



雲報專欄：大數據？當然不是！

—工研院巨量資訊科技中心主任/余孝先/技術專家委員會副主任委員、工研院行銷傳播處/李荼

Big Data 究竟該如何翻譯與詮釋才最到味，請參考本文中「必也正名乎」一節中加以分析。另外，如何掌握巨量資料的發展趨勢，以人才開創掌握台灣的「巨量」機會，則是本文主要剖析的重點。巨量資料以應用為主，看看日韓手機的應用，創造符合市民與觀光客的需求、提升人民對施政及商旅人士的滿意度。參考產業發展之趨勢與研究單位的預估資料，本文將歸納點出「挖寶」巨量資料應用需要什麼樣的人才，提供產學研等專家們參考。

必也正名乎

Big Data 在大陸上一直都譯為大數據，台灣是譯為巨量資料或大資料。到底 Data 一詞應該譯為資料或者是數據呢？教育部頒訂的國語字典上的解釋如下，資料：可供參考或研究的材料或記錄。數據：經由調查或實驗而得的數量化資料。所以，數據是資料中用數字表示的部分，不能包括文字資料。我們會說「準備會議資料」，不會說「準備會議數據」。早期資料處理的內容大多是數字，所以 Database 在台灣譯成資料庫是正確的，大陸上譯成數據庫也勉強辦得過去。不過這一波 Big data 的內容明顯加入了大量的文字資料或其他多媒體資料，不再以數字為主，再勉強配合大陸譯成數據就是中文沒讀通了。

掌握巨量資料發展趨勢，以人才創造台灣的「巨量」機會

在資料經濟時代，不僅資料量迅速成長，資料型態也變得更多元，再加上資料處理和分析技術的進步突破，使得過去很多無法利用的資料也變得有用，這些都是近年來巨量資料受到矚目的重要原因。如日本以手機漫遊資料取代過去的高成本低效率的問卷調查，分析出外國觀光客的旅遊偏好，並藉此規劃出觀光巴士的路線；韓國首爾市也同樣運用手機的發話資料來規劃夜間公車路線，以更符合市民需要、提高搭乘率、提升滿意度。

有人用「挖寶」來形容巨量資料的應用，資料愈經過挖掘、運用後，反而愈





能夠增加價值而且可重複的再利用。透過巨量科技快速發展，我們得以收集更多樣的資料、做更深入的分析、找出更多以前未發覺且具有意義的資訊、甚至導出與經驗直覺相悖的結論，並依此做出更快速與更準確的判斷與決策。根據 IDC 的預估，全球巨量資料應用已在各行各業加速發酵，巨量資料的市場預期將有驚人的成長，市場規模至 2016 年將成長至 507 億美元，年複合成長率達 47%，顯示巨量資料的市場商機無限。

巨量資料未來發展潛力無限，但要能夠從眾多資料中挖出寶、讓資料真正地加值，需要有相對應的人才來推動落實。從巨量資料的核心技術所需人才來看，可分為「資料處理技術」與「資料分析技術」兩大類。「資料處理技術」部分比較相關的技術是高速計算、平行計算與資料庫技術等，比較適合具有軟體方面專業的人才；而從推動台灣企業與政府廣泛使用巨量資料應用提升競爭力來看，更缺乏的人才會是在「資料分析技術」方面，包括所謂的資料分析師與資料科學家 (Data Scientist)。資料分析工作除了需要一般資訊工程與統計學的訓練外，「機器學習」(Machine Learning) 更是核心的技術，也就是如何讓機器或電腦能夠學會像人一般的自我學習成長，提升智慧與判斷能力。由於可以運用資料分析的產業包羅萬象，不僅需要同時懂得資訊與統計等專業領域的人才，這些人才也需要了解資料分析的特定產業，但這類的通才幾乎不存在，因此就需要透過團隊來達成，於是也就需要溝通協調與團隊合作的能力，才能成為好的資料科學家團隊。

美國求職網站 Indeed 的資料顯示，目前資料分析人才的需求已較五年前增加了近五倍之多，指名需要機器學習專長的職缺也較五年前成長超過 100%；最近國內 1111 人力銀行所公布的資料分析相關職缺，亦較四年前成長了三倍。由此看來，隨著巨量資料產業的發展，巨量資料的相關人才需求快速提升，已成了發展巨量資料產業的首要關鍵。台灣向來注重數理教育，高等教育體系過去二、三十年已培養出許多圖形識別 (Pattern Recognition) 與訊號處理的人才，能夠對應到資料分析人才所需的機器學習等專長和知識，若能適當加以導引，應能快速投入資料分析領域。最近東吳大學已獲得教育部同意，籌備成立巨量資料學院，就是一個很有眼光的重要行動。簡言之，在全球巨量資料產業加速成長的趨勢下，台灣應該積極運用人才策略，為產業創造新一波的競爭優勢。

