

資訊的完整性

謝清俊 931211 初稿
931221 一修
931222 二修
931224 三修

決定數位資訊產品價值的主要因素，是該產品承載的內容。這內容可能是各種科學數據，套裝軟體，文學、歷史或社會方面的資料、各行業相關的知識或記錄，數化的藝術品或是休閒遊戲等，種類繁多，難以儘舉。這些內容都不易衡量其客觀的價值，因為它相當依賴使用者想要解決的問題。

然而，除了內容的價值外，資訊產品的一些屬性，如資訊的完整性、正確性、精確度、時效、表現的形式、呈現的數量等，也都與使用者的問題息息相關。這些因素通常合稱為資訊的適用程度，其品質或價值則可以做較為客觀的衡量；其中，資訊的完整性〔**information integrity**〕是較特殊的，也是經常被忽略的。

所謂完整性主要是指資訊的來源是否交代清楚；有交代的完整性高，若完全沒有交代，則此資訊來源不明，資料不完整。來源交代清楚以後，才檢視其內容是否正確和完整無缺。所以，資訊是否完整，不僅僅視其內容，更需兼顧其來源。一則可信賴的資訊，其完整的信息應可追溯到第一手原始資料，或追溯到一個相當可靠的來源。完整性表現的是資訊的可信程度^①，所以它是數位資訊產品最基本、也是最重要的屬性。

學者或大眾傳播業使用資訊是最講求完整性的。若數位資訊產品的完整性沒有交代，許多學者寧可用紙本的資料。若消息的來源不明，或消息無法取得徵信，則負責的媒體寧可不登。有規模的企業使用數位資訊產品也有相似的要求。這樣的要求，固然是敬業、負責任、重視品質的表現，然而，也是因為數位資訊太容易篡改的緣故。

篡改或仿製紙本的資訊並不容易，可是，篡改或複製數位資訊則易如反掌。所以，紙本錄製的資訊(或任何以物質媒介錄製的資訊)，都被視為靜態的資訊；相對的，數位資訊(或是任何以能量形式錄製的資訊)則是動態的。比方說，紙本的書或雜誌，一旦付印出版，就定了形，無法更改。可是，網路上的書或雜誌，即使已出版了，仍可隨時更改——此即其動態的性質。

目前網路上的資訊，或是數位化工程的產出，大都未意識到完整性的重要。所以，有許多數位資訊產品未提供其原始資料的來源，或提供的信息不完整。因此，這些數位資訊產品的價值大幅滑落。例如，這些數位資訊不足以作為徵信的對象，不夠資格作為學術著作或是正式工作報告的參考資料。

許多有識之士，憂心網路資訊的可靠程度。要解決這個問題，唯有建立追溯資訊來源的制度；也就是建立資訊完整性的共同規範和制度。目前，世界上實施完整性最佳的範例，是度量衡的制度。無論是長度、重量、體積、時間、或是任何常用的基本量度，其標準值均可一級一級的回溯到國際標準的原始數據。這樣，便充分建立人們對度量衡的信賴，也完全消弭了對度量衡的爭議。

另一個可為典範的例子，是學術著作的引徵和參照。我們常說：「文章要寫得有憑有據」，學術著作的引徵和參照，正是此有憑有據的具體表現。對數位資訊產品而言，目前尚未建立產品完整性相關的標準和制度，應可比照度量衡的體制，或學術著作的引徵和參照體制。開創這樣的體制，未嘗不是建立數位化企業的新契機。

要建立數位資訊產品完整性的標準和制度，當然不是一蹴可及的。但是，治標的方法並不難：如果能為每份資料建立一組後設資料(metadata)交代其來源，則庶幾乎可以消除完整性的疑慮。如果能夠再針對每份資料建立防止篡改的設施，比如，建立其偵錯改錯碼(error detecting and correcting code)，如此便可提升資訊的可信程度；而這些技術，在國內都已相當成熟。

目前，無論是數位典藏國家型科技計畫，或是 C-Beta 的電子佛典計畫，似乎都忽略了「資訊完整性」這個議題。這當然影響到目前產出的可信賴程度及其價值，也會影響到往後的持續發展和各種可能的加值。這令人惋惜。

註：

- ① 可信程度涉及的問題較複雜，完整性僅為其中之一。可信程度可包含資料的安全性，如遭篡改、破壞、侵入、盜取、或植入病毒等。