國立臺灣大學文學院圖書資訊學系暨研究所

碩士論文

Department of Library and Information Science

College of Liberal Arts

National Taiwan University

Master Thesis

臺灣公民科學計畫資料徵集與品質維護：策略與實踐

Data Quality in Taiwan's Citizen Science Projects:

An Investigation of Current Practices in Data Acquisition

and Maintaining Strategies

甘婕人

Chieh-Jen Kan

指導教授：鄭瑋 博士

Advisor: Wei Jeng, Ph.D.

中華民國 112 年 1 月

Janurary 2023

摘要

科學研究在資料密集典範影響下產生大量研究資料，進一步促進了開放科學的發展。開放科學重視研究資料管理、研究通透性、資料的再利用並鼓勵公眾參與。作為開放科學最佳實踐之一的公民科學（citizen science），邀請公民科學家參與不同科學研究階段，共創知識並獲得研究成果。

然而由非研究人員所產生之資料，其資料品質是長期學術界關注的重點。臺灣公民科學計畫已發展超過 20 年，卻鮮少有針對公民科學計畫資料徵集、公開分享及資料品質維護策略之探討研究。為填補此一缺口，本研究旨在初探國內公民科學計畫發展現況以及執行計畫之研究團隊維護資料品質策略。

本研究為實徵研究，利用內容分析法檢視 127 件臺灣公民科學計畫網頁，透過四大面向 1) 公民科學計畫基本資料、2）計畫研究團隊徵集資料方式、3) 公民科學家紀錄資料類型及 4) 計畫資料分享與開放結果作為編碼框架，並細分為 13 個屬性 78 項變數觀察分析，後輔以質化研究方法，深度訪談國內具代表性之公民科學計畫執行研究團隊－路殺社，蒐集該計畫團隊實際規劃資料徵集工作流程及所面臨之資料品質難題與維護策略。

本研究發現臺灣公民科學計畫有著長期持續執行且負有監督與輔助政策之特色，計畫研究團隊主要使用 Web 應用程式或社群媒體平台 Facebook 作為資料徵集工具，在意資料品質的正確性（accuracy）、完整性（completeness）與效度（validity）。然而目前存在資料偏誤、資料重複、資料辨識困難及資料蒐集規範未被遵守等難題。研究團隊透過改善資料蒐集方式，如利用資訊科技工具自動檢核欄位、面對面指導或透過資深公民科學家協助新手等方式提升資料品質。本研究另以 FAIR 原則檢視國內公民科學計畫，超過 70% 的計畫符合可被查找（findable）指標，然多為有限制的公開研究資料，且以視覺化研究成果為主。

本研究透過了解國內公民科學計畫之資料徵集工作係為影響資料品質之關鍵，揭示各項徵集工具之優缺點及合適使用情境，並綜整國內公民科學計畫研究資料之分享與再利用情形。最後給予未來有興趣以公民科學作為研究途徑之單位或團體其維護資料品質的策略建議。

關鍵字：臺灣公民科學計畫、資料徵集、資料品質、臺灣路死動物觀察網

Abstract

Scientific research produces a substantial amount of research data in the data-intensive-paradigm, which further promotes the development of open science. Open science emphasizes research data management, research transparency, data reuse, while also encouraging public participation. Citizen science, a key practice of open science, invites citizen scientists to take part in different stages of scientific research, co-create knowledge, and access research results.

Citizen science projects in Taiwan has been developed for over two decades, , but there has been minimal research on strategies for acquiring, sharing, and maintaining data quality for these projects. To fill this gap, this study seeks to investigate the current state of citizen science projects in Taiwan, as well as the data quality maintenance strategies used by the research teams to implement these projects.

In this study, 127 Taiwan citizen science project web pages were analyzed using content analysis. Four dimensions were identified , including 1) basic information of citizen science projects, 2) data acquisition methods of project research teams, 3) types of data recorded by citizen scientists, and 4) project data sharing and open results, resulting in a total of 13 attributes and 78 variables. In addition, qualitative research methods were employed to conduct in-depth interviews with the Roadkill Observation Network, a representative citizen science project execution team in Taiwan. The interviews collected information on the project's data acquisition workflow, data quality issues, and maintenance strategies.

This study discovers that citizen science projects in Taiwan exhibit long-term and continuous implementation with well-defined supervision and assistance policies. The project research teams mainly use web applications or the social media platform Facebook as data acquisition tools, with a focus on the accuracy, completeness, and validity of the data. However, there are currently challenges such as data bias, data duplication, difficulty in data recognition, and non-compliance with data acquisition regulations. To address these challenges, the research team employed IT tools to automatically verify fields, offered face-to-face guidance, and enlisted experienced

citizen scientists to mentor newcomers. This study also examined domestic citizen science projects using the FAIR principles, and found that over 70% of the projects meet the "findable" indicators, but most of them are limited research data that focus on visualized research results.
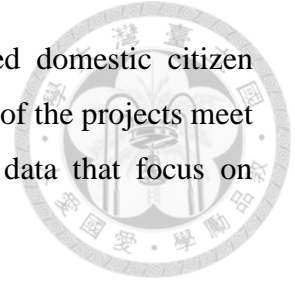
This study revealed that the data acquisition process is critical in affecting data quality for citizen science projects in Taiwan. It also identified the advantages and disadvantages of various collection tools and appropriate usage scenarios. Finally, the study offered strategic recommendations for maintaining data quality to future units or groups interested in using citizen science as a research method.

Keywords: citizen science project, data acquisition, data quality, Taiwan Roadkill Observation Network