

The Use of Radio Frequency Identification to Track Samples in Bio-repositories

Atieh Zarabzadeh[§], R. William G. Watson[#], Jane Grimson[§]

[§]Centre for Health Informatics, Department of Computer Science, Trinity College Dublin,
Dublin 2, Ireland

[#]UCD School of Medicine & Medical Science, UCD Conway Institute of Biomolecular &
Biomedical Research, University College Dublin

[§]zarabzaa@cs.tcd.ie; [§]jane.grimson@tcd.ie; [#]william.watson@ucd.ie

Abstract

Bio-repositories are resources for storing biological samples and data to support the discovery of biomarkers, therapeutic targets, and the underlying causes of diseases. The success of this knowledge discovery process depends critically on the quality of samples and their associated data. Biological samples are expensive to collect and store. The samples and their associated data pass through a number of processes, generally in multiple locations, from data collection from the participant, to clinical laboratory analyses, to processing in the research laboratory (proteomics, metabolomics, etc.) and ultimately to knowledge discovery. The samples are moved in and out of freezers at several different points in this workflow. The potential for errors in sample identification and linking the corresponding data to its sample is therefore very high. This paper describes a novel method based on Radio Frequency Identification (RFID) technology to support secure and reliable tracking of both samples and data which minimises the potential for error. The samples are labelled with RFID tags, which allow location-independent recording and updating of sample data as they move along the workflow. The system is linked to a web-based database, which provides the ability to identify and locate individual or group of samples rapidly.

1. Introduction

Understanding the underlying causes of diseases has been made possible by recent advancements in molecular science. Discovering the causes and prognosis of disease requires two types of data: molecular data and phenotypic data [1, 2]. Both types are obtained from the analyses of samples gathered

from generally large cohorts of individuals and from their medical records. Bio-repositories or biobanks are developed to store and maintain samples from large populations for several years. Winn et al has defined a Bio-repository as “a system which will store one or many types of biological samples for later analysis from single or multiple studies under conditions which permit efficient retrieval and optimum stability of the sample” [3]. Bio-repositories consist of multiple units and subsystems such as collection, processing and quality control units. Bio-repository Information Management Systems (BIMS) control and manage the communication between these units and the storage and retrieval of data [2, 4]. Sample identification and tracking is an important part of the BIMS.

Each sample collected is valuable and requires careful storage and management. Particular attention must be paid to ensure that these samples are not lost or mixed up. Furthermore, samples without their associated data are of little value to scientists. The risks of sample loss or mix up become even greater in the increasingly common context of large multi-institutional (or in some cases multi-functional) bio-repositories.

Identification and tracking technologies will be discussed in the next section and based on this discussion, Radio Frequency Identification (RFID) system specification will be covered in section 3. Finally section 4 describes how RFID technology will be used by the Irish Prostate Cancer Research Consortium (PCRC) to identify and track samples.

2. Sample identification and tracking

The ability to uniquely identify and track samples is an essential element of a BIMS. Traditional methods of hand-written labels are not a suitable approach since they are unreliable and error prone. More recent

automatic methods are being considered such as barcodes, RFID, memory spot, Bluetooth, Wi-Fi, ZigBee and infrared. Most of these technologies are suitable for network communications but some are only able to communicate with small groups of devices. The technical characteristics of barcodes and RFID make them particularly suitable to identification and tracking systems.

Barcodes, both in forms of linear barcodes and data matrices, are the traditional technique for tracking and identifying goods, especially in supply chain settings. The data that can be stored in a linear barcode is basically a string of static characters. Data matrices have larger capacity up to a maximum 3,116 Digits [5] depending on the label size, number of rows and columns, and type. Data stored on barcodes, regardless of the type, are static.

Radio Frequency Identification (RFID) covers “technologies and systems that use radio waves (wireless) to transmit and uniquely identify objects” [6]. RFID tags allow data storage on the memory they have onboard. This data propagates as radio waves and is received by the reader/writer antennae. Data is then sent to a computer for additional processing. RFID tags can be integrated with sensors to facilitate additional services, such as for example, temperature monitoring.

Barcodes and data matrices are cheaper than RFID; however RFID is capable not only of carrying more data on the tag but also allow the data to be updated dynamically [7]. Since biological samples are stored at extremely low temperatures, typically -80°C, frost crystals may cover the labels of barcodes or data matrices, making them unreadable by the laser scanner. RFID operates independent of line of sight of reader and is able to read data through frost crystals. RFID also allows multiple reads concurrently which is beneficial in terms of speed when several sample types and aliquots (i.e. samples derived by dividing a sample) arrive at the laboratory concurrently [8, 9].

Biological samples are valuable and must be carefully stored and treated in order to ensure accurate results when they are analyzed. For example, they must be maintained at the correct temperature throughout processing. Temperature sensors can be integrated into RFID tags which can be used to automatically monitor the temperature of the samples, hence helping to prevent loss of samples due to degradation.

There are many issues regarding the use of barcodes in bio-repositories, such as line of sight problems, inability to retrieve data independent of the databases, the static nature of label data and non-extensibility in terms of future system enhancements. As the number of bio-repositories grows, the ability to share data and samples across bio-repositories becomes important, and requires a more flexible and extensible system than

can be provided by barcodes. A comparison between RFID and barcode is given in Table 1.

RFID	Barcodes
Dynamic data	Static data
Memory on tag	Memory-less
No line of sight	Line of sight
More expensive	cheaper
Multiple reads at once	Single read at once
Integrateable with sensors	Not integrateable

Table 1: RFID and barcodes compared.

3. RFID

RFID technology comprises of three main hardware devices: a reader/writer, antenna and tags, and a software application running on a machine that displays the tag data and may in turn be connected to a database [10].

Tags can be passive, active or semi-passive. Passive tags are energized with radio waves emitted from reader/writer antenna, while active tags contain a power source on the tag itself. Semi-passive tags have a power supply onboard for running the tag’s circuit and use the antenna’s energy to transmit data [11]. Tags are also categorized as either read-only or rewriteable [10]. Samples in bio-repositories are stored for several years and their data is often updated, therefore rewriteable passive tags, which have longer life cycle than active tags, are the most appropriate.

Reader/writer devices are required to be set at a frequency that matches that of the tags. RFID tags may contain updateable data that can only be read by the reader/writer of the same frequency. Since some samples are liquid it must be possible to transmit data through liquids, and within a range of less than one meter, the expected reading distance. The frequency that matches these conditions is 13.56 MHz. The rate of data transmission at this frequency is 25 Kbits per second which is suitable for the text data stored on the tags [12]. The appropriate standard that matches 13.56 MHz frequency is ISO15693 [13].

4. Sample identification and tracking in PCRC

The Irish Prostate Cancer Research Consortium (PCRC) bio-repository consists of four major components: the participating hospitals with their Clinical Resource Centers (CRC), research institutes,

the identification and tracking system, and the central database.

Data processing takes place in two phases. The first is the initial processing phase, when the samples collected from participants (study subjects) are prepared in the collection site's CRC and the second, when the prepared samples go through "omic" techniques in the research institutes. After the initial processing phase, samples are stored in the designated freezers located at the hospital sites and are made available to the research institutes for the secondary processing phase. Sample, clinical and follow-up data are collected from participants every six months in accordance with the study protocol.

De-identified data is currently entered manually into a central database by the research nurse, and can be accessed from the research institutes. The central database houses the de-identified data from each site and the identification and tracking system is responsible for ensuring that data related to the samples accompanies them throughout the process from collection to storage, to retrieval, including the "omic" procedures. These procedures conform to the requirements of the Data Protection Commissioner [14].

There are three main categories of data that need to be maintained for each sample; namely "general" data, "collection" data and "procedure" data. "General" and "collection" data are gathered in the initial processing phase at the collection sites. "General" data applies to all samples regardless of type and includes PCRC Study Number (unique Study Number for each participant), age of participant at the time of surgery, race, medication taken, PSA (Prostate Specific Antigen- defined as "A protein in the blood that can be measured to detect or monitor prostate cancer" [15]) and Gleason grade (that is used along with other parameters to evaluate the prognosis of prostate cancer [15]).

The initial processing phase is common between samples of the same type, i.e. any sample of a particular type collected will go through identical initial procedures that are unified across collection sites using Standard Operating Procedures (SOPs). SOPs are standardized procedures written for every action carried out in a bio-repository. They are in unambiguous language to specify procedures, policies and processes across different sites and units [16]. Hence, the data stored on the tag is predefined depending on its type, i.e. tissue, blood and urine.

Data gathered at the secondary processing phase is procedure dependent. The "omic" procedures carried out in PCRC bio-repository are metabolomics, proteomics, glycobiology, methylation, transcritomic and genotyping. Each "omic" procedure will generate

its own data fields that are important for biomarker discovery.

Each sample is collected in one tube with capacity specified by the SOP for that particular type of sample. Each sample is then divided into about three aliquots in the collection hospital and stored for transfer to research institutes for further analysis. Each of these aliquots will undergo different "omic" procedures which might be in a different research institute. To carry out these "omic" techniques, scientists need the "general" and "collection" data. The sample identification and tracking system allows these data to be stored on the tag and be available to the scientists through the RFID reader. The "omic" procedures often generate a number of very small size samples. Tagging these samples individually would be a major challenge. However, since these samples have the same characteristics, they can be tagged in groups. Thus for a group of samples from the same origin and undergoing one unique procedure and with identical characteristics, one tag can be used to store all of the data. A data field is designated to store details of exceptions or sample removal. Hence, there will be three major stages of tube change for each sample as shown in Figure 1.

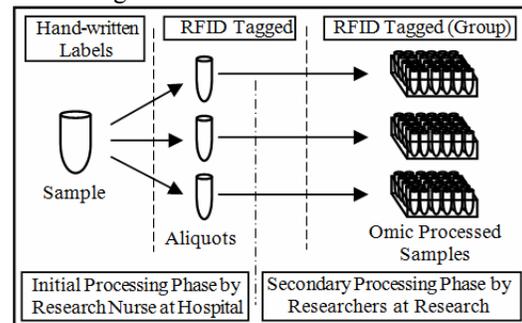


Figure 1: Sample division and processing

Original tubes are collected from one participant at a time therefore hand-written labels are sufficient to make sure the research nurse will not mix them up. The initial processing phase is carried out at the collection site and the original full capacity tube will not be RFID tagged until it is divided to aliquots. Aliquots are then tagged with the "general" and "collection" data and are stored in the designated freezers ready to be delivered to the research institutes. No identifiable data is stored on the tags. The de-identified "general" and "collection" data are also stored in the central database. Identifiable data are only stored on the local database of the original collection site [17].

The identification and tracking system will be integrated with the central database such that the data from the central database and tags are exchangeable via a web.

5. Conclusions

While some bio-repositories such as the Karolinska Institute bio-repository [18] in Sweden are using barcodes to identify their samples and prevent loss, PCRC currently uses hand-written labels as a temporary measure, but plans are underway to pilot the use of RFID technology.

Magellan Technology in Australia has developed a specimen slides tracking system based on RFID tagging [19]. These slides like the bio-repository's samples are small in size and large in population, but stored in containers of different shapes and go through different procedures. In this system an RFID antenna is implemented in the storage box containing the samples and the orientation of tags, their distance apart and other issues have been investigated.

Since biomarker discovery relies on large population cohorts, it is likely that there will be an increasing need to support sharing of samples between bio-repositories. To facilitate this, samples and data are required to be of the uniform quality and maintained through identical SOPs.

RFID offers the potential to share samples effectively and reliably across multiple sites. It also provides support for monitoring of samples quality, e.g. by incorporating temperature sensors into the tags, which is critical when samples are being moved from one location to another.

10. References

[1] G. Schmitz, K. Bhatia, M. Ferguson, and W. D.M., "Collaborating for Cures," in *Life Science Insights*, I. Melnikova, B. Reeve, and M. Swenson, Eds.: IBM, 2005.

[2] F. Betsou, M. Ferguson, B. Jallal, and J.-E. Litton, "Accelerating Molecular Medicine, Challenges Facing the Global Biobanking Community," in *Biobanks*, Z. Zimmerman, M. Swenson, and B. Reeve, Eds.: IBM, 2004.

[3] D. M. Winn, M. E. Reichman, and E. Gunter, "Epidemiologic issues in the design of biologic specimen banks," *Epidemiol Rev*, vol. 12, pp. 56-70, 1990.

[4] J.-E. Litton, "A Biobank Information Management System: The Importance of Data Integration," in *Life Science Insights*, Z. Zimmerman, M. Swenson, and B. Reeve, Eds. Stockholm: IBM, 2004.

[5] DataMatrix Barcode Reading SDK / ActiveX, Available: www.blackice.com/barcodeDataMatrix2DRead.htm visited on: 25/03/2008.

[6] S. Piramuthu, "Protocols for RFID tag/reader authentication," *Decision Support Systems*, vol. 43, pp. 897-914, 2007.

[7] B. K. Schuerenberg, "Bar Codes vs. RFID: A Battle Just Beginning," *Health Data Management*, vol. 14, pp. 32-4, 2006.

[8] D. Jacoby, "RFID comes of age," Economist Intelligence Unit, London March 2006 2006.

[9] H. Knospe and H. Pohl, "RFID security," *Information Security Technical Report*, vol. 9, pp. 39-50, 2004.

[10] G. Adams, "Pharmaceutical manufacturing: RFID - reducing errors and effort," *Filtration & Separation*, vol. 44, pp. 17-19, 2007.

[11] C. M. Roberts, "Radio frequency identification (RFID)," *Computers & Security*, vol. 25, pp. 18-26, 2006.

[12] K. Domdouzis, B. Kumar, and C. Anumba, "Radio-Frequency Identification (RFID) applications: A brief introduction," *Advanced Engineering Informatics*, vol. 21, pp. 350-355, 2007.

[13] D. Morgenroth and K. Fobes, "Another link in the chain," *Card Technology Today*, vol. 16, pp. 11-12, 2004.

[14] Data Protection Commissioner, Available: www.dataprotection.ie visited on: 24/03/2008.

[15] The Science of Proton Therapy, Available: <http://www.mpri.org/science/glossary.php> visited on: 25/03/2008.

[16] K. E. Pitt, L. D. Campbell, and A. P. N. Skubitz, "Best Practices for Repositories I: Collection, Storage, and Retrieval of Human Biological Materials for Research," *Cell Preservation Technology*, vol. 3, pp. 5-48, 2005.

[17] A. Zarabzadeh, R. W. G. Watson, G. Bradley, and J. Grimson, "Ensuring Participant Confidentiality In Bio-repositories," Trinity College Dublin, 2008, Submitted for publication.

[18] Karolinska Institutet Biobank, Available: www.kibiobank.se, visited on 25/03/2008.

[19] B. Bacheldor, "Magellan Creates Specimen-Tracking System," in *RFID Journal*, vol. 2008, B. Bacheldor, Ed.: RFID Journal, 2008.