

2013年美國FICO巨量資料分析研討會考察紀要

丁正中、李彥錚 / 金融聯合徵信中心資訊部、研究部

2013年美國FICO巨量資料分析研討會係由美國最早成立且規模最大的信用評分公司FICO (Fair Issac) 於美國邁阿密舉辦，主題為巨量資料分析 (BIG DATA ANALYTICS AND THE CUSTOMER REVOLUTION)。本次會議FICO除了介紹其核心業務產品—FICO SCORE及相關評分產品之近況外，相當大的重心放在近兩年最熱門的資訊議題—Big Data (下稱巨量資料)，以及如何應用巨量資料進行分析 (包含信用風險、行銷、客戶關係管理等層面)，亦有探討資料隱私保護之議題。

FICO在世界上是評分卡的創立者，關於評分卡建置及應用之理論幾乎是由其一手建立，其將評分技術亦提供跨領域的服務，除了信用之外，尚包含行銷、醫療、保險等層面。

此次參加會議，藉由瞭解其信用評分產品的最新近況，可作為聯徵中心信用評分改版及會員應用端之參考；除此之外，藉由觀察FICO其餘各類產品之內涵及應用，亦可探討開發新產品的空間。

而在評分領域之外的，則是本次會議的核心主題：巨量資料 (Big data)，隨著資料蒐集、儲存、運算的科技日益發達，各大公司及

政府機關在實質上皆是一個個不同領域的大資料庫，在各個資料庫下，資料要如何應用才能發揮最大效益？更有甚者，如果各個資料庫間有統合的可能性，在這樣的跨領域大資料集下，是否有足夠的能力駕馭？又是否可以產生對社會具效益的應用成果？皆是當今世界對於資訊管理趨勢的重要議題。

聯徵中心身為金融體系不可或缺的資訊成員，係台灣唯一的信用資料庫，除了善盡職責控管資訊品質及資訊安全之外，亦應就資料加值應用層次的提升再深入探討。

會議內容

本次會議由美國信用評分界的先驅FICO (Fair Issac) 公司於4月30日至5月3日間在美國邁阿密舉辦，參與人數約為700~800人之間，包含40個國家以上之成員，參與者多為銀行及信用資料分析機構；若就泛太平洋地區而言，參與國家包括我國、中國大陸、南韓、新加坡、印度、澳大利亞；而研討會的核心主題為巨量資料之分析應用。

本次研討會共有3個主講題 (keynote sessions)，分別為：

- 巨量資料 (Big Data)
- FICO的科技願景 (FICO's Technology Vision)
- 改變客戶的恆等式 (Changing the Customer Equation)

另有11個講座主題，每個主題下皆有5-6場相關講座，11個主題的名稱如下：

- 債權管理及回收 (Collections and Recovery)
- 客戶獲利性 (Customer Profitability)
- 偽冒管理 (Fraud Management)
- 房貸放款 (Mortgage Lending)
- 法規遵循 (Regulatory Compliance)
- 分析技術創新 (Analytic Innovation)
- 信用評分 (Credit Scoring)
- 決策管理科技 (Decision Management)
- 行銷及客戶體驗 (Marketing and the Customer Experience)
- 風險管理 (Risk Management)

- 小型企業放款 (Small Business Lending)

由於主題眾多，僅將參與的議題和其重要內容進行整合，彙整出四個面向加以說明，分別為「巨量資料」、「FICO信用評分產品」、「跨領域分析」、「資料隱私及保護」，詳述如下：

(一) 巨量資料

巨量資料 (Big Data)，主要係指針對大量的數據中找出有用的資訊，它是目前資訊管理走向雲端化下的必然產物，當個人資訊不再儲存於自己的電腦，而是儲存在供應商 (包括網路服務、百貨、銀行業等) 時，那麼擁有那麼多資訊的公司，勢必會從自己的資料庫中，去找尋能超越其資料儲存成本的利益。相較於2000年初盛極一時的科學—資料採礦 (data mining)，巨量資料具有更即時且更迅速的回應，並且將分析的觸角延伸至跨領域資料庫以及非格式化資料 (Unstructured Data，如email及社群網站之資訊)。

本次研討會邀請了撰寫Big Data—A Revolution That Will Transform How We Live, Work, And Think (2013) 一書的作者Kenneth Cukier為巨量資料主題的主講者，並於該場講座結束後，發送親筆簽名的新書給參加研討會的各界人士。由於此書內容生動而有趣，相較講座有限的時間能更深入的說明巨量資料，因此亦摘錄此書內容的一部份，並就簡單幾個小主題做為巨量資料的釋例。

· 禽流感

2009年新型禽流感疫情開始快速蔓延，美國疾病管制局(CDC)開始要求醫生告知新型流感的病例，但是這種報送作業需要約2個星期，這樣的速度會讓公共衛生機構錯失防疫的先機。而Google在科學期刊Nature發表一篇論文，說明如何利用美國各地區使用搜尋引擎所產生的資料，預測禽流感疫情，提供公共衛生機構決定防疫及醫療資源分配的決策。

如何從每天3百萬的查詢中找出特定的搜尋與禽流感相關，Google利用美國疾病管制局在2007年到2008年實際發生禽流感案例，並比較數學模型預測結果，最後找出45個搜尋項目的組合，與實際發生禽流感案例有很強的相關性，利用這些相關性可以發現，特定區域在搜尋特定的項目組合，發生禽流感的案例會增多。Google利用運算及線上查詢的優勢，提供更快速而精確的預測，協助公共衛生機構掌握疫情資訊。

· 亞馬遜

亞馬遜(Amazon)如何根據個別的消費習性，推薦特定的書給消費者？剛開始，亞馬遜採取傳統的分析方法，從大資料抽樣後分析客戶之間的相似度，然而給出的建議卻很粗糙，例如購買了一項嬰兒用品，然後就會被推薦一大堆相類似產品。任職於亞馬遜的Greg Linden找到了解決的辦法，他瞭解到系統不應該將客戶做比較，而是需要找出產品之間的關連。使用全部的資料，並在事前計算好，可以在客戶購物時很快給出建議，結果也比先前成

功，例如喜愛A作者的讀者未必會喜歡A作者全部的書，但經由關連分析後卻可能會發現喜愛A作者的人多半同時會喜歡B作者的書，這樣的交叉分析將帶來更大的效益。

· 風險

Amazon觀察我們的消費習慣、Google瀏覽我們的搜尋、Facebook能找出我們的社會關係、行動電話業者知道我們與誰通話及所在位置，隱私保護的議題從網際網路普及後就持續成為關注的話題(隱私保護議題詳第四段)，雖然巨量資料時代的來臨，可以藉由大量資料進行預測，但巨量資料僅能預測一些特定的事或行為，仍無法對人類社會的全體作完美的預測。巨量資料是一種資源，也是一種工具，它代表資訊而非解釋，如果運用失當，它也可能誤導我們。所以巨量資料可以提供好的答案，但並非最終的答案。

(二) FICO信用評分產品

FICO為信用報告機構(credit bureau)評分模型的先驅，並與美國及加拿大5家信用報告機構為夥伴關係，至今已賣出超過1千5百億個FICO評分，超過7萬家企業依靠FICO評分運作，FICO評分可運用於所有信用產業授信決策，包含核准或婉拒、行銷與信貸組合管理(portfolio management)。以下簡介FICO的產品內容及其近況。

FICO的信用風險評分，係利用統計程序將信用申請人及現有帳戶的資料轉換成數字，轉換後的數字可代表違約機率的高低。其評分

屬於風險評分(Risk Score)；但其另外有非風險評分的產品包含損耗(Attrition)評分、回應(Response)評分及獲利(Revenue)評分。

FICO本身並未擁有資料庫，其係透過美國與加拿大境內的信用報告機構 (Credit Bureau) 所提供的資料來建立模型。評分過程中係由FICO團隊定義其欲分析的特徵 (Characteristic，即變數)，再透過信用報告機構所提供評分對象的屬性 (Attributes，即資料結果)，接著即由統計程序依據歷史違約率給予分數，與聯徵中心自行蒐集資料、研發模型、產生分數於一身的模式有所不同。美國有3大信用報告機構，分別為EQUIFAX、TRANSUNION及EXPERIAN；加拿大為EQUIFAX及TRANSUNION。其所蒐集的消費者信用檔案，內容包含房貸(含二胎房貸)、車

貸、信用卡(包含銀行卡及百貨公司卡)、公開資料(Public Record)及被查詢紀錄(Inquiries)。以下列出FICO評分在世界各國除美國與加拿大以外的應用情形(表一)、以及與各信用報告機構間所合作開發的產品(表二)：

FICO評分為消費者信用檔案資訊的彙總，FICO評分為3位數字，範圍為300分到850分，分數越高代表未來發生違約機率越低。FICO評分不考慮年齡、性別、收入、地址及職業。評分區隔在FICO評分第8版仍與先前版本架構相同，如圖一。我們可以發現評分卡總共有12張，但評分卡區隔的原則都與先前相同，有嚴重延遲者會考慮有無其他的公開資訊及最差評等；無嚴重延遲者，則考慮檔案厚度、檔案存在的時間(帳齡)及最近新增帳戶。

表一 FICO評分於世界各國的應用概況

信用報告機構有使用FICO評分	授信機構有使用FICO評分
ICB—Ireland 伊朗	Japan 日本
ACSP—Brazil 巴西	Poland 波蘭
KCB—S. Korea 南韓	Sweden 瑞典
Benefit Co.—Bahrain 巴林	Taiwan 台灣
Biroul de Credit—Romania 羅馬尼亞	Thailand 泰國
CBS—Singapore 新加坡	
NBKI—Russia 俄羅斯	
Trans Union—South Africa 南非	
SIMAH—Saudi Arabia 沙烏地阿拉伯	
Circulo de Credito—Mexico 墨西哥	
iScore—Egypt 埃及	

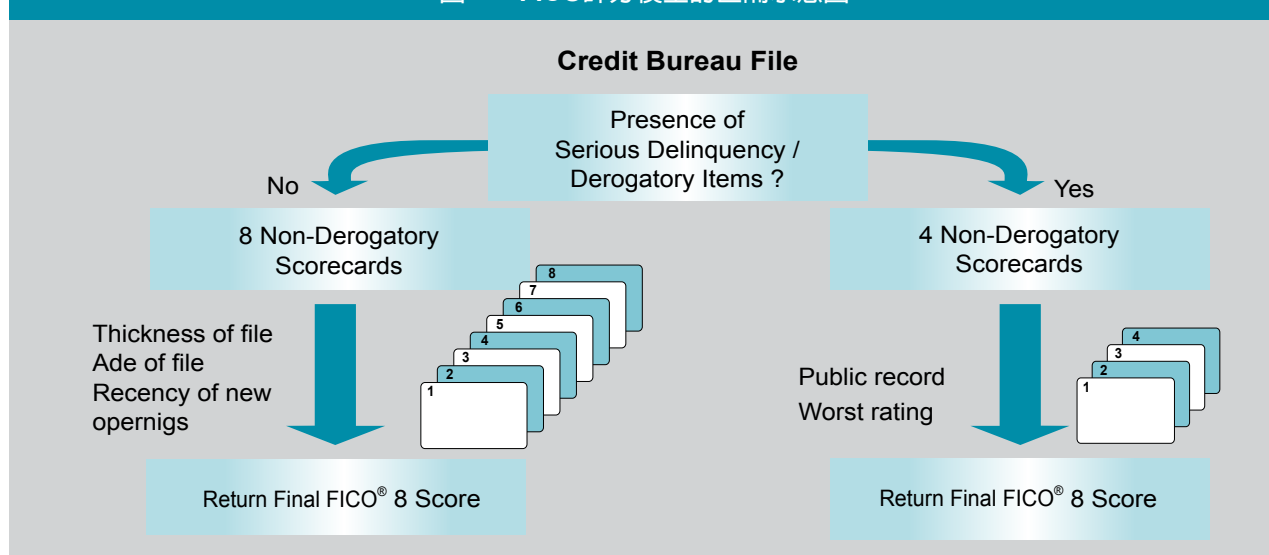
FICO評分長期都對外說明有5大類型的評分變數，各類型評分變數所佔的百分比也都類似。5大類型分別為繳款歷史 (Payment History)佔35%、負債總額 (Outstanding Debt) 佔30%、信用歷史長度 (Credit History Length)

佔15%、取得新信用 (Pursuit of New Credit) 佔10%、信用組合 (Credit Mix) 佔10%。評分的最低要件為被評分客戶仍然存活、至少有一信用交易帳戶開戶超過6個月及過去6個月至少有一筆沒有爭議的信用交易紀錄。

表二 FICO與各個信用報告機構所合作開發的評分產品

Product	EQUIFAX	TransUnion	Experian	MICROBILT	CoreLogic
FICO® Score	»BEACON®	»FICO® Risk Score, Classic	»Experian/FICO® Risk Model	»FICO® Expansion Score	—
FICO® Industry Score	»Auto »Bankcard »Mortgage	»Auto »Bankcard »Mortgage	»Auto »Bankcard »Mortgage	—	—
FICO® NextGen Score	»PinnacleSM	»FICO® Risk Score, NextGen	»Experian/FICO® Advanced Risk Score	—	—
FICO® Bankruptcy Score	—	»FICO® Bankruptcy Risk Score	»Experian/FICO® Bankruptcy Score	—	—
FICO® Credit Capacity Index™	»FICO® Credit Capacity Index™	—	—	—	—
FICO® Economic Impact Index	»FICO® Economic Impact Index	—	—	—	—
FICO® Mortgage Score 1.0 Powered by CoreLogic®	—	—	—	—	FICO® Mortgage Score 1.0 Powered by CoreLogic

圖一、FICO評分模型的區隔示意圖



FICO評分的一大特色，在於可替不同的需求者做「客製化評分」，藉由替需求者的量身訂作，不僅更符合客戶所需，評分對於風險預測也較精準，FICO提供客戶的選擇包含以下幾項：新的資料來源、最即時的信用報告、不同績效的定義 (例如：逾期60天或發生呆帳)、不同授信產品別 (房貸或信用貸款) 及地區區別。

(三) 跨領域分析

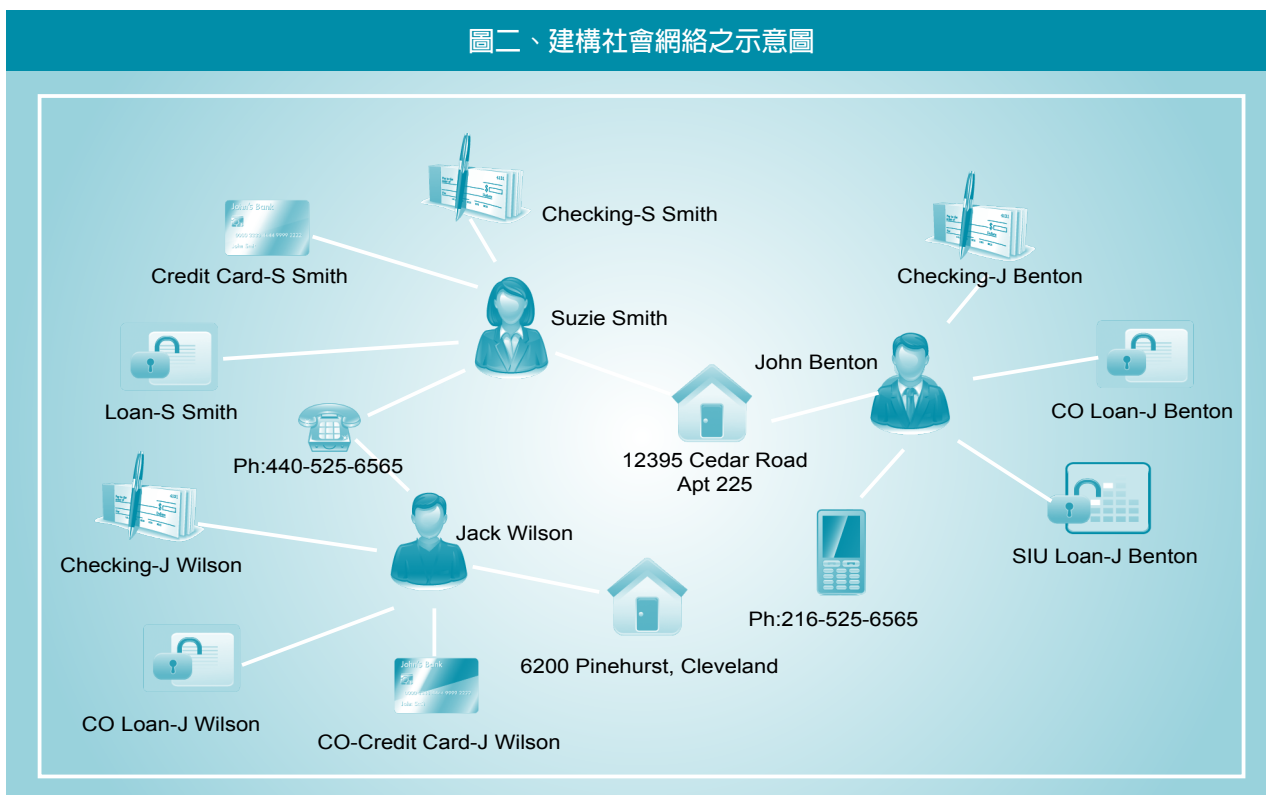
巨量資料所著重的在於利用龐大的資料庫建立更細緻的預測模型，但若我們試想另一種更宏觀的可能性，假如我們採用數個龐大資料庫的聯集來進行分析，那麼又會是什麼樣的情景？是否真的可以藉由資料的整合應用後，繼

而得以捕捉真實世界的完整全貌？

本次研討會即有數場講座在探討這樣的議題，即如何採用橫跨不同領域的巨量資料。如從信用分析的領域為出發點的話，最廣受討論的議題就是如何加入「社會網絡」(social network) 的資訊來加以分析。

「社會網絡」可藉由內部資料及外部資料來逐步建構，首先我們就內部資料來看，這指的是自身資料庫即有的資料，通常使用的是基本資料，例如：電話、地址、受雇公司等，我們可以將前述資訊任一相同的人視為是同一個社會網絡 (視為同住、鄰居或為同事關係，可見圖二之示意圖)，接著觀察這群社會網絡下的整體表現，再將這個整體表現作為評斷個人時的因素，此即以「物以類聚」的角度來出

圖二、建構社會網絡之示意圖



發給予評分，舉例而言，假如受評對象是住在一個犯罪率極高的區域，又或者其同居人或同事大多都有銀行違約的紀錄，我們可以藉此推斷，這個受評對象將因其社會網絡的負面因素而導致分數下降。在某講座中，Erste Group Bank (一家在東歐地區經營的銀行) 針對居住地區別好壞的因素進行實證，發現對於無延遲樣本的模型AUC可以增加11%，顯示其確實可提升模型預測力，間接地可使業務量增加。當然，該銀行亦承認這類相當敏感的資訊，可能會造成一些爭議，例如是否對特定地區或人士有歧視之虞。

至於利用外部資料建構社會網絡則是一種更大膽的嘗試，用以分析的資料來自社群網站如臉書 (Facebook)、推特 (Twitter)，主要可

應用在行銷面，此類資料來源雖無記載個人的信用行為表現，但仍反映了個人的人格特質：例如宗教信仰、政治傾向、感情關係、好友資料、喜好事物 (說讚的內容) 等，講座中提及劍橋大學曾對臉書資料與信用表現進行分析 (只針對小規模樣本，約5萬8千多筆使用者)，發現人種 (白人/黑人)、政治信仰 (民主黨/共和黨)、宗教信仰 (基督教/伊斯蘭教) 對於違約都具有相當高的關連性 (AUC>80)。然而該資訊來源所面臨的挑戰，是來自於社群網站的用戶是否會同意銀行取得他們的資料，畢竟這樣的資料已經過於隱私，所以雖然在學術裡的小規模研究中發現其有相當的預測效果，但何時會應用在實務，仍然有相當長的一段路要走。

圖三、取得個人臉書資訊之請求同意書



(四) 資料隱私保護

由於本次研討會的核心主題在於巨量資料的分析，因此主要的講題都聚焦在如何將資料做更有效的應用，而與之對立的一方即資料隱私保護之面向，相形之下則較少探討。然而此不單為資料分析科學最需注意的規範限制，亦是在國內通過個資法後所關切的重心，因此特別參與此議題之相關講座「Big Data and Global Privacy」。

我們從前段「跨領域分析」可以看到，資料的聯集應用可以涵蓋的面向既廣且深，假設真如與會者所言，個人於社群網站的喜好及點閱記錄都可以被加以利用分析，這門科學可說是應用到相當極致，相信在科學面會有相當的效益，應可藉此提昇模型的預測準確度，然而，這有可能過份地侵犯隱私，尤其是個人嗜好、價值觀認同這類足以表彰人格 (person-centric) 之資料。追求科學發展所帶來的經濟效益以及維護個人隱私之間，勢必得找出一個兩相平衡的中間點。而在本講座中，主辦單位邀請了美國及英國的信用分析機構CoreLogic及Equifax UK來分享他們對資料隱私議題的觀點。

若從資料保護的起源出發，其係源自1995年歐盟的規定，當時的考量為大量數位化資料的時代來臨，必須有一套完善的控管方式，以強化個人的權利及保護中小企業，而其核心的概念在於「當事人的同意」(data subject's consent)、

「當事人對於資料的請求權」(data subject may request information)、

「當事人對於資料的刪除權」(right to be forgotten)、

「當事人對於資料的複製權」(data portability)、

「當事人得拒絕資料被加以研究」(object to profiling)，我國個資法其實即參考歐盟的規定。然而，這套規定係在巨量資料時代未來臨前所制訂，與現今的環境是否能相符，容有疑義；就美國的角度來說，多數的公司並不太樂意巨量資料的創新使用方式被加以規範，尤其現今對於資料隱私的態度仍未持定論。

因此，無論是在美國或歐洲，都在著手對資料隱私的區塊進行一些新的嘗試，舉例而言，麻省理工學院(M.I.T Media Lab)正在研發可以控制並稽核資訊流的一些工具，讓未經授權的資料使用可以被偵測出來；而英國則是嘗試一種反向的方式，「既然這些資料分析的結果對於企業是這麼有價值，那麼何不免費回饋給當事人？或許這些結果對當事人來說同樣有價值」，這就是英國所提出的(圖四)歐盟的資料保護架構示意圖。

「midata」計畫，只要消費者有需要，企業必須免費提供他們的分析結果給當事人，可有助於他們尋找最便宜或最高性價比的商品(例如百貨公司的比價結果)，以更有效率地管理他們的生活。當然，這項計畫也引來了許多討論，例如這樣是否會加重企業成本，或是複雜且冗長的申請程序會降低消費者申請意

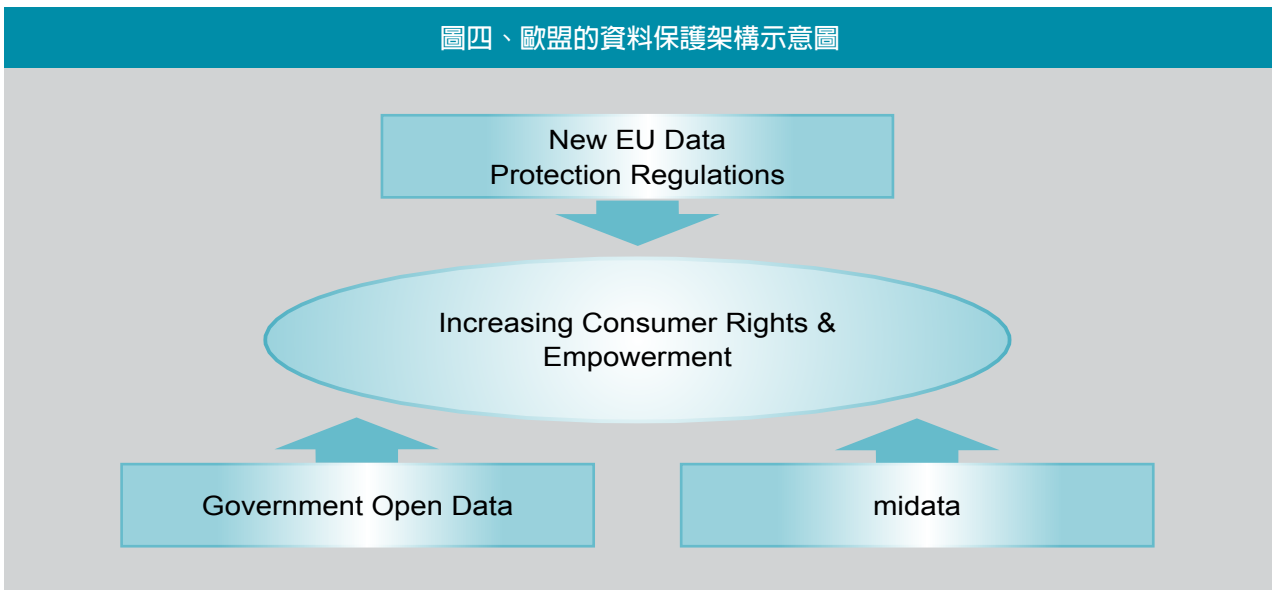
願，然而確實是一項相當創新的嘗試。前述「midata」計畫的內容是要求公開「企業的資料」，而無獨有偶的是，包括美國及英國在內的西方國家，已經陸續於2009年開始逐步公開「政府的資料」(Government Open Data)，同樣也是出於「將有價值的資料回饋給民衆」這樣的理念，內容包括交通運輸、醫療、氣象、房地產價格等資訊，這股風潮目前也已吹向我國，一些中央部會及地方政府，現也已將部分資料釋出公開於網路上，顯見新時代下的資料應用管理，未必皆是以保護使用為出發點，適度的分享可能也是一種互利的可能，歐盟在勾勒消費者權利的架構圖下，即以「新法規」、「midata」、「政府公開資料」為三個要素，見圖四。

另外，在講座中亦提及世界經濟論壇(The World Economic Forum)於2013年2

月發佈了一項報告，當中提出一種想法，那就是未來所蒐集的資料都應該透過密碼，個人即可藉由密碼來控制其被蒐集的資料要如何使用(例如：是高度使用或低度使用)；經會後查詢這份報告，當中並未詳述這項概念應如何落實，推測可能的作法是當資料蒐集者要使用資料時，必須擁有當事人所提供的密碼才能使用，以維護當事人對其個資被使用程度的偏好(preferences)，如此一來，未來的資料使用就需藉由如登入或註冊等手續，以保留軌跡來加以管制。

從上段所提出的建議或嘗試，我們可以看出資料保護的方式也在走向創新的方向，以因應巨量資料時代的來臨。或許資料保護的方式並不單單是禁止使用的一種思路，有效的分享或更有效率的管控，可以使資料隱私被妥善保護的同時，仍可創造有價值的利用結果。

圖四、歐盟的資料保護架構示意圖



心得與建議

參加本次研討會後，對於經歷超過50年的「FICO信用評分」及近年來持續探討的「巨量資料」兩個領域分別說明心得，再一併列出建議。

(一) FICO信用評分

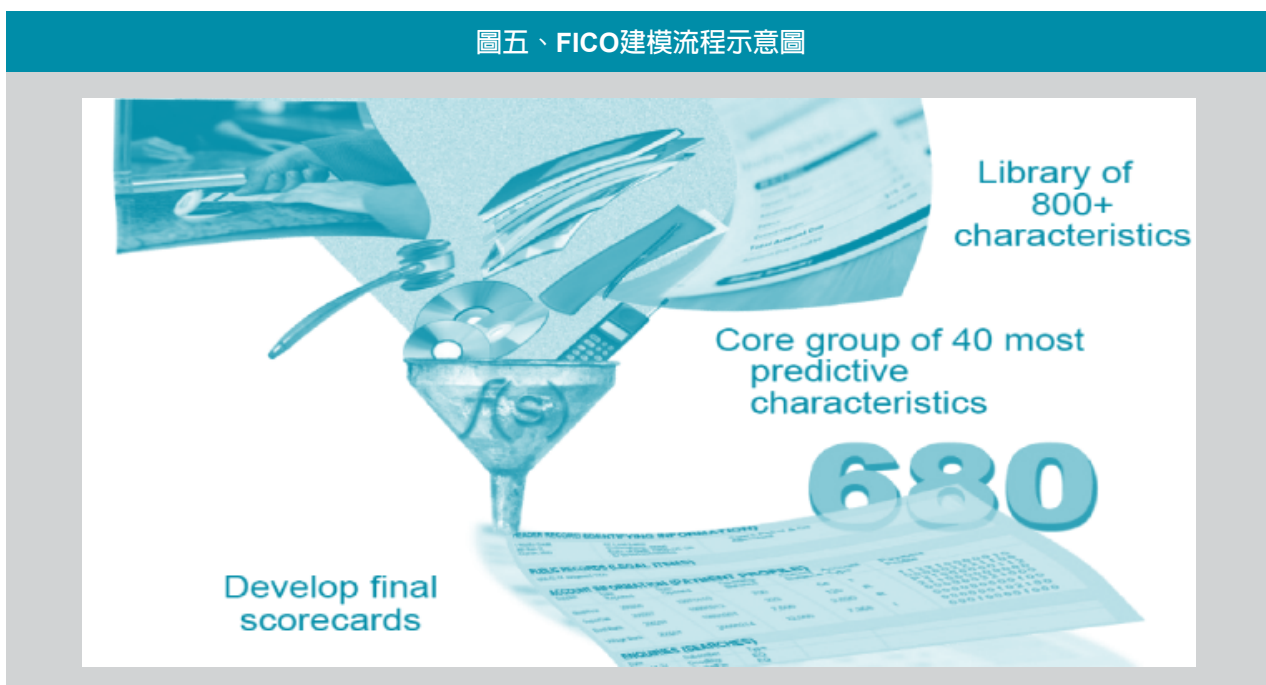
FICO Score在美國仍然是消費者信用狀況好壞的主要參考指標，即使美國三大信用報告機構，於2006年之後自行推出Vantage Score產品，仍然有許多美國銀行選擇傳統的FICO Score。另外，早在2002年FICO在舊金山研討會時，就已提出新版評分產品FICO Nextgen(下一個世代的評分)，但經過11年後，傳統的FICO Score依然存在其產品清單裏，顯見目前的使用者仍然需要傳統的FICO

Score，意味著該產品有相當強的競爭力。

維持競爭力的方式，除了持續維持它的預測有效性外，尚包含簡便有效率的評分上線過程，方能成為維持效率的一大關鍵，而此關鍵在於其核心變數的建立。參考Google預測禽流感的實際案例與FICO評分的簡報後，發現當今的資料累積都很快速(例如：每天Google有超過300萬筆查詢)，用來分析的資料量都很大，但都有一個特點，就是經過分析過程後，找出最具預測力的變數或特徵都不多，Google預測禽流感找出了45個，而從FICO的投影片中發現，最具有預測力的核心變數僅有40個(如圖五)，顯見核心變數在精而不在多。

利用核心變數產生評分模型後，預測力都可以到達統計上顯著性的要求，而應用不同類型的變數組合，因應客戶性質的不同做出客製

圖五、FICO建模流程示意圖



化產品，繼而將其產品的領域拓廣至信用分析以外(例如：保險及醫療)，此亦是其維持競爭力的方式，值得加以學習。

(二) 巨量資料

巨量資料的應用不僅限於獲取商業利潤，對於城市甚至整個國家的生活品質，也帶來進步的動能。例如當今世界正在嘗試的「智慧城市」概念，結合了交通、水電、治安…等公用資料庫來營造高生活品質的城市；至於我國也早已提出電子化政府的規劃，希冀能在不受各機關框架限制的模式下，可以將各政府機關的資料庫整合成巨量資料，以雲端技術管理來做加值應用，以提升政府體整施政政策之品質，但目前實務上的做法面臨許多限制，於是開放資料(Open Data)的方式讓有資料需求者加以應用。由此可知，資料保護與資料整合應用將會是未來持續探討的議題，兩者必須兼顧，才能顯現資料的價值。

(三) 建議

1. 藉由提供客製化服務以使產品多元化

我們可以看到FICO雖然是以信用評分起家，但FICO Score只是在它眾多產品清單下的其中一環，經過數十年的觀察可發現，FICO逐步擴大評分應用的範圍，提供客製化產品服務，創造更多應用價值。

對於台灣金融業而言，較大型的金融機構已有應用信用評分的經驗，篩選高風險客戶已是基本需求，而優質客戶的開發，新業務的推

展需要找出潛在客戶等應用，則是使服務多元化的動力，信用評分只是一個工具，經過一些重新組合與變化，可以讓推出的產品更符合客戶期待。

2. 資料保護議題的追蹤

資料隱私保護的議題在將來同樣仍然會繼續受到重視，尤其在巨量資料的年代，如何避免資料被濫用應會是討論的重點，目前國際上也為此問題在著手研究，找尋更能兼顧資料使用者及資料當事人兩者權益的方法。然而在作法上，國際上似乎有一些新的想法，或是採取資料分享或公開的方式、或是採取更先進的科技來管裡的方式，而非單單只是禁止使用而已。建議宜就此類議題的國際趨勢追蹤瞭解，以使資料能在最有效益的使用下又可以符合人性之價值，探討如何在未來的資料管理年代中獲得雙贏。

參考資料來源

1. 本次研討會 (FICO WORLD 2013 - BIG DATA ANALYTICS AND THE CUSTOMER REVOLUTION) 所提供之投影片。
2. Viktor Mayer-Schöberger and Kenneth Cukier(2013), Big Data – A Revolution That Will Transform How We Live Work, and Think, Houghton Mifflin Harcourt.