



莊庭瑞

中央研究院資訊科學研究所副研究員，
合聘於資訊科技創新研究中心以及
人文社會科學研究中心。

連結知識的持續識別碼

全面數位化的出版品搭配各種識別碼系統，開啟了知識探索的新路。

撰文／莊庭瑞

現今的學術出版品已經全面數位化了。雖然有些期刊還以紙本發行，但同時一定會有電子版；期刊論文就是電子檔，可在期刊網站閱覽或下載使用。

對於論文的引用，除了沿循傳統方式詳列作者、篇名、期刊名、刊出日期等，眾人也逐漸以「數位物件識別碼」（DOI）來指稱並取用論文。例如 doi: 10.1000/182 代表 *DOI Handbook* 這份出版品；如果讀者不知道這個數位識別碼所代表的物件，可以透過 DOI 網站查詢，或是直接造訪網頁 <https://doi.org/10.1000/182>，經由 DOI 的解析服務，取得 *DOI Handbook* 的資訊。類似的識別碼還有「國際標準期刊號」（ISSN），例如 ISSN 1682-2811 是台灣出版的一份月刊《科學人》雜誌，在封面的條碼上就可看到這個資訊。

論文的作者也有識別碼，可用來區分同名但不同人的學者。研究人員可申請「開放型研究者與貢獻者識別碼」（ORCID），把所發表的論文歸類在自己的 ORCID。研究機構也有全球的識別碼系統，常用的有「研究機構登錄」（ROR）以及「全球研究識別碼資料庫」（GRID）兩家。除了 DOI 以及 ISSN，期刊論文的「後設資料」（metadata）也開始記錄作者的 ORCID 以及其任職機構的識別碼。大量論文引用資訊在編碼後，論文的内容結構與彼此的連結更為精確，可用來了解研究人員之間的合作網絡、分析論文相互引用的群聚情形，或是探索研究機構在期刊發表管道上的偏好取向。

圖書館用來編目的「標題詞」（subject heading）也是一種識別碼系統，可標記專書及論文内容的主題。不同圖書館的標題詞系統或許不同，但只要能相互參照，就可跨館檢索書目。標題詞指涉範圍的上下位關係，通常會在系統中註明，方便進行廣義或狹義的搜尋。

這類識別碼有時稱為唯一識別碼，也常稱為「持續識別碼」（persistent identifier）。也就是個別識別碼與所指涉的概念或個體，保持唯一以及持續的關係，不會任

意變動。持續識別碼系統的維護機構非常多樣，有非營利民間組織例如 ORCID、國際標準組織（例如 ISO 3166「國家與地區代碼表」）或是商業公司，也有完全由使用社群自主維護的，最知名的應該就是「維基資料」這項由維基媒體基金會所支持的協作專案了。

我們從較熟悉的維基百科網站上，就能連結到維基資料。例如在「鄭成功」頁面點選左方「工具」選項的「維基數據項目」，就能得到鄭成功的維基資料頁面，其中 Q197615 就是它在維基資料的識別碼。而「蘭嶼」在維基資料的識別碼是 Q701647，透過編碼後的事實敘述，我們知道它是座島嶼，面積 45 平方公里，英文名為 Orchid Island，屬於台東縣（識別碼 Q249904）。

維基資料的各條目也連結到其他的識別碼系統（例如蘭嶼在美國國會圖書館的標題詞是 sh85074273），以及其在維基百科各種語言版本的詞條網址。人名與地名透過標準的編碼與關聯註記，讓我們知道鄭成功、國姓爺和 Koxinga 指稱的是同一位歷史人物，以及地理上的台灣包括澎湖群島和蘭嶼。對知識概念與個體持續進行編碼，並且維護不同編碼系統之間的對應，再透過不同來源但編碼過後的大量事實資訊，也就能以自動化方式來探索知識，這對眾人而言會是非常便利的事。

維基資料上約有三成是學術論文條目。這些論文條目都有 DOI，論文作者也連結到維基資料的人物條目以及 ORCID，論文主題也有編碼（例如 Q84263196 用來標記主題為 COVID-19 的論文）。丹麥技術大學的尼爾森（Finn Årup Nielsen）等人發起 Scholia 專案，以維基資料的條目為基礎，視覺化呈現研究主題、研究人員和研究論文的關聯。這項工作相當受到注目。

使用公眾維護的大量事實資訊，以持續識別碼進行系統性連結，並發展成眾人可自由使用的資訊工具與服務，這對於人類知識的共同生產與使用，再再指向令人嚮往的前景。SA