淡江大學資訊與圖書館學系碩士班

碩士論文

指導教授：陳亞寧 博士

以書目框架為基礎的鏈結資料
編目作業
The BIBFRAME Based Workflow for
Linked Data Cataloging

研究生：王薇涵 撰

中華民國 109 年 6 月

論文名稱：以書目框架為基礎的鏈結資料編目作業　　　頁數：153

校系(所)組別：淡江大學　　資訊與圖書館學系碩士班

畢業時間及提要別：　108 學年度第 2 學期　　　　　　　碩士學位論文提要

研究生：王薇涵　　　　　　指導教授：陳亞寧 博士

論文提要內容：

　　隨著語意網的興起與應用，近年來圖書資訊界已有許多個案嘗試導入鏈結資料（Linked Data, LD）技術，如何將既有書目資料 LD 化則是既新且複雜的議題。現有個案係依每一圖書館的環境需求與條件，而有不同的 LD 轉換策略，尚未有一致的標準規範或流程。截至目前為止，未有任何研究深入探討圖書資訊界有關 LD 的轉變方式及其作業流程。因此本研究主要目的是以八筆「書目紀錄」為研究樣本並採用使用個案法，探討 LD 的不同資訊組織作業流程，以提出適合圖資界的 LD 相關流程與作業，亦提供相關建議及館員知能，減少日後作業上的困難。

　　研究結果顯示將書目紀錄 LD 化的作業流程可分為有兩大類：

（一）既有書目紀錄轉變成 LD 有兩種方法，第一種是透過「MarcEdit—BIBFRAME TestBed」，作業流程分為「識別、清理資料、豐富、確認、命名、建模（選擇既有的知識本體）、重複使用詞彙及轉換」八個步驟，第二種則是「MarcEdit 及 OpenRefine 並用」，作業流程稍為複雜，共有 14 個步驟，分別是「識別、清理資料、豐富、確認、轉換檔案格式、識別、清理資料、調和、確認、URI 予以獨立、命名、建模（選擇既有的知識本體）、重複使用詞彙及轉換」。兩種作業流程皆符合 LD 的標準，若欲快速且由工具自動將記錄轉變為 LD，可參考第一種作業流程；若由人工進行辨識及對照使用詞彙，以轉變紀錄為 LD 形式，則可參考第二種作業流程，雖然作業流程較繁瑣且耗時，但彈性相對較高。

（二）在 LD 環境下，以 BIBFRAME 編輯平臺為 LD 編目，可分為原始和抄錄編目兩種，前者的作業流程包括「建模（選擇既有的知識本體）、識別、命名、重複使用詞彙、調和與豐富、加入書目關係與轉換及視覺化」七個步驟；後者則是「建模（選擇既有的知識本體）、識別、清理資料、重複使用詞彙、調和/豐富、加入書目關係與轉換及視覺化」七個步驟。

此外，在相關注意事項及館員知能方面，編目館員需注意的內容包括熟悉 MARC21 和 BIBFRAME 及其相關的編目規則與檔案格式、確認加入資料中 URIs 的正確性、熟悉各工具之使用方式與了解各種 RDF 序列化格式等。

根據研究結果提出之建議包括：編目館員應注意編目資料著錄的正確性、後續研究可探討加入發布後之完整編目作業流程、圖書館機構應制定書目紀錄 LD 化的作業流程、加強編目館員的教育訓練、建置在 LD 環境下的中文編目平臺及圖書館應建置完整的權威 LD 資料庫。

關鍵字：鏈結資料、書目框架、鏈結資料作業流程、書目知識本體

Title of Thesis：　　　　The BIBFRAME Based Workflow for　　　　Total pages:153
　　　　　　　　　　　　Linked Data Cataloging

Key word:　　　　　　　Linked Data (LD), Bibliographic
　　　　　　　　　　　　Framework (BIBFRAME), LD workflow,
　　　　　　　　　　　　bibliographic ontology

Name of Institute:　　　Graduate Institute of Information and
　　　　　　　　　　　　Library Science, Tamkang University

Graduate date:　　　　　June,2020　　　　　　　　Degree conferred: M.A.
Double space

Name of student:　　　　Wei-Han Wang　　　　　　Advisor: Dr. Ya-Ning Chen
Double space　　　　　　王薇涵　　　　　　　　　　陳亞寧 博士

Abstract:

　　　　With the growth of the Semantic Web and its applications, linked data (LD) applications have emerged in library and information science (LIS), but the transformation of legacy bibliographic data into LD is a new and complex problem. Different LD transformation strategies have been used according to the needs and conditions of each library, and no consistent standard or workflow has been established. Currently, no study has explored LD transformation strategies or workflows in LIS. Therefore, this study mainly used eight bibliographic records as research subjects and employed "use cases" as methodology to investigate information organization workflows for LD that are suitable for LIS and provide suggestions for cataloging and librarian's competency to reduce the difficulty of future operations.

　　　　The results revealed that the transformation of bibliographic records for LD can be divided into two categories.

(1). Legacy bibliographic records can be transformed into LD by using two approaches. The first one is through MarcEdit-BIBFRAME TestBed. The operation process has eight steps as follow: identification, data cleaning, enrichment, verification, naming, modeling (existing ontology selection), vocabularies reuse, and conversion. The second one is the combined MarcEdit and OpenRefine, which has a slightly more complicated workflow comprising 14 steps: identification, data cleaning, enrichment, verification, file format

conversion, identification, data cleaning, reconciliation, verification, independent uniform resource identifier (URI), naming, modeling (existing ontology selection), vocabularies reuse, and conversion.

Both workflows has proved feasible to be compliant with LD standards. The former can be used to rapidly and automatically transform records into LD, whereas the latter can be used to manually identify and map vocabulary to transform records. Although the latter workflow is cumbersome and time intensive, it is relatively flexible.

(2). Cataloging with BIBFRAME Editor (BFE) in the LD environment can be used for LD original and copy cataloging. The workflow for original cataloging consist of seven steps: modeling (existing ontology selection), identification, naming, vocabularies reuse, reconciliation and enrichment, the addition of bibliographic relationships, and conversion and visualization. Copy cataloging is comprised by the seven steps of modeling (existing ontology selection), identification, data cleaning, vocabularies reuse, reconciliation and enrichment, the addition of bibliographic relationships, and conversion and visualization.

In terms of competency, cataloging librarians must understand MARC21, BIBFRAME, and their related cataloging rules and file formats; confirm the accuracy of added URIs; and understand various tools and Resource Description Framework (RDF) serialization formats.

Suggestions based on the findings of this study are as follows:

(1). Cataloging librarians should focus on the accuracy of the descriptions of the cataloged material.

(2). Follow-up research can explore complete cataloging workflows with inclusion of LD publishing.

(3). Libraries should formulate the workflow for transforming bibliographic records into LD.

(4). Education for LD cataloging librarians should be provided.

(5). A Chinese cataloging platform should be established to fit for the LD environment.

(6). A complete and authoritative platform with URIs for individual Chinese LD is required.