

# 運用大數據研究的倫理議題

陳祖裕

大林慈濟醫院人體研究保護計畫總負責人

醫學研究倫理基金會董事暨聯合人體試驗委員會執行秘書

台中慈濟醫院人體試驗委員會委員

中山醫學大學附設醫院受試者保護諮議委員會委員

教育部人體研究倫理審查委員會查核作業諮議會議委員暨查核委員

財團法人醫院評鑑暨醫療品質策進會人體研究倫理審查會查核委員

前美國人體試驗委員會專業人員 (CIP)

# 報告大綱

---

- 前言
- 大數據研究
- 大數據研究之倫理考量
- 大數據研究之倫理審查
- 結語

# Belmont Report

---

- 三個必要的基本倫理原則：
    - 對人的尊重 (Respect for Persons)
    - 行善 (Beneficence)
    - 公平正義 (Justice)
  - 應用
    - 知情同意
    - 風險與利益評估
    - 受試者的選擇
- 

# 報告大綱

---

- 前言
- 大數據研究
- 大數據研究之倫理考量
- 大數據研究之倫理審查
- 結語



## U.S. Department of Health and Human Services

Enhancing the health and well-being of all Americans

### **Attachment A: Human Subjects Research Implications of “Big Data” Studies**

Specific recommendations include the following, the background and rationale for each of which is set forth in this SACHRP report:

#### **Recommendation One:**

- 許多研究人員和IRB對大數據研究的法規有明顯的焦慮
- 這種焦慮可能導致研究延遲
- OHRP澄清：這樣的研究可依循微小風險和實用性的同意豁免標準

**SACHRP = Secretary’s Advisory Committee of Human Research Protection**

DEBATE

Open Access

# Ethics review of big data research: What should stay and what should be reformed?

Agata Ferretti<sup>1\*</sup> , Marcello Ienca<sup>1</sup>, Mark Sheehan<sup>2</sup>, Alessandro Blasimme<sup>1</sup>, Edward S. Dove<sup>3</sup>, Bobbie Farsides<sup>4</sup>, Phoebe Friesen<sup>5</sup>, Jeff Kahn<sup>6</sup>, Walter Karlen<sup>7</sup>, Peter Kleist<sup>8</sup>, S. Matthew Liao<sup>9</sup>, Camille Nebeker<sup>10</sup>, Gabrielle Samuel<sup>11</sup>, Mahsa Shabani<sup>12</sup>, Minerva Rivas Velarde<sup>13</sup> and Effy Vayena<sup>1</sup>

## Abstract

**Background:** Ethics review is the process of assessing the ethics of research involving humans. The Ethics Review Committee (ERC) is the key oversight mechanism designated to ensure ethics review. Whether or not this governance mechanism is still fit for purpose in the data-driven research context remains a debated issue among research ethics experts.

**Main text:** In this article, we seek to address this issue in a twofold manner. First, we review the strengths and weaknesses of ERCs in ensuring ethical oversight. Second, we map these strengths and weaknesses onto specific challenges raised by big data research. We distinguish two categories of potential weakness. The first category concerns persistent weaknesses, i.e., those which are not specific to big data research, but may be exacerbated by it. The second category concerns novel weaknesses, i.e., those which are created by and inherent to big data projects. Within this second category, we further distinguish between purview weaknesses related to the ERC's scope (e.g., how big data projects may evade ERC review) and functional weaknesses, related to the ERC's way of operating. Based on this analysis, we propose reforms aimed at improving the oversight capacity of ERCs in the era of big data science.

**Conclusions:** We believe the oversight mechanism could benefit from these reforms because they will help to overcome data-intensive research challenges and consequently benefit research at large.

**Keywords:** Big data, Research ethics, Ethics, IRBs, RECs, Ethics review, Biomedical research

DEBATE

Open Access

# Ethics review of big data research: What should stay and what should be reformed?

Agata Ferretti<sup>1\*</sup> , Marcello Ienca<sup>1</sup>, Mark Sheehan<sup>2</sup>, Alessandro Blasimme<sup>1</sup>, Edward S. Dove<sup>3</sup>, Bobbie Farsides<sup>4</sup>, Phoebe Friesen<sup>5</sup>, Jeff Kahn<sup>6</sup>, Walter Karlen<sup>7</sup>, Peter Kleist<sup>8</sup>, S. Matthew Liao<sup>9</sup>, Camille Nebeker<sup>10</sup>, Gabrielle Samuel<sup>11</sup>, Mahsa Shabani<sup>12</sup>, Minerva Rivas Velarde<sup>13</sup> and Effy Vayena<sup>1</sup>

1. Health Ethics and Policy Lab, Department of Health Sciences and Technology, ETH Zurich, Hottingerstrasse 10 (HOA), 8092 Zurich, Switzerland.
2. The Ethox Centre, Department of Population Health, University of Oxford, Oxford, UK.
3. School of Law, University of Edinburgh, Edinburgh, UK.
4. Brighton and Sussex Medical School, Brighton, UK.
5. Biomedical Ethics Unit, Department of Social Studies of Medicine, McGill University, Montreal, Canada.
6. Johns Hopkins Berman Institute of Bioethics, Baltimore, USA.
7. Mobile Health Systems Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland.
8. Cantonal Ethics Committee Zurich, Zurich, Switzerland.
9. Center for Bioethics, Department of Philosophy, New York University, New York, USA.
10. Research Center for Optimal Digital Ethics in Health (ReCODE Health), Herbert Wertheim School of Public Health and Longevity Science, University of California, San Diego, USA.
11. Department of Global Health and Social Medicine, King's College London, London, UK.
12. Faculty of Law and Criminology, Ghent University, Ghent, Belgium.
13. Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva, Geneva, Switzerland.

- ❑ 利用非常規研究之資料來源（例如社群媒體）、部分難以理解的資料分析工具（例如機器學習）以及前所未有的資料量
- ❑ 研究方法的演變和事後資料探勘等新方法已模糊了「人類主體」的概念
- ❑ 個體和個體集合的研究資料儲存庫的規模正在顯著擴大
- ❑ 研究人員可以自遠端存取和使用大量潛在敏感數據，而無需與研究參與者進行溝通或積極互動
- ❑ 資料保護和隱私問題成為大數據研究審查的焦點
- ❑ 從身體或心理傷害的風險轉向資訊傷害的風險，例如隱私洩露或演算法歧視

# Experimental evidence of massive-scale emotional contagion through social networks

Proceedings of the National Academy of Sciences 2014

Adam D. I. Kramer<sup>a,1</sup>, Jamie E. Guillory<sup>b,2</sup>, and Jeffrey T. Hancock<sup>b,c</sup>

<sup>a</sup>Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of <sup>b</sup>Communication and <sup>c</sup>Information Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

情緒狀態可以透過情緒傳染傳遞給他人，導致人們在不知不覺中體驗到相同的情緒。情緒傳染在實驗室實驗中已被充分證實，人們會將正面和負面的情緒傳遞給他人。來自大型現實世界社交網絡歷時20年收集的數據顯示，持久的情緒（例如抑鬱、快樂）可以透過網絡傳遞 [Fowler JH, Christakis NA (2008)BMJ337:a2338]，儘管結果仍有爭議。在一項針對 Facebook 用戶的實驗中，我們透過減少動態消息中的情緒內容量來測試「人與人之間如果沒有面對面互動，情感傳染是否仍會發生？」。當正面的表達減少時，人們產生的正面貼文就會減少，而負面的貼文就會增加；當消極表達減少時，就會出現相反的模式。這些結果顯示，別人在Facebook上表達的情緒會影響我們自己的情緒，構成透過社群網路大規模傳播的實驗證據。這項研究還顯示，與普遍的假設相反，面對面的互動和非語言暗示對於情緒傳染並不是嚴格必要的，對他人積極經歷的觀察對人們來說也是一種積極的經歷。

# Experimental evidence of massive-scale emotional contagion through social networks

Proceedings of the National Academy of Sciences 2014

Adam D. I. Kramer<sup>a,1</sup>, Jamie E. Guillory<sup>b,2</sup>, and Jeffrey T. Hancock<sup>b,c</sup>

<sup>a</sup>Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of <sup>b</sup>Communication and <sup>c</sup>Information Science, Cornell University, Ithaca, NY 14853

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

## 重要發現

我們透過 Facebook 上的大規模實驗 (N=689,003)，發現情緒狀態可以透過情緒傳染傳遞給他人，導致人們在不知不覺中體驗到相同的情緒。

我們提供的實驗證據顯示，在人與人之間沒有直接互動（包括完全沒有非語言暗示）的情況下也可發生情緒傳染。

- 2014年的「情緒傳染研究」受到嚴厲批評，因為
  - 逃避IRB的倫理審查
  - 未獲得研究同意
  - 侵犯隱私
  - 造成情緒傷害
  - 歧視資料主體
  - 將弱勢參與者（例如兒童和青少年）置於危險之中

- 根據美國法律，該研究不需要經過倫理審查，原因有二：
  - 首先，分析的數據被認為是不可識別的，研究人員沒有直接與受試者接觸，從而使研究免於倫理審查
  - 其次，研究團隊包括公立大學（康乃爾大學）的科學家和 Facebook 員工，私人資助的研究不受與公共資助的研究相同的研究保護和倫理法規的約束

- ❑ 2015 年：英國國民保健署 (NHS) 與 Google DeepMind 共享 160 萬條可識別敏感資料的案例
- ❑ 這種從公眾到私人方的資料傳輸是合法進行的，不需要病人同意或倫理審查監督
- ❑ 這些案例顯示了研究人員如何在不屬於 ERC 職權範圍的情況下進行具有潛在風險的大數據研究

# Google and DeepMind face legal claim for unauthorised use of NHS medical records

The data on 1.6 million patients was allegedly used without their consent.

By [Tammy Lovell](#) | May 17, 2022 | 02:12 PM

Google因在 160 萬 NHS 病人不知情或未同意的情況下非法使用其機密醫療記錄而面臨集體訴訟。

由 Mishcon de Reya 律師事務所代表的 Andrew Prismall 在英格蘭和威爾斯高等法院提起新的代表訴訟，對 Google 提出索賠。

# Google and DeepMind face legal claim for unauthorised use of NHS medical records

The data on 1.6 million patients was allegedly used without their consent.

By [Tammy Lovell](#) | May 17, 2022 | 02:12 PM

為什麼它很重要？

「濫用私人資訊」的指控與Google子公司 DeepMind 和Royal Free London NHS 基金會信託公司於2015年達成的一項協議有關。DeepMind從Royal Free收到病人數據，用於對名為「Streams」的智慧型手機應用程式進行臨床安全測試，該應用程式是為檢測急性腎損傷而開發的。該應用程式隨後被Royal Free以折扣價使用。英國資料隱私組織資訊專員辦公室 (ICO) 隨後裁定 Royal Free 在提供病患資料時未遵守資料保護法的要求。

# Google and DeepMind face legal claim for unauthorised use of NHS medical records

The data on 1.6 million patients was allegedly used without their consent.

By [Tammy Lovell](#) | May 17, 2022 | 02:12 PM

## 更大的背景

這起訴訟發生之際，人們愈來愈擔心科技巨頭可能濫用健康數據。

2020年，歐洲資料保護委員會 (EDPB) 命令 Google 對其收購穿戴式裝置巨頭 Fitbit 的交易進行「資料保護要求和隱私影響的全面評估」。該協議最終簽署，並提出10年內必須遵循的一系列要求。

DeepMind 先前曾是針對Google和芝加哥大學醫學中心的集體訴訟的核心，該中心在兩者達成數據協議後指控侵犯患者隱私。

# Google and DeepMind face legal claim for unauthorised use of NHS medical records

The data on 1.6 million patients was allegedly used without their consent.

By [Tammy Lovell](#) | May 17, 2022 | 02:12 PM

## 更大的背景（續）

同時，科技司法公司Foxglove去年代表新聞網站openDemocracy就NHS新冠資料儲存與Palantir達成的價值2,300萬英鎊的交易提起訴訟。該訴訟稱，英格蘭國民保健署 (NHS England) 未能透過進行新的資料保護影響評估來考慮該交易對患者和公眾的影響。在面臨司法審查索賠後，政府承認了訴訟，並同意在未經協商的情況下不會將Palantir的合約延長到疫情之後。在2020年6月的上一次法律挑戰中，openDemocracy和Foxglove迫使英國政府公佈與大型科技公司的合同，認為公眾有權了解健康數據資產的轉移。

# Google and DeepMind face legal claim for unauthorised use of NHS medical records

The data on 1.6 million patients was allegedly used without their consent.

By [Tammy Lovell](#) | May 17, 2022 | 02:12 PM

## 記錄在案

Prismall表示：「我希望此案能為許多病人帶來公平的結果和結案，這些大型科技公司在病人不知情的情況下獲取並使用了他們的機密紀錄。」

Mishcon de Reya合夥人Ben Lasserson表示：「這一主張尤其重要，因為它應該提供一些急需的明確性，明確允許技術公司訪問和使用私人健康信息的適當參數。」

- ❑ 新的參與式治理模式已經出現，很大程度上是由於主體控制自己資料的要求增加⇒公司效法IRB，也建立研究倫理委員會
- ❑ 2020年5月：Facebook 迎來了第一批監督委員會成員，其目的是審查公司有關內容審核的決定
- ❑ 數據驅動的數位轉型對IRB的權限和效率提出了挑戰
- ❑ 它也提出了有關IRB作為科學研究道德和行為學健全性監督機構的作用和範圍的基本問題

# 報告大綱

---

- 前言
- 大數據研究
- 大數據研究之倫理考量
- 大數據研究之倫理審查
- 結語

# 大數據研究之倫理考量

---



## Big data and data sharing: Ethical issues

---

UK Data Service

---



Author: UK Data Service  
Updated: February 2017  
Version: 1

Libby Bishop (2017). Big data and data sharing: Ethical issues. UK Data Service, UK Data Archive.

# 大數據研究的主要倫理問題是什麼？

---

- 隱私
- 知情同意
- 去識別化
- 不平等—數位鴻溝
- 研究廉正
- 新出現的問題

- 隱私
- 知情同意
- 去識別化
- 不平等—數位鴻溝
- 研究廉正
- 新出現的問題

# 大數據研究之倫理考量—隱私

---

- 隱私權是為人權
- 研究對象的隱私可以透過多種方法來保護：限制收集哪些資料、更改資料以減少洩漏、規範對資料的存取...
- 但在大數據研究背景下，「私有」和「隱私」的定義含糊不清或有爭議：
  - 社群媒體空間是公共的還是私人的？
  - 遵守正式協議就足夠了嗎？
  - 能夠明確地區分公共和私人使用者或用途？
  - 資料科學是否應該被歸類為人類學科研究？

# 大數據研究之倫理考量——知情同意

- 隱私
- 知情同意
- 去識別化
- 不平等—數位鴻溝
- 研究廉正
- 新出現的問題

- 在大數據分析中，人們對收集資料時的未來預期用途知之甚少，故無論是利益或風險都無法被有意義地理解
- 在資料收集時（一次性）獲得的同意不太可能滿足「知情同意」的嚴格定義
- 例如，共享基因組數據的「廣泛」和「通用」同意程序受到批評，因為對未來基因技術未知的風險，這種同意沒有意義

# 大數據研究之倫理考量——去識別化

- 隱私
- 知情同意
- 去識別化
- 不平等—數位鴻溝
- 研究廉正
- 新出現的問題

- 傳統上：如果立即識別個人或組織的元素已被刪除或屏蔽，則表示資料集已被去識別化
- 大數據：可辨識性愈來愈被視為一個連續體，而不是二元體
- 洩露風險隨之增加之狀況：
  - 隨著維度（即變數數量）
  - 多個資料來源的連結
  - 資料分析的能力
- 洩露風險可以減輕，但不能消除

# 大數據研究之倫理考量——數位鴻溝

- 隱私
- 知情同意
- 去識別化
- 不平等——數位鴻溝
- 研究廉正
- 新出現的問題

- 誰可以存取資料？
- 原則上，任何研究人員都可以透過其應用程式介面（Application Programming Interface，API）存取Twitter 資料，但因所需的**成本**和**技能**而存在存取障礙
- 誰管理資料存取？真的是平等開放的嗎？這

# 大數據研究之倫理考量——研究廉正

- 隱私
- 知情同意
- 去識別化
- 不平等—數位鴻溝
- 研究廉正
- 新出現的問題

- 資料儲存庫的數據必須有明確的「來源」，研究人員須要了解、識別和記錄其來源和處理過程
- 資料管理者和資料生產者之間的關係減弱，使得這一點具有挑戰性，原因如下：
  - 許多並非為研究而收集的資料與研究資料有不同的標準
  - 對於某些通常具有商業價值的類型（例如Twitter 資料），複製資料存在法律限制
  - 資料儲存庫在與不定期為社會研究產生資料的資料生產者群體合作時，在履行其對透明度和可重複性標準的承諾方面面臨著挑戰

# 大數據研究之倫理考量——新的問題

- 隱私
- 知情同意
- 去識別化
- 不平等——數位鴻溝
- 研究廉正
- 新出現的問題

- 個人知情同意的替代方案：例如「社會同意」，從而採取足夠的保護措施，在倫理上允許在未經個人知情同意的情況下使用數據
- 在決定允許重複使用哪些資料時須要尊重資料的來源和出處，以及更廣泛的資料「上下文完整性」
- 大多數研究倫理都基於這樣的假設：面臨風險的實體是個人，因此去識別化可以提供保護；但若因具有群體成員身份便會受到傷害，則去識別化的保護就不再充分
- 如果不能清楚地區分公共和私人，有人建議根據結果評估資料使用情況，並允許出於「公共利益」或「公眾利益」使用資料

# 報告大綱

---

- 前言
- 大數據研究
- 大數據研究之倫理考量
- 大數據研究之倫理審查
- 結語

DEBATE

Open Access

# Ethics review of big data research: What should stay and what should be reformed?

Agata Ferretti<sup>1\*</sup> , Marcello Ienca<sup>1</sup>, Mark Sheehan<sup>2</sup>, Alessandro Blasimme<sup>1</sup>, Edward S. Dove<sup>3</sup>, Bobbie Farsides<sup>4</sup>, Phoebe Friesen<sup>5</sup>, Jeff Kahn<sup>6</sup>, Walter Karlen<sup>7</sup>, Peter Kleist<sup>8</sup>, S. Matthew Liao<sup>9</sup>, Camille Nebeker<sup>10</sup>, Gabrielle Samuel<sup>11</sup>, Mahsa Shabani<sup>12</sup>, Minerva Rivas Velarde<sup>13</sup> and Effy Vayena<sup>1</sup>

## Abstract

**Background:** Ethics review is the process of assessing the ethics of research involving humans. The Ethics Review Committee (ERC) is the key oversight mechanism designated to ensure ethics review. Whether or not this governance mechanism is still fit for purpose in the data-driven research context remains a debated issue among research ethics experts.

**Main text:** In this article, we seek to address this issue in a twofold manner. First, we review the strengths and weaknesses of ERCs in ensuring ethical oversight. Second, we map these strengths and weaknesses onto specific challenges raised by big data research. We distinguish two categories of potential weakness. The first category concerns persistent weaknesses, i.e., those which are not specific to big data research, but may be exacerbated by it. The second category concerns novel weaknesses, i.e., those which are created by and inherent to big data projects. Within this second category, we further distinguish between purview weaknesses related to the ERC's scope (e.g., how big data projects may evade ERC review) and functional weaknesses, related to the ERC's way of operating. Based on this analysis, we propose reforms aimed at improving the oversight capacity of ERCs in the era of big data science.

**Conclusions:** We believe the oversight mechanism could benefit from these reforms because they will help to overcome data-intensive research challenges and consequently benefit research at large.

**Keywords:** Big data, Research ethics, Ethics, IRBs, RECs, Ethics review, Biomedical research

# 大數據研究之倫理審查

---

- IRB在進行大數據研究之倫理審查時可能存在以下弱點：
  - 固有的弱點：IRB審查原本存在的弱點，大數據研究可能會加劇這些弱點
  - 新生的弱點：由大數據研究的性質帶來特有的弱點，可進一步區分為兩類：
    - 權限弱點：一些大數據項目可能繞過IRB的權限
    - 功能弱點：有些IRB可能不足以充分評估大數據研究因素

# IRB的弱點

## 固有的弱點



不同司法管轄  
區法規的衝突

ERC 內部和跨 ERC  
的審查標準不一致

不適合重新分配研究  
的風險和效益

缺乏透明度

待審查的項目  
遠超負荷

## 新生的弱點



### 範圍



大數據研究範圍廣泛

匿名資料的使用

新執行者進行研究

### 功能性



事前審查模式

不可預見風險評估

知情同意操作化

缺乏大數據專業知識

# 大數據研究之倫理審查—固有的弱點

---

- ❑ 大數據研究往往依賴多中心國際聯盟，而各國對研究監督的法規各不相同，各自要求進行不同類型的審查，加劇了當前監督機制的這一弱點
- ❑ 大數據研究可能會促進公共機構、大學、企業、基金會和科學學會之間的合作，在這個網絡中，每個利害關係人都有不同的優先級，並依賴自己的研究過程監管規則
- ❑ 學術期刊和資助機構對大數據倫理的看法存在分歧，導致缺乏統一方針
- ❑ 各國對人體研究倫理審查的要求各不相同，則一個國家的IRB認為可以接受的大數據項目可能會被另一個國家的IRB拒絕
- ❑ 此外，委員會之間所使用的評估標準也不一致

# 大數據研究之倫理審查—固有的弱點

---

- ❑ 大數據研究是一個較新的領域，加上沒有標準化的指標來評估國家內部和國家之間的風險，故可能會加劇分歧
- ❑ 目前的審查制度仍主要傾向於保護個別研究參與者的利益，IRB不會根據研究的整體進行評估社會風險和效益，配合個人風險與集體利益找到平衡點
- ❑ IRB通常仍採用針對傳統形式的生物醫學研究的審查方法（這些研究通常範圍狹窄，並且僅對所涉及的參與者帶來特定的風險）

# 大數據研究之倫理審查—固有的弱點

---

- ❑ 相較之下，大數據項目會更廣泛地影響社會（例如，電腦技術已能透過篩選臉部影像來確定個人性取向的潛力）
- ❑ 對此類研究產生的共同利益的評估不足可能會對社會造成損害
- ❑ 大數據計畫類似於公共衛生研究，其倫理重點是公共利益而不是個人自主權
- ❑ IRB承擔更大的責任，確保研究成果在人群中公平分配
- ❑ 準確地確定大數據研究的社會價值具有挑戰性，因為在研究開始之前可能很難發現其負面後果

# 大數據研究之倫理審查—固有的弱點

---

- ❑ 各國的共同點是IRB很少揭露其程序、政策和決策過程
- ❑ IRB的「保密性」可能會導致人們對倫理監督模式缺乏信任
- ❑ 在大數據研究中，隨著資料主體數量呈指數級增長，IRB缺乏問責制和不透明的審查過程可能會帶來更嚴重的公眾抵制
- ❑ 研究人員是另一類因IRB溝通和宣傳不善而受到負面影響的利害關係人
- ❑ IRB常常沒有明確提供有關研究審查中應用的倫理標準的指導
- ❑ 如果研究人員向資料主體提供不切實際的隱私和安全期望，IRB有責任標記這些承諾（例如，關於資料安全和主體資料的二次使用），特別是當研究涉及個人和高度敏感時數據

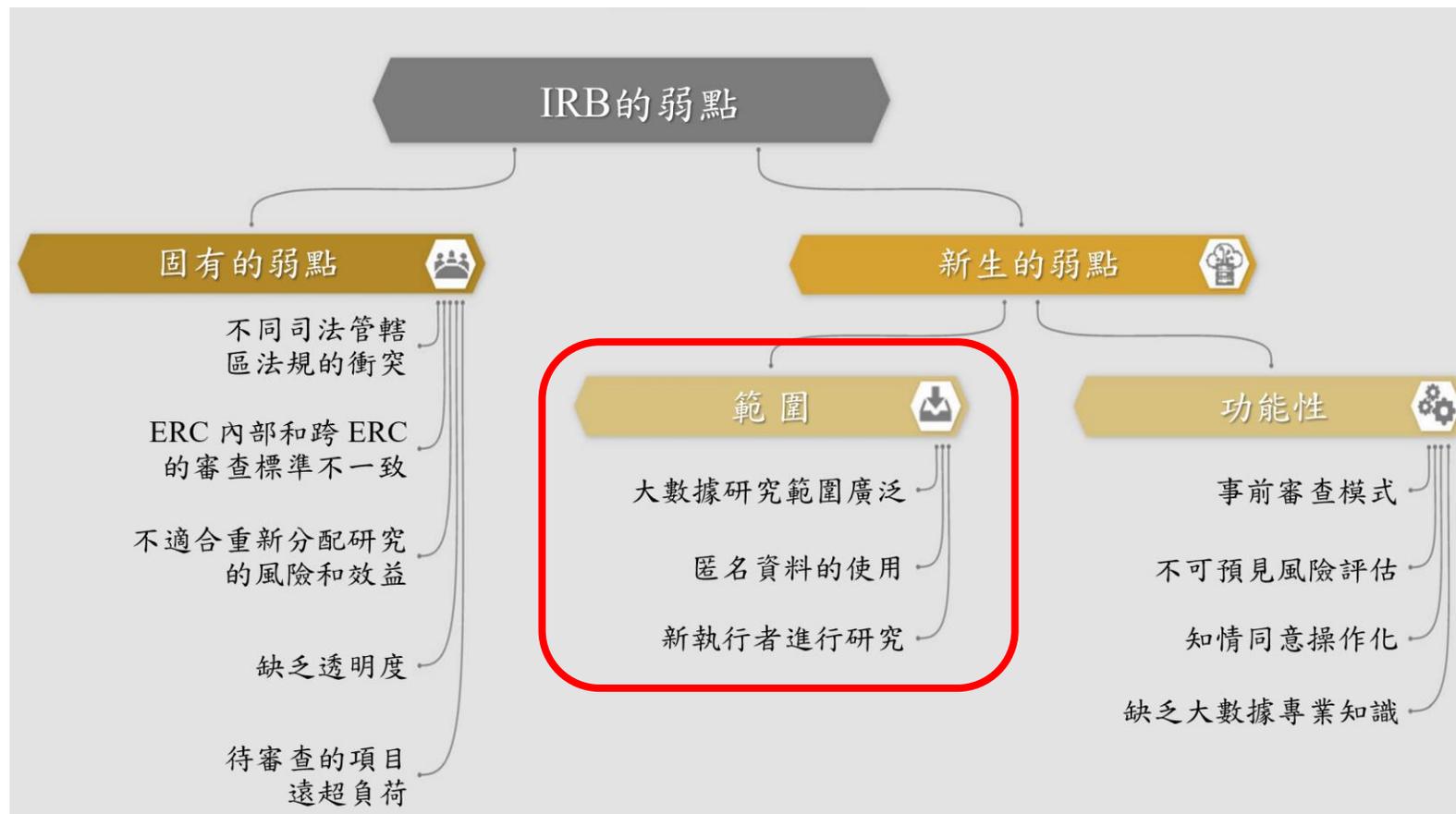
# 大數據研究之倫理審查—固有的弱點

---

- ❑ IRB正面臨因審查數量導致的不堪負荷的情況日益嚴重
- ❑ 廣泛的大數據研究可能會加劇這個現有問題
- ❑ 執行更多的研究審查和較少的時間完成任務可能會對IRB的審查品質產生負面影響，並延長審查所需的時間，因而導致：
  - ❑ 研究人員可能會執行具潛在風險的研究，因為這些研究的相關倫理問題被忽略
  - ❑ 研究本身可能會被大大推遲，直到失去其及時的科學價值

# 大數據研究之倫理審查——新生的弱點

- ❑ 範圍弱點
- ❑ 功能性弱點



# 大數據研究之倫理審查——範圍弱點 (1/5)

---

- ❑ 首先要確定大數據專案在哪些條件下屬於IRB的職權範圍
- ❑ 大數據計畫往往超越單一機構、跨學科、涉及從各種來源收集的數據、重複使用最初不是為研究目的收集的數據、結合不同的方法、指向人群層面的研究、依賴大數據聚合，並透過與私部門的合作而產生
- ❑ 鑑於這種情況，大數據項目可能不屬於IRB的權限範圍
- ❑ 大數據研究不屬於IRB權限的另一種情況是它依賴匿名數據
- ❑ 如果研究人員使用無法追溯到受試者的資料（匿名或非個人資料），那麼根據美國Common Rule和HIPAA法規，該專案被認為足夠安全，可以獲得倫理審查的豁免

## 大數據研究之倫理審查—範圍弱點 (2/5)

---

- ❑ 如果研究人員使用去識別化的數據，則必須申請研究倫理審查，因為原則上，將去識別化數據與受試者聯繫起來的密鑰可能會被洩露或被駭客攻擊，從而對受試者造成傷害
- ❑ 在歐盟，將由每個成員國（以及當地機構的國家法律或政策）來定義使用去名化資料的研究是否須要倫理審查
- ❑ 當前研究倫理法規相對寬鬆且跨司法管轄區脫節，並可能導致大數據研究有不受監管的領域
- ❑ 對去名化資料的特殊處理係來自於對個人層面的風險
- ❑ 到目前為止，在大數據討論中，危害的概念主要考量資料保護的脆弱性

# 大數據研究之倫理審查——範圍弱點 (3/5)

---

- ❑ 如果隱私法律得到尊重，並且在資料系統中內建保護，研究人員就可以防止有害的結果
- ❑ 然而，這種觀點是短視的，因為它不包括其他數據聚合的濫用，例如群體歧視和名人傷害
- ❑ 這些類型的危害已經在大數據生態系統中出現，其中去名化數據揭示了某個子群體的健康模式，或者計算技術包含強烈的種族偏見
- ❑ 使用去名化數據的研究不應被視為默認情況下不受監督，因為數據去名化化愈來愈困難——技術進步可能很快就能從聚合資料集中重新識別到個人

# 大數據研究之倫理審查—範圍弱點 (4/5)

---

- 由於與大學研究人員合作研究的團體多種多樣（例如私人公司、公民科學協會、生物公民團體、社區工作者合作社、基金會和非營利組織），與大數據項目相關的風險也有所增加
- 與傳統研究相比，健康相關大數據研究的新穎之處在於，任何能夠造訪有關個人的大量數據並基於這些數據構建預測模型的人，現在都可以確定並推斷一個人的健康狀況，而無需直接參與研究計畫中的人
- 例如，Facebook 進行一項自殺預測和預防項目，該項目完全依賴用戶在社群網路上發布的資訊

# 大數據研究之倫理審查—範圍弱點 (5/5)

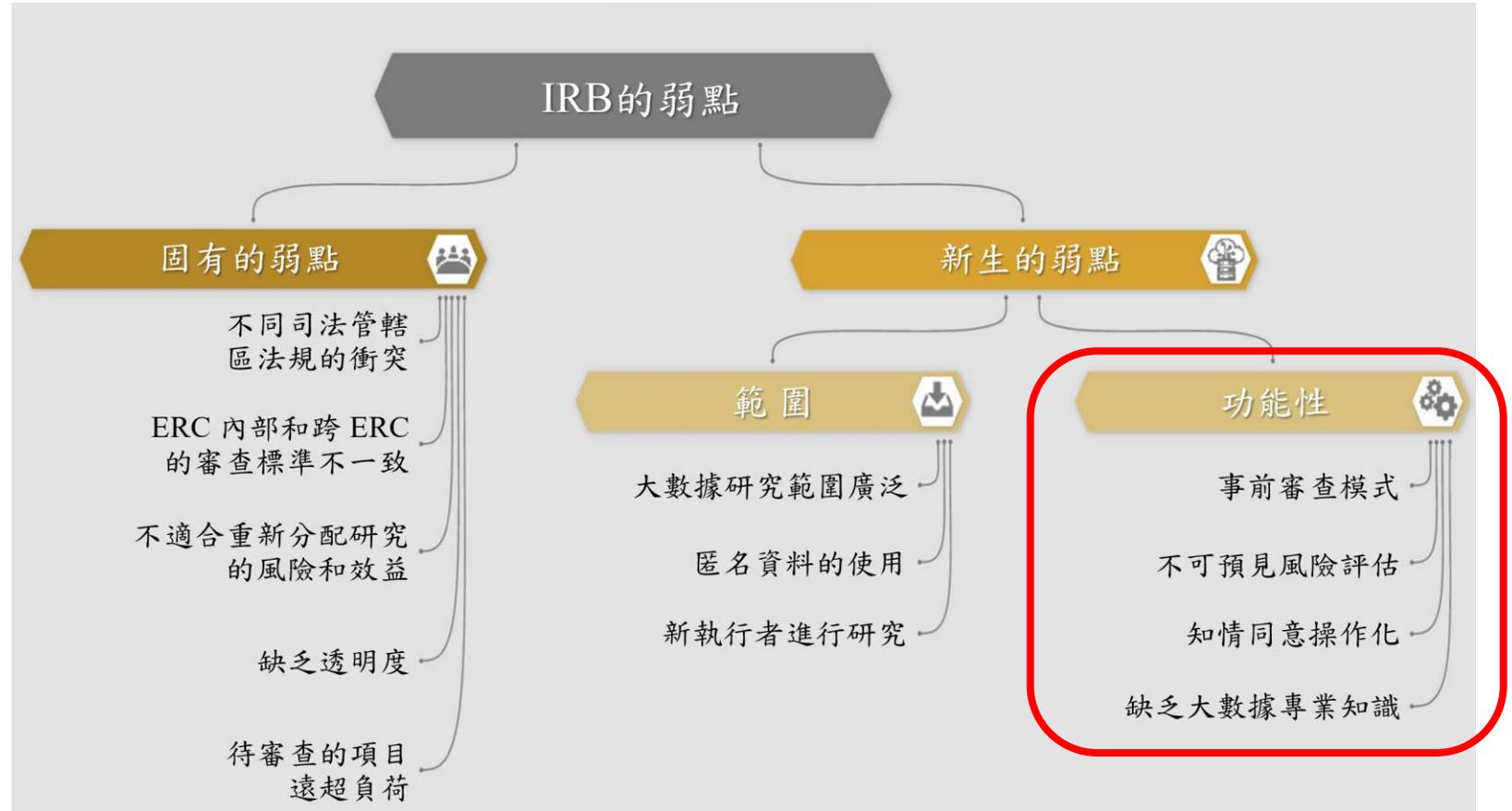
---

- ❑ 現有的倫理審查模式使許多大數據項目免除IRB審查，因此監管差距正在擴大
- ❑ 就像企業可以重複使用公開資料集（例如社群媒體資料）來確定人壽保險費一樣，公民科學計畫也可以在不尋求研究監督的情況下進行
- ❑ 事實上，參與者主導的大數據研究是傳統審委模式無效的另一個領域
- ❑ 此外，IRB並未嚴肅看待學術界或公共資助機構以外進行的研究，甚至漠視學術環境以外團體的審查請求

# 大數據研究之倫理審查——新生的弱點

- 範圍弱點
- 功能性弱點

功能性弱點是指IRB審查大數據研究時在技能、組成和作業等方而呈現的弱點。



# 大數據研究之倫理審查—功能弱點 (1/9)

---

- ❑ 事前審查模式可能不適合大數據研究
- ❑ 計畫設計階段或在資料收集層面的計畫評估不足以解決大數據研究所面臨的新挑戰，尤其是隨著時間的推移，資料可能會用於其他目的，從而重複使用或共享
- ❑ 事前審查模式的不足在基因研究領域已經顯而易見
- ❑ 在這種情況下，生物樣本庫通常必須進行第二次倫理評估，以授權對其原始資料樣本進行外顯子定序的特定研究用途
- ❑ 如果IRB核准了敏感個人資料的原始收集，資料存取委員會將確保二次使用符合原始同意和倫理審查的核准內容

## 大數據研究之倫理審查—功能弱點 (2/9)

---

- ❑ 然而，如果研究人員從公共存取平台收集數據，他們可以合法地和重複地使用數據進行研究，而無須尋求數據主體同意或IRB審查，在社群媒體研究中經常出現這種情況
- ❑ 研究人員或私人公司以廣泛同意的形式收集的社交媒體資料可以被研究人員重複使用以進行其他分析，而無須IRB核准
- ❑ 帶來不可預見風險的不僅是資料的重複使用
- ❑ 事前方法可能不適合評估資料生命週期的其他階段，例如部署機器學習（machine learning，ML）演算法

## 大數據研究之倫理審查—功能弱點 (3/9)

---

- ❑ 一些大數據研究不是重複使用數據，而是在現有數據上建立模式（使用數據挖掘和ML方法），創建新數據，然後用於進一步為演算法提供數據
- ❑ 有時無法預測研究中將利用哪些分析模式或工具（例如人工智慧〔artificial intelligence, AI〕）
- ❑ 從大數據中提取意義的計算技術的本質，使得預測分析中出現的所有相關性變得困難

# 大數據研究之倫理審查—功能弱點 (4/9)

---

- ❑ 大數據研究不易被清晰地框架，使得IRB更難預測不可預見的風險和潛力社會後果
- ❑ 鑑於現有法規和大數據項目固有的探索性，IRB的職責似乎無法保證做好對研究的監督
- ❑ 如果我們考慮大數據研究發表後可能出現的問題，例如再利用或雙重用途問題，情況似乎更不只如此

# 大數據研究之倫理審查—功能弱點 (5/9)

---

- ❑ 重新獲得研究對象的同意是不切實際的，特別是在將消費者產生的資料（例如社群媒體資料）用於研究目的時
- ❑ 在這些情況下，研究人員通常依賴概括同意和同意棄權
- ❑ 這使得資料主體不知道他們參與了特定的研究，因此使他們無法參與研究進展
- ❑ 因此，資料主體和他們所代表的社群很容易受到潛在負面研究結果的影響
- ❑ 同意工具在大數據研究中存在局限性—它無法披露數據未來所有可能的用途，部分原因是這些用途在數據生成時可能是未知的

# 大數據研究之倫理審查—功能弱點 (6/9)

---

- ❑ 此外，研究人員可以多次存取現有資料集，並將相同的資料用於其他目的
- ❑ 鑑於目前的知情同意模式在大數據研究中留下了倫理差距，IRB的策略應該是什麼？
- ❑ IRB可能專注於知情同意挑戰，反而忽略其他緊切的大數據研究問題
- ❑ 而文獻報導，愈來愈多的論文作者反對大數據研究須有新的同意書版本的想法

# 大數據研究之倫理審查—功能弱點 (7/9)

---

- ❑ 最後一個廣泛討論的問題是IRB在大數據研究領域的專業知識不足
- ❑ 大數據研究的獨特方法（基於資料聚合和挖掘）需要專門的技術專長（例如資訊系統、自學習演算法和匿名協定）
- ❑ 事實上，由於數據量和來源，大數據專案具有強大的技術成分，這帶來了特定的挑戰（例如，在社交媒體上的傳統協議之外收集數據）
- ❑ 其次，IRB可能不熟悉參與大數據研究的新參與者，例如公民科學參與者或私人公司
- ❑ 由於缺乏相關專業知識，IRB可能會要求對研究進行不合理的修改，甚至完全拒絕大數據專案

# 大數據研究之倫理審查—功能弱點 (8/9)

---

- ❑ 最後，IRB可能會失去作為能夠評估倫理違規和研究不當行為的監督機構的信譽
- ❑ 過去，IRB在審查某領域的研究計畫時透過諮詢相關領域的獨立專家來解決這項挑戰，然而，此解決方案並不總是實用，因為它取決於專家的可用性
- ❑ 此外，專家本身也可能是在該領域工作和發表論文的研究人員
- ❑ 這種情況會產生問題，因為由研究人員來定義專家必須遵守的規則，這損害了獨立審查的概念

# 大數據研究之倫理審查—功能弱點 (9/9)

---

- 儘管如此，這個問題並沒有否定專業審查的角色，但需要有關規則制定和合規性的高透明度標準
- 其他選項包括成立特設專家委員會或為現有委員會成員提供相關培訓

# 大數據研究之倫理審查之改革

---

- ❑ 法規改革 (regulatory reforms)
- ❑ 程序改革 (procedural reforms)
- ❑ 配套改革 (Complementary reforms)

# 大數據研究之倫理審查之改革——法規改革

---

- ❑ 人體研究的法規設計是第一個需要改革的面向
- ❑ IRB可以受益於有關大數據研究倫理的新指引（例如以流程圖的形式）
- ❑ 該指引可以建立在數據對社會運作的重要性、我們在社會中使用數據的方式，以及我們使用數據的理由進行深入反思的基礎上
- ❑ 例如，在英國，個人通常可以選擇不讓醫師系統或 NHS 數位服務儲存其資料（如醫院就診資料、健康紀錄、處方藥）
- ❑ 然而，當數據的使用對社會運作至關重要時（例如，在官方國家統計數據或壓倒性公共利益的情況下，例如 COVID-19 大流行），則適用此選擇退出政策的例外情況

# 大數據研究之倫理審查之改革——法規改革

---

- 我們認為這項新指引也將重新定義IRB審查的範圍，從保護個人利益到更廣泛地對研究衝擊的評估
- 它將允許IRB的範圍擴大並解決先前討論的權限問題
- 例如，不受監督的研究將會減少，因為更多的因素會先觸發IRB的權限
- 新的治理將對涉及去名化資料的研究或公私合作夥伴關係中的大數據研究實施IRB審查
- 此外，IRB的權限可以從研究的初始階段擴展到資料生命週期的其他階段

# 大數據研究之倫理審查之改革——法規改革

---

- ❑ 一種可行的做法是在研究得出結論後對其進行評估（就像製藥業的情況）：然後IRB可以決定研究結果是否應該發布並由科學界進一步使用
- ❑ 這項新的倫理指引不僅可以幫助IRB決定研究計畫是否須要審查，還可以幫助他們從過去的例子和最佳作業中學習如何進行最好的評估
- ❑ 該指引亦可提高IRB所使用的評估標準的透明度
- ❑ 可以透過定義倫理評估的最低全球標準來實現透明度，該標準允許基於開放資料和同質評估模式進行國際合作

# 大數據研究之倫理審查之改革——法規改革

---

- 接受全球標準也意味著相同的監督程序將適用於具有相似風險和研究路徑的研究項目，無論它們是由公共還是私人實體進行
- 增加澄清和透明度還可以簡化委員會內部和跨委員會的審查流程，從而使整個系統更有效率

# 大數據研究之倫理審查之改革——程序改革

---

- ❑ 程序改革可能針對IRB模式的特定方面，使其更適合大數據研究的審查
- ❑ 首先，IRB應開發新的作業工具來應對新出現的大數據挑戰
- ❑ 例如，可以使用 **Ainow** 演算法影響評估工具，該工具評估自動決策系統的倫理規範，並為是否在社會中部署該系統的決策提供資訊
- ❑ 概括同意和動態同意等形式也可以解決使用、重複使用和分享大數據（公開或不公開）所產生的一些問題
- ❑ 儘管如此，知情同意不應被視為解決大數據研究中所有倫理問題的靈丹妙藥，尤其是在公開的社群媒體數據的情況下

# 大數據研究之倫理審查之改革——程序改革

---

- ❑ 如果大數據研究的倫理衝擊影響到社會及其弱勢群體，則不能依賴個人的知情同意作為有效的保障
- ❑ IRB應採取更民主的審查程序，可能的策略包括讓研究參與者和社群參與決策過程或促進共同治理體系
- ❑ 最近的《蒙特婁負責任人工智慧宣言》是公眾參與下制定的倫理監督流程的一個例子，這種包容性方法可以提高倫理審查機制本身的可信度
- ❑ IRB愈多讓潛在資料主體參與有關大數據研究風險的透明對話，監督機制就愈具有社會責任感



# 大數據研究之倫理審查之改革——程序改革

---

- ❑ IRB還必須解決其缺乏大數據和一般電腦專業知識的問題
- ❑ 有幾種潛在的方法可以彌補這一差距：
  - ❑ IRB可以透過大數據的正式訓練來增強能力
  - ❑ IRB可以調整成員資格，納入來自所需領域的特定專家（例如電腦科學家、生物技術專家、生物資訊學家、資料保護專家）
  - ❑ IRB可以與外部專家進行具體諮商

# 大數據研究之倫理審查之改革——程序改革

---

- ❑ 在數據驅動的研究背景下，IRB必須承擔起監督管理者的角色，並引導研究人員完成倫理審查過程
- ❑ IRB應與研究者建立開放的溝通管道，傳達研究倫理的價值，同時明確地評估研究的標準
- ❑ 如果IRB和研究人員同意共同提高透明度，他們就會創造一個從過去的錯誤中學習並防止未來再次發生錯誤的機會
- ❑ 大學可能會尋求對研究人員進行有關進行數據驅動研究時可能出現的倫理問題的教育
- ❑ 一般來說，研究人員將受益於識別倫理問題或填寫倫理自我評估表的培訓，特別是如果他們負責提交計畫供審查

# 大數據研究之倫理審查之改革——程序改革

---

- ❑ 隨著生物醫學研究逐漸從醫院和臨床試驗轉向人們的家庭和私人公司，研究人員應努力提高清晰度、透明度和責任感
- ❑ 研究人員應揭露預期的風險和收益，以及對個人和人群層面的預期影響
- ❑ 然後，IRB可以更有效地評估大數據研究的影響，並確定公共利益是否得到保障
- ❑ 此外，他們可能會探討研究成果如何在整個社會中分配
- ❑ 本地化決策可以在這裡發揮作用
- ❑ IRB可能會考慮特定於社會背景的特徵，以評估研究是否尊重社會價值

# 大數據研究之倫理審查之改革——配套改革

---

- ❑ 解決大數據研究新穎性的另一項措施可能包括透過監督和程序工具來改革當前的研究倫理體系
- ❑ 然而，這一策略可能還不夠：當前系統可能需要其他形式的監督的額外支持來補充其工作
- ❑ 一種可能性是創建混合審查機制和規範，將傳統IRB審查模式的有價值的方面與參與研究的各種合作夥伴（例如企業、參與者、社區）所採用的更具創新性的模式相結合
- ❑ 這種綜合監督機制將涵蓋大數據研究的各個階段，並涉及所有相關利害關係人

# 大數據研究之倫理審查之改革——配套改革

---

- ❑ 期刊和出版業可以在這個混合生態系統中發揮作用，限制潛在的雙重使用的關切
- ❑ 例如，在研究發表階段，可以將資源分配給編輯，以評估研究誠信標準並僅推廣那些符合倫理的項目
- ❑ 然而，只有在監督生態系統內對最佳作業達成共識時，這些實施才能產生影響
- ❑ 另一個選擇是在IRB之外設立專門且獨特的倫理委員會，其目的是評估大數據研究並為研究人員提供部門認證
- ❑ 在這個模式中，IRB不會因為需要審查的研究提案數量而不知所措，並且可以外包需要大數據領域專業知識的評估

# 大數據研究之倫理審查之改革——配套改革

---

- ❑ 事實上專門委員會（資料安全監測委員會、資料存取委員會以及負責的研究和創新小組）早已存在，並支持大數據研究人員確保資料保護（例如系統安全、資料儲存、資料傳輸）
- ❑ 然而，像「資料審查委員會」是可以評估研究對個人和社會的影響，同時審查專案的技術特徵
- ❑ 同儕審查可以在這個模式中發揮關鍵作用：研究界保留進行倫理研究所需的專業知識，並在路徑不明確時相互支持
- ❑ 儘管可行，但這些場景都各至少有一個主要限制
- ❑ 走在前面者在試圖將不同利害關係人的優先事項和倫理價值整合到共同的研究規範中時，可能會面臨強烈反對

# 大數據研究之倫理審查之改革——配套改革

---

- ❑ 雖然分散的監督方法有助於解決難題，但也可能因過於分散而降低效率及導致程序混亂，同時增加工作量
- ❑ 例如，美國境內的研究監督委員會成倍增加，導致委員會之間的冗餘和不和諧
- ❑ 此外，當愈來愈多的研究將是大數據研究時，與當前IRB並行工作的專門大數據倫理委員會可能會引發對傳統IRB角色的質疑

# 報告大綱

---

- 前言
- 大數據研究
- 大數據研究之倫理考量
- 大數據研究之倫理審查
- 結語

# 大數據研究之倫理審查之改革——結論

---

- ❑ IRB在大數據研究背景下面臨一些挑戰
- ❑ IRB在大數據研究中顯得固有缺點的加劇及出現新生的缺點
- ❑ 儘管有這些缺點，當前的監督模式仍然能夠保證大部分研究的倫理評估，但仍迫切需要有新的倫理準則和新的倫理評估工具，以保護社會免受大數據研究帶來的新風險所影響
- ❑ 建議IRB：
  - ❑ IRB審查能力之提升：①委員接受大數據倫理方面的培訓；②納入具備專業知識之委員；③接受外部專家的諮詢
  - ❑ IRB讓研究者和資料主體共同參與大數據研究的評估
  - ❑ 承認其他形式的監督提供協調和補充支持行動的現有空間

# 大數據研究之倫理審查之改革——結論

---

- 然而，相關參與者必須對最佳作業和評估標準達成共識，以便有效補充現有的監督機制
- 這些適應性建議可以使IRB機制足夠靈活且配備精良，以克服數據密集型研究挑戰並造福整個研究