

## 成功大學人類研究倫理審查委員會 **南區研究倫理聯盟課程**

- 主辦:南區研究倫理聯盟、成功大學學術誠信推動辦公室
   成功大學人文社會科學中心人類研究倫理治理架構行政辦公室、
- 主題:研究倫理、學術倫理
- 時間: 108年1月24日(星期四)下午12:30-16:30
- 地點:國立成功大學光復校區國際會議廳第一演講室(台南市東區大學路一號)
- 流程: 13:00-13:10 開場致詞

13:10-14:40 大數據研究倫理議題 14:40-14:50 休息時間 14:50-16:20 統計分析的倫理議題 16:20-16:30 議題討論與Q&A



成功大學人類研究倫理審查委員會

時間	內 容	講者
12:30-13:00	報到&現場諮詢	
13:00-13:10	開場致詞	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
13:10-14:40	大數據研究倫理議題	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
14:40-14:50	休息時間	
14:50-16:20	統計分析的倫理議題	陸偉明 成功大學人類研究倫理審查委員會書審專家 暨教育研究所特聘教授
16:20-16:30	議題討論與Q&A	
16:30~	自由交流賦歸	



成功大學人類研究倫理審查委員會

時間	內 容	講者
12:30-13:00	報到&現場諮詢	
13:00-13:10	開場致詞	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
13:10-14:40	大數據研究倫理議題	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
14:40-14:50	休息時間	
14:50-16:20	統計分析的倫理議題	陸偉明 成功大學人類研究倫理審查委員會書審專家 暨教育研究所特聘教授
16:20-16:30	議題討論與Q&A	
16:30~	自由交流賦歸	

https://rec.chass.ncku.edu.tw/organization/southern\_consortium/universities

## 簽署大學校院及機構





### NCKU REC

![](_page_5_Figure_0.jpeg)

![](_page_6_Picture_0.jpeg)

時間	內 容	講者	
12:30-13:00		報到&現場諮詢	
13:00-13:10	開場致詞	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授	
13:10-14:40	大數據研究倫理議題	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授	
14:40-14:50	休息時間		
14:50-16:20	統計分析的倫理議題	陸偉明 成功大學人類研究倫理審查委員會書審專家 暨教育研究所特聘教授	
16:20-16:30	議題討論與Q&A		
16:30~	自由交流賦歸		

![](_page_7_Picture_0.jpeg)

### 大專院校研究倫理審查組織查核辦公室

University-based HRPP Certification and Auditing Office

		時 程	講題	講者
♠>最新消息>		09:40-10:00 報到		
類別・		10:00-11:00	什麼是人工智慧: AI的過去、現在與未來	王大為教授 中央研究院資訊科學研究所
類別	· 標題		(演講50分鐘、討論10分鐘)	1 2 4 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4
查核辦公 室	研究倫理系列演講:人工智慧(AI)與研究倫理	人工智	人工智慧:	邱文聰教授 中央研究院法律學研究所     主持人: 陳東升教授 國立臺灣大學社會學系
查核辦公 室	107年度研究倫理系列演講:原住民族研究與倫理專題	11 : 00-12 : 00	法令規範與研究倫理 (演講50分鐘、討論10分鐘)	
查核辦公 室	107年度申請教育部人體研究倫理審查委員會查核作業說明	12:00-12:30	綜合討論	
查核辦公	107年度教育部大專校院人體研究倫理審查委員會常規暨不定期追蹤查核說明會			
Ŧ		12:30 賦歸		歸
查核辦公 室	106年度教育部大專校院人體研究倫理審查委員會常規暨不定期追蹤查核說明會	(全程参加者,核發研究倫理教育訓練時數2.5小時) 投影片下載:		
1.20181212AI的過去、現在與未來(王大為教授)         1.20181212AI的過去、現在與未來(王大為教授)         2.法令規範與研究倫理:簡報檔案是否公開徽詢講者同意中				

	日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日	15 10.30 著		<image/> <image/>
13:40-14:0	0	報	到	
14:00-14:1	0 主持人致詞	范建得 教授 國立清華大學	显研究倫理專案辦公室主任	Ŧ

### Big Data (巨量資料,大數據)

...to refer to data sets that are too large or complex for traditional dataprocessing application software to adequately deal with. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. Big data was originally associated with three key concepts: volume, variety, and velocity. Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value...

By Wikipedia: <a href="https://en.wikipedia.org/wiki/Big\_data">https://en.wikipedia.org/wiki/Big\_data</a>

![](_page_10_Figure_0.jpeg)

Fernanda B. Viégas [CC BY 2.0 (https://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons

2009

1 KB	1024 B	B = byte
1 MB	1024 KB	KB = Kilobyte
1 GB	1024 MB	MB = Megabyte
1 TB	1024 GB	TB = Terabyte
1 PB	1024 TB	PB = Petabyte
1 EB	1024 PB	EB = Exabyte
1 ZB	1024 EB	ZB = Zettabyte
1 YB	1024 ZB	YB = Yottabyte

Proudly powered by WordPress | Theme: Yocto by Humble Themes. https://porkostournaments.info/relationship/bytes-mb-kb-gb-relationship.php

![](_page_12_Figure_0.jpeg)

By Myworkforwiki - Own work, CC BY-SA 3.0, <u>https://commons.wikimedia.org/w/index.php?curid=29452425</u>

"If one were able to store 175ZB onto BluRay discs, then you'd have a stack of discs that can get you to the moon — 23 times," by **IDC**'s *David Reinsel*.

Figure 1 - Annual Size of the Global Datasphere

![](_page_13_Figure_2.jpeg)

Source: Data Age 2025, sponsored by **Seagate** with data from **IDC** Global DataSphere, Nov 2018 https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

## How big is the data? By 2020...

- Amazon sells 600 items per second.
- MasterCard processes 74 billion transactions per year.
- Commercial airlines make about 5,800 flights per day.
- Walmart processes one million customer transactions per hour.
- On average, people send about 500 million tweets per day.

By Syncsort: <a href="https://blog.syncsort.com/2017/07/big-data/how-big-is-big-data-definition-examples/">https://blog.syncsort.com/2017/07/big-data/how-big-is-big-data-definition-examples/</a>

## How big is the data? By 2020...

- 5,200 gigabyte of data on every person in the world.
- On average, each person who uses email receives 88 emails per day and send 34.
- The average U.S. customer uses 1.8 gigabytes of data per month on his or her cell phone plan.

## (NCC: Taiwanese 4G cell phone users $\rightarrow$ 14 G / month)

https://www.businessweekly.com.tw/article.aspx?id=21546&type=Blog

By Syncsort: <a href="https://blog.syncsort.com/2017/07/big-data/how-big-is-big-data-definition-examples/">https://blog.syncsort.com/2017/07/big-data/how-big-is-big-data-definition-examples/</a>

## More about Future World... Evidence from AI Experts

![](_page_16_Picture_1.jpeg)

https://www.stylist.co.uk/people/sophia-therobot-meaning-life-secret-happiness-exclusiveinterview-five-minute-philosopher/185248

Robot (Sofia by David Hanson)

- Self-driving car (Google's Waymo)
- Al in medicine
  - Diabetic retinopathy [IDx-DR]
  - Skin cancer screening
- When will AI exceed human performance?
  - Translate languages by 2024
  - Writing high-school essays 2026
  - Driving a truck 2027
  - Working in retail 2031
  - Writing a bestselling book 2049
  - Working as a surgeon 2053

![](_page_16_Picture_15.jpeg)

https://www.cbc.ca/news/canada/windsor/google-self driving-spinoff-factory-michigan-1.4989204

![](_page_16_Picture_17.jpeg)

http://technews.tw/2018/04/12/fda-approvesai-powered-diagnostic-for-the-first-time/

## NORMAN

World's first psychopath AI.

EXPLORE WHAT NORMAN SEES

### Google成功預測2009年美國H1N1的流行與傳散範圍

- •2009年在美國的H1N1爆發幾周前,Google成功預測了H1N1在全美的傳播範圍 具體到了州還有特定地區,判斷非常及時。
  •CDC疾控中心通常只能在流感爆發一兩周之後才可以做到。
  •真正第一次利用搜尋引擎大數據,對疾病控制進行預測。

#### 方法:

▶ Google發現搜尋流感相關主題的使用者數 量與實際出現流感症狀的人數有密切關聯。 ▶Google將查詢次數與傳統流感監控系統數據 進行比較,發現某些搜尋關鍵字在流感季節 特別熱門。

▶只要統計使用者搜尋這些關鍵字的次數,便 能預測全球各個國家及地區的流感疫情發展。

20042005 2006 Google

#### Detecting influenza epidemics using search engine query data

Jeremy Ginsberg<sup>1</sup>, Matthew H. Mohebbl<sup>1</sup>, Rajan S. Patel<sup>9</sup>, Lynnette Brammer<sup>2</sup>, Mark S. Smolinski' & Larry Brilliant'

Google Inc. "Centers for Disease Control and Prevention

### Google的研究結果刊登在《自然》期刊

Nature 457, 1012-1014 (19 February 2009) http://static.googleusercontent.com/media/research.google.com/zh-TW//archive/papers/detecting-influenza-epidemics.pdf

從搜尋字眼(如:頭痛、止咳、退燒、醫 生)之使用量趨勢,預測流感的蔓延,讓 官方快速掌控疫情 Google 流感趨勢疫情預測 ● 美國的資料

學首首系習

![](_page_18_Figure_13.jpeg)

Facebook如何玩轉大數據----它會知道你戀愛了

 ▶ 利用Facebook網站的統計數據,用戶的post文數量、 內容判斷用戶是否、何時擦出了愛的火花。
 ▶ Facebook可能還比某些情侶早一步察覺他們之間萌生 了愛意。

#### 方法

元智大學資管系 曾淑

用統計方法分析時間軸雙方互動的內容,記錄表達積極情緒的詞彙(如"愛"、"愉悅"、"高興"等)
 和表達消極情緒的詞彙(如"討厭"、"受傷"、"糟糕"等)的數量,計算二者在貼文中所占比例的差值。
 在成為戀人前的100天,兩人相互post文數會緩慢逐步增加。

 > 戀情正式開始的那天起post文數會越來越少。
 > 如果通過Facebook發展的關係能维持三個月以上, 那這段關係可能可維繫至少四年以上。
 > 分手可能性最大的月份是5月、6月和7月。

From : http://www.kankanews.com/ICkengine/archives/112086.shtml

![](_page_19_Figure_6.jpeg)

![](_page_19_Figure_7.jpeg)

### 社群網路於醫療領域

- ▶ 加州大學洛杉磯分校(UCLA)以 Twitter 的訊息量、地點,來追蹤 性病擴散率與毒品濫用的行為。
- 大蒐集了5億5000萬條「推特文」,使用演算法篩檢出含有「性」 (Sex)、「快感」(get high)的字眼,並記錄發布內容的地區,最後用統計模型觀測這些區域是否有 HIV 新病例通報。
- 結果發現兩者之間有很顯著的關係, 當某地區的推文呈現很高的「性指數」
   HIV 的新感染病例也高。

資料來源健康達人網 <u>http://technews.tw/2014/06/15/big-</u> data%EF%BC%9A%E6%9C%AA%E4%BE%86%E9%86%AB%E7%99%82% E7%9A%84%E7%B5%82%E6%A5%B5%E6%AD%A6%E5%99%A8/

元智大學資管系 曾淑芬

![](_page_20_Figure_5.jpeg)

### **Beneficial / Questionable Uses**

### ▶ 醫療

- 。疾病預測、個人化醫療 vs
- 建立個人健康分數、醫療及保險的可能歧視
- ▶ 教育: 中輟預測、個人適性教學 vs 篩選入學學生

### ▶ 檢警執法

- 犯罪防制 vs 探勘搜尋行為尋找嫌疑犯、關鍵報告
   400Heat Lists in Chicago
- ▶ 零售: 精準行銷 vs 追蹤顧客行動
- ▶ 都市計畫:
  - 改善交通、路平 vs 全面監控、僅有利某些區域

An over-prediction could cause panic, misallocation of limited supplies of vaccines or medical resources, and damaging stigmatization of people or communities who don't pose a risk.

ScienceDaily, 9 February 2015 http://www.sciencedaily.com/releases/2015/02/15020 9143440.htm

## **Ethical Issues in the Big Data Industry**

Kirsten E. Martin George Washington University June 2015 (14:2) MIS Quarterly Executive

![](_page_22_Figure_0.jpeg)

Figure 1: Example of Information Supply Chain Within the Big Data Industry

![](_page_23_Picture_0.jpeg)

### Scenario: at the XYZ bank

Tavani, H. T. (2013).

Lee, a junior executive at ABE Marketing Inc., has recently applied for an automobile loan at the XYZ Bank. To secure the loan, Lee agrees to complete the usual forms required by the bank for loan transactions. He indicates that he has been employed at the ABE Marketing Company for more than 3 years and that his current annual salary is \$240,000. He also indicates that he has \$30,000 in a separate savings account, a portion of which he intends to use as a down payment for a new BMW. On the loan form, Lee also indicates that he is currently repaying a \$15,000 personal loan used to finance a family vacation to Europe the previous year. While examining Lee's loan form, the bank runs several programs. As a result, the bank refuses Lee's loan request.

Academic Research Credit Decision Consumer Surveys Criminal detection Social Advertising

## Scenario: DoubleClick and Abacus Tavani, H. T. (2013).

DoubleClick planned to purchase Abacus Direct Corporation, a database company, in late 1999. Abacus's databases contained not only records of consumer's catalogue purchases but also actual names and telephone numbers that had been collected by Abacus primarily from offline transactions. With this acquisition, DoubleClick could merge records in the Abacus database with its own database, which consisted of information gained primarily from Internet cookies files. And with its newly merged data, DoubleClick would have an information mosaic about individuals that included not merely anonymous and indirect information (such as IP addresses and ISP-related information) but also direct personal information. The Web profiles in DoubleClick's original database, gathered via cookies, included data about which Web sites that users (who are identified and tracked via an IP address) visit, how long they visit a particular site, and so on. That information would be able to be compared to and combined with explicit personal information (gathered offline and stored in Abacus's databases), including names, addresses, and phone numbers.

Academic Research Credit Decision Consumer Surveys Criminal detection Social Advertising

## Scenario: Shopping online/offline Tavani, H. T. (2013).

One day, you decide to shop for groceries at SuperMart. If I happen to see you enter or leave SuperMart, or if we are both shopping in this store at the same time, I now have information that you shop at SuperMart. If I also happen to pass by you in one of the aisles at SuperMart, I can observe the contents of your shopping basket; I may notice that you are purchasing several bottles of wine but relatively little food.

You visit an online bookstore called Nile.com to locate a particular book that you are considering purchasing. Because you are visiting this bookstore via a computer or electronic device located in your own home, you cannot be observed by people in physical space nor can you be seen by other customers on the Nile.com Web site. If you have visited this site before and have clicked on items that interested you, Nile can find a record of these items.

## Three Cybertechology-related Techniques that Threaten Privacy

- data-gathering techniques used to collect and record (monitoring, recording, and tracking techniques)personal information, often without the knowledge and consent of users. e.g., Cookies, RFID
- data-exchanging techniques (merging and matching) used to transfer and exchange personal data across and between computer databases, typically without the knowledge and consent of users.
- data-mining techniques used to search for patterns implicit in large databases in order to generate consumer profiles based on behavioral patterns discovered in certain groups.

Tavani, H. T. (2013). Ethics and technology: Controversies, questions, and strategies for ethical computing (4th ed.). Hoboken, N.J.: Wiley.

### How Companies Learn Your Secrets

![](_page_27_Picture_1.jpeg)

http://www.nytimes.com/2012/02/19/magazine/shoppinghabits.html?pagewanted=1&\_r=3&hp

元智大學資管系曾淑芬

Academic Research Credit Decision Consumer Surveys Criminal detection Social Advertising

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

![](_page_27_Picture_6.jpeg)

TARGET

**Kashmir Hill** Forbes Staff Welcome to The Not-So Private Parts where technology & privacy collide

http://www.forbes.com/sites/kashmirhill/2012/02/16/how-targetfigured-out-a-teen-girl-was-pregnant-before-her-father-did/

Academic Research Credit Decision Consumer Surveys Criminal detection Social Advertising

### Scenario: Public records available online Tavani, H. T. (2013).

In the late 1990s, information from the state of Oregon's Department of Motor Vehicle became accessible online. An independent computer consultant used the means available to any private citizen to purchase data from that state's department, which was already available offline to anyone willing to pay a small fee. Once he purchased the information and converted it to electronic format, the consultant set up a Web site where any Internet user could, for a small fee, enter an Oregon license plate number and obtain the name and address of the owner of the registered vehicle.

The city of Pleasantville has recently made all of its public records, including real estate records, available online; with a networked computer or electronic device, one can simply enter the address of any house in Pleasantville and retrieve the current tax assessment for the house, the price paid by the most recent owner, and a description of the physical layout of the house, including the location of doors and windows. The same information had previously been available as public records, stored in physical file cabinets at City Hall.

![](_page_29_Picture_0.jpeg)

## **Scenario: Before 911** Tavani, H. T. (2013).

At Super Bowl XXXV in January 2001, a facial recognition technology scanned the faces of individuals entering the stadium. The digitized facial images were then instantly matched against images in a centralized database of suspected criminals and terrorists. Those who attended the sporting advent were not told that their faces had been scanned.

## Information Ethics (Mason, 1986)

Privacy隱私: what information about one's self or one's associations must a person reveal to others, under what conditions and with what safeguards?

Access存取: What information does a person or an organization have a right or a privilege to obtain?

Property財產(IP = Intellectual Property): Who owns information? Who owns the channels? How should access to this scarce resource be allocated?

Accuracy精確: Who is responsible for the authenticity, fidelity and accuracy of information?

Mason, R. O. (1986). Four Ethical Issues of the Information Age. MIS Quarterly, 10, 5-12.

![](_page_31_Picture_0.jpeg)

![](_page_32_Picture_0.jpeg)

![](_page_32_Picture_1.jpeg)

France fines Google nearly \$57 million for first major violation of new European privacy regime

2019.01.21

By The Washington Post: <u>https://www.msn.com/en-us/news/technology/france-fines-google-</u>nearly-dollar57-million-for-first-major-violation-of-new-european-privacy-regime/ar-BBSxAkF

![](_page_33_Picture_0.jpeg)

"Google failed to fully disclose to users how their personal information is collected and what happens to it." mentioned by The French Data Protection Authority (the CNIL)

EU's General Data Protection Regulation (GDPR), legislated in 2016, and then implemented in 2018

"Google also did not properly obtain users' consent for the purpose of showing them personalized ads" the watchdog agency said.

![](_page_34_Picture_0.jpeg)

"People expect high standards of transparency and control from us. We're deeply committed to meeting those expectations and the consent requirements of the GDPR." Google said.

"Google is not the only one doing this...This is significant for Google as a company but also for other actors." said Estelle Massé, a data protection expert at the advocacy group Access Now.

"...why the Federal Trade Commission (FTC) failed to act against the tech firms over these many years," said Marc Rotenberg, the executive director of the Electronic Privacy Information Center.

### **BEST SEARCH ENGINES IN THE WORLD**

![](_page_35_Picture_1.jpeg)

## Yandex Aol. (Ask).com

![](_page_35_Picture_3.jpeg)

https://www.liveenhanced.com/worlds-top-15-search-engines-list-updated-list-of-2018/

https://www.forbes.com/sites/kalevleetaru/2018/08/13/social-science-one-and-how-top-journals-view-the-ethics-of-facebook-data-research/#417ee5ef5e38

Academic Research Credit Decision Consumer Surveys Criminal detection Social Advertising

## Facebook Beacon controversies

Tavani, H. T. (2013).

Facebook Beacon would let Facebook friends share information about what they do online, including the purchases they make. Beacon also allowed affiliate Web sites to send stories about a user's online activities to Facebook, which were then displayed to that user's "friends" in the form of news feeds and Social Ads. When Facebook introduced Beacon, it stated that it would not share any personally identifiable information in the Social Ads, and it claimed that users would only see those ads to the extent that they were willing to share that information with others.

- Facebook facing fines related to the Cambridge-Analytica scandal<sup>3</sup>
  - Massive amount of data accessed without consent for political use

Romm, T. & Dwoskin, E. (2018, January 18). U.S. regulators have met to discuss imposing a record-setting fine against Facebook for privacy violation. *The Washington Post*. Retrieved from <a href="https://www.washingtonpost.com/technology/2019/01/18/us-regulators-have-met-discuss-imposing-record-setting-fine-against-facebook-some-its-privacy-violations/?utm\_term=.d8237ec15765">https://www.washingtonpost.com/technology/2019/01/18/us-regulators-have-met-discuss-imposing-record-setting-fine-against-facebook-some-its-privacy-violations/?utm\_term=.d8237ec15765</a>

### **Research Review at Facebook**

![](_page_38_Figure_1.jpeg)

Note: Research involving any issues relating to privacy must also go through privacy review.

https://fpf.org/2016/06/14/14717/

![](_page_39_Picture_0.jpeg)

- Data sharing has benefits but how to obtain the consent?
- Data Ownership: Who owns the data? What does ownership mean? Is it even an applicable concept?
- Data minimization: Collect only what is useful and valuable?
- Big data and democracy:
  - Big nudging: Government using big data to steer the citizens do what it thinks is right
  - "Filter bubble" or "echo chamber effect" customization of data search ends up with repetition of one's opinion, leading to social polarization.
  - Centralized, top-down control of data vs. collective intelligence (citizen science, crowdsourcing and online discussion platforms)

## • Personal data: The "right to be forgotten" (information deleted upon request)

Hand DJ (2018) Aspects of data ethics in a changing world: where are we now? *Big Data* 6:3, 176–190.

Helbing D. et al. (2019) Will Democracy Survive Big Data and Artificial Intelligence?. In: Helbing D. (eds) Towards Digital Enlightenment. Springer, Cham

## Right to be forgotten (被遺忘權)

- It is a human right.
- It is distinct from the right to privacy.
- The right to silence on past events in life that are no longer occurring.
- Intended to secure potentially damaging, private information about individuals.

成功大學人類研究倫理審查委員會

- An individual has certain data (publicly known at a certain time) deleted so that third persons can no longer trace them.
- An individuals has information, videos or photographs about themselves deleted from certain internet records so that they cannot be found by search engines.
- There is no global framework to allow individuals control over their online image.

### https://en.wikipedia.org/wiki/Right\_to\_be\_forgotten

- Weber, Rolf H. "The right to be forgotten." More than a Pandora's Box 2 (2011)
- Pino, G. (2000). "The right to personal identity in Italian private law: Constitutional interpretation and judge-made rights".
   In: M. Van Hoecke; F. Ost (eds.). The harmonization of private law in Europe (pp. 225-237). Oxford: Hart Publishing. p. 237.

![](_page_41_Picture_0.jpeg)

![](_page_42_Picture_0.jpeg)

![](_page_43_Figure_0.jpeg)

TRANSPARENCY

PRIVACY BY

DESIGN

DIGNITY

**BIG DATA** 

**USER CONTROL** 

ACCOUNTABILITY

FREEDOM

![](_page_43_Figure_1.jpeg)

## **Towards a new digital ethics**

### Data, dignity and technology

https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11\_Data\_Ethics\_EN.pdf

EUROPEAN DATA PROTECTION SUPERVISOR

PRIVACY

![](_page_44_Picture_0.jpeg)

- 単調切入

2 禁止吸药

重研究管計

lied Health Research Data Integration Service Center

https://www.nytimes.com/2018/02/12/business/computer-science-ethics-courses.html Singer, N. (2018,February 12). Tech's ethical 'dark side': Harvard, Stanford and others want to address it. *The New York Times*. Retrieved from

• 學術界浮濫運用健保資料庫? 專家籲應嚴格把關 https://udn.com/news/story/7266/3015050

### Think Again: Big Data

海量資料萬歲?請三思!

2013/05/25 | 專欄 電腦資訊 | 標籤: Big data 大數據 數據 海量資料

### 譯者: Leonard Chien

http://pansci.asia/archives/42114

Why the rise of machines isn't all it's cracked up to be.

BY KATE CRAWFORD MAY 10, 2013 📑 🕤 🛇 🛞 💮 🗠

### Myths

- 「只要資料足夠,數字自會說話」
   Not a chance (量≠質)
- 「海量資料可提高城市智慧及效能」
  - Up to a point (Google Flu trends預測失靈; Boston street bump app排除少用者)
- ▶ 「海量資料對各個社會族群一視同仁」
  - Hardly (個人化 vs 污名化)英國劍橋大學海量研究58000人Facebook網站的按讚紀錄, 預測用戶敏感性個人資訊,如性向(準確率88%)、族群(95%)、政治傾向(85%)。
- 「海量資料屬匿名,不會侵犯隱私」
  - Flat-out wrong (研究歐洲150萬手機用戶資料發現四個參考點即可辨識95%的民眾)
- 「海量資料是科學的未來」
  - Partly true, but it has some growing up to do
  - 資料永遠不可能中立,也很難匿名,但我們可運用各項專業領域,以察覺種種偏見、落 差與假設,進而面對有關隱私及公平性的新挑戰

## What Big Data Needs: A Code of Ethical Practices

MIT Technology Review by Jeffrey Rayport 2011

- Four key principles that companies should follow if they hope to analyze customers' data without alienating them.
- ▶ Clarity on Practices: (明確目的)
  - When data is being collected, let users know about it in real time.
- ▶ Simplicity of Settings: (簡易設定)
  - Give users a chance to figure out for themselves what level of privacy they really want.

### ▶ Privacy by Design: (隱私設計)

• Organization incorporate privacy protections into everything they do.

## ▶ Exchange of Value: (價值交換)

 Radical transparency could make it easier for digital businesses to show customers what they will get in exchange for sharing their personal information.

## IBM Ethics for big data and analytics

### Mandy Chessell 2014

http://www.ibmbigdatahub.com/sites/default/files/whitepapers\_reports\_file/TCG%20Study%20Report%20-%20Ethics%20for%20BD%26A.pdf

## Ethical Awareness Framework

- Context For what purpose is the data being used?
- Consent & Choice What are the choices given to an affected party?
- Reasonable Is the depth and breadth of the data used reasonable for the application it is used for?
- Substantiated Are the sources of data used appropriate, authoritative, complete and timely for the application?
- Owned Who are responsible for the protection of resulting insight and the obligation to act?
- **Fair** How equitable are the results of the application to all parties?
- Considered-What are the consequences of the data collection and analysis?
- Access What access to data is given to the data subject?
- Accountable How are mistakes and unintended consequences detected and repaired?

## 美國政府2012年「消費者隱私保護法案」

### (Consumer Privacy Bill of Rights)

https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf 資策會科法所 https://stli.iii.org.tw/ContentPage.aspx?i=5662

- ▶ 1、透明度(Transparency): 消費者能容易的了解隱私及資訊安全的訊息。
- ▶ 2、獨立控制 (Individual control): 消費者有權了解自身資料被誰蒐集,以及 他們如何使用這些資料。
- ▶ 3、考慮內文(Respect for context): 消費者有權期待蒐集個人資料的組織, 處理個人資料的方式能提供消費者知悉並且言行一致。
- 4、限制蒐集(Focused Collection and Responsible Use): 企業僅能有 限度的蒐集、處理及使用消費者資訊。
- ▶ 5、安全性(Security): 消費者個資應受到安全可信任的保護。
- ▶ 6、近用與正確性 (Access and Accuracy): 消費者有權查詢與更正個人資料。
- ▶ 7、 責任 (Accountability): 消費者有權要求蒐集資訊的公司妥善保管個人資料 並遵循「消費者隱私保護法案」。

## Ten simple rules for responsible big data research

- 1. Acknowledge that data are people and can do harm
- 2. Recognize that privacy is more than a binary value
- 3. Guard against the re-identification of your data
- 4. Practice ethical data sharing
- 5. Consider the strengths and limitations of your data; big does not automatically mean better
- 6. Debate the tough, ethical choices
- 7. Develop a code of conduct for your organization, research community, or industry
- 8. Design your data and systems for auditability
- 9. Engage with the broader consequences of data and analysis practices
- 10. Know when to break these rules

## Anonymization (去識別化)

- 個人資料:指自然人之姓名、出生年月日、國民身分證統一編號、 護照號碼、特徵、指紋、婚姻、家庭、教育、職業、病歷、 醫療、 基因、性生活、健康檢查、犯罪前科、聯絡方式、財務 情況、社 會活動及其他得以直接或間接方式識別該個人之資料。
   去識別化:採取一組合理之步驟,移除識別資料與資料主體間之關
- <sup>影</sup> 一路的一路。 一路的過程。
- 非個人資料:無法以直接或間接方式識別該個人之資料
- 深度學習產生的類神經網路有自己的內部規則來分類
- 去識別化的資料即非個人資料??

the President. President's Council of Advisors on. Science and Technology. May 2014

## Anonymization (去識別化) - 2014年筆戰

- Cavoukian "Setting the Record Straight: De-Identfication Does Work"
  - 大多數的re-identification事件都是因為de-identification做得不好或沒有依照標準 指引 來做而造成
  - 許多媒體的報導多誇大或誤解了學術論文中的敘述
  - 引用Narayanan等人所寫的"Robust de-anonymization of large sparse datasets" 對 Netfilx公開資料所做的re-identification作為例子
  - 大多數人沒有進行複雜再識別演算的能力
- Narayanan,"No silver bullet: De-identification still doesn't work"
  - 進行再識別者所擁有的輔助資料很難在進行去識別化時候做一個合理的界定
  - 且認為再識別風險的評估需要許多假設頂多只能算是heuristic而不夠formal
  - 另外也指出擁有再識別所需的軟體能力者可能有百萬之譜
  - 文中提到了Heritage Health Prize的公開資料集當初請他進行再識別評估作為例子, 這引起了負責Heritage Health Prize去識別化的Emam為文回應。

the President. President's Council of Advisors on. Science and Technology. May 2014

## Anonymization (去識別化)

- 去識別化的效果在學術界仍有不同的看法
- 去識別化還沒有完整的數學模式
- 若以去識別化為基礎認為不再是個人的資料故無隱私問題需相當謹慎
- 去識別化仍有其價值但須嚴謹且明確說明所做的假設
- 去識別化認證在台灣

院 育計 日本

https://www.etc.org.tw/驗證服務/個人資料去識別化過程驗證.aspx

the President. President's Council of Advisors on. Science and Technology. May 2014

## Ten simple rules for responsible big data research

- This paper is a result of a two-year National Science Foundation (NSF)-funded project that established the Council for Big Data, Ethics, and Society, a group of 20 scholars from a wide range of social, natural, and computational sciences (http://bdes.datasociety.net/).
- The Council was charged with providing guidance to the NSF on how to best encourage ethical practices in scientific and engineering research, utilizing big data research methods and infrastructures.

![](_page_54_Picture_0.jpeg)

## User Informed Consent (同意)

- Declare clearly the purpose of the study
- Treat everyone fairly, especially to the minorities or those vulnerable groups.
- Be efficient: human-to-human bonding (building relationships), human judgement plus exercising empathy

### **Privacy and consent**

The requirements need to be adapted for a public health context as opposed to a commercial context

Vayena E, Salathé M, Madoff LC, Brownstein JS (2015) Ethical Challenges of Big Data in Public Health. PLoS Comput Biol 11(2): e1003904. doi:10.1371/journal.pcbi.1003904

![](_page_55_Picture_0.jpeg)

## Data Ownership (所有權)

- Others, not yours'
- Report study results to participants in a certain/formal form.
- Declare how the study results will be presented or published in which forms.

### Legitimacy

Digital disease detection needs codes of best practice to meet ethical requirement as well as clear communication to the public to prevent hype.

![](_page_56_Picture_0.jpeg)

## Data minimization (最小原則)

- Consider the necessity of carrying national-wide research.
- Show users how the study results are beneficial to them.
  - If not, why to carry out the study.
- Regarding the purchase of different kinds of databases.

### **Methodological robustness**

Methodology is evolving and requires constant adaptation to avoid false identification of outbreaks that could cause harm.

Vayena E, Salathé M, Madoff LC, Brownstein JS (2015) Ethical Challenges of Big Data in Public Health. PLoS Comput Biol 11(2): e1003904. doi:10.1371/journal.pcbi.1003904

![](_page_57_Picture_0.jpeg)

時間	內 容	講者
12:30-13:00		報到&現場諮詢
13:00-13:10	開場致詞	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
13:10-14:40	大數據研究倫理議題	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
14:40-14:50	休息時間	
14:50-16:20	統計分析的倫理議題	陸偉明 成功大學人類研究倫理審查委員會書審專家 暨教育研究所特聘教授
16:20-16:30	議題討論與Q&A	
16:30~	自由交流賦歸	

![](_page_58_Picture_0.jpeg)

時間	內 容	講者
12:30-13:00		報到&現場諮詢
13:00-13:10	開場致詞	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
13:10-14:40	大數據研究倫理議題	謝佩璇 成功大學人類研究倫理審查委員會副主任委員 暨工業與資訊管理學系副教授
14:40-14:50	休息時間	
14:50-16:20	統計分析的倫理議題	陸偉明 成功大學人類研究倫理審查委員會書審專家 暨教育研究所特聘教授
16:20-16:30	議題討論與Q&A	
16:30~	自由交流賦歸	