Privacy Concerns of the Australian My Health Record: Implications for Other Large-scale Opt-out Personal Health Records

Patrick Cheong-Iao Pang, Dana McKay, Shanton Chang, Qingyu Chen

- School of Computing and Information Systems, The University of Melbourne, Parkville, VIC, Australia

Xiuzhen Zhang, Lishan Cui

- School of Science (Computer Science and Information Technology), RMIT University, Melbourne, VIC, Australia

Corresponding Author:

Patrick Cheong-Iao Pang
School of Computing and Information Systems
The University of Melbourne
Parkville VIC 3010, Australia
mail@patrickpang.net

Abstract

Personal health records offer the convenience of accessing medical history and personal health information, but also raise a range of privacy concerns which affect their adoption. In 2018, the Australian nationwide personal health record, My Health Record (MHR), was changed to an opt-out model, meaning that users are automatically enrolled unless they opt out. This significant change sparked wide-ranging and vociferous discussions of the privacy concerns of MHR on Twitter thus provided a lens into people's concerns. This lesson offers useful insights for improving MHR and better implementing future large-scale health records. By using qualitative coding and topic modeling on Twitter data, we categorized the stakeholders who participated in the discussions and the privacy concerns expressed. We have identified 10 categories of stakeholders and 9 types of privacy concerns in the discussions, and our analysis finds that these stakeholder groups focused on different privacy aspects of MHR. This work implies that, for future provisions of similar systems, it is important to involve these stakeholders in the design and address their privacy concerns early, as they are interested in providing input and their strong opinions may influence the uptake of such systems. Based on the lesson gleaned from this case, we propose that system owners can proactively communicate the privacy and the security aspects of their PHRs with different parties on social media. We also highlight some suggestions for improving the consent model and third-party access to personal health records in this paper.

1. Introduction

Personal health records (PHRs) allow patients to access their medical history and personal health information (Tang et al., 2006). Many countries in the world have deployed PHR systems because of their anticipated benefits (Berge, 2016). In Australia, the Commonwealth Government has implemented a national PHR system called MHR. As a nationwide system, MHR is designed to connect different stakeholders to a single information system for providing better healthcare services by the collaborative use of personal health data. For instance, general practitioners and hospitals can upload data to the system, while pharmacists and allied health professionals can access patients' information to provide better services. The ability of multiple stakeholders to access, upload and share information among different stakeholders is one of many features of MHR (Australian Digital Health Agency, 2019a). As such, the acceptance by *all* stakeholders is crucial to the success of MHR.

The original MHR was introduced through the My Health Records Act 2012 (Cth), which initially stated that it was a voluntary national system for healthcare consumers to access their health information. In 2015, further legislation was introduced to allow the Government to implement an opt-out system in 2018. In the original opt-in model, people could choose to enroll if they wished. After the change, eligible Australians would have a record automatically created in MHR unless they explicitly excluded themselves during the opt-out period. After the transition to the opt-out model, new users to the healthcare system (e.g. new-born babies or new migrants) were forcibly included unless they applied to remove themselves from the system. In response, people who were not originally included in the system expressed high level of privacy concerns such as potential leaks and misuse of their data. Meanwhile, other stakeholders (such as doctors) concerned about the legal risks of adding and using patient data in this system after such a dramatic change (McCall, 2018). Privacy is one of the main barriers for low adoption of healthcare information technologies and PHR systems; understanding these concerns by different stakeholders is therefore key to adoption.

Previous work also suggests that privacy concerns affect the adoption of PHRs (Abdekhoda et al., 2019; Alyami and Song, 2016; Kenny and Connolly, 2015), however, most of the studies focus on the patient perspective and seldom investigate which stakeholders demonstrate privacy concerns and what their concerns are. According to our preliminary study (Pang and Chang, 2019), different categories of MHR stakeholders were willing to participate in online discussion regarding various privacy concerns, but further investigations are needed to understand what these privacy concerns are and how to address them. Without knowing the stakeholders and their concerns, it is hard to engage with them and draw on their feedback to improve the rollouts of PHRs. With the diversity of Twitter users and the multi-faceted content, Twitter can be a platform for researchers to understand these issues and derive insights for the implementations of systems. The findings not only are helpful for identifying the drawbacks of MHR and help the Australian digital health sector moving forward, but also provide guidance for designing and implementing similar large-scale PHR systems. Given this background, we intend to address these research questions in this study:

- *RQ1: What categories of stakeholders participated in the online discussion about the privacy concerns?*
- *RQ2: What are the privacy concerns of the stakeholders of PHRs?*
- *RQ3: What are the practical implications for the implementation of similar large-scale PHR systems?*

Our work highlights the needs of rethinking the design and the features of PHR systems in different perspectives. First of all, the case of MHR implies that the privacy of health records cannot be viewed as one-size-fits-all, instead, the role relative to the system is an important factor to predict the concerns. Secondly, for the provisions of similar large-scale PHR systems in the future, we suggest that the enthusiasm and the input of stakeholders can be better utilized by including them in the design stage, and social media is a potential venue for collect opinions for large-scale electronic health records. Finally, we propose some improvements in the areas of the consent model and the third-party access for addressing privacy concerns. These insights will benefit the practitioners who work in the digital health sector.

This paper is structured as follows. We start with the literature review and the details of our data collection and research design. Then, we present the results of our analysis and discuss our major findings, followed by the implications for the design and the implementations of future PHRs.

2. Literature Review

In this section, we report on the literature about PHRs and their relationship with the Australian MHR implementation. Secondly, we present an overview of privacy concerns in the context of PHRs, followed by a review of recent social media analyses and their applicability to our work. The search of relevant literature was carried out by all authors collectively, using journal databases such as Web of Science, PubMed, and the Association for Information Systems e-Library. Google Scholar was used to capture conference proceedings and gray literature.

2.1. Personal Health Records

PHRs are usually treated as a sub-category of electronic health record systems but the two types of health records have several differences. Electronic health records, which are intended to be used by healthcare professionals (Flaumenhaft and Ben-Assuli, 2018), are digitized health data of patients stored in a healthcare provider without the ability for patients to interact with their data (Roehrs et al., 2017). Their purposes include supporting decision making and improving the management of patient data among different organizations such as hospitals and laboratories (Roehrs et al., 2017). On the other hand, PHRs are patient-oriented systems that enable individuals to access, manage and share their health information (Tang et al., 2006). While the concept of PHR is not new, there is still no agreement on an exact definition

(Flaumenhaft and Ben-Assuli, 2018). Literature suggests that the common usages of PHRs include the use by patients (alongside clinicians and professionals) (Abdekhoda et al., 2019; Li, 2015), enhancing self-efficacy and self-management of health conditions (Abdekhoda et al., 2019; Kenny and Connolly, 2015; Li, 2015), and allowing other authorized personnel to access (e.g. patients of underage children) (Bourgeois et al., 2015; Mxoli et al., 2014). The putative patient-centric nature of PHRs means that there will be different privacy needs and expectations, and researchers and designers should consider the two types of systems separately.

PHRs can further be broken down into multiple sub-types. Li (2015) defines the architecture of PHRs as tethered and untethered. A tethered PHR refers to a closed solution which ties to the electronic health records of a particular healthcare provider, for their patients to access parts of their records stored in the records. Tethered PHRs have only a single data source and low interoperability with other systems and organizations. In contrast, an untethered PHR solution offers direct and portable access to many healthcare providers and personal health information can be shared and accessed among multiple organizations. PHRs can act as adaptive platforms in recent work, which can collect data from multiple electronic health records and data sources while preserving the context of data (Roehrs et al., 2017). MHR falls into the latter category of PHR, as it is nationwide and integrates with various information systems (such as medical histories, pathology, etc.) of different healthcare providers. A study also finds that the perspectives from PHR stakeholders, such as patients, health professionals, health technology organizations, government agencies and researchers, are important to the adoption of PHRs (Gagnon et al., 2016), particularly when the system involves many different parties. This supports our motivation to investigate the views of MHR's stakeholders.

2.2. My Health Record and Privacy

As currently implemented, the Australian MHR system is a secure online summary of key health information for individuals that theoretically allows them to manage the content and share necessary information with the stakeholders in the healthcare system (Australian Digital Health Agency, 2019a). The Australian Digital Health Agency (ADHA) is the governmental authority for the operations of MHR. MHR was introduced in 2012 as the Personally Controlled Electronic Health Record (PCEHR), which was in an opt-in basis and had only about 2.4 million registrations in 2015 (Parliament of Australia, 2015). After a review of the system in 2013, in order to boost the enrollment and maximize the benefits, the Government

renamed it to MHR and decided to move to an opt-out model, citing that health record systems in countries such as the UK and New Zealand had achieved better acceptance with opt-out models provided that safety and security issues were addressed (Royle et al., 2013).

Researchers have shown that the patients using MHR have concerns to the legal, privacy and security issues of the system (Haddad et al., 2016; Muhammad and Wickramasinghe, 2017), but it is not clear about the views of other MHR users, such as clinicians and community health organizations, in the privacy and security aspects. Although two opt-out trials in Northern Queensland and the Nepean Blue Mountains areas (Department of Health, 2017) were conducted by the Government, they failed to capture the major issues of the opt-out model, as experts believed that "the concept is good" but the implementation was shocking (McCall, 2018). While it is not easy to understand the underlying problems in the implementation, the discussions of MHR on Twitter allow us to peek at the core issues.

In addition, many studies about MHR were carried out before the opt-out model was introduced, therefore, there is a need to refresh our understanding and find out why the implementation went wrong, so that governments and health organizations can avoid the same mistakes in the future. A recent paper suggests that the change of the enrolment model has fueled public concerns about the privacy of MHR (Kariotis et al., 2019), and this is not without reason. According to a survey of World Health Organization (2016), Australia lacked policies or legislation to protect the privacy of personal identifiable data in digital format, comparing with other developed countries such as the United States, the UK and New Zealand. On the other hand, another study suggests that the reluctance from health professionals also contribute to the low rate of adoption (Almond et al., 2016). These clearly state that MHR is a complex information system that needs proper design, stakeholder management and legal support. Therefore, its case can provide valuable insights for future rollouts of other large-scale digital health records.

2.3. Privacy Concerns in the PHR Context

There are many reasons for concern about health information privacy. First of all, health discrimination is commonplace in employment (Dray-Spira et al., 2008; Roessler et al., 2011; Sharac et al., 2010), in education (Jung, 2003, 2002) and even in medical practice (Earnshaw and Quinn, 2012; King, 1989). It is even enacted through legislation: Australian consumers who access information about their own genetics are legally obliged to disclose it to insurers,

who may then use it to discriminate against them (Otlowski et al., 2019). Given this threat of discrimination, it is no surprise that health is one of the pieces of information characterized as "sensitive" under Australian law (Office of the Australian Information Commissioner, 2019), and one about which people are understandably concerned.

Privacy factors have affected the adoption of PHRs since their inception. Archer and Cocosila (2014) suggest that security, privacy and trust affect the perceived usefulness of PHRs. In addition, people highlight concerns regarding the access and the secondary use of data stored in PHRs (Abd-alrazaq et al., 2019; Angst and Agarwal, 2009; Bourgeois et al., 2015). These studies posit that privacy concerns are main barriers to prevent the adoption of PHRs. In Australia, people have similar concerns about unauthorized access, non-clinical use and data sharing, and they have adequate trust in the government agencies who run the MHR (Muhammad and Wickramasinghe, 2017). Notably, these concerns were measured when the MHR initiative was an opt-in model. The introduction of an opt-out model changes privacy perspectives. Compelling people to be included in the system will raise new privacy concerns: patients will make different risk and benefit calculations than healthy people (Rahman, 2019). Kariotis et al. (2019) have already shown that people have concerns, therefore, understanding what these concerns are and how they differ between groups could inform the design of a better and more acceptable PHR.

Some existing information privacy models can help us to understand the privacy concerns in context of MHR and PHRs. Smith et al. (2011) have reviewed research about information privacy in various disciplines. They propose the Antecedent-Privacy Concerns-Outcomes (APCO) model to indicate the causality of these factors, in which privacy can be analyzed on individual, group, organizational and societal levels. In another study, Hong and Thong (2013) have identified a number of factors, including data collection, secondary usage, errors, improper access, control and awareness, can cause privacy concerns with web-based information systems. In addition, the Health Information Privacy Concerns (HIPC) model provides another lens for studying the privacy issues around digital health information (Kenny and Connolly, 2016). The model draws suggests that individual characteristics (e.g. gender, age, health status and healthcare need), perceptions and past experience can affect the levels of privacy concerns. Similarly, recent work shows that confidentiality, privacy, trust have an impact on the perceived usefulness and the intention of use of PHRs (Abdekhoda et al., 2019; Adelmeyer et al., 2019). While these studies provide a comprehensive coverage about information privacy in the healthcare sector, none of them investigates into the privacy

concerns of a large-scale nationwide PHR system using an opt-out model. Nevertheless, the above literature can provide directions for us to understand, analyze and interpret our data.

2.4. Social Media and Twitter Analysis

Social media have been used for various purposes related to health, e.g. sharing personal experience (Lee et al., 2014), detecting and monitoring epidemics (Khatua et al., 2019), and interacting with other people with similar conditions (Sutton et al., 2018). More than 500 million tweets are published each day, and such a magnitude of content provides an opportunity to understand consumers' needs and behaviors with a holistic view in the health context (Mejova et al., 2015). In fact, the interests of users can be inferred from their posts on social media platforms (Zarrinkalam et al., 2018). Therefore, Twitter analysis can further be used to help policymakers and healthcare providers to obtain advice and feedback on the efficacy of their services (Mejova et al., 2015). As such, our approach to investigating our research questions using social media content is well-grounded in the literature.

We recognize that Twitter users are a distinct user group, and may not effectively represent the views of, for example, older adults or the digitally disconnected. Nonetheless, there are some 4 million Twitter accounts in Australia, which measures 20% of the population over 13 years old (Australian Bureau of Statistics, 2019). The ways in which people use those accounts have been demonstrated to be politically and topically diverse (Bruns, 2017). Twitter has also been the site of considerable political engagement in Australia for over a decade (Grant et al., 2010). While we cannot expect Twitter content to represent all concerns, we can make significant inroads into the question of different user groups' privacy concerns using this diverse and politically engaged dataset.

Social media analysis may pose ethical and privacy issues. A survey finds that most Twitter users do not know their tweets are used for research purposes and they feel uncomfortable if the content is interpreted without a specific context (Fiesler and Proferes, 2018). In fact, if the data is taken out of context and reduced into mathematical models, the meaning of the data may be distorted and therefore researchers need to take a holistic view to understand the socio-technical phenomena behind it (boyd and Crawford, 2012). Even though Twitter data is publicly available, extra considerations are needed to make the analysis ethical. For instance, direct quotes should obtain informed consent from relevant users (Webb et al., 2017). As this is not always feasible, quotes and user identification should be avoided in most cases to

maintain anonymity (Ayers et al., 2018; Rivers and Lewis, 2014). In our research, we used text mining techniques to aggregate tweets with similar meanings and maintain the context of similar tweets. Additionally, our results are reported aggregated and therefore no information about individual users is exposed.

2.5. Summary of Literature Review

In Table 1, we provide a summary of the literature reviewed in the above sub-sections and highlight its implications for our research.

Table 1. Summary of Literature Review

Theme	References	Description	Implications for Our Work
PHR	(Flaumenhaft and Ben-Assuli, 2018; Tang et al., 2006)	Definitions of PHRs	
	(Abdekhoda et al., 2019; Kenny and Connolly, 2015; Li, 2015)	Usage and applications of PHRs	
	(Gagnon et al., 2016)	Stakeholders of PHRs	• Provide a basis for classifying stakeholders in the MHR case
	(Abdekhoda et al., 2019; Adelmeyer et al., 2019; Angst and Agarwal, 2009; Archer and Cocosila, 2014; Bourgeois et al., 2015)	Privacy concerns of PHR systems	• Provide a basis for categorizing the topic modeling output into different privacy concerns
MHR	(Haddad et al., 2016; Muhammad and Wickramasinghe, 2017)	Legal, privacy and security are the main concerns of patients who use MHR	• Studies had performed before the opt-out model of MHR was introduced • Understand how the opt-out model affecting privacy concerns is needed
	(Almond et al., 2016)	Potential factors causing the low rate of adoption of MHR	• Factors are not generalized for other systems
	(Kariotis et al., 2019)	Study of privacy concerns based on the view of contextual integrity	• Mainly focused on the views of patients and clinicians • Our work extends the scope to other stakeholders
Privacy	(Rahman, 2019)	Healthy people have different risk and benefit calculations than patients	

	(Smith et al., 2011)	The APCO model	• Not specifically modeled after large-scale PHRs with an opt-out model • Provide tools for us to analyze and interpret the data • Provide lens for categorizing privacy concerns
	(Kenny and Connolly, 2016)	The HIPC model	
	(Hong and Thong, 2013)	Privacy factors of web-based information systems	
Social Media	(Khatua et al., 2019; Lee et al., 2014; Sutton et al., 2018)	The use of social media analytics in research	• Justify the use of social media data in this research • Twitter data is a diverse sample with rich information that can be analyzed • Guide the design and support the research methods of our work
	(Mejova et al., 2015; Zarrinkalam et al., 2018)	The ability of using social media to infer the interests of users and to obtain advice and feedback	
	(Bruns, 2017; Grant et al., 2010)	Tweets are politically and topically diverse	
	(Ayers et al., 2018; boyd and Crawford, 2012; Fiesler and Proferes, 2018; Rivers and Lewis, 2014; Webb et al., 2017)	Research methods and ethics of analyzing social media data	

3. Methods

In this study, we intended to use Twitter posts about MHR to analyze the privacy concerns stated by different users. We adopted both computational analysis and qualitative coding as a mixed research method. Computational algorithms enable the possibility of analyzing large quantity of data such as social media posts (Pang and Liu, 2020), but the output can be further improved by qualitative analysis (Chang et al., 2009; Vakulenko et al., 2014). A recent research commentary reiterates the importance of combining both computational and manual analyses, suggesting they are complement to each other (Berente et al., 2019). In light of the aforementioned work, we believe that this approach can take the best from both approaches and produce better results.

In the following sub-sections, we explain how our data was collected and cleansed. Then, we introduce topic modeling, which is a method to summarize latent topics from textual data, and our approach to select the most appropriate topic modeling algorithm. Finally, we describe how we categorize these Twitter users into types of stakeholders and the tweets into different privacy concerns. Figure 1 shows the overview of our research design.

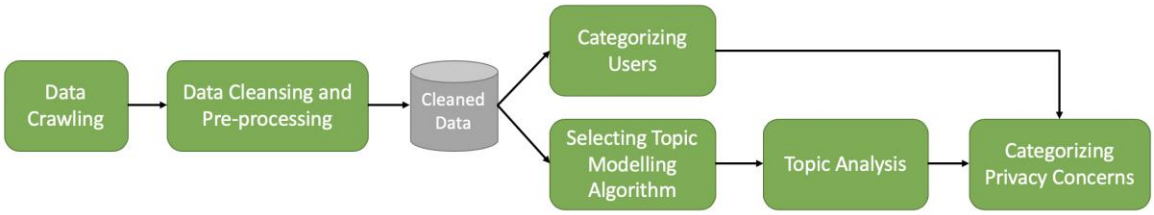

Figure 1. Overview of Research Design

3.1. Data Crawling

A Python script was used to connect to the Twitter Search API to collect tweets using 7 different keywords, including: "my health record," #myhealthrecord, #myhealthrecords, @MyHealthRec, #myhr, #MyHealthRecordFail, #mhr. These queries covered the tweets posted from the official MHR accounts, the tweets with relevant words, and the relevant hashtags. The data collection lasted for 213 days (31 weeks), which began in one week before the start of the opt-out period (8th July 2018), and ended one week after opt-out closed (9th Feb 2019). The University's human ethics committee confirmed that no ethics clearance was required for obtaining social media data from the public domain.

3.2. Data Cleansing and Pre-processing

After collecting data from the Twitter API, in line with Steinskog et al. (2017), we discarded retweets in order to understand the views of the original authors and avoid repetitive content producing noise in the topic modeling analysis. Inspired by other similar research (Bahja, 2018; Debortoli et al., 2016; Ma et al., 2016), we converted the text content of tweets to lowercase and removed English stopwords, i.e. contentless words like articles, using Python NLTK library version 3.4.4 (Bird et al., 2009). Additionally, we discarded web links (URLs), mentions of other users (i.e. words start with @) and symbols such as punctuations and emoticons. Since short sentences often do not have enough context to derive their topics, we further excluded tweets with less than 5 words from our analysis.

3.3. Categorizing Users into Stakeholders

We adopted qualitative axial coding (Creswell, 2014; Strauss and Corbin, 1998) to categorize users in our dataset into different groups of stakeholders. The classification was based on the self-provided description listed on their Twitter profiles. A researcher read the description of every user profile and assigned a code to each user for denoting their profession or expertise.

User categories were formed if there were a substantial number of users with similar codes. In this process, we referred to the definition of stakeholders used in Gagnon et al. (2016). When a category could not be found in the literature, a new category was created to allow capturing information that had not identified in prior work. Users who could not provide an understandable description (e.g. "Prisoner of Hope") or did not provide any profile description were grouped into a standalone *uncategorized* category. When a user could have more than one category identified from their profile, we took the first one in the list as their user category. For validation, another researcher independently redid the coding on a subset of these users and the results produced by both coders were compared.

3.4. Topic Modeling

Topic modeling algorithms are unsupervised machine learning methods to identify topics in a set of unstructured text (Blei, 2012), which are suitable for analysis a large amount of unstructured text information, for example, social media posts in this study. Using unsupervised algorithms researchers do not need to define topics in advance in order to use topic modeling. Topic modeling has a wide range of applications in many disciplines. Examples include categorizing themes of social media (Karami et al., 2018) and extracting concepts from electronic health records (Arnold et al., 2010). We used topic modeling for the first pass analysis in this work to identify themes from thousands of tweets. This not only allows for fast analysis but has the potential to identify themes that would not have noticed with manual inspection and less prone to human bias (Hagen, 2018).

Two topic modeling algorithms are used, namely Latent Dirichlet Allocation (LDA) and Biterm Topic Model (BTM). LDA is a probabilistic model of topic modeling commonly used in many topic modeling analyses (Blei et al., 2003), whereas BTM is another algorithm that specifically works better with short text such as social media posts (Yan et al., 2013). Recently a variant of BTM called online BTM (oBTM) has been introduced for handling a large amount of short text because the original version is slow when the data is large (Cheng et al., 2014). In this work, we tested both the LDA and oBTM algorithms and adopted the one with better performance for our study.

3.4.1. Selecting the Topic Modeling Algorithm

Topic modeling algorithms return a number of latent topics based on the distribution of words used in the corpus, i.e. the tweet dataset. Given the number of topics *K*, the algorithms can

generate a list of topics with keywords that are most relevant to individual topics. However, deciding the optimal number of topics is still an open problem. Recent papers used an experiential approach to determine the number of topics by evaluating these algorithms with different parameters (Kolini and Janczewski, 2017; Ma et al., 2016). In line with other research (Pang and Liu, 2020; Samtani et al., 2017), we started with K = 5 and tested various configurations of K in our study.

For the purpose of testing performance, we calculated the UMass coherence scores (Mimno et al., 2011) of each algorithm with two open-sourced Python machine learning implementations, "scikit-learn version 0.20.3" for LDA and "biterm version 0.1.5" for oBTM respectively. A higher value of UMass coherence score represents a topic model with better quality, and outputs are considered better and more interpretable when the scores are larger. This metric is recommended for research with qualitative components (Nikolenko et al., 2017), so is appropriate for our research.

3.5. Categorizing Privacy Concerns

For each of the categories identified in Section 3.3, we applied topic modeling to the tweets posted by the corresponding categories of stakeholders, in order to understand the topics discussed by them. Whenever there are multiple topics associated with a single tweet, we used the one with the highest probability, which is consistent with other work (Kolini and Janczewski, 2017). For deriving privacy concerns from these topics, we created a preliminary list of privacy concerns based on the literature review (Table 1). Then, we read the keywords of a topic and performed an open coding on 50 tweets from the topic, so that we could come up with the meaning of each topic. Two researchers performed the analysis and reconciled when there was a disagreement. After this process, a short description was given to briefly explain the meaning of each topic. Then, we further mapped these topics into the list of privacy concerns based on the related literature. This mapping process is similar to other topic modeling studies that investigated user satisfaction (Debortoli et al., 2016) and cybersecurity issues (Kolini and Janczewski, 2017).

4. Results

In this section, we present the results of data collection and stakeholder categorization, followed by the topic modeling results and the summary of privacy concerns.

4.1. Data Collection and Stakeholder Categorization

We downloaded 197,456 tweets including both original tweets and retweets. After discarding retweets that contained the exact content of the originals, 46,693 tweets were left. Next, we converted the text content to lowercase, removed English stopwords, and dropped other irrelevant information such as links, punctuations and emoticons. Also, we removed short tweets with fewer than 5 words because they did not provide enough information for topic modeling. Eventually 32,925 tweets from 14,081 users remained. Figure 2 illustrates the data pre-processing steps and the numbers of tweets remaining after each step.

Figure 2. Data Processing and Cleansing Pipeline

We further categorized these 14,081 users with qualitative coding techniques. The inter-coder agreement achieved 79% after a second researcher reviewed the user categories, and this reinforced the empirical validity of the categorization. Table 2 shows the list of stakeholders derived from the profiles of these users, which served as the basis for understanding the privacy concerns of different types of stakeholders in this paper.

Table 2. The Categories of Stakeholders Derived from Tweets

Category	Definition	Count	Percent
Academic	Academics, Lecturers, Professors or Researchers	696	4.9%
Clinician	Doctors, Nurses, Practitioners and Health Professionals	526	3.7%
IT	IT and Cybersecurity Professionals	1102	7.8%
Law	Lawyers and Law Groups	212	1.5%
Media	Media, News, Writers and Reporters	2099	14.9%
MHR	The official MHR account, accounts of ADHA and government departments	6	0.0%
PHN	Organizations in the Primary Health Networks – An Australian Government initiative to improve primary care access for patients	34	0.2%
Patient Group	Patients and Consumer Groups	147	1.0%
Politician	Politicians (both elected and non-elected)	73	0.5%
Privacy Advocate	Groups or individuals that advance the awareness of privacy	271	1.9%
Uncategorized	Individuals who cannot be grouped in the above categories	8915	63.3%

4.2. Determining the Topic Modeling Configuration

We adopted an experimental method to select the topic modeling algorithm with the best performance. In line with other research (Pang and Liu, 2020; Samtani et al., 2017), we started with the topic number K = 5 and measured the UMass conference scores when K = 5-20, 30, 50, 75, 100 topics with our dataset. As demonstrated in Figure 3, LDA performed consistently better than oBTM in various settings, and therefore LDA was chosen for the rest of our study. In addition, among these results, LDA with K = 5 had an average best coherence score. In this case, and this particular configuration was selected for our next phase of research. The full table of coherence scores of each topic and configuration is shown in Appendix 1.

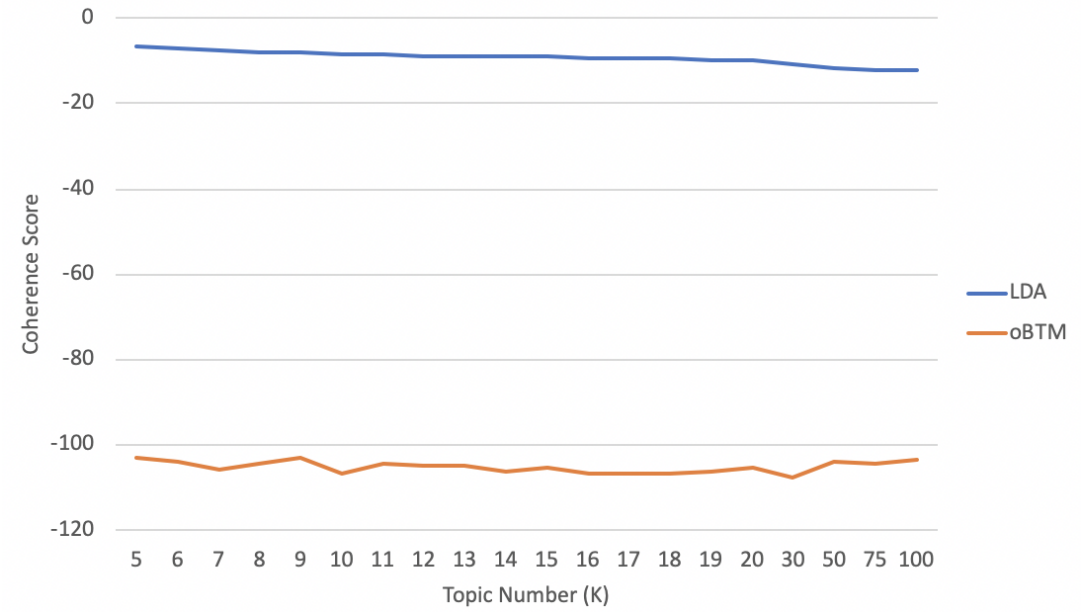

Figure 3. Average Coherence Scores of LDA and oBTM

4.3. Topic Analysis of Privacy Concerns

For every category shown in Section 4.1, we executed LDA algorithm with the configuration K = 5. As such, 5 topics were generated for each category and 55 topics in total were clustered for our dataset. Based on the keywords (listed in Appendix 2) and the sample tweets of each topic, we assigned a short description to represent the meaning of the topic, and then we mapped each topic to a privacy concern based on the lexicon of privacy concerns found in the existing literature. As listed in Table 3, we synthesized 9 types of privacy concerns after data analysis. Each privacy concern is presented with its definition, related literature and examples for illustrating the concern. In line with the research ethics of social media analysis, we have chosen not to provide original sample tweets of these topics, as this would leave the individuals

who tweeted vulnerable to identification by searching for direct quotes. Instead, we paraphrase the tweets or include only the key phrases in this table.

Table 3. Definitions of Privacy Concerns

Name	Meaning	Examples	Related Literature
Access	Who should/could have access to the data stored in MHR	• "worried that tens of thousands of people will have automatic access" • "afraid that a whole range of (government) agencies could have access to people's medical information" • "third-parties can access the data for making money"	(Angst and Agarwal, 2009; Kloss et al., 2018; Lafky and Horan, 2011; Mxoli et al., 2014)
Consent	How patients' consent should be used and whether consent was needed to create a medical record or to share information	• "data should be private unless a patient consents to release their medical records" • "my record created without consent" • "(a clinician) verbally obtained consent from a patient"	(Gagnon et al., 2016; Kloss et al., 2018)
Design	System design for addressing privacy concerns	• "the system could adopt record access codes as default instead of 'open to all'" • "the system should be free to choose to opt-in to a system"	(Flaumenhaft and Ben-Assuli, 2018; Li, 2015)
Governance	The governance by the owner of MHR (i.e. the government)	• "if (the government) has learned lessons from the UK's failure" • ADHA management issues, e.g. staff resignations	(Alyami and Song, 2016; OECD, 2015)
Legal	The legislation that could improve the privacy of the system	• "make a law to stop access without a court order" • "the law does not work because it is designed for an opt-in system"	(Bachiri et al., 2018; Flaumenhaft and Ben-Assuli, 2018; OECD, 2015)
Risk	The potential risks of using the MHR system	• "it's only a matter of time for MHR to be hacked" • Comparing the risks of MHR with the similar systems in other countries	(Li, 2015)
Security	Discussion of cybersecurity issues, particularly the potential problems of a centralized database with medical data	• "centralized personal records caused more damage than necessary when hacked" • "what are the data encryption and protection in place?"	(Adelmeyer et al., 2019; Heart et al., 2017; Mxoli et al., 2014)

Name	Meaning	Examples	Related Literature
Social	Social factors during the adoption of the system, e.g. cultural issues and the acceptance of the MHR model in communities	• "the system put vulnerable children at risk" • "the data could be used to against people with certain health or mental health conditions"	(Bourgeois et al., 2015; OECD, 2015)
Trust	The public trust in the owner (government) and the system operator (ADHA)	• Using the system means giving data to the government • Comparing the rollout of the MHR with other failed government IT projects • Outlining reasons that they do not trust the governments	(Adelmeyer et al., 2019; Heart et al., 2017)

If a topic was not related to privacy (e.g. a topic discussing the clinical benefits of MHR) and could not be assigned in one of the above types, we labelled the topic as *unrelated* and it was regarded irrelevant to the focus of this paper. 24 topics (out of 55; 44%) fell in this *unrelated* category, and these were used as a baseline to compare the ratios of privacy-related and non-privacy-related discussions.

Table 4 lists the results of the topic modeling. In this table, the topic description column includes our interpretations of the meanings of the topics after reading the keywords and sample tweets in the corresponding topics. The column of privacy concern shows our mapping of these topics to the privacy concerns listed in Table 3.

Table 4. Results of Topic Analysis Complemented by Qualitative Concept Mappings

User Category	Topic ID	No. of Tweets	% in User Category	Privacy Concern	Topic Description
Academic	A1	376	19%	Access	Discussion about who can access data
	A2	395	20%	Governance	Government needs to address privacy concerns
	A3	398	21%	Unrelated	---
	A4	307	16%	Access	Discussion about who can access data
	A5	457	24%	Design	Enhance system design to protect privacy
Clinician	C1	576	26%	Risk	Risk of using the system because of patients' privacy
	C2	348	16%	Unrelated	---
	C3	375	17%	Consent	Consent is needed to access patients' data
	C4	366	17%	Consent	General practitioners' consent of using patients' data

User Category	Topic ID	No. of Tweets	% in User Category	Privacy Concern	Topic Description
	C5	552	25%	Consent	Data access without consent and data security
IT	IT1	707	21%	Security	Online security of the system
	IT2	619	19%	Governance	Government denies privacy concerns
	IT3	498	15%	Unrelated	---
	IT4	648	20%	Unrelated	---
	IT5	849	26%	Governance	The administration and data breaches of ADHA
Law	L1	76	16%	Legal	The MHR legislation and its impact
	L2	113	24%	Unrelated	---
	L3	88	19%	Legal	Law to restrict access
	L4	86	19%	Trust	Low trust in government
	L5	101	22%	Unrelated	---
Media	M1	1,000	19%	Social	Deadlines of opt-out and public's privacy concerns
	M2	1,072	21%	Access	Discussion about who can access data
	M3	1,348	26%	Unrelated	---
	M4	739	14%	Unrelated	---
	M5	999	19%	Security	MHR issues and data security
MHR	MHR1	119	15%	Access	Information accessible by healthcare providers only
	MHR2	122	15%	Unrelated	---
	MHR3	189	23%	Unrelated	---
	MHR4	242	30%	Unrelated	---
	MHR5	145	18%	Unrelated	---
PHN	PHN1	188	19%	Unrelated	---
	PHN2	191	19%	Unrelated	---
	PHN3	218	22%	Unrelated	---
	PHN4	195	20%	Unrelated	---
	PHN5	203	20%	Unrelated	---
Patient Group	PG1	164	16%	Unrelated	---
	PG2	263	25%	Unrelated	---
	PG3	247	24%	Unrelated	---
	PG4	227	22%	Risk	Privacy risks of the system
	PG5	132	13%	Social	Consumers need to learn about privacy in MHR
Politician	PO1	44	21%	Legal	Legislation is needed to protect the security of records
	PO2	29	14%	Unrelated	---
	PO3	47	22%	Trust	Low trust in government
	PO4	48	23%	Unrelated	---

User Category	Topic ID	No. of Tweets	% in User Category	Privacy Concern	Topic Description
	PO5	42	20%	Governance	Government's track record on privacy
Privacy Advocate	PA1	408	19%	Design	Limit access to the data
	PA2	311	14%	Trust	More trust in systems owned by consumers
	PA3	707	33%	Trust	Low trust in government
	PA4	361	17%	Access	Obtain data from the system without authorization
	PA5	384	18%	Trust	Government's history of data abuse
Uncategorized	U1	1,816	12%	Social	Social perspectives of MHR privacy
	U2	4,165	29%	Unrelated	---
	U3	2,077	14%	Unrelated	---
	U4	4,400	30%	Security	Privacy concerns and data security
	U5	2,148	15%	Access	Police access without a warrant

4.4. Summary of Privacy Concerns

Figure 4 presents the percentages of each privacy concerns of all tweets in our dataset. The diagram shows that the majority of discussions (61% in total) were related to the themes of privacy, compared with the tweets which were *unrelated* to privacy concerns (39%). Apart from the *unrelated* category, the top three categories of concerns were *security* (18%), *access* (13%), and *social* (9%).

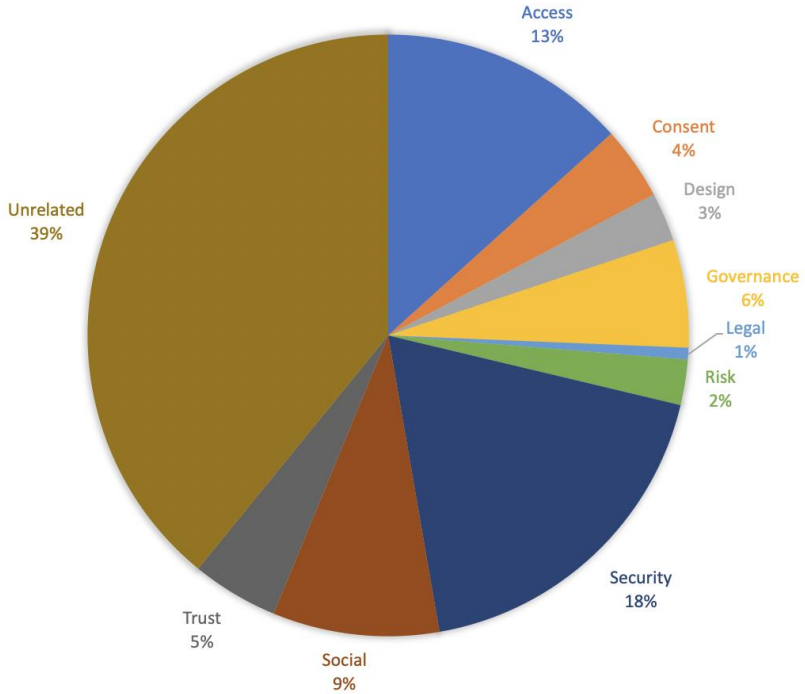

Figure 4. Privacy Concerns in Our Data

Figure 5 presents the percentages of different privacy concerns mentioned by each group of stakeholders. It is highly noticeable that each group had a different set of concerns. For instance, accounts such as *academic*, *clinician* and *IT* had the majority of their conversations related to privacy; the tweets by the *privacy advocate* group were entirely classified into privacy concerns. On the other hand, official accounts (i.e. *MHR* and *PHN*) had few discussions about privacy. For the *uncategorized* user cohort, which represents the users without specific professional backgrounds on Twitter, demonstrated a more balanced pattern: slightly more than half were related to privacy and security, and the rest of the posts were in regard to other aspects of MHR.

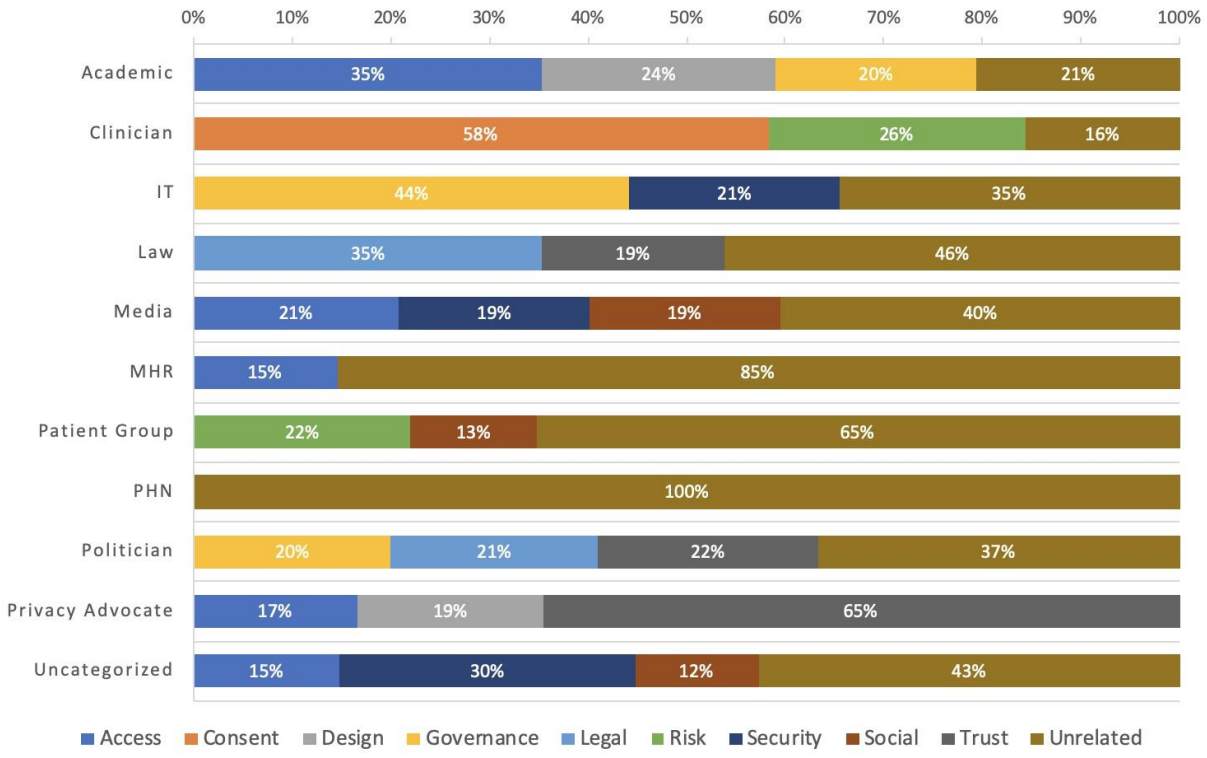

Figure 5. Ratios of Privacy Concerns among User Groups

In summary, we have been able to identify a range of different groups of stakeholders and users, and the figure shows that the interests of these groups in privacy are quite different. Some groups are not concerned with privacy, other groups have a range of privacy concerns. These concerns are likely to be related to the roles that the groups have in relation to MHR.

5. Discussion

In this section, we present the major findings to address our research questions, followed by the implications for future PHR systems and the limitations of this study.

5.1. Who Has Privacy Concerns?

As an answer to *RQ1*, we can infer that a variety of stakeholders are enthusiastic about discussing privacy concerns of MHR on social media, when the system was changed to an opt-out model. The stakeholders of the system, such as the owners (i.e. government departments), patients and patient groups are at the core of the conversation. Moreover, experts including academics, clinicians, IT professionals, legal personnel contribute to the discussion with their knowledge. This is largely consistent with the types of stakeholders listed in work by Gagnon et al. (2016), however the presence of legal professionals is new. We reckon that large-scale nationwide systems are usually governed by some legislation which affects many people; therefore, the scenario of MHR can attract legal professionals to participate in the debate. Furthermore, privacy advocate groups and activists show strong interests of MHR, and this is consistent with the deployment of other PHR systems worldwide (Lafky and Horan, 2011). The voices of these different parties will need to be considered in the design and the adoption of large PHRs. Mass media is also a part of the Twitter debate and demonstrates an impact on the dissemination of privacy topics, given their high numbers of tweets and large audience. As we do not study the effect of mass media in this study, future research can investigate how their social media accounts propagate information and facilitate the debate of privacy concerns in the context of PHRs and health information technologies.

As social media is an open discussion platform, users from different backgrounds can participate. Therefore, one can expect that the privacy topics around PHRs will attract opinions from a diversity of user cohorts on social media. As we observed that different categories of stakeholders showed different concerns, this implies the importance of calibrating the features and their expectations. After reading the data, we actually would like to pose a question: If the consultations of the MHR project had been effective when the Government started the project, why did so much discussion go on around the issues after the system was launched? This implies that many improvements can be made when consulting different stakeholders in the early stages of MHR. The Government referred to the Royle report (Royle et al., 2013) to support their decision of using the opt-out model in MHR that was similar to UK. However, the similar system in the UK was shut down due to privacy concerns and trust issues later (Boseley, 2016), and MHR failed to identify the major issues in the UK model (McCall, 2018). The similar issues around privacy concerns could be alleviated in MHR if the involvement of stakeholders could be done earlier.

As a general lesson for other PHR implementations, this work demonstrates the desire by a range of stakeholders to be involved in the discussions about privacy concerns associated with PHRs, and this is why co-designing with patients is being actively advocated for PHRs to ensure their patient-centered design meets the needs of users (Chung et al., 2017). Additionally, according to Palmer and Hemsley (2018), the rich insights collected from the social media discussion will be useful for addressing privacy and security issues for future systems, as well as engaging users for their ongoing use. Our work also demonstrates that, with the help from the computational methods that are capable to analyze a large quantity of posts, social media can be potentially used as one of the measures in the consultation phase of a project for collecting the views of different users.

5.2. What are the Privacy Concerns?

For *RQ2*, we have identified 9 types of privacy concerns (namely *access, consent, design, governance, legal, risk, security, social* and *trust*) discussed on social media regarding MHR, and as a theoretical contribution, our analysis has shown that different groups of stakeholders are concerned about different aspects. For example, medical professionals highlight concerns around obtaining patients' consents and the legal risk associated with the use of the system; whereas IT experts focus on the governance and the security aspects. Despite the large number of tweets related to privacy concerns of different parties, it is surprising that government departments and the system owners show little effort to alleviate these concerns. As noted in recent work, government-owned PHRs must strike a balance between providing the potential benefits and protecting the data and the rights of users (Flaumenhaft and Ben-Assuli, 2018). As part of engaging and interacting with stakeholders, the owners of PHRs should also consider responding to people's questions and address their concerns on social media, since this serves as an opportunity to increase their transparency (Mergel, 2013).

In addition to these differences, some stakeholder groups demonstrate a higher level of privacy concerns than the others. As shown in Figure 5, *privacy advocates* are an example of having a very high level of concern while *patient groups* show fewer concerns of privacy. Other research shows that older patients and patients with lower educational levels are more willing to use PHRs, as the potential benefits are more important to them (Nurgalieva et al., 2020). However, as large-scale PHR systems must be designed to be used by different people (Lafky and Horan, 2011), our work echoes the prior research that no users are alike (Dohan and Tan, 2013) and

the necessity of addressing different users' needs (Pang et al., 2016, 2015). The design of PHRs will therefore need to address privacy concerns separately for different user cohorts.

5.3. Practical Implications for MHR and Other PHRs

From the results we can observe that the major users of MHR, such as *clinicians* and *patient groups*, posted a large proportion of tweets regarding access and consent concerns and the risks of using the system. On the other hand, the PHNs and governmental accounts, who were responsible for promoting and supporting the rollout of the system, gave little coverage to these topics. This shows a disparity in focus between users and system owners. Many users appeared not to fully understand how the system worked, which is understandable as the privacy and security aspects of the system go beyond their knowledge. For the rollout of MHR, PHNs play a vital role to help frontline health practitioners to adopt the system and they should respond their concerns both on social media and offline. As an implication for other PHR systems, the awareness of privacy concerns is increasing among the users, therefore their owners need to take up the responsibility to communicate with them, clarify their concerns and assist with minimizing the risks. System owners can take further steps and use their online presence for conveying the caveats related to privacy and information security, in addition to pushing the adoption and highlighting the benefits as we have seen in the case of MHR.

Meanwhile, MHR creates an ethical dilemma in the Australian digital health space (Duckett, 2019) and demonstrates that the consent model is one of the most important decisions to make in implementing future PHRs. With an opt-out model, it is arguable that explicit consent is obtained even a user has not opted-out. Furthermore, there are even more implications for PHRs after users have been enrolled in the system. For instance, should the user's consent to create an account in the system be extended to all other operations, such as data uploads and clinician access? In the case of MHR, some settings were available to users for fine-tuning who could see the information, however the default permissions were applied and until users intervened (Office of the Australian Information Commissioner, n.d.). As reported in the tweets and also by McCall (2018), not everyone understands these settings and is aware to configure them. As a result, people may not notice that new data has been uploaded and are not aware the possibility of information leaks, and thus it this not a fully informed consent to the use of their data. Although implicit consent can be convenient for clinicians and healthcare providers to access information, as well as sharing information among treating clinicians and care providers (Kariotis et al., 2019), measures need to be employed to balance different privacy

needs and preferences. A decision needs to be made about how far the initial consent can be extended in the lifetime of the system. With the rapid development of mobile technologies, future PHRs can be linked to the mobile phones of individual users, so that they can consent or not to the actions in the system. Such approaches can increase the transparency of the systems and give control back to users.

Our study also provides insights into an area that is less discussed in the literature, that of third-party access. Under some cases, third-party access can be legitimate, for example, the police may access the records for investigations. However, people may have concerns because it may put some community members in disadvantage, which can also be seen in the experience of adopting PHRs in other countries (OECD, 2015). This prompted the Australian Government to strengthen the privacy laws regarding MHR (Australian Digital Health Agency, 2019b; McCall, 2018). Another better approach from the user perspective is to incorporate protection in the consent model. For future PHRs, the consent of secondary use and third-party access needs be well considered and defined, for instance, whether explicit consent from a user is required for the access from additional parties, for example, pharmacists and researchers. Dynamic Consent (Kaye et al., 2015; Prictor et al., 2019), which is an interactive and personalized consent that allows people to participate at different levels and alter the consent anytime, can shed light on improving the robustness of the consent model of PHRs.

5.4. Limitations

Several limitations exist in our work. Our approach of downloading tweets about MHR would not capture tweets where users did not explicitly use the keywords we searched on, or if they misspelled those words. Secondly, a sample of tweets does not represent all social media discussions and the views of the entire user population. In addition, we cannot guarantee that the users in our dataset were entirely from Australia. We also acknowledge that people with privacy concerns may not express their views on social media because they may worry about online privacy. For the categorization of stakeholders, we relied on the profiles of users but some of them did not provide their information or listed multiple occupations, and as a result some of the users might not be allocated to the correct group. Finally, in the topic modeling analysis, we only chose the topic with the highest probability for each tweet. For the tweets which could be classified into multiple topics, only the most predominant one was used.

6. Conclusion

We present an analysis of the Twitter debate about the rollout of the Australian national-wide PHR system and our contributions in this paper can be concluded in three points. Our first contribution comes through the stakeholder categorization and the topic analysis, resulting in a deeper understanding of privacy concerns of each type of stakeholders. This is helpful for the owners of PHRs to engage with stakeholders in the early stage of implementations and set up the expectation of communication. Secondly, in an open social media platform, a variety of other stakeholders such as experts, privacy advocates, patient groups can also demonstrate their concerns, which demonstrates the feasibility of using social media for collecting views and opinions in this context. In this research, we have shown that each of these user cohorts uses different lenses to view privacy. In this case, acknowledging their differences when designing and deploying large-scale PHRs is crucial. Our third contribution includes practical implications for future provisions of PHRs. Communicating with stakeholders on social media and accommodating the needs of various users in the design of the system are crucial. Additionally, we argue that a refined consent model and fine-grained control of authorizing third-party access can help to alleviate privacy concerns. Although the controversial episode of MHR enrolment has ended, the knowledge gleaned from this case remains helpful for developing user-centered PHR solutions and benefiting the digital health sector.

7. References

Abd-alrazaq, A.A., Bewick, B.M., Farragher, T., Gardner, P., 2019. Factors that affect the use of electronic personal health records among patients: A systematic review. Int. J. Med. Inf. 126, 164–175. https://doi.org/10.1016/j.ijmedinf.2019.03.014

Abdekhoda, M., Dehnad, A., Khezri, H., 2019. The effect of confidentiality and privacy concerns on adoption of personal health record from patient's perspective. Health Technol. 9, 463–469. https://doi.org/10.1007/s12553-018-00287-z

Adelmeyer, M., Meier, P., Teuteberg, F., 2019. Security and Privacy of Personal Health Records in Cloud Computing Environments – An Experimental Exploration of the Impact of Storage Solutions and Data Breaches, in: Wirtschaftsinformatik 2019.

Almond, H., Cummings, E., Turner, P., 2016. Avoiding failure for Australia's digital health record: the findings from a rural e-health participatory research project. Stud. Health Technol. Inform. 227, 8–13.

Alyami, M.A., Song, Y.-T., 2016. Removing barriers in using personal health record systems, in: 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). Presented at the 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), pp. 1–8. https://doi.org/10.1109/ICIS.2016.7550810

Angst, C.M., Agarwal, R., 2009. Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. MIS Q. 33, 339–370.
Archer, N., Cocosila, M., 2014. Canadian Patient Perceptions of Electronic Personal Health Records: An Empirical Investigation. Commun. Assoc. Inf. Syst. 34, 20. https://doi.org/10.17705/1CAIS.03420
Arnold, C.W., El-Saden, S.M., Bui, A.A.T., Taira, R., 2010. Clinical case-based retrieval using latent topic analysis, in: AMIA Annual Symposium Proceedings. American Medical Informatics Association, pp. 26–26.
Australian Bureau of Statistics, 2019. Australian Demographic Statistics [WWW Document]. URL https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Jun%202019?OpenDocument (accessed 1.19.20).
Australian Digital Health Agency, 2019a. Types of Digital Health Records [WWW Document]. URL https://www.digitalhealth.gov.au/get-started-with-digital-health/digital-health-evidence-review/types-of-digital-health-records#1
Australian Digital Health Agency, 2019b. Stronger My Health Record Privacy Laws [WWW Document]. URL https://www.myhealthrecord.gov.au/about/legislation-and-governance/summary-privacy-protections
Ayers, J.W., Caputi, T.L., Nebeker, C., Dredze, M., 2018. Don't quote me: reverse identification of research participants in social media studies. Npj Digit. Med. 1, 30. https://doi.org/10.1038/s41746-018-0036-2
Bachiri, M., Idri, A., Fernandez-Aleman, J.L., Toval, A., 2018. Evaluating the Privacy Policies of Mobile Personal Health Records for Pregnancy Monitoring. J. Med. Syst. 1. https://doi.org/10.1007/s10916-018-1002-x
Bahja, M., 2018. Identifying Patient Experience from Online Resources via Sentiment Analysis and Topic Modelling Approaches, in: Proceedings of International Conference on Information Systems (ICIS 2018). pp. 1–9.
Berente, N., Seidel, S., Safadi, H., 2019. Data-Driven Computationally Intensive Theory Development. Inf. Syst. Res. 30, 50–64. https://doi.org/10.1287/isre.2018.0774
Berge, G.T., 2016. Drivers and Barriers to Structuring Information in Electronic Health Records, in: Proceeding of the 20th Pacific Asia Conference on Information Systems (PACIS 2016). Chiayi, Taiwan, pp. 18–18.
Bird, S., Klein, E., Loper, E., 2009. Natural Language Processing with Python. O'Reilly Media Inc.
Blei, D.M., 2012. Probabilistic topic models. Commun. ACM 55, 77–84. https://doi.org/10.1145/2133806.2133826
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.
Boseley, S., 2016. NHS to scrap single database of patients' medical details. The Guardian.
Bourgeois, F.C., Nigrin, D.J., Harper, M.B., 2015. Preserving Patient Privacy and Confidentiality in the Era of Personal Health Records. Pediatrics 135, e1125. https://doi.org/10.1542/peds.2014-3754
boyd, danah, Crawford, K., 2012. Critical Questions for Big Data. Inf. Commun. Soc. 15, 662–679. https://doi.org/10.1080/1369118X.2012.678878
Bruns, A., 2017. Australian Twitter is more diverse than you think [WWW Document]. The Conversation. URL https://theconversation.com/australian-twitter-is-more-diverse-than-you-think-76864 (accessed 1.19.20).

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J.L., Blei, D.M., 2009. Reading tea leaves: How humans interpret topic models. Presented at the Advances in neural information processing systems, pp. 288–296.
Cheng, X., Yan, X., Lan, Y., Guo, J., 2014. BTM: Topic Modeling over Short Texts. IEEE Trans. Knowl. Data Eng. 26, 2928–2941. https://doi.org/10.1109/TKDE.2014.2313872
Chung, A., Chen, H., Shin, G., Mane, K., Kum, H.-C., 2017. The design of a patient-centered personal health record with patients as co-designers. J. Clin. Transl. Sci. 1, 18–18. https://doi.org/10.1017/cts.2017.77
Creswell, J.W., 2014. Research design: qualitative, quantitative, and mixed method approaches., Fourth edition, international student edition. ed. SAGE Publications, Inc., Thousand Oaks, California.
Debortoli, S., Müller, O., Junglas, I., vom Brocke, J., 2016. Text Mining for Information Systems Researchers: An Annotated Topic Modeling Tutorial. Commun. Assoc. Inf. Syst. 39, 110–135. https://doi.org/10.17705/1CAIS.03907
Department of Health, 2017. Evaluation of the My Health Record Participation Trials.
Dohan, M.S., Tan, J., 2013. Perceived Usefulness and Behavioral Intention to Use Consumer-oriented Web-based Health Tools: A Meta-analysis, in: Proceedings of the 19th Americas Conference on Information Systems (AMCIS). Chicago, Illinois.
Dray-Spira, R., Gueguen, A., Lert, F., 2008. Disease severity, self-reported experience of workplace discrimination and employment loss during the course of chronic HIV disease: differences according to gender and education. Occup. Environ. Med. 65, 112–119. https://doi.org/10.1136/oem.2007.034363
Earnshaw, V.A., Quinn, D.M., 2012. The impact of stigma in healthcare on people living with chronic illnesses. J. Health Psychol. 17, 157–168.
Fiesler, C., Proferes, N., 2018. "Participant" Perceptions of Twitter Research Ethics. Soc. Media Soc. 4, 205630511876336. https://doi.org/10.1177/2056305118763366
Flaumenhaft, Y., Ben-Assuli, O., 2018. Personal health records, global policy and regulation review. Health Policy 122, 815–826. https://doi.org/10.1016/j.healthpol.2018.05.002
Gagnon, M.-P., Payne-Gagnon, J., Breton, E., Fortin, J.-P., Khoury, L., Dolovich, L., Price, D., Wiljer, D., Bartlett, G., Archer, N., 2016. Adoption of Electronic Personal Health Records in Canada: Perceptions of Stakeholders. Int. J. Health Policy Manag. 5, 425–433. https://doi.org/10.15171/ijhpm.2016.36
Grant, W.J., Moon, B., Busby Grant, J., 2010. Digital Dialogue? Australian Politicians' use of the Social Network Tool Twitter. Aust. J. Polit. Sci. 45, 579–604. https://doi.org/10.1080/10361146.2010.517176
Haddad, P., Muhammad, I., Wickramasinghe, N., 2016. Assessing the Business Value of Australia's National e-health Solution, in: Proceedings of the 22nd Americas Conference on Information Systems (AMCIS). San Diego, CA, pp. 1–10.
Hagen, L., 2018. Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? Inf. Process. Manag. 54, 1292–1307. https://doi.org/10.1016/j.ipm.2018.05.006
Heart, T., Ben-Assuli, O., Shabtai, I., 2017. A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. Health Policy Technol. 6, 20–25. https://doi.org/10.1016/j.hlpt.2016.08.002
Hong, W., Thong, J.Y., 2013. Internet privacy concerns: An integrated conceptualization and four empirical studies. Mis Q. 275–298.
Jung, K.E., 2003. Chronic Illness and Academic Accommodation: Meeting Disables Students' Unique Needs and Preserving the Institutional Order of the University. J Soc Soc Welf. 30, 91.

Jung, K.E., 2002. Chronic illness and educational equity: The politics of visibility. NWSA J. 178–200.
Karami, A., Dahl, A.A., Turner-McGrievy, G., Kharrazi, H., Shaw, G., 2018. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. Int. J. Inf. Manag. 38, 1–6. https://doi.org/10.1016/j.ijinfomgt.2017.08.002
Kariotis, T., Prictor, M., Chang, S., Gray, K., 2019. Evaluating the Contextual Integrity of Australia's My Health Record. Stud. Health Technol. Inform. 265, 213–218. https://doi.org/10.3233/SHTI190166
Kaye, J., Whitley, E.A., Lund, D., Morrison, M., Teare, H., Melham, K., 2015. Dynamic consent: a patient interface for twenty-first century research networks. Eur. J. Hum. Genet. 23, 141.
Kenny, G., Connolly, R., 2016. Drivers of Health Information Privacy Concern: A Comparison Study, in: Proceedings of the 22nd Americas Conference on Information Systems (AMCIS). San Diego, CA, pp. 1–10.
Kenny, G., Connolly, R., 2015. Citizens' Health Information Privacy Concerns: A Multifaceted Approach., in: Proceedings of the Twenty-Third European Conference on Information Systems (ECIS 2015).
Khatua, Aparup, Khatua, Apalak, Cambria, E., 2019. A tale of two epidemics: Contextual Word2Vec for classifying twitter streams during outbreaks. Inf. Process. Manag. 56, 247–257. https://doi.org/10.1016/j.ipm.2018.10.010
King, M.B., 1989. Prejudice and AIDS: the views and experiences of people with HIV infection. AIDS Care 1, 137–143.
Kloss, L.L., Brodnik, M.S., Rinehart-Thompson, L.A., 2018. Access and Disclosure of Personal Health Information: A Challenging Privacy Landscape in 2016-2018. Yearb Med Inf. 27, 060–066. https://doi.org/10.1055/s-0038-1667071
Kolini, F., Janczewski, L., 2017. Clustering and Topic Modelling: A New Approach for Analysis of National Cyber Security Strategies, in: PACIS 2017 Proceedings. pp. 126–126.
Lafky, D.B., Horan, T.A., 2011. Personal health records: Consumer attitudes toward privacy and security of their personal health information. Health Informatics J. 17, 63–71. https://doi.org/10.1177/1460458211399403
Lee, J.L., DeCamp, M., Dredze, M., Chisolm, M.S., Berger, Z.D., 2014. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. J. Med. Internet Res. 16, e237.
Li, J., 2015. Ensuring Privacy in a Personal Health Record System. Computer 48, 24–31. https://doi.org/10.1109/MC.2015.43
Ma, B., Yuan, H., Wan, Y., Qian, Y., Zhang, N., 2016. Public opinion analysis based on probabilistic topic modeling and deep learning, in: Proceedings of Pacific Asia Conference on Information Systems 2016 (PACIS 2016).
McCall, C., 2018. Opt-out digital health records cause debate in Australia. The Lancet 392, 372. https://doi.org/10.1016/S0140-6736(18)31726-4
Mejova, Y., Weber, I., Macy, M.W., 2015. Twitter: a digital socioscope. Cambridge University Press.
Mergel, I., 2013. A framework for interpreting social media interactions in the public sector. Gov. Inf. Q. 30, 327–334. https://doi.org/10.1016/j.giq.2013.05.015
Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A., 2011. Optimizing semantic coherence in topic models. Presented at the Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics, pp. 262–272.

Muhammad, I., Wickramasinghe, N., 2017. User Perceptions and Expectations of the MyHealth Record: A Case Study of Australia's e-health Solution, in: Proceedings of the 50th Hawaii International Conference on System Sciences. pp. 3441–3450. https://doi.org/10.24251/HICSS.2017.416

Mxoli, A., Gerber, M., Mostert-Phipps, N., 2014. Information security risk measures for Cloud-based personal health records, in: International Conference on Information Society (i-Society 2014). Presented at the International Conference on Information Society (i-Society 2014), pp. 187–193. https://doi.org/10.1109/i-Society.2014.7009039

Nikolenko, S.I., Koltcov, S., Koltsova, O., 2017. Topic modelling for qualitative studies. J. Inf. Sci. 43, 88–102. https://doi.org/10.1177/0165551515617393

Nurgalieva, L., Cajander, Å., Moll, J., Åhlfeldt, R.-M., Huvila, I., Marchese, M., 2020. 'I do not share it with others. No, it's for me, it's my care': On sharing of patient accessible electronic health records. Health Informatics J. 1460458220912559. https://doi.org/10.1177/1460458220912559

OECD, 2015. Health Data Governance: Privacy, Monitoring and Research, OECD Health Policy Studies. OECD Publishing, Paris. https://doi.org/10.1787/9789264244566-en

Office of the Australian Information Commissioner, 2019. Australian Privacy Principles [WWW Document]. URL https://www.oaic.gov.au/privacy/australian-privacy-principles (accessed 8.29.19).

Office of the Australian Information Commissioner, n.d. Tips to protect your My Health Record [WWW Document]. URL https://www.oaic.gov.au/privacy/health-information/my-health-record/tips-to-protect-your-my-health-record/ (accessed 5.16.20).

Otlowski, M., Tiller, J., Barlow-Stewart, K., Lacaze, P., 2019. Genetic testing and insurance in Australia. Aust. J. Gen. Pract. 48, 96–99.

Palmer, S., Hemsley, B., 2018. Analysis of three Twitter hashtags for discussion of personal electronic health records. Presented at the ECSM 2018: Proceedings of the 5th European Conference on Social Media, Academic Conferences and Publishing International Limited, pp. 236–245.

Pang, P.C.-I., Chang, S., 2019. The Twitter Adventure of #MyHealthRecord: An Analysis of Different User Groups During the Opt-Out Period. Stud. Health Technol. Inform. 266, 142–148. https://doi.org/10.3233/SHTI190786

Pang, P.C.-I., Chang, S., Verspoor, K., Pearce, J., 2016. Designing Health Websites Based on Users' Online Information Seeking Behaviours: A Mixed-method Observational Study. J. Med. Internet Res. 18, e145. https://doi.org/10.2196/jmir.5661

Pang, P.C.-I., Liu, L., 2020. Why Do Consumers Review Doctors Online? Topic Modeling Analysis of Positive and Negative Reviews on an Online Health Community in China, in: Proceedings of the 53rd Hawaii International Conference on System Sciences. pp. 705–714.

Pang, P.C.-I., Verspoor, K., Pearce, J., Chang, S., 2015. Better Health Explorer: Designing for Health Information Seekers, in: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI'15). ACM, Melbourne, Australia, pp. 588–597. https://doi.org/10.1145/2838739.2838772

Parliament of Australia, 2015. Health Legislation Amendment (eHealth) Bill 2015 [Provisions].

Prictor, M., Lewis, M.A., Newson, A.J., Haas, M., Baba, S., Kim, H., Kokado, M., Minari, J., Molnár-Gábor, F., Yamamoto, B., Kaye, J., Teare, H.J.A., 2019. Dynamic Consent: An Evaluation and Reporting Framework. J. Empir. Res. Hum. Res. Ethics 1556264619887073. https://doi.org/10.1177/1556264619887073

Rahman, M.S., 2019. Does Privacy Matters When We are Sick? An Extended Privacy Calculus Model for Healthcare Technology Adoption Behavior, in: 2019 10th International Conference on Information and Communication Systems (ICICS). Presented at the 2019 10th International Conference on Information and Communication Systems (ICICS), pp. 41–46. https://doi.org/10.1109/IACS.2019.8809175

Rivers, C.M., Lewis, B.L., 2014. Ethical research standards in a world of big data. F1000Research 3. https://doi.org/10.12688/f1000research.3-38.v2

Roehrs, A., da Costa, C.A., Righi, R. da R., de Oliveira, K.S.F., 2017. Personal Health Records: A Systematic Literature Review. J Med Internet Res 19, e13. https://doi.org/10.2196/jmir.5876

Roessler, R., Hennessey, M., Neath, J., Rumrill, P., Nissen, S., 2011. The employment discrimination experiences of adults with multiple sclerosis. J. Rehabil. 77, 20.

Royle, R., Hambleton, S., Walduck, A., 2013. Review of the Personally Controlled Electronic Health Record.

Samtani, S., Chinn, R., Chen, H., Nunamaker, J.F., 2017. Exploring Emerging Hacker Assets and Key Hackers for Proactive Cyber Threat Intelligence. J. Manag. Inf. Syst. 34, 1023–1053. https://doi.org/10.1080/07421222.2017.1394049

Sharac, J., Mccrone, P., Clement, S., Thornicroft, G., 2010. The economic impact of mental health stigma and discrimination: a systematic review. Epidemiol. Psychiatr. Sci. 19, 223–232.

Smith, H.J., Dinev, T., Xu, H., 2011. Information Privacy Research: An Interdisciplinary Review. MIS Q. 35, 989–1015.

Steinskog, A., Therkelsen, J., Gambäck, B., 2017. Twitter topic modeling by tweet aggregation. Presented at the Proceedings of the 21st Nordic Conference on Computational Linguistics, pp. 77–86.

Strauss, A., Corbin, J., 1998. Basics of qualitative research techniques. Sage Publications, Thousand Oaks, CA.

Sutton, J., Vos, S.C., Olson, M.K., Woods, C., Cohen, E., Gibson, C.B., Phillips, N.E., Studts, J.L., Eberth, J.M., Butts, C.T., 2018. Lung Cancer Messages on Twitter: Content Analysis and Evaluation. J. Am. Coll. Radiol. 15, 210–217. https://doi.org/10.1016/j.jacr.2017.09.043

Tang, P.C., Ash, J.S., Bates, D.W., Overhage, J.M., Sands, D.Z., 2006. Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. J. Am. Med. Inform. Assoc. 13, 121–126. https://doi.org/10.1197/jamia.M2025

Vakulenko, S., Müller, O., vom Brocke, J., 2014. Enriching iTunes App Store Categories via Topic Modeling, in: Proceedings of 35th International Conference on Information Systems (ICIS 2014). Auckland, New Zealand, pp. 1–11.

Webb, H., Jirotka, M., Stahl, B.C., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O., Burnap, P., 2017. The Ethical Challenges of Publishing Twitter Data for Research Dissemination, in: Proceedings of the 2017 ACM on Web Science Conference, WebSci '17. ACM, New York, NY, USA, pp. 339–348. https://doi.org/10.1145/3091478.3091489

World Health Organization, 2016. Atlas of eHealth country profiles. WHO Document Production Services, Geneva, Switzerland.

Yan, X., Guo, J., Lan, Y., Cheng, X., 2013. A biterm topic model for short texts, in: Proceedings of the 22nd International Conference on World Wide Web (WWW '13). ACM Press, Rio de Janeiro, Brazil, pp. 1445–1456. https://doi.org/10.1145/2488388.2488514

Zarrinkalam, F., Kahani, M., Bagheri, E., 2018. Mining user interests over active topics on social networks. Inf. Process. Manag. 54, 339–357. https://doi.org/10.1016/j.ipm.2017.12.003

8. Appendix 1 – Matrix of Topic Coherence Scores

LDA

K	Academic	Clinician	IT	Law	Media	MHR	Patient Group	PHN	Politician	Privacy Advocate	Uncategorized
5	-4.68	-6.31	-6.32	-11.17	-6.19	-5.78	-6.68	-3.76	-11.28	-6.48	-4.73
6	-4.86	-7.27	-5.68	-11.12	-6.38	-5.71	-7.34	-3.90	-12.26	-7.03	-5.02
7	-4.80	-7.65	-6.50	-11.18	-6.29	-6.67	-7.17	-4.32	-11.80	-7.85	-6.32
8	-5.76	-9.01	-6.64	-11.79	-7.28	-6.67	-7.64	-4.31	-12.11	-8.45	-6.57
9	-5.28	-9.71	-7.80	-11.30	-7.92	-7.03	-7.58	-4.81	-12.80	-8.92	-6.31
10	-5.33	-10.80	-7.26	-11.97	-8.24	-6.91	-7.39	-4.51	-13.69	-10.18	-6.64
11	-5.45	-10.74	-7.71	-12.38	-8.24	-6.94	-7.46	-4.80	-13.36	-9.16	-7.10
12	-6.11	-11.01	-7.98	-12.00	-8.52	-6.88	-7.56	-4.99	-13.47	-10.25	-7.11
13	-5.39	-11.57	-8.19	-12.17	-8.99	-7.83	-7.90	-4.85	-13.28	-10.58	-8.01
14	-5.92	-11.73	-8.47	-13.00	-8.81	-7.22	-7.66	-5.02	-14.13	-10.26	-6.98
15	-5.96	-11.68	-8.07	-12.70	-8.53	-7.36	-8.02	-5.26	-13.85	-11.37	-7.12
16	-6.13	-12.19	-8.74	-12.83	-9.38	-7.37	-8.18	-5.20	-14.18	-11.18	-7.51
17	-5.75	-12.50	-8.93	-13.18	-9.86	-8.07	-8.48	-5.56	-14.28	-11.02	-6.92
18	-6.11	-12.06	-9.09	-12.83	-9.74	-7.52	-8.04	-5.75	-14.23	-12.18	-7.79
19	-6.48	-12.29	-9.37	-12.98	-10.02	-8.14	-8.34	-5.23	-14.62	-11.75	-8.64
20	-6.10	-12.38	-9.63	-13.26	-9.86	-7.77	-8.85	-5.52	-14.04	-13.10	-8.36
30	-7.02	-13.81	-9.80	-14.24	-11.41	-8.35	-9.41	-6.32	-14.23	-13.37	-9.87
50	-8.48	-15.39	-11.89	-13.97	-12.54	-9.57	-10.77	-7.71	-12.91	-13.88	-11.56
75	-10.05	-15.74	-12.90	-14.53	-13.38	-10.26	-11.48	-8.80	-11.07	-14.15	-12.35
100	-10.88	-14.97	-13.35	-12.74	-13.32	-10.69	-12.04	-9.91	-10.39	-13.22	-11.90

oBTM

K	Academic	Clinician	IT	Law	Media	MHR	Patient Group	PHN	Politician	Privacy Advocate	Uncategorized
5	-118.32	-128.26	-108.89	-69.54	-110.18	-84.94	-127.78	-92.91	-109.35	-77.30	-106.23
6	-111.66	-125.62	-120.72	-73.16	-113.37	-96.80	-121.61	-93.07	-102.10	-74.67	-111.70
7	-123.66	-126.79	-118.37	-69.15	-111.48	-95.03	-126.98	-94.70	-111.84	-69.79	-116.35
8	-126.18	-123.41	-115.76	-80.14	-114.00	-88.27	-129.40	-101.51	-100.64	-55.54	-114.45
9	-121.73	-120.43	-121.97	-70.17	-121.66	-79.87	-120.14	-99.46	-97.74	-69.13	-112.19
10	-120.44	-121.04	-113.74	-78.86	-113.47	-90.98	-106.83	-95.38	-72.96	-115.23	-123.54
11	-114.29	-121.91	-117.10	-78.02	-115.04	-93.42	-132.82	-96.56	-101.61	-62.33	-114.97
12	-117.45	-115.97	-114.39	-83.42	-117.99	-88.08	-123.18	-99.10	-107.70	-71.37	-116.99
13	-123.47	-124.23	-119.32	-76.58	-128.83	-82.93	-121.32	-96.41	-99.54	-59.65	-119.19
14	-121.42	-125.05	-123.64	-75.66	-125.37	-85.96	-128.04	-99.78	-95.18	-71.34	-118.41
15	-121.75	-127.65	-123.18	-77.49	-129.68	-87.60	-102.49	-103.04	-64.93	-118.54	-125.66

16	-121.33	-128.59	-126.95	-76.52	-130.40	-85.20	-126.87	-99.57	-91.70	-64.71	-122.07
17	-125.49	-119.93	-125.70	-73.69	-127.42	-84.67	-129.08	-102.42	-100.79	-63.93	-123.17
18	-120.25	-123.38	-127.78	-78.24	-135.20	-83.61	-128.47	-100.47	-95.00	-59.77	-123.36
19	-116.79	-124.47	-123.78	-72.28	-133.73	-80.81	-126.34	-104.72	-100.49	-62.65	-122.60
20	-120.76	-118.09	-121.78	-75.70	-130.74	-88.00	-99.91	-97.41	-60.29	-122.89	-124.70
30	-126.87	-124.40	-125.33	-72.25	-135.02	-82.74	-98.25	-103.06	-54.99	-125.62	-133.57
50	-121.97	-127.83	-129.20	-60.80	-133.08	-78.31	-90.85	-100.85	-41.81	-123.64	-135.80
75	-122.33	-126.97	-129.19	-50.43	-138.80	-83.50	-90.76	-97.03	-40.75	-127.44	-140.41
100	-122.16	-127.31	-128.91	-45.26	-139.56	-83.28	-89.61	-99.53	-31.11	-123.75	-145.99

9. Appendix 2 – Full Output of LDA Topic Modeling

User Category	**Topic ID**	**Top 5 Topic Keywords**
Academic	A1	data access people privacy government
	A2	data government think privacy concerns
	A3	information medical system use patient
	A4	data privacy records access issues
	A5	data privacy access system concerns
Clinician	C1	patients privacy information system care
	C2	information medical privacy interesting debate
	C3	access patient data need consent
	C4	information patients data consent gp
	C5	data access without consent security
IT	IT1	security data access system online
	IT2	data privacy government australian concerns
	IT3	people process want know work
	IT4	system privacy extended period time
	IT5	data access privacy adha government
Law	L1	privacy access law records people
	L2	law know free give offer
	L3	law privacy access safety time
	L4	data system australian government information
	L5	find packages calling busy law
Media	M1	privacy government concerns last today
	M2	australians privacy access advice people
	M3	system senate controversial government minster
	M4	government system new privately access
	M5	data government privacy issues security
MHR	MHR1	information providers access healthcare keep
	MHR2	healthcare access providers information help
	MHR3	help information line call cancel

	MHR4	access healthcare help online cancel
	MHR5	website information details help registration
PHN	PHN1	team week help questions people
	PHN2	learn know benefits team help
	PHN3	information team find community questions
	PHN4	today healthcare community learn visit
	PHN5	information learn help see care
Patient Group	PG1	people access webinar questions aboriginal
	PG2	information system learn benefits risks
	PG3	information people healthcare make decision
	PG4	privacy information data system risks
	PG5	consumers need information know privacy
Politician	PO1	data govt legislation records security
	PO2	government listening community use extend
	PO3	govt data trust public labors
	PO4	privacy system data security senate
	PO5	government time australians privacy records
Privacy Advocate	PA1	privacy access people data get
	PA2	consumer system data trust better
	PA3	trust gov't privacy system know
	PA4	get data system patients without
	PA5	medical gov't data people history
Uncategorized	U1	social family warned care privacy
	U2	time records system data people
	U3	uk scheme times failed system
	U4	data privacy records system concerns
	U5	access police government system without

Patrick Pang: Methodology, Software, Formal analysis, Writing – Original Draft; **Dana McKay**: Formal analysis, Validation, Writing – Reviewing & Editing; **Shanton Chang**: Conceptualization, Formal analysis, Writing – Reviewing & Editing; **Qingyu Chen**: Methodology, Software; **Xiuzhen Zhang**: Methodology, Resources; **Lishan Cui**: Software