

法國行紀（之二）——文獻分享與標準

熱訊 茶餘飯後

1991 年 12 月

謝清俊

看看計算機和通訊工程的發展史，你將發現它們實在就是為了能使「資訊」這個資源達到「物盡其用、貨暢其流」的奮鬥史。用句現代的詞來說，就是在追求一個人類能夠公平合理的「分享資訊」的環境。訂定標準的主要目的之一，也是要能夠分享資源。以上的信念，在法國的會議中感受特別強烈。

這次的會議是 ISO/IEC JTC1/SC18 召開的，討論內容都是和文件（document）和文書（text）相關的標準。這兒所指的文件或文書是相當廣義的，它泛指辦公室內可能接觸到的各種文獻；例如，公文、信件、報告、計劃書、簡報、書本、雜誌、圖表、甚至是包括集聲光之多媒體表達等等。在本文中，且讓我們用文獻一詞概括地表示這些資訊。據此，這裡所談的文獻和傳統的資料處理（data processing）所處理的格式化數據（data）是完全不一樣，它是真正我們日常工作和生活所接觸和用到的資訊形態。

文獻的電子化，可以使計算機能直接幫助我們處理這些文獻，也可能透過電信網路無弗遠近地分享文獻中記載的資訊。這麼做的意義非凡，它將造就一個史無前例的資訊分享的環境，幾乎可讓所有的資訊在網路上暢行無阻。這也難怪美國一些傳播界的人士已經斷言：個人電腦與網路是今後最有威力的傳播媒體，它比電視、廣播更有影響力。我們能想像這樣的世界將變成什麼樣子嗎？

文獻電子化之後，將使電腦處理文獻的威力和電信網路傳送資訊的威力延伸到需要文獻的每一個角落。透過這樣的環境和機制，每個國家在其社會、人文、文化、教育....等等甚為依賴文獻的立國傳統和基礎方面，均將受前所未有的衝擊和發生革命性的變化。如果說：以後的世界，將由電子化的文獻來主

幸；我想，也不會離譜。現在，已經可以感覺到，這樣的社會正一步一步地來到。這些感慨，也是參與這個會議後積鬱在心中的感受。

這一期的茶餘飯後，就讓我來介紹這次開會所談到的標準和文獻分享相關的一些問題吧。

SC18/WG8 標準應用範疇

這次會議中所研擬的標準，都是在 OSI 架構中第七層應用層（application layer）上的標準。其應用領域甚廣，包括：電子出版（electronic network）及各種選擇性的列印系統（information network）及各種選擇性的列印系統（demand printing）等。像這樣宏大廣闊的範疇，要條舉他們涉及的業界都不容易周延。要言之，則出版界、新聞界、各種傳播界、圖書界、電子業界、電腦業界、加值網路及各型資料庫系統（資訊供應者）、管理界、教育與學術界，尤其是以文獻為主的各領域，如文、史、法、商、及社會學門，都會深深地捲入。

如依功能而言，這一群標準將提供：文獻的自由交換和使用的能力，交付文獻給當地的設備處理的能力，交付給 OSI 網路上任何一套設備處理該文獻的能力，以及文獻儲存與檢索的能力。換言之，就是使文獻能完完全全地自由分享；無論是什麼程式、什麼機種、什麼通信設備，都能夠立即且沒有困難地了解，並且使用這些電子化的文書資料。這是要做到不依賴（independent）軟體、硬體和通訊設備的資訊分享！

怎麼才能做到像這樣理想的資訊分享呢？這也正是在擬定各文獻相關的標準要解決的事。

文獻分享的模式

現在，就讓我們談談如何少能做到理想中的資訊分享。坦白說，最早探討這個問題的不是電腦界，也不是電信界，而是傳播界及語言學界。早在 1948 年，Warren Weaver 在他的一篇文章 *Recent Contributions to the mathematical theory of communication* 中就指出：傳播界研究的問題可分為三個層次，即是傳送的技術問題、傳達的語意問題以及傳播的效果問題。傳送的技術問題是說用什麼技術（設備）可將訊號正確地傳輸。語意問題是指如何使接收者能準確地明白欲傳達的信息。而效果的問題是指此傳播行為是否能達到預期的效果。在他的文章裡，精闢地分析了 Claude E. Shannon 的 *Information Theory*（亦稱之為 *The mathematical theory of Communication*，有譯為消息理論者）對傳播界的影響。簡言之他的結論是 *Information Theory* 雖然探討的是在工程上對於資訊（*Information*）量的量測和編碼以及如何正確的傳送的問題，然而在本質上它卻對上述傳播界三個層次的問題，全部都有深入交織著的影響。

工程上的理論（如上所述的 *Information Theory*）對傳播學界的影響至此暫且不再談論。然而，欲建立沒有誤解的資訊分享，卻不能不澈底解決傳送的技術問題和傳達的語意問題。傳送的技術問題偏重於工程，這是計算機和通信工程師的專長，由他們大致可以解決。在 OSI 整個的規劃，都是為此而設，也可說，此問題已可解決。然而在語意問題上，涉及語言學及傳播行為，卻須依賴語言學及傳播行為，卻須依賴語言學及傳播理論之協助，加以仔細斟酌。

要使任何文獻資料，對任何一個電腦系統都能澈底地了解沒有誤會（這樣才能夠自由自在地分享資訊），那麼，唯一的辦法是：用一種大家都認可，且在語法和語意上都嚴格規定，使得不可能產生混淆語意的「語言」，來把文獻資料詳細地描述，然後，將這些對文獻資料的描述以及文獻本身當作一個整體來讓大家分享。

換言之，要達到自由自在的資訊分享有兩個先決條件：其一是大家要同意共用一個對「資訊結構及其性質」描述的人工語言，使在溝通時不致產生語意混淆的情況；其次是將欲分享之資訊和對它的描述一齊透過協議好的方式交換。前者在 OSI 模式中即通稱為 protocol 或資料描述語言，而後者即是其傳送時之多序化（serialization）規則。而在 OSI 之下的各個標準中，此二者都是相輔而成的。（關於此資料分享之模式之詳細說明，可以參考王安資訊 1991 年四月號拙著資訊的分享一文）。

各種文獻的描述語言

在這次會議中討論到的標準，大多是針對某種資料特性所設計的描述語言，諸如：

- SGML 是描述一般文件結構與性質的語言。
- DSSSL 是描述文件處理規格及其語意的語言。
- SMDL 是描述音樂（樂譜）的語言。
- Hytime 是描述超文書（hypertext）和多媒體（multimedia）的語言。

它包括對時間變化的敘述在內，因此，它甚至適用於電影、電視、動畫等結構關係之描述。

- SPDL 是描述排版處理規格的語言。
- DPA 是指揮印刷機（輸出設備）功能的語言。

這些語言純粹是描述性質，不干涉資訊本身的結構。透過他們的描述，各系統就可以免於語意上的混淆，完完全全了解接收到的資料，而達到充份分享資訊的目的。圖一中所呈現的，是一個在這種情況下的處理過程構想。在圖二中所顯示的則是各標準之間的關係以及其主要的應用。

談到這兒，這些標準的威力如何已不必再多言。可嘆的是過去三年之中，沒有中國人參與過這些標準制定的工作。這些已成交的標準草案是不是能適合我們的語言和漢字的應用呢？這著實是令人擔心的問題！現在積極的參與，或可補救一些以往的疏失。如若不然，豈不又變成了中文標準交換碼問題的重演了麼？（本文中用的縮寫名稱，請參閱上一期茶餘飯後之說明，又：法國之行未完，待續）