

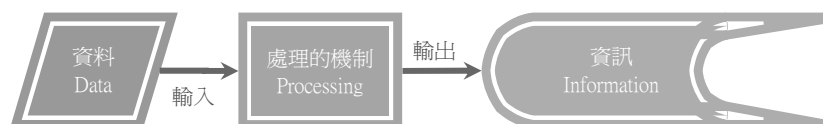
第2章 資料處理導讀

本章內容主軸在於：

- ✓ 資料處理的定義與特性。
- ✓ 資料、資訊與處理機制之間的關聯。
- ✓ 資料處理系統的作業方式與標準作業程序。
- ✓ 資料處理機制的常見類型與意義。
- ✓ 資料品質的重要性，與提升資料品質之驗證機制的意義與實例。

2.1 資料處理的基本觀念

資料處理係指「資料被轉換成爲資訊的處理機制」。爲了解資料處理機制的運作過程，首先必須理解何謂資料、資訊、處理機制。



- 資料：在電腦科學領域中，資料 (Data) 或被稱為數據，係指「適合透過電腦，來加以運用之任意形式的原始材料」，呈現著沒有特定意義之人事物相關的事實。在此，原始材料意指字元 (例：文字、數值、符號)、線條、形狀、顏色、音訊等等。
- 資訊：在電腦科學領域中，資訊係指「資料經過處理 (例：整理、分析) 而被轉變並輸出之有用的訊息」。在此，訊息意指有特殊意義的字串、報表、多媒體 (例：圖像、視訊、聲音) 等等。
- 處理機制：現今資料處理的機制相當繁多，凡舉資料的運算 (operation)、融合 (fusion)、分類 (classification)、取樣 (sampling)、存取 (access)、排序 (sorting)、搜尋 (searching)、編碼 (encoding)、解碼 (decoding)、傳輸 (transmission)、探勘 (mining) 等等，都是資料處理的機制之一。

現今任何類型的軟體 (例如：作業系統、應用程式) 都被建立在資料和程式碼的基礎之上，因此皆內建至少一種資料處理機制，以處理該軟體內

2-2 第 2 章 資料處理導讀

部所需的各項資料。也因為前述的特質 (性)，才使得該軟體具有特定價值與用途。廣泛來說，任意軟體大致可被視為多元化的資料處理系統。和資料處理的定義不同，**資料管理 (data management)** 係包含所有將資料視為寶貴資源，來加以管理的學科；而**資訊管理 (IM, information management)** 係指「聚集和管理來自單一或多個來源的資訊」與「散佈特定資訊給單一或多個對象」的學科。

資料管理、資料處理和資訊管理等 3 者有著許多研究領域是重疊或交錯的！未能被妥善歸納於資訊管理、資料處理領域之研究範圍的主題，則可被視為資料管理領域的主題，例如：資料資產 (data asset)、資料存取 (data access) 等等。

2.2 資料處理系統的作業方式與標準化

許多資料處理系統 (data processing system) 或被稱為資訊系統 (information system)，除了被設計了不同的處理機制，來滿足不同使用者欲獲取多種不同資訊的需求之外，也藉由不同的作業方式，來增加使用者的操作效率，以節省時間成本。

資料處理系統一旦被設計、更新與測試完成之後，應該由熟悉該系統的設計廠商或管理人員，為所有可能的操作人員，建立一套標準作業程序 (SOP, standard operating procedure)，以便操作人員可順暢地操作該系統與排除所遭遇的操作問題。

2.2.1 資料處理系統的作業方式

列舉如下資料處理系統的作業方式：

- 批次/整批處理 (batch processing) 具有以下特性：
 - 毋須立即執行每次已排定的工作，一當累積的工作件數到達一定數量時，操作人員才事先對特定資料處理系統，進行週期性自動執行的相關設定，以驅使該系統一次處理一連串程序。
 - 操作人員不會獲得其立即的結果資訊。
 - 有些資料處理的工作過於冗長、乏味、耗費人力資源；在充份利用電腦資源的情況下，當電腦處於較不繁忙的期間內 (例：凌晨)，可讓資料處理系統透過批次處理，自動進行多種資料處理。例如：龐大的公務人員進行工商普查之後，會留下大量新版本的工商資

料，可在這些資料被建檔之後，經由批次處理，由電腦代為處理其後續工作。

- 所需硬體與軟體都可以較簡單，其操作方式也比較單純，但不會降低單位時間內的電腦工作量。

➤ 即時處理 (real-time processing) ，其具有以下特性：

- 資料處理系統必須迅速進行特定資料處理，並立即輸出相關的結果資訊。
- 判斷特定系統是否足夠**及時迅速**的指標，即是透過測量「該系統被用戶操作而下達指令之後，用戶得到該系統之回應動作或訊息」的**反應時間 (response time)**。例如：飛機導航、緊急剎(煞)車、股票行情、銀行提存款等等系統，都是支援即時處理的作業方式。
- 能夠儘量保持其**結果資訊處於最新或最正確**的狀態。
- 能夠儘量**立即**處理已排定的工作事項。

➤ 分時處理 (time-sharing processing) ，其具有以下特性：

- 多項**程序**或多個**程式**共同分享該系統的**執行時間**，以便讓**多項**工作能在一段期間內，**同時被處理**。
- 能個別提升電腦內部之各硬體的效能。
- 能讓**份量較小**的工作，不必**一直等待**正在被處理之工作的完成，即可被分段處理而較早達成。
- 分時處理之大致的步驟為：
 1. 把即將被處理的各個處理程序 (process) ，全部載入到主記憶體佇列 (queue) 當中。
 2. 作業系統會把**特定長度**的**執行時間**，再細分成細小的**單位時間**。
 3. 在**每個**單位時間內，CPU 快速地『循環』執行各個**處理程序**。
 4. 遇到特定處理程序需要進行輸入或輸出工作的時候，CPU 會將其後續工作，交由 I/O 裝置去處理，並將此處理程序，再送回佇列，以等待下一次的執行；CPU 隨即會繼續執行下一個處理程序。
 5. 如此一來，這一段期間內，看起來就像同時執行許多不同的工作一般。
- 其優點：特定處理程序需要進行 I/O 工作時，CPU 的時間可被開放給下一個處理程序來使用，以節省 CPU 的時間。

2-4 第2章 資料處理導讀

- 其缺點：若一些冗長的處理程序完全不用進行任何的 I/O 工作時，則這些處理程序會持續『循環』佔用較多的 CPU 時間，導致其他處理程序不公平地少用了 CPU 時間。
- 分散處理 (distributed processing) ，其具有以下特性：
 - 在資料處理的過程中，資料處理系統會透過分散在網路各地之程式與資料，來逐一或同時進行個別的運算，最後再整合這些運算，並產出綜合性之結果資訊。
 - 可充份透過距離，來區隔操作人員與重要資料，以維護重要資料的安全性，亦可充份運用到分散在各地的電腦資源，以便讓過於複雜、繁瑣、困難的資料處理程序，能儘快被解決。
- 連線處理 (connection processing) ，泛指資料處理系統透過網路，進行任何形式之連線的作業方式。現今幾乎所有資料處理系統，大都支援使用者可連線至特定系統之資料庫，以存取資料。

特定資料處理系統被設計而支援的作業方式，與該系統之如下因素有關：

- 所需資料的數量
- 輸入資料的時間點
- 所在位置

大多數的資料處理系統可同時支援多種作業方式，例如：自動提款機 (ATM, automated teller machine) 允許多個使用者可同時使用，而且具備即時處理、連線處理等等作業方式的資料處理系統。

2.2.2 資料處理系統的標準作業程序

進行資料處理系統的作業標準化，係指建立該系統的標準作業程序 (SOP, standard operating procedure)：「透過步驟化或圖形化操作流程，讓資料處理系統的例行性工作，均能被呈現其操作步驟的前後關係與意義，以便讓操作人員，熟悉各項工作的操作細節，進而節省其時間成本。」

然而，隨著資料處理系統的改版與替換，相關人員也必須隨之修改對應的 SOP，以免反而阻礙或延宕各項工作的進行。