如何建立一個良好的農作資料庫

◎農試所農藝系　魏夢麗、呂椿棠、呂秀英

一、為何要建立農作資料庫

資訊是試驗者制定計畫的必備知識，透過電腦、區域網路甚至國際網際網路進行大量相關資料蒐集及分析的方式，可說是試驗者取得良好資訊的最快速途徑。針對不同農作物必須建立其完備的農作知識與技術，一個完整的農作(耕)資料庫從種子生產、種苗培育、植株生長、以至產量收穫與調配等一系列過程中，舉凡育種試驗、土壤測量、農業記錄、氣候資訊及種原資料等各項農耕及生態資料都應納入系統。以往許多農業試驗成果及生態資料，大都是以文字記錄或圖表描述方式分散於各期刊報告中，而試驗者也多以文件或不同的電腦檔案形式（如試算表、文書處理及統計分析等）予以保存，這些資料（訊）相當繁瑣且零散，也常因軟體版本之一再更新而造成資料的無法讀取及難以整合，同時分散的資料亦常導致遺漏，不利於資料的存取運用。因此如何將這些龐大的農作資料予以適當整理及歸納，使這些資料能有效地被應用並進而連結統計分析以形成資訊系統，首要工作是必須建立一個規劃完善之整合性的農作資料庫。

二、資料庫管理系統的發展及特性

所謂資料庫，就是針對特定目標進行資料的蒐集及組織後將其儲存在電腦內。傳統的資料處理方式常是每個資料庫各自獨立，而各資料庫都是因應不同的需求所設計，此種處理方式在程式設計上比較單純，但也因而衍生了許多困擾和問題，例如相同的資料會重複出現在不同的資料庫，不但造成記憶體空間的大量浪費，若想更新某筆資料就必須同時更正所有包含該資料的資料庫，否則就會造成資料的不一致，因此反而需要花費更多時間、精力和成本來維護資料庫；又由於不同資料庫之間的資料往往有密切關聯，倘資料分散存放於各個不同的資料庫，卻無法予以整合，資料之存取處理便會顯得非常複雜，同時更難保證所查詢到的資料是否完整，而資料的安全性也堪虞。

近幾年來，資訊的應用突飛猛進，為改進以往缺點，做到減少重複、維持資料完整性及資源共享的特性，同時實現安全性限制，而發展出資料庫管理系統（Database Management System，簡稱DBMS）。DBMS的觀念源自於1960年代開始，自1970年起相關的理論才蓬勃發展。DBMS不僅能維護資訊及依需求擷取資訊，更重要的是能對資料庫進行控制及管理。欲達到有效率的管理，

作者：魏助理研究員夢麗
聯絡電話：04-3302301轉129

在建置資料庫之前，應該清楚地決定哪些資料是我們需要記錄的，而哪些是可以被忽略的，它們之間的關係又是如何。這個重要程序稱爲資料模式化（data modeling），藉此產生一個資料模式（data model）。資料模式的設計宗旨是必須將所有訊息予以整合而不分散，對使用者提供有效且完整的資訊。爲了讓使用者可不經深思而很輕鬆地進行資料的查詢擷取，建立資料模式所使用的語言必須精確且簡單：精確的語言模式可避免因使用者之解釋不同而產生誤解，簡單的語言模式易懂且方便傳達聯繫。

從資料庫管理系統的歷史，可依照資料的結構及操作方式，將其加以分類。遠從資料庫之前的檔案(files)起，資料庫的系統模式歷經反轉串列(inverted-list)、階層式資料庫(hierarchy)、網路式資料庫 (network)，逐漸演進至目前最廣泛使用的關聯式資料庫（relational database）。例如時下流行的Sybase的Adaptive Server、Oracle的Universal Server、 Infomix的 Universal Server、Microsoft的SQL Server、Microsoft的Access 97及Visual FoxPro、IBM\DB2等均屬關聯式資料庫管理系統。關聯式資料庫管理系統（簡稱RDBMS）的基本概念爲闡述資料間的關係以關聯式的方式來組織，並且應用程式不應與資料的內部儲存結構有任何相依關係（此即資料獨立的概念－應用程式不依賴於任何一種特定的儲存結構或是存取策略）。在此模式下，應用程式不會因爲將來資料內部結構的修改而有所牽連，而會因應新的儲存結構而做大幅度的改變，因此應用程式的維護較爲簡單。關聯式資料庫的基本結構形式爲：

㈠使用者所看到的都是資料表（table）形式，整個資料庫分成數個資料表來儲存資料，這些資料表稱之爲「關聯資料表」，每個關聯資料表由一筆或多筆記錄（record）所集合而成，而每筆記錄可包含一至多個資料項目（item，相當於資料欄位）。

㈡各資料表之間存在某些限制條件，以維持資料完整性。

㈢使用者可使用的運算如資料新增、刪除、擷取，都是從舊有的資料表中產生新的資料表。

關聯式資料庫雖具有有效整合文數字資料，形成單純的表格及簡易的查詢語言(SQL, 爲一種世界標準的結構查詢語言－structure query language)的能力，但卻無法滿足圖形、影像及多媒體等各項資料型態的需求。最近由於物件導向技術（object-oriented program, 簡稱OOP）的開發，提供解決這項難題。所謂物件導向，是將資料的基本觀念抽象化，視資料處理的對象爲一個物件(object)，並允許使用者自定資料型態，除了文數字型態，包括圖形、程序及函數等都可依附在欄位或變數上，也可定義其他類別，因此豐沛了可用的型態。在資料庫管理系統內所有具備名稱的個體皆可被視爲是一個物件，如資料表、查詢、表單、報表、巨集、模組、視窗、控制按鈕等。根據物件導向程式設計原則所建立的關聯式資料庫，是目前資料庫管理系統的主流。

三、關聯式資料庫管理系統在農業上之應用

利用關聯式資料庫架構建立農業資訊系統的成功例子有：Fox *et al.*（1996）的國際

小麥資訊系統(IWIS)，以收集各種不同基因型小麥的基因譜系及雜交育種親緣資料；IRRI(1997)進而將IWIS擴大應用範圍至其他作物以發展出國際作物資訊系統（ICIS）；USDA-ARS的種原資訊網（GRIN）至少儲存450,000種重要經濟作物的分類學資訊及作物表徵等。其中ICIS-IWIS程式是利用SQL編寫程式，透過開放性資料庫連結標準（ODBC）來存取資料，因此可以在多種系統或平台（platform）上作業。

徹底瞭解所有資料之間的同質性與關聯性，是建立高效率資料庫管理系統的重要因素之一。建立一個關聯式農作資料庫之前，必須先將所收集的相關研究資料（訊）加以分類、分析與歸納，然後進行資料模式化，藉由資料模式化決定資料庫的邏輯資料結構（logical data structure）。邏輯係指完成「哪些」(what)工作方面的描述，和建置無關；相對地，實體係指「如何」(how)完成方面的描述，會隨著建置方法之不同而有所改變。RDBMS的邏輯資料結構爲元件實體（entities）、屬性（attributes）及關係（relationship）三者所組成。所謂元件實體就是需要記錄的資訊種類，該種資訊所包含的項目稱爲屬性，不同元件實體之間允許透過關聯屬性來存取檔案內的資料。若以資料的表格形式而言，必須針對不同的元件實體建立其各自的資料表，屬性便決定了表內的欄位種類及型態，各表之間的聯繫則靠關聯屬性來做爲橋樑。現就一個簡單的資料關聯圖，說明關聯性農作資料庫之邏輯資料結構的建立步驟：

㈠農作資料之產生通常來自於試驗研究的成果，因此可以先定義一個元件實體－「試驗」，例如圖一(a)所示之方塊圖，該方塊圖的名稱爲試驗。

㈡然後決定試驗內應該包含哪些屬性項目，例如圖一(b)中的試驗元件的屬性有名稱、假設、目的及描述等。

㈢由於大部份的農業試驗爲因子試驗，試驗的方法及設計也很重要，因此有必要加入另一個元件實體，其名稱定義爲「因子」。「因子」的屬性可能有名稱或描述等。以名稱屬性爲例，其欄位中輸入的資料可能有氮肥、品種等。「試驗」與「因子」兩個元件實體之間彼此的關係，是透過名稱這個關聯屬性來互相連結，一般以方塊圖間的連線來表達，稱爲關聯線，如圖一(c)所示；關係也可以用一個名稱來定義。建立資料連結的關聯屬性，可以一對一、一對多或多對多的關係。如圖一(c)所示，由於一個試驗可能是沒有、僅有單一或含多個處理因子，因此「試驗－因子」關係爲一對多的關聯性，因此連接到「因子」元件實體的關聯線端通常以分叉線來表示，意謂著該元件實體被關聯到的資料情況可能不只一種。

㈣當元件實體及其所含屬性愈來愈多且相互關係變得複雜時，爲加速查詢及其他相關之運算速度，必須在每一個元件實體內選擇其中一個或一個以上的屬性定義爲屬性索引（attribute identifier），如圖一(d)所示之下有橫線標記的屬性。此索引必須是唯一（unique）不可重複的，例如「因子」的屬性－名稱，被設定爲索引，則在因子資料表中的名稱欄位內所輸入的資料絕不可能有重複值。但有時候，資料表內沒有任何單一欄位所儲存的資料是絕對唯一，

此時可以創造出一個新的欄位，武斷地輸入一些編號，或由多個欄位所組成的運算式創造出絕對唯一的組合值，來做爲索引，例如圖一(d)中「試驗」的試驗代號。這個試驗代號的索引碼可以由人爲指定，或由電腦根據編碼原則自動產生。使用任何資料庫管理應用程式，一旦設定了索引欄位，鍵入資料時應用程式會拒絕主索引欄重複值之輸入，以確保資料之唯一性，記錄也會依照索引而排列。關係也可以建立索引（relationship identifier），如圖中連線上的十字記號。由圖中顯示，只要記得各試驗的索引碼就可以快速查詢到試驗的名稱、假設、目的、描述及因子，同時只要從因子的試驗和名稱即可知道有關因子方面的描述。

㈤當然，農作資料庫所包含的元件實體當不止於此，例如，若還考慮到試驗的來源，則可以進而加入更多的元件實體，比如「試驗者」（其屬性爲參與性質，輸入記錄可能爲主持人、研究人員、技術人員、學生或其他）、「人事通訊資料」（其屬性有身份證字號、姓名、所屬單位、聯絡地址、聯絡電話、電子郵件等）等，如圖一(e)所示，而這也不過是農作資料庫之資料模式的一小部份而已。

四、本土性農作資料庫的建置方向

一個完整的農作資料庫之資料模式應該包含哪些內容？倘若融合作物栽培、生理、植物營養、病蟲害、試驗統計、雜草管理及農業微氣象等資訊，一個農作資料模式所包含的元件實體至少可區分爲五大類別：

㈠**試驗**—該類元件實體通常爲農作試驗上比較熟悉的相關事物，比如上述所提的「試驗」、「試驗者」、「人事通訊資料」、「因子」，除此，從「因子」還可以聯結到「因子變級」、「試驗單位」、「試驗設計方法」等。

㈡**測量物件**—這類的元件實體有「物件」（處理、田間管理及測量的對象是何種作物、植物、動物或土地？）、「地點」（試驗執行的地點，但最好含有地理資訊系統的空間資料）等。

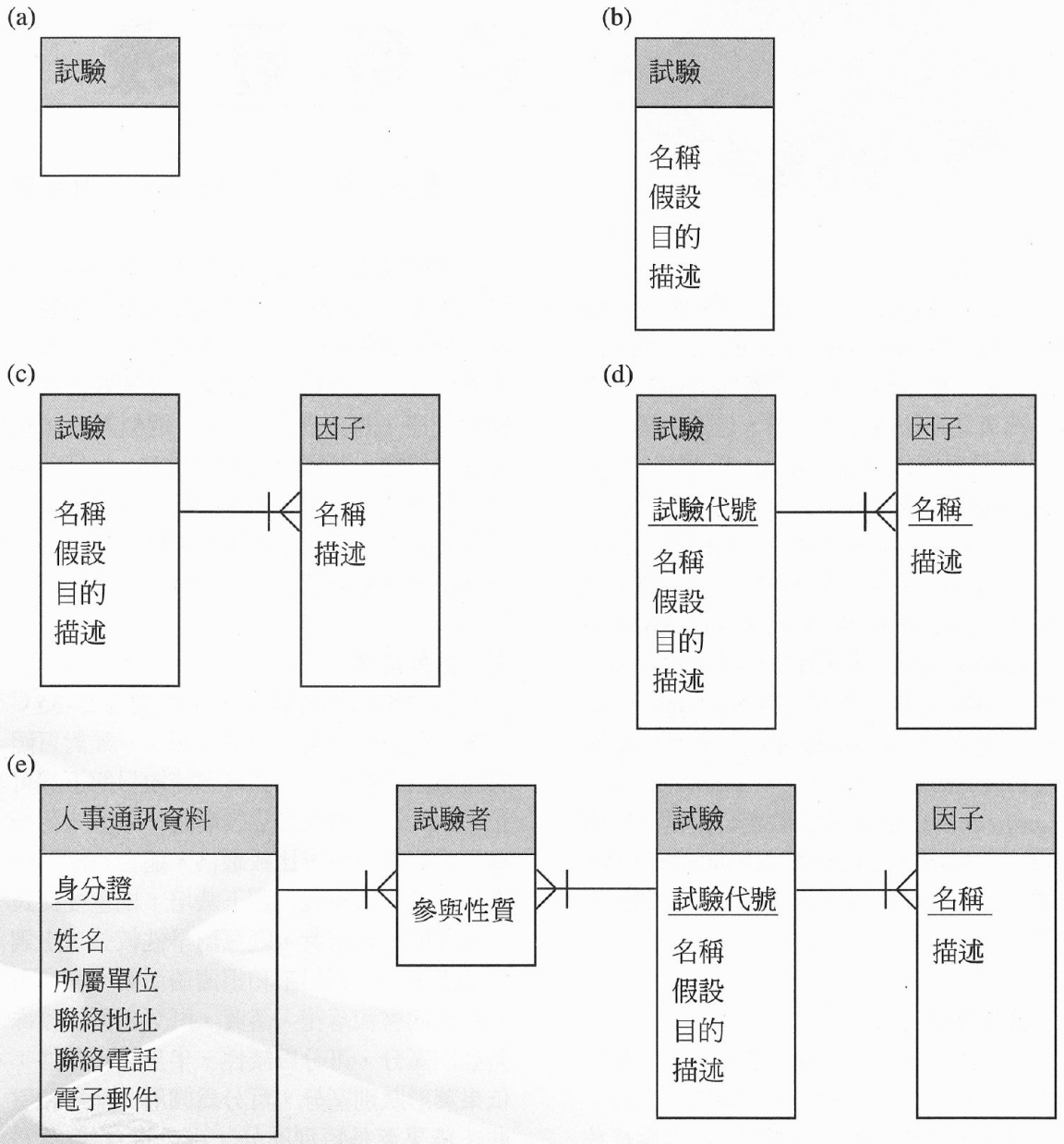
㈢**測量協定與設施**—這類的元件實體主要是在描述如何進行測量，像是「測量技術」、「測量設施」及其他可以確保資料品質的任何測量措施。

㈣**測量**—這類的元件實體是記錄在測量物件上測量些什麼（what）、測量哪裡（where）、何時測量（when）、如何測量（how）、測量多少（how much）之相關資訊。

㈤**田間操作**—有關耕犁、種植、施肥、病蟲害防治、灌溉、收穫等各種資訊。

近年來國內精準農業體系（precision farming system）正積極研發和建置中，精準農業爲一融合農耕知識及多種應用技術組成之農作物經營管理體系，更必須即時且確實掌握資訊，藉由完整資料庫進行模擬及擬定決策，期獲致最佳之收益與環境保育、確保農業之永續經營。而農作資料庫之建置，實爲精準農業作業系統重要的環節。在精準農業體系下，一個良好的本土性農作資料庫除了應包含上述五大類的元件實體和關係群外，有關處理環境變異（如逆境與災害等）之因應管理對策亦應廣泛納入資料庫，同時整合土壤肥力資料庫，配合地理資訊系統、全球衛星定位系統、遙測應用系統、自動化

全球衛星定位系統、遙測應用系統、自動化農機操作系統等之運作，共同發展成決策支援系統，以爲農業制定策略之重要資訊。



圖一、以資料關聯圖說明如何建立農作資料庫的邏輯資料架構：(a)定義元件實體的名稱，(b)決定元件實體內的屬性，(c)加入另一元件實體，並確立其間的關係，(d)建立索引，(e)加入更多的元件實體及關係群(圖中僅是農作資料庫之資料模式的一小部分)。