

概述資料品質管理的趨勢及方法

朱昌綸／金融聯合徵信中心資訊部資料品質小組

「資料管理」可分為資料資源管理（data resource management）及資料品質管理（data quality management）二種類型，雖然管理的對象都是資料，但管理的目標卻是大相逕庭；資料資源管理著重將資料做出策略性的整合運用，發揮資訊的價值（value）；而資料品質管理卻著重於資料轉移成為資訊的過程中，找出錯誤發生的因素，提升資訊的品質（quality）；本文就有關資料品質管理之趨勢及方法綜合引述說明。

資料品質的趨勢

西元1769年，瓦特發明蒸氣機，開啓了工業革命，人們開始懂得以機械所產生的動力取代人力，藉由機器自動化運轉動力大量製造產品；西元1776年，亞當斯密《國富論》中，為分工和專門化訂出規範準則；西元1950年，Edwards Deming提出Total Quality Management（TQM），就製造業訂出品質管理規範；約莫第二次世界大戰後期，在以德國及日本為主工業國家，開始重視製造業的品質管制，並調整製程將產品臻於精緻。

西元1946年，第一台可載入程式的電子計算機問世，開啓了資訊革命，人們開始懂得以電腦所產生的處理能力取代腦力，藉由電腦快速計算能力大量處理資料；然而，Edwards Deming之Total Quality Management

（TQM）中，並未就資訊產業訂定出品質管理規範，西元1993年，Larry English提出Total Information Quality Management（TIQM），就資料品質管理定出規範準則（principle），並預期未來隨著電腦設備效能日益增強及網路通訊速率日益快速，資訊產業將會面臨三種可能結局：一、若企業仍不訂定出資料品質管理規範，氾濫的資訊將牽制企業的效益及獲利，並加速企業的衰退；二、若企業將「發現錯誤後再更正」視為資料品質管理，該高成本的管理雖暫時能停減緩惡化，但潛在的危機終究導致企業邁入衰退；三、若企業訂定出資料品質管理規範，該項管理將為企業營運帶來績效。

資訊的生命周期

資訊的建置是長期性、連續性及延續性的工作，資料在經由定義、擷取、傳遞、檢核、建置、彙集、儲存、揭露等一連串流程後，才能達到資訊具體呈現，此將「資料」製作成為「資訊」的過程，稱為資訊生命周期，可分為：資料規劃、資料獲取、資料儲存、資料維護、資料應用、資料下線等六個階段；值得注意的是，資訊的價值僅於資料應用階段中產生，然而各項階段卻均擔負著資料品質的責任。

從資料收集到資訊呈現，可說是腦力與勞力、軟體與硬體、知識與技術協力配合的成

資訊生命週期



果；然而，影響資料品質之要素依複雜程度分別為：反映事實真相之資料（data）、處理該資料之程序（process）、參與該資料處理程序之人員（people）、參與該資料處理程序人員所使用之技術（technology）等四項。倘若以資訊生命週期六個階段為橫軸，再以影響資料品質四項要素為縱軸，所建構出來之矩陣稱為「資料品質架構」，如下圖。也就是說，任何資料品質的問題皆可由此架構中剖析出來。

資料品質架構

	資料 規劃 Plan	資料 獲取 Obtain	資料 儲存 Store	資料 維護 Maintain	資料 應用 Apply	資料 下線 Dispose
資料 Data	?	?	?	?	?	?
程序 Process	?	?	?	?	?	?
人員 People	?	?	?	?	?	?
技術 Technology	?	?	?	?	?	?

資料品質管理的方法

資料品質管理最棘手的部分，較不在於該使用何種方法處理，而在於處理的方法充斥如多頭馬車，尤其在不同階段中之不同經辦人員各自有其不同的處理方法，思考順序及採用技術欠缺組織，往往造成日後資訊整合或轉換時

產生資料品質問題；因此，企業應導入資料品質管理方法論（methodology），將各階段之處理方法（method）組織起來，並訂定出規範供經辦人員共同遵循；傳統上，製造業及服務業所採用之品質管理方法論有二：

一、P-D-C-A 循環：

Plan	計劃	5W1H：What、Why、Who、When、Where、How
Do	分析	Cause/effect analysis、cost/impact analysis
Check	檢驗	發現問題、控制問題、接受變通、改善問題、成長學習
Act	執行	修正/調整/確認、建立標準化、釐清殘留問題

二、「六個標準差」（Six Sigma）所採用之

D-A-M-I-C循環：

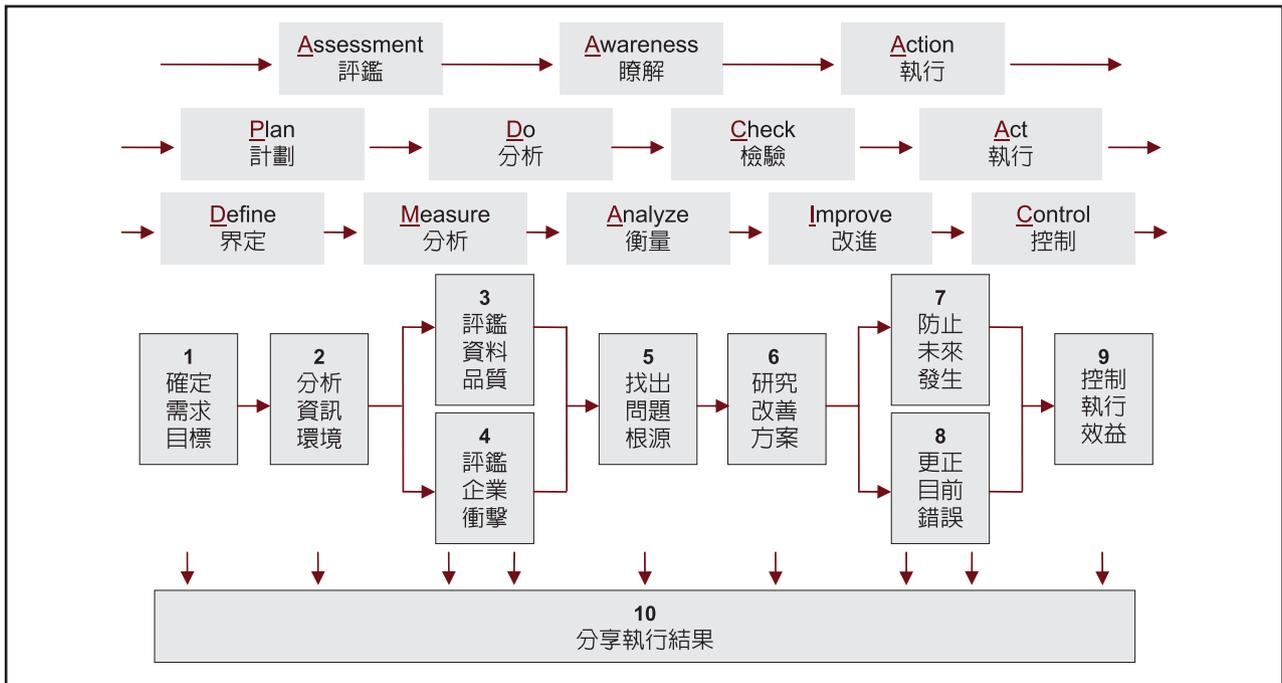
Define	界定	確認問題、界定需求、設定目標
Analyze	分析	因果假設、找出根源、驗證假設
Measure	衡量	驗證步驟流程、評估衝擊程度
Improve	改進	尋求解決方法、評鑑方案績效
Control	控制	維持績效、釐清細節、建立制度

然而，本文提出資訊業採用之品質管理方法論，稱A-A-A循環：

Assessment	評鑑	評鑑資料品質的優劣
Awareness	瞭解	瞭解資料品質的原因
Action	執行	執行資料品質的改善

資料品質管理的十項步驟

資料品質管理計有十項程序步驟，依序



為：確定需求目標、分析資訊環境、評鑑資料品質、評鑑企業衝擊、找出問題根源、研究改善方案、防止未來發生、更正目前錯誤、控制執行效益、分享執行結果，如下圖，上述十項步驟均符合前述三種品質管理循環（cycle），皆藉由不斷反覆修正的程序，將問題影響層面降低，藉以提升資料品質。

步驟一：確定需求目標

資料是企業提升績效的基礎，一組冷冰冰的資料實體（data），若穿上資料格式的外衣，就成為一則具體訊息（information），企業據以做出經營決策（action），藉以改善營運績效（performance）；反過來說，企業若想提升營運績效，必須做出正確的經營決策，而正確的經營決策來自於準確的市場資訊，而準確的市場資訊又來自於正確的資料。

步驟二：分析資訊環境

若要充分掌握資料品質，就必須先對整個資料處理環境有完整瞭解，資料處理環境包

括：資料品質的需求、資料內容的定義、資料建置的流程、資料處理的人員、資料存取的技術、資訊生命的期限、資訊評析的方式等七部分，其中，資料內容的定義、資料建置的流程、資料處理的人員、資料存取的技術四項程序，讀者應不陌生，係前述所提及「資料品質架構」中影響資料品質的四項要素。

步驟三：評鑑資料品質

在評鑑資料品質良莠之前，必須先解析資料品質好壞之維度，共分為正確性、真實性、新鮮度、即時性、順服度、完整性、連續性、一致性等八項：

正確性	以資料處理者角度評析資料內容是否符合定義？
真實性	以資訊應用者角度評析資料內容是否反映現實？
新鮮度	資料內容是否反映最時新之狀態？
即時性	資料於遞送及建置過程中有無延遲？
順服度	單筆資料中，各項欄位有否完整報送？
完整性	單項欄位中，資料筆數有否完整報送？
連續性	同一項目資料，前後期別之資料內容有否連續？
一致性	同一期別資料，不同項目之資料內容是否一致？

值得一提的是，「順服度」與「完整性」二項資料維度，若搭配檢核策略時，則往往出現彼此牽制，不易兼顧的現象；若資料的檢核策略寬鬆，換言之，單筆資料中之各欄位只要有一項欄位具有價值就視為檢核通過，則該寬鬆檢核策略將導致該項資料「順服度低，完整性高」；若資料的檢核策略嚴格，換言之，單筆資料中之各欄位只要有一項欄位檢核不符就視為錯誤剔退，則該嚴格檢核策略將導致該項資料「順服度高，完整性低」。

步驟四：評鑑企業衝擊

資料品質對於企業的衝擊共可分為四種型態：成本的增加、風險的增高、收入的減少、信認的降低；然而，如何準確評鑑出資料品質不良對於企業造成的衝擊，卻不容易，必須不斷地腦力激盪，培育出敏銳的智慧，才能做出準確的評估，本文提出「問5次為什麼」做為評鑑企業衝擊的標準程序，其重點不在問幾次？而在企業員工面臨陳述問題時，往往有避重就輕、實問虛答的習慣，導致在評鑑企業衝擊時失焦。

步驟五：找出問題根源

如何找出資料品質問題的根源，也不是容易的工作，必須不斷地腦力激盪，累積出豐富的經驗，才能做出準確的判斷，本文提出「問5次為什麼」做為找出問題根源的標準程序，其重點也不在問幾次？而在企業員工於面臨分析問題時，往往有頭痛醫頭、腳痛醫腳的毛病，導致在找出問題根源時迷失。

步驟六：研究改善方案

改善方案執行之前必須先進行評估分析，依其成本（cost）及衝擊（impact）的程度不同，可將改善方案區分為「高度衝擊，低度成

本」、「高度衝擊，高度成本」、「低度衝擊，低度成本」、「低度衝擊，高度成本」等四種類型，其處理優先順序亦是不同：若改善方案屬「高度衝擊，低度成本」，則第一優先，應該毫不猶豫，立刻執行；若改善方案屬「高度衝擊，高度成本」，則第二優先，可儘量將該改善方案分割成多個「高度衝擊，低度成本」次級方案後再逐項執行；若改善方案屬「低度衝擊，低度成本」，則第三優先，可以等到行有餘力時再去執行；若改善方案屬「低度衝擊，高度成本」，則其順位最低，可以考慮暫不執行。

步驟七：防止未來發生

要防止資料品質問題日後再度發生，進而將問題消弭於無形，必須就「步驟五：找出問題根源」中所找出問題發生的根源（root causes）徹底規劃改善方案，倘若問題未看到全面，思考未抓到重點，最後發現問題始終原地踏步，錯誤仍然持續發生。

步驟八：更正目前錯誤

對於資料品質問題，應建置個案處理措施因應，由於「步驟七：防止未來發生」中之改善方案必須經過系統規劃、程式設計、功能測試等一連串繁瑣程序後方可落實，而眼前的問題卻不可因此而等待，必須仰賴過渡性個案處理措施。此外，個案處理措施係一種臨時性的改善方案，通常僅能解決局部性問題，而無法將所衍生問題全面性解決，所以切切不可將此個案處理措施視為「步驟六：研究改善方案」，否則「資料品質管理」將不斷地在「步驟一：確定需求目標」與「步驟八：更正目前錯誤」二處無止境地往返。

步驟九：控制執行效益

成熟的控制機制涵蓋資料處理流程（data process flow）原始資料→資料檢核→資料庫儲存→資訊應用四個階段；而效益控制有「後知後覺型」及「先知先覺型」兩種角度；所謂後知後覺型控制（reactive control），其控制順序與資料處理流程相反，係當資訊於應用時發現資料品質不良時，再探討資料儲存之處有無問題，若確定資料庫儲存無問題，再研判資料檢核時有無問題，若確定檢核仍無問題，則再研究原始報送之資料有無問題，此被動控制方式必須付出相當高的代價；反之，先知先覺型控制（predictive control），其控制順序與資料處理流程相同，係當規劃原始資料之報送時，就預測出資料檢核時之可能問題，再預測出資料儲存之處之可能問題，再預測出資訊應用時可能發生資料品質的問題，此控制方式乍聽之下極完美，但實際執行時會發現其理想性過高，其效益與預期相差甚遠；本文提出「積極主動型」效益控制（proactive control）方式，係指：在原始資料↔資料檢核↔資料庫儲存↔資訊應用各個階段中，設計主動監控機制，往返查驗有無資料品質的問題，期盼資料品質問題能於資訊應用階段之前就被發現，並佐以處置作業機制，藉以降低衝擊。

步驟十：分享執行結果

上述九項資料品質管理程序的執行過程，不論是成功經驗或是失敗教訓，均為企業內部彼此分享的題材；愛因斯坦說：「聰明的人能解決問題，智慧的人能避免問題」，希臘哲人也說：「智慧的人乃是從別人所犯的錯誤中得到教訓，進而更正自己的錯誤」，再再提醒經驗分享及傳承的寶貴。

結語

西方企業較重視以方法論規範專案中之作業流程，舉凡P-D-C-A、D-A-M-I-C、A-A-A等循環，採用方法論之優點誠如前述，專案中各項程序作業之功能及權責界線均已劃分清楚，除了便於日後職務調整時，代理人能立即接替外，其各項程序之作業文件亦主動呈現；惟其缺點在於整個專案被分割成數個程序步驟，各程序步驟間必須藉由內部討論作為溝通介面，所以，各程序負責人彼此間必須具備嫻熟的溝通默契，否則，來自本位主義的計較，勢必耽誤整個專案落實的時效。

東方企業對於專案的推動，往往遷就於時效，省略或直接跳躍既定的程序步驟，此機動性變通作業，雖然能使專案在最速時效內完成，惟其缺點在於相關程序因行事倉促而不透明，而作業文件往往因「後補」而流於形式，此外，專案推動過程中各程序步驟之實務經驗也較不易分享流通。

在掌握資訊就掌握競爭優勢的金融環境中，資料品質管理在時間、人力等資源有限的情況下，是一項團隊工作，非一蹴可及，唯有聯徵中心與金融機構協力配合、凝聚共識、取得認同，並透過反覆溝通、檢討、修正、補充等循環，才能呈現精緻性的資訊內容品質，增進信用資訊的附加價值，落實資訊互惠共享的原則；既然「資料品質管理」已成為健康代名詞，就讓我們如吸呼空氣般的自在去實踐吧！