Enhancing client welfare through better communication of private mental health data between rural service providers

Oliver Burmeister
Charles Sturt University
oburmeister@csu.edu.au

Md Zahidul Islam
Charles Sturt University
zislam@csu.edu.au

Miriam Dayhew
Charles Sturt University
mdayhew@csu.edu.au

Merrilyn Crichton
Charles Sturt University
mcrichton@csu.edu.au

Abstract

Client welfare is detrimentally affected by poor communication of data between rural service providers, which in part is complicated by privacy legislation. A study of service provision involving interviews with mental health professionals, found challenges in communicative processes between agencies were exacerbated by the heavy workloads. Dependence on individual interpretations of legislation, and on manual handling, led to delays that detrimentally affected client welfare. The main recommendation arising from this article is the creation of an ehealth system that is able to negotiate differing levels of access to client data through centralised controls, where the administration of that system ensures that it stays current with changing legislative requirements. The main contribution of the proposed model is to combine two well-known concepts: data integration and generalisation. People with mental illness are amongst the most vulnerable members of society, and current ehealth systems that provide access to medical records inadequately cater to their needs.

Keywords: Confidentiality; privacy; ehealth; rurality; trust models

1 Introduction

In Australia there are local, state and federal laws that govern what public and private mental health providers are able to do. The study reported here found that many providers, fearing possible litigation, overly err on the side of caution about sharing client data with other services. That, in turn, frequently disadvantages their clients because in mental health, a client is rarely served by a single professional. More frequently they will be served by a case worker, a general practitioner (GP), a clinical psychologist, a psychiatrist and/or other professionals, all with different information needs. While privacy legislation is part of a mental health professional's training, changes to legislation as well as heavy workloads in rural areas, mean that it is difficult for service providers to keep up. As a result, information sharing between agencies becomes complicated. There is therefore no question of the importance of designing usable processes and technologies for interagency communication, an argument that has been documented previously (Asan & Montague, 2013; Friberger & Falkman, 2011). What is unique to this article is that it uses qualitative interviews with mental health service users and professionals to inform a model of communication that integrates privacy legislative changes into designs for communication between rural service providers.

Such communication has to navigate complex legislative restrictions, yet the burden of interpreting such restrictions falls to individual professionals in each service. Helpful to

interagency communication would be a means of systematising the complex negotiations involved, in a way that removes the onus from individuals and places it upon a centrally controlled, automated ehealth system. The model proposed here involves a central server control information that is classified, through encryption, in order to only make available relevant information to those who need it, while also complying with the legislation as it is revised. The model also ensures fast access to accurate data related to a mental health patient.

This article discusses privacy, confidentiality and trust within the current Australian legislative framework and how that relates to mental health service users and providers. It then discusses the methodology that influenced the development of the computer system model. In the final section, the proposed model is described.

2 The Principle of 'Keeping Confidence'

Privacy and confidentiality are closely related concepts and significantly affect the development of ehealth systems. Generally the difference can be described as that privacy is about people, whereas confidentiality is about data. Confidentiality, arguably the more significant of the two for ehealth systems, has been viewed as informational privacy, that is, it is about the storing of private data. Allen (2011) claimed that privacy is about controlling the access of health information that is required by governments, institutions and individuals, whereas "many medical professionals, hospitals, insurers and other entities with access to health information regard maintaining the confidentiality of medical communications and the security of medical records as paramount professional responsibilities".

There are justifiably instances when breaches ought to occur and therefore not all breaches are unethical. One instance occurs regularly, when a client in a hospital voluntarily waives their right to confidentiality, for third party access, such as care givers other than the medical personnel treating them. Other ethical issues can be involved, such as when a person does not have the mental capacity to make such a decision for themselves. In such a situation the surrogate decision maker needs to be fully informed about the situation, so that they can make an informed decision as to whether or not to waive the client's right to confidentiality. Yet another instance occurs when an institution or care giver is compelled through a court order to release confidential information. Confidentiality thus concerns issues of restricting the flow of information. Professional codes of ethics contain statements that explicitly deal with confidentiality in computing (Bowern, Burmeister, Gotterbarn, & Weckert, 2006; Burmeister, 2013; Burmeister & Weckert, 2003) and in professions providing patient care (Bernoth, Dietsch, Burmeister, & Schwartz, 2014). Crichton (2008) analysed professional codes of conduct and found that ethical service phrases, such as "shall avoid misleading clients", "due care", and "accurate information" were indicators of the professional values of honesty, integrity and expertise.

Keeping confidence is an important ethical principle because it promotes an environment of trust between a mentally ill client and the professional who is caring for them. As noted above, there are exceptions in which it is recognised that the professional has an obligation to breach confidentiality, or in other words, an obligation to betray the trust of their client. Such situations are typically for the attainment of a greater good or for the prevention of harm, such as in an instance of suicidal behaviours or some other situation which contravenes their duty to protect the client, or to protect colleagues. In a mental health context, confidentiality is about managing the protection of private client data that has been disclosed in a situation of trust.

3 Privacy Legislation

The privacy of health information is a difficult matter in law as well as in professional practice. The issue has even been the subject of civil court cases with the most notable matter being the ownership of patient records determined in Breen v Williams (Breen v Williams, 1996) where the High Court of Australia held that a patient had no general right to copy, nor to inspect medical records relating to them and held by a health care provider. Magnusson and Opie

(1998) identified that the High Court finding upheld the NSW Court of Appeal finding and was characterised by the unanimous agreement and significant consistency of the reasoning of the judicial members of the Court, consequently making this a powerful authority.

It is well recognised that within the existing legislation there is the potential for inconsistency and for difficulties in interpretation and application. The final report from the NSW Law Reform Commission (2010) makes 104 recommendations to simplify the law, improve the consistency and ensure adequate coverage and protection for individuals. Recognising this level of complexity at the level of the Law Reform Commission would suggest that the average practitioner might have difficulty remaining current in their knowledge and understanding of their statutory obligations under State and Commonwealth legislation. For mental health care professionals in NSW, a State jurisdiction within the Commonwealth of Australia, the significant pieces of legislation are the Privacy Act 1988 (Cth), the Privacy and Personal information Protection Act 1998 (NSW) known as the 'PPIPA', and the Health Records and Information Privacy Act 2002 (NSW) known as the 'HRIPA'. There is also the 'GIPA' which is the Government Information (Public Access) Act 2009 (NSW) to make up the legislative framework concerned with information privacy protection. Changes to the Commonwealth Privacy Act were brought into play in mid-2014 and there are now far greater requirements for transparency in the handling of personal information. The individual whose information is held by a private practitioner, community agency or public sector agency can expect a greater level of understanding regarding the disclosure of their personal information. All entities must have a clear privacy management policy that details how they handle personal information and this must be available for their clientele. The changes are designed to enhance protection of personal information and are linked to strengthened powers for the Australian Information Commissioner to investigate and resolve complaints and enforce compliance. A significant issue for mental health providers would be the receipt of unsolicited private information, which must now be destroyed or deidentified. Another major issue for mental health practitioners supporting their mobile clients is the impact of sharing information across borders and it becomes the responsibility of the information holder to confirm for themselves that the recipient of information will manage that information in accordance with the Australian Privacy Laws.

Each health practitioner will also be guided by the codes of conduct specific to their professional accreditation bodies. The PPIPA covers personal information, that information from which a person's identity may be confirmed, but the PPIPA doesn't cover 'health information' (PPIPA Section 4a), this belongs under the auspices of the HRIPA. The HRIPA applies to: "... every organisation that is a health service provider or that collects, holds or uses health information". The term "organisation" means a public sector agency or a private sector person. (Health Records and Information Privacy Act 2002 - Section 11). Furthermore, the Privacy Act 1988 (Cth) identifies that: A permitted general situation exists in relation to the collection, use or disclosure by an APP entity of personal information about and individual if specific conditions are met (Section 16A). The most likely of these specific conditions for a mental health professional, would be where there is a threat to the life, health or safety of an individual or the public.

In the Privacy Act quoted above, an 'APP entity' is identified as an agency or organisation. There are also Statutory Guidelines released by the NSW Privacy Commissioner (http://www.ipc.nsw.gov.au/privacy/ipc_index.html#15) that support the use and disclosure of health information. These address the need to release personal details and health information for the purposes of management of a health service, research, and training purposes, where consent may not have been obtained from an individual, and the use of de-identified data is insufficient to fulfil the information requirements. There is a further guideline for the use of disclosure of health information obtained from a third party.

All health information is considered sensitive and private and so mental health information is not specifically addressed as a discrete type of information under the HRIPA, but there are further impacts on the sharing of mental health information addressed by the Mental Health Act 2007 (NSW) and the Mental Health (Forensic Provisions) Act 1990 (NSW). The disclosure

of information is protected under both these Acts. No information is to be disclosed without the consent of the individual except where the information is necessary for the administration or execution of the Acts or other lawful purposes such as legal proceedings. There are special provisions under Section 76J of the Mental Health (Forensic Provisions) Act 1990 that allow the sharing of information between corrective services institutions and health service institutions, through the development of information sharing protocols that support the transfer of information, without the express consent of the individual where that information is necessary for the care and treatment of the individual as prescribed under the Act.

A person who accesses the services of their GP and is referred to a counsellor can expect that there will be information shared between these two practitioners and there is an implied consent to sharing the information as part of an ongoing care program. If, however, the person then seeks voluntary admission to the local public hospital mental health facility then the hospital is going to have to seek consent from the patient to request access to material held by the GP or counsellor. On discharge the GP may receive a copy of a Discharge Summary, provided that the patient nominated that GP as part of their health record. However, if they have not been nominated, then no information regarding the admission will be able to be made available. If the person has travelled interstate then the State-based privacy legislations will need to be considered, and although there is a relative sameness and the principles are relatively constant, there will be discrete differences. The primary key to transferring information between practitioners is written consent from the client.

4 Methodology

In order to understand the nature of mental health services in rural Australia, an interpretive study utilising qualitative techniques to interpret meaning and to analyse participant understandings was undertaken. The methodology was informed by Schwandt (2003) who argued that such a process helps to explore how people in a particular social context interpret mental health services and reveal meanings that constitute those perceptions. Interviews were conducted with mental health service users and practitioners in the western Murray Darling Basin (MDB) region of NSW between November 2011 to July 2012. The key question for this study was "What mental eHealth services will best meet the future needs of the western Murray Darling Basin?"

4.1 Sampling

The interviewees were chosen on the basis of criterion sampling based on age, location, and length of experience, in order to give an understanding of professional experience across the region. The sample size of practitioners was limited to 27 participants spread geographically across the area of the western MDB. Sampling of mental health service users reached saturation after only thirteen interviews when it became evident a limited range of servicing patterns were present, with no new data emerging very early in the interview stages of the project. Ten of the thirteen participants recorded their mental health servicing experiences over a period of one month (30 days), which was used to support and provide depth to consumer perceptions and analysis.

The study was conducted in an area of Australia where service providers cover large geographical territories. Additionally, Australian government funding has meant that there are limited numbers of mental health practitioners and only one psychiatrist living in the area. Figure 1 shows service provision that is based in one location, and reaches into other areas. For instance, the region is poorly serviced by psychiatrists and therefore there are two psychiatrists that service the region, who fly in from Sydney. Another example is that of a GP who is based in Wagga Wagga, but who also services the needs of prisoners at the Junee detention facility. Still another example is that of Griffith based mental health case workers who regularly service clients in the Hillston local government area (LGA). Although attempts were made to accurately depict all mental health services in the region, it is possible that some services have been omitted.

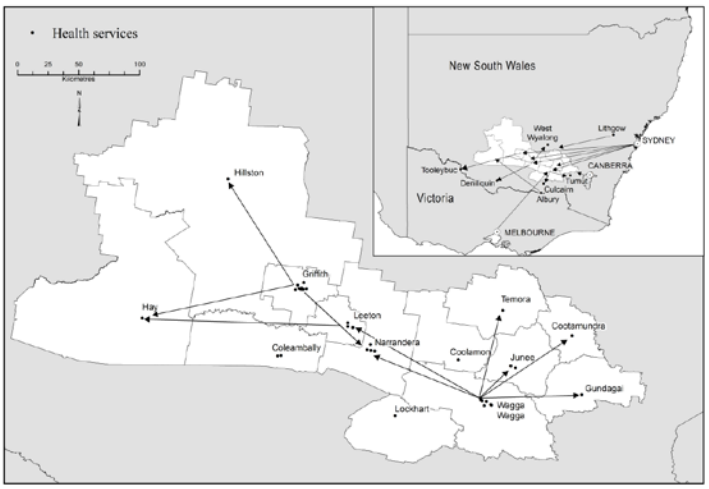

Figure 1: Mental health services available within the 13 Riverina Local Government Areas (Map produced by Spatial Data Analysis Network, Charles Sturt University, 2012).

5 Results

Regional mental health service providers deal with high case loads as well as large distances. This leads to difficulties in attracting and retaining staff, and to inadequate professional development opportunities, including training in changes to privacy legislation, knowledge of which directly impacted upon what professionals perceived they could or could not transfer about a client to another agency. The research revealed that clients at times experienced escalating problems because of the unavailability of their medical records, when transferring between rural services. In part this was due to the manual handling of records that could have been communicated electronically. These results are restricted to situations exemplifying issues of privacy in the western MDB. To protect participants the identifying details of people in these vignettes have been altered. The following examples illustrate the challenges of the distances involved. First there are exemplary quotations from rural service providers, then there is a vignette from a recipient of mental health care.

> *I have a client in Hay, Hillston, Goolgowi … Yanco … Barellan. I have had clients in West Wyalong … Narrandera … Ardlethan. … Hay is like a grey area, the case worker in Deniliquin normally services Hay, but if he's a bit overwhelmed then I pick it up.* (Case worker)

> *… way down on the Murray river, 40ks from Swan Hill. So nearly in Victoria. … We have such a huge volume of emails that come through our system and we expect clinicians to be able to follow policies and those policies are provided by email and they just cannot cope with the volume of work, when they're driving 4 hours they take, that takes a day out, sets them behind, you know 60 emails might come in that day.* (Service manager)

Jemma is a woman in her early 40s. She has four children, two are grown up and two are in primary school. Her early life was coloured by sexual and physical abuse. Jemma became a mother for the first time at 17. Jemma's diagnosis has been changed from Borderline Personality Disorder to that of Bipolar disorder. She sees a psychiatrist in Canberra, but lives in Wagga Wagga. With two small children, an active community life, and a limited capacity to work and therefore tight financial circumstances (she and her partner work as artists) travelling for treatment is a difficult process:

> *I drive over every 3 months to see my psychiatrist in Canberra … initially it was every 2 weeks … I drive for 2 and a half hours to get there, I have a 10 minute appointment with him, all he does is look at my bloods, order more blood tests "How are you going, everything okay?" "Good" "Yep see you later" that's it.*

5.1 Coordination between rural service providers

Lack of coordination is due to multiple reasons. For example, Jennie is a young, vibrant university student. Her mother (a mental health nurse) noticed that she was struggling to make friends and manage near the end of primary school. Jennie was put in touch with mental health services and has spent many years with the headspace team (headspace is an organisation providing treatment for people aged 12 to 25, who are experiencing mental illness) as well as community youth mental health. In the process of turning 18, Jennie was moved to the adult mental health system. She talked about the lack of communication between the organisations and having to repeat things many times to different people.

A mental health professional, addressed the difficulties on interagency coordination:

> *We've got a plethora of different software providers for different software but because they're only protecting their own little patch of territory they won't develop something that would interface into something else. … even with this building here we've got a headspace program and our national contract mandates we must use one form of software, but we've got a contract with another government department which mandates we must use this form of software. So you've constantly got staff juggling their way through two software systems and losing time there because the two software systems don't talk to each other.*

A mental health case manager with a government agency:

> *… if there was ways where we could access doctors a lot easier, or have the doctors understand the Centrelink process, and the information that we needed, we could probably cut down, a lot of red tape … to have a look at the clients patient record … see like the specialist letters or, just a specific part of it, not have full blown access maybe, but just a specific part where you can go in, and go okay, I need a bit more information about this, without having then, to contact the doctor. Because I'm busy, they're busy, often you play phone tag, there's a lot of doctors who don't want to talk to you … if you know that a client has seen a particular specialist, like a neurologist for example, if you can get access to their letters, because all specialists write a letter back to the GP ...I had a situation a couple of months ago, where I was trying to get, a psychiatrist letter … With this particular case I found out that the psychiatrist had written the report. It was sitting on the customers file for 3 weeks, while the case manager went on holidays, and nobody bothered to deal with it.*

6 A Privacy Preserving Data Sharing Model

The proposed model arises from the findings. It is apparent that a mental health client usually interacts with a number of service providers (parties). Each of these parties collects data, which can be useful for other rural service providers as well. Furthermore, a client often needs to provide a party with necessary information collected and stored by another party. However, due to various privacy issues, generally a hospital or a medical centre may not want to supply a client's sensitive information to a social worker. Therefore, for clients having serious mental health issues such as dementia, it can be very difficult to organize the transfer of his/her data from one party to another party. Electronically storing all information (about a client) in a password protected central server may not be an adequate solution given that mental health clients, such as those with dementia, may forget their password. Even if a mental health client can recollect the password, he/she may not be able to take the appropriate decision (due to their illness) on what information they need to release to a health service provider.

Different health service providers need to have access to different levels of personal health related information. That is, the level of health information required by a GP can be different to the level of information required by a case worker. For example, while a GP may need the exact disease of a patient (i.e. whether it is Dementia or Alzheimer's disease) a case worker may only need a higher level of information such as mental illness. Often there are also legal restrictions on the level of information that can be accessed by a party. A solution is a privacy

preserving data sharing model (see Figure 2), based on data integration and generalization/aggregation. Data integration is a well-known technique (Batini, Lenzerini, & Navathe, 1986; Hass et al., 1999; Xu, Zhang, & Dong, 2006) that integrates data from heterogeneous sources into one single system. Data integration is commonly applied when two or more organisations merge into one due to various reasons such as a government decision and one organisation acquiring another. Generalisation is another well-known approach that is used for a completely different purpose: to preserve individual privacy (Aggarwal & Yu, 2008; Verykios, 2004). For example, the age of a patient, say 34, is often generalised into an age category, say 30-40. A main goal of the generalisation is to prevent record re-identification and breaches of individual privacy.

While data integration and generalisation are two well-known concepts they are typically used in completely different contexts and purposes. The main contribution of our proposed data sharing model is combining them into a single model to allow communication between rural service providers, for solving the issues (as described before) of a mental health client. The proposed model allows all parties that are involved with a mental health patient to store their data in a trusted central server (instead of only storing them in their individual computers separately) and access the data in the appropriate permission level. Unlike the traditional generalisation techniques (Aggarwal & Yu, 2008; Verykios, 2004) our model applies different levels of generalisation for different parties to suit their access permission levels.

A problem in the interagency communication of the competing organisations is the absence of a trusted third party (a central server) that can make sure that all privacy sensitive data and logic rules (patterns) are appropriately handled. However, in the context of this study we propose a government run model where an advantage is the presence of a trusted third party, such as the Ministry of Health, that can protect the interests of all parties and follow the law of the land. A key issue of the proposed model is the dynamic data integration in the central server and automatic adaptation of the changes of privacy legislation in the data generalisation process. A truly dynamic data integration process allowing us to integrate the data from various sources that are unknown at the design time can be very challenging (Wang, Yu, & Zhang, 2009). If the data schemas of the various sources are unknown at the design time then it can be difficult to integrate them completely automatically. For example, in one schema the diagnosis of a disease can be labelled as “Disease” while in another schema the same attribute can be “Diagnosis”. The unit of the attribute “Weight” can be Kg in one schema whereas the unit of the same attribute can be Pound in another schema. An attribute value can be called differently (for example, Coke and Coca Cola) in different schemas. Sometimes different schemas can use different attributes where one attribute is a function of the other attribute. The integration of the schemas therefore may require human intervention and domain knowledge. Additionally, the automation of the generalisation process to adapt the changes in privacy legislation can be a difficult task since the type of changes are unpredictable. Therefore, in order to maintain a high accuracy and quality our proposed model uses a semi-automated data integration approach where the integration of the data can be done automatically for any additional data as long as the schemas of the sources are not changed. However, any existing dynamic data integration technique can be used if that is found suitable.

Although designing a suitable database schema for data integration and generalisation will require the involvement of a database administrator, it is required to be done only at the server level. Moreover, it needs to be re-designed only if a new party joins, an existing party changes its database schema or privacy legislation changes. In such cases typically a minor adjustment at the server end is sufficient.

6.1 A model to improve client welfare outcomes and legislative compliance

In the proposed model we have a central trusted server which stores all original information collected from the parties involved. The parties can be different health service providers such as hospitals, GPs and age care centres. Our model first analyses the properties (such as access permission levels) of the parties and then groups them into categories, where in each category the parties have exactly the same properties. It is not unlikely to have some categories having

only a single party where its property does not match with any other parties. Each party regularly uploads its data to the central trusted server. The data are first encrypted using an encryption filter as shown in Figure 2 before sending them to the server through the internet. In Figure 2 there are three parties (Party A, Party B and Party C) with exactly the same properties and grouped as Category 1. Party A is uploading its new data in the trusted central server. It first encrypts its data and then sends it to the server through the internet. The encrypted message contains the identification information of the sender party along with the actual data. Any suitable asymmetric encryption technique can be used (Rafaeli & Hutchison, 2003; Simmons, 1979). Note that each party only communicates with the trusted server directly and therefore, only needs to know the key of the server. The server of course needs to know the keys of all parties.

Collected data are then decrypted, integrated and pre-processed through the decryption, data integration and data pre-processing filter, respectively as shown in Figure 2. Data integration is required at the server since different parties can have different database schema. For example, as seen in the findings section above, public and private mental health providers frequently use differing software systems to record client data, and therefore, as shown in Figure 3, different parties can name the same table and the same attribute differently. Moreover, different parties can use different units for the same attribute. Some parties can have tables that are not stored by other parties. Figure 3 shows an example where parties belonging to Category 1 have two tables, namely Patient and Doctor, whereas parties belonging to Category 2 have two tables called Client and Social Worker. The same Patient table is called differently as Patient and Client by the categories. Moreover, the same attribute Diagnosis is also called differently as Diagnosis and Disease by the categories. In this example (Figure 3), both categories have a table that is not stored by the other category. The server takes all these facts into account and integrates the data into one consistent database schema. For example, the server creates a database from the tables of the two categories, where it finally contains three tables. It combines the attributes of the Patient and Client table and calls the table as Patient. It also calls the attributes Diagnosis and Disease as Diagnosis in its Patient table. In the Client table the identifier attribute is called C_ID (Client ID), whereas the same attribute is called P_ID (Patient ID) in the Patient table of the server. Records from different tables belonging to different categories can be matched in the combined table of the server through a global ID, like the driver license number or social security number.

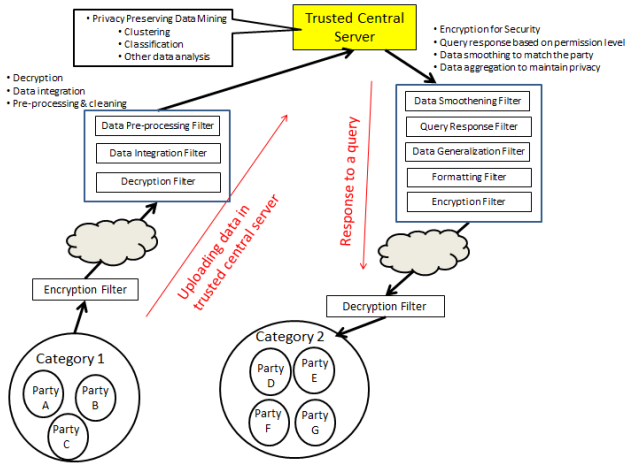

Figure 2: Overall block diagram of the privacy preserving data sharing model.

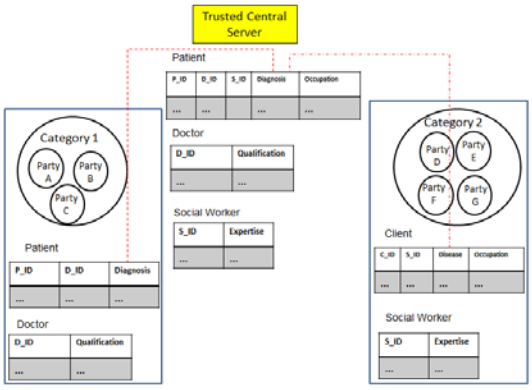

Figure 3: Data integration in the trusted central server.

Note that for each category, the server stores the necessary metadata describing the mappings (for the tables and attributes such as Disease to Diagnosis), and access permission levels for each attribute of the tables. Different types of data and database schema are often transformed and mapped into another type, for various reasons including building a data warehouse using heterogeneous data sources and generating a dynamic webpage using an underlying database (Batini et al., 1986; Hass et al., 1999; Xu et al., 2006). In our privacy preserving data sharing model we convert the data (uploaded in the server by a party) in order to suit the database schema of the server. For example, when Party D (Figure 3) uploads a record in the Client table it either inserts a new record or updates an existing record in the Patient table of the server. The C_ID attribute of the Client table and the P_ID attribute of the Patient table can be matched through an external ID. The metadata stored in the server for Category 2, helps the server to map the Client table to the Patient table, and the Disease attribute to the Diagnosis attribute.

Through the data pre-processing filter, our model then performs data pre-processing for data cleansing and missing value imputation (M. G. Rahman & Islam, 2014). Data cleansing techniques automatically identify any incorrect record or incorrect value in a table. The identified incorrect value is then typically amended by first considering it as missing and then making an educated guess using an missing value imputation technique. Generally the accuracy of the incorrect data identification and missing value imputation is very high. However, instead of automatically changing an uploaded data, our model reports an incorrect value with a suggested possible correction. The party then requires re-uploading of the data after necessary correction or ignoring the warning/suggestion. The pre-processing step aims to ensure a high quality of the data, which are stored in the central server.

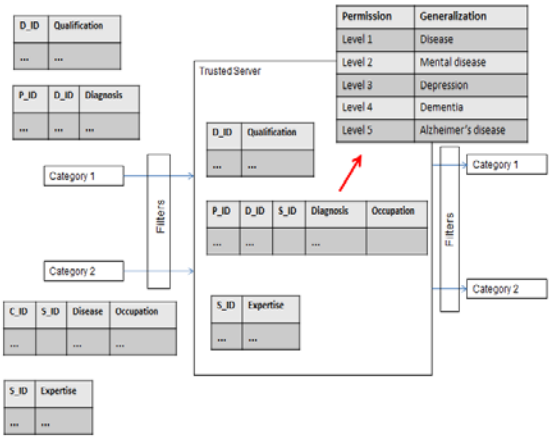

Figure 4: An example of generalization of data.

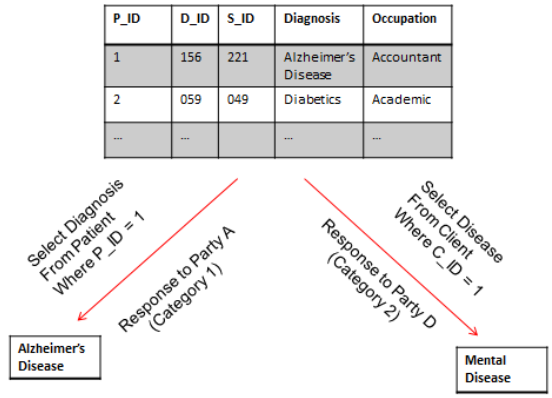

Figure 5: Different responses for a similar query

Once the data are stored in the trusted central server they can be used by any party (Figure 2 and Figure 4) in many different ways, including simple query response for providing information as requested by a party, and privacy preserving data mining through various data mining tasks such as clustering and classification for knowledge discovery. For query response, the server first uses the Data Smoothening Filter (see Figure 2) that changes a query made by a party into an appropriate query for the execution at the server. For example, a party may send a query as shown in Figure 5 where it asks for the disease name of the client having id = 1 from the Client table. However, the Client table is called the Patient table, and the attribute Disease is called the attribute Diagnosis in the server. The Data Smoothening Filter changes the table name from Client to Patient and the attribute name from Disease to Diagnosis knowing that the query is made by a party from, say Category 2, as reported by the Query Response filter in Figure 2.

While responding to a query the server applies necessary generalization functions, through the data generalisation filter, based on the access permission level of the party generating the query, as shown in Figure 2, Figure 4 and Figure 5. For example, based on the access permission level of a party the exact diagnosis ("Alzheimer's disease") of a patient can be generalized as "Mental disease". The level of generalization is related to the level of access permission. The more access permission results in the less aggregation/generalization. Each level of access permission will correspond to a level of generalization, where the lowest (minimum) level of access permission will have the maximum generalization. For example, a diagnosis "Alzheimer's disease" can be generalized as "Dementia", "Mental disease" or "Disease" depending on the access permission level of a party asking for the information. The party having the maximum access permission to the attribute can see the query result as "Alzheimer's Disease", whereas the party having a lower access permission can get a query response as "Mental Disease" (Figure 4 and Figure 5). Moreover, through the Formatting filter a query response is reformatted according to the data format of the party generating the query. For example, the weight of a patient can be returned as 50 kg or 110 pounds depending on the data format of the party. The whole process of data integrations, query smoothening, data generalization, query response based on access level and response reformatting remains transparent to a party.

The parties can also apply different data mining algorithms including decision tree (Quinlan, 1996), decision forest (Abellan & Masegosa, 2009) and clustering (Ji, Pang, Zhou, Han, & Wang, 2012; Kashef & Kamel, 2009; M. A. Rahman & Islam, 2011) on the dataset stored in the trusted server. While a party may not have full access to all data stored in the server it has access to complete results produced by the data mining algorithms. We consider the output of data mining algorithms as general patterns, which are not sensitive to individual privacy, and are public knowledge. For example, a decision tree algorithm (Islam, 2012) returns a tree containing a set of logic rules that explain the rules leading to a class value, say disease type.

An example of a logic rule can be "Diabetics = yes & Blood Pressure = yes & Education = PhD & Head Injury = yes → Alzheimer's (yes: 550, no: 20)" suggesting that people having diabetics, high blood pressure, previous history of head injury and a doctorate degree (i.e. people fulfilling the pre-condition) have high chance of Alzheimer's disease. Moreover, there are altogether (550+20) 570 patients, in the dataset, satisfying the pre-condition and 550 patients out of them have Alzheimer's disease. While disclosure of the diagnosis of a particular patient is considered as a breach of individual privacy, the discovery of the logic rule as explained above is not considered as a breach of privacy. The information revealed by the logic rule is generally considered as public knowledge involving a group of people (Islam & Brankovic, 2011). For example, the fact that smokers have higher risk of cancer is a public knowledge, and is not considered as a breach of individual privacy.

In order to facilitate data mining, a number of datasets can be manually or automatically prepared from the tables stored on the server. A party can only view the metadata (such as the names of the attributes and their domain values) of the datasets, and cannot view the actual records. However, a party can apply data mining algorithms on the datasets using the tools such as WEKA provided by the server. The algorithms are run on the datasets in the server, and the datasets are not required to be downloaded to the party's local machine. A party is also allowed to choose necessary parameters such as the attributes to use, any attribute as the class attribute (such as Diagnosis) for classification tasks, and the algorithm-specific parameters such as the minimum number of records per leaf of a tree. Several necessary restrictions such as the minimum number of records per cluster and the minimum number of records per leaf of a tree, are enforced in order to preserve individual privacy. Therefore, a user cannot learn a logic rule or cluster having a single patient (or a minimum number of patients), resulting in a protection of sensitive individual information from unauthorized parties.

An advantage of such a centralized data mining model is that it allows a party to extract general knowledge from the datasets without allowing them to learn sensitive information about an individual. Therefore, the proposed model capitalises on the advantages of privacy preserving data mining (Islam & Brankovic, 2011). There are also many other advantages of the proposed privacy preserving data sharing model. It allows the parties to collaborate and share data among themselves so that a party can access all the necessary information in order to provide an accurate and high quality service, without being required to ask a patient to arrange the transfer of his/her data from one service provider to another. Without the model it could be difficult for a party (service provider) to collect all data for a patient. For example, a party may not know which other party to talk to or how many parties to talk to in order to collect all data for a patient. Moreover, as seen in the example of the mental health nurse above, the party may not know what types of data are available about a patient. Similarly, the party may not have an idea on the number of available attributes and their names in the centralized dataset. However, when the data from all parties are stored in the central server, a party can get all information of a patient by using a simple query like select * from patient where patient.id=1. Such a query produces a response informing all attribute values stored in the Patient table for the patient, following the access permission level of the party. If a party has high access permission it gets low generalized data. Moreover, a party can explore the database schema for the central server to learn more about the available data.

Another advantage of the model is the simplicity of the encryption/decryption model. Each party only needs to talk to the central trusted server instead of all other parties. Therefore, each party only needs to know the key of the server resulting in a simple encryption/decryption process. It is more secure to communicate with a single trusted server (in a client server model) than to communicate with the peers in a peer to peer model (Rafaeli & Hutchison, 2003; Simmons, 1979). Still another advantage of a centralised model is that the onus of complying with legislative changes to privacy does not rely on individual, over-worked mental health professionals. Instead compliance can be centrally administered and controlled. Finally, from a managerial perspective, another advantage of data mining on a combined data set is that it may reveal interesting patterns that a party otherwise might not be able to extract. Various

data mining tasks can be performed resulting in useful pattern discovery involving all records and attributes – which otherwise would not be possible for a party to perform.

7 Conclusion

The study reported here involved qualitative, in-depth interviews with mental health professionals and recipients of mental health care, in the western MDB. One of the findings of the study was a need for better management of client records, when they are communicated between providers of services. Furthermore, it was found that the challenges experienced in the region were in part unique to the rurality of the services, exemplified in the excessive workloads of staff, and the long distances they needed to travel to service clients. A complicating factor is that privacy legislation is constantly being updated and it is different for the public and private mental health service providers and different again between states. Therefore it is unreasonable to expect rural health workers to also keep current with privacy legislation, which is not core business for them. Instead what is needed is a system that removes that responsibility from the health worker in a way that ensures compliance with legislative changes. The Privacy Preserving Data Sharing Model solves these challenges and can be transferred to other medical contexts, outside mental health and outside the rural Australian setting. It is possible to encrypt confidential data, smooth it, aggregate it or in other ways restrict access to it, such that centrally controlled data can be widely accessed by professionals servicing a particular client. The model facilitates accuracy and timely data access, whilst preserving confidential data and restricting access as appropriate to private information about the client. This speeds access to client data, improves the servicing of the client, and reduces lengthy visits in which client histories need to be repeated. The model allows all parties to access data on a patient (irrespective of who collected the data) according to the access permission level (following the privacy legislation) of a party. The model facilitates better professional practice in regards to compliance with privacy legislation, whilst simultaneously enhancing the welfare of clients through more efficient communication of their data between service providers.

8 Acknowledgements

This project was funded by a grant from the Strategic Projects Branch, Department of Regional Australia, Regional Development and Local Government, Canberra. The team included the now retired Dr Muenstermann.

References

Abellan, J., & Masegosa, A. (2009). An Ensemble Method Using Credal Decision Trees. *European Journal of Operational Research, 205*(1), 218-226.

Aggarwal, C., & Yu, P. S. (2008). A General Survey of Privacy Preserving Data Mining Models and Algorithms. *The Kluwer International Series on Advances in Database Systems, 34*, 11-52.

Allen, A. (2011). Privacy and Medicine. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Asan, O., & Montague, E. (2013). Technology-mediated information sharing between patients and clinicians in primary care encounters. *Behaviour & Information Technology*, 1-12. doi: 10.1080/0144929x.2013.780636

Batini, C., Lenzerini, M., & Navathe, S. B. (1986). A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys, 18*(4), 323-364.

Bernoth, M., Dietsch, E., Burmeister, O. K., & Schwartz, M. (2014). Information Management in Aged Care: Cases of Confidentiality and Elder Abuse. *Journal of Business Ethics, 122*, 453-460. doi: 10.1007/s10551-013-1770-7

Bowern, M., Burmeister, O. K., Gotterbarn, D., & Weckert, J. (2006). ICT Integrity: Bringing the ACS Code of Ethics up to date. *Australasian Journal of Information Systems, 13*(2), 168-181.

Medical Records Access case, HCA 57 C.F.R. (1996).

Burmeister, O. K. (2013). Achieving the goal of a global computing code of ethics through an international-localisation hybrid. *Ethical Space: The International Journal of Communication Ethics, 10*(4), 25-32.

Burmeister, O. K., & Weckert, J. (2003). Applying the new software engineering code of ethics to usability engineering: A study of 4 cases. *Journal of Information, Communication & Ethics in Society, 3*(3), 119-132.

Crichton, M. (2008). *The Praxis of Voluntary Service: An investigation of the logic of service in Rotary and Zonta*. (PhD PhD), Queensland University of Technology, Brisbane.

Friberger, M. G., & Falkman, G. (2011). Collaboration processes, outcomes, challenges and enablers of distributed clinical communities of practice. *Behaviour & Information Technology, 32*(6), 519-531. doi: 10.1080/0144929x.2011.602426

Hass, L. M., Miller, R. J., Niswonger, B., Tork Roth, M., Schwarz, P. M., & Wimmers, E. L. (1999). Transforming Heterogeneous Data with Database Middleware: Beyond Integration. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 22*(1), 31-36.

Islam, M. Z. (2012). *EXPLORE: A Novel Decision Tree Classification Algorithm*. Paper presented at the 27th British National Conference on Databases (BNCOD 2010), Dundee, Scotland.

Islam, M. Z., & Brankovic, L. (2011). Privacy Preserving Data Mining: A Noise Addition Framework Using a Novel Clustering Technique. *Knowledge-Based Systems, 24*(8), 1214-1223.

Ji, J., Pang, W., Zhou, C., Han, X., & Wang, Z. (2012). A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. *Journal of Knowledge-Based Systems, 30*, 129-135.

Kashef, R., & Kamel, M. S. (2009). Enhanced bisecting k-means clustering using intermediate cooperation. *Journal of Pattern Recognition, 42*(11), 2557-2569.

Magnusson, R., & Opie, H. (1998). Patient Access to Medical Records: Fiduciary Duties and Other Issues - A Classroom Interactive. *University of Tasmania Law Review 99, 17*(2).

NSW Law Reform Commission. (2010). *Protecting privacy in New South Wales*. NSW Law Reform Commission.

Quinlan, J. (1996). Improved use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research, 4*, 77-90.

Rafaeli, S., & Hutchison, D. (2003). A Survey of Key Management for Secure Group Communication. *ACM Computing Surveys, 35*(3), 309-329.

Rahman, M. A., & Islam, M. Z. (2011). *Seed-Detective: A Novel Clustering Technique Using High Quality Seed for K-Means on Categorical and Numerical Attributes*. Paper presented at the Ninth Australasian Data Mining Conference (AusDM 11).

Rahman, M. G., & Islam, M. Z. (2014). FIMUS: A Framework for Imputing Missing Values Using Co-appearance, Correlation and Similarity Analysis. *Knowledge-Based Systems, 56*, 311-327. doi: 10.1016/j.knosys.2013.12.005

Schwandt, T. A. (2003). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructionism. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research* (2 ed., pp. 292-331). Thousand Oaks, CA: SAGE.

Simmons, G. J. (1979). Symmetric and Asymmetric Encryption. *ACM Computing Surveys, 11*(4), 305-330.

Verykios, V. S. (2004). State of the Art in Privacy Preserving Data Mining. *ACM SIGMOD Record, 33*(1), 50-57.

Wang, J., Yu, A., & Zhang, L. Q. (2009, 8-9 August). *A dynamic data integration model based on SOA*. Paper presented at the ISECS International Colloquium on Computing, Communication, Control, and Management (CCCM 2009), Sanya, China.

Xu, Z., Zhang, S., & Dong, Y. (2006). *Mapping between Relational Database Schema and OWL Ontology for Deep Annotation*. Paper presented at the IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA.

An earlier version of this paper was presented at the Australasian Conference on Information Systems (ACIS) 2014 in Auckland, New Zealand.