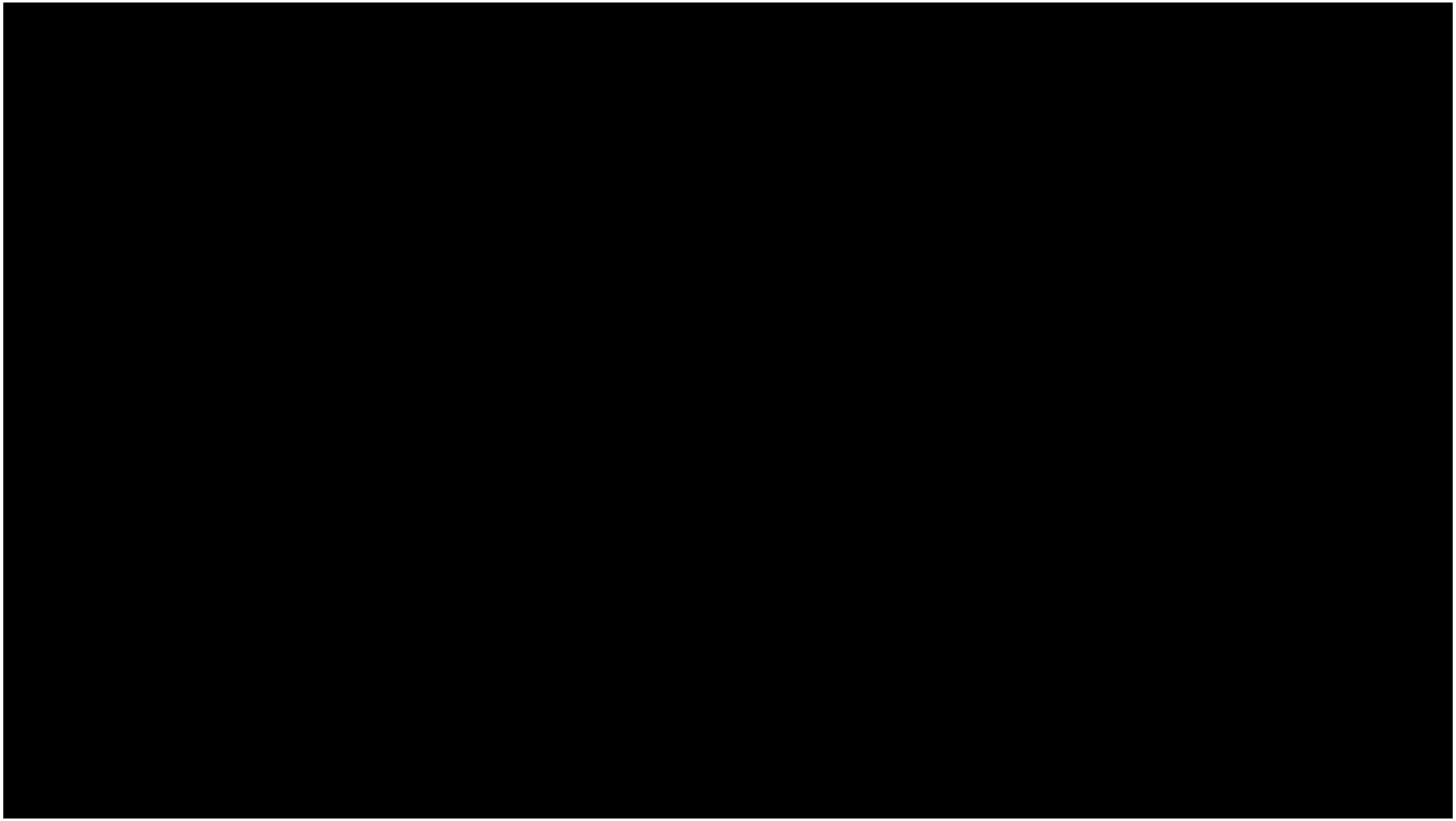
Big Data資訊安全現況與新型態威脅因應

大綱(Agenda)

- Big Data簡介
- 資訊安全概念與Big Data資安議題
- Big Data資安威脅與因應建議

大數據基本認知



HUAPLUS 2018年7月20日

巨量資料處理4V特性

Volume

- 處理大量的資料

Variety

- 處理結構化、非結構化等各種型態的多樣性資料

Velocity

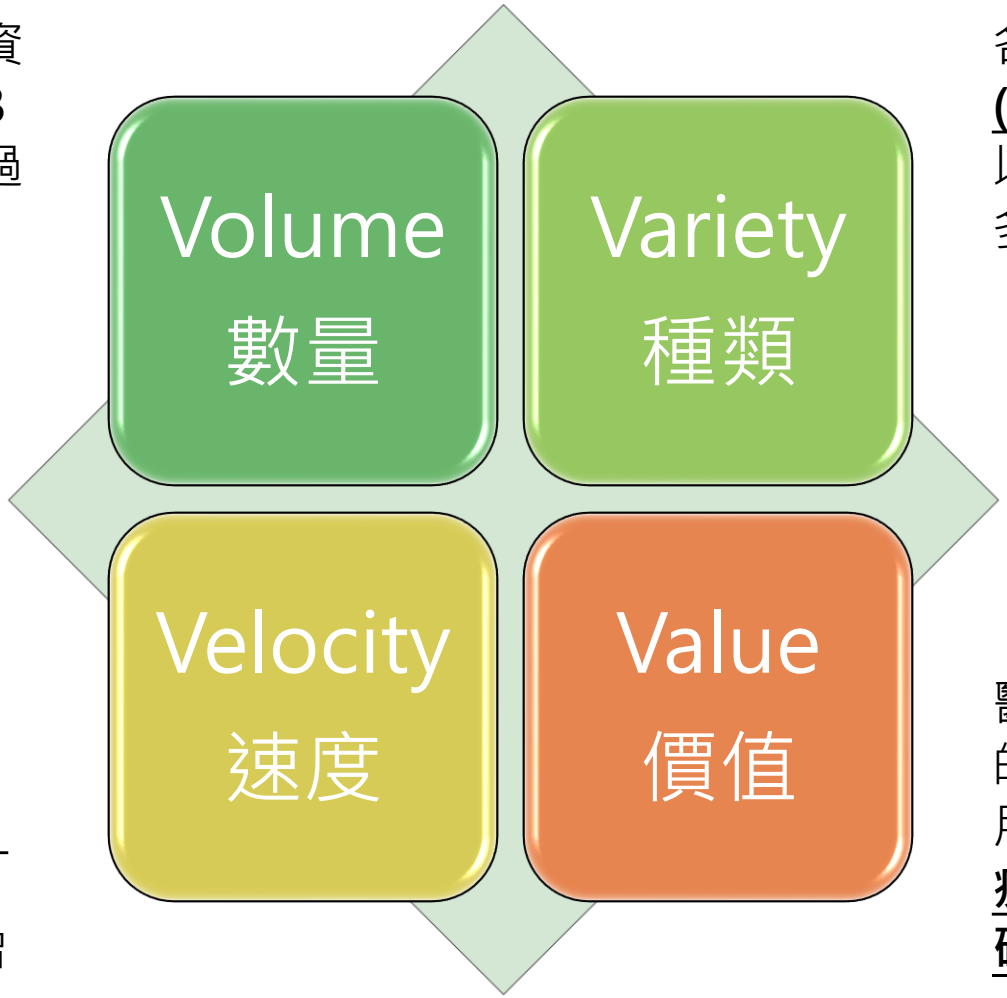
- 處理不同速度需求的資料

Value

- 權衡成本與價值

巨量醫療資料的特性

2012年全球醫療資料的規模為500PB →2022年後已超過**25000PB**



各種**結構化**與**非(半)結構化**資料以及**醫療影像**等多種多樣的資料

靜態醫療資料+**即時資料**不斷增加
重大疾病的爆發、氣候異常等等
導致醫療資料激增

醫療資料不僅與我們的生活息息相關，更用於國家甚至全球的**疾病控防、新藥物的研發**等等

巨量資料技術架構

- 基礎建設(Infrastructure)：伺服器、網路、儲存設備、虛擬監督器等
- 資料組織與管理(Data Organization and Management)：進行蒐集、整理、轉換、整合資料等處理工作的軟體
- 分析(Analytics)：將處理好的資料進行分析與視覺化的呈現
- 應用與服務(Applications and Services)：應用在適當的行業及應用情境

巨量資料基本分析類型

- **描述性分析**
  - 瞭解目前發生什麼事，將大量處理後數據進行概況、整理。
- **診斷性分析**
  - 瞭解問題發生的原因，分析數據的特徵，得出導致結果發生的原因。
- **預測性分析**
  - 預測未來事件，或發生問題的可能性。
- **最佳化分析**
  - 企業建立最佳化模型並自動化決策。

關聯式分析

發現資料子集

或資料屬性

的連結關係

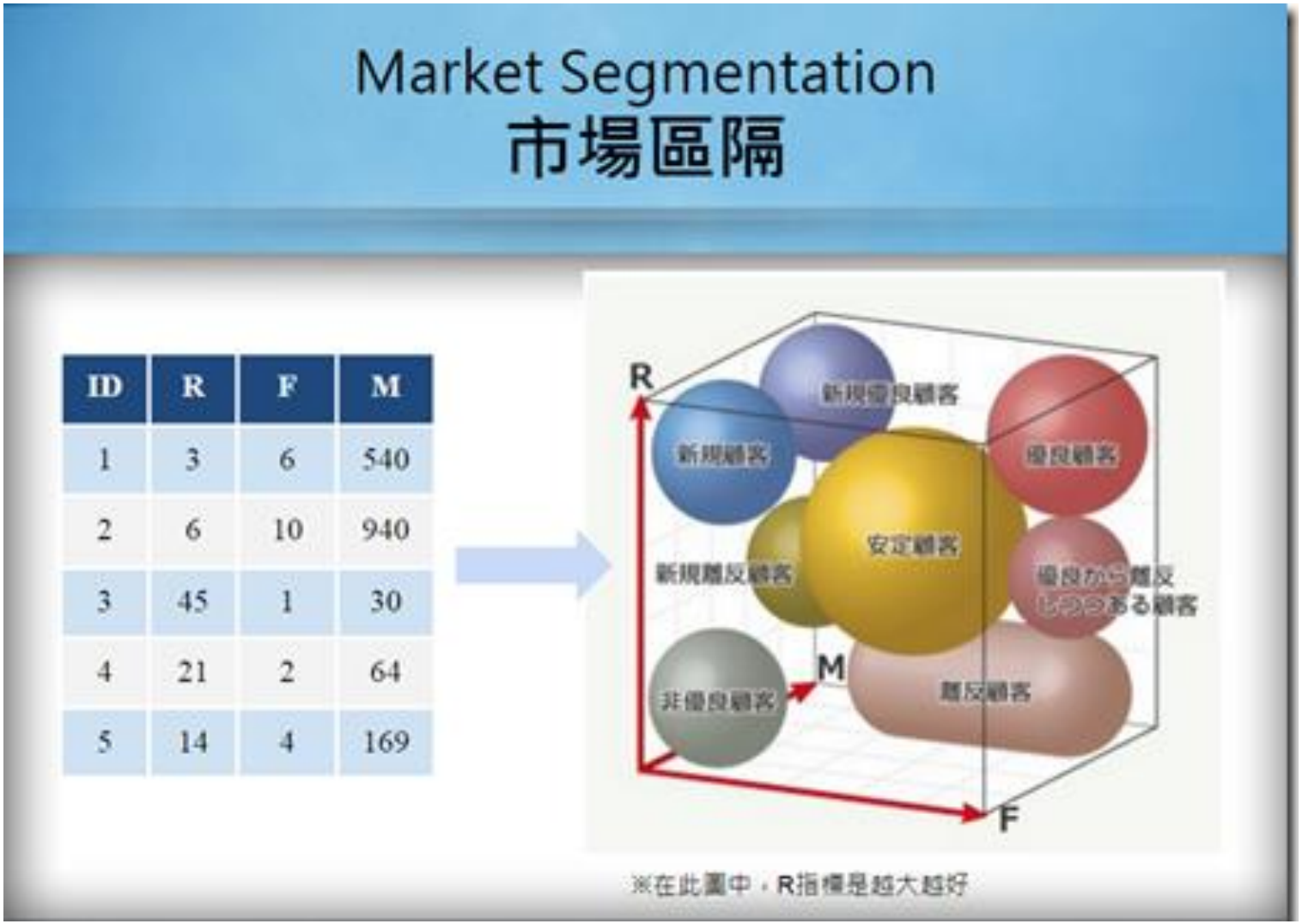
時間序列分析

根據歷史資料

找出變化規律

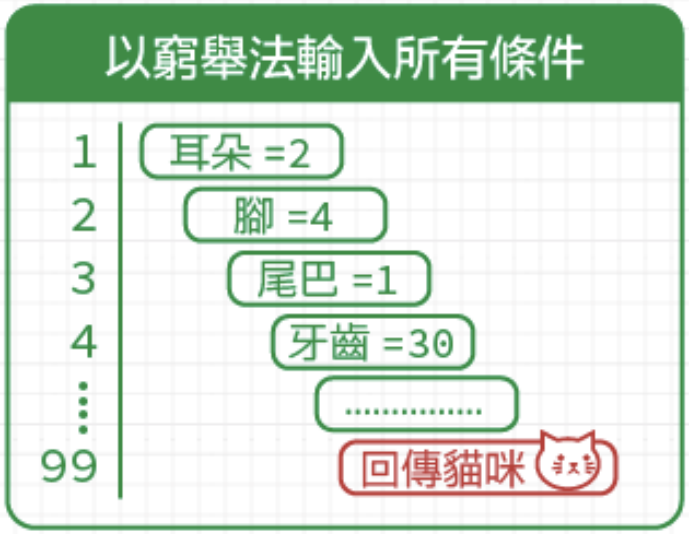
預測未來資料屬性的值

分群分析 Clustering



分類預測分析 Classification

機器學習概念



視覺化的結果

文字雲


東區
PTT
分鐘
民進黨
北屯
保護
天然氣
系統
花蓮
五分
分身
綠豆湯
律師
國外電腦
鄉民
OwO
執政
相信
吹冷氣
附近
滑鼠
HTC
哥哥
開除
國民黨
討論
發錢
北區
保證
貓貓
麻煩
行政
發電機組
嘻嘻
風向
關掉
士林
斑斑
山豬
前輩
正常
居民
智障
代表
錯誤
繼續
屏東
養媳
軟糖
警察
地區
核電
火力
蘋果
鍵盤
嚴重
綠豆
替死鬼
缺電
先知
瞬間
答案
危機
故意
員工
gg
呵呵
另外
ww
中華
媒體
損失
掰掰
雖然
電風扇
幾天
昨天
00
故障
me
遊戲
無腦
台南
民進
爽爽爽
疏失
wwww
全部
汐止
供氣
爽爽
文山
理由
回報
ptt
湖南
木栓
意外
松山
國民影響
準備
北部
檢查
誤操作
外面
選手
冬天
五十
反而
科技
xjp
市價
限電
屁事

網路購物行為分析

網路購物行為中，最令網路店家扼腕的是「不知道誰來過」，尤其是曾把商品放入購物車，卻沒有結帳的消費者，更是商家最想接觸的人。

AI應用無極限!

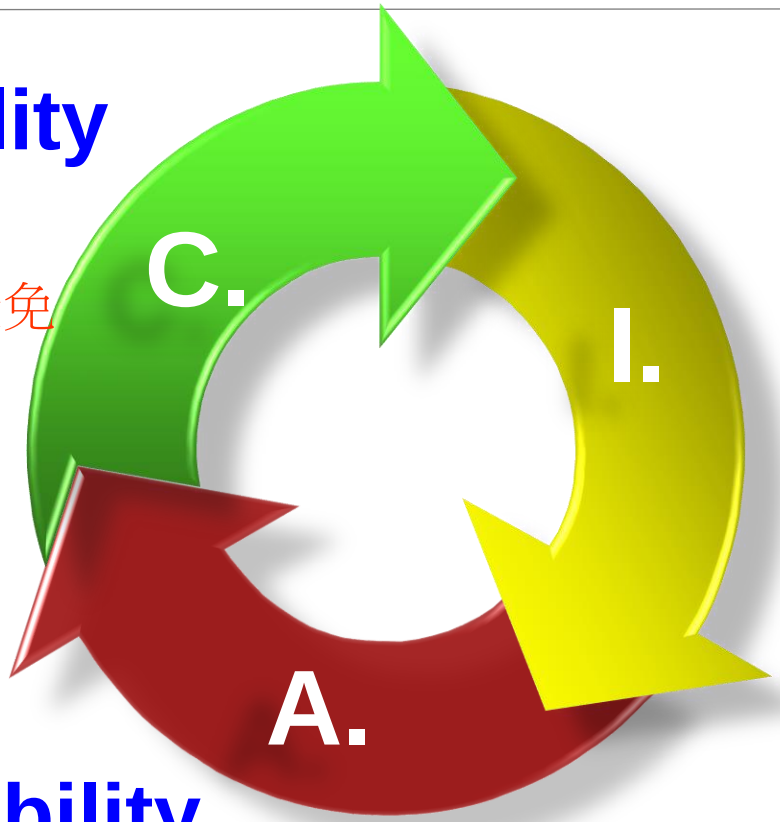

寰宇新聞
台北市
AI應用無極限！ 小巧智能音箱掀起飯店人力革命
財經訊息
就業報告超優！ 道指大漲701. 19點 那指數連漲6周

大綱(Agenda)

- Big Data簡介
- 資訊安全概念與Big Data資安議題
- Big Data資安威脅與因應建議

資訊安全三大要素

Confidentiality
(機密性)
確保資訊隱密性並避免遭到非法存取



Integrity
(完整性)
確保可提供正確與完整的資訊

Availability
(可用性)
確保可適時提供可用及正確之資訊

其他安全性要求：
- 不可否認性
- 身分鑑別
- 存取權限控制
- 可歸責性

大數據時代？

Google侵犯用戶隱私 面臨美國數州起訴

自由財經 2022/01/25

《路透》報導，Google遭德州、印第安納州、華盛頓州、哥倫比亞特區起訴，因Google以欺瞞方式收集用戶定位資料，侵犯了用戶隱私。

華盛頓特區檢察長拉辛（Karl Racine）聲明表示：「Google錯讓消費者相信，只要變更帳戶和設備設定就可以保護客戶個人隱私，還可以控管Google取得個人數據。」他認為，Google持續系統監控客戶，利用客戶資料牟利，這種做法明顯侵犯消費者隱私。

Google發言人卡斯塔尼達（Jose Castaneda）表示：「我們一直在產品中內建隱私功能，並為定位數據提供強大的控管，我們將大力捍衛自身權利。」

網路社群時代 個資大數據值錢

iPhone用戶小心！
你買房預算跟經期 App傳給臉書了

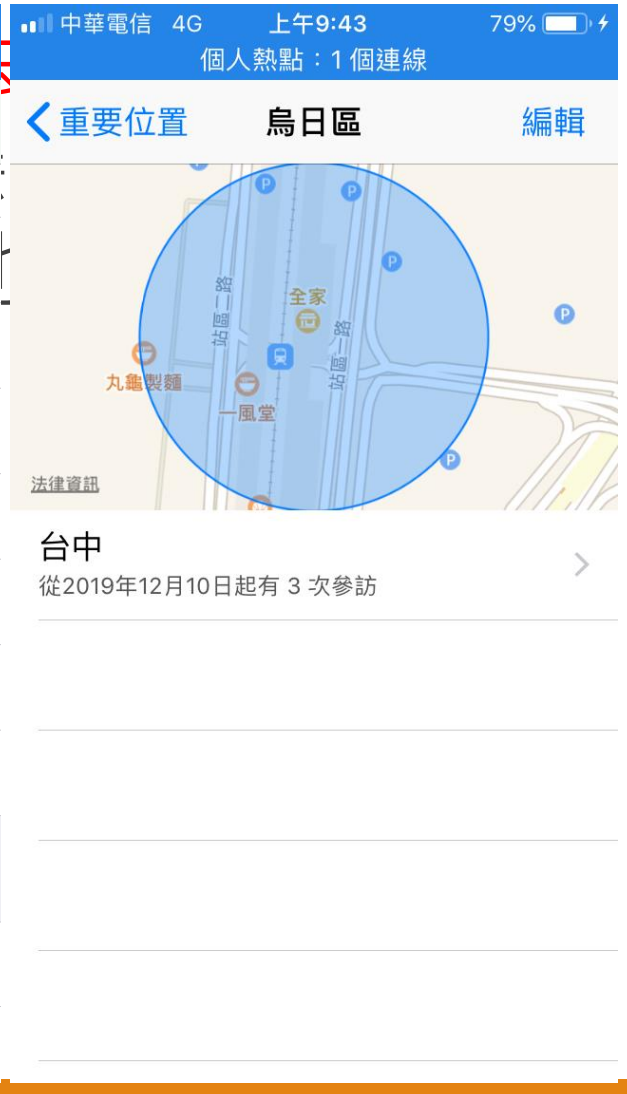
聯合新聞網 2019-02-24

手機用戶現在可能要擔心一件事情是：自己的私密資訊，如你的體脂肪、月經週期、懷孕狀態、甚至是買房子的預算，都有可能已經傳送到Facebook作為一系列的數據分析了。

據華爾街日報報導，他們針對iOS上70個最受歡迎的應用程式進行研究，居然發現有11個app向Facebook發送使用者個資的數據。

而這些應用程式開發商之所以分享數據是因為他們能夠利用Facebook的分析工具來了解自己應用程式的使用者各種喜好；並且透過廣告或促銷來吸引並定位這些使用者。

你知道iPhone一直偷偷在記錄你的位置嗎！



人臉辨識雙面刃用之不當恐引社會危機

民視新聞網 2019-2-12

防遭人臉識別紀錄 陸現戴全罩式頭盔看房

想監控又出招? "步態識別"技術

監聽通訊軟體? 重罪犯研擬植"木馬"蒐證

因應未來威脅，要先掌握資安關鍵議題

- 了解網路中誰持有哪些資料，且持有多久？
- 有哪些人濫用或合法使用資料？
- 許多資料使用是否與原先的使用目的相符？
- 執政當局提供的資料存取方式，可否確保資料安全？
- 如果資料有損失，由誰承擔資料損失的風險？
- 有誰提供金錢或資料復原的補救措施呢？
- 由誰負責網路、服務和應用程式之間的安全性？
- 哪些技術可確保網路、服務和AP在執行環境中的安全？

大綱(Agenda)

- Big Data簡介
- 資訊安全概念與Big Data資安議題
- Big Data資安威脅與因應建議

2023資安趨勢

iThome 新聞 產品&技術 專題 AI Cloud 醫療IT 資安 研討會 社群 IT EXPLAINED

2023資安趨勢

資安長 | CISO | 2023資安趨勢

【2023資安趨勢5：CISO】有獲利的公司，年底前都要設立資安長

法遵要求是所有資安長立足根本，但要懂得和公司業務連結，讓公司活下去；掌握資安治理和技術發展，公司有韌性

2023-01-16

2023資安趨勢 | 零信任 | 身分安全 | 可信任供應鏈 | 抗網釣MFA | Phishing-Resistant MFA

【2023資安趨勢1：零信任、身分安全、可信任供應鏈】零信任網路安全步入廣泛實踐的道路

身分識別成首要注目焦點，聚焦可抵抗網釣的MFA，其他領域也在探索ZTA可

2023-01-16

2023資安趨勢 | 軟體供應鏈安全 | 惡意開發軟體套件 | 程式碼外洩

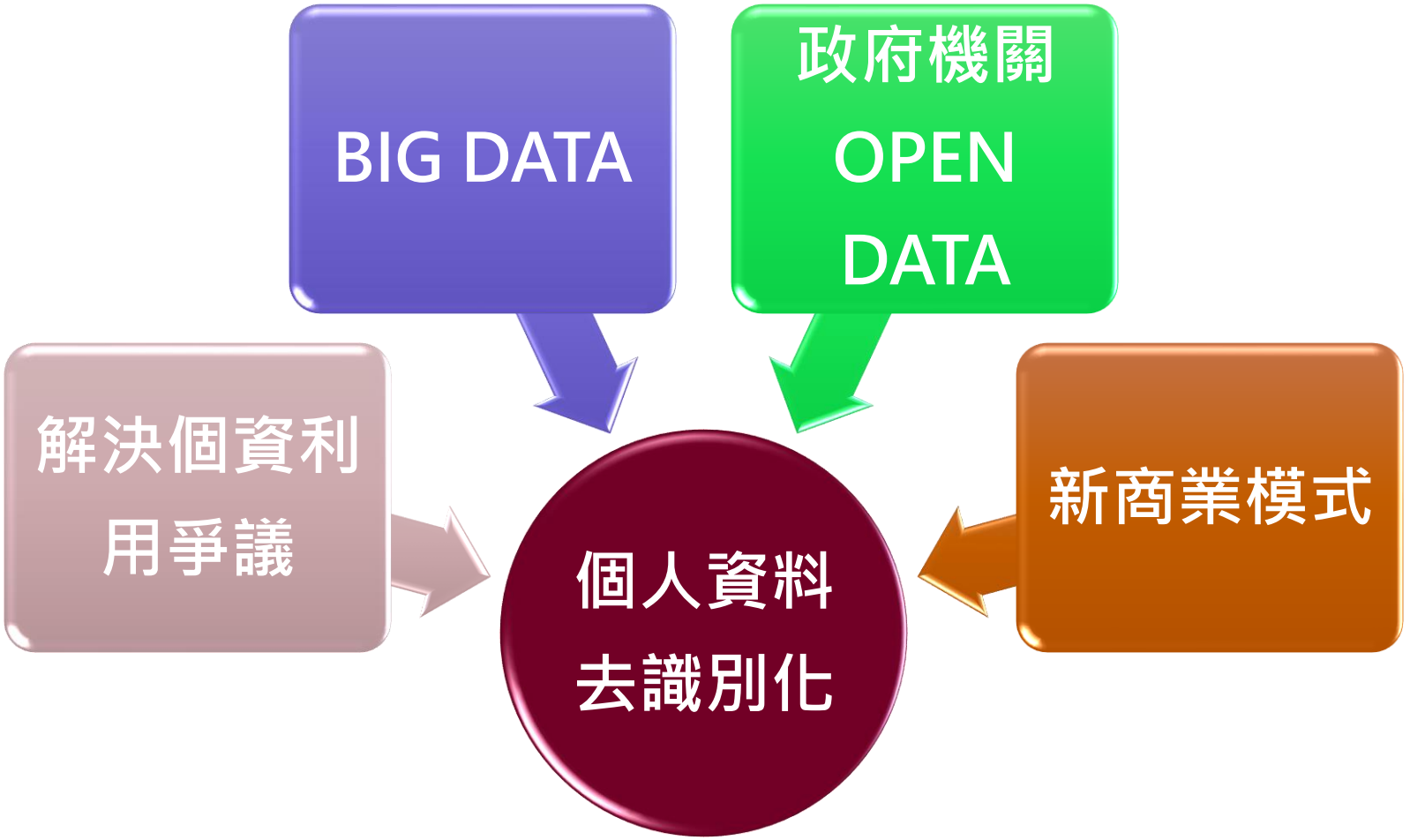
【2023資安趨勢3：軟體供應鏈安全】防堵惡意開發軟體套件與程式碼外洩成當務之急

積極發掘與修補軟體資安漏洞之餘，須設法因應惡意套件氾濫與不當設定造成的機密外洩

2023-01-16

<Source: iThome>

Big Data的隱私保護與資料去識別化



個人資料保護法判例(一)

健保局從1998年起委託**國家衛生研究院**建置「全民健康保險研究資料庫」，在未經個人同意的情況下，將個人健保資料身分欄加密後，轉提供研究使用，並以讀取1G資料200元至每張光碟片2500元不等之價碼收取工本費用。

2011年開始，衛生署進一步在「國民健康資訊建設計畫」的架構下，委由中國醫藥大學設立「**健康資料加值應用協作中心**」，將健保局、國健局、疾管局的國人健康資料，與其他公務機關為不同目的而蒐集的資料（例如戶籍資料、原住民檔、身心障礙檔、家庭收支調查檔等）串聯整合，以探挖原始資料下埋藏的豐富資訊。

個人資料保護法判例(二)

台灣人權促進會、台灣女人連線、民間健保監督聯盟 聯合聲明

長期以來，健保署在違反的業務使 用目的又無其他法令授權，亦無徵得當事人 同意的情況下，將全民健保資料釋出給「健 康資料加值應用協作中心」及「全民健保研 究資料庫」，透過這些機構再將全民健保資料出售給各研究單位、藥廠等。

2012年6月26日，台灣人權促進會以及關心 隱私權及個資保護的團體及個人，針對此事 以存證信函主張行使退出健保資料庫的「目的外利用」，遭到健保局拒絕。

個資利用爭議

- 本案的爭議點之一，就是提供研究的健保資料的去識別化是否已達到「無從識別特定當事人」的程度
- 人權團體蔡○○等8員主張健保署(原健保局)、衛福部、國衛院所謂經去識別化的資料還是有直接或間接識別的可能
- 健保署、衛福部、國衛院則辯稱資料已經去識別化，依法務部函釋意旨，並非個資法上的個資

最高行政法院106年度判字第54號

資料是否屬「個人資料」之判斷，為資料內容之**「去識別化」**作業是否完成？

- 本院發現**本案去識別化作業模式，實不足到達「完全切斷資料內容與特定主體間之連結線索」之程度，因此移轉至輔助參加人衛福部之健保資料，其「個人資料」屬性，尚未被全然排除（ 雖然已經大幅度降低）**，因此仍有繼續檢討該資料收受者輔助參加人衛福部對被上訴人所持有之健保資料有無蒐集處理職權之必要。
- 從被上訴人與輔助參加人衛福部自承之「去識別化」作業模式觀之，由輔助參加人衛福部派專人來執行「加密」作業，再攜回加密之個人資料建置資料庫。如此作業模式即表示輔助參加**人衛福部本也有「還原」資料與主體間連結之能力，此等結果顯然與由被上訴人「單方」掌握「還原」能力之「去識別化」標準不符。**

最高行政法院106年度判字第54號

- 被上訴人雖抗辯稱「『原始資料』仍然保留在被上訴人手中」云云，但此等抗辯毫無道理，因為:**資料實際上是一種訊息，應與訊息載體分離看待。而所謂「原始資料」其實僅是儲放資料（訊息）之載體，因此載體即使留在原處，載體內之訊息已透過轉換載體之方式，而脫逸被上訴人之控管，並為輔助參加人衛福部所控管，只要輔助參加人衛福部之內部單位公務員，有「還原」資料與主體連結之可能，對輔助參加人衛福部而言，該資料仍未「去識別化」，而屬「個人資料」**。
- 因為該等資料解密鑰匙，仍由輔助參加人衛福部內負責資訊秘密業務單位之公務員保有，而非輔助參加人衛福部內之任何成員均可取閱，則該等資料與特定主體連結之「可識別性」已大幅度降低，**但只要輔助參加人衛福部內部單位成員，有還原資料與主體連結之能力，即不符合「去識別化」之標準**。

去識別化只是用小毛巾遮臉？

2019-03-27 02:21聯合報

有個笑話：某甲洗澡時陌生人闖進來，某甲手上一塊小毛巾，不知要遮哪，就遮臉好了。邏輯是：至少被看光後，不知道光著身子的人是誰。

中研院資科所副研究員莊庭瑞指出，政府對公眾資料在自行聲明「已去個人識別處理」後，就轉為「特定目的以外的利用」。這作法實在太寬鬆。

更重要的是，目前去識別化作法仍可回溯追出你的身分。前立委黃淑英等人曾由健保資料庫申請出某群資料，在無姓名等資料下，仍能辨識出自己的健康資料。

台權會曾拿到研究後棄置的醫療個資光碟，發現**去識別化只是把姓名跟身分證字號加密**，其他資料「全都露」；再者，被抽出的民眾「根本沒有被告知將被研究，更別說同意了，就連退出的權利都沒有」。

Open Data VS 侵害個人隱私

行政院會決議指示

- 以開放為原則、不開放為例外
- 以免費為原則、收費為例外

法源依據

- 個資法規定要「去識別化」、無法識別特定當事人
- 識別：**直接**或**間接**方式識別該個人之資料

為何要做去識別化？

兼顧公益與隱私保護

- 公益
  - 醫院之間交換病人資料，可以減少重複的檢查
  - 公衛學者取得民眾就醫的資料，可以分析疾病的趨勢，並且協助政策訂定
  - 研究者找出患有某些疾病的人的共同特徵或基因，有助於解決該項疾病
- 隱私保護
  - 萬一老闆知道你得了某些疾病，可能不聘用你或想辦法解聘
  - 萬一醫生知道病患得了某些疾病，可能採取消極的作為

滿足法規要求

- 有些資料法律規定要釋出或是要收集，必須要匿名

行政院及所屬各級機關政府資料開放作業原則

第五條 第二款

開放資料有下列情形之一者，經機關首長核可，並於政府資料開放平臺公告停止提供及其理由後，得轉為依申請提供資料或不予提供資料：

(一)因情事變更或其他正當理由，致繼續提供該開放資料供公眾使用，不符合公共利益。

(二)有侵害第三人智慧財產權、隱私權等權利或其他法律上利益之虞。

https://theme.ndc.gov.tw/lawout/LawContent.aspx?id=GL000028

行政院及所屬各級機關政府資料開放作業原則

第十五條
各機關辦理政府資料開放，應依個人資料保護法及資通安全管理法等相關規定，辦理個人資料保護及資訊安全管理作業。

資訊安全管理系統 (ISMS)

個人資料去識別化管理制度 (PDMS)→
個資去識別化過程管理系統(personal information de-identification process management system, PIDIPMS)

https://theme.ndc.gov.tw/lawout/LawContent.aspx?id=GL000028

去識別化法源問題

各國法律對於具個人屬性資料處理至何程度（可被接受的風險）才不受法律規範的用語與定義內涵不同

- 美國HIPAA：De-identification
- 歐盟GDPR：Anonymization
- 我國個資法：**無從識別**特定（個資）當事人

歐盟

- **歐盟採用「匿名化」（anonymization）**
  - 各國雖對匿名仍有不同理解，但對於匿名處理後可被接受的風險具有共識
- **資料保護規則（GDPR）**
  - 聚焦於技術處理後之**資料狀態**，而未規範匿名化工具、標準與操作
  - 經匿名化之資料係指已達**無法連結（relate）特定個人**
    - 無法**SINGLE OUT** 個人
    - **以一切合理方法（all the means reasonably to be used）都無法與特定個人連結**的程度
- **歐盟Art 29 WP 第216號有關匿名化技術意見書**
  - 資料經處理後是否達到匿名化狀態之三項判斷依據（風險）
    - 是否仍可能**識別特定**個人（ “single out” an individual? ）
    - 是否仍可**連結至特定個人**相關紀錄（link records “relating” to an individual? ）**及**
    - 是否仍得從相關資訊**推斷至特定個人**（can information be “inferred” concerning an individual? ）

美國

- **原則禁止使用或揭露當事人之健康資訊（protected health information, PHI），除非有法律明文規範之以下目的，可不需當事人授權（authorization）：**
  - 提供予資料所有人、係基於診斷、付費及照護措施（treatment, payment, and health care operations, TPO）之目的、有提供當事人表示反對之機會…
  - 透過去除下列18 種個人資訊達成去識別（de-identified）

- 姓名
- 地理資訊
- 各種日期資訊
- 電話號碼
- 傳真號碼
- 電子郵件地址
- 社會安全號碼
- 病歷號碼
- 醫療計畫或健保號碼
- 帳號
- 各種證照編號
- 車牌、車籍資料
- 設備編號或序號
- 網路位址（URLs）
- IP位址
- 生物辨識資料（如指紋、聲紋）
- 臉部照片
- 其他可識別個人之編號或特徵

但在當事人第一次就醫並使用該機構提供之健康照護服務前，機構有義務以書面（notice letter）通知當事人資料可能之使用方式與對象（換言之，當事人似仍有選擇退出之機會）

台灣

法務部

- 法務部法律字第10303513040號函釋
  - 以運用技術去識別化而呈現方式**已無從直接或間接識別特定個人**，作為判斷資料是否屬個人資料之依據
- 法務部法律字第10503505760號函檢送「公務機關利用去識別化資料之合理風險控制及法律責任」之研析報告，**可採取「可逆之擬匿名化資料」方式進行去識別化**
  - 擬匿名化資料乃是以編碼或別名取代識別符（例如姓名、國民身分證統一編號等），使研究或統計人員得以針對個體資訊進行分析而無須識別個體身分，**且可回溯追蹤特定人**
- 「可逆之擬匿名化」與「代碼」是否不同？？

爭取歐盟GDPR適足性認定

爭取歐盟GDPR適足性認定 台歐盟展開技術性對話

2018-06-04 17:08

陳美伶說，GDPR號稱為史上最嚴格個資法，對於個資的輸出一定要獲得「明確同意」，且除非個資通通去識別化，使用代號或是假名都不符合規範，儘管歐盟沒有要求一定要修改當地法令，但在個資保護上必須符合歐盟要求，台灣個資法與歐盟1995個人資料保護指令大約同時，在個資輸出上並沒有明確規範，未來勢必要進行檢討，甚至不排除修法。

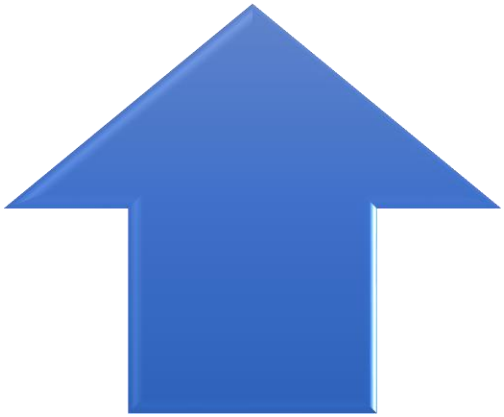
國發會主委陳美伶今日說明與歐盟達成開啟GDPR適足性認定工作（國發會提供）

個人資料再利用之需求與問題

隱私保護

法定職務
資訊安全

促進公益



學術研究
大數據分析

行為者及角色

PII當事人(PII principal)

- 個人可識別資訊(PII)所關聯之自然人

PII控制者(PII controller)

- 判定個人可識別資訊(PII)處理之目的及方法的隱私權相關者，而非就個人目的使用資料的自然人
- 備考：PII控制者有時委派他人(例：PII處理者)，代表其處理PII，而處理之責任仍由PII控制者承擔

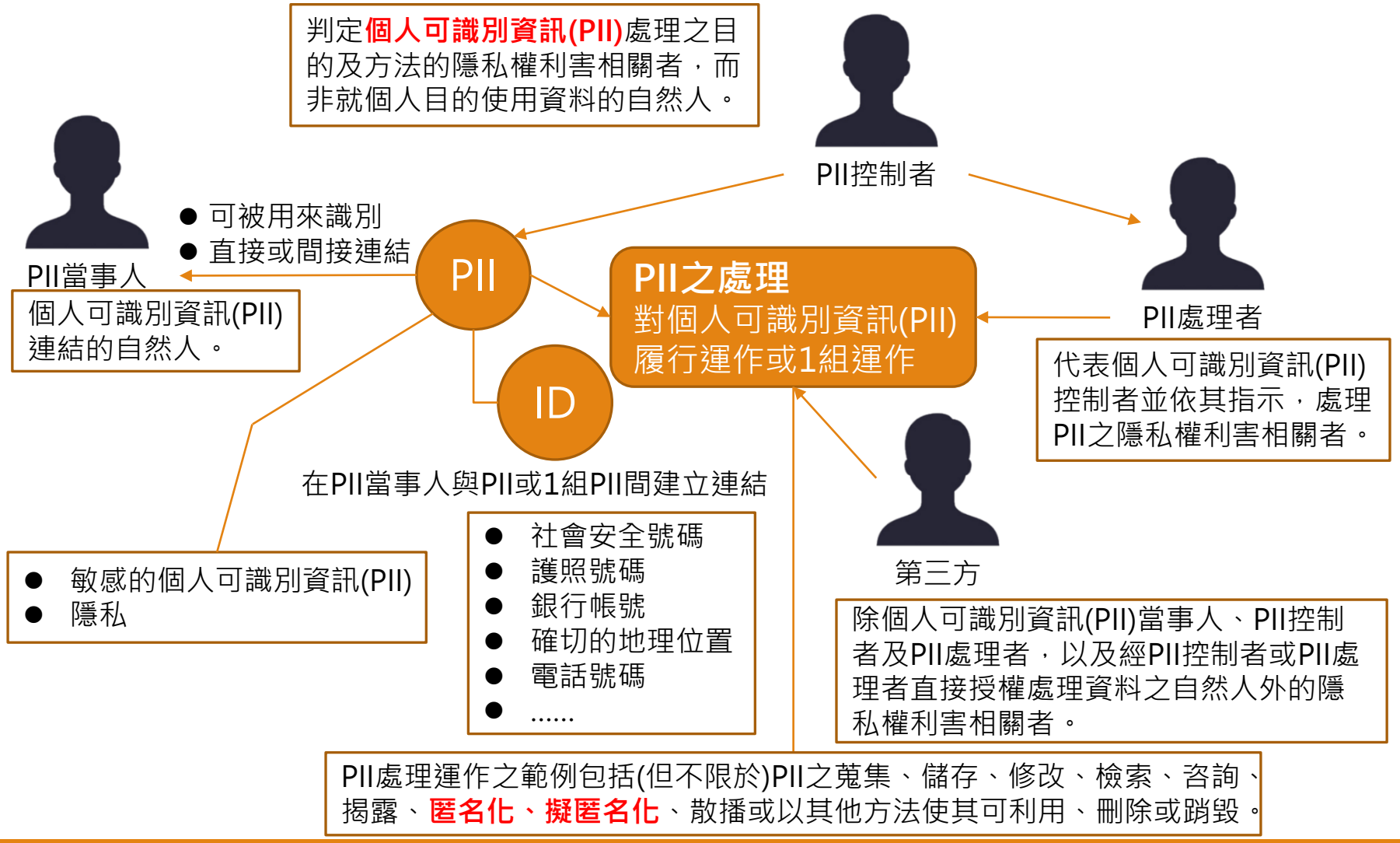
PII處理者(PII processor)

- 代表PII控制者並依其指示，處理個人可識別資訊(PII)之隱私權相關者

第三方(third party)

- 除PII當事人、PII控制者及PII處理者，以及經PII控制者或PII處理者直接授權處理資料之自然人外的隱私權相關者

PII之角色關係


判定個人可識別資訊(PII)處理之目的及方法的隱私權利害相關者，而非就個人目的使用資料的自然人。
PII控制者
PII當事人
個人可識別資訊(PII)連結的自然人。
● 可被用來識別
● 直接或間接連結
PII
PII之處理
對個人可識別資訊(PII)履行運作或1組運作
PII處理者
代表個人可識別資訊(PII)控制者並依其指示，處理PII之隱私權利害相關者。
ID
在PII當事人與PII或1組PII間建立連結
● 敏感的個人可識別資訊(PII)
● 隱私
● 社會安全號碼
● 護照號碼
● 銀行帳號
● 確切的地理位置
● 電話號碼
● .....
第三方
除個人可識別資訊(PII)當事人、PII控制者及PII處理者，以及經PII控制者或PII處理者直接授權處理資料之自然人外的隱私權利害相關者。
PII處理運作之範例包括(但不限於)PII之蒐集、儲存、修改、檢索、咨詢、揭露、匿名化、擬匿名化、散播或以其他方法使其可利用、刪除或銷毀。

常見去識別化技術

常見的【經驗法則】方式 (1/2)

- **對於區域的處理**
  - 例如 HIPAA Safe Harbor 中，要求 5 碼郵遞區號若人數超過兩萬人，可以只顯示前 3 碼。如果是不滿兩萬人，則顯示 000。
- **Top-Coding**
  - 如果資料超過一個數值，則只顯示到那個數值
  - 例如超過九十歲的人只把年齡顯示到九十歲
- **Threshold N+**
  - 針對統計的結果一定要超過一個數字

常見的【經驗法則】方式 (2/2)

- **(n, k) Rule**
  - 如果有少數的 n 個人佔據了全部的 k%，此時可能需要把這些異常值排除

- **降低精密度**
  - 例如不公布詳細的日期，只公布到年，或是身高等數字要四捨五入等。

學界對個人資料之隱私保護

隱私保護資料探勘（Privacy Preserving Data Mining, PPDM）

隱私保護資料公開（Privacy Preserving Data Publishing, PPDP）

保護隱私資料不因數據分析或公開而不當揭露

差分隱私(differential privacy)

針對資料集查詢注入隨機「雜訊」，以保證於數學上資料集之中的任一當事人個資之存在，將被遮蔽之方法。

- 備考：差分隱私係個資去識別化方法之一類。通常需以軟體估算查詢結果之隱私風險，並決定於釋出資料前，應注入查詢結果之雜訊等級。

蘋果的差分隱私

差分隱私是統計和數據分析領域的一個研究課題，指使用散列、子採樣和噪聲注入等方式，在每個用戶的信息仍然完全保密的情況下，使眾包形式的學習成為可能。

差分隱私的原理是用算法加擾個人用戶數據，使之無法回溯到個人，然後對數據進行批量分析，得出大規模的趨勢規律。其目標是保護用戶身份信息和數據細節，同時提取機器學習所需的通用信息。

https://kknews.cc/tech/38xqlyy.html

去識別化技術比較

	是否仍存在被識別風險？	是否仍存在關聯風險？	是否仍存在推論風險？
擬匿名化 (pseudonymization)	是	是	是
添加雜訊(noise addition)	是	不太會	不太會
替換(substitution)	是	是	不太會
聚合或*k*-匿名(aggregation or *k*-anonymity)	否	是	是
l-多樣(*l*-diversity)	否	是	不太會
差分隱私	不太會	不太會	不太會
雜湊化/符記化 (hashing/tokenization)	是	是	不太會

資料來源：European Commission, Opinion on Anonymisation Techniques

ChatGPT議題-1

網路溫度計

乾脆拜AI為師！教授用ChatGPT寫論文評語被抓包 女大生笑死：反正我也用AI寫的

Dailyview

2023年5月17日 週三 上午9:24

image source:截自TikTok

文 / 李羿嫻

ChatGPT議題-2


TVBS新聞台 HD
更多新聞在這裡
企業新寵!ChatGPT增工作效率 導入有利有弊
掌握新聞脈動 ▶ 訂閱TVBS NEWS頻道

AI偽造問題...


TikTok
@deeptomcruise
Tom Cruise
TVBS新聞台 HD
普遍級
更多新聞在這裡
恐怖! 3秒能仿人聲 詐騙電話恐"電腦講的"
掌握新聞脈動 ▶ 訂閱TVBS NEWS頻道

AI安全議題

除蘋果外，摩根大通及美國威瑞森電信（Verizon）也禁止員工使用ChatGPT。亞馬遜也限制員工使用ChatGPT，

<Source: iThome>

AI可以做這些事嗎？

- 幽默感...
- 善意的謊言...
- 開玩笑...
- 任性...
- 犯錯...

廣島G7峰會
呼籲為AI發展研擬國際技術標準

廣島G7峰會 呼籲為AI發展研擬國際技術標準

路透社 中文新聞

2023年5月20日 週六 下午8:19

（路透東京20日電）七大工業國集團（G7）領袖今天呼籲，為了發展值得信賴的人工智慧（AI），應開發並施行一套國際技術標準。一些富裕國家的國會目前都針對這項新科技展開討論。

G7領袖正在日本廣島舉行峰會，他們在聲明中表示，為了達成「能有值得信賴的AI技術這項共同的願景與目標，或許會有不同做法」，但「對於數位經濟的管理，應該依照我們共有的民主價值與時俱進」。

該做的是什麼？

- 制定標準？
- 遵循標準？
- 期望法規？
- ….

是人在運用科技？還是人被科技左右？

聲控取代了一切的生活會是什麼模樣呢？

Thank You

Thank you