

# MITCDOIQ Symposium 2016

朱昌綸/金融聯合徵信中心 徵信部  
蘇柏鳴/金融聯合徵信中心 資訊部

MIT Chief Data Officer & Information Quality (MITCDOIQ) 係美國麻省理工學院 (MIT) 資訊工業技術研發中心對於資料品質管理之學術研究平台，自2007年起，每年定期邀集各產業之菁英參與研討會，期盼將更多智慧元素導入資訊管理自動化流程，藉以提昇資訊管理。

資料長 (Chief Data Officer, CDO) 不同於科學家、分析師、工程師...等專注於技術層面之職務，乃是隨著企業不斷發展而誕生的一個新型管理者，其資訊管理職務可分資料資源管理 (Data Resource Management) 及資料品質管理 (Data Quality Management) 二種，雖然二者管理的對象都是資料，但是管理的目標及規範卻是不同，資料資源管理著重在將資料做出策略性的整合運用，盼能發揮資訊最佳的價值 (Value)；而資料品質管理著重在將資料轉移成為資訊的過程中，找出影響品質的因素並降低錯誤發生的機率，盼能提升資訊產出的品質 (Quality)。

本次研討會邀集各國際企業CDO參與，經由經驗分享與交流，讓我們可以瞭解資訊管理相關領域先行者目前的研究方向及趨勢。

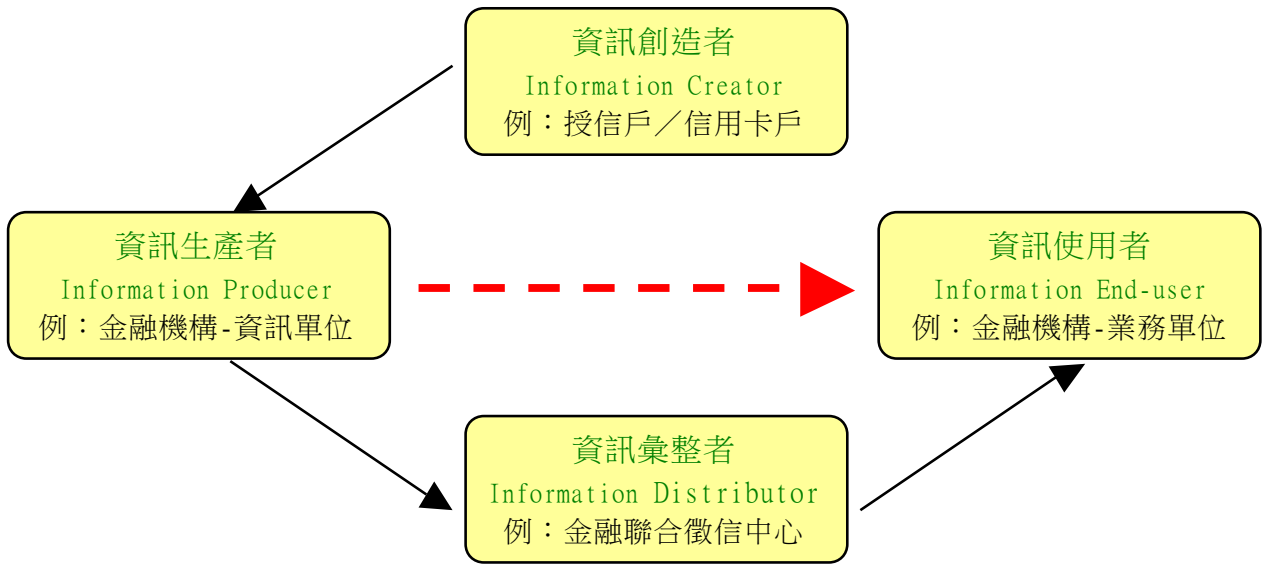
## 壹、參加研討會目的

資料成為資訊的過程稱為資訊供應鏈，其組成包括四項角色：

- 1、資訊創造者 (Information Creator)：係指向金融機構借貸之客戶，例如：授信/信用卡戶。
- 2、資訊生產者 (Information Producer)：將資訊創造者之借貸資訊，依格式製出檔案，再報送至資訊彙整者，例如：金融機

構資訊單位。

- 3、資訊彙整者 (Information Distributor)：將來自各金融機構之資料彙整於資料庫中，並設計介面供金融機構業務單位查詢利用，例如：金融聯合徵信中心。
- 4、資訊使用者 (Information End-user)：藉由資訊彙整者所提供之資訊，獲悉其客戶之信用狀況，例如：金融機構業務單位。



在此供應鏈上，任何一角色都必須依賴其他角色提供無錯誤的品質服務，任一角色必須承擔應有的品質控管責任。然而，資訊使用者認為資料品質控管應是資訊彙整者的責任，資訊彙整者又認為資料品質控管應是資訊生產者的責任，是以，資料品質控管往往在供應鏈上角色的本位主義考量下被忽視。

近年來，由於資訊處理技術日益普及、電腦效能日益增強、通訊速率日益快速、設備成本價格日益滑落，資訊生產者開始思考如何兼具資訊彙整者的角色，藉由雲端資料庫直接提供資訊分享服務，供應鏈上各角色既合作也競爭，原本角色間明顯界線也日漸模糊，未來對於本中心信用資料庫之永續經營與發展，必將會是一項嚴竣的挑戰。

## 貳、參加研討會議程

本次MITCDOIQ研討會於105年7月12日至14日間在美國波士頓麻省理工學院舉辦，今

年已邁入第10屆，由MIT資訊工業技術研發中心Dr. Stuart Madnick及Dr. Richard Wang 二位教授策劃及主持。

本次研討會參與人員包括：資訊產業（IBM、SAP、EMC、Intel、PWC...），金融產業（SunTrust Bank、Citibank、US Bank、TD Bank、JP Morgan、AIG...），政府部門（紐約州政府、英國陸軍、美國白宮辦公室、美國運輸部、美國陸軍、美國衛生部、美國公眾服務部...），學術機構（哈佛商學院、MIT...），共計約有250人參加，我國除本中心徵信部及資訊部各推派一員參加，景文科技大學旅遊管理系亦有一教授出席本次研討會。

本次研討會進行計3日（2016.7.12 ~ 2016.7.14），規劃10場次及30項討論議程（session）包括：資料長（CDO）角色、資料品質管理、資訊風險管理、區塊鏈資訊管理...等，資訊研討領域包含：金融、保險、電信、零售、醫療、藥品...等產業。

## 參、參加研討會內容

時間：2016年7月12日，議程：Session 1A

主題：CDO Best Practices: Creating Truth, Trust and Traceability – By MITCDOIQ Symposium Valued Partners, Collibra, Global ID' s and Sandhill

引言人：Robert Lutton, Donaid Soulsby

主講人：Stan Christiaens, Bill Winkler

內容摘要及感想：

DIKW Pyramid模型，能夠幫助我們理解資料（Data）、資訊（Information）、知識（Knowledge）、智慧（Wisdom）間關係的模型，「資料」是對人、事、時、地、物等基本單位描述之記錄與儲存，通過某種方式組織和處理，資料就有了意義，這就是「資訊」，資訊經過使用者思考的過程，將資訊與自身經驗結合後產生「知識」，將知識實際運用於決策並產生價值就是「智慧」。

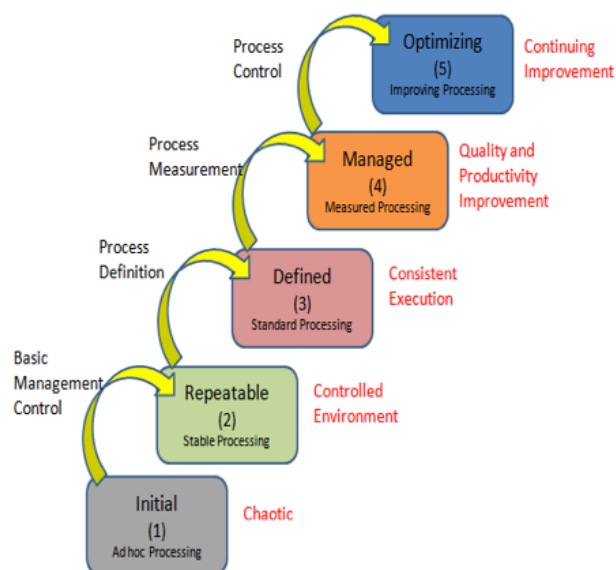
"You can't manage what you can't measure, you can't measure what you can't define", 是資訊管理常見的問題，如果無法估計資料蒐集的數量、種類...等metadata，不管在技術或管理層面都很難有效落實，資訊爆炸的時代，資料數量估計及資料型態的擷取愈趨複雜，造成了資訊管理上的問題，在尚未決定資料蒐集資料用途之後前，資料應用流程是很難做出評估（圖二）。

CMMI（Capability Maturity Model Integrated）協會提供資訊管理成熟度（Data

圖一：DIKW Pyramid 模型



圖二：資訊管理流程



Management Maturity, DMM<sup>1</sup>) 模型 (圖三), 希望在共同的架構下縮短業務與技術對資訊管理認知的差距, 會議中講員特別提出4種資料品質評估方式, 包括: 資料定義、資料策略、資料清理、資料評鑑:

#### 1、資料定義 (Data Profiling)

透過資料定義的過程使資料收集者能夠瞭解資料內容以及診斷資料的問題, 主要是發掘組織中的資源的特性, 有效的資料定義可以瞭解組織資料的內容、結構、關係, 許多企業投入相當多資源在顧客關係資訊系統, 嘗試整合各部門之客戶與產品資料, 藉由資料定義亦可檢視各類資料結構, 發現資料不正確、不完整、不一致的狀況, 再加以統合修正, 產生更高的價值。

#### 2、資料品質策略 (Data Strategy)

資料品質的好壞不只單純影響資料應用的效能, 也會影響相關業務的發展, 如何提高資料品質的方式很多面向, 從資料擷取、處理、利用...每個階段, 組織內各部門所考量不同, 組織在落實資料品質提升的目標前, 應考量各面向所預想到的瓶頸, 統整後提出系統性策略, 並全方面落實。

#### 3、資料清理 (Data Cleansing)

資料清理是指發現並糾正資料中可識別的

錯誤最後一道程序, 包括檢查資料一致性, 處理無效值和缺失值等, 主要分成三個步驟: 1定義錯誤資料類型, 2探索錯誤資料來源, 3修正錯誤資料紀錄。資料清理的過程非常費時, 惟資料清理的優先順序與處理方式攸關資料清理成敗, 企業常因忽略資料清理的順序, 導致錯誤的分析結果。

#### 4、資料評鑑 (Data Assessment)

評定資料內容代表意義之準確性, 其主要評估指標包括:

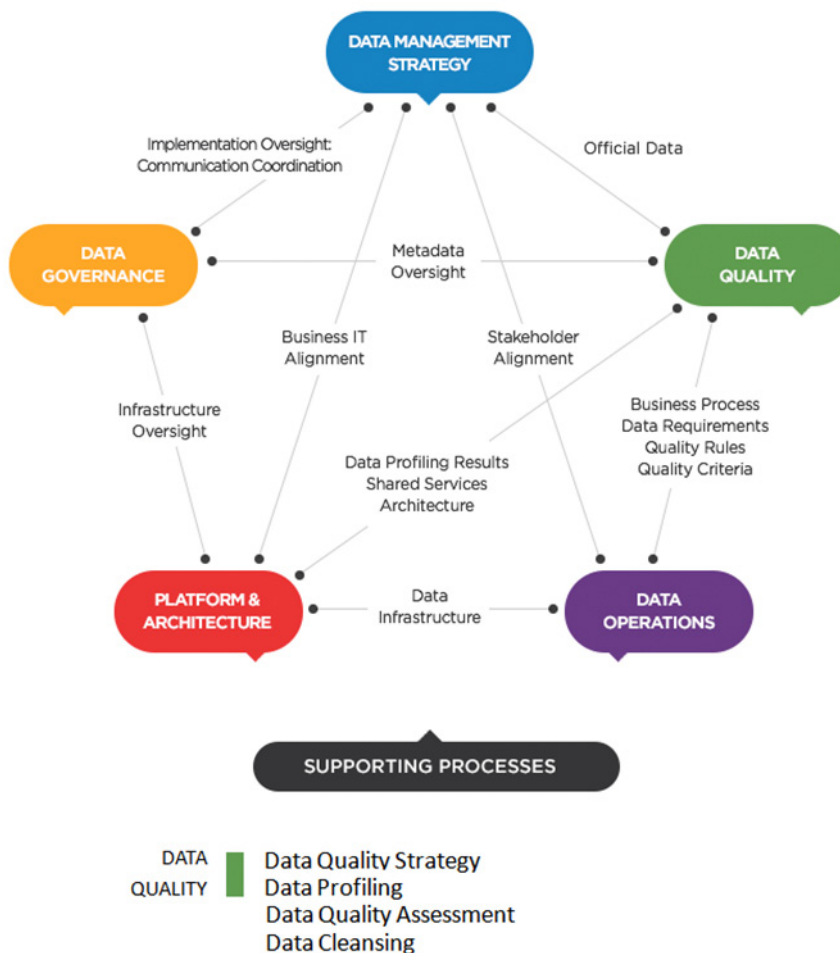
- \* 有效性 (Availability) : 資料是否有取得的管道跟介面, 並且符合時效。
- \* 可用性 (Usability) : 資料來源是可信的, 來自一個國家, 領域或行業, 並且有專家定期審計和檢查資料內容的正確性。
- \* 可靠性 (Reliability) : 資料的內容是正確的, 經過一段時間後能保有一致性, 且資料格式是符合標準。
- \* 關聯性 (Relevance) : 收集的資料不完全是一致的主題, 但它們闡述的大方向是相同的。
- \* 呈現度 (Presentation Quality) : 資料的描述、分類和編碼內容滿足規範, 呈現方式容易理解。

時間: 2016年7月12日, 議程: Session 1C

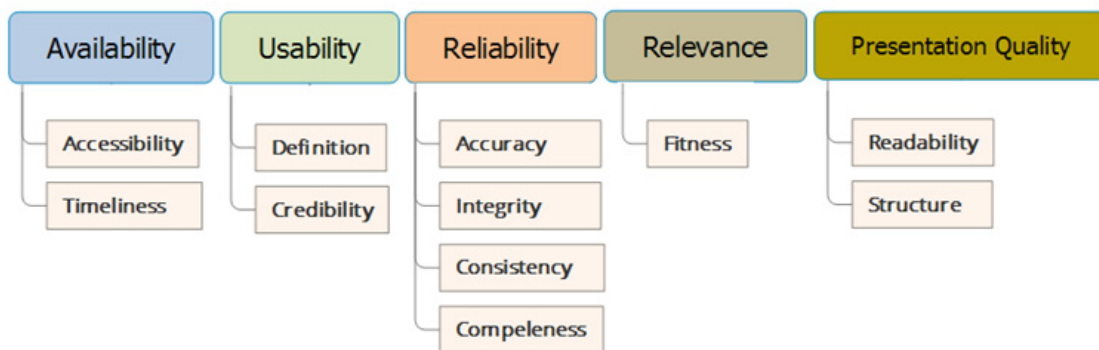
主題: Big Data Means Big Change

1 <http://cmmiinstitute.com/>

圖三：資料管理成熟度模型



圖四：資料評鑑指標



主講人：Peter Anlyan, Amanda Marko

內容摘要及感想：

大數據 (Big Data) 係指儲存、處理、分析這些在3V (Volume、Variety、Velocity) 層面難以管理的資料之技術，也包含「分析過往」與「預測未來」的能力；大數據並非斬新的概念，如果單純只從量的觀點來看，過去就已經存在，例如：飛機噴射引擎產生的數據、生物領域染色體分析、NASA太空領域的探索…，在過去，是用相當昂貴的超級電腦來進行龐大資料分析，現在談及的大數據與以往不同之處主要有以下3點：

- 1、隨著社群媒體與感測網路 (Wireless sensor network) 的發展，從日常生活中產出大量且多樣的資料。
- 2、隨著資訊設備軟、硬體的進步，資料的儲存與處理成本大幅下降。
- 3、隨著雲端運算的盛行，已無必要自行準備巨量資料的儲存與處理環境。

在觀念上，大數據萃取出智慧的作法，和傳統由人工智慧 (artificial intelligence) 的作法有很大的區別。人工智慧技術重視演算法 (algorithm) 和模型 (model) 的開發，例如，自然語言處理 (Natural Language Processing, NLP) 偏重於文法、語意…等規則，總是想盡辦法把諸多的規則放進電腦，希望讓電腦成為有能力分析人類語言的專家系統，無奈，雖然基本的文法規則很簡單，人類語言與文字並不見得一定會按照文法與語意規則呈現，總有層出不窮的「例外」，大數據對於自然語言處理的方法，不講究文法與語意

規則，因為雲端網路就是一個巨量的資料庫，擁有龐大數量的文章和各類的知識，電腦收到一個詢問，就瞬間將此詢問與這個雲端資料庫進行比對，找到相近的資訊回覆，從中得到答案，目前Google翻譯就是將網路上所有的文章當成自己的資料庫應用。

對於在大數據運用的模式我們可大略分成4項層次：

#### 1、掌握過去與現狀：

運用大數據，就要先從蒐集資料開始，資料有可能是：顧客購買商品紀錄、Facebook留言內容、Line語音通話紀錄、血壓心跳體重健康資料、就診醫療紀錄…等，除了人類的資料以外，像是：伺服器或網路設備的日誌檔 (Log)、家用智慧電表、車用感測器紀錄的資訊、氣象資料、IoT (Internet of Things) …等由事物或大自然產生的資料，這些資料並不一定要做深入的分析，也可以只是用來「即時監控當下」的情況，進行異常管理 (如：於在發現異常時才發出警訊)。

#### 2、發現行為模式，洞察先機：

累積大量資料後，就運用資料探勘 (Data Mining) 或機器學習 (Machine Learning) 等技術，勾稽出對業務有幫助的行為模式，例如，針對企業或家用智慧電表資料與氣象資料等外部資料進行綜合分析、針對顧客電洽客服中心的内容與Facebook留言内容進行情感分析 (Sentiment Analysis) …等行為模式分析應用。

#### 3、預測未來：

大數據一旦能夠分析出成功或失敗等模



式，接下來，只要使用者將資料輸入，就能進行未來預測。例如，針對當日的天候、氣溫、濕度…等天氣資料與家用智慧電表的資料進行綜合分析，就可能高精確度做出「今天電量使用率會來到90%」的預測。

#### 4、發展優化策略：

如果電力公司能夠精確的預測出電力的使用量，就可以提出離峰電價下降或高峰用電電價提高的優化策略，將用電需求趨於平衡，降低發電成本，這就是一種優化策略發展。

時間：2016年7月13日，議程：Session 5D

主題：Data Management on the Blockchain

主講人：Justin Magruder, Diane Schmidt

內容摘要及感想：

區塊鏈（BlockChain）技術是一種通過自身分散式節點（不依賴第三方），進行網路數據的存儲、驗證、傳遞和交流的技術。可以把區塊鏈系統中的每個區塊想像成是一個節點，所有節點功能權限平等，數據資料在各節點上都彼此完整複製，亦即，當任一節點上的紀錄變動，系統上的每一個節點都會跟著同步變動，區塊鏈技術係由電腦系統按照預先設定好的規則，自動化執行一切處理流程，因此，交易紀錄一旦發生，就不可逆，提升資訊的正確性與安全性，區塊鏈技術且具可追溯性及透明性，不像是現在大部分的商業模式，由各產業的中央主管機構或由系統支援廠商自行居中運籌帷幄，掌控各階段交易過程，並保存所有數據紀錄資料。

透過區塊鏈技術發送訊息，系統上的每一個節點都會接收到該訊息，惟僅有擁有私鑰

（private key）的使用者才能解密看到訊息內容，也就是說，在區塊鏈上儲存的訊息都是加密的，雖然訊息共享但是非常安全，不需要擔心訊息內容遭竄改、遺失，也不用擔心交易過程會出差錯。

區塊鏈技術的特性可以解決原有金融服務交易過程中的漏洞或是弊端，並減少交易時的成本，對銀行而言，運用區塊鏈技術來記錄帳戶資料的更改，遠比保留內部的帳冊資料來的更有效率及安全，因此，許多外國企業和金融機構都已開始投資區塊鏈技術並進行研究發展，希望可以創造出繼比特幣之後的另一個新商機。

「區塊鏈」如「大數據」一樣，只是一種底層技術，若沒有應用在現有生活場景或是商業模式中，就無法產生價值，區塊鏈技術只是一種手段，當運用區塊鏈技術產生新的商業模式時，相對應的配套工具和規範是必要的。

區塊鏈技術擁有「去中心化」（decentralization）特性，將信用紀錄儲存於區塊鏈機制中，金額機構（資訊生產者）可隨時更新其客戶之信用資訊，該客戶往來之金額機構（資訊使用者）才具查詢該戶資訊的權限，再搭配使用者付費的機制，就足以改變信用資訊集中於聯合徵信中心（資訊彙整者）的現況，本中心應密切關注此資訊供應鏈上角色既合作也競爭的變化。

時間：2016年7月14日，議程：Session 7

主題：Four Eras of Analytics, Four CDO

## Roles

引言人：Robert Lutton

主講人：Tom Davenport

內容摘要及感想：

"Data is the new oil."，不只表示資料已成為有價值的東西，亦表示資料是人類驅動進步的燃料，資料一直產生於日常活當中，只是近年來的科技發展，使我們可更容易紀錄或收集資料，主講人Davenport博士將資料分析的演進分為4個階段（圖五），說明如下：

#### 1、Data 1.0：Transactional Data & DSS

資料主要以交易資料型態呈現，其特性就是：欄位內容結構化、單筆資料長度小，通常單純用於協助企業的內部決策，以決策支援系統（Decision Support Systems，DSS）為代表，所需的技能：資料整合及規劃應用、商業敏銳度、統計學。

#### 2、Data 2.0：Big Data Era

數位化進展和普及之下，資料的型態與數量快速增加，帶動巨量資料（Big Data）、開放資料（Open Data）等應用，資訊科學家（Data Scientist）、資料長（Chief Data Officer）等專職領域因應出現，支援資料密集分布的開源軟體框架（如：Hadoop）成為資料儲存分析主要工具，資訊之呈現

也由冷冰冰的數字進化為以視覺化（Data Visualization）方式呈現，所需的技能：資料重組（Data Restructure）、開放資料程式開發、資料視覺化呈現。

#### 3、Data 3.0：The Data Economy

以政府的開放資料與產、政學界的巨量資料為核心，透過各種科技分析工具規劃與應用，將資料附加價值最大化來創造經濟效益，如：提升營運及管理效能、協助決策或創新…等，所需的技能：敏捷式專案開發、預測分析能力、變革管理<sup>2</sup>（Change Management）。

#### 4、Data 4.0：The Age of Automation

藉由Machine Learning及Deep Learning技術，電腦自我學習資料分析過程，資料蒐集完全自動化，超越資料分析僅供人類參考階段，進入人工智慧（AI）自主判斷階段，例如2016年3月Google DeepMind開發的電腦圍棋軟體（AlphaGo），在南韓首爾擊敗世界排名第五的南韓棋王李世石，所需的技能：自動化機器學習（Machine Learning）、深度學習、自然語言處理（NLP）<sup>3</sup>、類神經網路系統。

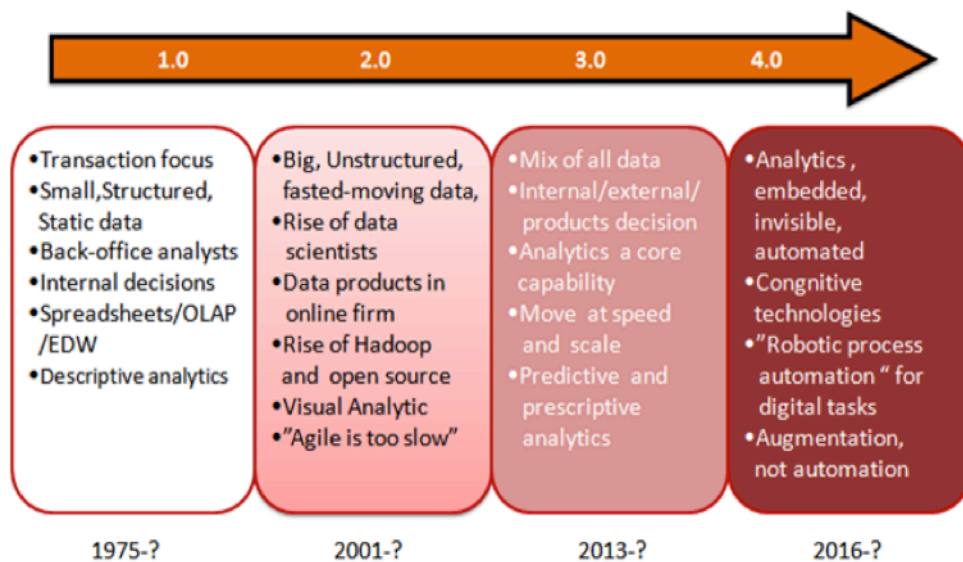
資訊管理已成為企業營運成長之關鍵所在，由前揭資料分析4個階段演進過程中

2 變革管理（Change Management）意即當組織成長遲緩，內部不良問題產生，愈無法因應經營環境的變化時，企業必須做出組織變革策略，以達企業順利轉型。

3 自然語言處理(Natural Language Processing ,NLP)是人工智慧和語言學領域的分支學科。在這此領域中探討如何處理及運用自然語言；自然語言認知則是指讓電腦「懂」人類的語言。



圖五：資料分析的 4 個階段



發現，資料處理與分析層面已經愈來愈複雜，企業管理階層若未能即時正確聚焦，將無法獲得良好的營運成果；當企業的研發、行銷…等業務單位忙於推展相關數位資訊商機之際，企業的法務、人資…等支援單位亦應同步關注如何確保相關作業符合法規，並保護企業資訊資產與企業商譽。

為順應新的數位環境需求，資料長（CDO）不但扮演攻擊（Offense）的角色，協助企業相關研究發展，例如：研發資料分析模型、督導數位創新方案、轉型既有的商業模式與流程、提升企業與顧客關係，亦扮演防守（Defense）的角色，協

助企業建立相關控管機制，例如：強化資訊安全管理、深化資料品質管理、落實資訊法規遵循，資料長角色扮演多元化，責任負擔亦日趨沉重。

#### 肆、參加研討會心得

謹將本次研討會相關案例研討（Case Study）議程中，CDO以其實務經驗所歸納出企業於資訊管理上常犯的錯誤，進行心得說明：

- 1、No Knowledge Workers be Involved：資訊使用者未參與規劃  
資訊系統開發者倘若對業務不甚瞭解，對於資訊問題的描述，屢以主觀角度判斷，

流程設計時易產生偏見。應體認：唯有資訊使用者才具客觀立場評鑑資訊流程，應於系統設計之前，就掌握日後應用之需。

#### 2、No Chief Officers be Involved：領導階層未參與決策

資訊管理係屬一連串複雜的抉擇，必須在安全、效率、成本三者間取得平衡，常因各部門間本位主義充斥，資訊管理責任歸屬不夠明確。企業領導階層應參與資訊管理之建構過程，將管理原則融入企業文化。

#### 3、Rely on Tools Rather than Principle：過度依賴工具

資訊管理乃是人員及程序的綜合體，工具僅扮演輔助角色，企業若過度聚焦或依賴工具，將落入不斷購置工具後，又再尋求工具整合的痛苦循環。資訊管理應致力於管理的原則及方法，將管理工具需求交給支援廠商鑽研。

#### 4、Security / Accessibility Trade-Off：安全性與實用性無法兼顧

安全性與實用性無法兼顧，是資訊管理與生具備的困境，然而資訊使用者卻往往無法體認這一點，歸咎於資訊管理不當。管理階層應先將有關資訊安全性政策透明公告於資訊使用者，。

#### 5、External Data Beyond Control：企業夥伴未分擔品質管理責任

忽略合作夥伴也應對分享的資料品質負起責任，放任外部資料品質不顧，必將連累

內部資料品質，且日後將付出加倍的代價。企業應提高資料品質的警覺性，將品質管理的經驗與合作夥伴分享，要求分擔資訊管理的責任，共同提升資料的品質。

#### 6、Policies Not be Carried Out：管理政策未確實遵循

員工缺乏誘因或害怕遭受抱怨而去改善資料品質問題，導致企業徒具資訊管理政策，卻未落實遵循。企業管理階層應善用管理策略，找出資料品質改善後的受益者，由受益者主導改善計畫，並建立改善績效指標。

### 伍、參加研討會建議

過往，企業業務單位與資訊單位之間的職務分工清楚明確，然而，隨著硬體設備成本下降、應用軟體介面普及、網路通訊傳輸快速、雲端儲存技術提升…等因素，原本業務單位與資訊單位間職務分工的界線日漸模糊，加上業務單位與資訊單位間充斥著本位主義，「資訊管理」係企業面臨永續經營時的嚴峻挑戰，資料長（CDO）這種新型管理者角色乃隨之誕生，系統支援廠商亦隨之蓬勃發展相關資訊管理工具。

本次研討會除了企業資料長參與外，相關資訊科技系統支援廠商亦設立攤位，介紹相關資訊管理工具，主辦單位亦規劃若干討論議題，由系統廠商分享其系統開發之經驗，討論會中，SAP系統開發專案經理提及：

「No Measure, No Manage, No Define, No Measure；沒有量化就無法管理，沒有定義就無法量化」，並分享：資訊管理系統最棘手的部分，較不在技術的開發，而在問題的溝通，業務單位與資訊單位雙方主觀的認定，往往是多頭馬車，其效果卻彼此牽制，系統開發專案經理表示：約60%的專案期程耗費在資訊定義（Define）釐清，且往往發現：當資訊定的義釐訂明確後，原本約20%的系統需求可不用繼續開發，由此可見，將資訊明確定義佔資訊管理中非常重要的比重。「明確定義資訊」簡言之，就是「使用相同的話語，描述相同的事件」，溝通時應避免「使用不同的話語，描述相同的事件」（synonym）或「使用相同的話語，描述不同的事件」（homonym）。

聯徵中心信用資料庫係各金融機構客戶信用資訊的彙集，聯徵中心其實就扮演著各金融機構間信用資訊部分CDO的角色，各金融機構資訊單位依據「資料報送作業要點」中的定義及規範（使用相同的話語，描述相同的事件），將資訊報送至本中心，再由本中心設計信用資訊查詢產品，供各金融機構業務單位參

考利用，協助做出准駁決策。

聯徵中心組織結構中雖無資料長（CDO）職銜，但已逐漸導入相關資訊管理工具（如：檔案傳輸監控系統、資料檢核稽催機制…），本中心仍應將「使用相同的話語，描述相同的事件」理念融入員工作業習慣及企業文化，藉以強化業務單位與資訊單位相互間的溝通，凝聚組織共識、提升作業效率，維持正確的資訊管理方向。

## 六、研討會會場照片

