Data Architecture for Telehealth Services Research: A Case Study of Home Tele-Monitoring

(invited paper)

Surya Nepal, Julian Jang-Jaccard, Branko Celler, Bo Yan, Leila Alem
CSIRO Computational Informatics
Sydney, Australia
{Fname.Lname}@csiro.au

***Abstract*—Telehealth services research projects often require to access a variety of data sources under different data access policies and privacy constrains. There is a need to link these clinical and administrative records from different data custodians and produce a research data for analytics. One of the challenges is that the research data must meet the data access policies and privacy constraints of all data custodians participating in the project. These data custodians often operate in different jurisdictions. In this paper, we present our practical experience through the design and implementation of a service-oriented data architecture for extracting research data for telehealth services research in the context of a tele-home monitoring project. This project is being carried out at six locations in five different states in Australia. Each site represents a different model of care for the management of chronic disease in the community ranging from community-based, nurse-led programs to a hospital-focused, chronic-disease management program. The aims of this project are wide ranging and investigate many aspects of deploying at home telehealth services to better manage chronic disease. This paper however focuses on data architecture. We highlight the underlying issues, our experience and explain a practical health data linkage protocol adopted in the project. We also explain the features of the research data service portal in operation.**

Keywords- health data linkage, data architecture, data privacy, data services, telehealth research data portal

I. INTRODUCTION

Telehealth promises to revolutionise the way health services are delivered to patients at home [10]. A number of projects have been conducted to study the feasibility of implementing tealehealth services. Example projects include chronic disease management [1], mental health care [2], ophthalmology [3], diabetes [11] [4] and allied health [5]. Ageing populations and growing cost of healthcare are some of the main reasons governments around the world are attracted to telehealth services as a way to reduce overall health cost [19] [5]. Furthermore, advancements in information and communication technologies, such as ubiquitous broadband networks, cloud computing, and social networks, have also played a key role in developing further interests from governments in telehealth. As a result, a number of government-funded telehealth services research projects have been established. CSIRO Health Services research program was recently awarded a project under the National Broadband Network (NBN) Enabled Telehealth Pilots Program to demonstrate how telehealth services for chronic disease management in the community can be deployed nationally to reduce the rate of hospitalisation and improve health outcomes of chronically ill patients living at home [25]. In addition, the project aims to develop advanced modelling and data analytics to risk stratify patients on a daily basis to automatically identify exacerbations of their chronic conditions.

Telehealth research projects typically deal with two types of data: (a) administrative and clinical data and non clinical data created as part of the research project, and (b) administrative and clinical data created as part of the normal care provided to patients (including historical data). The first type of data is collected during the project with the aim of understanding in detail the progression of the patients' chronic conditions and their wellbeing (e.g., quality of life, mental health etc.) as outlined in the research project objectives, whereas the second type of data comes from different sources such as hospital records, government records, nursing home records, and community care records. One of the first requirements for these research projects is to get access to these administrative and clinical data, and to generate data that can be used for both clinical and research purposes

We identify the following three steps for the generation of research data: (a) access to data from different data sources, (b) linking the different data sets, and (c) generating research data in a suitable format. The first step involves applying for ethics clearance for data access from each organization that is the custodian of the data set. Each organization often releases the data in an anonymised form by assigning a de-identifiable unique identity for each patient. These unique identities for different data sets are different. The second step involves linking the data sets collected from different sources. The third step is generating the linked data as requested by researchers for the purpose of analysis. Therefore, it is necessary to develop a service that deals with the lifecycle of data in telehealth services research, from collection to integration

COLLABORATECOM 2013, October 20-23, Austin, United States

DOI 10.4108/icst.collaboratecom.2013.254220

to generation. In this paper, we describe our experience, challenges and our approaches in each of these steps.

Each of the data set collected from its source is subjected to a certain jurisdiction under which it is collected. The data set is released after the approval of the ethics application. Each ethics application provides strict guidelines on how to deal with the data such as who can access the data and how can such data be stored, distributed, and destroyed. These guidelines define the data access policies. These access policies should be followed throughout the lifecycle of the data in the project as well as after the project as long as the data is retained for the research purposes. The question is then how to integrate data from different sources under different policy constraints in such a way that the integrated data set meets all the constraints as specified in the ethics approval process and corresponding data laws from the custodians' jurisdictions. In the database community, the general schema matching approach [20] has been proposed and used in many application domains, but they cannot be applied in health services where privacy is a major issue [18].

The problem specified above is not new and has been dealt with in a number of different ways in the literature. CSIRO HDI^{TM} tool is designed to integrate health data in [6]. The tool enables linking of patient records in different data sets while maintaining the privacy of the patients. This tool utilizes the integration server as third party protocol [7][8]. These approaches are generally applicable to integrate health data sets but are not specifically designed for telehealth services research. This means these tools are designed to integrate second type of data. However, telehealth services research projects often deal with two different types of data as described earlier. The health data maps in [9] also raise a lot of problems that are relevant to telehealth research project such as maintaining user involvement in system development, understanding and integrating data, bringing disparate data sources together, and making use of assembled data.

Data integration issues are also dealt with in other domains such as bio-informatics [12] and public health surveillance [14]. Semantic Web and Web 2.0 technologies are the latest to find traction from the bio-informatics community [22]. Privacy has been one of the major concerns in health data integration [18]. O'Keefe has summarise the problem of enabling the use of health data for research on clinical practice and policy analysis while protecting the privacy and confidentiality of individuals, health care providers, health care facilities and data custodians [13]. The benefits of integrated health data has been amply reported in the literature [15][16], but the data integration problem is harder and the devised solutions are tailored to a specific context. Some of the probabilistic based approaches discussed in [21] are not suitable in telehealth scenario as the application demands the exact match of records from different data sources. Alternative approaches have been investigated [23], such as proposed by Daniel et al. suggesting the need for an information accountability approach to deal with privacy in the Web [23].

Each of these approaches proposed in the literature has their own merits and applications to specific data sets. However, applying them in a real life field trial always brings practical difficulties. Our study imposes the following requirements to data services:

- In general, different sets of clinical and administrative data are collected using different ethics applications. Hence, the data services must be in compliance with all data access policy constraints specified in different ethics applications. The question is how to perform this in an optimal way?
- Data services must not only provide the functionality to collect pre-existing data set, but also the mechanisms to collect project specific data such as usability.
- Data services must be able to generate de-identifiable research data in real time as the study progress so that it can support the data analytics to risk stratify patients on a daily basis to automatically identify exacerbations of their chronic conditions.
- Data services must support different data sources and data output formats.
- Data services must also provide a mechanism to provide analysis results back to care providers and patients.

In order to meet these requirements, we have designed and developed service-oriented data service architecture [17] for our telehealth services research. The architecture has been implemented and has been used in the study for the last six months. The complex process of ethics clearance and data handling protocol in the system is based on a best practice protocol proposed in [24]. This paper describes our practical experience of designing and developing data architecture for telehealth services research, and provides unique insights into the complexity of applying research outcomes in practice.

The rest of the paper is organised as follows. We explain the motivation and background information on our home tele-monitoring project in Section II. Various data types and data sources involved in the project and their corresponding privacy policies emanating from ethics applications are explained in Section III. We then present the design and implementation of our data architecture in Section IV. We have implemented a data service portal to realize the underlying data models and data architecture, which we describe in Section V. Finally, we present a brief discussion and conclude the paper in Section VI.

II. BACKGROUND AND MOTIVATION

CSIRO through its health services theme has carried out a comprehensive analysis of the international literature on telehealth trials with a focus on the factors impacting

the success of telehealth programs associated with the ageing of the population and increasing burden of the chronic diseases [26][27][28][29]. It is found that a strong primary health care system is critical to the future success and sustainability of the healthcare system. The importance of primary care has emerged as a recuurent theme in government reports and strategy documents [26][27][28].

Chronic and aged care accounted for over 70% of Australia's $103.6bn expenditure on healthcare during 2007-08, and is projected to dramatically increase into the future [29]. Australia's hospital-centric public health system is unnecessarily burdened by the management of chronic diseases, which should preferably occur in home and community settings. Telehealth services delivered through home monitoring have been demonstrated to deliver cost-effective, timely and improved access to quality care [30]. They also reduce social isolation n and enhance the quality of life and sustanability of these communities by allowing chronically ill and aged members to stay in their homes and within their communities longer. These statements are also supported by recently released findings of the Whole System Demonstrators (WSD) by the UK department of health [31]. Similar outcomes were observed in the telehealth programs administered by Veterans Health Adminsitration (VHA) in US [32].

However, experience in Australia with the deployment of telehealth services is extremely limited, with most deployment of small scale and detailed analysis of the following key success factors: health care outcomes, health economic benefits, impact of clinical workforce availability and deployement, human factors, workplace culture and organisational change management and business processes.The development of robust business case and business models for large scale commercial deployment of telehealth services based on solid socio-economic evidence is therefore essential if these services are to be scaled up nationally to have an impact on escalating costs of health services delivery and increasing deficit in clinical workforce.

With the aim of addressing the above issues, a project "Home monitoring of chronic disease for aged care" is being carried out by CSIRO since Januaray 2013 for the duration of 18 months under the National Broadband Network Enabled Teleheath Pilot program administered by the Department of Health and Ageing (DoHA). The project is the first multi-state, multi-site telehealth trial performed in Australia and includes six nodes in five states. The project will also develop advanced modelling and data analytics to risk stratify patients on a daily basis to automatically identify exacerbations of their chronic condition. These capabilities will in turn be used as the basis of a decision support system to ensure the optimal orchestration of health services so that patients receive the best available care at the right time to avoid unnecessary hospital admission.

As the project involves six different organisations in five different states, the clinical and administrative data of patients that need to be collected for modelling and analytics is subject to clearance from a number of different ethics committees. The management of data is further complicated by the fact that these organisations operate under different juricdictions, and we therefore need to consider different data privacy laws and acts. The paper provides our experience in managing such process through the lense of data service architecture and data management protocols.

III. Data Types and Data Sources

The project as a whole is still currently enrolling patients with some sites having completed their enrollment process. Once the enrollment process is completed, the project will have 25 test patients and 50 control patients on each site, i.e., 150 test patient and 300 control patients in total. We need to collect a variety of patient data from different sources for modeling and analytics. In the following, we explain different data types and data sources we use in the project followed by corresponding protocols and data models developed and implemented.

A. Hospital Records

Since our aim is to reduce hospital admissions via tele-monitoring of patients at home, we need to understand patient's medical history. How often were they hospitalised in the past? What were the main causes of hospitalisation? Hospital data is collected at different stages. First, we collect hospital data for all potential patients in a site to match and select the test and control patients. The potential test patients in the trial are determined by whether the patient's residence is connectable via high bandwidth national broadband network (NBN). Second, we collect historical hospital data for both test and control patients in order to compute the probability of hospital admission. Third, we collect hospital data of all participants in the trial for the duration of the trial, in order to evaluate the effects of the tele-monitoring intervention.

Hospitals records differ from one hospital to another. Furthermore, hospital data gets even more complicated if a patient visits more than one hospital within or outside their community. In such scenarios, the hospital data is less likely to be complete. One of the major issues is that these hospital records do not have similar schemas and the best of the available technologies for schema matching only yields approximate results, which are not suitable for analysis purpose. The project requires exact and complete hospital data. To minimize the complexity of this problem, we use data from Health Roundtable[1] as much as possible. However, it is important to note that, not all hospitals participate in the Health Roundtable.

[1] www.healthroundtable.org

B. PBS and MBS

Patients' full medical history may not be reflected in hospital records. Furthermore, hospital records do not provide any information about patients' visits to their local GPs or their use of medications. We use Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) records from the Department of Human Services (DHS) (formerly known as Medicare) to capture such visits. The PBS provides all Australian residents and eligible overseas visitors' access to subsidised prescription medicine in a way that is affordable, reliable and timely. Similarly, the MBS provides the subsidized medical services. The PBS/MBS data provides information about patients when they visit medical practitioners and accessed subsidised services or purchase medicines under subsidized schemes. The recorded medical service or prescription of the medicine may either give new information about patients or missing information from hospital records.

C. PCEHR

The Personally Controlled Electronic Health Record (PCEHR) System is the next step in using eHealth to enhance the Australian healthcare system. The PCEHR System enables the secure sharing of health information between an individual's healthcare providers, while enabling the individual to control who can access their PCEHR. Clinical documents, such as Shared Health Summaries, Discharge Summaries, Event Summaries, Pathology Result Reports and Specialist Letters will be collected from a range of participating organisations, and stored within a number of secure repositories in the PCEHR System. The PCEHR System may also share key health information entered by the individual (such as over-the-counter medications and allergies), and access information from the Department of Human Services. This includes Medicare information, such as an individual's organ donor status, dispensed medications funded under the Pharmaceutical Benefits Scheme (PBS), information about healthcare events from an individual's Medicare claiming history and a child's immunization history. The PCEHR System will also collect information about the location of an individual's advance care directive (if they have one). However at this stage the PCEHR contains mainly PBS and MBS data and following ethics approval from the Department of Human Services, we have chosen to access this data directly rather than through the PCEHR. Furthermore, our aim is to conduct the feasibility of integrating telehealth trial with eHealth system so that we can upload the summary report from the analysis to PCEHR and share with care providers such as local GPs.

These three data sources provide the patients' health records that have been collected as part of the national health systems. In the longer term, we believe that the PCEHR may supersede all other data. However, at the time of the project implementation, a small number of patients have registered for PCEHR, and therefore we need to consider other data sources. We next describe the administrative and clinical data sources collected as part of the project.

D. TeleMedCare (TMC)

There are many home monitoring devices and systems currently available in the market. After extensive evaluation of alternative systems by an expert panel against our project objectives, the panel selected the TeleMedCare[2] (TMC) as a home monitoring system for the trial. A unique feature of this system is that the signal traces for all vital signs recorded are retained and are uploaded to the server. Fig. 1 shows the overview of the TMC system. It consists of a device at patient's home that monitors the vital signs of a patient as per schedule. The vital signs include weight, blood pressure (auscultatory and oscillometric), temperature, single lead ECG, spirometry, glucometry, etc. The patient can be reminded to take their vital signs as per schedule. The measured vital signs are stored in the local computer and are sent to the server at regular intervals. The TMC system can also deliver a comprehensive set of clinical questionnaires and record the results. The TMC system supports both a push and pull model for accessing data. In the push model, the TMC server can send de-identified data to authenticated services at regular intervals. Similarly, in the pull model, a service can authenticate and retrieve de-identified data at anytime. The data at server is made accessible to care team including care nurses and local GPs. The TMC provides a data service to access vital signs as well as generated reports, but it is limited to only data collected by TMC. We refer readers to TMC device manual for further information about TMC.

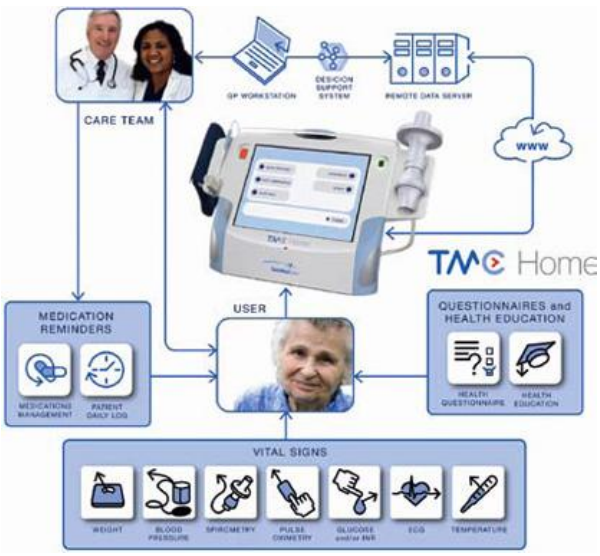

Figure 1. TMC Overview

E. OpenClinica (OC)

TMC supports a number of standard questionnaires. However, for the purpose of the project some needed to be modified and others added. Even when questionnaires are

[2] www.telemedcare.com.au

approved by ethics committees they may be subjected to constraints. As a result, the data collected through such questionnaires may not accessible through the TMC server. Hence, we use OpenClinica [3], an open source clinical trial software, to model and capture various types of questionnaires at different stages of the project implementation. In our project, we have three types of questionnaires: entry, exit, and intermediate. Entry questionnaire capture the following nine types of patient data: demographic information, medical information and use of health services, behavior information, physical activities, anxiety and depression, living with and managing medical conditions, quality of life, social isolation, and medical adherence. These questions are designed using established questionnaires. Some questionnaires are administered during the enrollment process by the project officer in each site, whilst others are administered through the TMC home portal.

F. Care Provider Data (CPD)

The care providers conduct a number of activities as part of providing daily care to patients. These include home visits, phone calls to family members, and calls to patients or their carers. As the project will be studying business processes involved in the delivery of care, it is important to log all these activities. All activities from care providers during the trial are recoded as activity logs. Care providers may typically use different software package to record these activities. There is no established uniform and standard way of recording such activities. Therefore, we have developed and provided a standard data model for capturing activity log.

G. Patient Personal Data (PPD)

The entry questionnaire has two parts. The de-identified part contains information under nine categories explained earlier and is stored in OpenClinica. The identifiable part including patient's name, address, contact details, doctor's details and next of kin details is stored in a as a patient personal data. The patient personal data is very sensitive and need to be treated differently. We store this data in a secure encrypted database. We explain it further in the implementation section below.

H. Project Specific Data (PSD)

In addition to all data sources explained above, there are some project specific data that needs to be collected from different sites. More specifically, the data needed to conduct impact of clinical workforce availability and deployment, human factors, workplace culture and organisational change management and business processes. It includes observation data such as specific incidents that are not captured in the activity log.

The characteristics of the above data sources can be summarized as follows:

- **Federated Schemas**: Patient data is recorded in a variety of databases using different types of database schemas. Even the same type of data is recorded differently in different data sources such as hospital data and care provider data.
- **Federated Identities**: Different identifiers are used to identify patients in different data sources. Even within the same data type, the identifiers are different (for example, the way a patient is identified within a system in each care provider is different).
- **Variety of Formats**: Each data source exports data in a different format. There is no uniform way of accessing data.
- **Different Jurisdictions**: As the project is being run on six different sites in five states, the data coming from a source needs clearance from its own ethics committee and may be subject to different privacy constrains.

These characteristics pose a challenge to collecting data from different sources, linking them together and generating a research data which meets a set of constraints imposed by different ethics committees and local jurisdictions. In this paper, we describe a practical solution used in our home tele-monitoring project. We have adopted the best available practical protocols and tools to implement the solution. In the following section, we describe our solutions using data models, data architecture and corresponding protocols.

IV. DATA ARCHITECURE

Implementation and roll out of our research project required a number of people playing a number of roles. We first define different roles for the people who are involved in the project. Fig. 2 shows different roles and their relationships. The project director (PD) oversights the overall project team consisting of a number of researchers, a project manager (PM) and a clinical trial coordinator (CTC). Each trial site has a project officer (PO) and one or more clinical care coordinators (CCC).The PO and CTC works with PM and CCCs to run the project on each site. Researchers work with PD to conduct different research activities. Three groups of researchers are involved in the project. The data management group deals with collection, integration, and generation of data. The HCI group deals with workforce availability and deployement, human factors, workplace culture and organisational change management and business processes. The data analytics group deals with developing advanced modelling and data analytics to risk stratify patients on a daily basis to automatically identify exacerbations of their chronic conditions. All three groups have different data requirements.

[3] www.openclinica.com

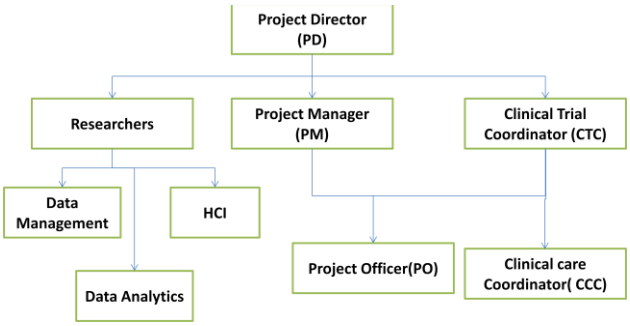

Figure 2. Project Role Allocation

We next explain the data architecture and services. Fig. 4 shows the data architecture for collecting, entering and generating different types of data. Who can access the data service? The access to a data service depends on the following factors: (a) roles, (b) organizational boundaries, and (c) ethics applications for sites (ethics clearance is necessary to access data) and (d) jurisdictions (each site is subject to state law). To satisfy these requirements we designed and implemented a privacy-aware role based access control mechanism.

The lowest level in the architecture is the data sources. Accessing data from different data sources requires its own ethics approval. We define RBAC mechanism for each data source rather than each patient. For example, TMC data of a patient can be accessed by CCC for the site, but PBS/MBS data of the same patient is not accessible to CCC, as it is subjected to different ethics application. Once the collected data is stored at the research project host organization (in our case CSIRO), the access to collected data is subjected to two ethics applications: data custodian and research project host organization. If there is access denied due to conflict in these two ethics applications, the access to data is denied until one of the ethics applications is amended to resolve the conflict. Our data service access policies often trigger the amendment of ethics application if there is a need to access certain data by a particular role. For practical reasons, we create a new role and assign to the person, as a person can have multiple roles. This is a practical approach to modeling potential conflicts on the system.

The top layer is a data service layer. There are three types of data services: collection, integration, and generation. Some of these services can access only linked data, whereas others can access linked data along with Patient Personal Data (PPD). One of the most important services is the research data generation service that outputs privacy preserving linked data to researchers for modeling and analysis. Fig. 3 shows a high level schematic of the research data analytics service. The data is made accessible based on sites, data sources and different fields within database schemas of data sources. Some fields are directly taken from database schemas, whereas some fields are computed on the fly. The data is generated in two formats: XML and CSV.

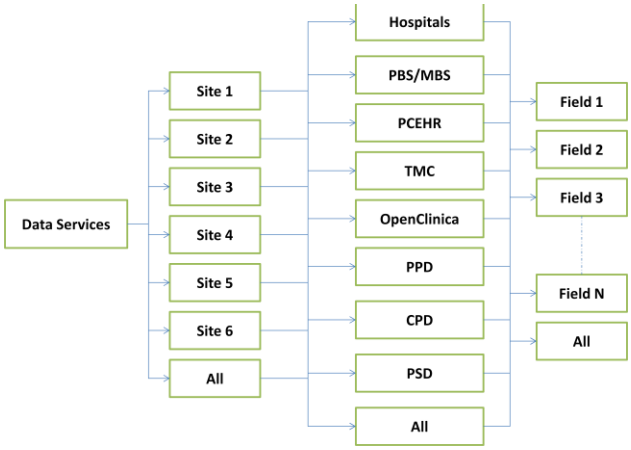

Figure 3. Data Service for Research Data Generation

The second layer of the architecture supports two major functionalities: resolving access conflicts, and creating linked data using a data link file. We have briefly covered the access conflicts resolution mechanism above. We next describe the protocols for creating linked data and corresponding data link file.

One of the important tasks in telehealth research projects is linking data coming from various data sources. We use a link file for this purpose. The link file has a unique identifier for all patients involved in the project. Since the patients are enrolled using OpenClinica, we use the OpenClinica Identifier (OCID) as a unique identifier for the patient in the project. Fig. 7 shows the data model for the link file. Note that ActivityLog is a PPD and ObservationLog is a PSD. PBS and MBS data are kept separately. There are two different mechanisms for creating link file. The corresponding protocols are shown in Fig. 5.

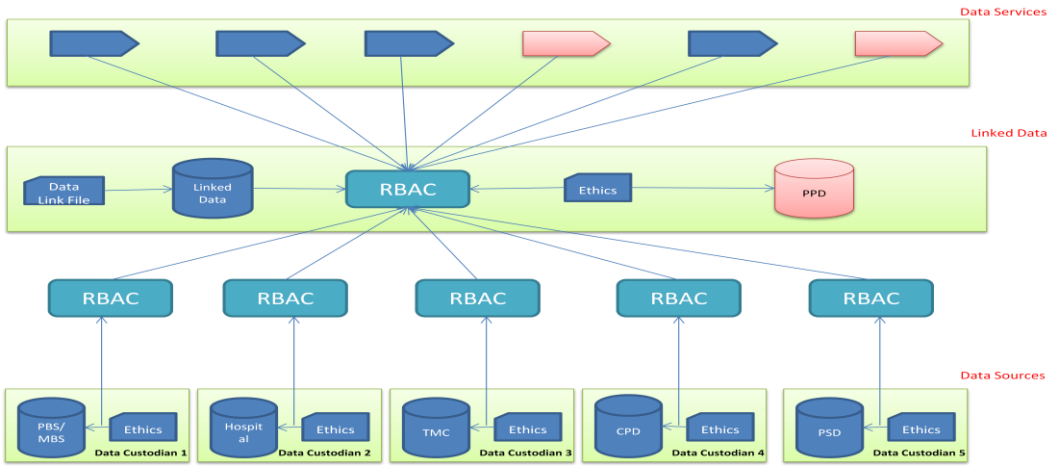

Figure 4. Data Architecture with Access Controls

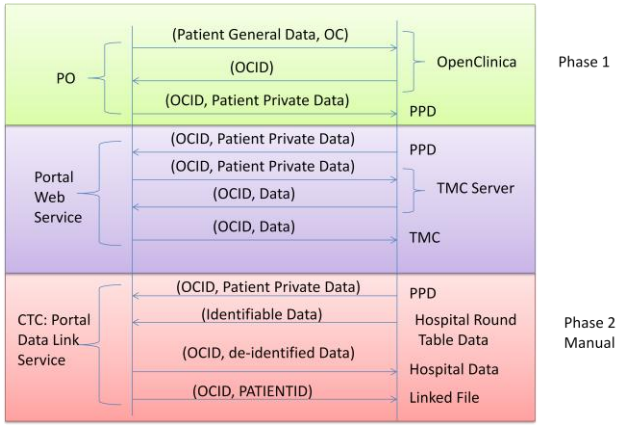

Figure 5. Protocol for creating a link file

The data link protocol has two phases. In the first phase, the PO gets the consent form signed by patient to be part of the study. The PO then enrolls the patients for the trial. The PO collects the patients' general data (which is de-identifiable) and private data (which is identifiable) as part of the enrolment process. The PO then enters the de-identified general data into OpenClinica. It assigns a unique identifier for the patient, called OpenClinicaID (OCID in Fig. 7). We use this identity as a unique identifier for creating the link file and linked data. The private part of the data is then stored to a secure encrypted database (PPD) along with OCID. This process linked the private and public parts of the entry questionnaire through OCID. In the second phase, we link data coming from different sources using OCID. We have implemented two different mechanisms for collecting and linking other data: automatic and manual.

We explain the phase 2 automatic process using the TMC server as an example as shown in Fig. 5. TMC server provides an API to access TMC data through a Web Service. The Web service retrieves OCID and identifiable patient data from PPD, and sends to TMC server along with authentication details. Upon the successful authentication, the external server first anonymises the data and sends the requested data along with OCID. The Web service then stores the data along with OCID in the corresponding local TMC data server. This represents a pull mechanism for automatically linking data from different sources using OCID. We have also implemented the push model, where the external data sources regularly dump the anonymised data along with OCID.

Manual processes involve the clinical trial coordinator (CTC), as s/he has ethics clearances from all data custodians and can access the identifiable data. Fig. 5 shows the manual process in phase 2. In this process identifiable data collected from different sources are stored in a secure temporary database before creating the linked data. The CTC uses the data link service provided by the portal to perform the linking task. The data link service retrieves the private patient data along with OCID from PPD. It also retrieves the identifiable data from the temporary database. The portal then allows the CTC to link the identifiable data with OCID by manually matching the data. The process can be automated. However, the data linking through schema matching process relies on the probabilistic model. As the research project requires exact matching of the data, we used a manual process in cases where the automatic process described earlier is not possible. The CTC is able to perform this action as OCID is displayed along with patient personal data retrieved from PPD including name and address. Once the linking process is completed, the links are stored in the LinkFile, whereas the identifiable data is de-identified and stored with OCID into the corresponding databases. At the end of the process, the temporary identifiable data is destroyed.

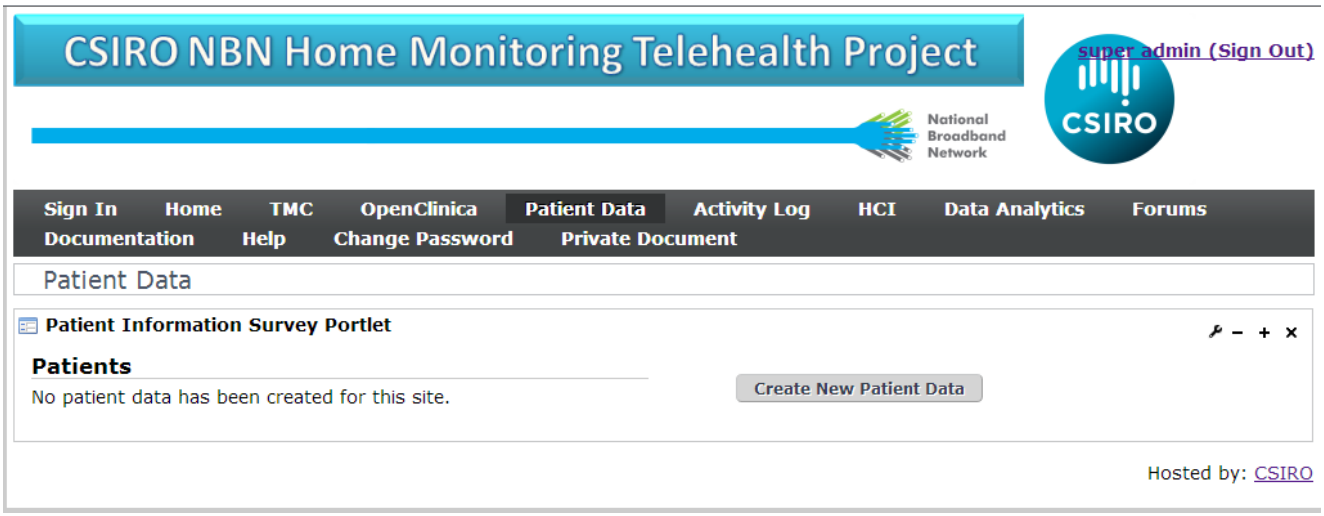

Figure 6. A screenshot of the implemented system with data services

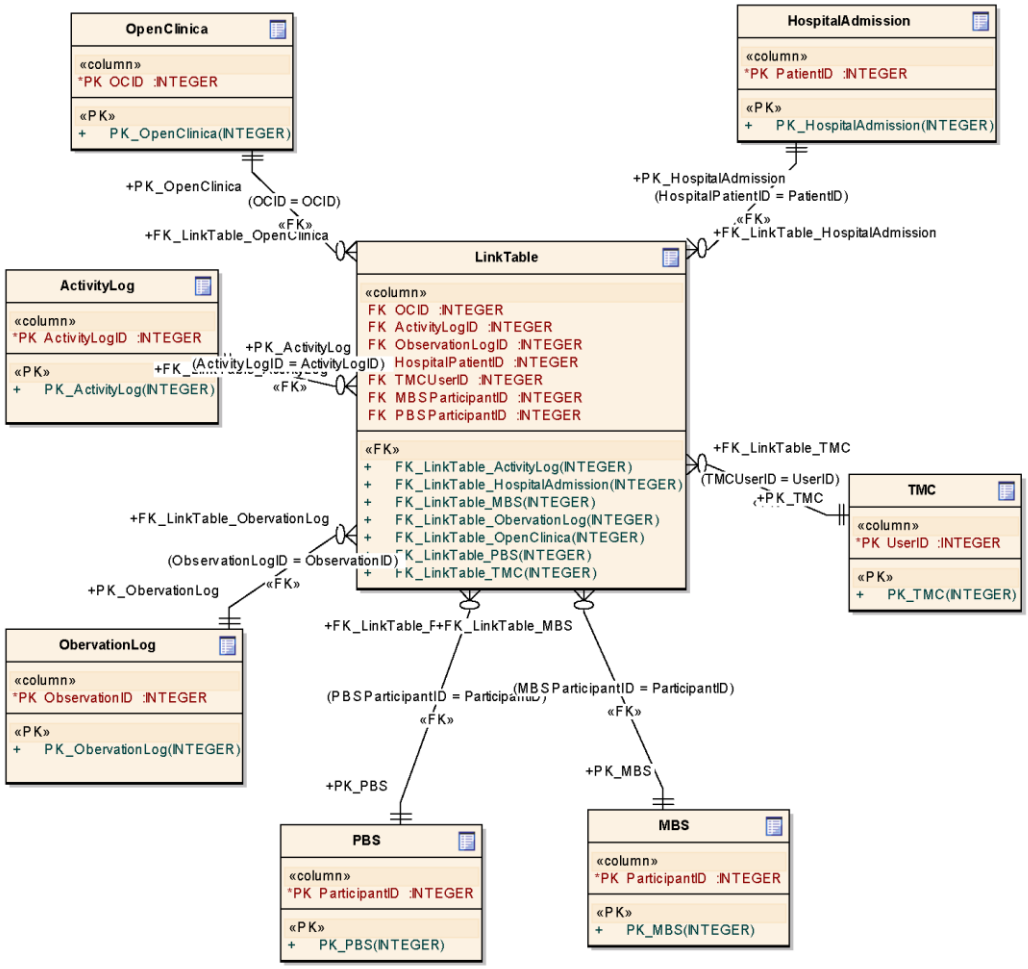

Figure 7. Data model for the link file

V. IMPLEMENTATION

We have implemented a telehealth services research portal using the data architecture and corresponding data services described in earlier sections. The portal has been in use since February 2013, and a number of services have been added and deployed. The system architecture is shown in Fig. 8. We use Liferay[4], an open source enterprise portal, offering content management, collaboration, and social networking functionalities, along with enterprise databases and document management solutions. Fig. 8 shows some of the key components used in our implementation. We provide three key services to users: authentication service (SignIn in Fig. 6), social network service (forum in Fig. 6) and data services (activity log, data analytics, patient data, TMC, OpenClinica, private document in Fig. 6). The data services are supported by three types of underlying data management systems. Data from services such as TMC and activity log are stored in the MySQL databases as linked data behind the enterprise firewall.

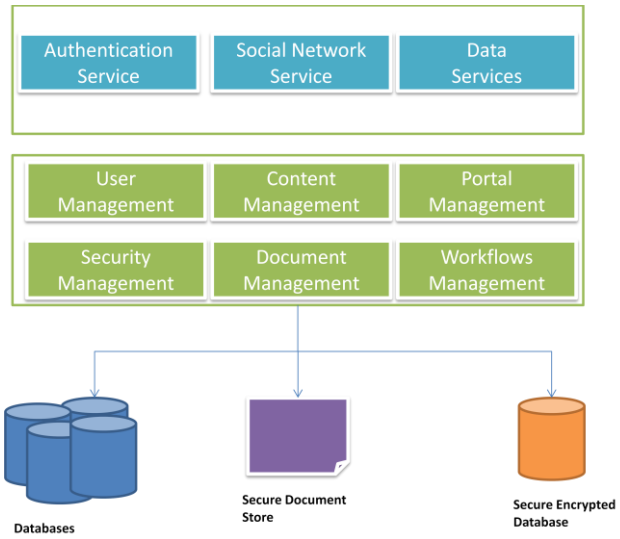

Figure 8. System Architecture

There are two types of data that contain patient personal information: consent documents signed by patients, and patient personal data used for creating linked data. The document data is managed using Microsoft SharePoint, whereas we use the secure encrypted database to manage patient personal data. We next describe briefly how we have implemented the secure encrypted database. We encrypt all attributes of a patient personal data (e.g., first name, surname, date of birth, contact details, emergency details, contact GP/nurses, etc.) using the AES (Advanced Encryption Standard) algorithm with a key length 128-bit. The database schema is designed in such a way that all attributes of a patient except patient identity is declared with VARBINARY in the database. This is to avoid potential problems with trailing space removal or character set conversation that would change data values as many encryption and compression functions return strings for which the result might contain arbitrary byte values. We then used MySQL function AES_ENCRYPT(str, key_str) to implement encryption of the patient private information. AES_Encrypt () encrypts the string str using a key key_str and returns a binary string containing the encrypted output.

To secure the passing of a key to the key_str argument, we create a truly random 128-bit value in a hex format and pass it as a binary value. The key is securely stored in a file on a secure location. Only users with the correct permission can access the key file. Similarly, we utilize an internal function of MySQL AES_DECRYPT (crypt_str, key_str) to decrypt the encrypted string crypt_str using a key key_str and return the original plaintext string. If AES_DECRYPT() detects invalid data or incorrect padding, it returns NULL. It returns a garbage value if the input data or the key is invalid.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we have presented a practical approach to privacy aware data services for a telehealth services research that is being used in a home tele-monitoring trial conducted by CSIRO in Australia. Designing and implementing data services for telehealth services research encounters a number of challenges that are different to that of telehealth services. First, telehealth services are often operated by a single organization under a single jurisdiction. On the other hand, a complex telehealth services research, like ours that has multiple sites in multiple states, involves multiple organizations across multiple jurisdictions. Secondly, telehealth services often have well defined data access rules that can be modeled using standard role-based access control mechanism, whereas the data access in telehealth services is driven by ethics clearances obtained from data custodians. For example, a patient's PBS/MBS or PCEHR data cannot be accessed by care provider even though they can access the relevant data if it was present in their own records. Third, telehealth services are designed with an aim of providing care to patients, whereas telehealth services research also involves researchers who model and analyze the data. The outcomes of the research are normally disseminated to wider research communities and policy makers. As the data goes beyond the control of health services providers, it is important to ensure that published results do not identify any individual patients thus ensuring their privacy.

This paper proposes a practical service-oriented approach to deal with telehealth services research data. We have described the system architecture, data architecture, data services, and corresponding protocols. In the future, we plan to develop a formal model, similar to models presented in [33][34], and corresponding data services authoring tool for telehealth services research. The authoring tool should enable researchers conducting telehealth services trials to model, analyse and resolve data privacy issues by considering obligations resulting from ethics clearance and data access laws in each data custodian's jurisdictions. The tool should also trigger relevant actions that need to be taken to meet obligations specified in ethics clearance. We also plan to develop a privacy preserving protocol for integrating results from analysis to various data sources, more specifically to the PCEHR.

[4] www.liferay.com

ACKNOWLEDGMENT

The research is funded by Australian Government Department of Health and Ageing under the broadband-enabled telehealth pilots program. We also would like to acknowledge and thank our partners: Department of Human Services, Nepean Blue Mountains Medicare Local and LHD, NSW, Anglican Retirement Villages, NSW, ACT Health, ACT, Northern District Health Service, TAS , Townville Mackay Medicare Local and Health District, QLD, Ballarat Health Services and Grampians Health Alliance, VIC, TelMedCare, iiNet, and National Brodband Network (NBN) Co.

REFERENCES

[1] J. Polisena, D. Coyle, K. Coyle, and S. McGill, "Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations," *International journal of technology assessment in health care* 25, no. 03 (2009): 339-349.

[2] B. H. Stamm, "Clinical applications of telehealth in mental health care,"*Professional Psychology: Research and Practice* 29, no. 6 (1998): 536.

[3] K. Bahaadinbeigy, and K. Yogesan, "A Literature Review of Teleophthalmology Projects from Around the Globe," In *Digital Teleretinal Screening*, pp. 3-10. Springer Berlin Heidelberg, 2012.

[4] J. Polisena, K. Tran, K. Cimon, B. Hutton, S. McGill, and K. Palmer, "Home telehealth for diabetes management: a systematic review and meta-analysis," *Diabetes, Obesity and Metabolism* 11, no. 10 (2009): 913-930.

[5] S. M. Finkelstein, S. M. Speedie, and S. Potthoff, "Home telehealth improves clinical outcomes at lower cost for home healthcare," *Telemedicine Journal & e-Health* 12, no. 2 (2006): 128-136.

[6] D. P. Hansen, C. Pang, and A. Maeder, "HDI: integrating health data and tools," *Soft Computing* 11, no. 4 (2007): 361-367.

[7] C. M. O'Keefe, M. Yung, L. Gu, and R. Baxter, "Privacy-preserving data linkage protocols," In *Proceedings of the 2004 ACM workshop on Privacy in the electronic society*, pp. 94-102. ACM, 2004.

[8] T. Churches, and P. Christen, "Some methods for blindfolded record linkage," *BMC Medical Informatics and Decision Making* 4, no. 1 (2004): 9.

[9] D. L. Buckeridge, R. Mason, A. Robertson, J. Frank, R. Glazier, L. Purdon, C. G. Amrhein, et al., "Making health data maps: a case study of a community/university research collaboration," *Social Science & Medicine* 55, no. 7 (2002): 1189-1206

[10] J. Mohan, and R. R. R. Yaacob, "The Malaysian Telehealth Flagship Application: a national approach to health data protection and utilisation and consumer rights," *International Journal of Medical Informatics* 73, no. 3 (2004): 217-227.

[11] M. A Hebert, B. Korabek, and R. E. Scott, "Moving research into practice: a decision framework for integrating home telehealth into chronic illness care," *International Journal of Medical Informatics* 75, no. 12 (2006): 786-794.

[12] C. Goble, and R. Stevens, "State of the nation in data integration for bioinformatics," *Journal of biomedical informatics* 41, no. 5 (2008): 687-693.

[13] C. M. O'Keefe, "Privacy and the use of health data-reducing disclosure risk," *electronic Journal of Health Informatics* 3, no. 1 (2008): e5

[14] F. C. Tsui, J. U. Espino, V. M. Dato, P. H. Gesteland, J. Hutman, and M. M. Wagner, "Technical description of RODS: a real-time public health surveillance system." *Journal of the American Medical Informatics Association* 10, no. 5 (2003): 399-408.

[15] D. Detmer, M. Bloomrosen, B. Raymond, and P. Tang, "Integrated personal health records: transformative tools for consumer-centric care," *BMC medical informatics and decision making* 8, no. 1 (2008): 45.

[16] J. Grimson, G. Stephens, B. Jung, W. Grimson, D. Berry, and S. Pardon, "Sharing health-care records over the internet," *Internet Computing, IEEE* 5, no. 3 (2001): 49-58

[17] K. L. Taylor, C. M. O'Keefe, J. Colton, R. Baxter, R. Sparks, U. Srinivasan, M. A. Cameron, and L. Lefort, "A service oriented architecture for a health research data network," In *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*, pp. 443-444. IEEE, 2004

[18] D. J. Weitzner, "Beyond secrecy: New privacy protection strategies for open information spaces," *Internet Computing, IEEE* 11, no. 5 (2007): 96-95.

[19] S. R. Bird, W. Kurowski, G. K. Dickman, and I. Kronborg, "Integrated care facilitation for older patients with complex health care needs reduces hospital demand," *Australian Health Review* 31, no. 3 (2007): 451-461.

[20] P. A. Bernstein, J. Madhavan, and E. Rahm, "Generic schema matching, ten years later," *Proceedings of the VLDB Endowment* 4, no. 11 (2011): 695-701.

[21] L. Gu , R. Baxter, D. Vickers, and C. Rainsford, "Record linkage: Current practice and future directions," *CSIRO Mathematical and Information Sciences Technical Report* 3 (2003): 83.

[22] O. Bodenreider, "Biomedical ontologies in action: role in knowledge management, data integration and decision support," *Yearb Med Inform* 47 (2008): 67-79.

[23] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman, "Information accountability,"*Communications of the ACM* 51, no. 6 (2008): 82-87.

[24] C. W. Kelman, A. J. Bass, and C. D. J. Holman, "Research use of linked health data—a best practice protocol," *Australian and New Zealand journal of public health* 26, no. 3 (2002): 251-255.

[25] http://health.gov.au/ehealth-nbntelehealth (Accessed 27 September 2013)

[26] Australian Government, Department of Health and Ageing, "Building a 21st Century Primary Health Care System. Australia's First NationalPrimary Health Care Strategy", http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/Content/nphc-draft-report-toc/$FILE/NPHC-Draft.pdf (Accessed 27 September 2013)

[27] NHHRC, "A Healthier Future For All Australians – Final Report of the National Health and Hospitals Reform Commission – June 2009", http://www.health.gov.au/internet/nhhrc/publishing.nsf/content/nhhrc-report (Accessed 27 September 2013)

[28] Preventive Health Taskforce, "Taking Preventative Action - the Government's response to the report of the National Preventative Health Taskforce," http://www.preventativehealth.org.au/internet/preventativehealth/publishing.nsf/Content/taking-preventative-action (Accessed 27 September 2013)

[29] Australian Institute of Health and Welfare, "Australia's Health 2010," http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442452962 (Accessed 27 September 2013)

[30] R. E. Litan, "vital signs via broadband: remote health monitoring transmits savings," *enhances lives*, October 24, 2008

[31] Whole System Demonstrator Programme Headline Findings – December 2011. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/215264/dh_131689.pdf (Accessed 27 September 2013)

[32] A. Darkins, P. Ryan, R. Kobb, L. Foster, E. Edmonson, B. Wakefield, and A. E. Lancaster, "Care coordination/home telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions," *Telemedicine and e-Health* 14, no. 10 (2008): 1118-1126.

[33] J. W. Byun , E. Bertino, and N. Li, "Purpose based access control of complex data for privacy protection," In *Proceedings of the tenth ACM symposium on Access control models and technologies*, pp. 102-110. ACM, 2005

[34] Q. Ni, E. Bertino, J. Lobo, C. Brodie, C. M. Karat, J. Karat, and A. Trombeta, "Privacy-aware role-based access control," *ACM Transactions on Information and System Security (TISSEC)* 13, no. 3 (2010): 24.