

大數據時代的個人資料隱私 與去識別化之探討

劉宜君*

元智大學社會暨政策科學學系教授兼主任



摘要

大數據演算系統的特色是從龐大雜亂的資料中，挖掘出無法預見的關聯，甚至可能重新識別個資當事人，進而連結個人敏感的資訊。為此，資料運用者難以預見未來可能的應用目的，往往無法明確告知當事人其蒐集目的及利用範圍。隨著越來越多的資料被保留，以往大部份使用者未意識到大量個資被組織因不同目的保有，以及資料隱私意識的興起，此一情況正在改變。因應大數據的發展，有必要在隱私保障與資料運用的便利性之間取得平衡，本研究主要說明大數據的衝擊、隱私保障，以及去識別化的意涵與社會影響。研究發現大數據雲端運算服務可能涉及的個人資料保護問題既深且廣，從原則性的問題（例如個資法效力範圍、去識別化與加密資料是否屬於個人資料、如何區分資料管理者與受託者），到較為細節性、技術層面的事項（例如對受託處理個人資料的雲端服務提供者監督的具體內容、因應雲端服務國際傳輸頻繁採取的措施），以及雲端環境下特殊的資料安全維護義務內涵。

關鍵詞：K匿名、大數據、去識別化、個人資料、隱私保障

* 通訊作者：劉宜君

電子郵件：chun0820@saturn.yzu.edu.tw

Personal Data Privacy and De-identification in the Age of Big Data

I-Chun Liu *

Professor & Director, Department of Social and Policy Sciences, Yuan Ze University

Abstract

In large-number calculation systems, unforeseen links can be mined from amid a chaotic mass of data. It is even possible to re-identify sensitive personal data of people involved and thereby make a connection to sensitive information about individuals. Therefore, data users find it difficult to envision possible future applications and purposes, and often it is impossible to give the persons involved clear information about the purpose of its collection and the scope of its utilization. In the past, most users were unaware that a large quantity of personal data was organized and retained for different purposes; however, this situation is now changing because with the retention of ever-increasing quantities of data, there has also been an upsurge in awareness of data privacy. To respond to the development of big data, it is essential to achieve a balance between the safeguarding of privacy and the convenient use of data. Our study primarily aims to present the implications and social effects of the impact of big data and the safeguarding of privacy and de-identification. This study found that personal data protection issues involved in big data cloud operations services are both deep and broad, and range from broad issues of principle (for example, the scope of the effectiveness of the Personal Data Protection Act, whether de-identified and encrypted data is classified as personal data, and how to differentiate between data managers and trustees), to comparatively detailed, technical matters (for example, the specific nature of the supervision of cloud service providers who are entrusted with handling personal data and the measures frequently

* Corresponding Author: I-Chun Liu
E-mail: chun0820@saturn.yzu.edu.tw

劉宜君

taken in response to international transmissions by cloud services), as well as the meaning of special obligations for the protection of data in the cloud.

Keywords: k-Anonymity, big data, de-identification, personal data, privacy protection