

巨量資料與統計分析

政治大學統計系余清祥

2020年12月24日

第十五週：人性與大數據

<http://csyue.nccu.edu.tw>

大數據衍生的特性



<http://h30507.www3.hp.com/t5/Journey-through-Enterprise-IT/Analyze-This-Big-Data-is-insurance-against-losing-a-competitive/ba-p/143577#.UgZmpLQVEqQ>

什麼是大數據？

□ 大數據2010年由IBM所提出，涵蓋四個V：

→ 大量化(Volume)：至少TB及PB以上

→ 多樣化(Variety)：視頻、GIS等多樣性

→ 快速化(Velocity)：即時處理

→ 真實性(Veracity)：資料品質（2014年提出）

註：加入資料分享與傳遞(Visible)後成為「大、快、雜、疑、轉」；也有人加入價值(Value)成為新的5V。

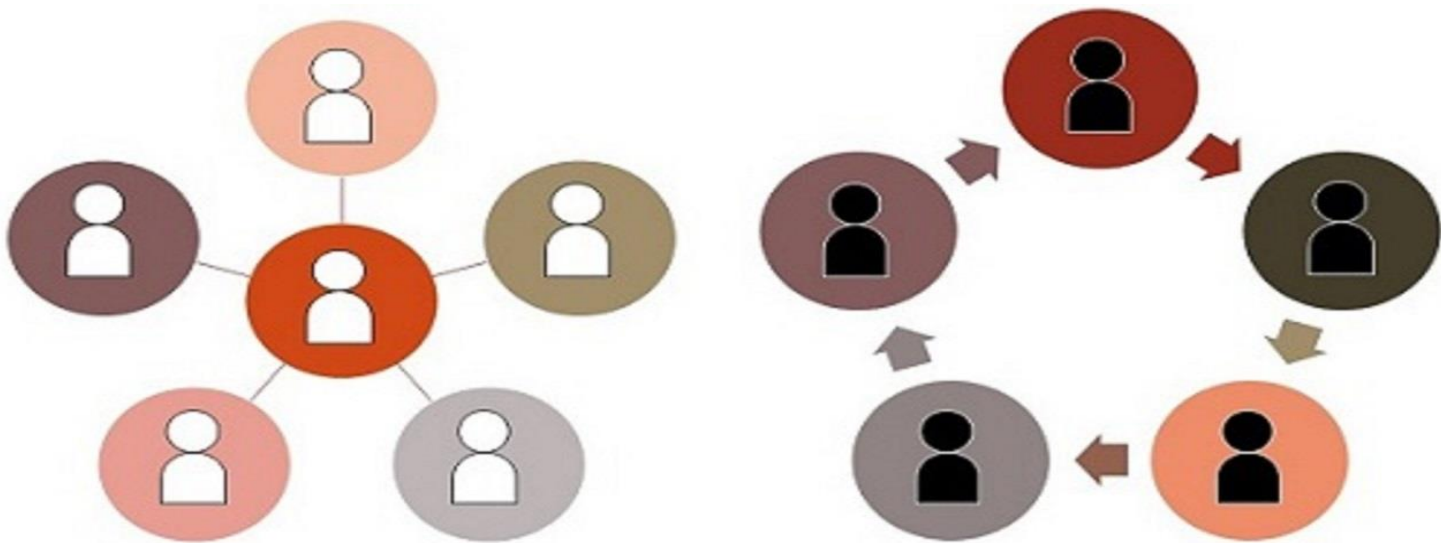


資料分享的副作用

- Facebook社群軟體盛行也帶來負面效應，除了個人資料安全疑慮，害怕與人接觸的社交恐懼症 (sociophobia)日趨嚴重。
 - 擔心自己遭受社會遺忘或淘汰；
「滿紙荒唐言，一把辛酸淚！都云作者痴，誰解其中味？」
 - 人際關係的疏離（虛擬vs.實際）；
「假作真時真亦假，無為有處有還無」
 - 私刑（人肉搜索）、扮演法官。

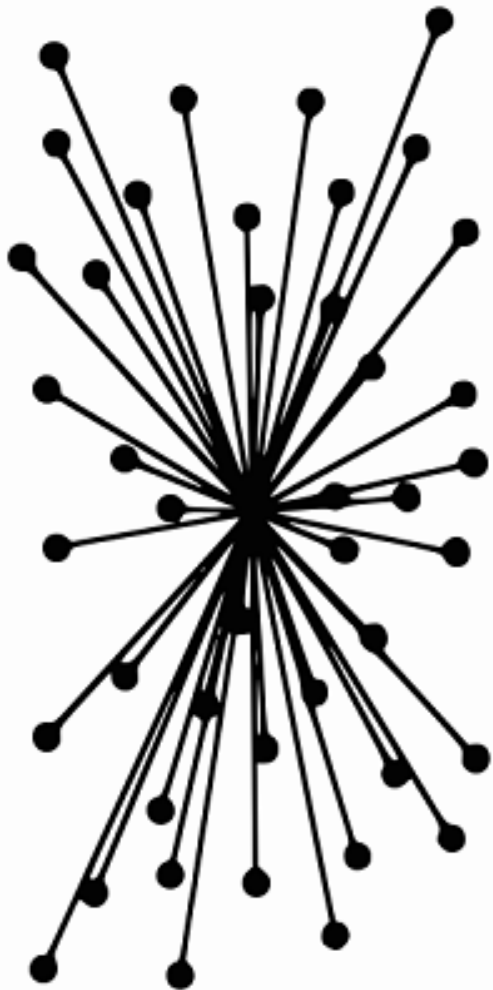
另一個特性：去中心化

- ❑ 去中心化 (Decentralization) 是一種現象或結構，出現在眾多用戶或眾多節點的系統，每個用戶都可連接並影響其他節點。

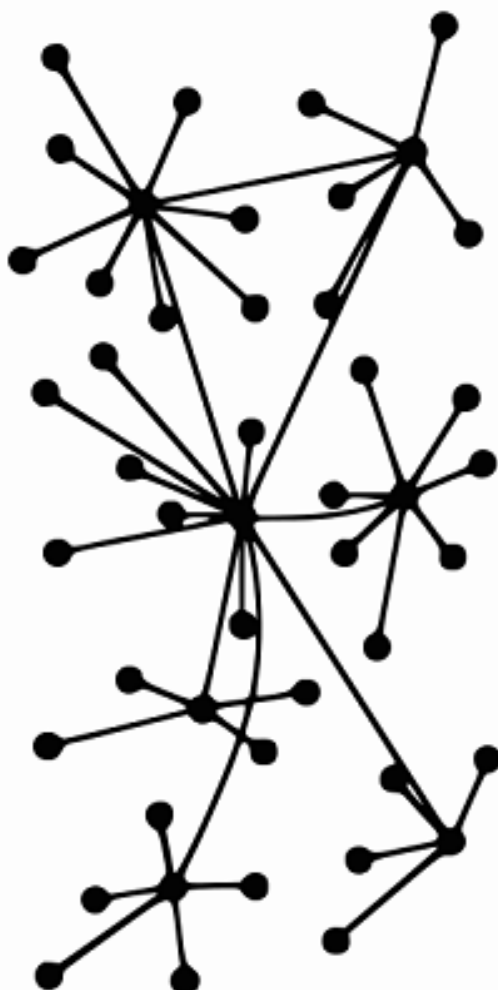


CENTRALIZATION VS DECENTRALIZATION

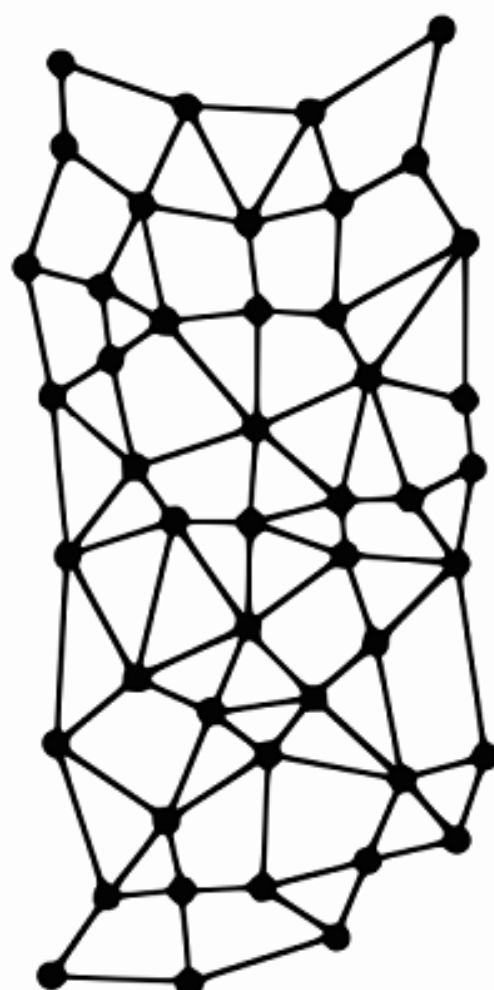
如何測量「去中心化」？



Centralized



Decentralized



Distributed

金融交易的服務流程



金融交易是「中心化」的典範

■ 關於金融交易的幾個問題：

→ 誰賦予貨幣（如：新台幣1000元）價值？

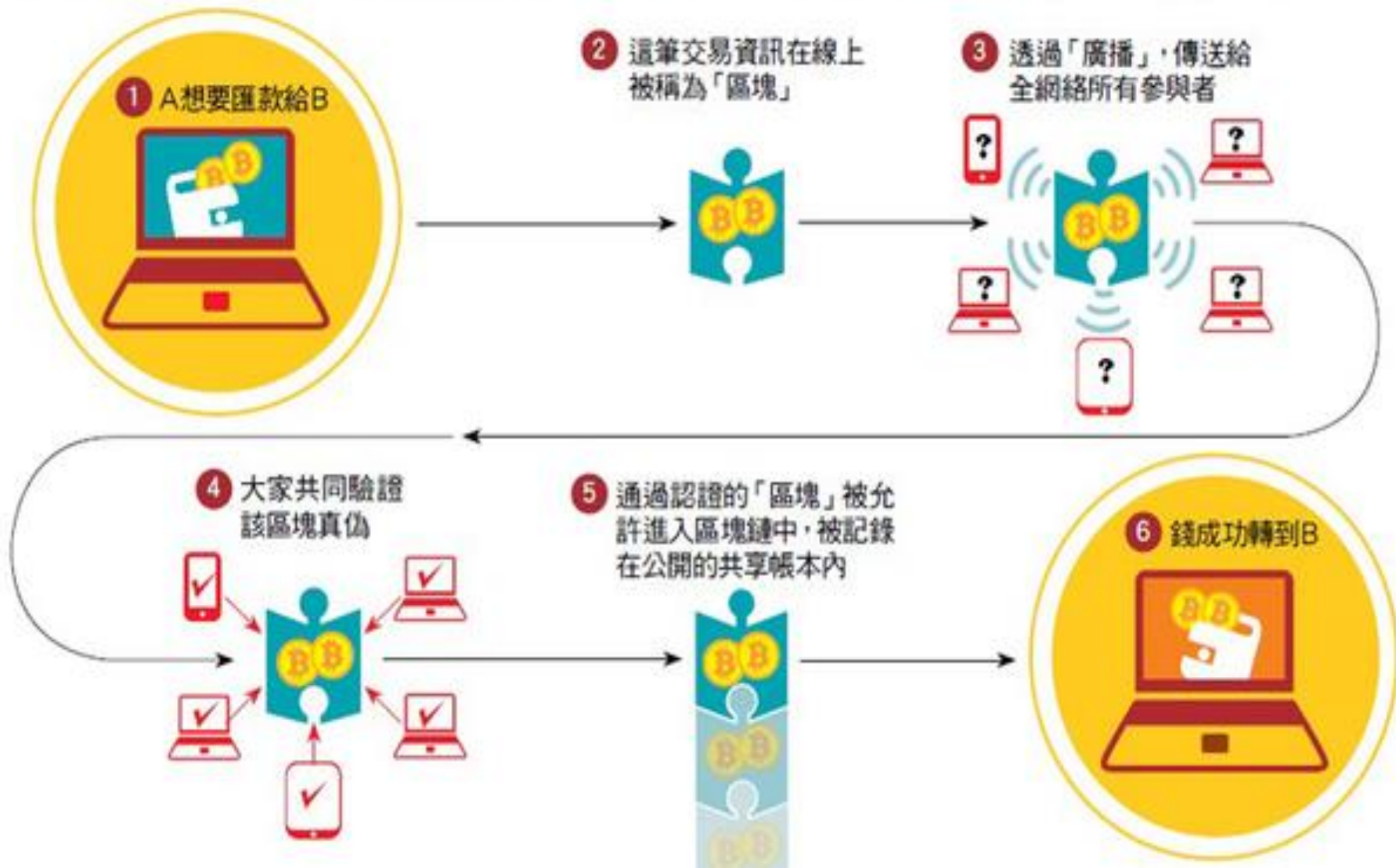
→ 貨幣價值由誰保障？

單筆付款流程 100100



區塊鏈這樣運作——以金融轉帳流程為例

資料來源：高盛、微拓、金融時報



虛擬貨幣的挑戰

- 貨幣有金錢價值，通常具備以下功能：
 - 交易工具；
 - 記帳單位；
 - 儲藏價值。
 - 根據這種標準，以下哪些可做為貨幣？
 - 信用卡（或現金卡）；
 - 悠遊卡、支付寶等非現金交易工具；
 - 比特幣（或以太幣等虛擬貨幣）。
-

去中心化與金融保險的發展

- 金融科技(Fintech)與保險科技(Insurtech)
 - 金融科技指技術帶來的金融創新，它能創造新的模式、業務、流程與產品，既可以包括前端產業也包含後臺技術。例如：互聯網和移動支付、網路信貸、區塊鏈。
 - 保險科技是科技進步帶來的保險創新。無論是產品、銷售通路、核保、理賠、後台作業與客服等傳統價值鏈，都將被保險科技帶來的創新徹底顛覆。
-

傳統的製造、行銷、消費通路



Q：生產及消費者的額外負擔？
→ 效率市場？



大數據代表一切？

■ 無限母體 vs. 有限母體

→ 再多的資料還是樣本！ 如Google Flu Trend比美國疾管局(Center of Disease Control; CDC)更早偵測流感盛行（消費者行為）。

→ 族群及關鍵詞、從眾效應(Bandwagon effect)

■ 資料品質仍是主要關鍵！

→ 人工智慧（AI）需仰賴好的訓練資料，尋找背後的規則及趨勢（如：alphago）。

[Google.org home](#)

[Dengue Trends](#)

Flu Trends

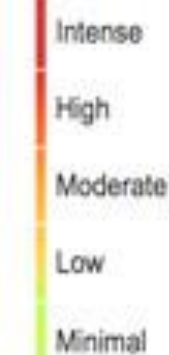
Home

Select country/region ▾

[How does this work?](#)

[FAQ](#)

Flu activity



Explore flu trends around the world

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)



從眾（或一窩蜂）效應的典範「寶可夢現象」



自由時報

第4章 實數

4.1 平方根

2. $(-\frac{1}{2})$ 的算術平方根是 _____

3. 下列說法中正確的是 _____

A. 0 的平方根是 0

C. -1 的平方根是 -1

B. 1 的平方根是 1

D. $(-1)^2$ 的平方根是 -1

課堂演練

4. $\sqrt{1000} = 10\sqrt{10}$, $-\sqrt{144} = -12$; $= \sqrt{\frac{5}{121}} = \pm \frac{5}{11}$

5. 用計算器求下列各式的值。

$\pm \sqrt{7.41} =$ _____

$\sqrt{11.7} =$ _____

6. 平方根和算術平方根相等的數是 _____

A. 1

B. -1

C. 0

D. _____

7. 若正數 a 的算術平方根是 \sqrt{a} , 那麼 a 的平方根是 _____

學習及考試方式?

「作業神器」APP讓學生按鍵一按便能獲得解答

大數據與人性



<http://cdn.marketingtechblog.com/wp-content/uploads/2013/05/Screen-Shot-2013-05-28-at-11.22.05-AM.png>

未來發展的幾個考量因素

- 未來各項產業的發展會與大數據愈來愈有
關連，但仍須注意以下事項：
- 資料的取得與更新：公司內部、同業合作
（財產權？）、政府及開放資料；
- 資料的維護與使用：安全性（雲端？）、
IRB（個資法？）、個體vs.大眾利益；
- 資料的分析與整合：R&D、產學合作、
Bayesian Credibility。

人體試驗委員會 (IRB)

- 人體試驗委員會

IRB : Institutional Review Board

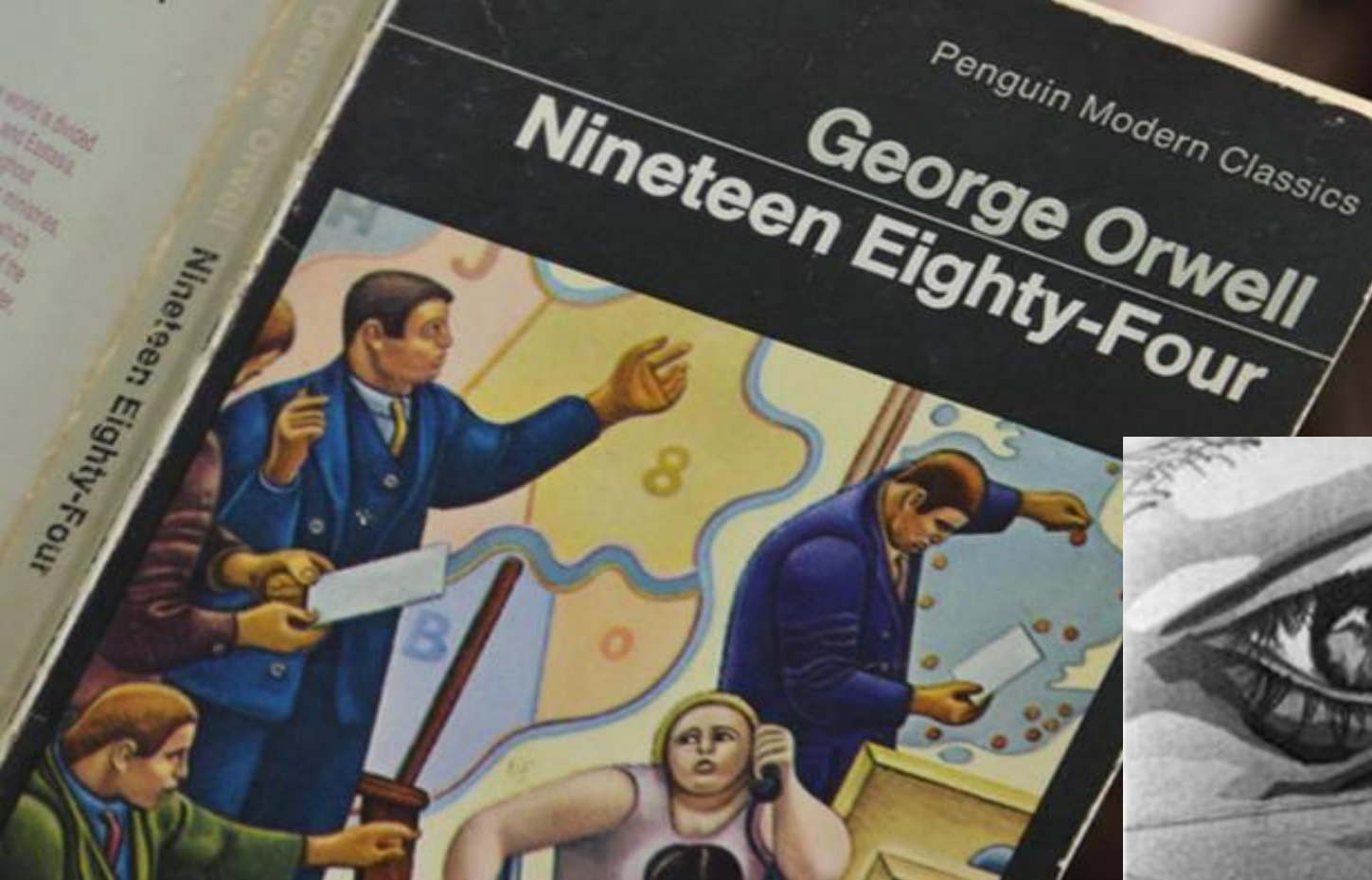
- 研究倫理委員會

REC : Research Ethics Committee

- 倫理審查委員會 - 人體研究法

赫爾辛基宣言 (Declaration of Helsinki)

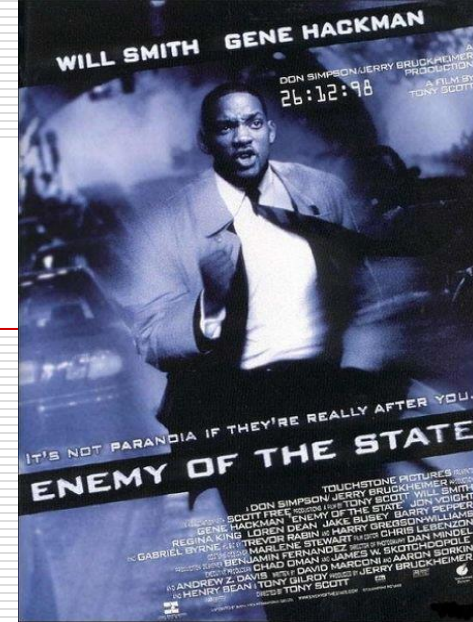
- 國際醫學協會針對人體試驗，在1964年提出了赫爾辛基宣言。
- 倫理相關規定類似紐倫堡公約
 - 研究計畫的設計與執行需經獨立的委員會審查
 - 著名雜誌的編輯開始要求研究需依赫爾辛基宣言，結果發表需有審查委員會同意函



「老大哥在看著你」
(Big Brother is Watching You!)

大數據也有黑暗面

- 資訊蒐集對隱私造成威脅（類似歐威爾的小說「1984」）
 - 避免迷信因果關係而做出「預防性的懲罰」（例如：「關鍵報告」）
 - 迷戀資料分析而濫用，形成「資料獨裁」或「大數據的傲慢與偏見」。
- 「考試成績並不等於真正的學識」！





資料獨裁 vs. 資料驅動

To be or not to be, that is the question...

大數據研究的建議

- 統計學家（&資料科學家）的專業技能
 - 與人溝通、與資料溝通(data sense)、與專業溝通、與電腦（機器）溝通。
 - 增加專業領域間的溝通機會（產學合作？）
 - 如何結合人與機器？
 - 多樣性（物種、知識）是重要資產，不少發現源自其他角度的思維。
- （馬爾薩斯《人口論》與達爾文、馬克斯）

大數據的未來發展？

- 生活會因為大數據而改變，但顛覆既有一切的可能性不高，人性依然是關鍵！
 - 例如：自動駕駛可顯著降低傷亡，是否規定部分/全部的自動駕駛？
 - 自動駕駛衍生傷亡，責任歸屬？
- 資料的取得與使用、財產權的考量？
 - 監視器及公眾資料的使用規範，有限度鬆綁個資法，公眾大數據的財產權（捐血？）。

使用大數據的建議



<http://solutionsreview.com/data-integration/hard-vs-soft-data-whats-the-difference/>

■ 監管大數據的三個策略：

→ 個資保護由「個人同意制」，

轉成資料使用者的「使用責任制」；

→ 運用大數據時，需尊重個人的「能動性」

（Human Agency; 政府對人民的判斷是基於實際的

行動，而非大數據分析的預測）；

→ 培養「演算學家」以處理大數據的審計師。

IBM對大數據的建議

- IBM認為大數據不只是一項挑戰，更是絕佳機會洞悉新興的資料類型、使企業運作更加靈敏、並為過往所無法企及的問題提供解答，為世界開啟一扇大門。
- 成功的關鍵因素，至少包括：
 - 資料的可獲得性、取得成本與使用技能；
 - 以大數據達成設定目標所需要的時間；
 - 能夠克服障礙而快速有效的處理大數據；
 - 提出有效解決方案以達成客戶要求。

可行的大數據發展方向？！

- 除了知識萃取與傳播外，需求與商機也是大數據必須考量的因素，其中異業結盟或許是可行方向之一（「從核心出發」？）。
- Google以搜尋引擎起家，之後發展3D及虛擬地圖、電子化圖書館，以及其他的大數據應用（例如：H5N1流感、去中央化）。
- Amazon等公司可挾其在物流業的資源，發展其他可能的經營方向或組合？