A PROPOSAL OF "IDENTITY COMMONS" FOR SERVICE CREATION USING COMMUNICATION LOG

Position Paper

Memiko Otsuki
The Graduate University for Advanced Studies (Sokendai), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan

Noboru Sonehara
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan

Keywords: Privacy, Anonymization, Information loss, Utility.

Abstract: In this paper we consider the possibility of introducing Identity Commons (IdC) system. This will work as an ID bank to balance the utilization and protection of personal data. This paper first describes current problems on the utilization and protection of personal ID. It then proposes the IdC System. Finally, it makes some considerations regarding some of the institutional and technical challenges that need to be addressed in order to share personal data, combine them with related services and ensure the flexible use of them.

1 GENERAL CONSIDERATIONS AND CURRESNT PROBLEMS

Along with the development of advanced ICT network and terminals, electronic commerce and marketing research using digitalized personal identity information have drawn attention. Here we define two types of identity information. One is personal identification information which specifies an individual such as name, gender, address and date of birth. The other is attribute information which describes an individual such as shopping history, communication log, position information, clinical history and financial information.

Recently, consumers can buy goods and services easily and receive customized advertising through Internet. Companies can obtain consumers data and deliver goods and services with less cost and analyze consumers' information for more effective targeting using digitalized data. However, there is a problem that such service innovation is not advancing smoothly because of the following problems.

1.1 Problems of Consumers

Although online services can make people's life more convenient, consumers have several problems and anxieties of providing their personal information online. First, they have concern about the protection of individual information and privacy. Although technologies like data anonymization have actually developed, such technologies are still difficult for consumers to understand and are not able to untangle their concern about data leaks and hacking.

Secondly, it is hard for consumers to figure out how and by whom their information is used. Most of the countries have laws or regulations to obligate online service providers to display their privacy policy/notice on the website, it is sometime difficult to find or hard to understand. In general, consumers do not read the policy/notice carefully because it is too long. We picked up some websites of bookshops, banks and online shopping to examine how long the privacy policy is and found that it is three to five pages on average in A4 sized paper.

Thirdly, whenever consumers want to purchase goods and services, they have to input information such as a residential address, a telephone number, email account and their interests and hobbies over and over again on website of each company. This is quite time consuming. Also, they have to remember each log-in ID and password to enter those websites.

Finally, consumers have little option for providing information. They cannot receive goods and services unless they give all information the

Otsuki M. and Sonehara N..
A PROPOSAL OF "IDENTITY COMMONS" FOR SERVICE CREATION USING COMMUNICATION LOG - Position Paper.
DOI: 10.5220/0003605802310234
In *Proceedings of the International Conference on e-Business* (ICE-B-2011), pages 231-234
ISBN: 978-989-8425-70-6

service provider asks. Even if consumers think that some of the asked information items are not essential to receive the certain service, they do not have a choice not to provide them. Sometimes they do by contacting a customer service using telephone or email, but it needs so much effort that they cannot do that every time they shop online.

1.2 Problems of Service Providers

It is important for service providers to collect, accumulate and analyze information of consumers while protecting their security and privacy in order for effective targeting, advertisement and service delivery. However service providers also have difficulties to deal with such information. First, it is costly to obtain information securely and maintain them safely. Service providers need to install software to avoid virus and hacking. If they want the system more secure, they have to pay more cost. Also, if they want to acquire the "privacy mark", a certification by third party organization to guarantee service providers activities to keep privacy information safe, they have to pay more. In the case of Japan, when service providers newly apply for obtaining the mark, it costs JPY 300,000 for small sized companies and JPY 1,200,000 for large sized companies (which is approximately 0.3 to 1.2 million in US dollars).

Secondly, there is a possibility that service providers cannot collect appropriate information due to consumers' wariness. Service providers tend to collect as much information as possible to conduct a marketing research but consumers sometimes input false information when they think the information is not essential for the service (e.g. gender, age, telephone number, and hobbies and interest), which can lead to wrong result of the service marketing.

Lastly, but not least, service providers might lose possible consumers due to the requirement of providing information. Consumers might think that it is easier, safer and more convenient if they go shopping to a real shop instead of shopping online when possible even if a service provider offers better price and campaigns.

1.3 Problems of Other Institutions

There are several institutions conducting commercial and academic research. Some help companies develop marketing strategies and others serve for universities, government or governmental agencies by providing results of academic research in the fields of social studies, social psychology and engineering. However, current environment for data usage is hindering their ability of conducting more effective and accurate researches for the following reasons.

One problem they face is about data quality. Statistical analysis for marketing, academic and educational purpose generally uses a huge amount of data. However, to collect such data, it should be needed to make those data anonymous to ensure secure protection of personal data. Sonehara, et. al. investigated the data anonymization technology that guarantees the utility of the data depending on the purpose, and analyzed assessment criteria to maintain data anonymity and data accuracy (Huda et al., 2011a, Huda et al., 2011b). This study revealed that publishing human and social information using anonymization technology enhances the transparency of empirical study in the human and social studies. However, this study also found that research using data with more information tied with each other (which might identify individuals more easily) would extract more accurate results, which can lead to a result of higher quality than the research result using anonymous data.

The other problem is about a data collection. There are several companies in one industry field and each company has its unique set of data. However, it is still hard to integrate them for more effective research focusing on one industrial field. For example, communication logs and positional information collected by mobile phone companies are very useful for social studies including city planning. However, there are a few such companies and there is no way to integrate each company's data in one place due to the constraint of laws or regulations regarding personal data protection. If research organization, educational institution or even a company can obtain integrated data, regardless of whether they are anonymous or not, they can make use of nationwide data, which has more samples and can lead to more accurate and useful research results.

As described above, it is important to collect, accumulate and analyze dispersed personal data while protecting their security and making them available for various service providers for future service innovation. Therefore, it is necessary to establish a new data sharing platform which integrates a large amount of personal information that exists ubiquitously. Constructing the platform will create a cycle which generates great value for the whole society and economy. The next section proposes the possibility of establishing Identity Commons (IdC) noting its similarity with the copyright in the point of flexible use of information.

2 IDENTITY COMMONS: A PERSONAL DATA SHARING PLATFORM

To solve the abovementioned problems and promote an effective use of digitalized data, we propose to establish an organization called Identity Commons (IdC). This translates the idea of Creative Commons (CC) proposed by L. Lessig (Soenahara, et. al 2004) that ensures flexible use of copyrights by letting people attach license prepared by CC to their work. CC gives some conditions to each licence and people can choose how they want to release their copyright based on those conditions. The details of CC are referred at http://creativecommons.org/.

IdC has three main functions; integrating data in one place, letting consumers control the disclosure of their personal information by licensing and providing data to entities such as service providers and researchers in need.

2.1 Personal Data Integration

Figure 1 shows the general information flow from consumers to service providers. In the current commercial practice, system, consumers give service providers a different set of information. Each service provider maintains information and it is hard to be shared legally and technically.

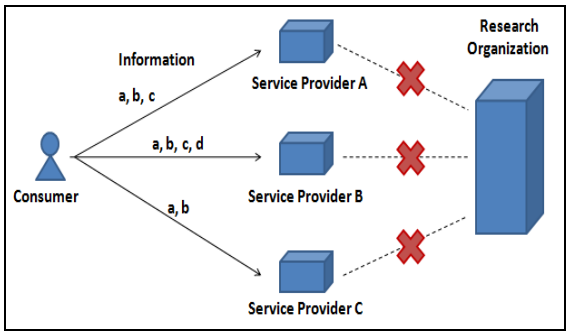

Figure 1: Current information flow.

Figure 2 describes a basic idea of IdC. Once consumers register information into IdC, they do not have to provide information to each service provider every time they purchase online. Service providers ask IdC to disclose consumers' information, which reduce costs for both consumers and service providers.

To be independent and reliable, IdC should be a non-profit organization consisting of staff with various specializations such as law, commerce, computer systems, security and privacy.

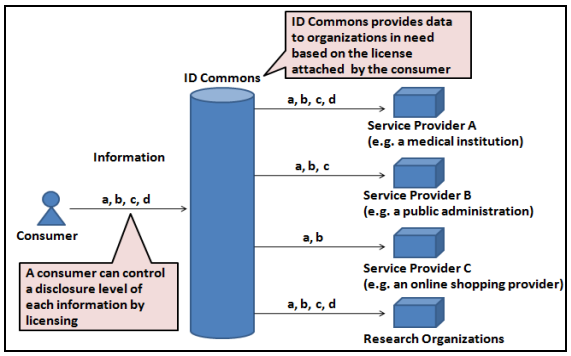

Figure 2: The idea of ID Commons.

2.2 Licensing

Another main feature of IdC is that it lets consumers control their information disclosure. It will solve many problems that consumers are currently facing.

Firstly, as shown in Figure 2, consumers can decide what information to be disclosed when they register their information to IdC.

Secondly, comnsumers can choose what information can be disclosed to whom. For example, medical institutions might need more private information such as the consumer's clinical history. Consumers agree with providing more data to such institution, but can provide less data to other service providers like online shopping company.

Thirdly, consumers can manage how the information is treated. It needs more detailed design, but we are currently considering conditions similar to the ones in CC. For example, consumers can choose to allow or not to allow service providers to pass their information to the third party. They can also choose to allow the information disclosure to the third party only if the service provider can keep it anonymous.

Finally, IdC system will enhance a transparency of information flow. Currently consumers cannot track who holds their information, and what information items it contains. Because IdC system is a one-stop platform of information, consumers can easily obtain how their information is treated by only accessing the IdC system.

These features are achieved by attaching a prototype of licenses that consumers can use free of charge when they want to manage their information. As employed by CC, IdC provides a format of Meta Data by RDF (XML) that make search and machine process easy on the Internet. In this proposal, the design of IdC License is introduced following the idea of CC. It is a mechanism that gives a license called IdCPL (ID Commons Public License) in order to permit the use of personal data. This license

works to enable users and service providers to easily access and securely use personal data.

IdCPL is composed of three layer structure; (a) Commons Deed that human can read, (b) Legal Code that lawyers can read and (c) Machine Code (RDF: Resource Description Framework) that machines can read.

2.3 Data Distribution

Finally, IdC system provides data to service providers and researchers in need based on the licenses consumers attached. It will enable to gather integrated information in one industry and to promote more accurate and useful research by using more number of samples with correct information.

It can keep identity information anonymous when it is exchanged between service providers, which technologies were developed by Huda et al. (Huda et al., 2011a, Huda et al., 2011b).

3 IMPLEMENTATION DESIGN

To achieve the effective circulation of personal data, this section considers the information licensing mechanisms and future prospects.

3.1 Transformative Identity

Sonehara et al. proposed a project called TEAM DigitalCommons that aims at the development of the contents circulation network (Sonehara et al., 2004). This is an approach to establish a framework using CC License to form a spiral of contents circulation.

In this project, contents open to the public of Commons have flexible use conditions such as file sharing and modification, and the production of derivation work, which is called Transformative Contents. In this proposal of IdC, we employ an idea of Transformative Identity, identity open to the public of Commons that have flexible use conditions such as marketing and service mash up, and the production of innovative services, which is achieved by aforementioned licensing system.

3.2 Future Prospect

IdC project aims at deterring an excessive statutory regulation concerning the protection of personal data and supporting the development of new information services. For future development, there are many things to be considered. The followings are the main categories of challenges for future prospect.

First, it is important to consider an institutional design. The challenges are to decide; who should manage the organization; how to make it reliable for the consumers and service providers; how to motivate consumers to register their information to IdC and; how to follow the existing laws and regulations.

The other important challenge is to consider a technical design. As the IdC becomes larger and as it has more information, the risk of being targeted by criminal activities such as hacking will also increase. Appropriate and cost-effective security measures should be needed when the IdC is designed.

To disclose personal data under the license of IdCPL does not necessarily mean the abandonment of the privacy right. It is rather a mechanism to offer users a choice to provide the right occasionally or only under certain conditions to form a rising spiral of information for more effective use.

REFERENCES

Sonehara, N., Minami, K., Abe, T. and Lessig, L. (2004). "TEAM Digital Commons ―A Plan for Market Activation by Revolution of Network Contents Circulation." *NTT Technical Review*, April 2004, 30-35.

Huda, M. N., Yamada, S., and Sonehara, N. (2011a). An Efficient k-anonymization Algorithm with Low Information Loss. A conference paper presented at *IEEE Data Engineering and Internet Technology (DEIT 2011)*.

Huda, M. N., Yamada, S., and Sonehara, N. (2011b). On Enhancing Utility in k-anonymization. A conference paper presented at *International Conference on Database and Data Mining (ICDDM 2011)*.