

中正大學

National Chung Cheng University



大數據發展趨勢以及在 犯罪防治領域之應用

國立中正大學犯罪防治系
教授兼系主任、所長
許華孚





許華孚 教授簡歷

英國艾賽克斯大學 (Essex University)

犯罪社會學碩士、博士

國立中正大學犯罪防治學系專任教授

暨系主任、所長

國立中正大學犯罪研究中心執行長

國立中正大學防制藥物濫用教育中心執行秘書

台灣青少年犯罪防治研究學會理事長

中華民國犯罪學學會副理事長

亞洲犯罪學會理事



前言

人類史上的幾項重大發明，屢屢引領文明躍進，隨著廿世紀末電腦問市，引領全球邁向另一場資訊革命以及帶動網路資訊蓬勃發展，當傳輸速度與儲存技術不斷向上躍升，一場跨越政府與民間企業的全球大數據 (Big Data) 應用革命正襲捲而來，不僅走入你我世界、改變眾人生活，更影響交易模式與政府決策思維。

前言cont.

大數據的發展不僅代表巨量數據、資訊先進技術、龐大商機，它更是一套解決問題、改造世界、探索未來的新穎觀念與科學應用的有效方法。因此，若從解決當前犯罪問題的思維觀之，大數據亦代表改造犯罪環境、矯治潛在犯罪人、減少犯罪機會、解決犯罪問題，以及研議未來犯罪防治對策的無限機會與可能。

大數據的意涵

- 「Big Data」一詞的應用最早出現於2010年IBM電腦公司所提出，係指一套系統化、標準化及科學化的複雜概念，含括商業與管理的專業，需具備電腦儲存的位元組概念，以及網路應用、系統管理等專業知識。
- Jaap Bloem(2013)解讀大數據之公式如下：
大數據（Big Data） = **事務處理（Transaction）**
+ **互動（Interactions）** + **觀測（Observations）**

大數據的意涵cont.

大數據資料產生的來源 (Mark Myers ,2010):

1.雲端資料 (Cloud)

2.移動通訊 (Mobile)

3.社會資料 (Social)

4.消費資料 (consumerization)

5.大數據 (Big Data)

大數據的產生，主要仰賴左揭
四種資料的匯流

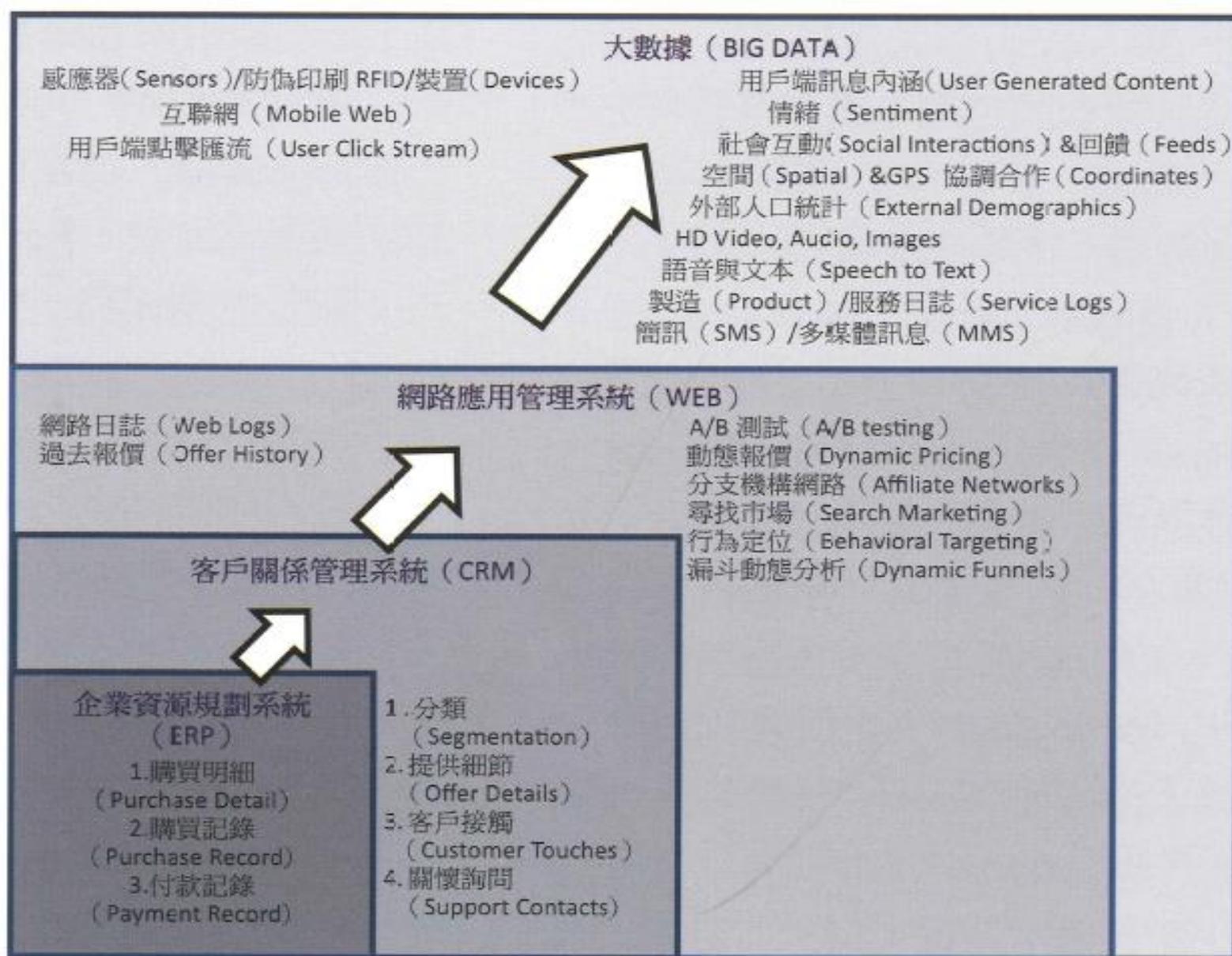
位元組的次方單位

拍位元組 10^{15}
(Petabytes)
1PB = 1,000 TB

兆位元組 10^{12}
(Terabytes)

吉位元組 10^9
(Gigabytes)
1GB = 1024MB

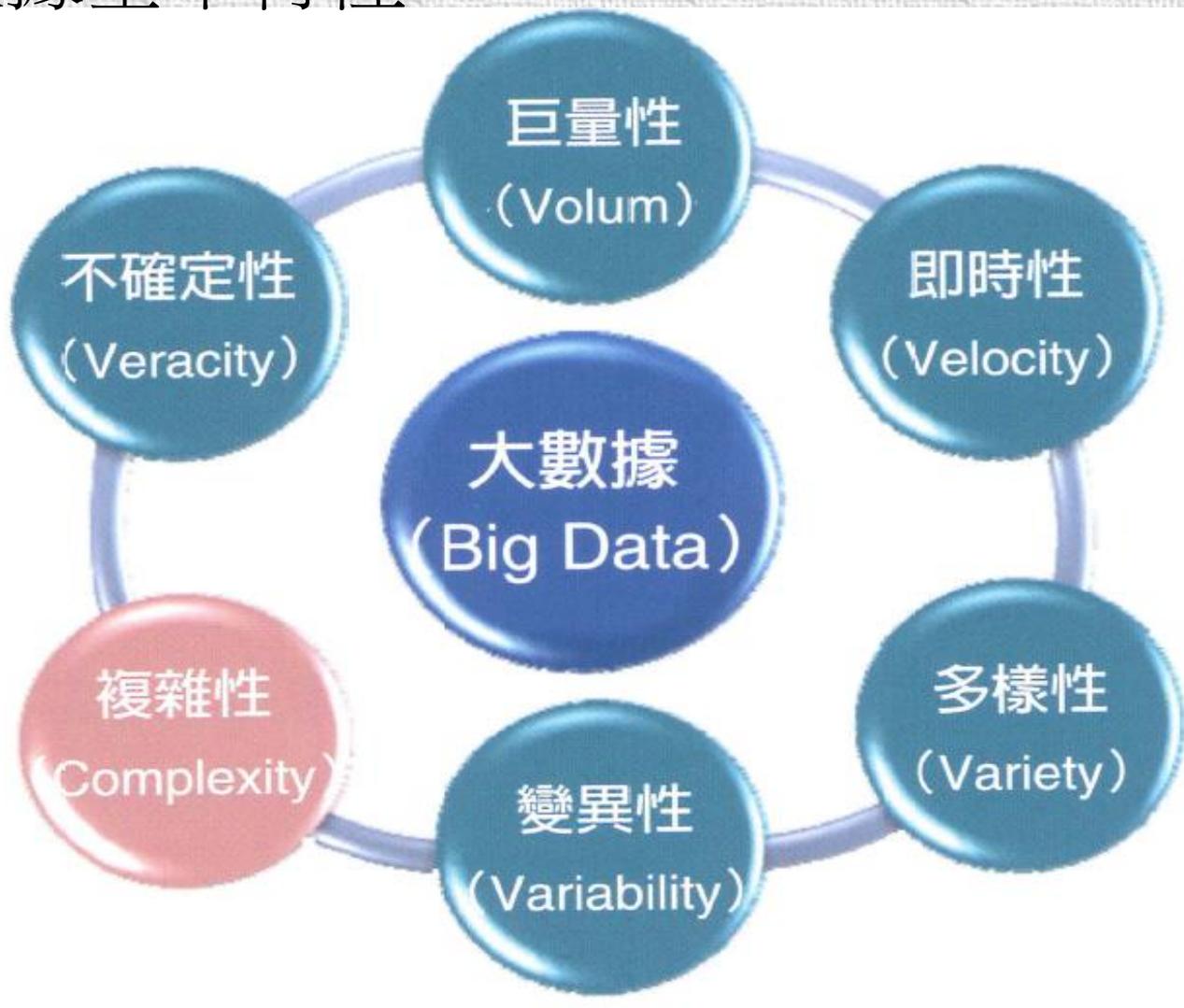
百萬位元組 10^6
(Megabytes)



資料逐漸多元性與複雜性 (Increasing Data Variety and Complexity)

大數據的意涵cont.

- 大數據基本特性：



大數據的意涵cont.

- ✓ 巨量性 (volume) : 大數據特色在於「龐大」，無論政府機構、民間企業以及即時性感應偵測器等資料多元而巨大，也由於資料龐大複雜，致難以利用傳統方式處理，必須透過新的硬體與演算系統加以管理分析，以為人類所能解讀（陳柏宇，2014）。
- ✓ 即時性 (Velocity) : 大數據通常具有時效性，一旦串流到運算伺服器就須立即使用，即時才能發揮高價值。由於資料產生的速度不斷躍升，越多的資料會透過網路或感測器即時記錄、傳遞與儲存，必須應用高效率及自動化運算系統才能分析。

大數據的意涵cont.

- ✓ 多樣性 (Variety) : 大數據資料範疇不僅止於結構化資料，尚含括非結構化資料，如文字、音訊、視訊、點擊串流、日誌檔等，而數據依組合性質可分結構化 (structured)、半結構化 (semi-structured)、準結構化 (quasi-structured) 和非結構化 (unstructured) 等四類型。
- ✓ 變異性 (Variability) : 此關於由「誰」來主導分析資料的關鍵問題，不同目的與背景的人，在相同時間所做分析會產生南轅北轍的結果，會嚴重妨礙資料管理與分析的正確性。

大數據的意涵cont.

- ✓ 不確定性 (Veracity) :由於資料在蒐集、儲存與分析過程，可能因參雜人為欺騙或操作錯誤，抑或因演算系統的失準等，致使大數據分析存在不準確的特性（胡世忠，2013）。
- ✓ 複雜性 (Complexity) :資料管理會是一個非常複雜的過程，特別是當大量的數據分別來自多個來源時。相關數據資料必須經由截取連結分類傳輸，最後被適當歸類儲存，這是龐大且複雜的處理過程。

大數據發展與應用現況

• 各國政府部門之發展：

(一)、美國

- ✓ 2009年聯邦政府即公布大量資料於中央信息交換庫(Data.gov)公民眾閱讀，制定「證據導向」的公共政策，以提升決策品質與效率。
- ✓ 2012年3月歐巴馬政府投資2億美元實施「大資料研究和發展計畫」，84個子計畫涵蓋國防、醫療、教育、能源、交通運輸、商業、科學、工業、國土安全等應用領域。
- ✓ 2014年5月1日美國白宮發表「2014年大數據白皮書」視大數據為**未來的新石油**。

大數據發展與應用現況cont.

(二)、亞太地區

- ✓ 新加坡政府與GE共同成立「大數據發展中心」，協助企業發展大數據應用。
- ✓ 韓國積極制定關於大數據發展策略，視其為「國家意志」之科技專案，並投資兩億美金預算，打造韓國首都成為資料匯流中心。
- ✓ 日本「創建最尖端IT國家宣言」闡述將以開放公共數據與大數據為核心，將日本建設成「世界高水準的信息產業應用技術社會」。
- ✓ 中國政府核批「十二五國家政務資訊化建設工程規劃」，發展計有人口、法人、空間、宏觀經濟和文化等五大資料庫的建設工程。

大數據發展與應用現況cont.

(三)、歐洲地區

- ✓英國實施「開放資料」專案，建立「資料英國」網站用於資料公開，2014年更宣布建立圖靈大數據研究院，確保英國未來大數據發展在經濟及社會中的領導地位。
- ✓法國政府「數位化路線圖」計畫，大力支持大數據戰略性新技術，投入3億歐元資金於推動大數據領域的發展。

大數據發展與應用現況cont.

(四)、我國發展現況

➤ 政府行動:

- ✓ 行政院長毛治國祭出科技三箭，其中兩件與數據相關，一是開放資料，二是發展大數據，並指派資訊專長之副閣揆規劃。
- ✓ 工研院成立巨量資訊科技中心—「國家高速網路與計算中心」，並與資策會同步著手大數據的研究

➤ 學術發展

- ✓ 東吳大學成立校級「海量資料分析研究中心」
- ✓ 元智大學成立「大數據與數位匯流創新中心」
- ✓ 交大、政大、輔大相繼成立大數據相關分析學程，極力架構一個國內與國際交流平台，主導國內外整體大數據發展態勢。

大數據發展與應用現況cont.

- 應用領域之發展：

(一)、疾病預防

- ✓ 美國官方傳統追蹤疫情擴散方式，乃靠蒐集醫師初診資料來推估，通報流程亦需時間，故掌握疫情速度約慢1至2周
- ✓ Google發表另一種追蹤方法，挑出美國最常使用的5000萬個搜尋字眼，並與疫管局流感疫情比對，發現可利用其中45個搜尋字眼，搭配某數值模型，成功預測2009年冬季流感散播。

大數據發展與應用現況cont.

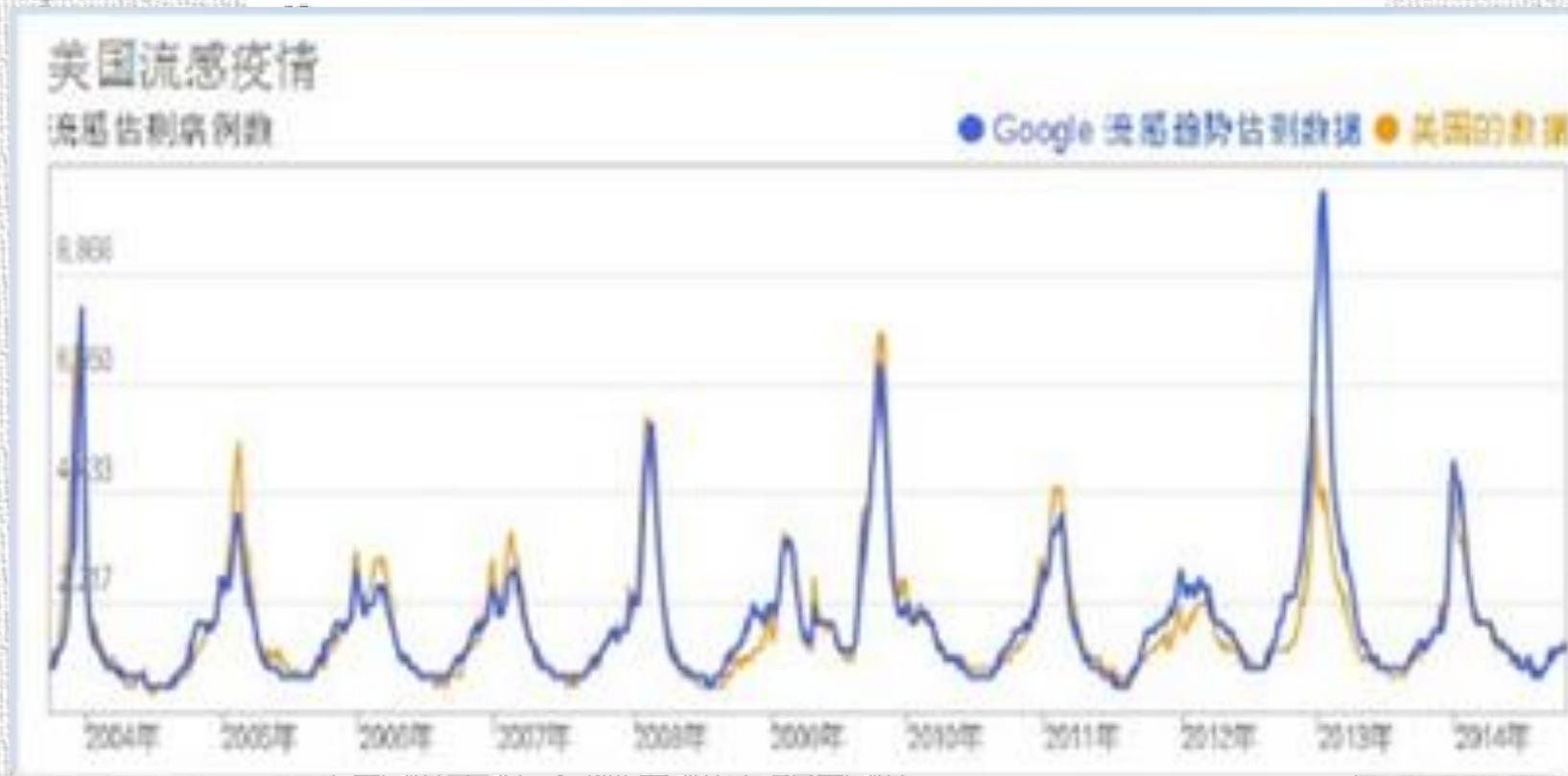


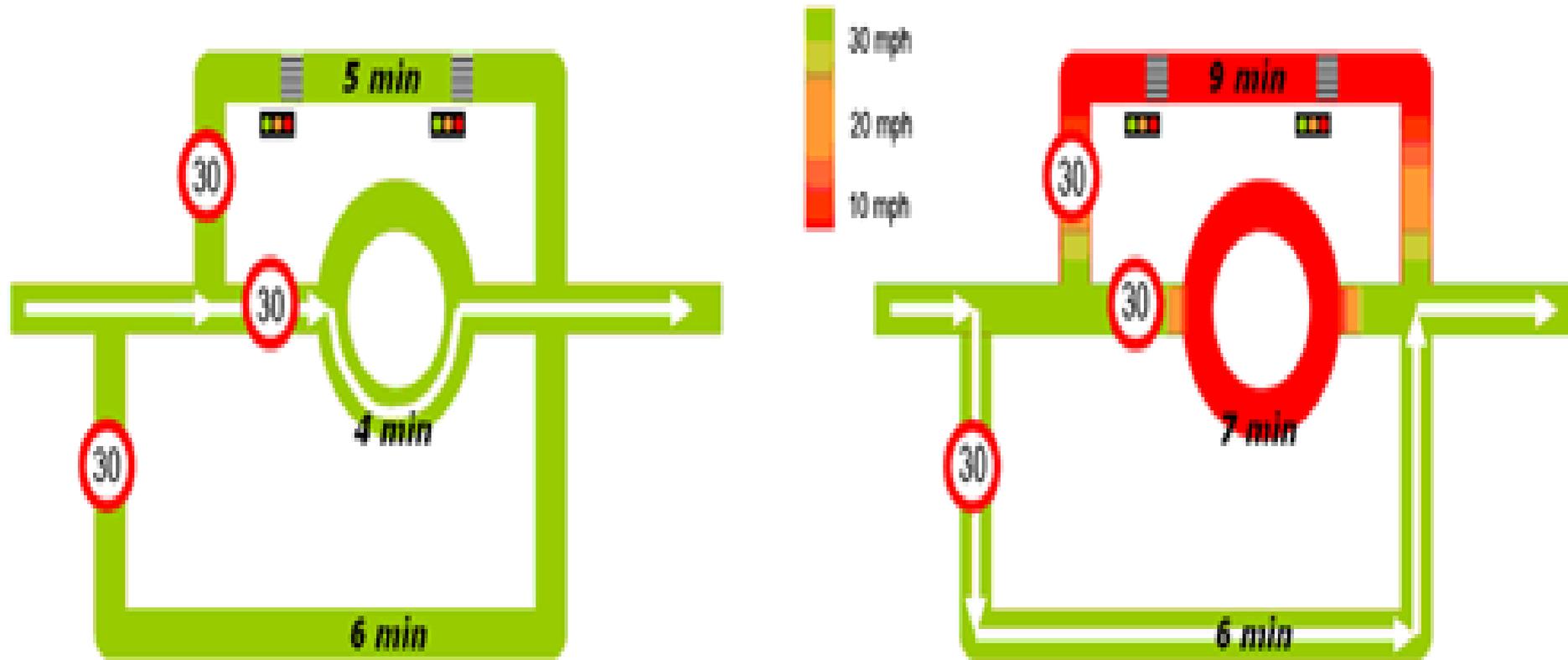
圖 4 美國流感疫情-google 估測病例數(資源來源: Baidu 文庫。網址: <http://wenku.baidu.com/view/ec55c144482fb4daa58d4b73.html>)。

大數據發展與應用現況cont.

(二)、交通運輸

- ✓ 美國衛星領導品牌TomTom利用監測的八千萬支行動電話，一百萬台TomTom Live 衛星導航積在路面移動速度，搭配RDS-TMC道路資訊系統建構即時資料庫，並透過GPRS即時訊息，例如某路段平均速度、道路施工及事故等，以每兩分鐘一次的頻率及時推播衛星導航器。
- ✓ 客戶只需使用Apple、iphone及Anddroid上的TomTom導航app，即可提供駕駛人最佳化路徑、節省時間，即使身陷車陣亦可準確預測延誤時間，根據統計平均可節省15%的行車時間。

大數據發展與應用現況cont.



最短路径與過去實際路径時間比較表

壅塞或施工左圖最短路径規劃；右圖為真實用路時間歷史紀錄，可清楚看到紅綠燈的多寡頻率以及真實行駛速度

大數據發展與應用現況cont.

(三)、醫療保健

- ✓ 以對岸為例，目前已有十餘座城市開展數字醫療，醫療的病例、影像遠程醫療等都製造大量數據，形成電子病歷及健康管理檔案
- ✓ 基於這些海量數據，醫院能精準分析病症、治療費用和療效數據，避免過度及副作用較高的治療，並用於遠程監護，針對慢性病進行管理。

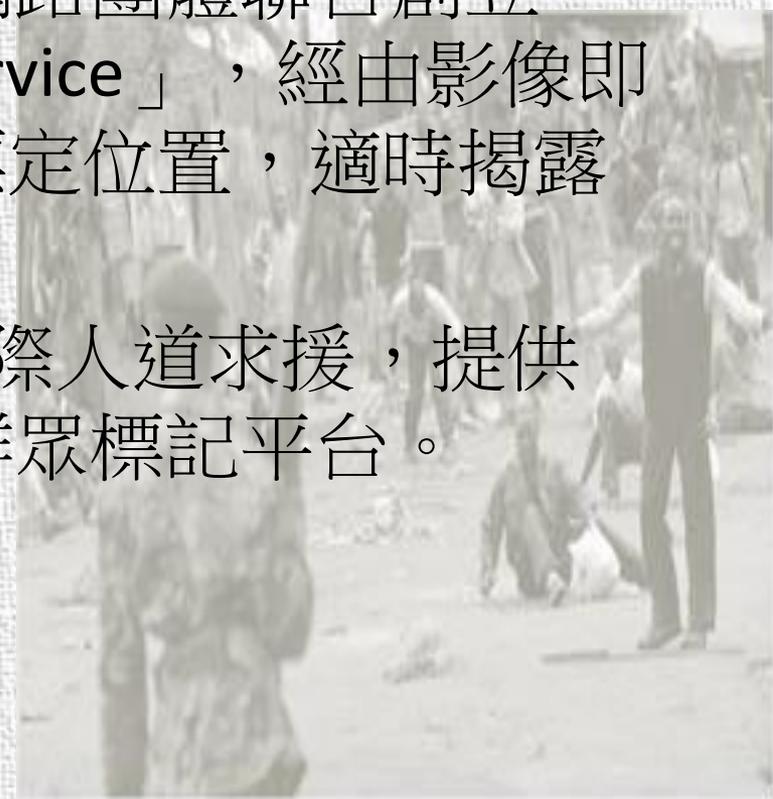
(四)、監督施政

- ✓ 美國「太陽哨兵報(Sun-Sentinel)」記者為調查員警超速，向當地交通管理部門申請110萬筆警車通過高速公路收費站紀錄。資料經大數據系統分析，竟有3900輛警車在13月內發生5100次超速紀錄。

大數據發展與應用現況cont.

(五)、人道救援

- ✓ 2007年東非肯亞共和國總統大選爭議引發全國暴動，一群程式設計師與網路團體聯合創立「Ushahidi hort Message Service」，經由影像即時上傳，復經Googla map標定位置，適時揭露地區實況，尋求國際援助。
- ✓ 隨後Ushahidi計劃常用於國際人道求援，提供多國用於重大事件通報與群眾標記平台。



大數據發展與應用現況cont.

• 大數據的發展未來，呈現八大趨勢：

- (一) 數據資源化，將成為最有價值的資產。
- (二) 大數據在更多傳統行業的企業管理落地。
- (三) 大數據和傳統商業智慧化融合，行業訂製化解決方案將湧現。
- (四) 數據將越來越開放，數據共用聯盟將出現。
- (五) 大數據安全越來越受重視，大數據安全市場將愈發重要。
- (六) 大數據促進智慧化城市發展，為智慧化城市的引擎。
- (七) 大數據將催生一批新的工作崗位和相應的專業。
- (八) 大數據在多方面改善我們的生活。

犯罪防治領域的相關應用

• 犯罪防治研究的應用:

(一)、國內資料庫建置與應用

中華民國刑案統計

台灣高等法院判決書

監獄個案資料

刑事司法部門
施政滿意分析

缺乏統計效度、
無法即時提供因應策略

透過大數據將左揭的資料庫之官方文件蒐集與調查，以了解犯罪模式概況，且針對個案訪談了解受刑人之犯罪模式、電話調查機構施政滿意與效能，最後對於以上的調查分析提出適切的防制犯罪對策與政策建議。

刑事司法系統大數據分析架構cont.

• 刑事判決書資料數位化及分析

- ✓ **刑事判決資料庫**：按裁判有罪者為犯罪之定義，以及司法院法學資料庫檢索系統所公開的刑事判決書，已對犯罪人的犯罪動機、手法及犯後態度等資訊有詳細的記載，則該資料庫具備豐富資訊可供探勘。
- ✓ **效益**：透過刑事判決書資料數位化後，可將大量非結構性資料轉換成可判讀之量化資料，建立各犯罪類型的模式與被害情境，供刑事司法人員運用，作為擬訂政策之參考。
- ✓ **限制**：現行收錄之判決書多為純文字資料，未經字串編碼、過濾的數位化處理，無法發掘各犯罪類型的犯罪者犯罪動機、手法與被害情境。

刑事司法系統大數據分析架構cont.

• 犯罪矯正部門資料統計與分析

✓ 獄政統計資料包含：

- 1) 矯正統計指標、性別統計（依照人數、罪名、前科情形、教育程度等基本人口變項進行統計）
- 2) 業務成效：依照參與教化處遇、技能訓練、更生保護等獄政業務之人數進行統計
- 3) 專題分析：由業務項目、犯罪類型中抽出專項進行分析，或針對特殊議題進行描述性統計等。

✓ 限制：

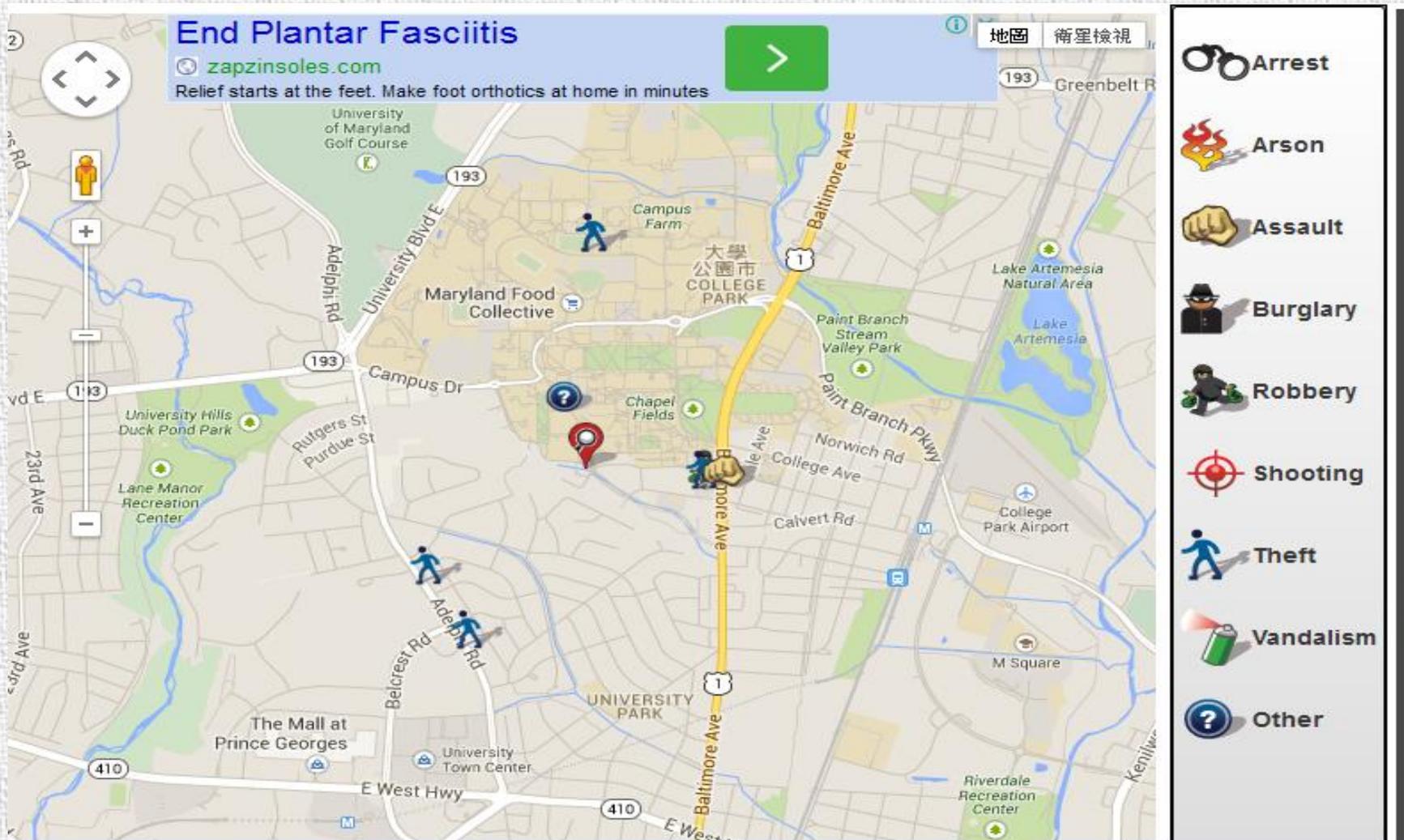
- 現有矯正統計資料庫龐大，但欠缺統合各變項關係及未來趨勢之分析，也無法深入討論專項獄政業務對收容人之影響，以及處遇感受、生活適應、心理健康、人際互動等脈絡。

犯罪防治領域的相關應用cont.

(二)、國外資料庫建置與應用

- ✓目前美國、英國、澳洲等國家皆提供線上調查犯罪現況的資料，並採用治安氣象圖分析犯罪問題，將犯罪現況呈現於地圖上，以利於警政單位了解熱點所在，並掌握治安趨勢。
- ✓因此，藉由大數據資料的分析應用，可改善我國現有犯罪資料庫無法即時提供因應策略的缺陷，並深入了解犯罪資料之防治意義，掌握犯罪之動機及未來犯罪趨勢。

犯罪防治領域的相關應用cont.



資料來源:美國Spotcrime.com 線上分析治安氣象圖

犯罪防治領域的相關應用cont.

(三)、大數據資料的應用－犯罪趨勢的分析

- 依據警政署、法務部、衛福部等機關，在違反相關刑事法律所定義之犯罪發生數、犯罪率、犯罪嫌疑人數、破獲數等相關統計資料，可發現近十年犯罪趨勢如下：
 - ✓ 公共危險罪(酒醉駕駛及肇事逃逸)人數呈遞增趨勢
 - ✓ 青少年與兒童犯罪人口大幅增加
 - ✓ 犯罪者高齡化比率上升
 - ✓ 女性再累犯上升趨勢較男性明顯

➡ 採用上述相關統計資料，透過大數據分析將可預測犯罪者的特性、預防可能的犯罪事件，以減少社會危險，並能及時提出最適切的犯罪對策。

犯罪防治領域的相關應用cont.

(四)、綜合建言

- 綜合國內現有犯罪資料庫、地區性資料、過去五年施政滿意度調查，發現未來犯罪防治政策建議如下：
 - ✓ 提倡以知識和證據為基礎的刑事、矯正政策
 - ✓ 提供適宜的性別意識與需求之處遇措施
 - ✓ 高齡化犯罪者和受監禁之高齡化犯罪者之處遇措施
 - ✓ 毒品相關政策
 - ✓ 酒後駕車和肇事逃逸的防治政策
 - ✓ 少年犯罪的預防及處遇措施
 - ✓ 以司法院所上網公告之司法判決書為主要分析資料，在現有的基礎上，選定非性侵暴力犯罪為主要探討類型，就非性侵暴力犯罪之被告選定基礎特徵，追蹤其再犯情形。

犯罪防治領域的相關應用cont.

• 犯罪防治實務的應用

(一)、**犯罪活動線上即時偵測**

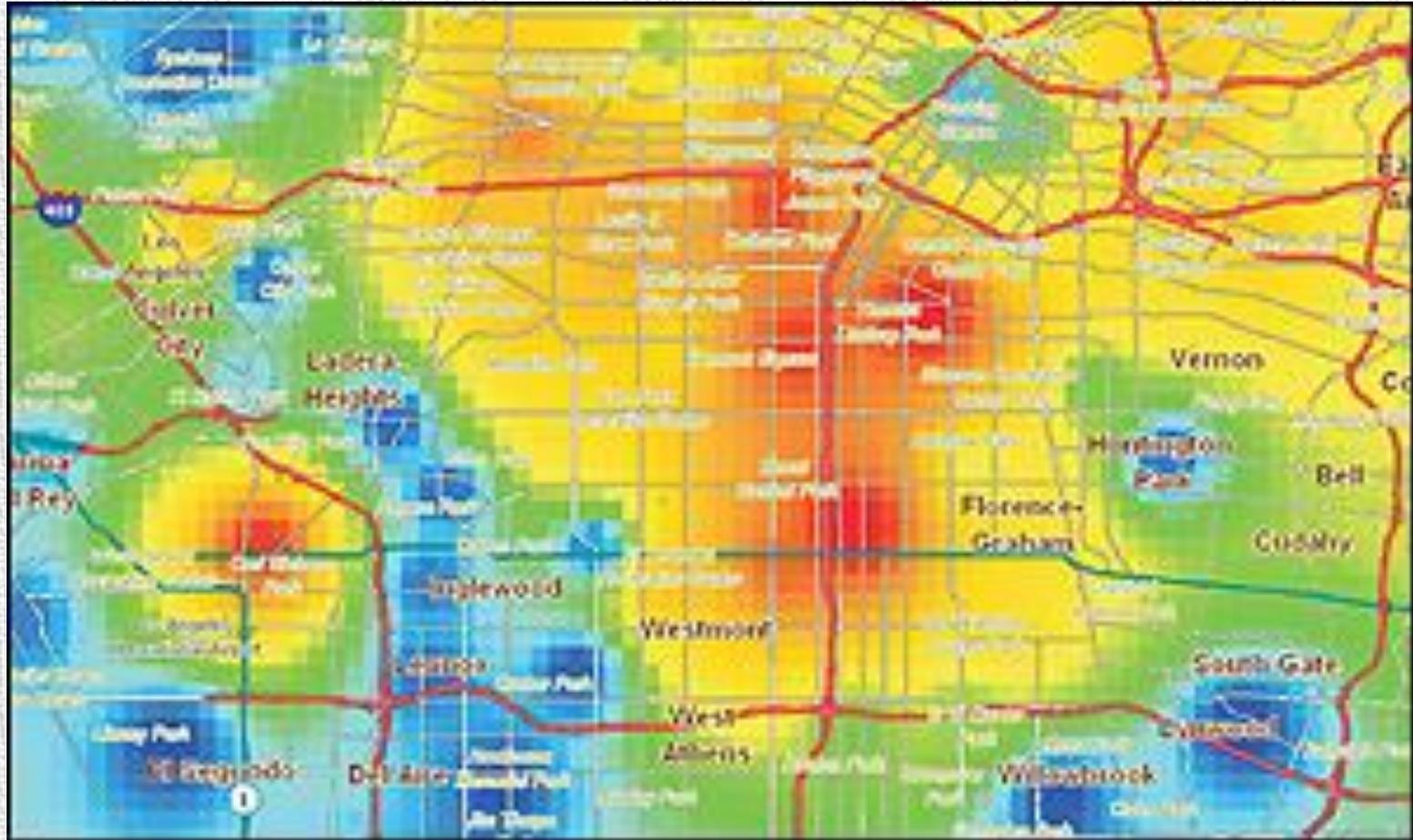
- ✓ 借助現代科技警察能夠在犯罪前正確預測，並輔助現代犯罪偵查工具包括電話定位監控、生物鑑別等設備，此皆是犯罪人個別身分的鑑別與監控，其受惠於個人資料蒐集與取得便利影響，讓犯罪偵查技術得以掌握犯罪的前行為階段，可謂大數據時代犯罪偵查的一大躍進。
- ✓ 美國Palantir數據預測公司開發相關應用程式，使得美國情治單位得用以對付恐怖攻擊
- ✓ 費城賓州大學犯罪防治所所開發的數據演算推測系統，藉以預測特定對象可能成為謀殺案件的被害人；同時，亦開發「假釋委員審核假釋」評估軟體，用以評估個案假釋出獄後再犯機率。

犯罪防治領域的相關應用cont.

(二)、犯罪地點與型態的預測(GIS)

- ✓國內警政署利用現有警政治安資料庫所含各項資料，透過GIS圖資應用與視訊影像分析技術，從現有資料加以萃取應用，進而提升案件偵防能量
- ✓自2008年起警政署與各縣市警察局陸續規劃「犯罪地圖資訊管理系統」，結合犯罪資料庫與地理資訊系統，並以「強盜」、「搶奪」、「竊盜」三項案類作分析比較，製作犯罪地圖，以強化犯罪熱區勤務佈署及警力巡邏，提升犯罪防治成效。

犯罪防治領域的相關應用cont.



圖為洛杉磯汽車竊案強度分佈圖

(資料來源：<http://www.ipam.ucla.edu/programs/chs2007/>)

犯罪防治領域的相關應用cont.

(三)、監視錄影系統

- ✓ 監視系統因具備「事前嚇阻」、「事中監控」、「事後調閱」功能，逐漸成為打擊犯罪利器。除傳統監控功能，未來系統更可進行即時性的智慧分析，如人臉、車牌辨識等，並具備「通報設定功能」，當監控環境發生異狀，系統可自動撥號到指定號碼，如連接警衛室、110，讓監控力量即時介入。
- ✓ 鑑此，2011年新北市警局宣布建構「科技防衛成」啟動「情資整合中心」，其包括整合勤務指揮、專家資訊、智慧影像辨識、交通執法、犯罪資料庫、治安人口影像監控系統、身分證相片影像、刑事犯罪資料影像、查贓系統等。

犯罪防治領域的相關應用cont.



圖11 新北市設置「數位式遠端路口監錄系統」
(資料來源：2011.08.25自由時報)。

大數據發展潛藏危機



大數據發展潛藏危機cont.

(一)、傳統專家權威的勢微

- ✓ 大數據作為「統計分析改變」資料巨量
- ✓ 蓋德藍雲家析與
藍芽雲端醫家析與
可透過上傳到三十多
護分
- ✓ 因此
監視病功
對於傳統專家權威之依賴必然降低



可透過
上傳到
三十多
護分
隨時
測疾
民眾

大數據發展潛藏危機cont.

(二)、大數據人才培育缺口

- ✓ 大數據發展必須仰賴一群有技術、懂管理、有大數據應用經驗的專業團隊，欠缺人才的環境將阻礙市場發展。
- ✓ 根據Gartne 預測，至2015年全球將新增440萬個與大數據相關職務，且會有25%的組織設立首席資料官職位元。其需要的是複合型人才，能夠對數學、統計學、資料分析、機器學習和自然語言處理等多方面知識綜合掌控
- ✓ 以對岸為例，2015年IBM將推出「IBMU-100」合作計畫，預計於2015年底與40所大學完成合作，**每年培養4萬名數據專才**，反觀國內相關基礎教育仍有很大的努力空間。

大數據發展潛藏危機cont.

(三)、個人隱私權侵害

- ✓ 大數據年代的個資保護已從「蒐集階段」轉移「資料利用階段」。原先「對消費者告知」的防線，轉趨「協力廠商參與」與「去識別化」為主軸。
- ✓ 現行個資法保護機制注重「告知同意」，使擁有雲端運算技術的電信業者，憑藉著資金、市佔率與資料規模等優勢，讓告知同意的機制備受挑戰。
- ✓ 而且，大數據時代即便去識別化，仍常殘留足以重新連結的訊息，資料可能被重新辨識，因此不宜直接排除個資法適用。

大數據發展潛藏危機cont.

(四)、網路犯罪問題

- ✓ 線上數據愈來愈多，駭客犯罪動機較以往強烈且一些知名網站密碼洩漏、系統漏洞導致使用者資料被盜，再加上近來個人敏感資訊洩漏事件頻傳，皆顯示在大數據時代建構網路安全的重要性。
- ✓ 另外，隨著大數據不斷的增加，對於數據儲存的物理安全性要求會愈來愈高，從而數據多副本與容災機制理應提高，但是目前許多傳統企業的資料環境安全仍令人憂心

大數據發展潛藏危機cont.

(五)、國家安全危害

- ✓ 各國情報機構以國家安全為名，侵害整體隱私的現象存在多年。縱使防護網絡如何固若金湯，專業駭客仍能迎刃而解，讓資安門戶洞開。

Ex:美國「稜鏡計畫」、中共間諜入侵

- ✓ 因此，隨著大數據快速發展，更讓國家安全承受空前挑戰，惟回到國家資安問題，追求公共資料安全保護的最高利益，仍是未來大數據發展的嚴肅課題。

結論

- 大數據的全球性革命已悄然成形，這是一場攸關國家、社會、民生以及科技的巨大變革。其逐步顛覆傳統思維，影響各種專業，遍及世界各地。基於犯罪防治研究立場，亦當無以迴避的以嶄新思維，面對大數據時代的來臨，並及時調整因應。
- 在犯罪防治領域的未來研究，我們更加體驗在數據爆發增長的年代，如何善用現存數據的重要性，如何強化各項軟硬體的數據截取與分析能力、提升研究人員在數據應用領域的相關訓練，提高大型資料的獲取、處理及分析能力，使得犯罪防治研究工作能儘快與新的科技接軌，讓研究更為精確、更具預見性及全面性，才能持續保持研究的前端領先。

中正大學

National Chung Cheng University



感謝聆聽，謝謝指教

Thanks for your listening.

