Integrating Data Masking Standards and Applications into Open Government Data

Yumei Chen[a, 1], Yuan Hong[b] and Theresa A. Pardo[c]
[a] *School of Public Administration and Emergency Management, Jinan University, China, Email: tchenym@jnu.edu.cn*
[b] *Department of Information Technology Management, University at Albany, SUNY, NY, Email: hong@albany.edu*
[c] *Center for Technology in Government, University at Albany, SUNY, NY, E-Mail: tpardo@ctg.albany.edu*

Abstract. This poster will present our work in progress in the development of a series of data masking standards and applications for the compliance of security policies in the context of open government data. Our current efforts focus on the feasibility analysis to use data masking technologies in datasets processing, access and download.

Keywords. Open Government Data, Data Masking, Standard, Data Security

1. INTRODUTION

Open Government Data (OGD) has recently become a worldwide hallmark of modern government. Of concern in efforts to make government data open is that a wide variety of government data includes sensitive information, thus utilizing government data to create value-added benefits via different OGD actions may explicitly compromise the government and/or individuals' privacy. Therefore, alternative solutions for desensitizing government data must be explored.

Data masking aims at identifying and removing the sensitive information in the "raw" data to make the data publishable where the utility of the published data can be maximized.

In this poster, we explore connections between OGD and data masking standards and applications. Given barriers to OGD progress, we focus on data quality and security. In this poster description we provide a brief review on the concepts of data masking technology and its standard uses and preliminary ideas of the application of these concepts in OGD. We finish the paper by providing our current plans and next steps.

2. OBSTACLES TO OPEN GOVERNMENT DATA

While OGD efforts can potentially provide numerous benefits, such efforts face a number of barriers. From the data provider's viewpoint, governments have concerns for privacy, confidentiality and liability [1] as major obstacles to the progress of OGD. Barriers are categorized into institutional, task complexity, use and participation, legislation, information quality and technical levels [2]. Specifically, at the legislation level, privacy violation and security are mentioned most frequently. At technical level, absence of standards and lack of meta data standards are pointed out. Data quality and security are mentioned at all levels. In fact, data is spread and fragmented across different agencies, each of which is responsible for just some of the data. TNO conducted a survey among policy-makers and experts in five different countries (AU, DK, ES, UK, US) [3], and identified top 10 barriers in which the second is the tension between open data and the privacy of their citizens. In the rest security and privacy threats and lack of standardization are mentioned.

Data masking technologies are one opportunity for protecting data from abuse and preserving data privacy during data sharing and exchanging inter-organizations. Such approaches should be introduced to the processing of raw datasets for use by the public.

[1] Corresponding Author.

3. DATA MASKING MEANING TO OGD

Data masking standards and applications are key to OGD efforts. As OGD becomes a priority for at all levels of government it is necessary understand the legal aspects of protecting privacy of citizens.

Traditionally each agency is responsible for ensuring compliance with legal requirements for data usage. While in the OGD ecosystem data will be shared and exchanged across agencies and organizations, and data will be opened to citizens, business companies and NGOs, it seems that data the capability of each agency to ensure compliance is very difficult if not impossible. Data masking is one approach to protecting privacy data, in that masking would be carried out before the data becomes "open". The suggestion described as bellowing Figure 1:

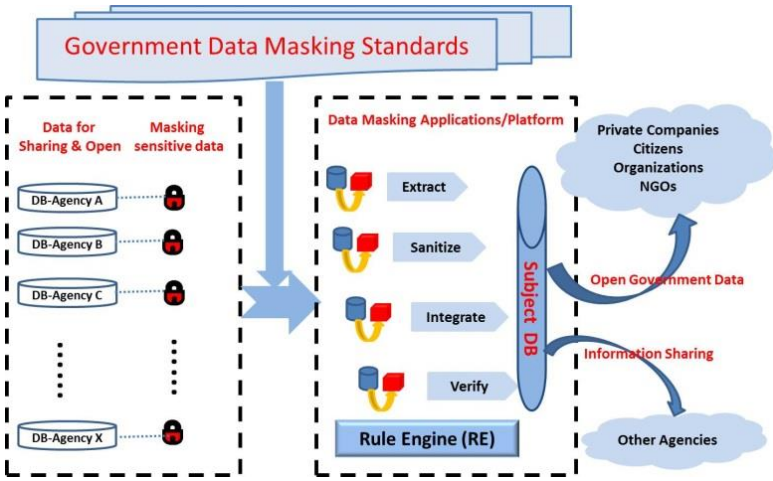

Figure 1. Government Data Masking Process.

Under FOIA [4], nine different categories of information are defined as sensitive information, and are therefore, exempt from requests for disclosure. In the next generation OGD actions/framework, integrated novel data masking techniques should suppress the sensitive information mentioned above while maximizing the output utility of the desensitize data, and new data masking standards will be proposed to measure the privacy protection in the context of OGD. More specifically, data can be generalized or suppressed to satisfy a predefined privacy notion (e.g., k-anonymity [5]); a randomization mechanism can be developed to publish probabilistic OGD while the randomization satisfies a "differential privacy notion" [6]. The privacy notions or novel privacy standards can be extended to formal Data masking standards.

3.1. CONCLUSION AND FUTURE WORK

Our current efforts focus on the development of a framework that uses the concepts of Data masking to take advantage of current Open Government Data experience. The concept will integrate standards, applications and governance structures to facilitate collaboration among government agencies in open government data practices.

References

[1] Dawes, S.S., et al., Planning and designing open government data programs: An ecosystemapproach, Government Information Quarterly (2016) ;
[2] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29(4), 258–268.
[3] Noor Huijboom, Tijs Van den Broek (2001). Open data- an international comparison of strategies.pdf. European journal of ePractice, 12(1), 1-13.
[4] FOIA, Freedom of Information Act, http://www.foia.gov/
[5] L. Sweeney (2002), k-Anonymity: A Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5), 557-570.
[6] Cynthia Dwork (2006), Differential Privacy. ICALP (2), 1-12