

# 巨量資料相關應用的規範省思

劉靜怡\*

## 一、前言

個人資料保護與隱私權保護相關爭議，是資訊時代裡無法避免的課題，眾所周知。在巨量資料 (big data，或稱「大數據」) 相關應用和討論日益受到重視的今日，政府或民間所擁有的各種資料，無論產官學界來說，自然是想像空間無窮的珍貴「國家寶藏」。除了政府治理、學術研究和各種產業應用之外，由於基於促進政府運作的透明、課責與創新等目的考量，「公開資訊」或「開放資料」的壓力與需求也日趨龐大，以國際發展趨勢來看，例如美國歐巴馬總統任內提出「開放政府」的倡議<sup>1</sup>，歐盟在 2013 年也修正了「公部門資訊再利用指令」(Directive 2013/37/EU)<sup>2</sup>，足見政府資料開放與再利用受各國政府重視之程度。不過，政府資料開放可能涉及哪些法律爭議，許多具有公開或開放必要性的資料，是否可能引發個人隱私保護的疑慮，這些爭議與疑慮是否和大數據應用無法脫離關係？在現行個人資料保護法制的架構下，是否會帶來一些嚴峻的挑戰？另外，不難想像的是，某些類型的大數據應用很可能涉及敏感性個人資料，那麼，這類應用因涉及個資如何處理與利用，是否應有事前的研究倫理審查機制，扮演把關甚至煞車角色，同時，應該採行怎樣的風險控制措施，方足以回應資訊安全的疑慮，均屬難以迴避的課題。

換言之，確認個資保護法制下的主要基本原則，在大數據時代中會遭遇哪些挑戰，是否仍有適用餘地，尤其是理解規範研究方向的發展趨勢，可以帶給我們哪些思考反省的線索，恐怕都是研究大數據應用以及大數據應用所涉及的

---

\* 國立臺灣大學國家發展研究所教授

<sup>1</sup> The White House, Open Government Initiative, <https://obamawhitehouse.archives.gov/open> (last visited Nov. 25, 2017).

<sup>2</sup> Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 Amending Directive 2003/98/EC on the Re-Use of Public Sector Information, 2013 O.J. (L 175), available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>. (last visited Nov. 25, 2017).

個資保護與隱私保護爭議，甚至處理大數據應用的倫理爭議時，應該踏實耕耘的研究領域。

## 二、巨量資料的概念內涵與應用潛力

與其說現代生活已經離不開「凡走過必留下痕跡」的網路和資訊科技，或者目前最炙手可熱的研究者是「資料科學家」，不如說長期以來就存在於公私部門並且不斷累積的各種「大規模資料」或「巨量資料」，其實早已埋下當前大數據風潮的前奏曲：今天的研究者——至少同時包括資料科學家和專業領域知識研究者在內——可以不斷透過效能更佳的技术進行大數據分析，以便更深入理解人類行為背後的內在因素和外顯特徵的來龍去脈，甚至協助預測未來發展趨勢，使得公私部門能據以做成公共治理決策或商業決策。

以近年來美國政府與民間企業的大數據分析應用為例，其乃是被視為無論是在提升政府與民間服務品質、商業創新、醫療研究突破、治安維護和國家安全提升……等方面，均會帶來無窮可能性的新機會。以 2013 年美國歐巴馬總統所提出的「大數據研究與發展倡議」(Big Data Research and Development Initiative) 為例，即以匯聚多個聯邦部會與機構(國家科學基金會、國家衛生研究院、國防部、能源部、地質調查局等)之力，投入兩億美金，推動大數據的基礎研究和教育，並且專注在醫療、國防、氣候變遷、地震研究等領域的應用上<sup>3</sup>。接著，總統顧問 John Podesta 則領導一個高層的跨部會小組，研究大數據科技所帶來的影響<sup>4</sup>。總統科學技術顧問委員會也組成工作小組，同步研究大數據可能帶來的隱私影響<sup>5</sup>。

這兩個小組在 2014 年分別提出報告，一致認為大數據分析在各個層面的應用，加上開放政府資料與物聯網科技的配合，不論是對公部門或私部門來說，都會帶來新的發展契機。例如，第一個小組報告即指出透過大量醫療紀錄、基因資訊與臨床試驗等資料進行比對分析的結果，可能有助於發現並歸納特定類型病人對不同治療的反應，除了可以改善醫療品質外，也將有助於達成個人化醫療(personalized medicine)的理想。再者，透過線上學習課程紀錄，例如學生

<sup>3</sup> Office of Science and Technology, Executive Office of the President, Obama Administration Unveils “Big Data” Initiative: Announces \$ 200 Million in New R&D Investment, Mar. 29, 2012.

<sup>4</sup> Executive Office of the President, Big Data: Seizing Opportunities, Preserving Values, May 2014.

<sup>5</sup> President’s Council of Advisors on Science and Technology, Executive Office of the President, Big Data and Privacy: A Technological Perspective, May 2014.

對不同教學方法的反應、各種情境下的學生注意集中度和尋求協助紀錄……等資料進行分析，其結果將有助於教育成效的提升。不過，這兩份報告在預測大數據前景的同時，也嚴肅指出：大數據相關應用對既有的價值（例如隱私和平等）造成衝擊，對現行個人資料保護法律的基本架構及其原則帶來挑戰，都是不可避免的結果，因此，在規範層面應該如何因應，也就成了不可避免的課題。

同時值得注意的是，政府運作的「開放透明」，向來被視為民主治理的基石，其不僅可以促進透明度以改善政府效能和落實責任政治，政府資料的開放，也有助於各種決策和科技創新，有助於促進市場競爭和經濟發展。例如，在公共衛生領域頗為知名的 Google Flu Trends，可以快速有效地追蹤流感，有助於公共衛生措施的落實。因此，近年來美國不管是聯邦、州或市政府各種鼓勵機關以合法為前提，釋出可用資訊的政府資料開放計畫，不勝枚舉，也就不足為奇了。

除了以上所述之外，大數據的分析應用領域，還有許多其他的想像可能性。例如，以個人健康資訊為例，雇主其實具有很高的動機，透過大數據的運用，找到最健康的受雇者，市場上的廠商和廣告業者，也會希望透過個人健康資料的大數據分析，找到最合乎其獲利要求的消費者。在這些需求下，不難想像的是，除了市場上的資料仲介商所發揮的功能外，各級政府部門所建置的大型資料庫，也可能是上述資料的來源。

換言之，無論是政府、雇主或廠商，都會希望透過大數據分析所得的各種人口統計學資料，協助其做出最佳決策。甚至，醫療研究人員在為了發現健康預測指標而挖掘資料來源這方面，也展現出前所未有的豐富創造力。不過，因為政府提供的開放資料中，包含幾可謂為無窮的個人資訊，因此，如何透過有效的規範來處理各種個資的「再利用」(secondary use) 可能衍生的爭議，是關鍵所在。究諸實際，根據美國各級政府的執行經驗，在資訊隱私保護和資訊安全維護上的無效率，正是公眾疑慮的根源，反而會成為政府開放資料與政府資訊使用的主要障礙，因此，以平衡但有效的模式解決規範難題，應該是大數據應用能否充分發揮潛力的關鍵所在。

### 三、巨量資料應用下的規範難題

大數據應用的特性，往往就是其規範難題起源。首先，在大數據應用可以改善人類生活條件的前景中，也同時創造了新的歧視 (discrimination) 風險。因此，如何在規範上適當處理根據大數據分析結果而「預先採行」差別待遇措施所

引發的歧視現象，有效禁止公私部門基於當事人「未來可能出現」特定身心缺陷而予以差別待遇，並且有效禁止雇主與保險公司不得基於大數據分析所得的基因資訊，採行歧視措施，是必須正視的規範難題。

舉例言之，已經有不少研究指出，在大數據的時代裡，現行法中關於身心障礙的定義恐怕已經不足以處理歧視問題。因為，現行法的規範對象，並未及於對「未來」身心障礙的預測，也就無法阻止決策者透過大數據分析結果認定某些人可能因為其健康習慣、壓力測試、曝露於環境汙染或其他公害下等原因，因此會在未來出現身心障礙，並且以此為基礎做出決定。

或者，美國現行法即《反基因歧視法》禁止雇主與健康保險公司基於基因資訊而對個人採取差別待遇，因此禁止雇主和健康保險公司搜尋包含個人家族健康史在內的基因資料，也不可基於基因異常和家族健康史而預測個人將來可能生病，而對該個人做出不利決定。然而，此一《反基因歧視法》乃是以基因資訊歧視當作規範對象，並不包括其他基於預測而得的資料，自然也不包括經過大數據分析所得的資訊，因此現行法令可能有修正調整之必要。同樣地，利用大數據技術進行資料探勘所得結果，究竟是否應該納入修法範疇的討論，也因為大數據所提供的「預測未來」功能而衍生出來。反歧視法令為了提供有意義的保護，或許需要將「使用大數據分析」則「應予以揭露」的要求納入規範，例如要求雇主如果想以除了傳統醫療檢驗和調查以外的大數據分析方式，獲得申請人和受雇人的健康醫療相關資料時，必須以書面通知申請者與受雇人。換言之，在大數據時代裡，是否應該重新建構一套反歧視法令，將未來可能因為其經過大數據分析所得的習癖、購物紀錄、醫療健康紀錄、生物標記等指標之故，而遭受差別待遇並因此受有損害的個人，無論在學術上或實務上均有思考價值。

對於研究者來說，環繞在大數據分析應用的「歧視」，不僅是法律問題，也可能隱含倫理風險，同樣值得關注。例如，大數據分析在教育上的應用，固然有助於因材施教，但也可能因為誤用、濫用，導致將受教育者予以不當甚或歧視性分類的結果，進而影響其未來發展機會。同樣地，其他諸如房屋租賃、商業消費、保險、信用評等……種種領域，也可能因為利用大數據分析進行預測而導致歧視的發生。

究諸實際，因為個資而衍生的反歧視爭議，早非新聞，在個資所涉及的法律和倫理雙重意涵之下，我們不難發現，敏感性較高的個人資料，例如健康醫療資料，向來是個資保護領域的討論焦點，大數據時代也不例外。由於個人健康醫療資料具有相當高的敏感性，不管是不當使用或者外洩，都將可能讓個人

隱私造成難以回復的傷害。所以，上述大數據相關應用不免衍生下列規範困境：現行個資保護法制賴以存續的「足以識別個人資料」前提，是否依然適用？畢竟，不管公私部門，進行大規模而且日益複雜的資料分析比對之後，其結果可能是原本去識別化的資料，或本已不足以識別特定個人的資料，帶來了「再識別」可能性大幅提高的風險，個資保護規範究竟應將其歸類為足以識別特定個人的資料，適用個資保護規定，抑或將在大數據分析下具有高度再識別風險的資料歸類為非個人資料，不受《個資法》之保護？若果如此，是否無異於為大數據時代量身訂做了開啟隱私保護漏洞的體制？若非如此，而是設法找出可以平衡大數據運用與隱私保護兩者的規範架構，那麼，針對現行個資法制中的匿名化或去識別化規範，應該朝怎樣的方向調整？

其次，傳統個資保護法制是以「告知後同意」(notice and consent)原則為核心，亦即必須在資料蒐集之「當時」，告知資料當事人蒐集與利用其個資的「特定目的」，並取得當事人的同意，而取得同意後的個資利用，也受到該告知內容亦即特定目的限制，以落實保護個人對於自己個資之「自主控制權」此一原則。

不可否認地，現行「告知後同意」的機制，由於告知內容往往範圍廣泛且冗長複雜，當事人不見得充分瞭解，一旦同意則喪失對資料的控制權，事實上並未真正發揮維護資料當事人權益的功能。此一現行法制問題在大數據時代裡並未消失，而且，此一時代適用告知後同意原則，困難度大增。因為，大數據時代的特徵，便是資料蒐集的頻率與數量大幅提高，不但必然導致當事人將更無法仔細審閱告知內容的現象產生，當事人同意也將更流於形式化，更重要的是，大數據分析的特性，便是希望能夠藉此發現嶄新的、蒐集資料時未預期其存在的資料使用目的，所以極可能無從在資料蒐集當時即告知當事人蒐集目的和所有潛在使用目的；換言之，「告知後同意」此一現行機制難以運作，是大數據時代的必然命運。因此，如果我們肯認「隱私」依然是大數據時代的重要核心價值，那麼，「告知後同意」機制是否必須調整，如何調整才能確保資料當事人對自身資料的「自主控制權」，自然是當今重要研究課題。

此處值得一提的，是在大數據時代中應該如何看待「資料刪除」原則。傳統個資保護原則要求基於特定目的蒐集並使用之後，已經不需再利用的資料，應該予以刪除，一則用以落實當事人自主原則，他方面也藉此降低資料儲存維護的成本與風險。不過，隨著大數據分析的盛行，此一原則勢必遭遇挑戰。畢竟，各種大數據應用模式，正是以自各種類型的巨量資料（甚至是傳統上視為毫無價值的資料）中，透過大數據分析技術，找出各種新的應用可能性。因此，大數據時代中是否仍有「資料刪除」原則的存活空間？此一原則是否及如何調整，

也成為兼具理論與實務價值的研究熱區。

最後，無論公私部門，「資料控制」是權力集中的主要原因，不僅政府和私人大企業具有這種權力，資訊流的「中介者」（例如搜尋引擎、網路服務供應商、徵信機構、行銷公司等等）雖然不盡然具有生產資訊的功能，但因為在傳遞上扮演關鍵角色，也是值得關注的資料權力掌控者。在大數據時代裡，這些資料控制者不僅扮演資料近用的守門人角色，在資料使用下所衍生的知識生產、解釋和預測上，都扮演影響深遠的角色。因此，對於這些資料控制權力的行使，應該如何予以監督節制，如何設計出適當的管制架構，不僅涉及大數據時代中的資訊社會透明度高低問題，而且很可能是處理大數據時代下的「資訊生活憲政架構」何去何從的基礎，也是對於大數據時代的公共治理與公法爭議有興趣的研究者，不該忽視的題目。

#### 四、結語

大數據時代的規範研究課題，正如同大數據分析的應用一樣，很可能是無窮無盡的，有待研究者齊心合力不斷挖掘。而在這個規範研究的試驗場裡，不管是探索大數據應用帶來的法律和倫理爭議，或是檢討現行法律的不足之處，其終極問題，可能依然不脫「我們」作為一個「資訊生活共同體」，究竟應該接受怎樣的「資訊生活憲章」與「資訊生活管制架構」此一範疇的提問。在此一理解下，關於政府資料開放所涉及的隱私管制，是不是應該針對不同類型資料予以不同的隱私保護程度；關於個人資訊的再識別風險評估標準及其管制模式，應該如何進行類型化的處理，並且落實在具體的管制立法和措施中。

最後，大數據時代的政策制定者，在政策制定的各個面向，包括法律、科技、經濟、程序甚或教育層面，應該如何建構更為完善的開放政策和相應管制措施，特別注意各種資料應用模式可能衍生的威脅或風險，尤其是對個資保護潛在風險予以有效回應與處理，而非一味將個資保護法令當成阻礙大數據發展的絆腳石看待，或許才是比較健康而且可長可久的發展策略。

#### 參考文獻

- Ajunwa, Ifeoma (2014). Genetic Testing Meets Big Data: Tort and Contract Law Issues. 75 Ohio State Law Journal 1225.
- Altman, Micah et al. (2015). Towards a Modern Approach to Privacy-Aware Government Data Releases, 30 Berkeley Technology Law Journal 1967.

- Crawford, Kate & Jason Schultz (2014). *Big Data and Due Process: Toward A Framework to Redress Predictive Privacy Harms*, 55 *Boston College Law Review* 93.
- Ferguson, Andrew Guthrie (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press.
- Hoffman, Sahrona (2016). *Electronic Health Records and Medical Big Data: Law and Policy*, Cambridge University Press.
- Lane, Julia (2014). *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press.
- Lohr, Steve (2015). *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*, HarperCollins.
- Lynskey, Orla (2016). *The Foundations of EU Data Protection Law*, Oxford University Press.
- Miller, Kevin (2014). *Total Surveillance, Big Data, and Predictive Crime Technology: Privacy's Perfect Storm*, 19 *Journal of Technology Law & Policy* 105.
- O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increase Inequality and Threatens Democracy*, Crown.
- Pasquale, Frank (2013). *Grand Bargains for Big Data: The Emerging Law of Health Information*, 72 *Maryland Law Review* 682.
- Pasquale, Frank (2014). *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press.
- Pasquale, Frank & Tara Adams Ragone (2014). *Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing*, 17 *Stanford Technology Law Review* 595.
- Porat, Ariel & Lior Jacob Strahilevitz (2014). *Personalizing Default Rules and Disclosure With Big Data*, 112 *Michigan Law Review* 1417.
- Richards, Neil & Jonathan King (2014). *Big Data Ethics*, 49 *Wake Forest Law Review* 393.
- Rubinstein, Ira S. (2014). *Voter Privacy in the Age of Big Data*, 2014 *Wisconsin Law Review* 861.
- Schneider, Bruce (2016). *Data and Goliath: The Hidden Battle to Collect Your Data and Control Your World*, W. W. Norton Company.
- Schünemann, Wolf J. & Max-Otto Baumann (2017). *Privacy, Data Protection and Cybersecurity in Europe*, Springer.
- Adam Tanner (2014). *What Stays in Vegas: The World of Personal Data-Lifeblood of Big Business-and the End of Privacy as We Know It*, PublicAffairs.
- Tene, Omer & Jules Polonetsky (2013). *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 *Northwestern Journal of Technology & Intellectual Property* 239.
- Tene, Omer & Jules Polonetsky. (2013). *Judged by The Tin Man: Individual Rights in the Age of Big Data*. 111 *Journal on Telecommunications & High Technology Law* 351.
- Voigt, Paul & Axel von dem Bussche (2017). *The EU General Data Protection Regulation (GDPR): A Practical Guide*, Springer.