

# 第 2 課：數據控制及數據組織 (Data Control and Organization)

## 2.1 數據分級 (Data Hierarchy)

數據通常以分級形式儲存，每個較高層的數據都是較低層的集合。

表 2.1 數據層中不同層級的示例

層級	說明
欄 (field)	儲存實體的特定屬性，例如學生的姓名。它可儲存何種類型的數據，例如數字或文本。
記錄 / 列 (record)	集合某實體的所有欄位，例如學生 ID、姓名、性別、班級和學號。
檔案 / 數據表 (file/table)	集合結構相同的記錄，例如一張儲存了 100 名學生資料的數據表。
數據庫 (database)	是相關檔案的集合。例如，一間學校的數據庫可能包含：儲存學生的數據表；儲存教師資料的數據表；儲存教室資料的數據表；及儲存課程時間表資料的數據表。

## 2.2 數據控制 (Data Control)

- 「無用輸入，無用輸出」 (Garbage in, garbage out, GIGO): 當不可靠或不協調的數據被輸入到數據分析中時，便不能得出有用的結果。

表 2.2 常見錯誤列表

錯誤類型	說明	例子	解決方法
數據來源錯誤	數據來源提供了不正確的數據。	用戶錯誤地提供了一個不存在的電子郵件地址。	向用戶提供的電子郵件地址發送認證碼，然後要求用戶輸入認證碼。
轉錄錯誤	數據被誤讀、誤聽或誤植。	<ul style="list-style-type: none"><li>光符識別 (Optical Character Recognition, 簡稱 OCR) 程式將大楷字母「I」誤識為小楷字母「l」或數字「1」。</li><li>打字員將「泊」打錯為「拍」。</li></ul>	<ul style="list-style-type: none"><li>通過檢查單詞的拼寫來檢驗數據的有效性。</li><li>要求用戶再次輸入數據。</li><li>校對數據。</li></ul>
換位錯誤	數據位置不正確。	<ul style="list-style-type: none"><li>蘋果的價格被儲存為西瓜的價格。</li><li>某人於記錄一項交易時誤將兩個相鄰數字調轉，令「\$15」變成「\$51」。</li></ul>	<ul style="list-style-type: none"><li>校對數據。</li></ul>

## 數據控制方式：

- 數據有效性檢驗 (Data Validation): 由電腦進行，目的確保輸入數據合理且符合規則。
  - 常見類型：完整性檢查、固定值檢查、類型檢查、範圍檢查、長度檢查、格式檢查、一致性檢查、唯一性檢查。
    - 檢查數位 (Check Digit): 附加於數據末尾（例如香港身份證號碼）。
  - 奇偶檢測 (Parity Check): 檢查數據傳輸過程中錯誤的方法。

表 2.3 數據有效性檢驗方法一覽

類型	功能	例子
完整性檢查	確保必填欄位中有數據。	申請人必須提供姓名。
固定值檢查	確保數據只能來自一系列特定的值。	申請人的性別必須是「M」或「F」。
類型檢查	確保數據必須為特定類型。	申請人的年齡必須是數字。
範圍檢查	確保數據必須位於指定範圍內。	申請人的年齡須大於 12 歲。
長度檢查	確保數據的字數在指定範圍內。	申請人需提供 8 位電話號碼。
格式檢查	確保數據必須採特定格式。	申請人電郵地址必須有「@」符號。
一致性檢查	確保數據與其他欄位無不合理之處。	結束日期需在開始日期後。
唯一性檢查	該欄位數據必須獨一無二。	學生 ID 不可重複。

表 2.4 用戶界面設計一覽（部分內容）

類型	有效性檢驗	例子（示例控制項）
下拉式選單	固定值檢查：只能輸入特定值	[性別下拉選單]
單選按鈕	固定值檢查：只能輸入特定值	[參與活動選項]
複選框	固定值檢查：可選取多個指定值	[同意條款複選框]
日期選擇器	類型和格式檢查	[可出席日期日曆控制項]

表 2.5 數據類型一覽

數據類型	說明	數據值例子	用途例子
整數	可為正/負	74、-29、0	學生人數、年齡
浮點數/實數	可有小數位	74.44、-2.09、1.00	分數、身高
字串/文字	由字符組成	"RA^1s"、「日」	姓名、地址
布爾	TRUE 或 FALSE	TRUE、FALSE	是否合格

表 2.6 奇偶檢測的類型

原始二進制代碼	1 的總數	偶數檢測	奇數檢測
0000 000	0	0000 0000	0000 0001
1010 101	4	1010 1010	1010 1011
0100 101	3	0100 1011	0100 1010
1110 011	5	1110 0111	1110 0110

- 數據驗證 (Data Verification): 由人手進行，幫助驗證數據在媒體之間傳輸時是否準確。

表 2.7 數據驗證方法一覽

類型	說明	例子
輸入數據兩次	操作員須輸入同一數據兩次	設置新密碼時需輸入兩次
雙重數據輸入	兩位操作員各自輸入並比對	兩人輸入同一問卷數據
校對數據	將輸入數據與原始檔案比較	打印成績單並與試卷核對

## 2.3 數據庫的功能 (Database Functions)

- 組織：數據根據結構組織，結構決定欄位的數據類型。
- 儲存：用 結構化查詢語言（SQL）或 圖形用戶界面（GUI）插入/更改/刪除記錄。
- 提取：用 SQL 或 GUI 提取特定數據。
- 索引 (Index)：為欄位建立索引提升提取速度。

表 2.8 數據表結構示例

欄位名稱	數據類型	說明
SID	字串	學生獨一無二編號。數字無法儲存前置零，須用文本儲存。
NAME	字串	學生姓名
SEX	字串	學生性別
CLASS	字串	學生班級
CNO	整數	學生學號

## 2.4 檔案存取方式 (File Access Methods)

數據存取方式有兩種：

- 順序存取 (Sequential Access): 儲存設備按順序從開頭讀入數據。
- 直接存取 (Direct Access): 儲存設備每位址唯一，可直接定位。

表 2.9 磁帶和硬碟之間的比較

屬性	磁帶	硬碟
檔案存取方式	順序存取	直接存取
數據存取速度	難以預測，通常較慢	穩定，通常較快
容量單位價格	較低	較高
輕便性	較高	較低
壽命	較長	較短
常見用法	儲存備份	儲存常用數據