its associated data to improve rate of data loss and reduce sample mix-up during the procedures. To address this objective, technologies that are available for identification and tracking purposes or have potential of such applications were studied in detail. The outcome of the study revealed that RFID technology has the potential to deliver such functionality through its unique tag identifier and its capability of storing dynamic data on its tags.

To link a sample to its associated data each container that a sample is stored in is tagged with a RFID tag - tubes and well-plates. Data associated with the physical sample that was stored in that particular container are recorded on the tag attached to the container. This allows storage of sample data on the tube containing the actual physical sample. Data on the tag memory can be updated by the user as changes are made to the sample. For example, exposure of the sample to a certain temperature change can be logged on its tag memory. In addition to the data being stored on the tag memory, they are also stored on a database accessible through the web. This

functionality has been provided to support users with access to the data if the sample is not available, to run queries, and to locate a particular sample.

9.2.6 Confidentiality of donors

To address the sixth objective of ensuring confidentiality of donors including consenting, consent management, concealing of data to prevent identification of donors, an integral part of the research carried out focused on reviewing state of the art approaches for confidentiality maintenance, which is a crucial prerequisite of sample identification and tracking activities.

The outcomes of the review identified various methods of concealing personal identifiable information, various types of consents and rules and regulations imposed by different countries and by the governing body of biobanks. De-identification was found to be the most suitable approach for concealing donors' data as it supports tracking of longitudinal studies and allows the locating and identifying samples in the event that consent is withdrawn. The PCRC biobank approach for concealing donors' information while maintaining a link between the donor and his/her donated samples and data, was illustrated in this thesis. In addition to informed consents collected from donors, the PCRC biobank deploys a double layer data warehouse infrastructure in which personal identifiable information are maintained in the secure local database of the collection hospital and the de-identified data is stored on the BIMS central database. As a result of the analysis carried out on various aspects of confidentiality maintenance, a guideline for best practice was developed and adherence of the PCRC biobanks to it was investigated.

9.2.7 Role based system

The intended environment for SITS that is described in this thesis is a multi-institutional biobank that involves a number of physically distributed individuals collecting, processing and using samples and data. Thus, providing a role based system of granting access privileges to users is recognised as an important objective of this thesis that must be achieved as a requirement to ensuring donor confidentiality and security.

Five main categories of users were identified in multi-institutional biobanks, and based on their needs, they could be granted access rights to the system. Administrator, researcher, research nurse, guest and root users are the roles defined for accessing SITS. Each role can only access the data about the sample that is within the domain of his/her operations. These operations include adding, updating, deleting, and viewing initial samples from the initial processing phase and derived samples (aliquots) from the secondary processing phase. While the research nurse role deals with samples collected in the initial processing phase and have no other interaction with the aliquots, this role is not allowed to carry out any operation on the aliquots data. On the other hand, researchers need to view data about samples collected in the initial processing phase in addition to full access rights to the data about aliquots in the secondary processing phase. The administrator account is allocated to the PIs who are granted full access rights to all types of samples. A guest account is also defined for guest users who are allocated temporary access to the system. The root user is at a higher level to administrator account holders and there is only one root account holder. The root user can delete data in the log file of the system.

9.2.8 Validating the resulting SITS in the context of the PCRC biobank

Finally, in order to meet the objective of validating the resulting SITS in the context of the PCRC biobank the outcome of the research carried out in this thesis was evaluated in the context of the Irish PCRC biobank, a multi-institutional biobank. A SITS prototype, PCRC-SITS, was developed to fulfil this objective.

Based on the requirements a web-based SITS prototype was designed and implemented for the PCRC biobank using RFID technology. The RFID tags were tested against extreme temperatures and number of freeze-thaw cycles by storing them at -80°C freezers and in -190°C LN vapour tanks. The prototype was evaluated in real-life practice by two users, a research nurse and a researcher. These users carried out initial and secondary processing phases using SITS, and were satisfied with the system. The feasibility of integrating SITS in BIMS was investigated and proved by integrating the PCRC-SITS into the PCRC BIMS using OLE Automation. Constraints imposed by the PCRC BIMS limited the functionality of the integration. Finally, the opinions and perceptions of users about the method used in the PCRC biobank prior to

the introduction of the SITS prototype and of the electronic tagging samples were analysed through a questionnaire.

9.3 Contribution to the State of the Art

A novel approach for technology-based sample identification and tracking in multi-institutional biobanks is the primary contribution to the state of the art made by this thesis and the research described therein. This approach is significantly different to those already deployed in biobanks.

The patented system for tracking biological samples (Torre-Bueno, 2007) that was described in Section 2.6.1, offers unique identifiers for samples that can be read by input equipment and offers a workflow for processing samples. Input devices in this system may be any equipment or method of data entry to the system. Although there are a number of issues with the input equipment, the major drawback of this system is that it is not applicable to biobanks. The system does not support biobank activities that involve complicated workflows that samples go through, long-term storage of samples, and exchange of samples and data across various locations which must guarantee the security and confidentiality of donors.

A more comprehensive approach to sample identification and tracking is offered as part of the GHRC project (IBMT, 2009). Although in this system storage of dynamic data on the tubes containing the sample is facilitated, there are a number of issues that limit its operations and adoption by other biobanks. The system supports identification, location and documentation of samples in LN racks that are dry and no frost is involved. However in many biobanks, laboratory freezers are used in which moisture and frost are to be expected. This system cannot operate in wet laboratories as this will cause problems when plugging the tubes in to the racks or higher level structure of the hierarchy. The hierarchical system provided by GHRC is very detailed and highly structured and would be virtually impossible to incorporate into an existing biobanks. In the case of new biobanks, the GHRC approach could be a very expensive option. An interesting point with this system is that in addition to having Flash-memory implemented in the tubes, barcodes and RFID tags are also attached to the tubes. Using RFID tags with larger memory sizes will assist in preventing the need for

having Flash-memory chips and will also improve the system by supporting operation in ordinary freezers. Unlike the Flash-memory chips, RFID tags are not required to be plugged in to any other devices and can be read from a distance. In that case the system could be implemented without the need for making modifications to the racks or freezers. By contrast, the SITS described in this thesis requires little modification to the infrastructure of the biobank and can be incorporated into existing biobanks relatively easily.

Biobanks such as the KI biobank (KI, 2007) and the CPCTR (CPCTR, 2007) take advantage of using databases and barcodes to identify and track samples within their biobanks. Also, the UKBiobank (UKBiobank, 2008) has a customised LIMS system for storing, managing and integrating data, along with automated sample handling that reduces the need for human interaction with the samples. This biobank also uses barcodes for identifying samples. Examples of the shortcomings that barcodes entail include the need for a line-of-sight when reading labels, supplying static data that cannot be updated, and the need for being supported by a database system which houses dynamic data. With the approach illustrated throughout this thesis, dynamic data are stored on the tags attached to the tubes containing samples. The fundamental and foremost advantage that RFID offers to SITS that cannot be achieved by barcodes is the ability to store dynamic sample information with the actual physical sample. This increases the robustness of the system and allows many operations to be carried out without the need to access the database.

The Paoli Calmettes (Paoli-Calmettes, 2009) biobank has taken advantage of RFID for storing data on the tubes containing the sample. Data stored on the RFID tags are protected by unique identifiers and passwords. The limitation of this system in comparison to the SITS described in this thesis is that only partial sample data are stored on the tag. Workflows are not implemented and tracking information is not maintained.

The influence of this research on the state of the art is reflected by its direct contribution to the publications in an edited book chapter, presentations and proceedings of national and international conferences.

9.4 Lessons Learned

The research described in this thesis has identified several key lessons in the development of robust and reliable methods for sample identification and tracking. In particular, two areas are highlighted as being particularly significant. The first relates to the importance of ensuring the privacy and confidentiality of donors. If biobanking is to reach its full potential in continuing to the development of personalised medicine, then potential donors must have confidence that their information will be kept secure and that it and their samples will be used only for the purposes for which consent has been given. The twin issues of consent and confidentiality must be considered as if fundamental importance not just in order to receive ethical approval at the start but also when designing and implementing the SITS all the way through to long-term stage.

The second important lesson to be learned from this research relates to maturity and robustness of RFID technology. While the technology may only be at the start of the *Slope of Enlightenment* according to Gartner Hype Cycle (see Figure 4-11) it is capable of handling the complex and demanding environment of sample identification and tracking in biobank. The tags are capable of very low temperatures and successive freeze-thaw cycles. They are small enough to fit on the lids of sample tubes and can store sufficient information to meet the requirements of SITS.

9.5 Future Work

The research carried out in this thesis can be extended and advanced in a number of potential areas. The PCRC-SITS can be extended to cover extra functionalities such as maintaining full history of records on the SITS database, identification of individual wells in a well-plate and customisable roles. One of these areas is integration with sensors such as temperature sensors and temperature loggers as identified by users during the evaluation of the prototype. Also security risk assessment of SITS is another area of extension to this research. With the system illustrated in this research, more comprehensive and standardised networked biobanks can be developed as SITS will assist in standardised collection and processing that will lead to collection and maintenance of comparable data.

9.5.1 PCRC-SITS Added Functionalities

Evaluation of the PCRC-SITS in real setting revealed the need of certain functionalities for full deployment of the system. Functionalities such as being able to identify each well in well-plates and associating records with each individual well need to be taken into consideration. This can be done by allocating storage space for data associated with each well on the RFID tag and on the SITS database. Since the tags have limited memory and there are usually about 96 wells in each well-plate, only partial data can be stored on the tag and full record can be maintained on the SITS database. One tag can be attached to each well-plate and then the row and column numbers can be used to identify each well on the well-plate.

Maintaining a full history of records on the SITS database needs to be facilitated for SITS. The full history of records is now maintained by the audit trail, however this functionality is needed to be added to the SITS database as well.

SITS needs to facilitate customisable roles for accessing data, such that an individual can access certain data used for a certain research project regardless of his/her location or the administrator can decide the level of access granted to each individual. The system administrator role should be allocated carefully and guest accounts need to have much limited rights.

9.5.2 Integration with Temperature Sensors and Temperature Loggers

From the results of the questionnaire one can note the attention of users to monitoring temperatures that samples are exposed to during procedures and transit across locations. Temperature changes or climate changes, to include humidity, of the sample storage environment play an important role in the quality of the sample and its degradation process.

Monitoring the temperature of each aliquot will require RFID tags be either integrated with temperature sensors, or in case of not using RFID tags, will require temperature sensors that log temperature changes to be attached to the tube. These temperature sensors that are capable of keeping a log are called temperature loggers. An issue that rises is the size of the sensors that should be small enough to fit the body or the lid of

the tube. Although, the need to have such sensors is reduced by freezers alarm system that take care of temperature changes while samples are within the biobank freezer, monitoring temperature changes and duration while transferring samples from one location to another is still a valuable resource to researchers. The freezers alarm system does not facilitate monitoring of temperature at sample level. The PCRC biobank storage freezers are equipped with alarms that are triggered under certain circumstances.

In the PCRC biobank workflow the highest risk of samples being exposed to temperature changes is when they are being transferred from one institute to another. Aliquots are transferred from a location to another in boxes of ice; however the duration of transit plays an important role on the temperature and quality of the environment. Logging temperatures while the sample is in transit can be carried out by deploying temperature loggers. There are a number of products that can be chosen for monitoring temperature (DeltaT, 2009, Microsensys, 2009). The characteristics that are important when choosing a product depend on the following:

- The temperature range that the sensor operates at and survives
- The resolution of the sensor is important to measure its specificity
- Size is also important especially when using small boxes
- The reader that will be used to read the sensor's data. The best scenario would be to have a reader that is capable of also reading the RFID tags or barcodes that are used for SITS.

The comments from the researcher who was testing the system for the secondary process phase and the comment made on the questionnaire revealed the vital need of integrating SITS with temperature sensors or temperature loggers. Integrating sensors with the proposed SITS will require modification to the SITS Win32 Controller to allow extraction of certain data that are logged and then uploading those data to the "Temperature Changes and Duration" fields. In the case of additional sensor data e.g. humidity data, related data fields should be created on the database and the Win32 Application will then need to be modified to allow transfer of those data to the RFID tag memory. Also related pages, depending on whether an initial sample or a derived sample is involved will need to be updated.

9.5.3 SITS Security

Security and privacy concerns regarding the use of RFID technology were discussed in Sections 4.4.4 and 4.4.5. In addition, the approach taken by the current PCRC BIMS to ensure security and confidentiality was presented in Section 7.3. However, it is necessary to conduct a detailed security risk assessment for the PCRC-SITS. For example, it would be necessary to ensure that the SITS database and the associated tag data conform to the requirements of Data Protection Commissioner.

It would be preferable to deploy Client Certificates for pre-identified IP addresses for access to the SITS web interface. Other security risks particularly in relation to data transfer require detailed assessment.

9.5.4 Network of Biobanks

Developing a federated network of biobanks represents the first step towards worldwide sample collection, and linking, integrating or merging BIMS from different biobanks is the next challenge if we are to fully harness the value of such biobanks (Ölund et al., 2007). A network of biobanks will allow exchange of large amount of data and samples from a variety of biological and geographical backgrounds and hence more complete collections of samples and data will be possible. This offers the potential to accelerate knowledge discovery significantly by greatly increasing the study population size for both discovery and most importantly validation of novel biomarkers which is required before clinical utilisation and impacting on participant care. Multi-institutional biobanks can be considered as the first step towards a network of biobanks where standardisation becomes a major challenge.

The research reported in this thesis forms a sound basis for the development of a sample identification and tracking system for such a network.

References

ACTIVEBARCODE (2008a) *Barcode types*, http://www.activebarcode.com/codes/, accessed on 22/12/2008.

ACTIVEBARCODE (2008b) *Data Matrix*, http://www.activebarcode.com/codes/datamatrix.html, accessed on 27/12/2008.

ACTIVEBARCODE (2008c) *EAN-2 Addon / EAN-5 Addon*, http://www.activebarcode.com/codes/ean5_ean2.html, accessed on 22/12/2008.

ACTIVEBARCODE (2008d) *EAN-13*, http://www.activebarcode.com/codes/ean13.html, accessed on 22/12/2008.

ACTIVEBARCODE (2008e) *ITF-14 / SCC-14*, http://www.activebarcode.com/codes/itf14.html, accessed on 22/12/2008.

ACTIVEBARCODE (2008f) *UPC-A / UPC-E*, http://www.activebarcode.com/codes/upca_upce.html, accessed on 22/12/2008.

ADAMS, G. (2007) Pharmaceutical manufacturing: RFID - reducing errors and effort. *Filtration & Separation,* 44**,** 17-19.

ADAMS, R. (2008) *Specifications For Popular 2D Bar Codes*, Adams Communications, http://www.adams1.com/stack.html, accessed on 27/12/2008.

ADLER, R., FLANIGAN, M., HUANG, J., KLING, R., KUSHALNAGAR, N., NACHMAN, L., WAN, C.-Y. & YARVIS, M. (2005) Intel Mote 2: an advanced platform for demanding sensor network applications. *SenSys '05: Proceedings of the 3rd international conference on Embedded networked sensor systems.* San Diego, California, USA, ACM.

AIMGLOBAL (2008) *Two dimensional (2D) Bar Code Symbologies*, http://www.aimglobal.org/technologies/barcode/2d_symbologies.asp#stacked, accessed on 22/12/2008.

ALLABOUTGEMSTONES (2009) *The Diamond Trade: The 'Diamond Pipeline'*, AllAboutGemstones.com, http://www.allaboutgemstones.com/diamond_pipeline.html, accessed on 01/05/2009.

ALMUDEVAR, A., LEIBOVICI, A. & TENTLER, A. (2008) Home monitoring using wearable radio frequency transmitters. *Artificial Intelligence in Medicine,* 42**,** 109-120.

ALTO, P. (2006) *HP Unveils Revolutionary Wireless Chip that Links the Digital and Physical Worlds*, Hewlett-Packard Development Company,

http://www.hp.com/hpinfo/newsroom/press/2006/060717a.html, accessed on 28/09/2009.

ANDERSON, N. L. & ANDERSON, N. G. (1998) Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*, 19**,** 1853-1861.

ASHG *American Society of Human Genetics*, http://www.ashg.org/, accessed on 04/11/2008.

AYOADE, J. (2006) Security implications in RFID and authentication processing framework. *Computers & Security*, 25**,** 207-212.

AYOADE, J. (2007) Roadmap to solving security and privacy concerns in RFID systems. *Computer Law & Security Report*, 23**,** 555-561.

AZURA *AZURA*, AZURA Group, http://www.azuragroup.com/, accessed on 27/08/2009.

BACHELDOR, B. (2007) *N.J. Medical Center Uses LF Tags To Protect Patient Records*, http://www.rfidsolutionsonline.com/article.mvc/NJ-Medical-Center-Uses-LF-Tags-To-Protect-Pat-0001, accessed on 13/10/2008.

BACHELDOR, B. (2008) *Magellan Creates Specimen-Tracking System*, RFID Journal, http://www.rfidjournal.com/article/articleview/3892/1/1/, accessed on 12/11/2008.

BAJIC, E. & CHAXEL, F. (2002) Auto-ID mobile information system for vehicle life cycle data management. *IEEE International Conference on Systems, Man and Cybernetics.* Tunisia, IEEE.

BALL, E., CHADWICK, D. W. & MUNDY, D. (2003) Patient privacy in electronic prescription transfer. *Security & Privacy, IEEE,* 1**,** 77-80.

BASHAN, O. (2003) *ISO 14443 - An introduction to the contactless standard for smart cards and its relevance to customers*, OTI - On Track Innovations LTD., http://www.otiglobal.com/objects/ISO%2014443%20WP%204.11.pdf, accessed on 26/10/2009.

BBMRI *Biobanking and Biomolecular Resources Research Infrastructure* Biobanking and Biomolecular Resources Research Infrastructure http://www.bbmri.eu/, accessed on 26/09/2009.

BESKOW, A. (2009) The Swedish Biobank Model and the Biobank Organization at Karolinska Institutet. *Clinical Chemistry and Laboratory Medicine,* 47**,** S19.

BETSOU, F., FERGUSON, M., JALLAL, B. & LITTON, J.-E. (2004) *BIOBANK: Accelerating Molecular Medicine, Challenges Facing the Global Biobanking Community*, IBM, http://www-03.ibm.com/industries/global/files/Biobanks_Accelerating_Molecular_Medicine.pdf, accessed on 26/10/2009.

BETTENDORF, E., MALENFANT, C. & CHABANNON, C. (2005) RFID Technology and Electronic Tags to Identify Cryopreserved Materials. *Cell Preservation Technology,* 3**,** 112-115.

BIRF (2009) *Brain Injury Glossary*, The Brain Injury Resource Foundation, http://www.birf.info/home/bi-tools/qlinks_c.html, accessed on 25/10/2009.

BOTTANI, E. & RIZZI, A. (2008) Economical assessment of the impact of RFID technology and EPC system on the fast-moving consumer goods supply chain. *International Journal of Production Economics,* 112**,** 548-569.

BOUKERCHE, A. (2005) *Handbook of algorithms for wireless networking and mobile computing*, Chapman & Hall.

BRITISH MEDICAL ASSOCIATION (2005) *Confidentiality as part of a bigger picture*, British Medical Association, http://www.bma.org.uk/images/confidentiality_tcm41-20480.pdf, accessed on 24/02/2009.

BROWN, I. & RUSSELL, J. (2007) Radio frequency identification technology: An exploratory study on adoption in the South African retail sector. *International Journal of Information Management,* 27**,** 250-265.

BURNETT, M. & KLEIMAN, D. (2005) Ten Password Pointers: Building Strong Passwords. *Perfect Password: Selection, Protection, Authentication* Burlington, Syngress.

BYLUND, J. & JIRHEDE, F. (2003) Visual Barcode Reading. *Communications.* Stockholm, Sweden, KTH Royal Institute of Technology.

CANCERLINEUK (2009a) *CancerlineUK* AstraZeneca, http://www.cancerlineuk.net/, accessed on 26/10/2009.

CANCERLINEUK (2009b) *Glossary*, AstraZeneca, http://www.cancerlineuk.net/517210/?selectedPage=2&letter=I, accessed on 26/10/2009.

CASPIAN (2004) *Consumers Against Supermarket Privacy Invasion And Numbering* CASPIAN, http://www.nocards.org/, accessed on 08/08/2009.

CATCHPOOLE , D., DEFAZIO, A., DEVEREUX, L., FLEMING, M., HOF, M., SCHMIDT, C., THORNE, H. & NIKOLAS ZEPS, N. (2007) *Biorepository Protocols*, Australasian Biospecimen Network, http://www.abrn.net/pdf/ABN_SOPs_Review_Mar07_final.pdf, accessed on 26/10/2009.

CHAN, J. C., CHU, R. W., YOUNG, B. W., CHAN, F., CHOW, C. C., PANG, W. C., CHAN, C., YEUNG, S. H., CHOW, P. K., LAU, J. & LEUNG, P. M. (2004) Use of an electronic barcode system for patient identification during blood

transfusion: 3-year experience in a regional hospital. *Hong Kong Med J,* 10**,** 166-71.

CHOW, H. K. H., CHOY, K. L., LEE, W. B. & LAU, K. C. (2006) Design of a RFID case-based resource management system for warehouse operations. *Expert Systems with Applications,* 30**,** 561-576.

COLLIN, S. M. H. (2004) *Dictionary of Computing*, Bloomsbury Publishing Plc.

COMPTON, C., DESCHÊNES, M., FURRER, B., GAUCH, S., GRIZZLE, W. & POTTER, J. D. (2005) *Biobanks: Strategies to Address Next-Generation Challenges for Biobanking*, IBM, http://www-03.ibm.com/industries/global/files/StrategiestoAddressNextGenerationChallengesforBiobanking.pdf?re=healthcare&sa_message=title=biobanks:_strategies_to_address_next-generation_challenges_for_biobanking, accessed on 26/10/2009.

CORCHADO, J. M., BAJO, J., DE PAZ, Y. & TAPIA, D. I. (2008) Intelligent environment for monitoring Alzheimer patients, agent technology for health care. *Decision Support Systems,* 44**,** 382-396.

COYLE, K. (2005) Management of RFID in Libraries. *The Journal of Academic Librarianship,* 31**,** 486-489.

CPCTR (2007) *Cooperative Prostate Cancer Tissue Resource*, Cooperative Prostate Cancer Tissue Resource, http://www.cpctr.info/, accessed on 12/12/2008.

DATA PROTECTION COMMISSIONER IRELAND (2007) *The Data Protection Rules*, http://www.dataprotection.ie/ViewDoc.asp?DocId=21, accessed on 26/11/2007.

DATTA, M. (2003) *Instructions for CPCTR TMA Block Production*, University of Pittsburgh School of Medicine Department of Pathology, http://www.pathology.pitt.edu/pdf/cpctr/block.htm, accessed on 26/10/2009.

DAVIS, F. D. (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of lnformation Technology. *MIS Quatterly,* 13**,** 318-340.

DAVISS, B. (2005) Growing pains for metabolomics. *The Scientist,* 19**,** 25.

DECODE *deCODE*, deCODE, http://www.decode.com/, accessed on 15/09/2009.

DELTAT (2009) *Safe and efficient solutions for temperature stable transportation*, DeltaT, http://www.deltat.de/products/dataloggers.html, accessed on 26/10/2009.

DIRECTBARCODES *Barcode Symbology*, http://www.directbarcodes.com/page1.php, accessed on 22/12/2008.

DLSOFT (2009) *GS1 Composite symbols*, DL Technology Ltd, http://www.dlsoft.com/barcode_types/gs1_composite_barcode_symbols.htm, accessed on 26/10/2009.

DOBSON, T. & TODD, E. (2006) Radio frequency identification technology. *Computer Law & Security Report,* 22**,** 313-315.

DODEK, D. Y. & DODEK, A. (1997) From Hippocrates to facsimile: Protecting patient confidentiality is more difficult and more important than ever before. *CMAJ,* 156**,** 847-852.

DOMDOUZIS, K., KUMAR, B. & ANUMBA, C. (2007) Radio-Frequency Identification (RFID) applications: A brief introduction. *Advanced Engineering Informatics,* 21**,** 350-355.

DURST, C. H. P., IHMIG, F. R., BIEL, M., DAFFERTSHOFER, M., ZIMMERMANN, H. & KRYOBIOPHYSIK, A. (2005) XML based Process Management in Cryo-Biotechnology. *GI-Workshop XML4BPM - XML for Business Process Management.* Karlsruhe, Germany.

DURST, C. H. P., IHMIG, F. R., FUCHS, C. C., BIEL, M., DAFFERTSHOFER, M. & ZIMMERMANN, H. (2006) A Method and Infrastructure for Long-Term Managing of Sample Preparation Knowledge for Cryo-Biomedical Applications. *I-KNOW.* Graz, Austria.

DURST, C. H. P., IHMIG, F. R., SHIRLEY, S. G. & ZIMMERMANN, H. (2007) Management, Interchange and Reproducible Execution of Sample Preparation Knowledge in Collaborative Research Scenarios. *I-KNOW.* Graz, Austria.

EDWARDS, C. (2008) *Temperature control*, The Institution of Engineering and Technology, http://kn.theiet.org/magazine/issues/0813/temp-control0813.cfm accessed on 26/10/2009.

EGP *Estonian Genome Project*, Tartu, Estonian Genome Project, http://www.geenivaramu.ee/index.php?lang=eng, accessed on 15/09/2009.

EISEMAN, E., BLOOM, G., BROWER, J., CLANCY, N. & OLMSTED, S. S. (2003) *Case Studies of Existing Human Tissue Repositories "Best Practices" for a Biospecimen Resource for the Genomic and Proteomic Era*, RAND Corporation.

ELECTRONIC PRIVACY INFORMATION CENTER (EPIC) Comments submitted in consideration of the Article 29 Data Protection Working Party "Working Document on Data Protection Issues related to RFID Technology".

ELLIOTT, P. & PEAKMAN, T. C. (2008) The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *International Journal of Epidemiology,* 37**,** 234-44.

EMEA (2002) Position paper on terminology in pharmacogenetics.

EMEA (2009) *European Agency for the Evaluation of Medicinal Products*, EMEA, www.emea.europa.eu, accessed on 26/10/2009.

EPIC (2009) *Electronic Privacy Information Center* EPIC, http://epic.org/, accessed on 08/08/2009.

EUROPEAN TELECOMMUNICATIONS STANDARD INSTITUTE (2009) *Telecommunications Standard Institute*, www.etsi.org, accessed on 26/10/2009.

FARAHANI, S. (2008) *ZigBee Wireless Networks and Transceivers*, Newnes.

FDA (2000) CFR - Part 11 Code of Federal Regulations Title 21. IN DEPARTMENT OF HEALTH AND HUMAN SERVICES, U. (Ed.), U.S. Food and Drug Administation.

FENN, J. & LINDEN, A. (2005) *Gartner's Hype Cycle Special Report for 2005*, Gartner, Inc., http://www.gartner.com/resources/130100/130115/gartners_hype_c.pdf, accessed on 26/10/2009.

FENN, J. & RASKINO, M. (2009) *Understanding Gartner's Hype Cycles, 2009*, Gartner, Inc., http://www.gartner.com/resources/169500/169518/understanding_gartners_hype__169518.pdf, accessed on 26/10/2009.

FIELDING, R. T. (2000) Architectural Styles and the Design of Network-based Software Architectures. *Information and Computer Science.* Irvine, University of California.

FISHER, J. A. & MONAHAN, T. (2008) Tracking the social dimensions of RFID systems in hospitals. *International Journal of Medical Informatics,* 77**,** 176-183.

FLETCHER, A. (2004) Field of genes: the politics of science and identity in the Estonian Genome Project. *New Genetics and Society* 23**,** 3-14.

FLINT, D. (2006) RFID tags, security and the individual. *Computer Law & Security Report,* 22**,** 165-168.

FRANK HEALTHCARE (2008) *EDTA Blood Collection Tube*, Frank Healthcare Co., http://www.tootoo.com/d-rp13606982-EDTA_Blood_Collection_Tube/, accessed on 19/10/2009.

FRAZELLE, E. (2001) Implementing logistics systems. *Supply chain strategy: the logistics of supply chain management.* McGraw-Hill.

FURLANI, K. M. & STONE, W. C. (1999) Architecture for discrete construction component tracking. *16th IAARC/IFAC/IEEE International Symposium on Automation and Robotics in Construction.* Spain.

GARFINKEL, S. L. (2002) *An RFID Bill of Rights*, www.technologyreview.com/articles/02/10/garfinkel1002.asp, accessed on 26/10/2009.

GARFINKEL, S. L., JUELS, A. & PAPPU, R. (2005) RFID privacy: an overview of problems and proposed solutions. *IEEE Security & Privacy,* 3**,** 34-43.

GARTNER (2009a) *Gartner Hype Cycles*, Gartner Inc., http://www.gartner.com/it/products/hc/hc.jsp, accessed on 19/09/2009.

GARTNER (2009b) *Hype Cycle for Emerging Technologies*, Gartner Inc., http://www.gartner.com/DisplayDocument?id=1085912&ref=g_sitelink, accessed on 19/09/2009.

GENOMEUTWIN (2009) *GenomEUtwin*, GenomEUtwin, http://www.genomeutwin.org/, accessed on 15/09/2009.

GENUTH, I. (2006) *HP's Memory Spot Chip is Spot On*, The Future of Things, http://thefutureofthings.com/articles.php?itemId=27/59/, accessed on 28/09/2009.

GLOBAL HIV VACCINE ENTERPRISE (2009) *Global HIV Vaccine Enterprise*, Fraunhofer, http://www.hivvaccineenterprise.org/index.aspx, accessed on 26/09/2009.

GLOVER, B. & BHATT, H. (2006) *RFID Essentials*, O'Reilly Media.

GODARD, B., SCHMIDTKE, J., CASSIMAN, J.-J. & AYME, S. (2003) Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. *European Journal of Human Genetics,* 11**,** S88-S122.

GOODRUM, P. M., MCLAREN, M. A. & DURFEE, A. (2006) The application of active radio frequency identification technology for tool tracking on construction job sites. *Automation in Construction,* 15**,** 292-302.

GOU, Q., LIANG, L., HUANG, Z. & XU, C. (2008) A joint inventory model for an open-loop reverse supply chain. *International Journal of Production Economics,* 116**,** 28-42.

GRAPHLOGIC (2009) *GraphLogic Proven Solutions*, GraphLogic Inc., www.graphlogic.com, accessed on 26/10/2009.

GRAUER, Z. (2003) Laboratory Information Management Systems and Traceability of Quality Systems. *American Laboratory,* 35**,** 15-18.

GRAVES, S. (2002) An Automated System for Processing Sample Vials Stored at Ultra-Low Temperatures. *Journal of the Association for Laboratory Automation,* 7**,** 32-35.

GRAY, J. (2008) *Magellan Technology tracks diamonds for New York Diamond Merchant*, http://www.magellan-rfid.com/news/magellantechnologytracksdiamondsfornewyorkdiamondmerchant, accessed on 02/05/2009.

GROOVER, M. P. (2007) Automatic Identification and Data Capture. *Automation, Production Systems, and Computer-integrated Manufacturing.* 3 ed., Prentice Hall.

GS1-128 (2008) *GS1-128 (formerly UCC/EAN-128) Barcodes*, Bar Code Graphics, Inc., http://www.gs1-128.info/, accessed on 22/12/2008.

GS1 (2008a) *Composite Component*, http://barcodes.gs1us.org/dnn_bcec/Standards/Barcodes/CompositeComponentSymbols/tabid/181/Default.aspx, accessed on 22/12/2008.

GS1 (2008b) *GS1*, http://www.gs1.org/, accessed on 19/12/2008.

GS1 UK (2009) *Case Study: MayDay Healthcare NHS Trust Electronic Blood Tracking pilot*, GS1 UK, http://www.gs1uk.org/downloads/about/Case%20Studies/Case%20study_Mayday.pdf, accessed on 26/10/2009.

HANSSON, M. G. (2005) Building on relationships of trust in biobank research. *Journal of Medical Ethics,* 31**,** 415-418.

HANSSON, M. G. (2006) Combining efficiency and concerns about integrity when using human biobanks. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences,* 37**,** 520-532.

HANSSON, M. G., DILLNER, J., BARTRAM, C. R., CARLSON, J. A. & HELGESSON, G. (2006) Should donors be allowed to give broad consent to future biobank research? *The Lancet Oncology,* 7**,** 266-269.

HAWKES, B. (2007) *Data Protection Guidelines on research in the Health Sector*, Data Protection Commissioner, http://www.dataprotection.ie/documents/guidance/Health_research.pdf, accessed on 24/02/2009.

HIBCCAU (2005) *Tagsys*, http://www.hibcc-au.com.au/RFID%20Website/Tagsys_Page.htm, accessed on 27/12/2008.

HITACHI (2009) *Hitachi*, Hitachi, www.hitachi.co.jp, accessed on 12/08/2009.

HO, L., MOH, M., WALKER, Z., HAMADA, T. & SU, C. F. (2005) A Prototype on RFID and Sensor Networks for Elder Healthcare: Progress Report. *ACM SIGCOMM workshop on Experimental approaches to wireless network design and analysis.* USA, ACM.

HOEYER, K., OLOFSSON, B.-O., MJORNDAL, T. & LYNOE, N. (2004) Informed consent and biobanks: a population-based study of attitudes towards tissue donation for genetic research. *Scandinavian Journal of Public Health,* 32**,** 224-229.

HOLLAND, N. T., PFLEGER, L., BERGER, E., HO, A. & BASTAKI, M. (2005) Molecular epidemiology biomarkers--Sample collection and processing considerations. *Toxicology and Applied Pharmacology,* 206**,** 261-268.

HOLLAND, N. T., SMITH, M. T., ESKENAZI, B. & BASTAKI, M. (2003) Biological sample collection and processing for molecular epidemiological studies. *Mutation Research/Reviews in Mutation Research,* 543**,** 217-234.

HUANG, C.-L., CHUNG, P.-C., TSAI, M.-H., YANG, Y.-K. & HSU, Y.-C. (2008) Reliability Improvement for an RFID-based Psychiatric Patient Localization System. *Computer Communications,* 31**,** 2039-2048.

IA SORENSEN, T., BELL, J., HUTTUNEN, J., PIKANI, J. & PRINGLE, M. (2005) The International Evaluation of Swedish Biobanks.

IBMT (2009) *Global HIV Vaccine Research Cryorepository,* Fraunhofer, www.ghrc.info/, accessed on 26/09/2009.

IEEE (1998) *IEEE Std 830-1998 IEEE Recommended Practice for Software Requirements Specifications*, IEEE, http://www.cis.njit.edu/~campbell/CIS673/IEEE.pdf, accessed on 26/10/2009.

IHMIG, F. R., SHIRLEY, S. G., DURST, C. H. P., SCHULZ, J. C., BRIESEN, H. V. & ZIMMERMANN, H. (2009) The technology of the Global HIV Vaccine Research Cryorepository. *Engineering in Life Sciences,* 9**,** 1-8.

IHMIG, F. R., SHIRLEY, S. G., DURST, C. H. P. & ZIMMERMANN, H. (2006) Cryogenic electronic memory infrastructure for physically related "continuity of care records" of frozen cells. *Cryogenics,* 46**,** 312-320.

INTEL CORPORATION *Intel Motes Overview*, Intel Corporation, http://techresearch.intel.com/articles/Exploratory/1503.htm, accessed on 27/09/2009.

ISABELLE, M., TEODOROVIC, I., MORENTE, M. M., JAMINE, D., PASSIOUKOV, A., LEJEUNE, S., THERASSE, P., DINJENS, W. N. M., OOSTERHUIS, J. W., LAM, K. H., OOMEN, M. H. A., SPATZ, A., RATCLIFFE, C., KNOX, K., MAGER, R., KERR, D., PEZZELLA, F., VAN DE VIJVER, M., VAN BOVEN, H., ALONSO, S., KERJASCHKI, D., PAMMER, J., LOPEZ-GUERRERO, J. A., LLOMBART BOSCH, A., CARBONE, A., GLOGHINI, A., VAN VEEN, E. B., VAN DAMME, B. & RIEGMAN, P. H. J. (2006) TuBaFrost 5: Multifunctional central database application for a European tumor bank. *European Journal of Cancer,* 42**,** 3103-3109.

ISO (1999a) *ISO/IEC FCD 15693-2*, ISO/IEC, http://www.waazaa.org/download/fcd-15693-2.pdf, accessed on 26/10/2009.

ISO (1999b) *ISO/IEC FINAL COMMITTEE DRAFT 15693-1*, ISO/IEC, http://www.waazaa.org/download/fcd-15693-1.pdf, accessed on 26/10/2009.

ISO (2000) *ISO/IEC FCD 15693-3*, ISO/IEC, http://www.waazaa.org/download/fcd-15693-3.pdf, accessed on 26/10/2009.

JACOBY, D. (2006) RFID comes of age. IN LOFTHOUSE, G. (Ed.) *The Economist Intelligence Unit.* London, Economist Intelligence Unit.

JASELSKIS, E. J., ANDERSON, M. R., JAHREN, C. T., RODRIGUEZ, Y. & NJOS, S. (1995) Radio-Frequency Identification Applications in Construction Industry. *Journal of Construction Engineering and Management,* 121**,** 189-196.

JOHNSTON, C. & KAYE, J. (2004) Does the UK Biobank have a Legal Obligation to Feedback Individual Findings to Participants? *Medical Law Review,* 12**,** 239-267.

JUELS, A., RIVEST, R. L. & SZYDLO, M. (2003) The Blocker Tag: Selective Blocking of RFID Tags for Consumer Privacy. *Proceedings of the 10th ACM Conference on Computer and Communications Security.* USA, ACM Press.

KAISER, J. (2002) BIOBANKS: Population Databases Boom, From Iceland to the U.S. *Science,* 298**,** 1158-1161.

KAROLINSKA INSTITUTET BIOBANK *KI BIMS*, Karolinska Institutet biobank,, http://www.kibiobank.se/wps/portal, accessed on 08/10/2009.

KERRY, J. P., O'GRADY, M. N. & HOGAN, S. A. (2006) Past, current and potential utilisation of active and intelligent packaging systems for meat and muscle-based products: A review. *Meat Science,* 74**,** 113-130.

KEUCHEL*, J., SHOLTZ, L., MARION, N. & RUPP, M. E. (2004) Saving Dollars through Innovative Strategies in Biohazardous Waste Reduction. *American Journal of Infection Control,* 32**,** E32-E33.

KI (2007) *Karolinska Institutet*, Karolinska Institutet, http://www.ki.se/kibiobank, accessed on 10/12/2008.

KIMBERLEYPROCESS.COM (2009a) *Frequently Asked Questions*, kimberleyprocess.com, http://www.kimberleyprocess.com/faqs/index_en.html, accessed on 01/05/2009.

KIMBERLEYPROCESS.COM (2009b) *What is the Kimberley Process?*, 2009, http://www.kimberleyprocess.com/, accessed on 01/05/2009.

KLING, R., ADLER, R., HUANG, J., HUMMEL, V. & NACHMAN, L. (2004) Intel Mote: using Bluetooth in sensor networks. *Proceedings of the 2nd international conference on Embedded networked sensor systems.* USA, ACM.

KNIGHT, W. (2006) RFID - another technology, another security mess? *Infosecurity Today,* 3**,** 35-37.

KNOSPE, H. & POHL, H. (2004) RFID security. *Information Security Technical Report,* 9**,** 39-50.

KRASOVEC, E. (2007) *Limber Labs*, http://www.starlims.com/ICT_Autumn_2007.pdf, accessed on 19/09/2009.

KVARNSTROM, B. & OGHAZI, P. (2008) Methods for traceability in continuous processes-Experience from an iron ore refinement process. *Minerals Engineering,* 21**,** 720-730.

LABVANTAGE *LabVantage*, LabVantage, www.labvantage.com, accessed on 26/10/2009.

LEE, L. S., FIEDLER, K. D. & SMITH, J. S. (2008) Radio frequency identification (RFID) implementation in the service sector: A customer-facing diffusion model. *International Journal of Production Economics,* 112**,** 587-600.

LENK, C., HOPPE, N. & ANDORNO, R. (2007) *Ethics and law of intellectual property: current problems in politics, science and technology - Applied legal philosophy*, Ashgate Publishing, Ltd.

LI, C., YANG, Y.-X. & NIU, X.-X. (2006) Biometric-based personal identity-authentication system and security analysis. *The Journal of China Universities of Posts and Telecommunications,* 13**,** 43-47.

LINDEN, A. & FENN, J. (2003) *Understanding Gartner's hype cycles*, http://carbon.cudenver.edu/~jgerlach/emergingtechnologyOL/FirstReadings/HypeCycleIntro.pdf, accessed on 26/10/2009.

LITTON, J.-E. (2004) *A Biobank Information Management System: The Importance of Data Integration,* Stockholm, IBM, http://www-03.ibm.com/industries/global/files/Biobanks_Accelerating_Molecular_Medicine.pdf, accessed on 26/10/2009.

MAGELLANTECHNOLOGY (2009) *Magellan Technology*, http://www.magellan-rfid.com/, accessed on 02/05/2009.

MAHMOUD, Q. H. (2005) *Service-Oriented Architecture (SOA) and Web Services: The Road to Enterprise Application Integration (EAI)*, Sun Microsystems, Inc., http://java.sun.com/developer/technicalArticles/WebServices/soa/, accessed on 26/10/2009.

MAINTAG (2008) *BIOtrack® RFID*, MAINtag, http://www.maintag.com/images/ficheproduit_uk/10010058-060608%20BIOtrack.pdf, accessed on 04/10/2009.

MARTIN, P. R., BERDYCHEVSKI, R. E., SUBRAMANIAN, U., BLAKELY, W. F. & PRASANNA, P. G. S. (2007) Sample tracking in an automated cytogenetic biodosimetry laboratory for radiation mass casualties. *Radiation Measurements,* 42**,** 1119-1124.

MCFARLANE, D. & SHEFFI, Y. (2003) The Impact of Automatic Identification on Supply Chain Operations. *The International Journal of Logistics Management,* 14**,** 1-17.

MCGIVEN, J., HENDRY, L., BROWN, D., GILLARD, K., KATIYAR, A., STACK, J., PERRETT, L., ROLLINS, M., HUGHES, S., READY, R., KEMP, P., MORRIS, P. & MAWHINNEY, I. (2007) The Development of an Efficient Test Automation and Sample Tracking System in a High-Throughput Veterinary Laboratory to Improve the Emergency Response to Outbreaks of Highly Infectious Disease in Great Britain. *Journal of the Association for Laboratory Automation,* 12**,** 73-82.

MEDICINENET (2009) *Definition of Hippocratic Oath*, MedicineNet.com, http://www.medterms.com/script/main/art.asp?articlekey=20909, accessed on 26/10/2009.

MELAMED, J., DATTA, M., BECICH, M., ORENSTEIN, J., DHIR, R., SILVER, S., FIDELIA-LAMBERT, M., KADJACSY-BALLA, A., MACIAS, V., PATEL, A., WALDEN, P., BOSLAND, M. & BERMAN, J. (2004) The Cooperative Prostate Cancer Tissue Resource: A Specimen and Data Resource for Cancer Researchers *Clinical Cancer Research,* 10**,** 4614-21.

MERRIAM-WEBSTER (2002) *Merriam-Webster's medical desk dictionary*, Springfield.

METSPALU, A. (2004) The Estonian Genome Project. *Drug Development Research,* 62**,** 97–101.

METSPALU, A., KOHLER, F., LASCHINSKI, G., GANTEN, D. & ROOTS, I. (2004) The Estonian Genome Project in the context of European genome research. *Deutsche medizinische Wochenschrift,* 129 S25-8.

MEYER, H. J., CHANSUE, N. & MONTICELLI, F. (2006) Implantation of radio frequency identification device (RFID) microchip in disaster victim identification (DVI). *Forensic Science International,* 157**,** 168-171.

MHRA (2009) *Medicines and Healthcare products Regulatory Agency*, Medicines and Healthcare products Regulatory Agency, an executive agency of the Department of Health, http://www.mhra.gov.uk/index.htm, accessed on 28/10/2009.

MICROSENSYS (2009) *TELID® sensor devices* microsensys GmbH, http://www.microsensys.de/products/sensors.html, accessed on 26/10/2009.

MICROSOFT (2009a) *About Dynamic Data Exchange*, Microsoft Corporation, http://msdn.microsoft.com/en-us/library/ms648774(VS.85).aspx, accessed on 26/10/2009.

MICROSOFT (2009b) *MFC Library Reference - Automation*, Microsoft Corporation, http://msdn.microsoft.com/en-us/library/dt80be78.aspx, accessed on 26/10/2009.

MICROTOME (2009) *Microtome*, Microtome, http://www.microtome.com, accessed on 23/10/2009.

MIN, Z., LI, W., ZHONGYUN, W., LI BIN, A. L. B. & XIA, R. A. X. R. (2007) A RFID-based Material Tracking Information System. IN LI, W. (Ed.) *IEEE International Conference on Automation and Logistics.* China, IEEE.

MINISTRY OF HEALTH AND SOCIAL AFFAIRS, S. (2002) Biobanks in Medical Care Act. IN MINISTRY OF HEALTH AND SOCIAL AFFAIRS, S. (Ed.), MHSA.

MORGENROTH, D. & FOBES, K. (2004) Another link in the chain. *Card Technology Today,* 16**,** 11-12.

MOURTZIS, D., PAPAKOSTAS, N., MAKRIS, S., XANTHAKIS, V. & CHRYSSOLOURIS, G. (2008) Supply chain modeling and control for producing highly customized products. *CIRP Annals - Manufacturing Technology,* 57**,** 451-454.

MULLER, M. (2003) *Essentials of Inventory Management*, AMACOM.

NACHMAN, L., KLING, R., ADLER, R., HUANG, J. & HUMMEL, V. (2005) The Intel® Mote platform: a Bluetooth-based sensor network for industrial monitoring. *Proceedings of the 4th international symposium on Information processing in sensor networks.* USA, IEEE Press.

NATHANSON, V., ROMANO-CRITCHLEY, G., SOMMERVILLE. A & FRASER. P (1999) *Confidentiality and disclosure of health information, Guidance from the BMA's Medical Ethics Department*, BMA, http://www.eastmidlandsdeanery.nhs.uk/document_store/12142148621_confidentiality.pdf, accessed on 24/02/2009.

NATIONAL BIOETHICS ADVISORY COMMISSION (1999) Research Involving Human Biological Materials: Ethical Issues and Policy Guidelines. National Bioethics Advisory Commission.

NATIONAL HEALTH SERVICES (2008) *Pilot site: Mayday NHS Trust*, NHS, http://www.connectingforhealth.nhs.uk/systemsandservices/bloodpilot/mayday, accessed on 26/10/2009.

NATIONAL HEALTH SERVICES (2009a) *A project manager's implementation diary*, National Health Services, http://www.connectingforhealth.nhs.uk/systemsandservices/bloodpilot/mayday/diary, accessed on 27/10/2009.

NATIONAL HEALTH SERVICES (2009b) *Right Patient, Right Blood: What is the blood tracking pilot all about?*, National Health Services, http://www.connectingforhealth.nhs.uk/systemsandservices/bloodpilot/about, accessed on 27/10/2009.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (2008) *Doubly linked list*, National Institute of Standards and Technology,, http://www.itl.nist.gov/div897/sqg/dads/HTML/doublyLinkedList.html, accessed on 03/10/2009.

NATIONAL PATIENT SAFETY AGENCY (2006) Electronic Clinical Transfusion Management System. IN THE CHIEF MEDICAL OFFICER'S NATIONAL BLOOD TRANSFUSION COMMITTEE (Ed.), National Health Services.

NCI (2009) *Dictionary of Cancer Terms*, National Cancer Institute, http://www.cancer.gov/dictionary/?CdrID=44149, accessed on 15/09/2009.

NGAI, E. W. T., MOON, K. K. L., RIGGINS, F. J. & YI, C. Y. (2008) RFID research: An academic literature review (1995-2005) and future research directions. *International Journal of Production Economics*, 112**,** 510-520.

NIETFELD, J. J. (2007) What is anonymous? *EMBO reports*, 8**,** 518.

NPSA (2008) *National Patient Safety Agency*, National Patient Safety Agency,, www.npsa.nhs.uk/ accessed on 28/10/2009.

O'CONNOR, M. C. (2006) *Testing Ultrasound to Track, Monitor Patients*, RFIDJournal, http://www.rfidjournal.com/article/articleview/2199/1/131/, accessed on 26/10/2009.

OFFICE OF PUBLIC SECTOR INFORMATION (2005) Blood Safety and Quality Regulations 2005. The Stationery Office Limited.

ÖLUND, G., LINDQVIST, P. & LITTON, J.-E. (2007) BIMS: An information management system for biobanking in the 21st century. *IBM Systems Journal*, 46**,** 171-182.

PAOLI-CALMETTES, I. (2009) *Institut Paoli-Calmettes*, http://www.institutpaolicalmettes.fr/accueil.html, http://www.institutpaolicalmettes.fr/accueil.html, accessed on 28/10/2009.

PARDUS (2009) *Pardus*, Pardu, www.pardus.hr, accessed on 26/10/2009.

PARLIKAD, A. K. & MCFARLANE, D. (2007) RFID-based product information in end-of-life decision making. *Control Engineering Practice*, 15**,** 1348-1363.

PATEL, A., GILBERTSON, J., PARWANI, A., DHIR, R., DATTA, M., GUPTA, R., BERMAN, J., MELAMED, J., KAJDACSY-BALLA, A., ORENSTEIN, J. & BECICH, M. (2006) An informatics model for tissue banks--lessons learned from the Cooperative Prostate Cancer Tissue Resource. *BMC Cancer,* 6.

PATEL, A., KAJDACSY-BALLA, A., BERMAN, J., BOSLAND, M., DATTA, M., DHIR, R., GILBERTSON, J., MELAMED, J., ORENSTEIN, J., TAI, K. F. & BECICH, M. (2005) The development of common data elements for a multi-institute prostate cancer tissue bank: The Cooperative Prostate Cancer Tissue Resource (CPCTR) experience. *BMC Cancer,* 5.

PAUTASSO, C., ZIMMERMANN, O. & LEYMANN, F. (2008) RESTful Web Services vs. Big Web Services: Making the Right Architectural Decision. *17th International World Wide Web Conference (WWW2008).* China.

PCRC (2009) *Irish Prostate Cancer Consortium biobank*, Irish Prostate Cancer Consortium, https://pcrc.tchpc.tcd.ie, accessed on 11/09/2009.

PIRAMUTHU, S. (2007) Protocols for RFID tag/reader authentication. *Decision Support Systems,* 43**,** 897-914.

PITT, K. E., CAMPBELL, L. D. & SKUBITZ, A. P. N. (2005) Best Practices for Repositories I: Collection, Storage, and Retrieval of Human Biological Materials for Research. *Cell Preservation Technology,* 3**,** 5-48.

PORTERI, C. & BORRY, P. (2008) A proposal for a model of informed consent for the collection, storage and use of biological materials for research purposes. *Patient Education and Counseling,* 71**,** 136-142.

RASILANTTECHNOLOGIES (2009) *White Paper: Radio Frequency Identification Technology (RFID) In The Diamond Industry*, Rasilant Technologies Pvt. Ltd., http://www.rfidinfo.jp/whitepaper/802.pdf, accessed on 26/10/2009.

REGATTIERI, A., GAMBERI, M. & MANZINI, R. (2007) Traceability of food products: General framework and experimental evidence. *Journal of Food Engineering,* 81**,** 347-356.

REKIK, Y., SAHIN, E. & DALLERY, Y. (2008) Analysis of the impact of the RFID technology on reducing product misplacement errors at retail stores. *International Journal of Production Economics,* 112**,** 264-278.

RIEBACK, M. R., SIMPSON, P. N. D., CRISPO, B. & TANENBAUM, A. S. (2006) RFID malware: Design principles and examples. *Pervasive and Mobile Computing,* 2**,** 405-426.

ROBERTS, C. M. (2006) Radio frequency identification (RFID). *Computers & Security,* 25**,** 18-26.

ROSKIN, G. (2001) *Diamond Inscription Enters the Matrix*, Reed Elsevier Inc., http://www.jckonline.com/article/291845-Diamond_Inscription_Enters_the_Matrix.php, accessed on 26/10/2009.

ROTONDI, A. J., BRINDIS, C., CANTEES, K. K., DERISO, B. M., ILKIN, H. M., PALMER, J. S., GUNNERSON, H. B. & DAVID WATKINS, W. (1997) Benchmarking the perioperative process. I. Patient routing systems: A method for continual improvement of patient flow and resource utilization. *Journal of Clinical Anesthesia,* 9**,** 159-169.

RUBINSTEIN, W. (2006) Gemstone inventory and detection system. United States, 7119683.

SANGWAN, R. S., QIU, R. G. & JESSEN, D. (2005) Using RFID tags for tracking patients, charts and medical equipment within an integrated health delivery network. *Proceedings of the IEEE Networking, Sensing and Control Conference.*

SCHMITZ, G., BHATIA, K., FERGUSON, M. & D.M., W. (2005) *Biobanks: Collaborating for Cures*, IBM, http://www-03.ibm.com/industries/global/files/Biobanks.pdf, accessed on 26/10/2009.

SCHREIER, P. (2008) *Biosample storage*, Scientific Computing World, http://www.scientific-computing.com/features/feature.php?feature_id=195, accessed on 11/12/2008.

SERIOUS HAZARDS OF TRANSFUSION (2007) *Annual report 2007 Summary*, Manchester, http://www.shotuk.org/Summary%202007.pdf, accessed on 26/10/2009.

SHANG-WEI, W., WUN-HWA, C., CHORNG-SHYONG, O., LI, L. & YUN-WEN, C. (2006) RFID Application in Hospitals: A Case Study on a Demonstration RFID Project in a Taiwan Hospital. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences HICSS '06.* . Kauai, IEEE.

SHICKLE, D. (2006) The consent problem within DNA biobanks. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences,* 37**,** 503-519.

SHIRLEY, S. G., DURST, C. H. P., FUCHS, C. C., ZIMMERMANN, H. & IHMIG, F. R. (2009) A large-scale cryoelectronic system for biological sample banking. *Cryogenics,* In Press, Corrected Proof.

SHOT *Serious Hazards of Transfusion*, http://www.shotuk.org/, accessed on 28/10/2009.

SIGMA-ALDRICH (2009) *RNAlater™ Storage Solution*, Sigma-Aldrich Co., http://www.sigmaaldrich.com/life-science/molecular-biology/dna-and-rna-purification/rnalater-storage-solution.html, accessed on 19/10/2009.

SLIDEPATH *SlidePath Digital Slide Solutions*, Slidepath, http://slidepath.com/, accessed on 27/08/2009.

SMART MEDICAL TECHNOLOGIES (2005) *Smart Test Tube™* http://www.smartmedtech.com/Products/product1.php, accessed on 04/10/2009.

SOGA, S., HIROSHIGE, Y., DOBASHI, A., OKUMURA, M. & KUSUZAKI, T. (1999) Products lifecycle management system using radio frequency identification technology. *Proceedings of 7th IEEE International Conference onEmerging Technologies and Factory Automation, ETFA '99* IEEE.

SONG, J., HAAS, C. T., CALDAS, C., ERGEN, E. & AKINCI, B. (2006) Automating the task of tracking the delivery and receipt of fabricated pipe spools in industrial projects. *Automation in Construction,* 15**,** 166-177.

SORENSEN, T. I. A., BELL, J., HUTTUNEN, J., PIKANI, J. & PRINGLE, M. (2005) *The International Evaluation of Swedish Biobanks*, http://www.biobanks.se/documents/Evaluation%20of%20Swedish%20biobanks%20Final.pdf, accessed on 24/02/2009.

SPACECODE-RFID (2007) *RFID In The Diamond Industry*, Spacecode, http://www.spacecode-rfid.com/site/french/markets/pdf/diamond.pdf, accessed on 01/05/2009.

SPACECODE-RFID (2009) *Spacecode*, Spacecode, http://www.spacecode-rfid.com, accessed on 01/05/2009.

SPYCHIPS (2003) *Position Statement on the Use of RFID on Consumer Products* Spychips, http://www.spychips.com/jointrfid_position_paper.html, accessed on 8/08/2009.

STAFFORD, J. E. H. (1998) LIMS: an automating or informating technology? *Laboratory Automation & Information Management,* 33**,** 163-168.

STARLIMS (2008) *STARLIMS version 10*, STARLIMS, http://www.starlims.com/STARLIMS_V10_Brochure.pdf, accessed on 26/10/2009.

STARLIMS (2009) *STARLIMS*, STARLIMS, www.starlims.com/, accessed on 26/10/2009.

STATISTICS ESTONIA (2009) *Most requested statistics*, Statistics Estonia,, http://www.stat.ee/main-indicators, accessed on 28/10/2009.

STOCKMAN, H. (1948) Communication by Means of Reflected Power. *Proceedings of the Institute of Radio Engineering.* IEEE.

STOPRFID (2005) *stopRFID*, stopRFID http://www.stoprfid.de/en/, accessed on 08/08/2009.

SUTROP, M. & SIMM, K. (2004) The Estonian healthcare system and the genetic database project: from limited resources to big hopes. *Cambridge Quarterly of Healthcare Ethics*, 13**,** 254-62.

SZARVAS, G., FARKAS, R. & BUSA-FEKETE, R. (2007) State-of-the-art Anonymization of Medical Records Using an Iterative Machine Learning Framework. *Journal of the American Medical Informatics Association*, 14**,** 574-580.

SZMEREKOVSKY, J. G. & ZHANG, J. (2008) Coordination and adoption of item-level RFID with vendor managed inventory. *International Journal of Production Economics*, 114**,** 388-398.

TAGSYS RFID (2008) *TagSYS RFID e-connecting goods*, TagSYS-RFID, http://www.tagsysrfid.com/, accessed on 27/12/2008.

TAILBY, R. (2002) The Illicit Market in Diamonds. IN AUSTRALIAN INSTITUTE OF CRIMINOLOGY (Ed.), Australian Institute Of Criminology,.

TAJIMA, M. (2007) Strategic value of RFID in supply chain management. *Journal of Purchasing and Supply Management*, 13**,** 261-273.

TEC-IT *2D Barcode Symbologies Overview*, TEC-IT Datenverarbeitung GmbH, http://www.tec-it.com/en/support/knowbase/symbologies/barcode-overview/2d-barcodes/Default.aspx, accessed on 22/12/2008.

TEKLYNX *2D BARCODE SYMBOLOGIES*, Braton Groupe sarl., http://www.teklynx.com/barcodes/article_43.html, accessed on 22/12/2008.

TEKLYNX *CODE 49*, Braton Groupe sarl., http://www.teklynx.com/barcodes/article_47.html, accessed on 22/12/2008.

TEXASINSTRUMENT (2008), Texas Instruments Incorporated, http://www.ti.com/rfid/, accessed on 27/12/2008.

THERMO FISHER SCIENTIFIC INC (2009) *Nautilus LIMS*, Thermo Fisher Scientific Inc., www.thermo.com/nautilus, accessed on 21/09/2009.

THERMO SCIENTIFIC (2009) *Shandon Grosslab? Senior Grossing Workstation*, Thermo Fisher Scientific Inc., http://www.thermo.com/com/cda/product/detail/1,,1000005630393,00.html, accessed on 23/10/2009.

TI (2008), http://focus.ti.com/paramsearch/docs/parametricsearch.tsp?family=rfid&sectionId=475&tabId=2102&familyId=1352, accessed on 27/12/2008.

TORRE-BUENO, J. D. L. (2007) System for tracking biological samples. United States.

TZENG, C.-T., CHIANG, Y.-C., CHIANG, C.-M. & LAI, C.-M. (2008a) Combination of radio frequency identification (RFID) and field verification tests of interior decorating materials. *Automation in Construction,* 18**,** 16-23.

TZENG, S.-F., CHEN, W.-H. & PAI, F.-Y. (2008b) Evaluating the business value of RFID: Evidence from five case studies. *International Journal of Production Economics,* 112**,** 601-613.

U.S. NATIONAL INSTITUTES OF HEALTH (2008) *NIA Glossary of Clinical Research Terms*, www.nia.nih.gov/NR/rdonlyres/D95A1337-6340-4C95-96F9-ABD3FD3C0F03/0/NIAGlossaryofClinicalResearchTermsFINAL.doc, accessed on 14/05/2010.

UK BLOOD TRANSFUSION SERVICES *National Blood Transfusion Committee*, UK Blood Transfusion & Tissue Transplantation Services, http://www.transfusionguidelines.org/Index.aspx?Publication=NTC&Section=27&pageid=999, accessed on 28/10/2009.

UK HUMAN TISSUE BANK (2006) *UK Human Tissue Bank*, UK Human Tissue Bank, http://www.ukhtb.org, accessed on 19/10/2009.

UKBIOBANK (2008) *UKBiobank*, http://www.ukbiobank.ac.uk/, accessed on 11/12/2008.

UMETANI, T., ARAI, T., MAE, Y., INOUE, K. & MAEDA, J.-I. (2006) Construction automation based on parts and packets unification. *Automation in Construction,* 15**,** 777-784.

UNESCO (1997) *Glossary*, United Nations Educational, Scientific and Cultural Organization, http://www.unesco.org/webworld/mdm/administ/en/guide/guide011.htm, accessed on 25/10/2008.

UNITED STATES DEPARTMENT OF HEALTH AND HUMAN SERVICES (2003) Summary of the HIPAA Privacy Rule. IN UNITED STATES DEPARTMENT OF HEALTH AND HUMAN SERVICES (Ed.), HHS.

USAMI, M. & OHKI, M. (2003) The μ -Chip: An Ultra-Small 2.45 GHz RFID Chip for Ubiquitous Recognition Applications. *Institute of Electronic Information Communication Engineering (IEICE) Transactions on Electronics,* E86-C**,** 521-528.

UZUNER, O., LUO, Y. & SZOLOVITS, P. (2007) Evaluating the State-of-the-Art in Automatic De-identification. *Journal of the American Medical Informatics Association,* 14**,** 550-563.

VENKATESAN, M. & GRAUER, Z. (2004) Leveraging Radio Frequency Identification (RFID) Technology to Improve Laboratory Information Management. *American Laboratory,* 36**,** 11-15.

VERGARA, A., LLOBET, E., RAMIREZ, J. L., IVANOV, P., FONSECA, L., ZAMPOLLI, S., SCORZONI, A., BECKER, T., MARCO, S. & WOLLENSTEIN, J. (2007) An RFID reader with onboard sensing capability for monitoring fruit quality. *Sensors and Actuators B: Chemical,* 127**,** 143-149.

WAGAR, E. A., STANKOVIC, A. K., RAAB, S., NAKHLEH, R. E. & WALSH, M. K. (2008) Specimen labeling errors: a Q-probes analysis of 147 clinical laboratories. *Archives of Pathology and Laboratory Medicine,* 132**,** 1617-22.

WALLACE, H. M. (2005) The development of UK Biobank: Excluding scientific controversy from ethical debate. *Critical Public Health,* 15**,** 323 - 333.

WATSON, J. & CYRANOSKI, D. (2005) Beset by practical hurdles, UK Biobank moves at sluggish pace. *Nat Med,* 11**,** 696-696.

WEBB, W. (2008) *RFID in embedded designs: Your move,* Electronic Design, Strategy, News (EDN), http://www.edn.com/article/CA6615601.html, accessed on 23/12/2008.

WENDLER, D. & EMANUEL, E. (2002) The Debate Over Research on Stored Biological Samples: What Do Sources Think? *Archives of Internal Medicine,* 162**,** 1457-1462.

WINN, D. M., REICHMAN, M. E. & GUNTER, E. (1990) Epidemiologic issues in the design of biologic specimen banks. *Epidemiologic Reviews,* 12**,** 56-70.

WMA (2009) *The World Medical Association*, WMA, www.wma.net, accessed on 26/10/2009.

WOOD, S. (2007) Comprehensive Laboratory Informatics: A Multilayer Approach. *American Laboratory,* 39**,** 20-23.

WOODWARD, B. (1995) The computer-based patient record and confidentiality. *The New England Journal of Medicine,* 333**,** 1419-22.

WORLD HEALTH ORGANIZATION (2009) *Blood transfusion,* WHO, http://www.who.int/topics/blood_transfusion/en/, accessed on 11/06/2009.

ZARABZADEH, A., WATSON, R. W. G. & GRIMSON, J. (2008) The use of radio frequency identification to track samples in bio-repositories. *1st International Conference on Information Technology, IT 2008.* Poland, IEEE.

ZHANG, X.-D., YUE, S.-J. & WANG, W.-M. (2006) The review of RFID applications in global postal and courier services. *The Journal of China Universities of Posts and Telecommunications,* 13**,** 106-110.

Appendix A. The SOP for Collecting Blood in EDTA Tube

Standard Operating Procedures

Prostate Cancer Research Consortium 1/11/04

SOP 10 : **Total DNA from peripheral blood**

4ml EDTA blood sample (purple lid)

For genomic and mitochondrial DNA isolation

Need before starting

1. QIAamp DNA blood mini kit (Qiagen cat no. 51104) for volumes up to 400µl.
2. Set heating block to 56^{o}C
3. Equilibrate Buffer AE to RT
4. If Buffer AL has a precipitate, dissolve by incubating at 56^{o}C

1. Equilibrate sample to RT

2. Spin sample at 2,500x*g* for 10 minutes at RT. 3 layers are now visible:
upper clear layer (plasma (inc. serum & fibrinogen))
middle white layer (buffy coat containing concentrated leukocytes)
lower red layer (concentrated erythrocytes)

Note: buffy coat isolation is not necessary but does yield ~ 5-10 times more DNA than an equivalent volume of whole blood.

3. Aspirate upper plasma layer into a 1.5ml eppendorf and label clearly. Store in prostate cancer biobank –80ºC freezer.

4. Carefully transfer buffy coat layer in 200µl volumes to a fresh 1.5ml eppendorf.

Note: If buffy coat volume is <200µl, bring volume up to 200µl with PBS. If volume is >200ul but < than 400ul, increase volumes of protease and Buffer AL accordingly.

5. Add 20µl of protease & 20µl RNAseA (20mg/ml).

6. Add 200µl of Buffer AL. mix by pulse vortexing for 15secs.

Note: mix thoroughly, solution should be homogeneous.

7. Incubate tube at 56ºC for 10 mins, spin down to remove drops from lid.

8. Add 200µl of ethanol (96-100%) to the sample, & vortex for 15 secs. Spin down to remove drops from lid. If sample volume larger than 200µl, increase ethanol volume proportionally.

9. Carefully apply sample to QIAamp spin column in a 2ml collection tube, without wetting the rim.

10. Centrifuge at full speed for 1 min.
Note: if lysate has not completely passed through column after centrifugation, centrifuge again until spin column is completely empty.

11. Transfer spin column to a fresh 2ml collection tube. Add 500µl of buffer AW1.

12. Centrifuge at 6000x*g* for 1 minute.

13. Transfer spin column to a fresh collection tube. Add 500ul of buffer AW2.

14. Centrifuge at full speed (20,000x*g*) for 3 minutes.

15. Transfer spin column to a fresh collection tube and spin again at 20,000x*g* for 1 minute.

16. Transfer spin column to a fresh, labelled 1.5ml eppendorf tube. Add 200ul of Buffer AE. Incubate at RT for 5 min.

17. Centrifuge at 6000x*g* for 1 min.

18. Repeat step 16 and 17 & combine samples.

19. Spec DNA, typical yield should by ~ 6ug, A260/A280 ratio of 1.7-1.9.

Appendix B. Workflow Documents for Blood and Serum Proteomics, and Urinary DiGE Analysis

Work Flow – Proteomics

1. Serum or Plasma (Collected from patients using SOP and stored at –80°C)

2. Patient groups identified and samples requested from Bio-Resource (we currently request a 500ul aliquot from one of the three 1500ul aliquots stored) (see "Serum Pilot Study n=6 gleason 5 v 7 for work flow" for example of the clinical and collection data)

3. Fractionation of samples (HPLC removal of abundant proteins – results in the generation of two samples

(a) high abundant proteins which needs to be stored and tracked for future analysis

(b) low abundant proteins used for further DIGE analysis

4. DIGE analysis – two samples stained either with Cy3 or Cy5 and combined with pooled standard. Pooled protein is separation by 2-D gel electrophoresis based on size and charge, gel is then scanned (using a Typhoon 9410 Variable Mode Imager (Amersham BioScience – which creates images of 27-20MB in size) and uploaded into Progenesis (see Work Flow – DIGE) (see attached manual for Progenesis software). This fluorescence image needs to be saved for future analysis using more up to date software (average size) as it becomes available. Spot ID generated using Progenesis software and correlated to clinical parameters samples were chosen from (see "Progenesis Data for Work Flow" for example of generated data set of differentially expressed spots between Gleason 5 vs. 7).

5. None-cy labelled samples are then run on a 2-D gel and stained with silver stain to visualise protein spots. Spots of interest (identified in 4) are identified and cut by a robot and identified by MALDI-Tof/Tof analysis. List of differentially expressed proteins is generated based on specific clinical question

Workflow for Proteomic Analysis of Human Serum

1. Sample collection and storage
    a. Clinical serum sample received from hospital lab is thawed and aliquotted out into 100µl samples. These are stored at -80°C.
2. Serum depletion
    a. A 100µl aliquot of crude serum is removed from the -80°C freezer and allowed to thaw.
    b. 40µl of each serum sample is diluted 1:5 in Agilent buffer A. 200µl of this solution is injected into the Vision HPLC system to undergo depletion using the MARS column.
    c. 2ml of diluted depleted serum is collected off the column and stored immediately at -80°C.
    d. 2ml of diluted abundant proteins is collected off the column and stored immediately at -80°C.
    e. For each sample, the depletion is done in duplicate i.e. two injections of 40µl crude serum.
3. Concentration and clean up of depleted serum
    a. To 2ml of diluted depleted serum, 4X volume of ice-cold acetone is added and mixed well.
    b. This is kept at -20°C for at least 2 hours and then centrifuged at 4°C at 1000 g for 15 mins.
    c. The resulting protein pellet is resuspended in lysis buffer.
    d. Samples are cleaned using GE Healthcare 2D clean up kit.
4. Protein estimation
    a. Bradford assay incorporating a BSA standard curve is used to determine the protein content of each sample (~3µl of each sample used).
5. 2-DE and DIGE
    a. Depending on whether 2-DE or DIGE is being run, a certain amount (µg) of depleted serum is removed from each sample and either labelled with fluorescent dye (DIGE) or applied directly to IPG strips.

Database Protocol for Urinary DiGE

1. Urine thawed from –80C freezer. 40ml removed and remainder refrozen.
2. 40ml centrifuged for 5 mins to remove debris.
3. Supernatant recovered and 50% TCA added. Centrifuged for 15 mins.
4. Small pellet visible at bottom of tube- supernatant removed and ice cold acetone added. Centrifuged for 25 mins.
5. Supernatant removed and pellet left to air dry for < 5mins. DiGE lysis buffer added to resuspend (80-100uL depending on difficulty in getting pellet to resuspend).
6. Sample can be placed back in –80C at this point.
7. Sample desalted using 2D Cleanup Kit (GE Healthcare) as per protocol.
8. Protein estimation performed using 2D Quant Kit (GE Healthcare).
9. 25ug protein labelled with 200pmol of Cy 3 & 5. Internal standard labelled with Cy 2. Samples combined so total of 75ug run per gel.
10. Sample combined with rehydration buffer to total volume of 450uL to rehydrate 24cm pH 4-7 IPG strip.
11. Samples focused in first dimension and separated in second dimension as per non DiGE gels.
12. Images scanned on Typhoon Scanner and image analysis performed using Progenesis software package (Non Linear Dynamics).

Appendix C. Workflow Diagrams for Tissue and Urine Samples

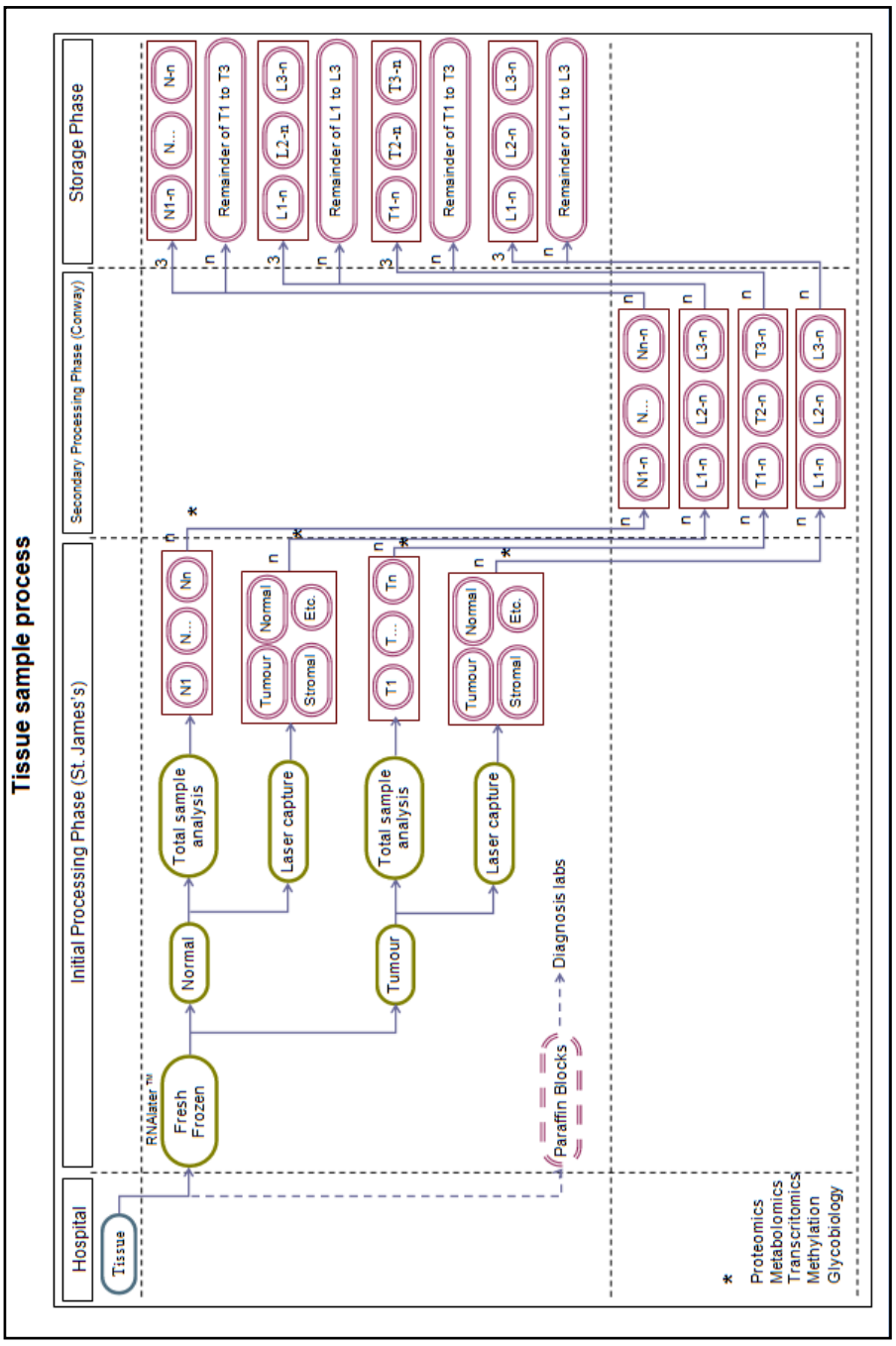


Figure App. C-1: Workflow Diagram for Tissue

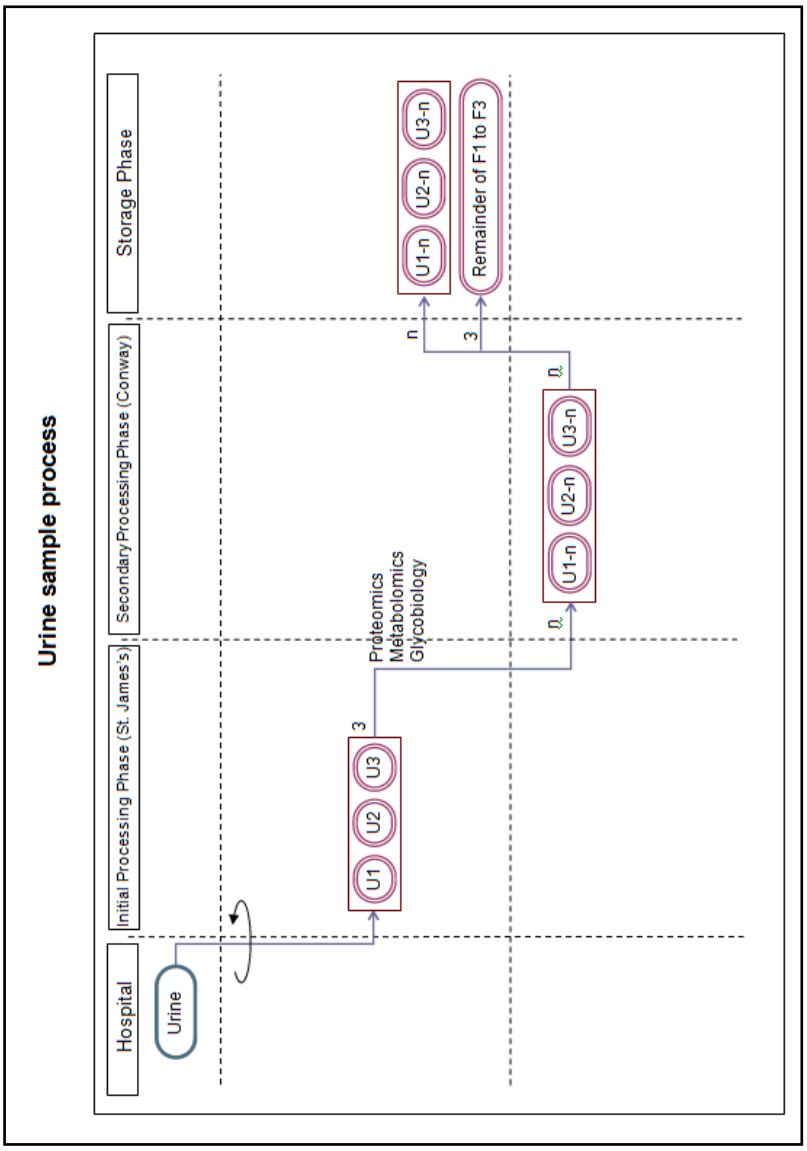


Figure App. C-2: Workflow Diagram for Urine

Appendix D. PCRC Biobank Patient Information Leaflet and Consent Form

Patient Information Leaflet and Consent Form

Molecular and Cellular Mechanisms Leading to Prostate Cancer Cell Development

What is the purpose of this research study?

The purpose of this leaflet is to explain as clearly as possible all the procedures involved in this study before you decide whether or not to take part. Prostate cancer is now the most common non cutaneous cancer among men in Ireland. Risk factors include a positive family history, advanced age and diet: however prevention and screening is difficult as there is a marked disease diversity with some disease progressing quickly and some disease progressing slowly. Determining which patients to treat aggressively remains a significant dilemma for the clinician, indicating prostate cancer as an appropriate disease to pursue novel markers of disease and progression. The aim of this study is to characterise prostate cancer cells and investigate the mechanisms and markers of tumour progression. Key to understanding prostate cancer is the collection of appropriate samples from prostate cancer patients at various stages of their disease, these include: tissue, blood and urine. The Prostate Cancer Research Consortium (see "Who is organising the study?") which is co-ordinating this study will use these samples in combinations with your anonymised clinical data to better understand prostate cancer development and progression. The consortium has utilised the significant experience of the Trinity Centre for High Performance Computing in order to develop a central database across the different sites of the consortium allowing central access to all anonymised patient and experimental data, as well as tracking all the patient samples.

Why have I been chosen?

All patients undergoing TURP or Radical Prostatectomy will be approached to participate in this study.

Who is organising this study?

This study is being organised by *(Name of lead consultant for the respective hospital site*) from the Prostate Cancer Research Consortium. Created in October 2003 with funding from the Irish Cancer Society and under the Dublin Molecular Medicine Centre, the Prostate Cancer Research Consortium is a multi-disciplinary, trans-institutional collaboration. The consortium is currently made up of five teaching hospital sites, Mater Misericordiae University Hospital, St Vincent's University Hospital, Beaumont Hospital, St James' Hospital and AMNCH as well as three research institutions Conway Institute, University College Dublin, Institute of Molecular Medicine, Trinity College Dublin and Clinical Research Centre, Royal College of Surgeons in Ireland. Additional Irish academic sites may join the consortium. Each hospital site maintains and organises its own biobank and clinical data but samples tracking and anonymised clinical information is centrally stored on the database.

What will happen to me if I take part?

Prior to your surgery, you will be invited to take part in this study. You will be given as much time as necessary to think this over and to ask questions. Prior to your surgery, a 10ml blood sample and 50ml urine sample will be obtained. During the operation the surgeon will remove some or all of your prostate gland. This will undergo routine analysis in the hospital laboratory. If you decide to take part in the study, both the material used for routine analysis and a fresh portion of the prostate gland will be available to the consortium for the characterisation of prostate cells and discovery of biomarkers. This material will be characterised at the level of DNA, RNA and protein, for genes and proteins responsible for tumour progression and survival. DNA, RNA protein and anonymised clinical data may be shared with other groups within or outside the consortium in Ireland and abroad working to discover markers of prostate cancer and progression. The clinical material and clinical information will be stored for 15 years, at the respective hospital biobank and central database.

Are there any disadvantages in taking part in this study?

You may experience discomfort at the site where the blood is taken. Where possible, the research blood sample will be taken at the same time as routine bloods to minimise this discomfort.

Are there any possible risks of taking part?

No.

What are the possible benefits of taking part?

You will not receive any direct benefit from taking part in the study. However, information obtained during the course of the study may help us to more clearly understand prostate cancer and to treat patients with prostate cancer in the future. You will not benefit financially in any way, should this research lead to the development of a new treatment or medical tests.

What happens after the study?

Once your surgery has been performed, your doctors will maintain the normal postoperative follow-up.

Confidentiality- who will know I am taking part in this study?

All information, which is collected, about you during the course of the research will be kept strictly confidential. Your tissue, blood, urine sample and clinical data as well as the products of this material, including DNA, RNA and Protein, will be made anonymous prior to being made available to consortium researchers or other collaborating academic or biotechnology research groups but a code will be maintained on the hospital system to allow appropriate follow up. Any information about you, which leaves the hospital, will contain no information as to your identity so that you cannot be recognised from it. Your medical records may be reviewed and information taken from them by (*Name of lead consultant for the respective hospital site*) team in the strictest confidence and your GP may be contacted to follow up your progress.

In order to verify that the study meets best international research standards and practices, your file may be accessed for audit purposes by the hospital Research Ethics Committee or relevant regulatory

authorities such as the Data Protection Agency, Irish Medicines Board or European Commission, who will be bound by a strict code of confidentiality and your identify will never be revealed to the auditors.

If the research is to differ in any significant material way from the purposes currently envisaged, you will be asked for further consent.

Hospital Research Ethics Committee Approval.

This study has been approved by the (*Name of respective Hospital*) Research Ethics Committee.

What will happen to the results of the study?

The tissue, blood and urine sample you have given for this research and the information gathered may be stored in computer or manual format. The results obtained from your participation in this study will be combined with the results from other patients. The findings will be presented at research meetings and may be published in medical journals.

Any information about you, which leaves the hospital, will contain no information as to your identity so that you cannot be recognised from it. The results of this study will not be made known to you.

Voluntary participation.

It is up to you to decide whether to take part or not. If you do decide to take part in the study following reading this information leaflet and talking with the research nurse you will be given a consent form to sign. Even if you do decide to take part, you are free to withdraw at any time and without giving a reason. This will not affect the standard of care you will receive. Your doctor will not be upset if you decide not to take part. If you withdraw your consent, your samples will be destroyed.

Further Research

Further research, using the sample that you give, may include genetic research, which will be aimed at understanding genetic influences and risks related to prostate cancer and its treatment. The results of these further investigations are unlikely to have any implications for you personally. This research may involve sharing of your DNA, RNA and protein extracted from your donated tissue sample and medical data with other academic or biotechnology research groups.

Contact details.

If you would like further information about any aspect of this study, please feel free to contact either of the following.

Thank you for agreeing to take part in this study.

Consent Form

Reference Number:……. **Protocol Number:**

Title of Study: Molecular and cellular mechanisms leading to prostate cancer cell development.

Patient name:……..

Name of Doctor and telephone no:

1. I confirm that I have read and understood the information leaflet dated xxxxx for the above study and received an explanation of the nature, purpose, duration, and foreseeable effects and risks of the study and what my involvement will be.
2. I have had time to consider whether to take part in this study. My questions have been answered satisfactorily and I have received a copy of the patient information leaflet.
3. I understand that my participation is voluntary (by choice) and that I am free to withdraw at any time without my medical care or legal rights being affected.
4. I have to the best of my knowledge informed the investigator of my previous or present illnesses and medications and of any consultation I have had with a doctor for the last 4 months.
5. In order to verify that the study meets best international research standards and practices, I understand that my file may be accessed (at random) for audit purposes by the hospital Research Ethics Committee or relevant regulatory authorities, who will be bound by a strict code of confidentiality.
6. I understand that my blood, urine and tissue samples and clinical data will be collected and stored as in the information leaflet and will be coded and made anonymous prior to being made available to consortium researchers but a code will be maintained on the hospital system to allow appropriate follow up by the research nurse only.
7. I understand that this research may involve sharing my anonymised DNA, RNA, protein and clinical data with other academic or biotechnology research groups within or outside the consortium in Ireland and abroad.
8. I agree to take part in the above study as detailed in the information leaflet.

……	……..	……
Name of patient	Date	Signature
……	……..	……
Name of person taking consent	Date	Signature
……	……..	……
		Doctor/Researcher ***Date***

Appendix E. Pilot and Studies

Examples for different scenarios depending on the environment, amount of items to be tracked, nature of the items and range of operation have been observed. For each example the technology used for what type of application has been listed in tabular form. The extent that the application has been developed, advantages and limitations that are experienced by the developer have been investigated.

Construction Environment

Construction environments are often noisy both by sound and electromagnetic waves, as well as being harsh by the nature of the jobs carried out in there. Asset management is often a major time consuming task. A piece of equipment is used by multiple people in different part of the site. Studies have investigated methods of asset tracking, tools and components. Other areas of activities carried out on construction sites are the subject of other studies, for instance quality control, concrete processing and tracking, and end-of-life decision making. Examples of applications in construction environment are given in Table App. E-1.

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Goodrum et al., 2006)	RFID	Tool Tracking	Field trials	Readability over walls Low temperature survival (-12°C) Long reading range Reliable data transfer	Reading range shortening in extreme temperatures Cost Lack of standardisation Lack of directional and range data
2	(Jaselskis et al., 1995)	RFID + others	Concrete processing and handling Cost coding for labour and equipment Materials control	Proposal	Reading from distance Multiple reads at once Data storage on tags Eliminating the need for optical scan of barcodes	Interoperability across system suppliers Psychological barriers Functionality issues by metal and other RF systems
3	(Furlani and Stone, 1999)	RFID, barcodes, PDA and GPS	Component tracking	Prototype/to be fully deployed	Data storage of high value components Tracking assembled and sub-assembled components	
4	(Min et al., 2007)	RFID and PDA	Quality inspection and management of concrete specimens	Field trials	Improvement of work efficiency Eliminating operation cost Improvement of customer satisfaction Time saving	Reading range of 3cm when placed inside the concrete, i.e. short reading range

No	Paper	Technology	Application	Scope	Advantages	Limitations
5	(Song et al., 2006)	RFID	Pipe spools tracking		Operational in metal intense and congested environment Long read range Improvement in efficiency for receiving and inventory purposes	Reading difficulties when low RF surrounded by solid metal Reading issues when tags in full touch with a surface Accurate and time-efficient Reliable reads when moving with a speed less than 2 mph
6	(Kvarnstrom and Oghazi, 2008)		Tracing continuous processes-Iron ore refinement		Prevent lot mismatching Lot-end-mixing Lot-sequence mixing	
7	(Min et al., 2007)	RFID	Material tracking	Prototype	Traceability	
8	(Umetani et al., 2006)	RFID	Parts and packets unification process automation	Experimented	Robustness against environment, e.g. temperature and humidity Assembly and disassembly data inheritance and management	Not operational in metal, water and noise surrounding environments Reader antenna not mountable on machines
9	(Parlikad and McFarlane, 2007)	Networked RFID	Product end-of-life decision making	Under development and refinement	Unique product identifier Wireless communication Networked readers	

No	Paper	Technology	Application	Scope	Advantages	Limitations
10	(Tzeng et al., 2008a)		Interior decorating	Prototype		Metal affecting RFID functionality that might lead to complete failure Accuracy issues depending on the distance and reading angle Expensive

Table App. E-1: Pilots and Studies on Construction Environment

Kvarnstrom and Oghazi (Kvarnstrom and Oghazi, 2008) have talked about an iron ore refinement process. During this process material with different concentration, i.e. different characteristics will pass different surroundings environments. The process cannot be defined as a linear flow; hence it is not possible to recognise the path in advance.

From Table A2-1following lessons can be learnt:

- Although RFID provides distance learning, the reading range and reading rate is reduced in low temperatures
- RFID tags can be unreliable in metal intense environments
- RFID can be combined with other technologies such as barcodes to provide a better solution

Retail Supply Chain Management

Supply chain can be considered a closed loop system that a certain path is defined for each product. However maintaining a track of products lifecycle will help improve quality of products as well as dealing with management issues, such as preventing empty shelves. Table App. E-2 briefly describes the areas that have been investigated in supply chain management.

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Mourtzis et al., 2008)	RFID	Control on highly customised products	Design and implementation	Real or near real time information provided	
2	(Brown and Russell, 2007)	RFID	Retail sector	Study		Lack of standards Costs
3	(Tajima, 2007)	RFID	Retail	Study	Quality control Production tracking Product/material handling Increased data accuracy Reduced stock-outs Customer service ...	Lack of Return on Investment (ROI) Technical risks Popularity of barcodes Privacy concerns
4	(Soga et al., 1999)	RFID	Product life-cycle management	Prototype	Data storage on the product Lifecycle management	Small memory size
5	(Chow et al., 2006)	RFID	Warehouse operations management	Pilot	enhanced the warehousing operating performance: operation level enhancement, operating cost reduction, customer satisfaction and resource management	

No	Paper	Technology	Application	Scope	Advantages	Limitations
6	(Bajic and Chaxel, 2002)	escort memories/ electronic tags	Vehicle lifecycle data management	Elementary implementation	Extremely high temperature (120C)	Small memory size Tags surviving extreme temperature being expensive Temperature affecting performance and reliability

Table App. E-2: Pilots and Studies on Retail Supply Chain Management

Rekik et al. talks about reducing misplacement errors in stores by a system keeping track of products (Rekik et al., 2008). They find RFID a suitable technology only for more expensive products due to its costs. Szmerekovsky and Zhang on the other hand proposed an item-level inventory that shares the costs between different parties to overcome this issue (Szmerekovsky and Zhang, 2008).

Lessons learnt:

- Cost is a major barrier for RFID adoption in retail. Return On Investment (ROI) should be guaranteed
- RFID can be used for more expensive products
- However, it can improve operations and maintain customers satisfactions

Library and Other Applications

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Coyle, 2005)	RFID	Library	Study	Multiple action done at once in a closed-loop environment Fourteen areas in which RFID will benefit libraries	Tags being larger than items in some cases Read range issues Long life tags Reprogrammable and inexpensive tags
2	(Lee et al., 2008)	RFID	Library	Partly implemented	Unique identifier for each item Multiple reads at once Speed In the future: inventory maintenance	
			Road running race	Partly implemented	Quicker availability of race results Each runner individually timed	
			Healthcare system vendor: Patient tracking Asset tracking	Partly implemented	Minimised search time Faster response to patients	

Table App. E-3: Pilots and Studies on Library and Other Applications

Food

Food products often pass through a complicated process and they should be treated more carefully. In a study by Vorst et al. they have found that RFID passive tags do not influence the beef muscle (Vorst et al., 2004). Maintaining a tracking record of food products through their production as well as active packaging of meat products have been investigated briefly in Table App. E-4.

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Regattieri et al., 2007)	RFID/barcode/alpha numerical code	Food management (cheese)	Developed	Compatible with food Easy link between the food and tag No electromagnetic interaction Reduction in labour cost Speeding the physical flow Eliminating profit loss Improved quality control and monitoring Improved knowledge on customer behaviour Improved care of perishable food Improved of product recalls	Tag cost Lack of standardised RFID protocols Scanning issued in electromagnetic intense environments Barcodes and alphanumerical codes are more suitable
2	(Kerry et al., 2006)	sensor technologies, indicators and RFID	Active and intelligent packaging of meat and muscle-based products	Review	Monitor the condition of the food during transport and storage	Cost

No	Paper	Technology	Application	Scope	Advantages	Limitations
3	(Vergara et al., 2007)	RFID reader with onboard gas sensor	Fruit quality monitoring	Prototype	be low cost Small size Sensing capability to measure physical and chemical parameters Data storage and communications	

Table App. E-4: Pilots and Studies on Food

Postal and Cargo Services

The domain covered by postal and cargo services can be considered worldwide. Also the types of packaging and processing methods vary from country to country. Table App. E-5 observes how RFID has been adopted to these services.

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Zhang et al., 2006)	RFID	Postal and courier services	TNT: 6 projects implemented	Improve processing Enhancing accuracy Improved efficiency Transparency in transport and logistics	
				Australian Post: deploying		
				Finland Post: Pilot	Maximise efficiency of supply chain Effective to track re-useable assets Improved financial performance by improving customer service, profitability and reducing costs Optimising asset inventory	
				Post Denmark: Deployed	Reduce loss of cages Improve security of mail during transport Tracking roll-cage usage Improved workflow and so reduced inefficiencies	
				Swedish Postal Posten: Deployed	Reduce internal theft Enhanced security of valuables and/or expensive items	

No	Paper	Technology	Application	Scope	Advantages	Limitations
				Italy Post: Deployed	Speed sorting of parcels Accurate tracking and delivery of items	
				China Post Logistics: Deployed	Increased efficiency of express delivery Faster disposal sorting	
				UK Royal Mail and US Postal Services (USPS): Test	To improve efficiency To reduce costs and improve profit To improve assets storage	
				UPS and FedEx: Trials		Already invested in barcodes Widely using barcodes that are working fine
2	(Huang et al., 2008)	RFID	Import cargo	Study	Reduction of total inventory cost for import cargos Reduction of operators' labour costs	

Table App. E-5: Pilots and Studies on Postal and Cargo Services

Individuals' Care

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Almudevar et al., 2008)	RF, infrared motion detector, magnetic switch and pressure pad	Home monitoring	Study	Outcome validated by the home sensor network	
2	(Corchado et al., 2008)	RFID	Monitoring Alzheimer patients	Prototype	Reduced time on supervision and control and attending false alarms Nurses left with more time to spend with patients	Nurses and technology interactions Access point installation issues interfering with other signals Collocation of RFID door readers
3	(Keuchel* et al., 2004)	Barcode	Biohazardous waste reduction		Savings are expected	
4	(Meyer et al., 2006)	RFID/microchip	Disaster Victim Identification (DVI)	Prototype	Low temperature storage (-18°C) with no effect on readability Unique identification of bodies Shortening stay time in cold room, opening bags Long life spam Costs feasible with thousands	

Table App. E-6: Pilots and Studies on Individuals' Care

Healthcare Setting

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Ball et al., 2003)	barcode	Electronic prescription transfer			
2	(Bacheldor, 2007)	RFID	Patient data protection	Deploying	Ease of use	
3	(Rotondi et al., 1997)	Barcode and LAN	track the progress of patients during the preoperative process	Development	More accurate assessment of Operation Room (OR) utilisation and modified to allow the recording of OR delays Providing users with data on surgical patients	
4	(Shang-Wei et al., 2006)	RFID	Location-based tracking against SARS	Deployed		
5	(Sangwan et al., 2005)	RFID	Tracking patients, charts and medical equipment	Prototype	Faster patient check-in Locating patients while waiting Automatic tracking of patients Right drug to the right patient Infection control	
6	(O'Connor, 2006)	Ultrasound, RFID and GPS	Tracking and monitoring patients	Trial	Save staff Improve the speed and quality	
7	(Ho et al., 2005)	RFID sensor networks	Elder care	Prototype		

No	Paper	Technology	Application	Scope	Advantages	Limitations
8	(Tzeng et al., 2008b)	RFID, GPS	Location based tracking of patients Physiological monitoring and infectious control Access to and download of medical information through an ID Monitor patients' temperatures Access control of workers Patients and medical worker tracking Tracking the movement of hospital waste in transport Access control Drug dispensing SOPs being adhered to Outpatient and newborns management	Deployed	Effective communication Increased asset utilisation Enhanced patient-care process Active patient management Virtual integration of the supply chain New service strategy New business opportunities	Dramatic change both in the business processes and personnel Emergency rooms low usage Limitation of wireless technology in surgery rooms

Table App. E-7: Pilots and Studies in Healthcare Setting

In addition to the abovementioned approaches, as has been discussed by Fisher and Monahan hospital infrastructure is more complicated than what has been previously described (Fisher and Monahan, 2008).

Clinical and Laboratory Settings

No	Paper	Technology	Application	Scope	Advantages	Limitations
1	(Martin et al., 2007)	Barcode and LIMS	Sample tracking in an automated cytogenetic biodosimetry laboratory	Beta version	Improve efficiency, confidence, speed, and throughput Decreasing data-transcription errors	Large volumes of samples and their data stored in laboratory notebooks causes QC and QA issues
2	(McGiven et al., 2007)	Linear and 2D barcode, robots and LIMS	Efficient test automation and sample tracking	Development	Eliminate human error Reduced bottlenecks Increase throughput Invested capital received in three years by labour saving	Difficulties to design such that allows future developments possible
3	(Graves, 2002)	Barcode and freezers	Automated system for sample storage	Development	Sample integrity maintenance Sample quality maintenance Enhanced security and access control Improved sample and data visibility	
4	(Bettendorf et al., 2005)	RFID	Identification of Cryopreserved materials		Extreme temperature survival	
5	(Bacheldor, 2008)	RFID	Sample storage	Manufactured		

Table App. E-8: Pilots and Studies in Clinical and Laboratory Settings

Appendix F. Workflow Diagrams for Omic Procedures

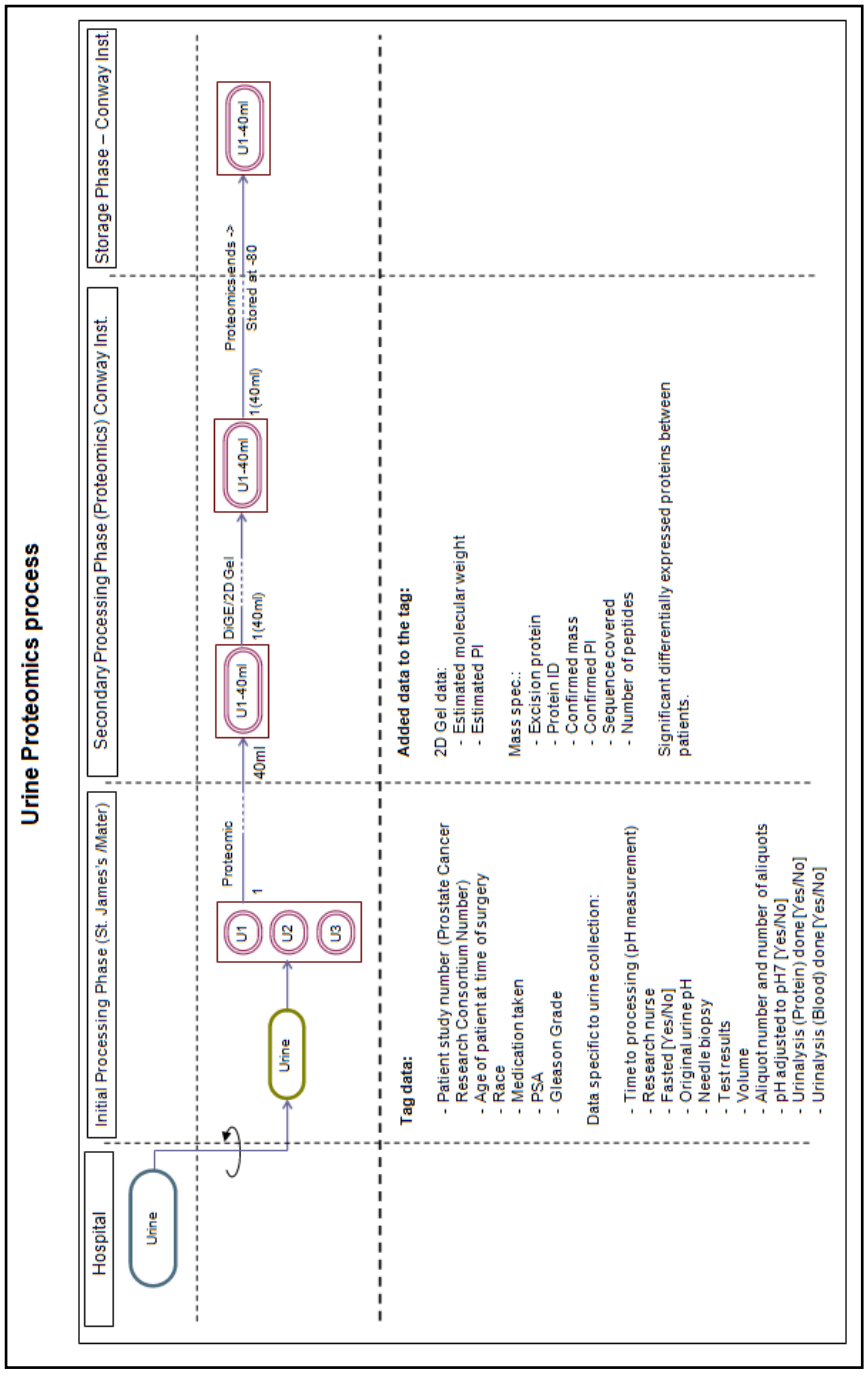


Figure App. F-1: Urine Proteomic Workflow

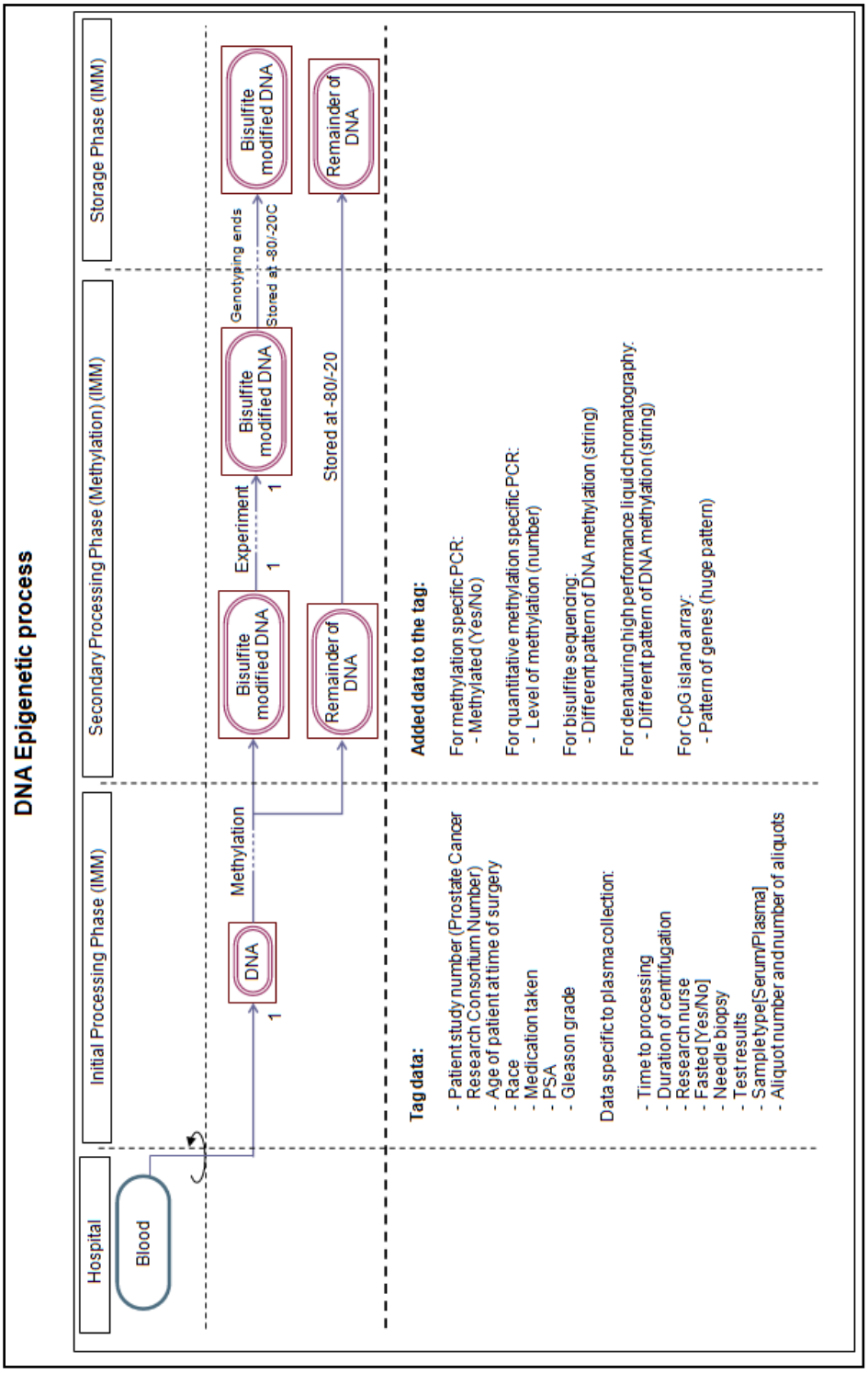


Figure App. F-2: DNA Epigenetic Workflow

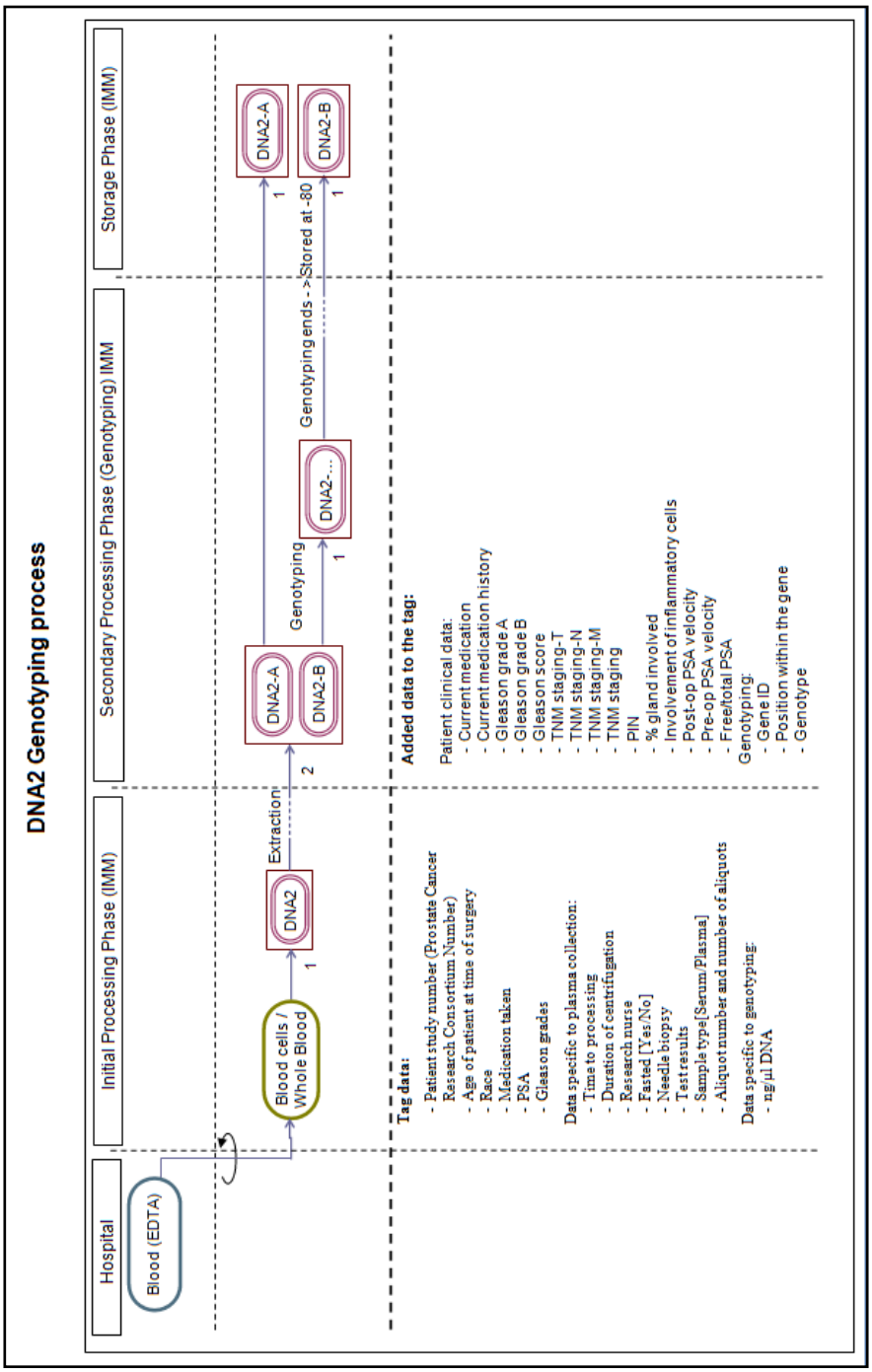


Figure App. F-3: DNA Genotyping Workflow

Appendix G. Initial Attempt to Extraction Data Elements from Procedures

Field Name	Type
PCRC number	Unique number
Age of patient at the time of surgery	2 digit number
Race	Text chosen from options
Medication taken	Text
PSA	Number
Gleason Grade	Number

Table App. G-1: General Data

Field Name	Type
Time to processing	2 digit number
Duration of centrifugation	1-2 digit number
Research nurse	Text
Fasted	Yes/No
Needle biopsy	Yes/No
Test results	Text
Sample type	Serum/Plasma
Aliquot number and number of aliquots	Number/Number

Table App. G-2: Serum and Plasma Collection Data

Field Name	Type
Time to processing (pH measurement)	2 digit number
Research nurse	Text
Fasted	Yes/No
Original urine pH	Number
Needle biopsy	Yes/No
Test results	Text
Volume	Number
Aliquot number and number of aliquots	Number/Number
pH adjusted to pH7	Yes/No
Urinalysis (Protein) done	Yes/No
Urinalysis (Blood) done	Yes/No

Table App. G-3: Urine Collection Data

Field Name	Type
2D Gel data	
Estimated molecular weight	Number
Estimated PI	Number
Mass spec.	
Excision protein	Text
Protein ID	Number
Confirmed mass	Number
Confirmed PI	Number
Sequence covered	Text
Number of peptides	Number
Significant differentially expressed proteins between patients	

Table App. G-4: Urine and Serum Proteomic Process Data

Field Name	Type
For methylation specific PCR	
Methylated	Yes/No
For quantitative methylation specific PCR	
Level of methylation	Number
For bisulfite sequencing	
Different pattern of DNA methylation	Text
For denaturing high performance liquid chromatography	
Different pattern of DNA methylation	Text
For CpG island array	
Pattern of genes	Large text

Table App. G-5: DNA Epigenetic Process Data

Field Name	Type
Patient clinical data	
Current medication	Text
Current medication history	Text
Gleason grade A	Text
Gleason grade B	Text
Gleason score	Text
TNM staging-T	Text
TNM staging-N	Text
TNM staging-M	Text
TNM staging	Text
PIN	Text
% gland involved	Number
Involvement of inflammatory cells	Text
Post-op PSA velocity	Text
Pre-op PSA velocity	Text
Free/total PSA	Text
Genotyping	
ng/µl DNA	Text
Gene ID	Text
Position within the gene	Text
Genotype	Text

Table App. G-6: DNA2 Genotyping Process Data

Appendix H. Final Database Tables for SITS

ISampleA

Field Name	Data Type	Description
AID	AutoNumber	Indexing Number
TimeStamp	Date/Time	Time of Last Sync
Version	Number	Indicate Memory Format
UID	Number	User who updated/edited the tag
Synced	Yes/No	will be Yes if it's synchronized
IsNew	Yes/No	will be Yes if the Tag is new
TAGID	Text	RF-TAG ID
InfoMem	Memo	Entire RF-TAG Memory
PCRCNO	Number	PCRC NO
SOP	Number	SOP version
TTP	Number	Time to process
TCAD	Text	Temperature change and duration
CT	Number	Control temperature
STSU	Text	Sample type, serum/urine
ANNOA	Number	Aliquot number/number of aliquots
VOL	Number	Volume
HID	Text	Handler ID
LOC	Text	Location
DT	Number	Date

Figure App. H-1: “ISampleA” Table

ISampleB

Field Name	Data Type	Description
AID	AutoNumber	
TAGID	Text	Tag ID
DTT	Number	Date taken
SOP	Number	SOP version
VTL	Text	Volume taken/left
TCAD	Text	Temperature change and duration
HID	Text	Handler ID
NTID	Text	New tag ID

Figure App. H-2: “ISampleB” Table

DSample

Field Name	Data Type	Description
AID	AutoNumber	Indexing Number
TimeStamp	Date/Time	Time of Last Sync
Version	Number	Indicate Memory Format
UID	Number	User who updated/edited the tag
Synced	Yes/No	will be Yes if it's synchronized
IsNew	Yes/No	will be Yes if the Tag is new
TAGID	Text	RF-TAG ID
InfoMem	Memo	Entire RF-TAG Memory
PCRCNO	Number	PCRC NO
PTID	Text	Previous tag ID
NTID	Text	New tag ID
SOP	Number	SOP version
TTP	Number	Time to process
TCAD	Text	Temperature change and duration
CT	Number	Control temperature
STSU	Text	Temperature change and duration
ANNOA	Number	Aliquot number/number of aliquots
VOL	Number	Volume
HID	Text	Handler ID
LOC	Text	Location
CDT	Number	Date (Collection)
SDT	Number	Date (Sample Data)
SOPV	Number	SOP version (Sample Data)
OA	Number	Omic Analysis
STCAD	Text	Temperature change and duration
SCT	Number	Control temperature
VTL	Text	Volume taken/left
SLOC	Text	Location (Sample Data)
SHID	Text	Handler ID (Sample Data)
SRFRC	Memo	Sample removed from row,column;row,column;....

Figure App. H-3: “DSample” Table

TagIDs

Field Name	Data Type	Description
TAGID	Text	
IsInitial	Yes/No	
ToDelete	Yes/No	
ToUpdate	Yes/No	
WrtID	Number	
DelID	Text	

Figure App. H-4: “TagIDs” Table

Users

Field Name	Data Type	Description
AID	AutoNumber	Indexing Number
AccessLevel	Number	Access Rights
User	Text	Login Username
Pass	Text	Login Password
PID	Text	Nurse ID
Location	Text	

Figure App. H-5: “Users”

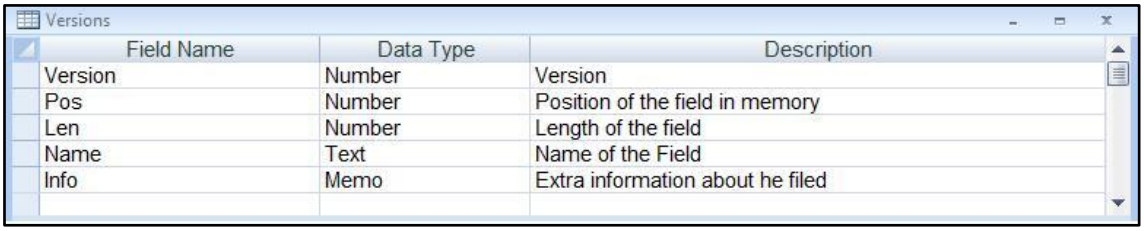


Figure App. H-6: “Versions”

Appendix I. Tags Survival Experiment

Experiment 1:

Aim: Testing the adhesive operation in extreme cold and number of freeze-thaw cycles at -190°C freezers.

Method: Tags attached to tubes using the adhesive. Tubes are filled with water to maintain the cold from freezer to the reader device. Actions are carried out as described in Table App. I-1.

Results: Adhesive found to be persistent against cold and freeze-thaw cycles as described in its specification document.

Date	Action	Condition
23/07/2008 - 16:30	4 tags attached to the adhesive	All 4 tags attached
24/07/2008 - 16:40	4 tags and adhesive attached to tube lids (adhesives are double sided)	1 tag is a little loose
25/07/2008 - 16:40	4 tagged tubes stored at -80° C freezer	They all contain water
28/07/2008 - 16:40	4 tagged tubes removed from freezer to room temperature	1 tag is a little loose
	4 tagged tubes washed in hot water	All tags readable and attached
28/07/2008 - 16:45	4 tagged tubes stored at -80° C freezer	They all contain water
05/08/2008 - 16:45	4 tagged tubes removed from freezer to room temperature	1 tag is a little loose, all tags readable and attached

Table App. I-1: Details of experiment 1

Experiment 2:

Aim: Testing the adhesive operation and tags survival in extreme cold and number of freeze-thaw cycles at -190°C Liquid Nitrogen vapour.

Method: 5 tagged tubes were stored in LN vapour tank from 23rd of July 2009 to 30th.

Results: All 5 tags were affixed to tubes. All 5 tags were readable and rewriteable.

Experiment 3:

Aim: Testing the adhesive operation and tags survival in extreme cold and number of freeze-thaw cycles at -190°C Liquid Nitrogen vapour.

Method: 6 sets of 4 tagged tubes were stored for various periods of times and different number of freeze-thaw cycles as described below:

Set 1 - 1 month LN2 Vapour

Set 2 - Placed in freezer Monday and freeze/thawed Tue, Wed, Thr, Fri (4 cycles)

Set 3 - Placed in freezer Monday and freeze/thawed Wed, Thr, Fri (3 cycles)

Set 4 - Placed in freezer Monday and freeze/thawed Thr, Fri (2 cycles)

Set 5 - Placed in freezer Monday and freeze/thawed Fri (1 cycle)

Results: The result of each set is outlined in Table App. E-1. Set 1 and set 2 are still stored in the tank. Tags retrieved from set 3-6 are all readable. Tags from set 3 with 4 freeze-thaw cycles are still rewriteable.

Date	Action	Outcomes
Monday 17/08/2009	Sets 1 – 5 stored in LN vapour tank	All 4 tags for each set attached
Tuesday 18/08/2009	Set 2 thawed and frozen	All 4 tags for each set attached
Wednesday 19/08/2009	Set 2 and 3 thawed and frozen	All 4 tags for each set attached
Thursday 20/08/2009	Set 2, 3 and 4 thawed and frozen	All 4 tags for each set attached
Friday 21/08/2009	Set 2, 3, 4 and 5 thawed and frozen	All 4 tags for each set attached
Monday 24/08/2009	Set 2, 3, 4 and 5 removed from the tank	All 4 tags for each set attached, except for one tag fell off from Set 4. All readable and rewriteable
Thursday 17/09/2009	Set 1 removed from the tank	All 4 tags attached, readable and rewriteable

Table App. I-2: Details of Experiment 3

Appendix J. Images from SITS

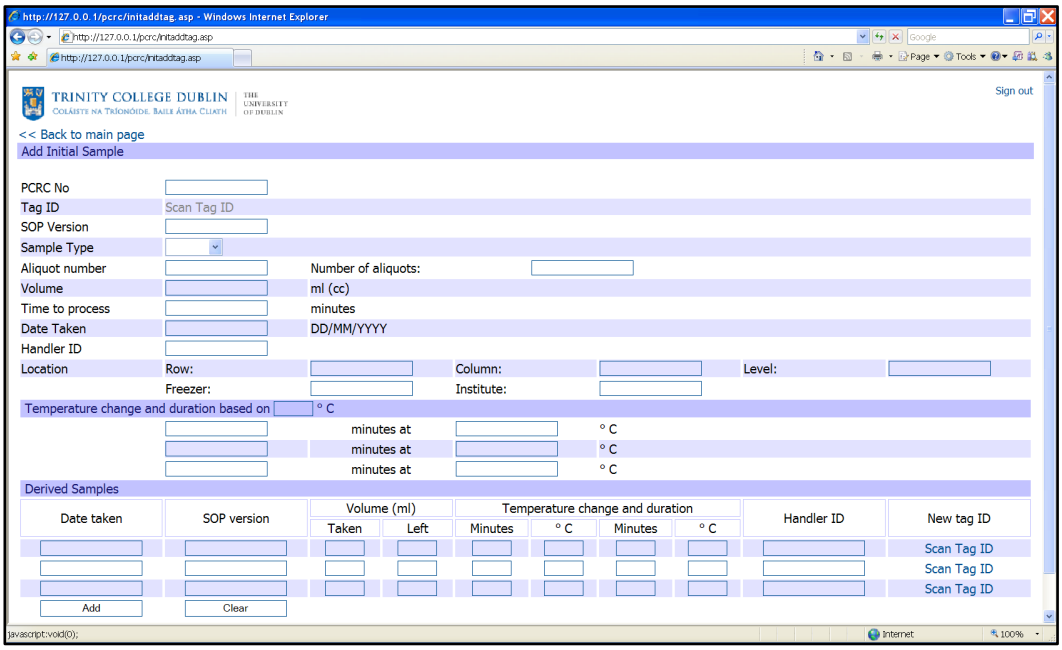


Table App. J-1: Adding Initial Sample Page

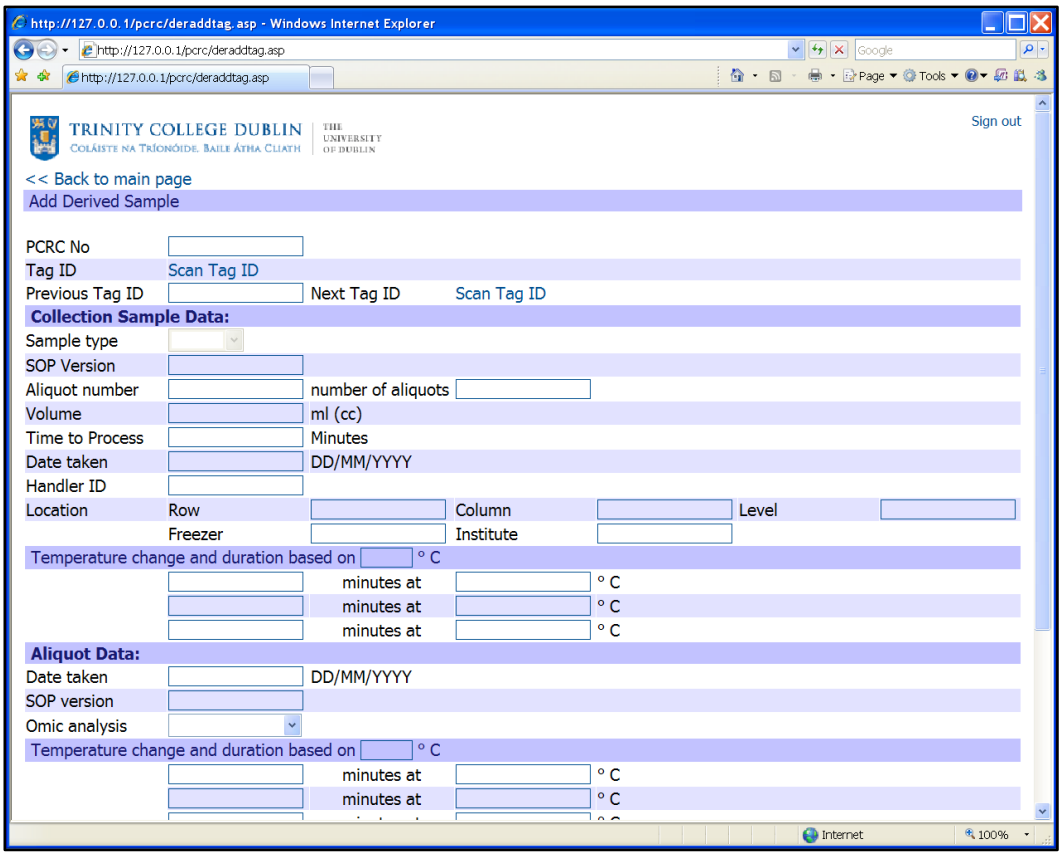


Table App. J-2: Adding Derived Sample Page

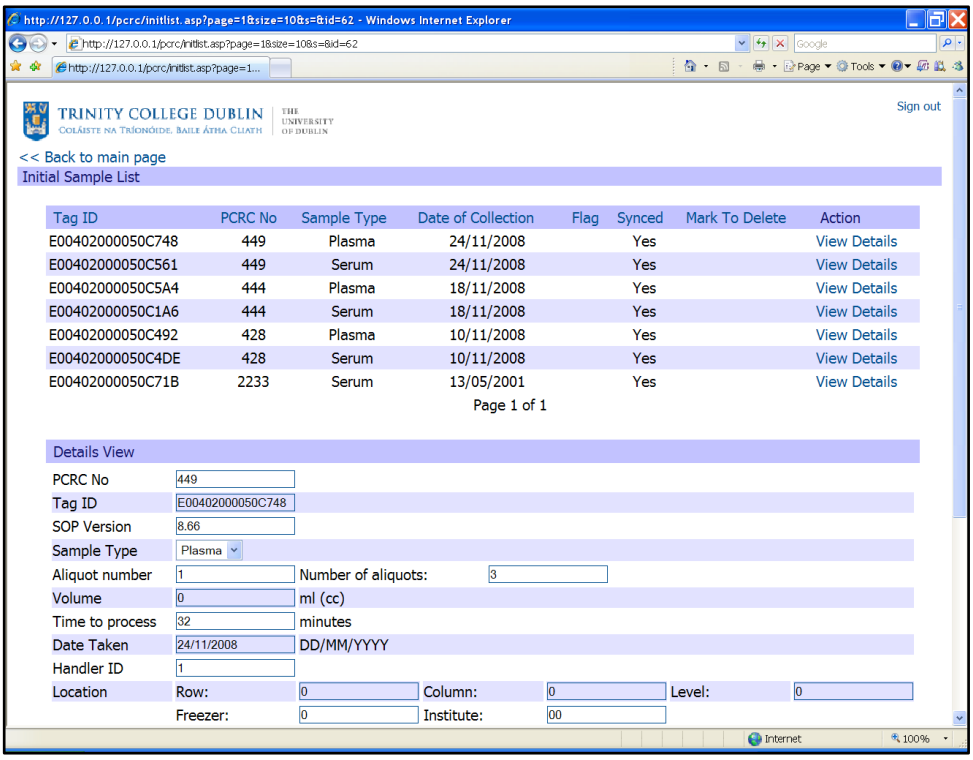


Table App. J-3: Viewing Initial Sample Record Page

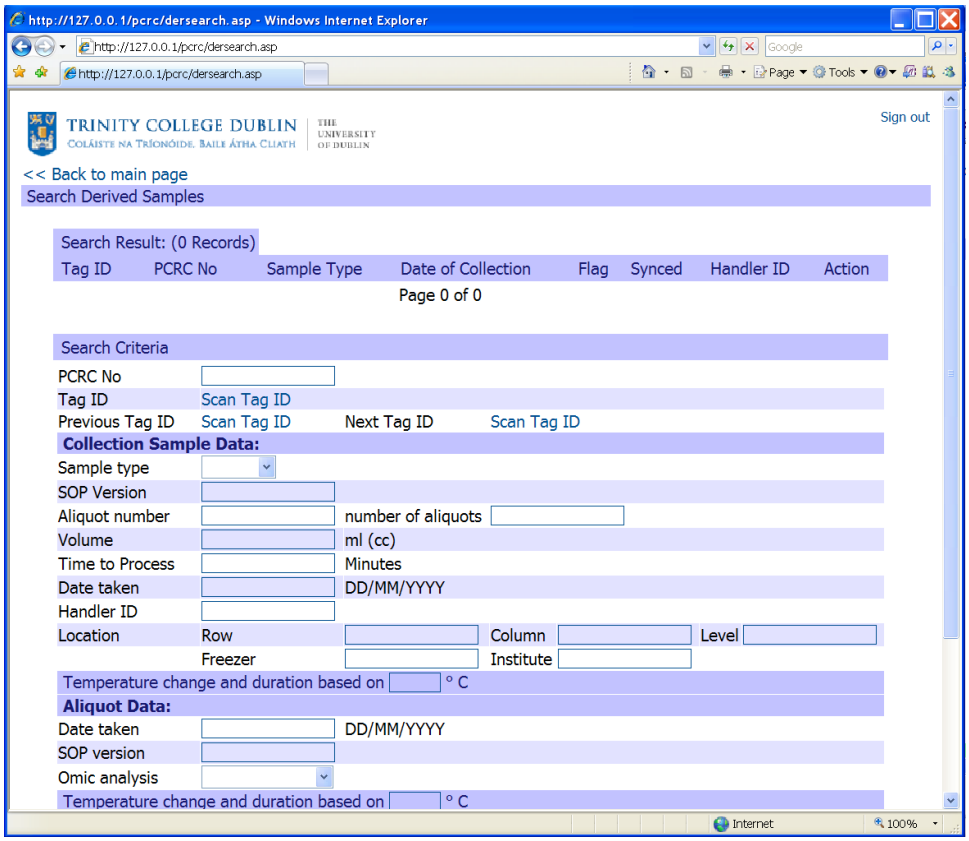


Table App. J-4: Searching Derived Sample Page

Appendix K. Data Fields

	Name	Length (Byte)	Position			Format	DB Field Type	Searchable	
Header	Fields Structure	1	1			D	N/A		Protocol
		1	2			D	Integer		Version
	Time Stamp [Time]	2	3			HHMM	Date/Time		TimeStamp
	Time Stamp [Date]	4	5			YYYYMMDD			TimeStamp
	Person who wrote on tag	2	9			DD	Integer		UID
Compulsory Data	PCRC NO	4	11			DDDD	Integer	x	PCRCNO
	Tag ID	8	N/A			XXXXXXXXXXXXXXXX	Text:20	x	TAGID
Collection Data	SOP version	3	15			D[.]DD	Single	x	SOP
	Time to process	2	18			DD	Integer	x	TTP
	Temp. change & duration	12	20			TT±DDTT±DDTT±DD	Text:20		TCAD
	Control temperature	1	32			±D	Byte	x	CT
	Sample type: S/U/P	1	33			C (S/P/U)	Text:02	x	STSU
	Aliq. number /number of aliq.	3	34			D/D	Single	x	ANNOA
	Volume	3	37			DDD	Integer	x	VOL
	Handler ID	3	40			ALPHANEUMERICAL	Text:05	x	HID
	Location	6	43			DDDDDD	Text:10	x	LOC
	Date	6	49			[YY]YYMMDD	Long Integer	x	DT
Derived Samples Data (x N)	Date taken	6	55	90	125	[YY]YYMMDD	Long Integer	x	DTT
	SOP version	3	61	96	131	D[.]DD	Single	x	SOP
	Volume taken/left	7	64	99	134	DDD/DDD	Text:10		VTL
	Temp. change and duration	8	71	106	141	TT±DDTT±DD	Text:10		TCAD
	Handler ID	3	79	114	149	ALPHANEUMERICAL	Text:05	x	HID
	New tag ID	8	82	117	152	XXXXXXXXXXXXXXXX	Text:20	x	NTID
	Sum:	35							
	x:	3							
Total		158							

Figure App. K-1: Initial Sample

	Name	Length (Byte)	Position			Format	Dbase Field Type	Searchable	
Header	Fields Structure	1	1			D	N/A		Protocol
		1	2			D	Integer		Version
	Time Stamp [Time]	2	3			HHMM	Date/Time		TimeStamp
	Time Stamp [Date]	4	5			YYYYMMDD			TimeStamp
	Person who wrote on tag	2	9			DD	Integer		N/A
Compulsory Data	PCRC NO	4	11			DDDD	Integer	x	PCRCNO
	Tag ID	8	N/A			XXXXXXXXXXXXXXXX	Text:20	x	TAGID
	Previous tag ID	8	15			XXXXXXXXXXXXXXXX	Text:20	x	PTID
	New tag ID	8	23			XXXXXXXXXXXXXXXX	Text:20	x	NTID
Collection (can be omitted)	SOP version	3	31			D[.]DD	Single	x	SOP
	Time to process	2	34			DD	Integer	x	TTP
	Temp. change and duration	12	36			TT±DDTT±DDTT±DD	Text:20		TCAD
	Control temperature	1	48			±D	Byte	x	CT
	Sample type, S/U/P	1	49			C (S/P/U)	Text:02	x	STSU
	Aliq. number /number of aliq.	3	50			D/D	Single	x	ANNOA
	Volume	3	53			DDD	Integer		VOL
	Handler ID	3	56			ALPHANEUMERICAL	Text:05	x	HID
	Location	6	59			DDDDDD	Text:10	x	LOC
	Date	6	65			[YY]YYMMDD	Long Integer	x	CDT
Sample Data (Aliquot)	Date	6	71			[YY]YYMMDD	Long Integer	x	SDT
	SOP version	3	77			D[.]DD	Single	x	SOPV
	Omic Analysis	1	80			D	Byte	x	OA
	Temp. change and duration	12	81			TT±DDTT±DDTT±DD	Text:20	x	STCAD
	Control temperature	1	93			±D	Byte		SCT
	Volume taken/left	7	94			DDD/DDD	Text:10		VTL
	Location	6	101			DDDDDD	Text:10	x	SLOC
	Handler ID	3	107			ALPHANEUMERICAL	Text:05	x	SHID
Group Data	Sample removed from row	1	110	112	114	D	Byte		SRFRC
	Sample removed from column	1	111	113	115	D	Byte		SRFRC
	Sum:	2							
	x:	20							
Total		148							

Figure App. K-2: Derived Sample

Appendix L. User Guides for the Research Nurse and Scientist

Sample Identification and Tracking System (SITS)

USER GUIDE FOR RESEARCH NURSES

May 2009

1. BACKGROUND

The Sample Identification and Tracking System (SITS) is based on Radio Frequency Identification (RFID) technology. RFID uses electronic tags to store data on the physical object, in this application sample tubes.

Apart from tags, reader/writer device and its antenna is also needed to communicate with the tag and transmit data. This communication is made possible by an application running in the background of the computer that the reader/writer is connected. However, users only need to use the web interface as will be described in this document.

2. LOGIN PAGE

Access to SITS is made available through web on http://yuriko.cs.tcd.ie/pcrc/ by username and passwords that are provided to users in advance. Figure 1-1 shows the log in section.

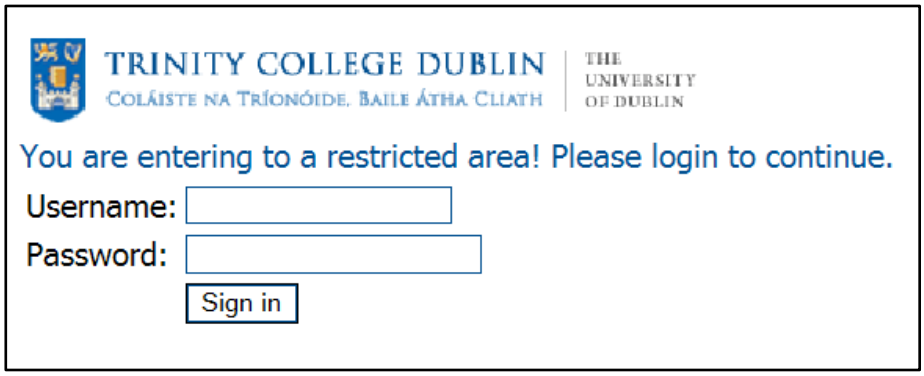


Figure 1-1: log in section

3. HOMEPAGE

The three main sections on the homepage after the user has successfully logged in are: *User Management*, *Initial Samples* and *Status Monitor*, as shown in Figure 2-1.

User Management: There is only one page available in this section, i.e. *Change Password*. In this page the user can change his/her password by giving his/her current password.

Initial Samples: This section allows operations to be carried out on samples that are just collected from the participant. The main operations supported are *List Samples*, *Search*

Samples, *Read Sample tag*, *Add New Sample*, *Update Sample tag* and *Edit/Delete Samples.*

Status Monitor: This section displays the status online reader devices, i.e. *Available Readers* and *Online Users* with their roles. If a reader device is available *Online* is displayed beside its name.

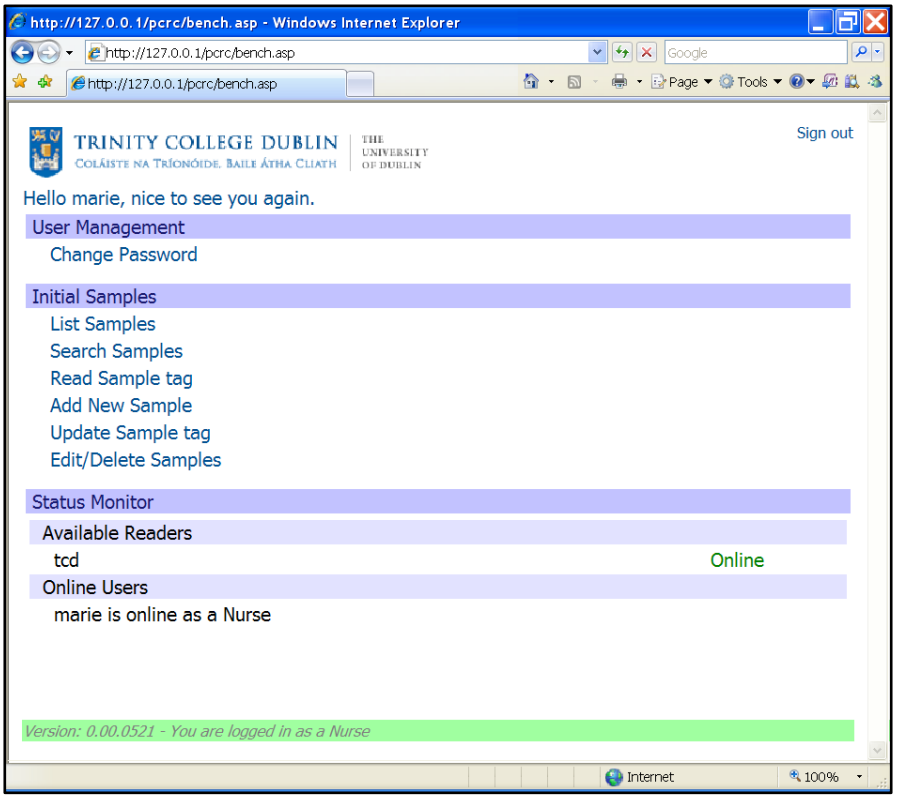


Figure 2-1: Homepage when the user has logged in

4. *Initial Samples* SECTION

The main six operations that can be done on an initial sample are:

List Samples: This page displays all the available samples in the SITS database and is used for browsing or searching the database.

Search Samples: This page allows search on any data field of the database. If a tagged tube is available its ID can be scanned to be searched for in the database.

Read Sample tag: This page allows reading the memory of the tag

Add New Sample: When a new sample is collected from the participant, a new record should be created on the database and sample's data should be stored on it.

Update Sample tag: This page maintains the tag ID and data, if the tag is not available, for future.

Edit/Delete Samples: This page allows editing or deleting a record. In case of deleting the tag should be in the reading range of the antenna. If not available, it will be updated later through *Update Sample tag* page. In case of deleting a record, the operation will be confirmed by the administrator of the SITS.

Figure 3-1 shows the screenshot of *Add New Sample* page. Similar style is maintained for the rest of the pages. In this page the tag ID of the original tube is scanned by clicking on the *Scan Tag ID*. Other data fields need to be typed in by the user. These fields will be looked up for the three aliquots taken from this tube of the sample. Note: The collection tube is not tagged in the current settings of PCRC biobank, so a tag is scanned to fill this field, but is not attached to any tube. The tag IDs of the aliquots should also be automatically captured by clicking the *Scan Tag ID* links in their relevant row. Date taken, SOP version and other data for each aliquot can be stored separately.

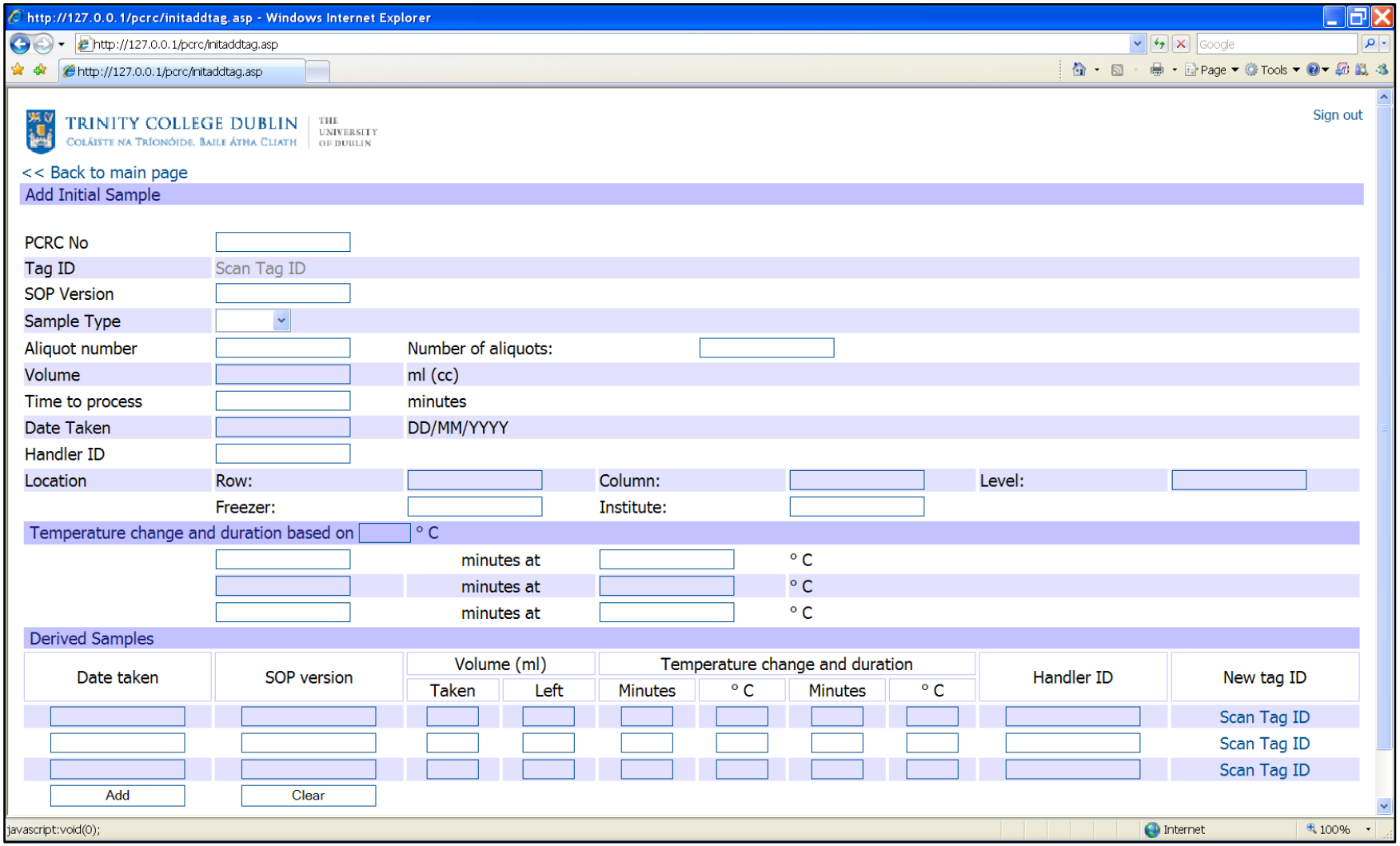


Figure 3-1: *Add Initial Sample* page

Sample Identification and Tracking System (SITS)

USER GUIDE FOR RESEARCHERS

May 2009

1. BACKGROUND

The Sample Identification and Tracking System (SITS) is based on Radio Frequency Identification (RFID) technology. RFID uses electronic tags to store data on the physical object, in this application sample tubes.

Apart from tags, two other equipment are needed to communicate with the tag and transmit data. They are reader/writer device and its antenna. This communication is made possible by an application running in the background of the computer that the reader/writer is connected to. However, users only need to use the web interface as will be described in this document.

2. LOGIN PAGE

Access to SITS is made available through web on http://yuriko.cs.tcd.ie/pcrc/ by username and passwords that are provided to users in advance. Figure 2-1 shows the log in page.

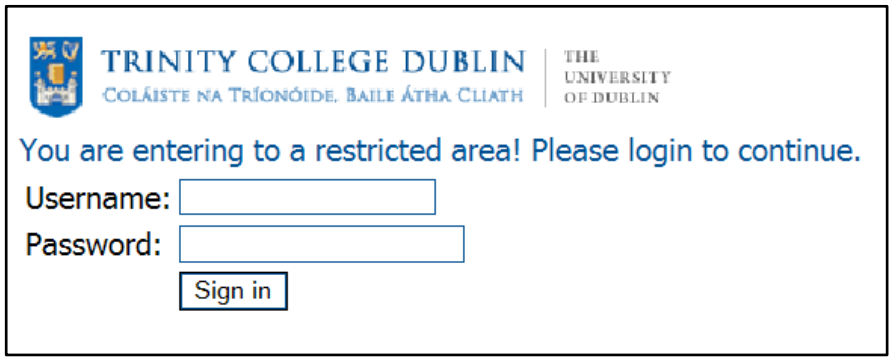


Figure 2-1: log in section

3. HOMEPAGE

The four main sections on the homepage after the user has successfully logged in are: *User Management*, *Initial Samples*, *Derived Samples* and *Status Monitor*, as shown in Figure 3-1.

User Management: There is only one page available in this section, i.e. *Change Password*. In this page the user can change his/her password by giving his/her current password (Figure 3-2).

Initial Samples: This section allows operations to be carried out on samples that are just collected from the participant. Researchers can only access the *List Samples*, *Search Samples*, *Read Sample tag* and *Update Sample tag* pages.

Derived Samples: This section is not available to research nurses and is designed to allow operations on derived samples or aliquots. These operations are *List Samples*, *Search Samples*, *Read Sample tag*, *Add New Sample*, *Update Sample tag* and *Edi/Delete Samples*.

Status Monitor: This section displays the status of the online reader devices, i.e. *Available Readers* and *Online Users* with their roles. If a reader device is available and is turned on, *Online* is displayed beside its name.

Please note that by clicking on the blue heading of each section the menu for that section will be minimised.

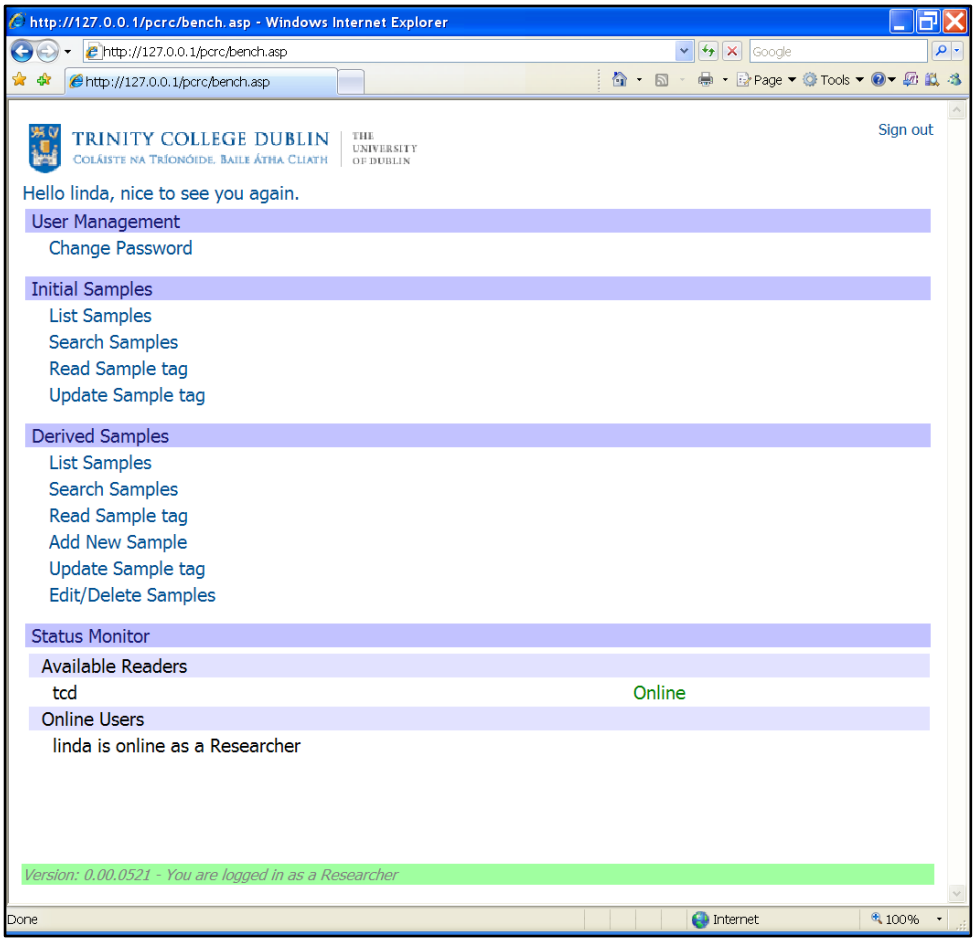


Figure 3-1: Homepage when the user has logged in

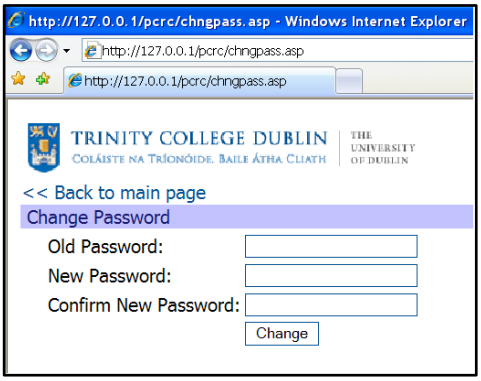


Figure 3-2: Change password form

4. Initial Samples SECTION

The main four operations that can be done by researchers on an initial sample are described as follows.

List Samples: This page displays all the available samples in the SITS database and is used for browsing or searching the database. By clicking on *View Details* data that are already stored for each record will be displayed below the list (Figure 4-1).

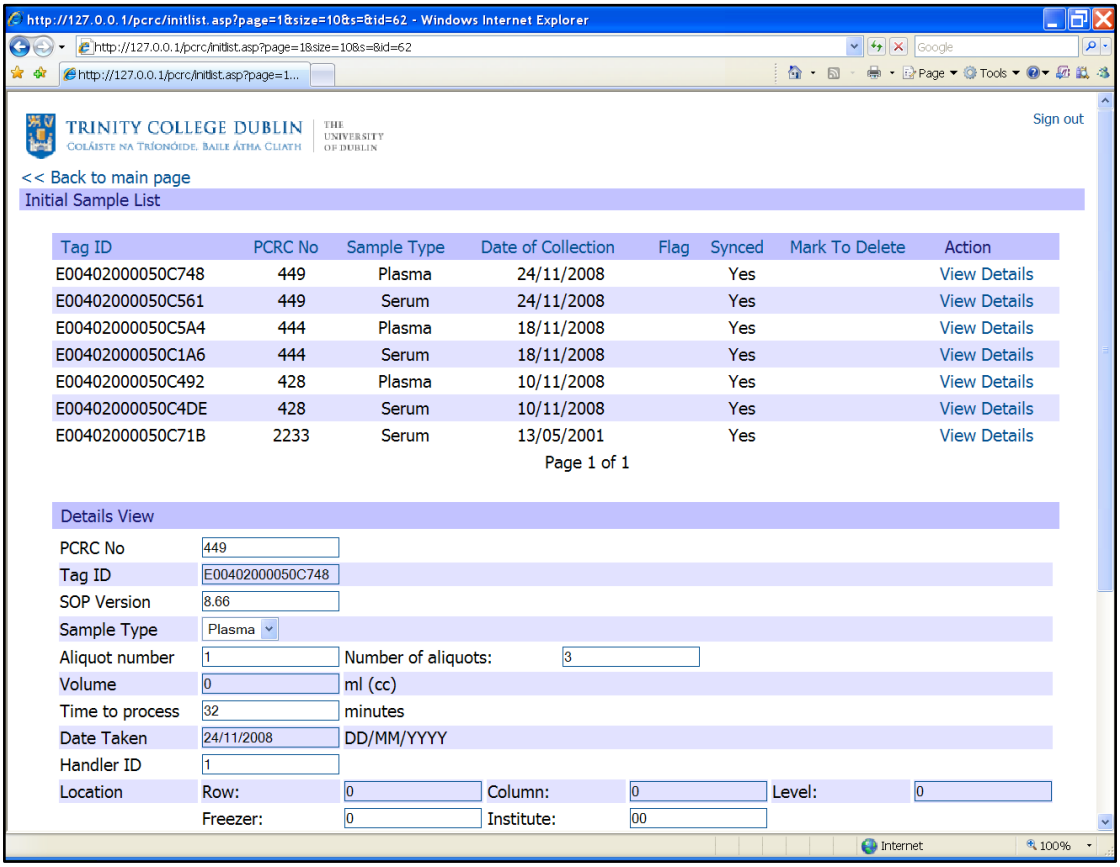


Figure 4-1: *Initial Sample List* page

Search Samples: This page allows search queries to be carried out on any data field of the database. A list of samples is displayed followed by the form for search query. If a tagged tube is available its ID can be scanned to be searched for in the database (Figure 4-2).

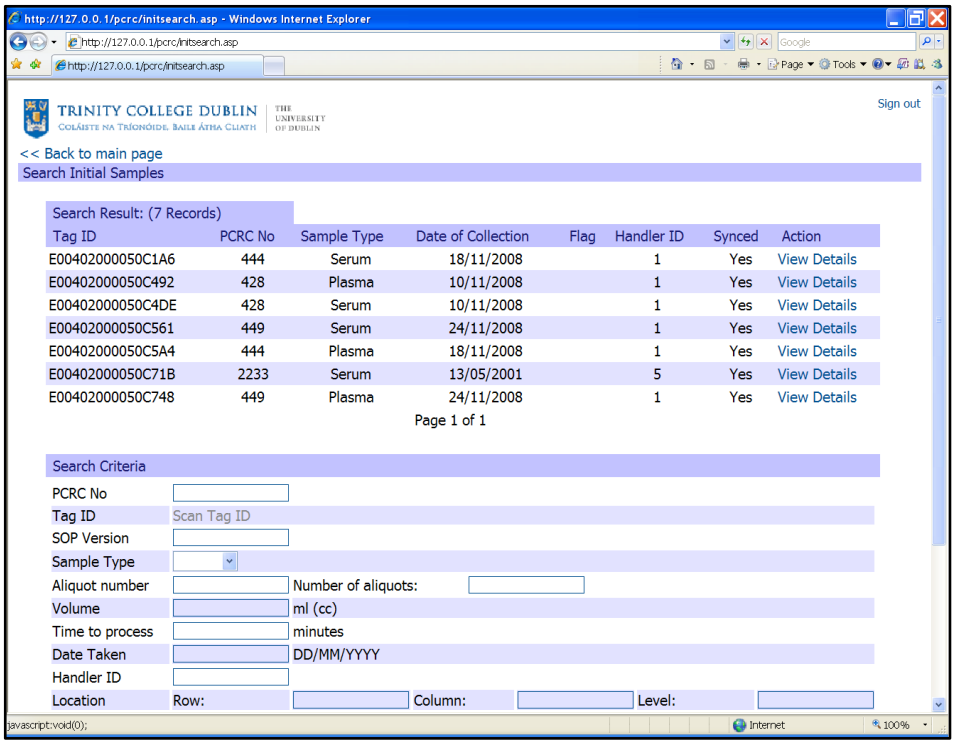


Figure 4-2: *Search Initial Samples* page

Read Sample tag: This page allows reading the data stored on the tag (Figure 4-3). Depending on the availability of the data on the database and other criteria an appropriate message and data records will be displayed. It also identifies if a sample is initial or derived.

Update Sample tag: If the tag is not available at the time of writing data, its information is maintained in this page. In the future whenever the tag is in within the reading range, its data are updated. This page operates the same, regardless of the type of the tag and sample, whether initial or derived sample.

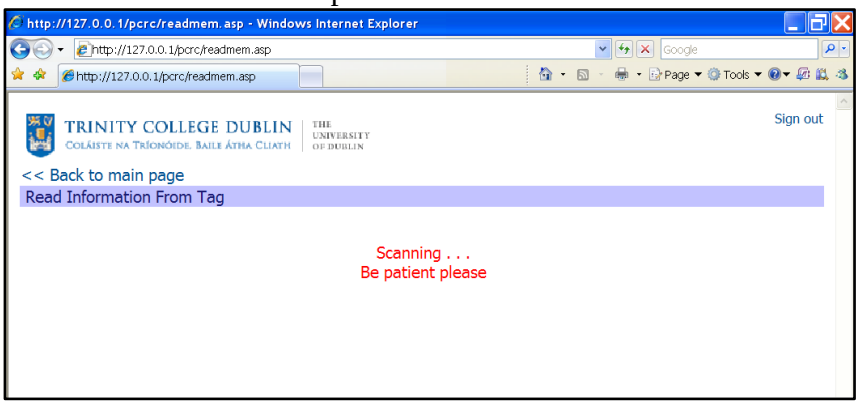


Figure 4-3: *Read Information From Tag* page

5. *Derived Samples* SECTION

The main six operations that can be done on a derived sample are:

List Samples: This page displays all the available derived samples (aliquots) in the SITS database and is used for browsing or searching the database. It is important to note that no records of these samples is made available in this page, until they reach research institute and a researcher reads data from the tag and adds it to the database. The layout of this page is very similar to Figure 4-1, however since none of the tags in that Figure is processed yet, this list is empty in Figure 5-1.

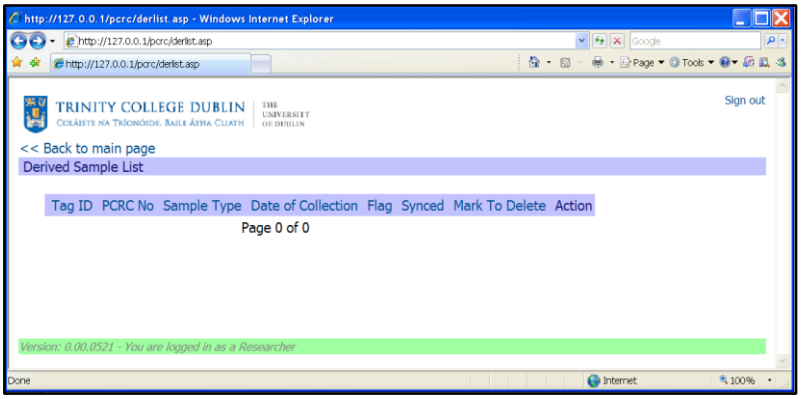


Figure 5-1: *Derived Sample List* page

Search Samples: This page allows search on any data field of the database. If a tagged tube is available its ID can be scanned to be searched for in the database. Please note that in the context of derived samples, *Previous Tag ID* and *Next Tag ID* refer to the tag IDs of the tube that the sample had previously being stored in, and the tag ID of the tube that the sample will be moved to, respectively. The search query form is shown in Figure 5-2.

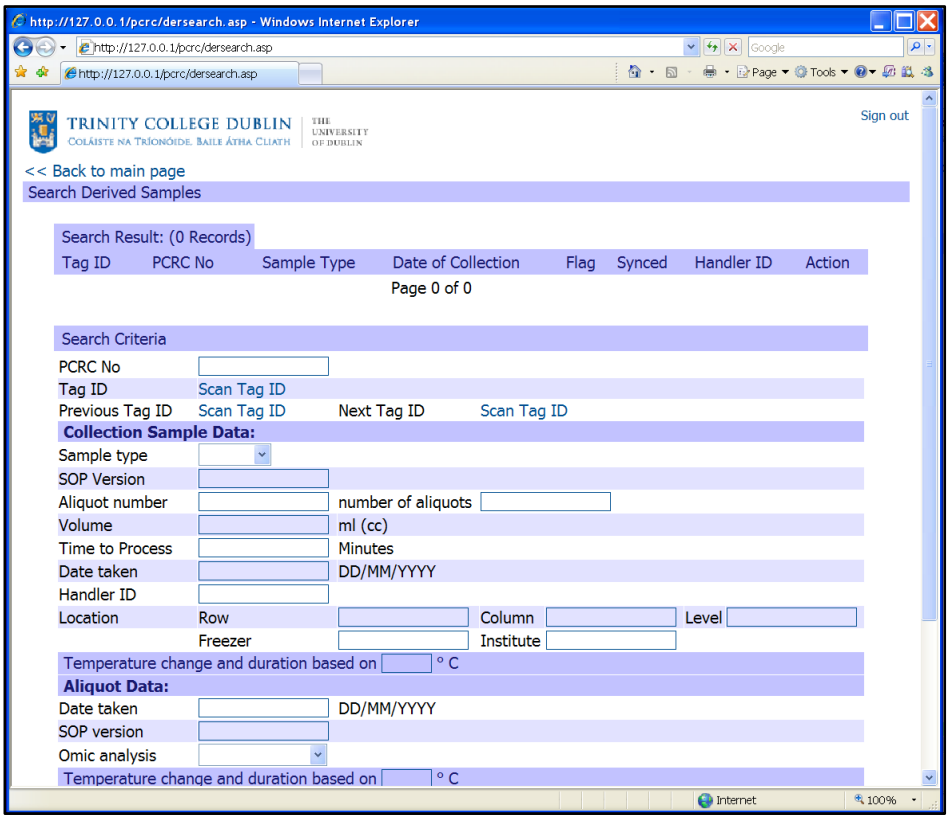


Figure 5-2: *Search Derived Samples* page

Read Sample tag: This page allows reading the memory of the tag and its operation is identical to the page for initial samples.

Update Sample tag: If the tag is not available at the time of writing data, its information is maintained in this page. In the future whenever the tag is in within the reading range, its data are updated. This page operates in exactly the same way as it would do in *Initial Samples* Section.

Edit/Delete Samples: This page allows editing or deleting a record and needs the tag to be in the reading range of the antenna. If not available, it will be updated later through *Update Sample tag* page. In case of deleting a record, the operation will be confirmed by the administrator of the SITS.

Add New Sample: In this page, Figure 5-3, the tag ID of the tube is scanned by clicking on the *Scan Tag ID* opposite the *Tag ID* label. *Scan Tag ID* opposite the *Next Tag ID* label is used when the sample is moved to another tube. Data from the initial sample matching this aliquot will be looked up automatically. Derived sample related data

fields need to be typed in by the user. These fields will be looked up for the aliquots taken from this tube of the sample.

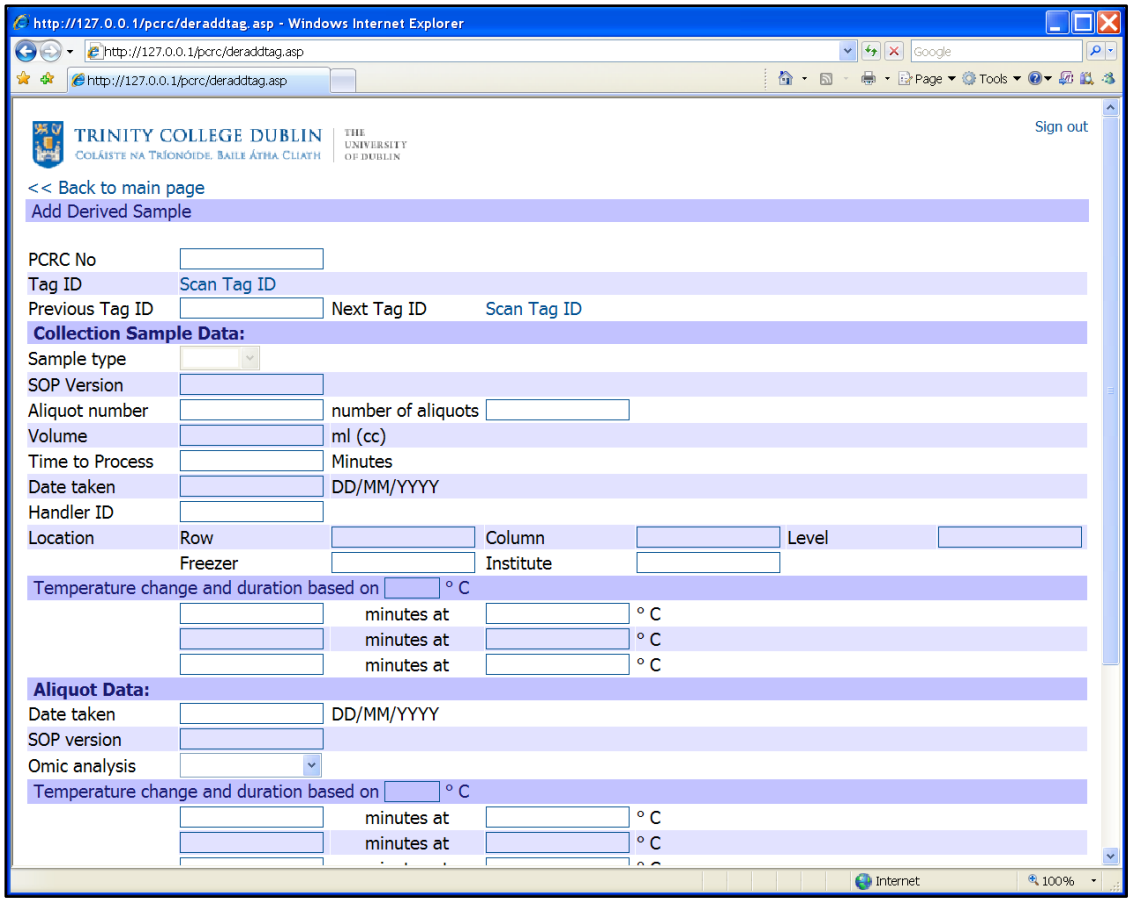


Figure 5-3: *Add Derived Sample* page

Appendix M. Form for Evaluation of the Initial Processing Phase

How would you rate the following		
Criteria	**score out of 10**	**Comments**
Interface		
Ease of use (user-friendlyness)		
Covering what is needed		
Meeting requirements		
Usability		
Functionality		
Any comments/suggestions		

Thank you

Appendix N. Secondary Processing Phase Evaluation

STANDARD OPERATING PROCEDURE

Title: Performing Real-time PCR Analysis

Materials Required Before Starting

1. High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, part number – 4368814)
2. Molecular grade nuclease free water (Sigma, part number – W4502)
3. Gene expression universal master mix (Applied Biosystems – 4304437)

Keep RNA samples and all reagents on ice at all times.

Preparation of cDNA

1. Thaw the cDNA reverse transcription kit on ice. While the kit is thawing label the PCR tube and clean the bench with ethanol/RNaseZap.

2. Prepare RNA dilution.

   200-500ng of RNA made up to a final volume of 25µl is required. Calculate this by dividing 200ng/500ng by the RNA yield (from Nanodrop Spectrophotometer reading). For example if the RNA yield is 100.2ng/µl the final dilution for a 200ng concentration is 200/100.2 which equates to 2.0µl diluted with 23µl of nuclease free water)

3. Briefly vortex and centrifuge the RNA sample before combining with dH_20.

4. Denature the RNA dilution by incubating at 70°C for 10 minutes (PCR programme: [ELAINA]/RNA-DNTR]

5. Cool sample on ice for 5 minutes.

6. Prepare the reverse transcription master-mix as follows

Reagent	1 x (for 25µl vol)	1 x (For a 10µl vol)
10 x RT reaction buffer	5	2.0
25 x dNTP Mix (100mM)	2	0.8
10 x RT Random Primers	5	2.0
MultiScribe RT	2.5	1.0
Nuclease Free Water	10.5	3.2
	25.0	**10.0**

7. For a 50µl final volume combine 25µl of master-mix with the 25µl RNA dilution (For a 20µl final volume conmine 10µl of master-mix with 10µl of RNA dilution)

8. Vortex the master-mix/RNA dilution briefly and retain on ice.

9. PCR the sample under the following conditions:

	Step 1	Step 2	Step 3	Step 4
Temperature	25°C	37°C	85°C	4°C
Time	10 min	120 min	5 sec	-

10. cDNA samples can be stored at -20°C until required.

Preparation of RT-PCR Assay to Quantify Gene Expression

7. Thaw all reagents and cDNA on ice before use. Vortex and briefly centrifuge cDNA.

8. Prepare the appropriate volume of PCR master mix (for either 10µl or 50µl final volume) according to the following:

Reagent	For 10µl final rxn
Universal Mastermix	5
PDAR (the gene!)	0.5
SterileWater	2
cDNA	**2.5**

9. Transfer 7.5µl of reaction mix to a 96well Taqman plate and record where each sample is placed.

10. Add 2.5µl cDNA to each well. Note: perform each reaction in triplicate.

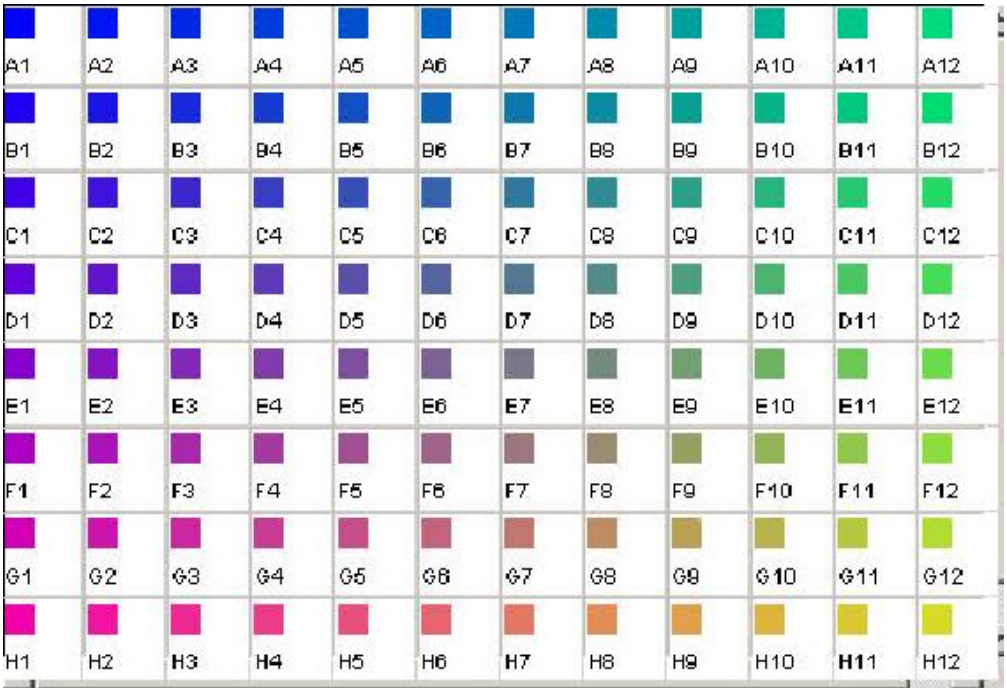
11. Cover the plate with an optical adhesive cover and ensure it is secured tightly to the edges of the plate by pressing down firmly with a plastic card.

12. Centrifuge the plate briefly to spin the contents and eliminate any air bubbles from solution.

13. Place the 96 well reaction plate on the 7500 Taqman system.

14. Perform Absolute Quantitation/Relative Quantitation as required.

96 well plate format



Appendix O. Distiller Database

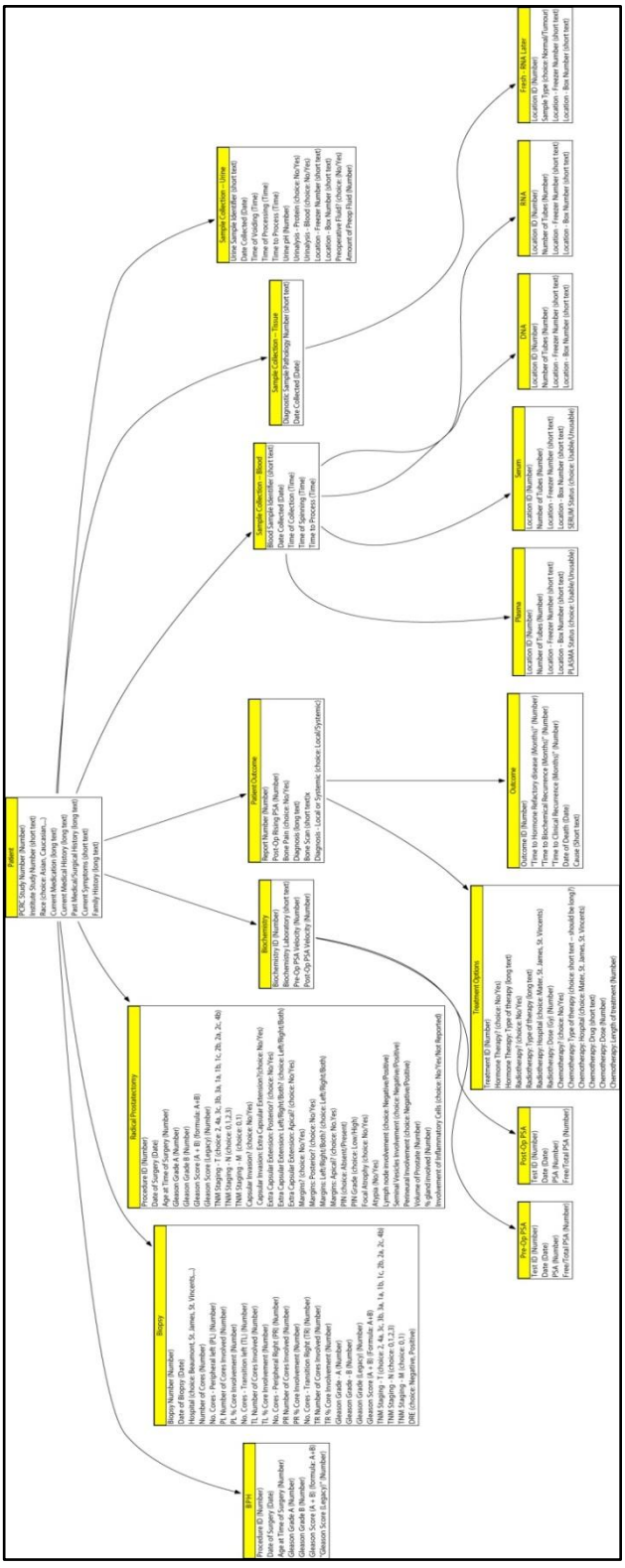


Figure App. O-1: Distiller Database Layout

Appendix P. Questionnaire

Questionnaire:

This questionnaire focuses on Sample Identification and Tracking System (SITS) in biobanks. The aim of this study is to:
- Learn about the opinion of the users of the current BIMS identification and tracking method
- Learn about individuals perception of electronic tagging to identify and locate samples
- Evaluate the necessity of deploying of electronic tagging system

Completing the questionnaire is expected to take about 10 minutes.

Note: As this is an anonymous questionnaire, completion of this questionnaire indicates your willingness to take part in this study.

Which one of the following best describes your position?

- ○ Researcher/Scientist (R)
- ○ Research Nurse (N)
- ○ Principal Investigator (PI)
- ○ Other, Please specify....

Part A: *Opinion of the users of the current BIMS identification and tracking method.*

Question 1. a) As part of your activities in the PCRC biobank are you involved in transferring a sample from a tube to another tube?

- ○ Yes
- ○ No

b) If yes, how do you rate the likelihood of mixing samples up, while aliquotting them?

very unlikely				very likely
○	○	○	○	

Question 2. How do you rate the current tracking system of samples using BIMS?

not very useful				very useful
○	○	○	○	

Question 3. a) Are you involved in storing samples?

- ○ Yes
- ○ No

b) If yes, how do you rate the procedure?

very easy				very difficult
○	○	○	○	

very quick				very time taking
○	○	○	○	

Question 4. a) Are you involved in uploading the location data on the BIMS?

- ○ Yes
- ○ No

b) If yes, how do you rate the procedure?

very easy				very difficult
○	○	○	○	

very quick				very time taking
○	○	○	○	

Question 5. How would you rate your confidence in the current BIMS system to prevent the following?

- Sample mix up during aliquotting

very unconfident				very confident
○	○	○	○	

- Sample degradation

very unconfident				very confident
○	○	○	○	

- Sample misplacement

very unconfident				very confident
○	○	○	○	

Question 6. Have you ever come across a sample that you cannot identify?

- ○ Yes
- ○ No
- ○ Not applicable

Question 7. Have you ever not been able to find a sample in its expected location indentified by BIMS?

- ○ Yes
- ○ No
- ○ Not applicable

Question 8. Have you ever had difficulties reading the labels on the tubes for example due to being covered by frost layers?

- ○ Yes
- ○ No
- ○ Not applicable

Please include any comments on the current system for indentifying and locating samples in the box below:

Part B: *Individuals perception of electronic tagging to identify and locate samples.*

Question 1. Would a system for monitoring temperature changes during transfer of location of samples be an advantage to your research?

- ○ Yes
- ○ Maybe
- ○ No

Question 2. Would you think it is useful for quality control purposes to know what processes the sample has gone through? e.g. Number of freeze-thaw cycles, centrifugation.

- ○ Yes
- ○ Maybe
- ○ No

Question 3. Do you think storing sample information (including sample identifier, tracking information and temperature changes) on a tube would prevent incidents below?

- Sample mix up during aliquotting
  - ○ Yes
  - ○ No
  - ○ Not sure

- Sample degradation
  - ○ Yes
  - ○ No
  - ○ Not sure

- Sample misplacement
  - ○ Yes
  - ○ No
  - ○ Not sure

Question 4. Have you ever heard about electronic or RFID (Radio Frequency Identification) tags?

- ○ Yes
- ○ No

Please include any comments on electronic tagging samples for identification and tracking in the box below:

Thanks for taking part in this questionnaire.