

【傳播媒體新趨勢】

新聞可以更好： 運用鉅量資料提升台灣的新聞品質

林照真^{*}

時下幾乎每個人對新聞內容都有很多抱怨，包括內容太淺碟、煽色腥與娛樂化。更嚴重的是，每一台播的、每一家報紙登的新聞，幾乎都一樣。之所以如此，是因為目前新聞記者還是按照傳統新聞的路線主義分工跑新聞，甚至哪一個政治人物有助於收視率，就可以大肆報導，完成不顧新聞該有的全面思維。因為這樣，國內新聞的公信力愈來愈低，對未來想從事新聞工作的年輕人，實是很大的衝擊。

身為新聞教育的一份子，我最大的心願就是我的學生，未來可以成為擁有成就感的新聞工作者。成就感得來不易，目前可應用的新聞技巧愈來愈平民化，加上公共新聞、公民記者、社群網站等時代趨勢，受過專業訓練的新聞記者與公民記者的界限愈來愈模糊。這其中的原因很多，較關鍵的因素可能在於記者不明白自己的專業為何，或是專業訓練門檻太低，以致業餘者很容易就可以超越。

好的新聞工作者並不是擔心會被別人趕過去，而是我們真的要思考，在網路已成主要媒體平台，以及大數據時代來臨時，新聞教育該做什麼樣的改變？

這時，資料新聞學的出現，應是數位時代最明顯的新聞挑戰，這已是我開設資料新聞相關課程的第三年。資料新聞學包含兩個階段，第一個階段是蒐集資料、分析資料、從資料中找到新聞故事。第二階段因為資料數量龐大，於是要試著以視覺化方式呈現，才能說明新聞故事的全貌。這兩個部分，與早期的新聞教育都有一定差異。資料新聞學與調查報導中的「電腦輔助報導」（computer-assisted reporting）

^{*} 作者林照真為台灣大學新聞研究所教授，E-mail: carolinelin@ntu.edu.tw。

非常相近，但眼前的整體配套條件已成熟許多。不少第一個階段要蒐集的資料已非機密資料，而是公開、甚至多數是免費的原始資料（raw data）。同時，分析資料使用的軟體有不少也是完全免費，分析軟體等工具也愈來愈多元，這些都與全球資料開放運動的蓬勃息息相關。

進行資料的蒐集與分析時，因為多數是量化資料，新聞工作者首先要克服懼怕數字的心理障礙。過去的新聞教育，多半強調採訪技巧與寫作拍攝，較少接觸數字，這個情形現在會改變，因為數據資料動輒數千筆、數萬筆、數十萬筆，新聞記者必須不厭其煩地檢驗這些數字，並比對其他資料，才可能找到好故事。

想要檢驗這些資料，就必須確定這些資料已做好極佳的數位典藏，並且具有清楚、前後一致的類目，才可以進行試算。台灣在這方面還在急起直追，因為台灣約有 65% 的資料，是以 PDF 的形式出現，若真要運算，只能一筆一筆填上去，台灣的選舉資料便遇上這樣的困境。

去年（2014）九合一選舉，涉及村里長與鄉鎮市民代表等基層選舉，有許多資料到現在，仍只有紙本的形式。為了得到完整的選舉資料，我與台大新聞所同學在一個多禮拜內，天天到中選會把紙本資料印出來。中選會資料不足的，再到政大選舉研究中心找，還缺少的，就打電話到各地方選委會要。到最後因為我們蒐集的資料最完整，才能做出比主流媒體更完整的台灣基層選舉報導（相關作品請見網站：datajournalism.ntu.edu.tw），由此可見資料的重要性。

又，這學期（2015 年）的第一個資料分析主題，我建議同學關注 PM2.5 的問題。PM2.5 並非專屬北京的特殊現象，其實已是國際間空氣品質不佳的普遍問題。相關資料不只國內，舉凡國內環保署、中國大陸、國際衛生組織、綠色和平組織等都有資料可查。另外，因為 PM2.5 的問題與能源有關，也可從美國能源局等單位查尋。各地方城市如台北、倫敦、紐約等市政府的資料平台，同樣有一定的相關資料公開，但是各地的資料整理方法不一。例如倫敦提供的資料最為詳細，新加坡受印尼影響的 PM2.5 霾害問題，則找不到細部資料。這些資料都需要新聞工作者非常耐心地去找，有問題時也要主動與各政府單位聯絡，就有可能掌握相關資料。

得到這些資料後，便需要試算。如果資料整理的類目清楚，新聞工作者本身已具有統計等基本概念後，就可以使用「Excel」、「Google Spreadsheets」等試算軟體，自行進行大量資料的運用。因為「Google Spreadsheets」免費，又可以多人一起編輯，因此多數新聞單位也是以此為試算工具。這個試算工具可以協助加總、比較、過濾資料，但必須確保資料試算一定正確。為了提供使用者檢驗的機會，試算

者也應交代資料來源，才符合資料新聞學的報導倫理。

如果資料量過大，必須試算數萬筆、或數十萬筆資料時，新聞背景的人多半已無法獨自進行資料檢索，就需要結合資工背景的人一起合作。這時，新聞工作者要學習如何與資工人員對話，這也是過去新聞教育缺乏的經驗。我的經驗是，我已多次邀請資工所與新聞所同學一同合作、成為朋友，可以一起聊天、談問題，並一同完成作品。我相信，建立跨領域的合作關係，是確保資料新聞成功非常重要的一環。至於龐大資料的新聞重點為何，依然有賴新聞工作者找新聞的功夫而定，這是新聞工作者在大數據時代的重要角色，他人無法取代。

資料的運用，在於新聞工作者啟動了什麼問題意識？再以 PM2.5 為例，有的會關心城市的比較；有的關心 PM2.5 與呼吸道感染、癌症等關聯性；有的則關心大陸的風向，造成台灣部分城市受 PM2.5 影響等問題；有的分析境內、境外形成 PM2.5 的主因。這些問題意識，多數無人研究。因此，新聞工作者掌握資料後，是唯一有能力從枯燥資料中，找出新聞故事的人。

完成資料試算後，新聞工作者會有所發現，接著是要呈現資料中的故事。這時，當然可以用文字來寫。然而，如果可以選用視覺圖表，就可以表現出資料的情境與相互關係，讀者更可以一目了然。資訊的視覺化，也有另一個專有名詞即「視覺新聞」，意即人的雙眼具有辨認資訊的能力，因此資訊視覺化更講究讀者可以一眼就看到資訊，也就是新聞。目前我將主要的繪圖軟體「Illustrator」，當作視覺新聞的必備學習內容。有了這樣的要求，我發覺同學的作品，在呈現上比以前更純熟了。

有關資料新聞與視覺新聞的應用，已構成今日新聞界因應數據時代的重要工作。想做好這些工作的新聞記者，必須比傳統記者學會資料試算與資訊視覺化兩項工作。即使如此，傳統新聞中跑破鞋的採訪功夫還是非常重要的，文字也一樣重要，但從資料中發展新聞，可以發現跑破鞋也找不到的新聞，一定是無可取代的。

記得曾在網站中，看過這麼一段故事。發明網際網路的英籍科學家 Tim Berners-Lee 被稱為網路之父，他同時也協助英國政府公開資料等工作。曾有記者問他：「政府花那麼多錢，整理那些資料是要給誰用呢？」Berners-Lee 回答：「是給新聞界用的。我們需要具備資訊技能的記者，來幫我們從資料堆中找故事。」我真是同意他說的話，才知道世界很大，卻是德不孤。

雖然數位時代，新聞記者已了解大數據可以作為新聞來源，但因為資料品質參差不齊，有時未必可以找到完整的故事。然而，隨著這波全球趨勢，各種資料正進

行跨國界的流通，所以新聞工作者更需要強化自己找資料的能力。此外，分析的軟體工具愈來愈普遍，在開放原始碼的共同價值下，毫無疑問，資料新聞一定會愈來愈見發展。

很慶幸我們恭逢大數據時代，值此時刻，新聞界若能善用大數據，一定可以擴大新聞來源，使新聞力量更強大。這並不是說我們要放棄傳統的、其他的找新聞的方式，而是資料新聞將會是當代新聞記者監督政府、認識全球、把新聞故事講得更精確的新穎方式。

如今新聞產業問題百出，很多記者只能進行淺碟式新聞，更覺得自己沒機會做一個好記者，心中非常挫折。坦白說，資料新聞無法讓一個壞記者變成好記者，卻可以讓一個好記者，變成更好的記者。新聞記者若能學會從資料中找新聞的技能，一定有助於挽救新聞、提升台灣的新聞品質，讓台灣新聞變得更好。