

# **From Government Gazettes to Personnel Changes Databank: Digital Data Production for Retrieval and Analysis**

Jyi-Shane Liu\*

## **Abstract**

This paper reports the databank development of government personnel changes based on content of government gazettes. The work serves a dual purpose of digital archiving historical documents and producing a subject databank for encoding entity meta-information in the textual content. The set of documents are Taiwan government personnel order published on government gazettes and are dated from the year of 1896 to present day. Digital images are created by scanning the printed text documents. Meta information of appointing and dismissing government officials are further extracted and encoded in a subject databank. The product is a Taiwan government personnel changes databank that contains up to one million entity instances over a time period of more than one hundred years.

The essential meaning of the databank is to record personnel change information in a structured data set and, therefore, enable the provision of higher granularity data for more effective information use. In other words, the databank provides fine-grained information that can be integrated and analyzed as needed to help reveal unknown aspects and trends of subject matters. Such a data analysis utility is not possible with full-text document digitization and keyword-matched retrieval. The construction of the databank opens up tremendous opportunities for empirical and validation research that were previously hindered or limited by extreme difficulty of data access to thousands of printed materials. Yet another contribution of the databank is to build a comprehensive personnel history for the Taiwan government who has not attempted any retrospective data input work after implementing computerized administrative operations during the 1980s. This databank construction work demonstrates the potential of library as data producer in exploiting primary textual materials, especially in the area of subject-based document collections, and developing valuable online information resources for public interest.

**Keywords: databank development, textual data extraction,  
government personnel information, data retrieval and analysis**

---

\* Director, University Libraries & Professor, Department of Computer Science, National Chengchi University

# 從政府公報到人事異動資料庫 ——數位資料產出及檢索分析

劉吉軒\*

## 摘要

本研究提出以政府公報內容為基礎之官員人事異動資料庫發展，並討論其雙重目的：(1) 對歷史文件的數位化；(2) 摘錄文件內容的重要資訊而製作成主題資料庫。研究的文件對象為《中華民國政府公報》及《台灣總督府府報》中刊載之官員任免命令，時間範圍可回溯至 1896 年。這些刊載官員任免命令之紙本公報以掃描取得影像檔，再針對其中官員的任命及免職資訊，以後設資料摘錄蒐集，而成為官員異動之主題資料庫。此台灣政府官員異動資料庫將涵蓋超過一百年期間之近百萬筆官員異動資料。

此資料庫之重要意涵為以結構化方式，記錄人事異動資訊，而能提供高度精細的資料檢索及更有效的資訊取用。這些從大量文件中摘取出的精細資訊能被依需求檢索分析，而有助於發掘未知的主題面向與趨勢。這種資料分析能力是文件全文數位化與關鍵字比對檢索無法提供的。過去大量紙本文件中的資料難以取得使用，而限制了許多人文社會學科的實證研究。主題資料庫的發展將解除資料取得的限制，進而開啓許多新的研究機會。此主題資料庫的另一貢獻是完整涵蓋記錄中華民國與台灣地區政府的官員異動資訊。此資料庫的建置工作展示了圖書館扮演資料產製者角色的可能性，尤其是在掌握主題性的文件資料館藏上，進而發展出有價值及有公共利益的數位資訊資源。

**關鍵字：**資料庫發展、文字資料摘取、政府人事異動資訊、資料檢索分析

---

\* 政治大學資訊科學系教授兼圖書館館長。