Personal Data Management Services to Empower Individuals

Sujata Jindal*
Computer Science Engineering Department,
SGT Institute of Engineering and Technology,
Gurgaon, India

Dr. Ritu Sindhu
Associate Professor, CSE Department,
SGT Institute of Engineering and Technology,
Gurgaon, India

***Abstract* Our objective is to leverage the personal big data to improve decisions and empower people. Till now big companies are making money from individuals' data. Individuals are creating contents on Facebook, Google, YouTube, Twitter etc. The more content individuals are creating, the more valuable these are becoming. But that data effectively belongs to respective companies. Individuals are creating data, but the profits are earned by these big organizations. Individuals are sowing and they are reaping. There is need to tools and services to help individuals to have fine grained control over their data so that they can share their data to third parties under their control. Apart from sharing individuals should be able to make better decisions based on their data. In this paper our work is twofold. First we have introduced a new personal data management framework with architecture and second we have added privacy feature to that framework to protect the privacy of personal data at an individual level.**

Keywords- Big Data, Personal Big Data, Privacy, Personal Data Management

I. INTRODUCTION

Personal data is the new class of data that is emerging due to this digital world [1]. It is any data related directly or indirectly to an individual. It is in the form of thousands of emails, phone logs, location information, your searches, preferences while shopping, transactions data, medical history etc. You are leaving footprints while browsing the digital world.

Till now it has been used by big organizations to provide you with various services to ease your life. Various apps use these data to provide us with smart services and personalized experiences. These help users to become more connected, productive and entertained. You are giving control of your data knowingly or unknowingly in return of such services. These types of services are required to ease our life. We need to share our data in order to run this digital world. That is the need of the hour. But the question here is who has control of all that data.

Personal data has great potential but still has not gained its full potential. It is fragmented as it is collected and gathered by various different services/applications and companies. Due to this reason even personal data is inaccessible to individuals who have created it directly or indirectly. Individuals don't have full control over personal data. Due to privacy concerns and the associated risk of re-identification [2-4], it is not broadly accessible.

In this paper we introduce a personal data management framework which will allow users to collect, store and use/analyze their data and share it with third parties under their control. Along with the integrated personal data management, privacy/security concern has also been taken into consideration so that data can be shared safely under the control of individual.

II. PERSONAL DATA MANAGEMENT

Every individual want easy access to data, information and services. Personal data Management help individuals gather, store, analyze and use their own data. PDM consists of apps and tools that save time provide insight and understanding. It helps individuals to keep them safe and to get basic tasks done easily. Individuals will get an integrated environment to manage their personal data from one point and to share it under their control.

Personal data management services are different due to

- Volunteered Personal Data
- Individuals in the control of their data
- Trust
- Data analysis for individuals
- Data tools for individuals

III. DIMENSIONS OF PERSONAL BIG DATA

Big data describes a massive volume of structured and unstructured data that is so large that it's difficult to process using traditional database techniques. [17-18] .Personal Data means any information that could identify a person directly or indirectly. The kind of personal data available have increased in recent times due to emergence of social media and growth in mobile devices. It is any information relating to a data subject, being an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by a data controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person. [17,19]

Dimensions of individuals personal data that can be stored, used, and shared under individuals control is

- Credentials management
- Relationships
- Employment career
- Transactions Repository
- Preference intentions
- Contracts
- Social Stream

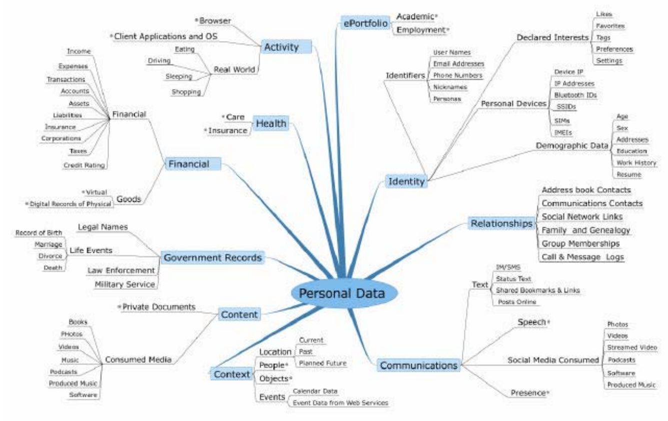

Figure 1 : Dimensions of Personal Data
Source: http://pimcoach.com/personal-data-management/world-economic-forum-graphic-3/

IV. INDIVIDUAL'S PERSONAL DATA STORE

Individual's data store is secure centralized location to store individual's data. Individuals will fully control usage of their data. Users will be able to view and manage their data under their control. Individuals will decide whether to provide their data in return of services or not. They can have better services and algorithms in return of their data. Two types of storage systems for such data store are possible.

First is cloud storage systems e.g. Dropbox[5], Carbonite[6] and second is personal data repositories but both the options have only basic type of access control and sometimes privacy is compromised. Personal Data Store we are presenting here is different from existing ones due to its privacy part.

V. PRIVACY CONCERN

Any digital data that individuals create in this digital connected world has sensitive private information about an individual. Users are now concerned about the use of their personal data. There are many revelations about this [7-8]. Numerous risks are associated with personal data and it is very hard to protect the privacy of the data. Anonymization is lost in this digital world. Many works have exposed the risks of re-identification or de-anonymization [2-4] [9-10]. There are many risks associated with data specially geospatial data which is recorded by smartphone applications [11-12]

Anonymity is hard but numerous methods have been proposed to solve this problem of de-anonymization [13-16]. Some methods are for very specific type of data and can't be extended for other types. While some methods are for very specific data applications and can provide privacy against some specific set of attacks.

A new approach we are providing here is to provide the privacy personal data by minimizing the dimensionality of data. Our solution will not give access to raw data but the computation that is done on the personal data to get information will be transferred from requesting application to individual's environment. In this way requesting application will get the final output not the raw data.

VI. ARCHITECTURE/PERSONAL DATA MANAGEMENT FRAMEWORK

The approach taken in our work is that any computations or algorithms are executed in the safe environment of user through personal data services. Raw data is not shared with requesting parties. User has complete control over that processing as only limited data or summarized results after the computations are returned.

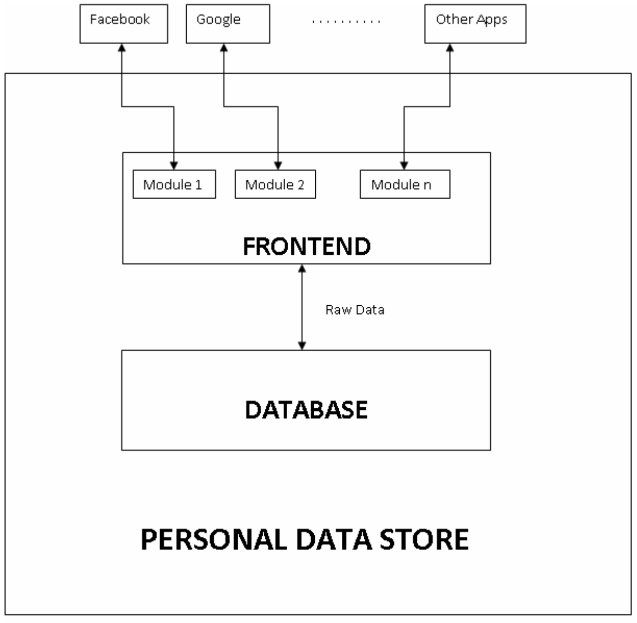

Figure 2: Personal Data Management System's Architecture

Let us explain this with example. Rather than exporting your all shopping details with products bought, it could be sufficient for an app to know which type of stuff you are looking for or will buy in the future. So users are not exposing all the data but only limited data that is actually required by the app. Users are aware of what data is being used for computation of such results.

Main component of the architecture would be:

A. FrontEnd

Browser and web client will form the frontend. Front-end will make sure that no unauthorized operations are carried out on the data. It will interact with requesting applications. It will consist of access control mechanism. Applications will be allowed to compute using only that data for which they have permissions. Each application/app will be registered as a user and will access the data through permission system.

B. BackEnd

Personal Data Service and Data Storage will form the Backend. Personal data will be stored in a database. All the modules will use the same database. Personal data service will let the user control how individual's personal data is shared with others.

VII. CONCLUSION AND FUTURE WORK

Finally, we know that personal big data has great potential. It is the new class of data that is emerging as an asset. But the benefits we are getting from personal data should be balanced with various types of risks associated with it. Review of most common data miners, security concerns and methods they provide to control personal big data on various social networks and many related works have always been studied. [17,20]. Our work open up a new way for individuals to have higher level of control over their data and to use it to take better decisions.

Approach discussed in this paper also has some challenges that open up ways for future research work. We need to have new set of data intensive services and privacy enabled algorithms to implement all this. Interactive and better User interfaces to allow users to better understand the risks associated and monitoring and visualizing the data used by other applications needs to be designed and implemented.

ACKNOWLEDGMENT

The work presented here could not have been completed without the contributions of many supportive and knowledgeable people. Foremost I want to express my deep gratitude to my guide Dr. Ritu Sindhu, Associate Professor (Department of CSE, SGT Institute of Engineering & Technology) for the continuous support and mentorship during my project. Her mentorship was paramount in providing a well rounded experience consistent with my long term career goals. She encouraged me to not only grow as an experimentalist and a personal data scientist but also as an individual thinker.

I would like to thank faculty members and fellows at Department of CSE, SGT Institute of Engineering & Technology, for their inputs, valuable discussions and accessibility.

REFERENCES

[1] "Rethinking Personal Data | World Economic Forum ..." 2014. 9 Feb. 2015 <http://www.weforum.org/projects/rethinking-personal-data>

[2] http://pimcoach.com/personal-data-management/world-economic-forum-graphic-3/de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the crowd: The privacy bounds of human mobility. Nature SRep 3.

[3] Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparsedatasets. In: Security and Privacy, 2008. SP 2008. IEEE Symposium on. IEEE,pp. 111–125.

[4] Sweeney L (2002) k-anonymity: A model for protecting privacy. InternationalJournal of Uncertainty, Fuzziness and Knowledge-Based Systems 10: 557–570.

[5] "Dropbox." 2009. 9 Feb. 2015 <https://www.dropbox.com/>

[6] "Carbonite Cloud Backup Services - Online Backup." 2005. 9 Feb. 2015 <http://www.carbonite.com/>

[7] Gellman B, Soltani A (2013) NSA tracking cellphone locations worldwide,snowden documents show. The Washington Post.

[8] Greenwald G, MacAskill E (2013) NSA prism program taps in to user data of apple, google and others. The Guardian.

[9] Solomon A, Hill R, Janssen E, Sanders SA, Heiman JR (2012) Uniqueness andhow it impacts privacy in health-related social science datasets. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. ACM,pp. 523–532.

[10] Butler D (2007) Data sharing threatens privacy. Nature 449: 644.

[11] Thurm S, Kane YI (2014) Your Apps Are Watching You. The Wall Street Journal.

[12] Stopczynski A, Pietri R, Pentland A, Lazer D, Lehmann S (2014) Privacy in sensor-driven human data collection: A guide for practitioners. arXiv preprint arXiv:14035299.

[13] Sweeney L (2002) k-Anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10: 557–570.

[14] Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M (2006) l-Diversity: privacy beyond k-anonymity. In: Proceedings of the 22nd International Conference on Data Engineering (ICDE'06). p. 24.

[15] Cao J, Karras P, Kalnis P, Tan K (2011) Sabre: a sensitive attribute bucketization and redistribution framework for t-closeness. The VLDB Journal 20: 59–81.

[16] Li N, Li T, Venkatasubramanian S (2010) Closeness: A new privacy measure for data publishing. IEEE Transactions on Knowledge and Data Engineering 22: 943–956.

[17] Jindal S, Sindhu R (2014) Personal Big Data Usage and Controls – Review, International Journal of Advanced Research in Computer Science and Software Engineering 4(10), October- 2014, pp. 351 – 354

[18] "What is Big Data? Webopedia." 2011. 29 Sep. 2014 http://www.webopedia.com/TERM/B/big_data.html

[19] European Commission, Proposal for a Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data protection regulation), COM (2012) 11/4 draft.

[20] de Montjoye Y-A, Shmueli E, Wang SS, Pentland AS (2014) openPDS: Protecting the Privacy of Metadata through SafeAnswers. PLoS ONE 9(7): e98790. doi:10.1371/journal.pone.0098790

AUTHORS PROFILE

Sujata Jindal, pursuing M.Tech. in Computer Science Engineering , SGT Institute of Engineering and Technology, Gurgaon, Haryana, India.

Dr. Ritu Sindhu, Associate Professor and Head, Computer Science Engineering Department, SGT Institute of Engineering and Technology, Gurgaon, Haryana, India.