

FRBR、RDA 与 eXtensible Catalog

金 晶^{1,2} 姜恩波¹

(1 中国科学院国家科学图书馆成都分馆 成都 610041

2 中国科学院研究生院 北京 100190)

文 摘 通过对 FRBR/RDA 相关背景和发展概况的介绍,以 eXtensible Catalog (XC) 元数据管理发布平台为例,讲述了 XC 在传统图书馆元数据转化方面的相关细节和优势,以期为我国图书馆关联数据化的发展提供一些借鉴与参考。

关键词 eXtensible Catalog FRBR RDA 关联数据

FRBR、RDA and eXtensible Catalog

Jin Jing^{1,2} Jiang Enbo¹

(1. Chengdu Branch of National Science Library, CAS, Chengdu 610041;

2. Graduate University of Chinese Academy of Science, Beijing, 100190)

Abstract: This article introduces the background and development situations about FRBR/RDA, takes the eXtensible Catalog (XC) as an example to represent the details and strengths of the XC in metadata transformation services. Hope it will provide reference for the development of library linked data in China.

Key words: eXtensible Catalog FRBR RDA linked data

1 FRBR/RDA

1.1 FRBR 的产生与发展

由于信息资源的海量增长和文献类型的推陈出新,传统文献编目工作面临巨大压力,为了提高编目的质量和效率,1997 年 IFLA 成立了国际研究小组对“书目记录的功能需求 (FRBR)”进行深入探讨与研究。1998 年,该小组发布了 FRBR 的最终研究报告——《FRBR: Final Report》。该报告指出 FRBR 是用 E-R 模型 (如图 1) 来分析和定义作品之间实体与属性关系的框架,通过该模型可以清晰定义书目记录中的实体、实体中的属性及各实体之间的关系。FRBR 的研究目的有两个:一是提供一个清晰定义的结构化框架,使书目记录中的数据与用户需求之间产生关联;二是推荐由国家书目机构创建的记录的基本功能级别^[1]。

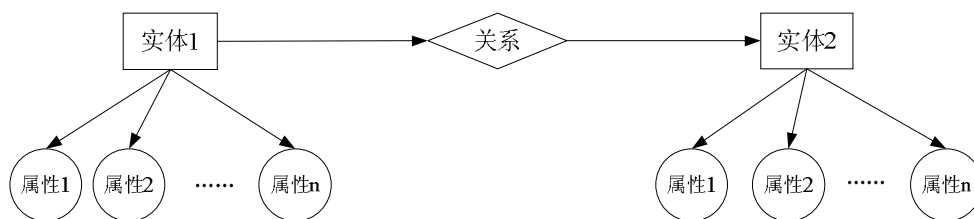


图 1 FRBR 的 E-R 概念模型框架

我们通常使用的 MARC 元数据是平面的线性结构,只注重文献的物理形态,无法同时展现作品不同载体、不同内容表达之间的关系;而 FRBR 则将编目层次上升到以文献主题为核心,更注重书目之间的关系描述。它从一个全新的视角审视书目记录的结构和关系,改变了图书馆界描述文献的基本方法。在 FRBR 提出之前,人们对于“作品”、“单件”的概念理解是非常模糊的,对于所有的纸质资源都可以统称为“书籍”,这极大的阻碍了图书馆员和用户对于信息的查找、分类。而 FRBR 则将书目资源分为四个层次进行描述,把作品的不同内容表达和载体表现相关联,避免了用户在检索某一作品的相关版本时进行不必要的重复查找。

FRBR 的四层书目结构包括“作品 (work)”、“内容表达 (manifestation)”、“载体表现 (expression)”和“单件 (item)”。作品 (work) 是一个抽象的实体, 指的是资料对象中包含的内容, 不涉及载体形态。内容表达 (expression) 是通过文字、声音、图像等形式实现作品的方式。对原作品在内容上的任何改动都被视为一种新的“内容表达”。载体表现 (manifestation) 是对于“内容表达”的实现, 每一种“内容表达”都必须有物理实体承载, 原作品载体形式一旦发生变化, 就会产生新的“载体表现”。目前 MARC 字段多局限于描述作品的“载体表现”, 对于“内容表达”层面的揭示并不多。单件 (item) 是一种具体物理资料的指代, 有时“单件”可理解为“复本 (copy)”^[2]。以上四者之间的关系如图 2 所示, 从“作品”到“单件”是一个由抽象到具体的渐进过程。

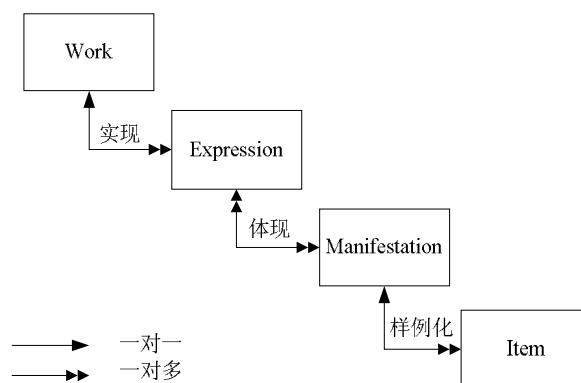


图 2 FRBR 四层书目结构模型

虽然 FRBR 提出了图书馆界描述文献的新方法, 但它毕竟只是一个概念架构, 必须有相应的编目规则予以支撑。在 FRBR 的推动下, 一种新的编目规则——RDA 应运而生。

1.2 RDA 概述

进入新世纪以来互联网发展迅速, 人们查找信息时更愿意跳过图书馆转而使用网络搜索引擎; 检索的资源范围也不再仅仅局限于图书馆的馆藏资源, 而是扩展到整个网络; 所需要的信息也不再是一本书、一篇文章, 而是经过深层次知识挖掘后形成的理论。为了跟上时代的步伐, 图书馆应当增加所收录信息资源的种类, 同时对相关编目准则和规范进行修订。1978 年颁布的 AACR2 在经过多次修改后, 依然无法满足当今时代的要求, 所以 JSC 决定重新制定一套能够在网络时代指导图书馆员编目的规范, 并将其命名为 RDA (Resource Description Access, 资源描述与检索)。

RDA 在 AACR2 的基础上做了很大改动, 包括删除过时、冗余的术语, 保证一致性; 修正含糊不清和前后矛盾的术语; 修订资源类型的分类方式和一般资料标识 (GMD) 等^{[3][4]}。它的主要目的是为了反映与 FRBR 和 FRAR 模型中定义的实体属性和关系有关的记录数据提供指导^[5]。2008 年颁布的 RDA 草案^[6]将 RDA 分为 10 个部分, 共 37 章进行介绍, 其内容按照 FRBR 的属性和关系排列。RDA 在用户任务的规定上也与 FRBR 相一致——“发现”与用户请求相一致的资源, “识别”资源的载体表现形式, “选择”所需要的资源, 最终通过借阅、访问、购买等方式“获得”资源^[7]。表 1 揭示了 RDA 各个部分与 FRBR 的对应关系。

如今, 图书馆数字化的结果就是越来越多的纸本资源转化为电子资源存储在互联网上, 这样的转变为图书馆的数据库建设和维护以及馆藏资源的著录和检索带来了便利。RDA 的开发就是为了应对这些新兴的数据库技术, 并且利用这些新兴技术组织有效、灵活地数据获取、检索和显示。eXtensible Catalog 系统平台就很好地应用了 RDA 来完成这些工作。

表 1 RDA 各部分与 FRBR 的对应关系表

RDA 各部分标题	与之对应的 FRBR 实体属性
-----------	-----------------

1.记录的载体表现/单件属性	第 1 组实体的属性
2.记录的作品/内容表达属性	
3.记录的个人/家族/团体属性	第 2 组实体的属性
4.记录的概念/实物/事件/地点属性 (缺)	第 3 组实体的属性
5.记录与作品/内容表达/载体表现/单件的基本关系	第 1 组实体间的主要关系
6.记录与资源相关的个人/家族/团体的基本关系	第 1、2 组实体间的关系
7.记录主题关系 (缺)	第 1、2 组实体间的关系
8.记录与作品/内容表达/载体表现/单件的相关关系	第 1 组实体间的其他关系
9.记录与资源相关的个人/家族/团体的相关关系	第 2 组实体间的其他关系
10.记录与概念/实物/事件/地点的相关关系 (缺)	第 3 组实体间的其他关系

2 RDA 在 XC 中的实现

2.1 eXtensible Catalog (XC)

自 20 世纪 50 年代以来,计算机开始逐渐渗透到各个学科,图书馆界也开始探索计算机的各种应用。1969 年由美国国会馆发行的机读目录 MARC 正式将计算机应用扩展到整个图书馆领域,实现了图书馆的全面自动化,大大提升了图书编目的速度。但由于如今越来越多网络、多媒体资源的出现,给传统编目理论带来了巨大挑战。尤其是在关联数据提出之后,以 MARC 元数据编码的书目信息必须要经过数据转换方可加入关联数据云。基于此 eXtensible Catalog Organization 在 A.W.Mellon 基金会的支持下开发了 eXtensible Catalog 元数据管理发布平台。

eXtensible Catalog (XC) 是一个面向图书馆的、开源的元数据管理发布平台,由 XC 项目小组的工作人员开发并维护,其目标之一就是非 MARC 系统中允许 MARC 元数据的重用。XC 包括四个软件工具包^[8]:

- (1) Drupal Toolkit: 为 Web 网站提供统一的检索界面;
- (2) Metadata Service Toolkit (MST): 是 XC 的核心,为元数据提供各种处理与转化服务;
- (3) OAI Toolkit: 负责定期收割机构数据库的数据资源;
- (4) NCIP Toolkit: 完成馆际互借和流通应用之间的数据交流。

本文将主要对 MST 中的 MARC 元数据转化服务进行介绍。

2.2 XC Schema

MST 模块提供 MARC 元数据转化服务,即依据 MARC 与 XC Schema 的转化映射表 (<http://code.google.com/p/xcmetadata servicestoolkit/wiki/TransformationServiceStepsBib>) 将 MARC XML 转化为以 FRBR 实体为描述框架的 XC XML 格式。XC Schema 是 XC 项目开发小组自行定义的一种元数据模式,它尽可能多地引用各种目前正在使用的内容描述框架中的数据元素和属性,自己只定义其中的少部分属性。这种方式不仅简化了 XC Schema 的定义过程,也帮助用户和机器能够更好的理解相关属性信息。XC Schema 大体上由四部分构成: 22 个 RDA 元素(如表 2)、11 个 RDA 角色指示符(如表 3)、所有的“dcterms”元素以及专为 XC 系统功能设计的 XC 元素,以上所有的词表元素都已在 NSDL 元数据登记中心注册^[9]。

XC Schema 之所以选择 RDA 是因为 RDA 基于 FRBR 三元组定义,且可以灵活的与其他元数据 schema 兼容。为了能够将转换后的 FRBR 记录发布为关联数据,XC 开发小组还另外定义了一些 XC 独有的元素属性,这是为了在实现关联数据的相关应用时可以方便的获取资源 URI 标识。XC Schema 在设计之初使用实际词汇作为元素标识,随后便改为采用 URI,这样做的好处是: ①URI 的作用明确被限定为资源标识符; ②减少 schema 的维护成本,无需在词汇本身形态改变的情况下做大范围的更新; ③NSDL 元数据登记系统可以通过版本控

制机进行更新^[10]。

表 2 XC Schema 中所使用的 22 个 RDA Elements^[11]

XC Schema 中所使用的 22 个 RDA Elements	
rdvocab:coordinatesOfCartographicContent	rdvocab:scale
rdvocab:dissertationOrThesesInformation	rdvocab:dimensions
rdvocab:identifierForTheWork	rdvocab:editionStatement
rdvocab:natureOfTheContent	rdvocab:frequency
rdvocab:titleOfTheWork	rdvocab:modeOfIssuance
rdvocab:artisticAndOrTechnicalCredits	rdvocab:numberingOfSerials
rdvocab:awards	rdvocab:placeOfProduction
rdvocab:identifierForTheExpression	rdvocab:plateNumber
rdvocab:illustration	rdvocab:publisherNumber
rdvocab:performerNarratorAndOrPresenter	rdvocab:soundCharacteristics
rdvocab:placeAndDateOfCapture	rdvocab:statementOfResponsibilityRelatingToTitle

表 3 XC Schema 中所使用的 11 个 RDA Role Designators^[11]

XC Schema 中所使用的 11 个 RDA Role Designators
rdarole:artist
rdarole:author
rdarole:compiler
rdarole:composer
rdarole:speaker
rdarole:director
rdarole:editor
rdarole:illustrator
rdarole:performer
rdarole:producer
rdarole:translator

2.3 MARC 记录的 FRBR 化

传统的图书馆元数据大都是为了某个特定的应用程序而设计,为了提高系统的灵活性和重用性,有必要将无法显示记录间关系的传统 MARC 记录转化为能够在各种资源之间建立关联的 FRBR 记录。一条 MARC 记录包含的信息往往需要映射到 FRBR 的不同实体,从而形成多条 FRBR 记录,但 MARC 著录的时候往往缺少一些必要的说明性信息来让后续应用准确地映射,对于包含了丰富内容的字段(比如 245 字段)在映射时也会比较复杂,这些都为 XC 的转化服务带来了很大困难。XC 开发小组制订了 MARC 与 XC 的映射表,帮助 XC 转化服务的顺利进行。

MARC 数据和 FRBR 实体映射时,不仅仅是属性元素之间的映射,更重要的是发现 MARC 记录和 FRBR 实体之间的关系。很多时候这种关系是显而易见的,但还是存在 MARC 元素可能会和一个 FRBR 中未定义的实体相关联的情况,这就需要编目人员手工操作进行分类转化。在转化过程中,XC 会为每一条 FRBR 记录赋予一个全网唯一的 URI 标识,同时在记录的最后还标识出与之相关的其他实体的 URI,通过该语句用户可以很容易的查找到与此记录相关联的其他记录信息,实现了书目数据的内部关联。

例如,有两条 FRBR 记录的 URI 分别为: <xc:entity id="oai:mst.rochester.edu:Metadata-ServicesToolkit/marctoxttransformation/392" type="work">和<xc:entity id="oai:mst.rochester-

edu:MetadataServicesToolkit/marctoxttransformation/393" type="expression">, 在 ID=393 的记录末尾有这样一条语句: <xc:workExpressed>oai:mst.rochester.edu:MetadataServicesToolkit/marctoxttransformation/392</xc:workExpressed>, 用户通过这条语句就可以形成如图 3 所示的内部关联。



图 3 FRBR 记录的内部关联图^[12]

2.4 XC 的优势

XC 为图书馆遗留数据提供了一个转换为关联数据的平台, 同时解决了元数据在转换过程中存在的各种潜在问题, 对图书馆关联数据的发展起到了举足轻重的作用。在 MARC 元数据转换前, XC 还会对 MARC 记录进行必要的预处理:

- (1) 将 MARC ISO2709 转化为 MARC XML, 便于用户通过普通浏览器查询检索馆藏资源;
- (2) 删除 MARC 与 XC Schema 映射过程中一些不必要的字段;
- (3) 将 MARC 数据中的字符代码转换为更适合用户理解的名称表达, 例如若 leader06=a, 则在新增的 932 字段中将资料类型标明为“language material”。

XC 的预处理操作不仅便于用户理解 MARC 记录各字段的具体含义, 同时也避免了转化服务过程中一些不必要字段可能带来的复杂操作。

另外, 关联数据之所以近几年间才在图书馆界发展起来, 主要是因为各大图书馆都不愿意放弃现有成熟的编目技术和 OPAC 系统平台, 担心新技术可能带来的未知风险。然而, XC 为图书馆提供了一个安全、高效的关联数据实验平台:

- (1) 自带 OAI 数据收割功能, 将机构知识库中的数据采集回来后隔离实验, 不会对原有数据资源的完整性、准确性造成破坏;
- (2) 数据批处理, 大大缩短转换时间, 提高编目人员的工作效率;
- (3) 提供对元数据的存储、检索和显示;
- (4) XC Schema 对于 RDA 的使用, 推动了 RDA 的发展, RDA Toolkit 已经将 XC 作为其唯一的应用实例。

3 结语

互联网的出现及迅速兴起, 不仅改变了人类的生产、生活方式, 而且极大的改变了人类获取、存储、传播信息的方式。在这一背景下, 图书馆界需要各个部门的研究人员通力配合, 包括 MARC 编目人员、schema 编制人员, 软件开发人员和语义网领域专家等, 制订新的编目规范、编制配套的 schema、设计并测试各类转化算法、发布图书馆关联数据。如今 RDA 的实施就将极大地促进语义技术在图书馆的应用, 使图书馆基于网络的数字资源组织、整合和服务全面进入规范控制时代^[13]; 本文所介绍的 eXtensible Catalog 元数据管理发布平台将传统图书馆元数据转化为关联的 FRBR 记录, 为即将进行关联数据化的图情机构提供了一个可供借鉴的范本。如今我国还鲜有图书馆将自己的馆藏资源加入语义网并发布为关联数据, 本文也旨在为我国的图书馆关联数据化发展提供一些参考。

参考文献:

- [1] IFLA. 书目记录的功能需求中文版 [EB/OL]. [2010-02-11]. <http://archive.ifla.org/VII/s13/frbr/frbr-zh.pdf>
- [2] 刘素清. IFLA 书目记录功能需求 (FRBR) 初探[J]. 大学图书馆学报. 2004 (6): 65-69
- [3] JSC. 2002 Annual Report [EB/OL]. [2010-02-12]. <http://www.rda-jsc.org/docs/annrep02.pdf>

- [4] JSC. 2003 Annual Report [EB/OL]. [2010-02-12]. <http://www.rda-jsc.org/docs/annrep03.pdf>
- [5] JSC. RDA: 资源著录与检索内容说明书中文版[EB/OL]. [2007-03-13].http://www.rda-jsc.org/docs/5rda-prospectusrev3_chi.pdf
- [6] RDA Toolkit. RDA Constituency Review Draft [EB/OL]. [2008-11-24]. <http://www.rdatoolkit.org/constituencydraft>
- [7] RDA Constituency Review Draft, 0 Introduction [EB/OL]. [2011-07-22]. http://www.rdatoolkit.org/constituencyreviewfiles/constituency_review.zip
- [8] eXtensible Catalog Organization. XC Software Suite [EB/OL]. <http://www.extensiblecatalog.org/sites/default/files/home/eXtensible%20Catalog%20Information%20Sheet.pdf>
- [9] Jennifer Bowen, David Lindahl. RDA and the eXtensible Catalog [EB/OL]. [2010-04-08]. <https://urresearch.rochester.edu/institutionalPublicationPublicView.action?institutionalItemId=14154&versionNumber=1>
- [10] Jennifer Bowen. Moving Library Metadata toward Linked Data: Opportunities Provided by the eXtensible Catalog [EB/OL]. [2010-10-20]. <http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/1010>
- [11] eXtensible Catalog Organization. eXtensible Catalog Project XC Schema Definition[EB/OL]. [2009-01-07]. <http://www.extensiblecatalog.org/sites/default/files/fulltext/XCSchemaProperties20090107.pdf>
- [12] 胡小菁. RDA 与关联数据[EB/OL]. [2010-08-23]. <http://www.kevenlw.name/downloads/shld/%E7%BC%96%E7%9B%AE%E7%B2%BE%E7%81%B5--RDA%E4%B8%8E%E5%85%B3%E8%81%94%E6%95%B0%E6%8D%AE.pdf>
- [13] 刘炜, 胡小菁等. RDA 与关联数据[J]. 中国图书馆学报. 2012 (1): 34-42

金晶 女, 1987年生, 硕士研究生。

姜恩波 男, 1972年生, 副研究员, 已发表论文数篇。

(收稿日期: 2012-05-13 编发: 许桂菊)