

基于 BIBFRAME 模型的书目资源关联数据化^{*}

贾君枝 崔西燕

摘要 BIBFRAME 可以实现书目数据的关联数据化以及图书馆数据与外部资源的链接,为智慧图书馆建设提供书目的数据化支持。在 BIBFRAME 模型特点及词表重用方式的基础上,利用采集到的国家图书馆书目数据,着重探讨了基于 BIBFRAME 中文书目关联数据化的映射设计过程,并归纳相应的映射处理方式;同时利用拟定义 URI,分析书目数据内外部实体资源链接方式,以完善关联数据发布过程。通过此研究路径,提出了书目数据映射过程中的相关问题,构建了书目数据实体资源关联模型,为实现中文书目数据 CNMARC 语义关联数据化提供借鉴。图 5。表 1。参考文献 23。

关键词 书目数据 BIBFRAME 模型 映射 实体关联模型 关联数据

Bibliographic Resources Linked Datafication Based on BIBFRAME Model

Jia Junzhi Cui Xiyan

Abstract: BIBFRAME can realize the association and digitization of bibliographic data and the link between library data and external resources, and provide bibliographic datafication support for the construction of the smart library. Based on the characteristics of BIBFRAME model and the way of thesaurus reuse, using the collected bibliographic data of the National Library of China, this paper focuses on the mapping design process of Chinese bibliographic linked data based on BIBFRAME, and summarizes the corresponding mapping processing methods. At the same time, the proposed URI is used to analyze the link mode of internal and external entity resources of bibliographic data, so as to improve the publishing process of linked data. Through this research path, this paper puts forward the related problems in the process of bibliographic data mapping, and constructs the entity resource link model of bibliographic data, which provides a reference for the realization of Chinese bibliographic data CNMARC semantic link data. 5 figs. 1 tab. 23 refs.

Keywords: Bibliographic Data; BIBFRAME Model; Mapping; Entity Link Model; Linked Data

智慧图书馆的发展建立在数据化馆藏的基础上,而以文献处理为核心的 MARC 数据不具有细粒度、语义性、开放性,影响了馆藏资源由数字化向数据化的转化实现。关联数据技术是一种在数据级别支持分布式 Web、更容易为机器所理解的技术,可作为图书馆之间、图书馆与 Web 之间资源共享的沟通桥梁,旨在通过丰富的链接实现信息共享。

书目资源数据化是将 MARC 格式的书目记录中有意義的知识元转换为机器可识别、理解、计算和分析的结构化数据的过程^[1]。关联数据技术是实现书目资源数据化的有效手段,有助于书

目数据网络的形成。美国国会图书馆于 2011 年启动了将书目描述规则设计为关联数据的书目框架计划 BIBFRAME,为书目数据关联数据化提供了一种国际通用的层次化模型及语义丰富的词汇表,可确保书目信息以数据化的语义三元组形式链入网络社区,在图书馆领域内外得到更广泛的应用。目前基于 BIBFRAME 2.0 开展书目数据化项目的相关机构或平台共有 9 个,包括伊利诺伊大学分校图书馆、Ex Libris - Alma、Reasonable-Graph、@ CULT、匈牙利国家博物馆图书馆、科罗拉多学院、美国陆军工程兵研究开发中心图书馆、斯坦福大学、德国国家图书馆^[2]等,充分验证了

^{*} 本文系国家自然科学基金项目“数据开放环境中的词表重用问题研究”(项目编号:19BTQ023)、中国人民大学科学研究基金项目“数据驱动下的知识组织发展研究”(项目编号:202230018)的研究成果之一。

基于 BIBFRAME 2.0 实现书目数据关联开放的可行性。

国内外针对 BIBFRAME 书目数据化实践项目进行了研究,如 Tharani 等^[3]以伊斯迈利穆斯林(Ismaili Muslim)社区的藏品进行案例研究,认为 BIBFRAME 是图书馆在网上获取和共享书目元数据的可行方法,并介绍了图书馆实现 BIBFRAME 的具体的和可重复的步骤。Jin^[4]将 7829 本 19 世纪英国小说的数字藏品的元数据,通过数据建模、转换和链接等过程从 MARCXML 转换为 BIBFRAME 2.0,指出 BIBFRAME 2.0 模型可以增强图书馆社区内外数字馆藏的可发现性。李勇文^[5]从理论层面对 CNMARC 模型和格式进行剖析,从格式、模型、互操作、技术环境等方面探讨了 BIBFRAME 在中文书目数据中应用的实践范式。夏翠娟等^[6]以上海市文献联合编目中心的中文书目数据为例,基于 BIBFRAME 初步探索了中文关联书目数据发布的数据建模、本体词表设计、CNMARC 与关联书目数据模型的映射设计、中文关联书目数据发布的内容整合方案和技术实现方案。许磊^[7]在梳理元数据映射、元数据元素与知识本体之间的关系基础上,对 CNMARC 字段与 BIBFRAME 本体词表的语义映射关系进行了总结。周小萍^[8]建立了 CNMARC 相关字段、子字段与 BIBFRAME 2.0 词汇的映射,并选取 CNMARC 实体记录进行了 BIBFRAME 模型的转化。国内外学者从不同角度对 MARC 到 BIBFRAME 的转换过程及步骤进行了较为丰富的研究,但现有研究在理论方面对 BIBFRAME 重用情况及映射处理方式等分析不足,实践层面也未进行映射质量检验。

鉴于 MARC 格式当前的应用普遍性,国内对于书目数据关联数据化的转换和开发存在实践、成本、技术等诸多方面的考虑,对图书馆关联数据的实现尚处于构思探索阶段。基于前人研究,本文较为全面地分析了语义网环境下 BIBFRAME 模型的特征及重用方式,以 BIBFRAME 模型为核心分析并归纳 MARC 到 BIBFRAME 的映射处理方式及书目实体内外部语义关联等,通过映射检

验发现存在的问题,以弥补以往研究中存在的不足,为快速推动我国图书馆书目关联数据化提供参考路径。

1 BIBFRAME 模型特点及词表重用方式

1.1 BIBFRAME 模型特点

针对开放的语义网环境设计的 BIBFRAME,提供了图书馆书目资源实现数据化的描述模型和本体框架,是图书馆书目资源发布为关联数据的最佳实践模型。当前以 MARC 格式为主的书目数据,描述了便于用户通过检索系统识别和获取书目的所有著录项目。但其仅能实现简单的检索功能,并具有高度结构化特点,这难以满足开放互联的网络环境。通过与 MARC 的比较,BIBFRAME 模型超越性主要体现在以下三个方面。

(1) 基于实体的层次化结构检索。MARC 作为描述性元数据,每条记录都结构化地描述了文档的书目信息,混合了作品、实例、单件的属性,呈现扁平化的单一线性结构,无法揭示书目各个字段的层次关系。而 BIBFRAME 2.0 围绕书目资源建立了作品、实例、单件三个层次结构模型,揭示了文献资源的实体要素及关系,有利于实现书目数据分层导航浏览与检索。如基于 BIBFRAME 模型的 Share-VDE 项目配备了 3 个分层导航工具,以提供人/作品、出版物(实例)、单件层面发现相关资源^[9],其利用关联数据实现了基于实体对象检索的语义链接并呈现图书馆的书目信息,提高了图书馆资源的可见性。

(2) 细粒度语义标签。MARC 格式以记录为单位,每条记录对应一个文档,以揭示资源的外在特征为主,缺乏对书目数据内各个知识单元及关系的细粒度揭示,每条记录采用图书馆专用的包括字段、指示符、子字段等的标识系统来表示书目数据,字段描述含义有重复,仅供机器可读,语义不明确,对于非专业用户并不友好。而 BIBFRAME 利用万维网联盟的 RDF 三元组数据模型进行数据编码,以细粒度的语义数据为单位,提供了明确

定义的语义概念,并采用209个属性以实现资源的描述。

(3)数据的开放与关联。目前,书目资源主要储存于图书馆内,并用高度结构化的MARC格式进行资源描述,使其只能在图书馆专门的软件系统如书目检索系统OPAC中运行,或通过Z39.50客户端进行查询,利用率较低,数据价值有限,具有“孤岛式”封闭性。尽管MARC可通过§2及§3子字段在6XX字段中来实现系统内记录与外部数据之间(如词表等规范档)的关联^[10],但对于实现书目数据在Web环境中与其他数据的开放关联来说还有较大差距。为适应关联数据环境,BIBFRAME 2.0采取了如下措施^[11]:采用URI标识Web资源,提高链接的灵活性,突破规范“名称”检索的局限;复用FOAF、SKOS、DC、RDF、RDFS等标准成熟词表来扩展Web数据,实现与其他权威开放关联词表的链接,便于多个词表之间的互操作,使书目信息在图书馆内部和外部都能被访问和利用。

1.2 BIBFRAME 词表的重用

数据模型的选取依赖于数据集的特点及其描述任务。BIBFRAME模型确定后,需重用或定义词表对书目数据进行描述表示。关联数据原则强调重用已有词表术语描述数据,这样更有利于实现数据间的互操作与共享,增强词表之间的链接。Duval等指出,没有哪一个元数据标准可以符合多个应用需求,应从一个或多个元数据模式中选择元素组装形成复杂模式以适应需求^[12]。

BIBFRAME词表中“类”及“属性”语义丰富,为书目资源提供了多种表达类型及载体表现形式,为用户精准查找资源提供了多种导航方式,其中“类”用于表达事物对象,且每个“类”都有其独有的属性以及从父类继承的属性。目前,BIBFRAME仍处于不断完善的状态,其最新版本中包括3个核心类的193个类,如文本、地图、音频、乐谱、舞谱等作品类,印刷、手稿、档案、电子、触觉材料等实例类。“属性”用于表达所描述资源的特征以及资源之间的关系,根据属性关联对象及性

质可分为142个对象属性、62个数据属性、5个对称属性,如作者、日期、题名、主题、语种、代理及管理元数据等描述属性,以“relatedTo”为父属性的丛编、替代、合并、翻译、相同等子属性^[13],以及新增的11个逆属性以提供关联资源的双向链接,充分体现了本体构建的可扩展性原则。BIBFRAME词表基本覆盖了CNMARC中0—9全部字段的书目记录信息,一定程度上保证了书目数据在全网域范围内的访问、定位和关联,可作为图书馆书目数据化优先选用的数据模型。

BIBFRAME词表重用可以分为三种方式:完全重用、生成扩展词表、部分重用。“完全重用”意味着完全采用BIBFRAME词表中的术语进行描述,如PCC(Program for Cooperative Cataloging,合作编目计划)“CONSER”关联数据项目中仅以BIBFRAME词表为核心实现RDA到BIBFRAME映射^[14],如rdaw:dateOfWork映射为bf:originDate,rdac:contributor映射为bf:contribution等。“生成扩展词表”是在BIBFRAME词表基础上重用其他词表的术语而生成新词表。如2016—2018年LD4P^[15]项目为涵盖更多大学图书馆的多种资源格式,由斯坦福大学、哥伦比亚大学、康奈尔大学、哈佛大学、普林斯顿大学、美国国会图书馆等联合开发。项目以BIBFRAME词表为核心,重用foaf、schema、prov、rdau、oa等的成熟词表,形成了扩展本体bibliotek-o。此外,LD4P项目针对《艺术和稀有材料BIBFRAME本体扩展》提出了对BIBFRAME属性及类的修改,如建议bf:dimensions属性重用ARM测量本体(Art and Rare Materials Measurement Ontology)描述,以增强描述的规范性及可表达性;上海图书馆关联书目数据描述^[16]基于BIBFRAME词表,重用rdf、dc、dct、bibo、foaf、madsrdf、void等成熟词表,构建了上海图书馆关联书目数据本体词表。“部分重用”意味着BIBFRAME与其他词表共同用来描述数据模型。

重用的程度取决于BIBFRAME词表能否完全表达资源描述对象。因此,结合BIBFRAME已有实践及为取代MARC而设计的特性,中文书目

数据可根据本地馆藏书目特点采用“生成扩展词表”方式,即以 BIBFRAME 词表为核心,部分引用其他词表或自定义相关术语而形成扩展词表,以为未来的中文资源提供重用基础。如中文资源通常采用《中国图书馆分类法》进行分类标引,可在 Classification 类下自定义 ClassificationClec 子类。

2 书目数据映射实现

确定 BIBFRAME 模型为书目数据模型后,需建立重用词表与 CNMARC 数据的对应关系,以实现书目资源的 RDF 三元组数据转换。针对 BIBFRAME 模型中“作品层”“实例层”“单件层”,对 CNMARC 数据元素进行层次化解构,识别其中各种书目实体,建立 BIBFRAME 词表的类、属性与 CNMARC 字段对应,并建立不同实体“类”层次之间的关系。

2.1 书目数据实体识别

BIBFRAME 2.0 模型基于 FRBR 实体关系模型建立了三个层次关系,并围绕三个核心“类”描述了相应的类和属性,揭示了资源的实体要素及关系。基于 BIBFRAME 模型中每一层次概念的具体含义,对各个字段进行解构,实现 CNMARC 平面数据立体化描述,有助于后期书目分层检索应用。以美国国会图书馆 MARC 21 书目数据到 BIBFRAME 2.0 映射为参考^[17,18],建立 BIBFRAME 词表类及属性与 CNMARC 记录的字段信息的对应关系,以实现语义化数据描述。

“作品层”反映了编目资源的概念本质,是一个抽象概念,包含代理(agent)、主题(subject)、事件(event)等属性特征。其中代理指通过作者、编辑、艺术家等角色与作品或实例相关联的人、机构、家族等,如对应书目记录的 701-702 人名字段、801 编目机构字段、711-712 团体名称字段、721-722 家族名称字段等;主题指表达一部作品的相关概念,主要对应书目记录的 6-主题分析块。而 CNMARC 书目记录是基于书目资源的载体表现信息编制,作品信息隐含在形成的书目记录中,从而使作品信息抽取具有复杂性^[19]。作品

层关注于资源的内容,当一部作品经过改编、翻译、修订等操作时,只要该作品的责任者和主题不变,即可指定是相同作品^[20]。因此,作品层关联的属性有管理元数据、题名、责任者、主题、分类、体裁、语种、受众等要素。

“实例层”是作品的具体化表达,反映了一部作品的不同物理介质或载体形态,即一个作品对应一个或多个实例,包含出版机构(publisher)和格式(format)等属性特征,如对应书目 210 出版发行项。当作品的载体发生改变时,即产生新的实例。不同的出版机构也会产生不同的载体表现。因此,出版机构、出版地点、出版日期、载体形态、标识符、版本说明等属于实例层。

“单件层”是实例的物理或电子版的单一副本,反映了诸如位置(物理或虚拟)、书架标记和条形码等信息^[21],可以帮助用户找到书目作品所在的物理位置,即包含收藏机构(heldby)和条形码(barcode)等属性特征,分别对应书目的收藏馆和条形码(905 馆藏信息)。当作品的存放位置发生改变时,即产生新的书架标记或馆藏信息,因此,所对应 CNMARC 字段即为描述书目馆藏信息和条目信息的条形码、索书号、馆藏信息。

基于此,可将平面化结构的 CNMARC 书目记录转换为基于实体的层次结构,如图 1 所示,以“代理”实体为例,将书目记录的代理实体分别映射于作品层、实例层、单件层。

2.2 各层次关系映射

实体与实体之间通过关系属性进行关联,如图 2 所示,包括三个层面的实体关系:一是书目之间的关系,主要体现为作品与作品之间、实例与实例之间的横向关系,通过“subPropertyOf”建立层层关联,直至顶级属性“relatedTo”,并在 BIBFRAME 中明确指明了属性使用范围,其中各具体属性对应明确的 CNMARC 字段。如图 2 左部“书目与书目之间横向关系”所示,包括翻译(字段 453、454)、合并(字段 436、447)、拆分(字段 437、446)、替换(字段 432、442)、丛编(字段 410、

225)、合订(字段423)等关系,既可用于作品与作品之间,亦可用于实例与实例之间。而不同载体的其他版本(字段452)关系则限定其取值为实例层,关于丛编关系详见2.3.4。二是书目内部的层次关系,表现为BIBFRAME模型描述的纵向层次关系,即作品有实例(hasInstance)、实例有单件(hasItem)。三是书目与相关实体之间的关系,表现为作品与人之间的代理关系(agent-701\$a/702\$a)、作品与主题之间的主题描述(subject-600/605/606\$x;classification-690\$a)关系、实例

与出版机构关系(agent-210\$c)、单件与收藏机构关系(heldBy-905\$a)等。

2.3 映射处理方式

2.3.1 精确对应

为构建映射的一对一关系,通常以CNMARC数据的字段与子字段联合表达相应的类或属性。其中字段指明了书目所映射的类,子字段描述了书目的对象,并揭示了类(字段)所拥有的描述和连接的实例化的层次属性。当以指示符表示对象

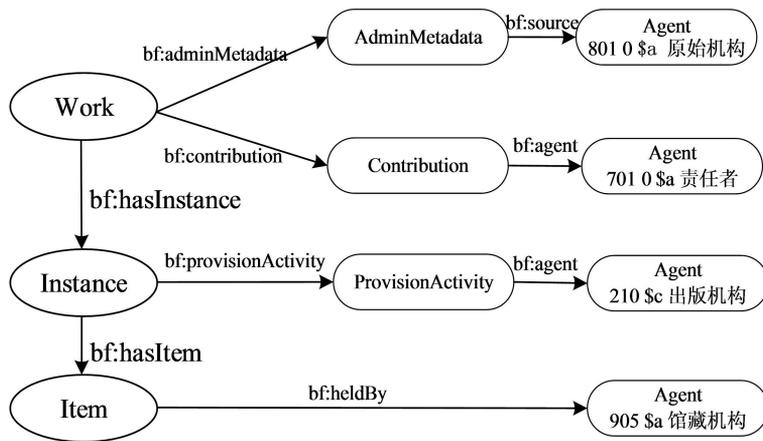


图1 书目“代理”实体的层次结构字段映射

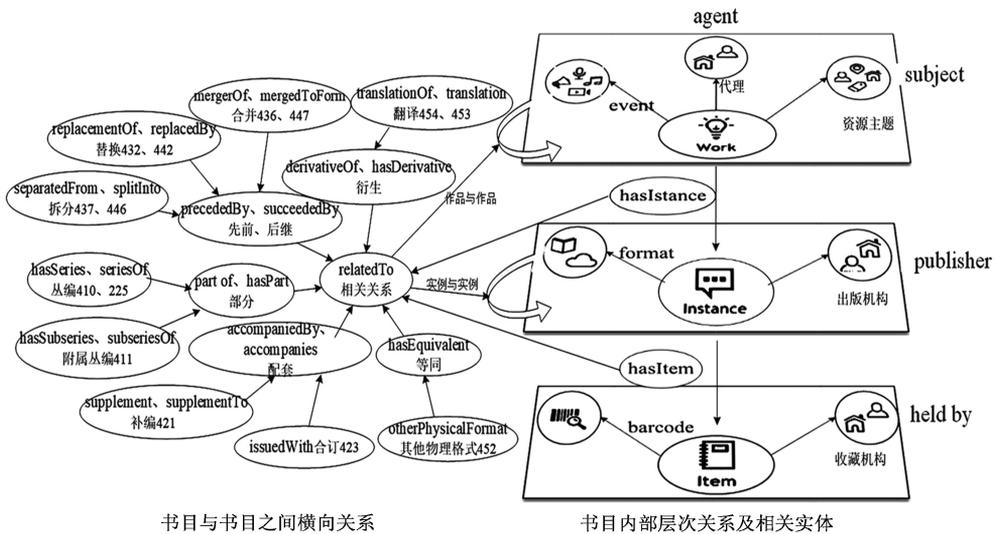


图2 CNMARC 实体及关系属性映射

所描述的数据内容及作用时,会影响映射的属性及属性值,需进一步辨别指示符的含义,如 711 标识团体名称的指示符(0 团体/1 会议)、801 标识机构作用的指示符(0 原始机构/1 转换机构/2 修改机构)等,当 801 记录来源字段指示符为 0 时,映射到属性 source,指示符为 1 对应属性 assigner,指示符为 2 对应属性 descriptionModifier。

2.3.2 字段集中式映射

BIBFRAME 利用统筹性功能属性“ adminMeta-

data”(管理元数据)描述资源的编目信息,如记录号、创建时间、来源等,与连接类“AdminMetadata”、核心类“Work”形成三元组陈述,并新增逆属性 adminMetadataFor 形成双向链接,进而利用 AdminMetadata 类将如 001 记录标识号、005 记录处理时间标识、100 编目时间、801 编目机构等编目元数据属性相关字段集中统一管理,如图 3 所示。与之相类似的属性还有责任者、分类、题名等。

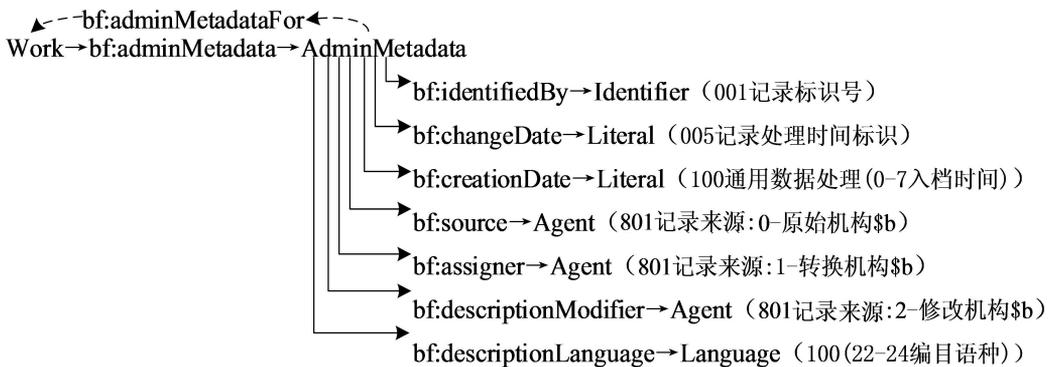


图3 属性集中映射示意图

2.3.3 优先级设定

当多个字段如题名、责任者、语种表达相同含义时,需设置优先级。CNMARC 中的著录必备字段与主标目检索字段表示的含义相同,则优先采用 CNMARC 的主标目检索功能字段,在无检索点功能块的情况下则采用著录必备字段,其余说明字段放于实例层。如作品层中题名字段有:著录必备字段 200 和可用于检索字段 500,其均有子字段 \$a 为书目的正题名、\$h 与 \$i 分别代表书目的分册(辑)号/名,则优先将 500 统一题名 \$a \$h \$i 用于作品层题名描述;若无 500 字段,则采用 200 \$a 字段,其余 200 \$d \$e 等副题名及 510 并列题名至 517 等其它题名归入实例层。这样作品层中仅包含用于识别书目内容的 \$a \$h \$i 题名信息,实例层中题名涵盖了书目所涉及的全部题名信息,包括正题名、副题名、变体题名、封面题名等。责任者优先采用 701/711 \$a 责任者检索

字段用于作品层,其 200 \$f \$g 责任者著录字段用于实例层。语种信息包括 100(22-24)、101 \$a、500 \$m 等字段,其中 100(22-24)表示编目所使用的语种,用于元数据语种,而 101 \$a 作品语种字段,500 \$m 为题名信息中表示作品语种字段,则可优先采用 101 \$a 正文语种字段,101 \$d 等文摘语种用于实例层。

2.3.4 丛编关系的明确

BIBFRAME 定义丛编关系既可用于作品与作品之间,亦可用于实例与实例之间。225 字段表示国际标准书目著录的丛编项,410 丛编关系字段表示丛编款目的检索点形式。410 字段与 225 字段均可采用“seriesOf/hasSeries”关系属性建立关联。

2.3.5 值词表数据化

书目数据所涉及人名、机构名、主题、分类、地点、文学体裁、载体形式、语种等,其取值要求来自于具体列表集合中的元素。其中,名称规范档、

《中国图书馆分类法》、《中国分类主题词表》中定义了人名、机构名、主题词、分类号等实体资源及关系,在 SKOS 的基础上将其转换为 RDF 表示,定义其相应命名空间,明确词表术语 URI,以关联数据的形式进行发布,以便于与书目数据进行数

据关联。如表 1 所示,对于体裁类型、载体类型、语种、图表类型等具有明确取值的,需对每个取值定义 URI,以 URI 链接的形式管理,如语种值词表中“汉语”可定义为“http://nlc.cn/vocab/language/chi”,以便于直接参引属性值 URI。

表 1 CNMARC 值词表元素及字段

值词表	元素	字段
发行方式 Issuance	集成性资源、专著、连续出版物	记录头标(7)
受众 IntendedAudience	普通青少年、学龄前儿童(0-5岁)、小学生(5-10岁)、少年(9-14岁)、青年(14-20岁)、研究人员、普通成人、特殊读者(如盲人)	100(17-19)
体裁类型 GenreForm	政府出版物、书目、学位论文、目录、法律和法规、索引、数字表格、文摘、技术报告、字、词典、试题、百科全书、文学评论、人名录、指南、条约、统计资料、卡通或连环画、成套教科书、专利文献书、标准或规定、会议出版物、纪念文集、小说、短篇小说、戏剧、诗词、散文、演说集、幽默、讽刺小品、非文学作品、书信、自传、个人传记、集体传记、包含传记资料、其它	100(20)、 105(4-9,11-12)
载体类型 Carriers	大型印刷品、报纸、盲文、微型印刷品、手稿、多种载体、小型印刷品、一般印刷品、其它多种载体	106
图表类型 Illustration	图表(包括画册)、地图、肖像、海图、设计图、单页插图、乐谱、摹真本、徽章图、世系表、表格、样品、录音资料、透明图片、彩饰	105(0-3)
语种 Language	汉语、阿尔巴尼亚语、冰岛语、波斯语(近代)、阿拉伯语、印度语、波兰语、孟加拉语等	100(22-24)、 101 \$ a、500 \$ m

2.3.6 作品层数据合并

作品层识别过程中,题名和责任者是识别作品的主要对象,当不同书目具有同一题名和责任者时,则涉及基于实例信息的作品层数据合并。即当不同出版社提供同一作品时,需去重保留唯一作品,同时聚合实例信息。如作家出版社、人民教育出版社各自出版的《红楼梦》属于同一作品,为避免重复编目,减少数据冗余及转换成本,对作品层 CNMARC 数据先合并再定义 URI。通过识别并抽取不同书目记录作品层的题名和责任者进行相似度匹配,去重后保留唯一数据值。

2.4 映射质量评估

BIBFRAME 是为书目资源关联数据化而定制的,其与 MARC 各个字段建立了对应关系,由于国内图书馆书目数据主要采用 CNMARC 的必备

字段及主要字段著录,映射质量基本可以得到保证。现采集国家图书馆 100 条 CNMARC 书目数据进行字段分析及检验。各书目记录字段分布如图 4 所示,其中 001、100、200、210、215、606、096 等关键字段均在每条书目里出现。

排除重复字段,共有 100 个子字段及字符代码需建立映射,97 个子字段及字符代码可实现完全匹配,其中 100 字段中 21 修改记录代码、25 音译代码、26-29 字符集代码等没有建立对应,造成一定程度的语义损失。这三个字符代码的设置考虑了现有字符集能否充分满足转录出版物文字数据的需求,以防止在数据交换过程中产生乱码。同时,映射效果取决于书目记录描述的完整性,比如各书目记录 4XX 字段普遍缺乏,导致其不能充分建立书目记录之间的语义关系。

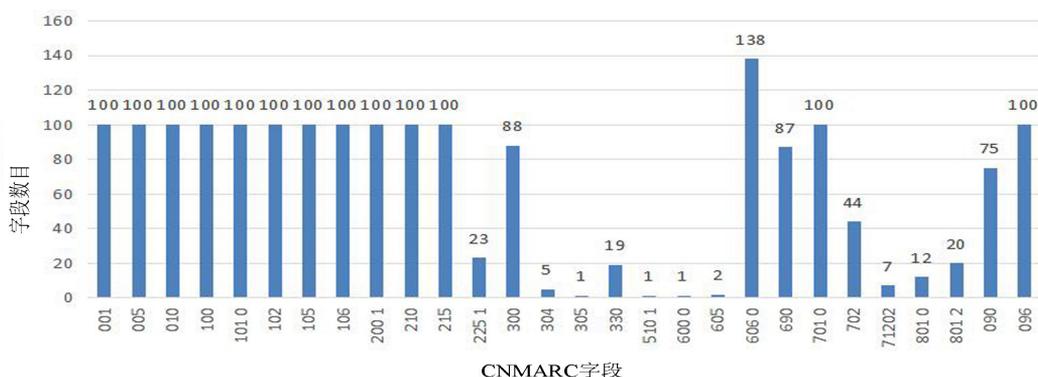


图4 国家图书馆 100 条书目记录的字段信息

3 书目内外部数据的 URI 关联

书目数据要实现真正意义上的关联,需对 CNMARC 实体资源定义唯一标识的 URI,通过 HTTP URI 使得书目记录在开放的互联网环境中以 URI 链接的形式被广泛使用和定位。属性作为实体资源的连接谓词,包括数据类型和对象属性,数据类型属性的取值始终是文本(字符串),对象属性的取值始终是资源,可分配 URI 作为资源访问的入口。为实现根据 URI 灵活定位到资源集合或者任一特定资源^[22,23],可以采用 303 重定向响应机制,以国家图书馆书目数据为例,定义 URI 为 <http://nlc.cn/bib>。基于多层次映射关系及表 1 值词表的分析,利用本

地定义的 URI 为语义化的 CNMARC 数据构建关联,构建如图 5 所示的 URI 实体资源关联关系模型,包括书目内外部数据关联。

3.1 书目内部关联

书目内部关联包括同一书目的纵向的层次结构关联,即“hasInstance”“hasItem”;不同书目之间的横向关系属性关联,如“hasSeries”关系属性等;以及不同书目数据匹配合并给以同一 URI 链接,如不同书目具有相同题名“title”。利用关系属性并为其分配对应的作品、实例、单件等 URI 链接,以建立内部数据关联,其 RDF 描述如下:

书目 1 作品与实例之间层次关联:

```
<bf:Work rdf:about="http://nlc.cn/bib/work/01">
  <bf:hasInstance rdf:resource="http://nlc.cn/bib/instance/01"/>
```

书目 1 有丛编关系为书目 2:

```
<bf:Work rdf:about="http://nlc.cn/bib/work/01">
  <bf:hasSeries rdf:resource="http://nlc.cn/bib/work/02"/>
```

书目 1 与书目 2 有同一题名 URI:

```
<bf:Work rdf:about="http://nlc.cn/bib/work/01">
  <bf:title>
    <bf:Title rdf:about="http://nlc.cn/bib/title/01"/>
<bf:Work rdf:about="http://nlc.cn/bib/work/02">
  <bf:title>
    <bf:Title rdf:about="http://nlc.cn/bib/title/01"/>
```

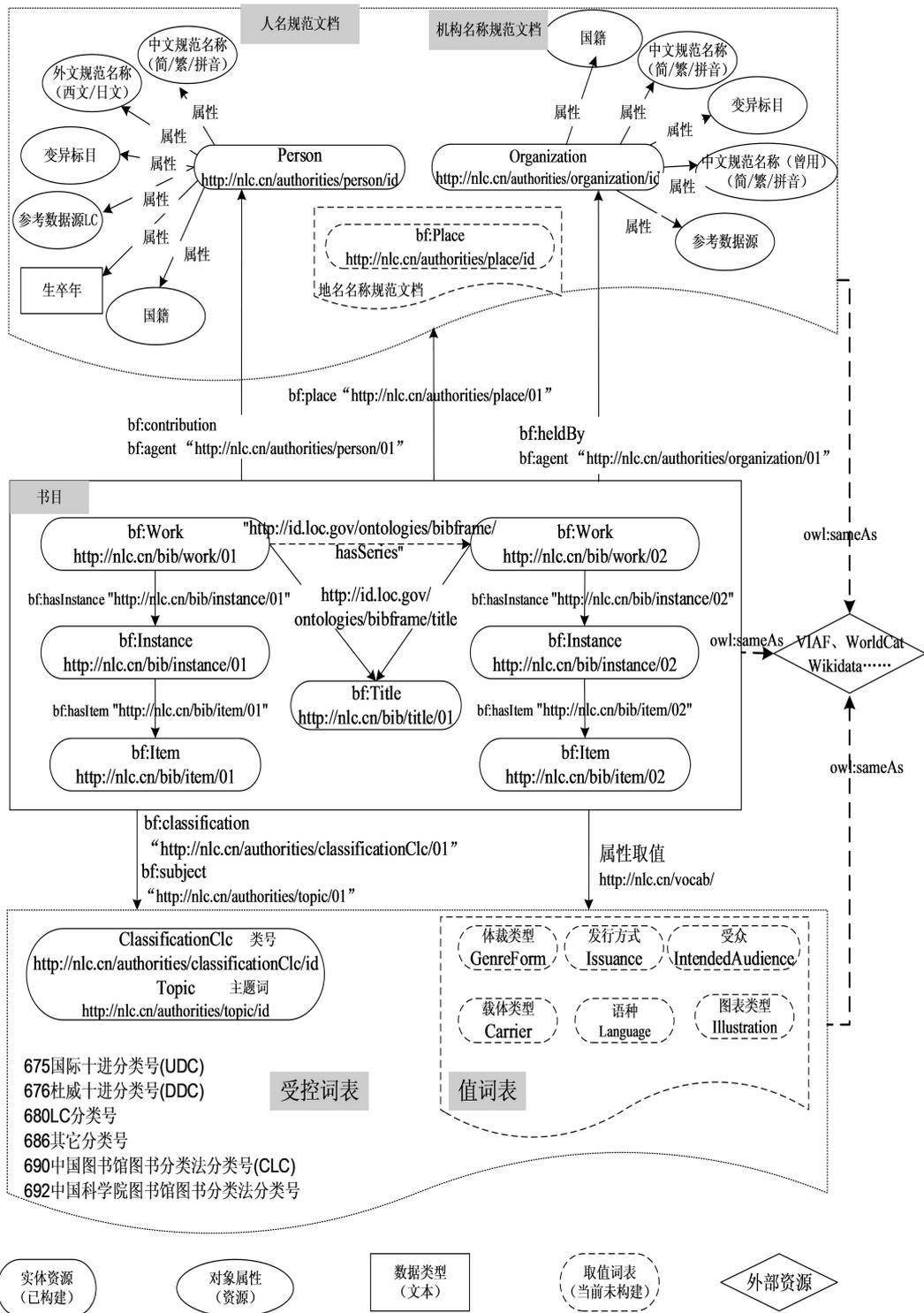


图5 URI 实体资源关联关系模型

3.2 书目外部关联

书目外部关联是将书目涉及的实体资源通过引用 URI 建立与不同知识库之间的关联,从而达到利用外部资源丰富本地书目数据的目的。如图 4 所示,链接的资源库有名称规范档、《中国分类主题词表》及其外部资源 VIAF、WorldCat、Wikidata 等。实体资源发布为关联数据后可使其在全局

域语义网内被广泛引用,即可通过相应的值链接进入相应的外部网站。通常内部资源直接引用实体资源的 URI,如各种取值词表或名称规范档,书目内部资源可直接引用相应属性值 URI;外部资源采用“owl:sameAs”将内外部数据相匹配的资源建立关联,如与 VIAF 建立链接,以“owl:sameAs”形式建立关联及定位。RDF 表示方式描述如下:

```
<bf:Work rdf:about="http://nlc.cn/bib/work/01"/>
  <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Work"/>
  <bf:contribution>
    <bf:Contribution>
      <bf:agent>
        <bf:Agent rdf:about="http://nlc.cn/authorities/person/01">
          <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person"/>
          <rdfs:label>作者</rdfs:label>
          <owl:sameAs rdf:resource="http://viaf.org/viaf/id"/>
        </bf:Agent>
      </bf:agent>
    </bf:Contribution>
  </bf:contribution>
</bf:Work>
```

书目数据内外部数据关联的建立,便于形成多种数据源集成的图书馆数据网络,便于实现对各类资源的无缝浏览。

4 总结

本文通过探究 CNMARC 书目数据化中 BIBFRAME 模型的层次化描述及关联分析,以厘清中文书目数据应用 BIBFRAME 模型实现关联数据化的最有效方式。首先需确定 BIBFRAME 词表的重用方式,再分析书目实体及对应关系识别,从精确对应、字段集中映射管理、优先级设定、丛编关系、值词表、作品层数据合并等层面探讨了具体映射的处理方式,并结合实例进行映射质量评估,分析了映射字段损失情况。BIBFRAME 的层次结构打破了 CNMARC 记录的单一线性平面化结构,建立了实体之间立体的关联关系网络。通过数据

关联机制,使书目数据以 URI 定义的方式与语义网中的任何资源建立链接关系,从而实现书目数据的开放性、语义性、关联性、可扩展性。本文不足之处在于,借鉴已有 MARC 数据转换为 BIBFRAME 的映射成果,对国家图书馆普通书目数据中若干常用字段、子字段、指示符进行实践,未对多样性馆藏资源的 CNMARC 数据转换作详细探讨。未来研究中将进一步探讨图书馆馆藏资源数据化实现所引起的业务流程的变化及发展。

参考文献

- 1 索传军,王新. 书目数据 RDF/XML 序列化方法研究[J]. 国家图书馆学报,2016(3):3-14.
- 2 BIBFRAME 2.0 Implementation Register [EB/OL]. [2021-05-15]. <https://www.loc.gov/bibframe/implementation/register.html>.
- 3 Tharani K. Linked Data in Libraries: A Case

- Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME[J]. *Information Technology and Libraries*, 2015, 34(1): 5-19.
- 4 Jin, Q. Enhanced Discovery with Linked Open Data for Library Digital Collections[J]. *Technical Services Quarterly*, 2021, 38(1): 17-32.
 - 5 李勇文. 书目框架(BIBFRAME)在中文书目数据中的应用范式探讨[J]. *图书情报工作*, 2016(2): 101-105, 145.
 - 6 夏翠娟, 许磊. 中文关联书目数据发布方案研究[J]. *数字图书馆论坛*, 2018(1): 8-16.
 - 7 许磊. CNMARC 与 BIBFRAME 的映射基础及其实现[J]. *数字图书馆论坛*, 2018(11): 26-32.
 - 8 周小萍. CNMARC 与 BIBFRAME 的映射及转换研究[J]. *图书馆杂志*, 2018(8): 21-29.
 - 9 SHARE - Virtual Discovery Environment [EB/OL]. [2022-05-17]. <https://share-vde.org/sharevde/clusters?l=en>.
 - 10 周小萍, 唐维. 基于 RDA 关系的扩展中文关联书目数据模型研究[J]. *图书馆学刊*, 2018(2): 45-49.
 - 11 胡小菁. BIBFRAME 核心类演变分析[J]. *中国图书馆学报*, 2016(3): 20-26.
 - 12 Duval E, et al. Metadata Principles and Practicalities[J]. *D-Lib Magazine*, 2002, 8(4): 1-16.
 - 13 BIBFRAME Ontology: BIBFRAME 2 List View [EB/OL]. [2021-11-10]. <https://id.loc.gov/ontologies/bibframe.html>.
 - 14 编目精灵. RDA 到 BIBFRAME 的映射报告 [EB/OL]. [2021-05-18]. <https://catwizard.net/posts/20171112192236.html>.
 - 15 LD4P Outputs[EB/OL]. [2021-05-26]. <https://wiki.lyrasis.org/display/LD4P/LD4P+Outputs>.
 - 16 基于 BibFrame 的上海图书馆关联书目数据本体词表[EB/OL]. [2021-06-05]. <http://bib.library.sh.cn/ontology/view/list>.
 - 17 Compare MARC converted to BIBFRAME[EB/OL]. [2021-06-10]. <https://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=5226>.
 - 18 Xu A, et al. From MARC to BIBFRAME 2.0: Crosswalks [J]. *Cataloging & Classification Quarterly: CCQ*, 2018(2-3): 224-250.
 - 19 胡小菁, 高红. CNMARC 书目记录的关联数据转换: 作品层分析[J]. *图书馆杂志*, 2019(1): 74-82.
 - 20 胡晓鹰. 中文编目 FRBR 化之各层次的识别与著录[J]. *图书馆杂志*, 2009(3): 24-26, 63.
 - 21 Overview of the BIBFRAME 2.0 Model [EB/OL]. [2021-06-26]. <https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
 - 22 贾君枝, 石燕青. 中文个人名称规范文档的关联数据化研究[J]. *情报学报*, 2016(7): 696-703.
 - 23 Cool URIs for the Semantic Web [EB/OL]. [2021-06-25]. <https://www.w3.org/TR/cooluris/>.
- (贾君枝 教授 中国人民大学信息资源管理学院, 崔西燕 中国人民大学信息资源管理学院情报学专业 2020 级博士研究生)

收稿日期: 2022-01-21