

聯合徵信中心 資料品質管理實務

朱昌綸 / 金融聯合徵信中心資訊部資料品質小組

前言

資料品質分析實驗時，分析人員經常面對一種方法上的惰性，即根據既有的知識而產生期望，如果得到的結果不符合期望，則總是想辦法找出各種理由去排除那個結果，而不會去深究那個不符期望的結果為何會產生。在資料品質分析時，若存有這種心態，將只會活在期望的甜蜜陷阱裡，它只會反芻與保守，不可能發現與創新。

本文謹就聯徵中心資料品質管理的角色，資料品質檢核的範圍、資料品質檢核的類型，綜合引述說明。

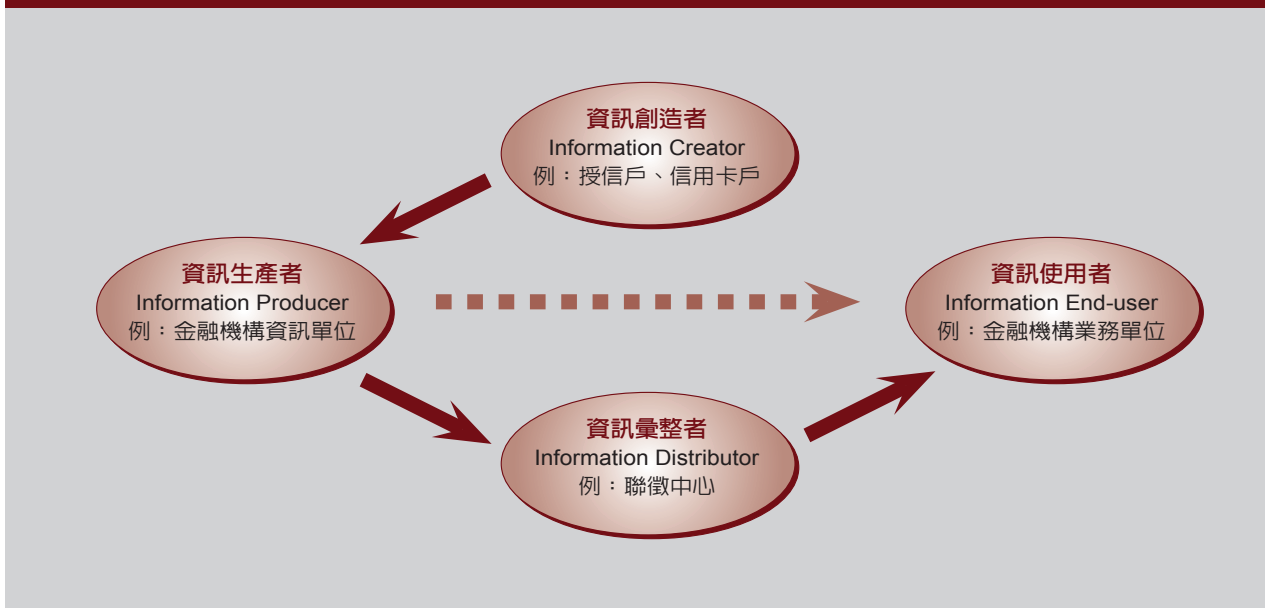
聯徵中心資料品質管理的角色

「資料」轉為「資訊」的過程，稱為「資訊供應鏈」（Information Supply Chain，如圖一所示），共有四項角色，任一角色都必須依賴其他角色提供品質服務，亦即，任一角色必須承擔應有品質管理的責任。

1. 資訊創造者（Information Creator），係指與金融機構發生借貸交易之往來客戶，如授信貸款戶或信用卡持卡戶。
2. 資訊生產者（Information Producer），係指將資訊創造者之資料製成檔案，再傳遞至資訊彙整者處理，如金融機構之資訊單位。
3. 資訊彙整者（Information Distributor），係指將資訊生產者之報送資料彙集，再提供資訊使用者查詢，如聯徵中心。
4. 資訊使用者（Information End-user），係指參考利用資訊彙整者所揭露資訊，如金融機構之業務單位。

在過去，資訊生產者與資訊彙整者的角色容易區別，資訊生產者擷取資訊後，交付資訊彙整者彙集，再提供資訊使用者查詢；因此，資料品質管理的責任，若從資訊使用者的角度來看，因為不清楚角色的分工，僅以查詢結果評斷資料品質的優劣，直覺上認為資料品質管理應是資訊彙整者的責任；反之，對於資訊彙

圖一、資訊供應鏈 (Information Supply Chain)



整者而言，資料既是由資訊生產者提供，資料品質管理當然是資訊生產者的責任；囿於資訊供應鏈上角色的模糊，資訊使用者的資料品質需求，屢屢在資訊彙整者本位主義考量下遭到忽視。

近年來，由於電腦效能增強、通訊速率快速、技術成本滑落，資訊生產者開始思考兼具資訊彙整者的角色，資訊供應鏈上的角色既相互合作也相互競爭，使得原本角色間的界線日漸模糊，此一趨勢，對於聯徵中心的經營與發展，將是嚴竣的挑戰。

聯徵中心資料品質檢核的範圍

1. 資料收件檢核 (authentic check)：

資料若以媒體報送，則處理人員必須確認媒體遞送單內容；資料若以檔案傳輸

系統報送，則系統必須自動確認檔案發送來源。聯徵中心處理資料之前，系統仍須就資料檔案首筆 (header) 及末筆 (trailer) 之資料屬性再進行確認，若有不符，則逕行退回資料檔案。

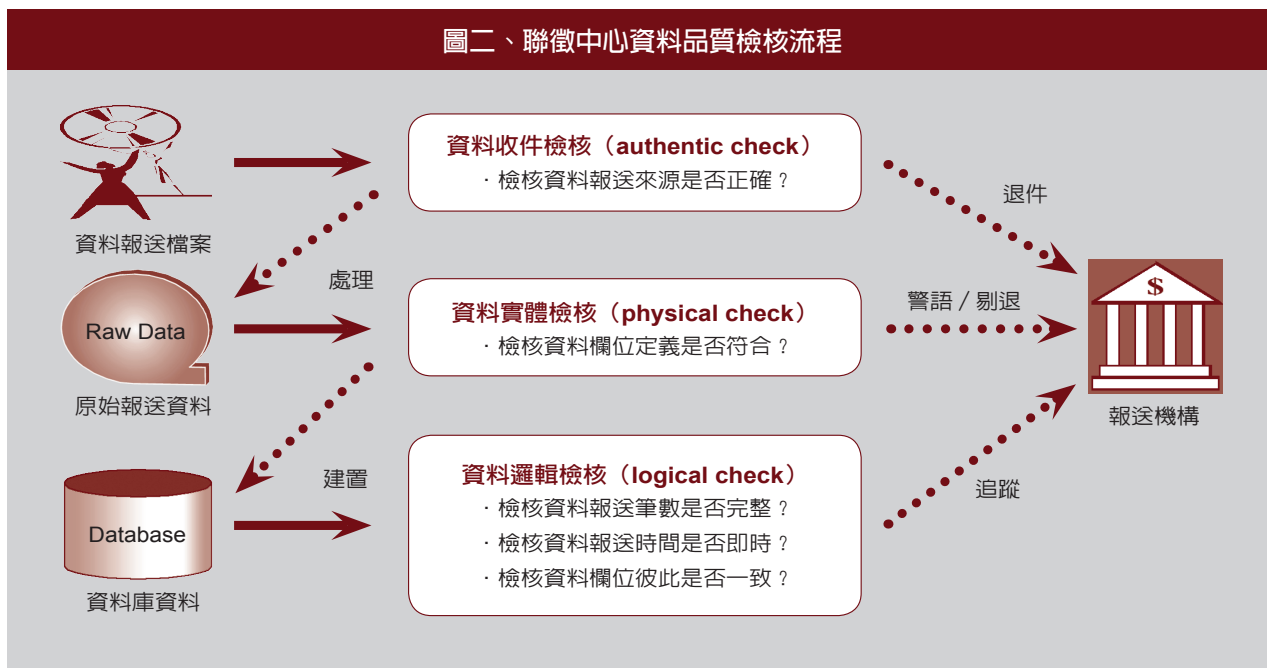
2. 資料實體檢核 (physical check)：

資料實體檢核係屬資料建入資料庫前，先行的檢核作業，當聯徵中心建檔程式發現報送資料內容不符合定義時，系統旋即剔退該錯誤資料，並產生錯誤明細檔案回傳原報送機構，要求更正後重新報送。

3. 資料邏輯檢核 (logical check)：

資料邏輯檢核係屬資料建入資料庫後，再行的檢核作業，聯徵中心定期勾稽比對資料庫中各項資料，當發現資料間彼此矛盾時，經辦人員將通知資料報送機構要求改善，並持續追蹤改善進度。

圖二、聯徵中心資料品質檢核流程



聯徵中心資料品質檢核的類型

聯徵中心定期就資料之報送來源、項目、所屬期別、欄位內容、報送筆數、報送日期等屬性，進行勾稽比對分析，資料檢核的類型歸納如下。

1. 檢核資料欄位內容之正確性 (Is data accurate?)，包括：

- (1) 檢核報送機構所報送資料項目，單筆資料中個別欄位之內容是否符合定義？
- (2) 檢核報送機構所報送資料項目，單筆資料中相異欄位之內容是否彼此矛盾？
- (3) 檢核報送機構所報送資料項目，全部資料中相同欄位之筆數分佈有否異常？

2. 檢核資料報送時間之即時性 (Is data immediate?)，包括：

- (1) 檢核尚未報送－檢核新設立金融機構或會員有否依規定報送資料？
- (2) 檢核忘記報送－檢核資料報送機構有否依規定確實定期報送資料？
- (3) 檢核延遲報送－比較資料發生日期與資料報送日期有無明顯差異？

3. 檢核資料報送筆數之完整性 (Is data complete?)，包括：

- (1) 比較同一報送機構、同一資料項目、前後期別之資料筆數有無明顯差異？
- (2) 比較同一報送機構、不同資料項目、同一期別之資料筆數有無明顯差異？
- (3) 比較不同彙集機構、同一資料項目、同一期別之資料筆數有無明顯差異？

4. 檢核資料彼此相關之一致性 (Is data consistent?)，包括：

- (1) 比對同一戶、同一報送機構、同一資料項目、前後期別之資料內容有否連續？
- (2) 比對同一戶、同一報送機構、不同資料項目、同一期別之資料內容是否一貫？
- (3) 比對同一戶、不同報送機構、同一資料項目、同一期別之資料內容是否一致？

茲列舉聯徵中心資料檢核實務做為案例，分別說明如下：

1. 資料正確性檢核

(1) 檢核報送機構所報送資料項目，單筆資料中個別欄位之內容是否符合定義？

案例：A銀行「授信餘額月報資料」中，欄位「用途別代號」值為9。

說明：「用途別代號表」中有效值為1,2,3,4，該資料欄位定義不符。

(2) 檢核報送機構所報送資料項目，單筆資料中相異欄位之內容是否彼此矛盾？

案例：B銀行「信用卡資料」中，欄位「啓用日期」值為0990401，且欄位「停用日期」值為0990130。

說明：信用卡之啓用日期應早於停用日期，該資料欄位之內容矛盾。

(3) 檢核報送機構所報送資料項目，全部資料中相同欄位之筆數有否分佈異常？

案例：

| | | | | |
|-----|----------|--------|----------------------|-------|
| C銀行 | 信用卡戶帳款資料 | 99年04月 | 繳款代號：0 (無延遲) | 2345筆 |
| C銀行 | 信用卡戶帳款資料 | 99年04月 | 繳款代號：1 (延遲未滿1個月) | 0筆 |
| C銀行 | 信用卡戶帳款資料 | 99年04月 | 繳款代號：2 (延遲1個月至未滿2個月) | 1234筆 |
| C銀行 | 信用卡戶帳款資料 | 99年04月 | 繳款代號：3 (延遲2個月至未滿3個月) | 5678筆 |

說明：C銀行99年04月資料中，繳款延遲未滿一個月之筆數分佈異常。

2. 資料即時性檢核

(1) 檢核尚未報送－檢核新設立金融機構或會員有否依規定報送資料？

案例：A農會信用部於99年01月01日加入聯徵中心會員；但截至99年04月10日止，尚未報送「授信餘額月報資料」至聯徵中心。

說明：A農會必須依會員規約，報送「授信餘額月報資料」至聯徵中心。

(2) 檢核忘記報送－檢核資料報送機構有否依規定確實定期報送資料？

案例：

| | | | |
|-----|-------|--------|-------|
| B銀行 | 信用卡資料 | 99年03月 | 1234筆 |
| B銀行 | 信用卡資料 | 99年04月 | 0筆 |
| B銀行 | 信用卡資料 | 99年05月 | 1308筆 |

說明：B銀行99年04月末依規定，定期報送資料至聯徵中心。

(3) 檢核延遲報送－比較資料發生日期與資料報送日期有無明顯差異？

案例：C銀行99年05月12日報送「信用卡戶帳款資料」中，欄位「結帳日期」值為0990409。

說明：「信用卡戶帳款資料」必須於結帳日後5日內報送至聯徵中心。

3. 資料完整性檢核

(1) 比較同一報送機構、同一資料項目、前後期別之報送筆數有無明顯差異？

案例：

| | | | |
|-----|----------|--------|--------|
| A銀行 | 信用卡戶帳款資料 | 99年03月 | 39875筆 |
| A銀行 | 信用卡戶帳款資料 | 99年04月 | 39421筆 |
| A銀行 | 信用卡戶帳款資料 | 99年05月 | 10278筆 |

說明：A銀行98年05月之資料報送筆數明顯減少。

(2) 比較同一報送機構、不同資料項目、同一期別之報送筆數有無明顯差異？

案例：

| | | | |
|-----|----------|--------|--------|
| B銀行 | 信用卡戶帳款資料 | 99年03月 | 57129筆 |
| B銀行 | 信用卡戶基本資料 | 99年03月 | 14469筆 |

說明：B銀行98年03月之資料報送筆數明顯有差異。

(3) 比較不同彙集機構、同一資料項目、同一期別之資料筆數有無明顯差異？

案例：

| | | | |
|---------|---------|--------|-------|
| 聯徵中心 | 信用卡發卡數量 | 99年04月 | 2345筆 |
| C銀行網站公告 | 信用卡發卡數量 | 99年04月 | 6789筆 |

說明：C銀行99年04月之資料報送筆數明顯有差異。

4. 資料一致性檢核

(1) 比對同一報送機構、同一戶、同一資料項目、前後期別之資料內容有否連續？

案例：

| | | | | |
|-----|----|----------|--------|-----------------------|
| A銀行 | 甲君 | 授信餘額月報資料 | 99年03月 | 本月還款紀錄：1 (還款遲延30~59天) |
| A銀行 | 甲君 | 授信餘額月報資料 | 99年04月 | 本月還款紀錄：2 (還款遲延60~89天) |
| A銀行 | 甲君 | 授信餘額月報資料 | 99年05月 | 本月還款紀錄：6 (還款遲延180天以上) |

說明：A銀行甲君之「還款紀錄」欄位報送前後期明顯不連續。

(2) 比對同一報送機構、同一戶、不同資料項目、同一期別之報送內容是否一貫？

案例：

| | | | | |
|-----|----|----------|-----|---------|
| B銀行 | 甲君 | 信用卡戶基本資料 | 98年 | 職業類別：學生 |
| B銀行 | 甲君 | 授信戶基本資料 | 98年 | 職業類別：醫生 |

說明：B銀行甲君之「職業類別」欄位報送明顯不一貫。

(3) 比對不同彙集機構、同一戶、同一資料項目、同一期別之報送內容是否一致？

案例：

| | | | | |
|-----|----|----------|-----|-------------|
| A銀行 | 甲君 | 信用卡戶基本資料 | 98年 | 教育程度：1 (博士) |
| B銀行 | 甲君 | 信用卡戶基本資料 | 98年 | 教育程度：1 (博士) |
| C銀行 | 甲君 | 信用卡戶基本資料 | 98年 | 教育程度：5 (高中) |

說明：C銀行甲君之「教育程度」欄位報送與其他報送機構明顯不一致。

結語

聯徵中心資料品質提升的工作重點：

1. 加強資料定義的品質 (Definition Quality)

對內，藉由「資料格式規劃諮商會議」平台，評估資料蒐集策略及規劃欄位格式；對外，藉由「資料報送規範講習會議」平台，宣導資料報送規範及溝通報送機構。

2. 落實資料內容的品質 (Content Quality)

對內，藉由「資料品質檢核專案會議」平台，討論資料檢核邏輯及建立追蹤機制；對外，藉由「資料回饋方式說明會議」平台，透

明資料檢核機制及宣導回饋方式。

3. 提升資訊呈現的品質 (Presentation Quality)

對內，藉由「信用資訊產品開發會議」平台，檢討資訊產品內容精確度及易讀性；對外，藉由「資訊產品應用講習會議」平台，提升查詢機構服務及輔導資訊應用。

誠如前言，聯徵中心必須積極加強資料品質的管理，方可超越既往被賦予「垃圾進，垃圾出」(garbage in & garbage out)的刻板印象，也唯有如此，聯徵中心的信用資料庫才能獲得金融機構的信賴及依賴。