

BIBFRAME介绍及实践 现状研究

国家图书馆中文采编部 柏林

2020.11

目录

- 1 BIBFRAME概况
- 2 BIBFRAME研究热点
- 3 BIBFRAME实践现状
- 4 可行性分析
- 5 总结与展望

BIBFRAME概况

PART ONE

01

- 概念、总体情况
- 理论渊源及演进过程
- BIBFRAME1.0与2.0对比分析
- BIBFRAME2.0模型架构
- 用例及实现

产生背景

- MARC的优点：操作简便、字段完善、著录详细、使用灵活、广泛；
- 缺陷：随着网络资源的指数化增长，其语义表达能力的局限性日益突出；

产生背景

- 互联网发展，关联数据、语义网等概念的提出
- 传统的 MARC 格式
- MARC must die
- 新的书目格式

BIBFRAME (Bibliographic Framework Initiative , 书目框架行动)

- 美国国会图书馆 (LC) 2011年发起的一项将书目描述规则设计为关联数据的运动
- 由图书馆专业人员设计
- 为图书馆专业人员使用
- 期望在未来工作中取代MARC机读目录
- 使书目信息在图书馆领域内外得到广泛的应用

- 采用资源描述框架 (RDF) 三元组模型对数据建模
- 书目数据以数据元素为单位
- 面向语义网和关联数据进行架构，对书目资源进行语义化组织与关联，促进图书馆资源在关联数据环境中与其他资源产生联系
- 有利于实现网络环境下书目信息交流与共享
- 针对以链接数据技术为基础的更广泛的网络中
- 构建可容纳图书馆、博物馆、档案馆等多种机构文献资源的宏观模型
- 为资源描述提供了基础

- LC发布了BIBFRAME模型的框架、MARC记录转换为BIBFRAME的方法、元数据方案、用例等
- 丰富的类和属性，其中所包含的属性值多为结构化数据，一个属性又可包含多个子属性，它的专业度更高，并且对书目描述更加细致准确
- 开放性和包容性

理论渊源及演进过程

- 2011年10月，LC发布了BIBFRAME计划，确定MARC 21格式的转换路径，保留近几十年来支持资源共享和编目成本节约的强大数据交换
- 考察与借鉴已有关联数据项目的成功案例及建模经验
- 采用FRBR的实体-关系（E-R）模型分析法，对所涉及的实体、实体的属性及其关系等进行分析 and 明确标识，使机器能够理解和处理实体及其关系

理论渊源及演进过程

- 2012年11月，LC发布的BIBFRAME模型草案（即BIBFRAME1.0），设计了“作品”（Work）、“实例”（Instance）、“规范”（Authority）和“注释”（Annotation）4个核心类
- “作品”和“实例”被视为FRBR模型的简化版

理论渊源及演进过程

- 2013年初，发布遵循BIBFRAME模型的词表
- 2014年，联合美国国立医学图书馆、大英图书馆、德国国家图书馆、联机计算机图书馆中心等多家机构对BIBFRAME的功能进行研究和测试
- 2015年6月起，发布了5个修订建议，取消规范类，增减单件类
- 2015年10月，以BIBFRAME2.0为名发布7个修订草案，明确将取消规范类和注释类，增加单件核心类
- 2016年4月，正式推出BIBFRAME 2.0，发布转换工具

BIBFRAME 2.0的变化

- BIBFRAME 1.0 的4个核心类（作品、实例、规范、注释）有子类，其他类无子类
- BIBFRAME 2.0 的3个核心类（作品、实例、单件）中，单件没有子类，其他13个非核心类有子类

BIBFRAME 2.0的变化

1、类增加属性减少

- BIBFRAME 1.0 类 53 个，属性 289 个，合计 342 个；
- BIBFRAME 2.0 类 176 个，属性 197 个，合计 373 个；
- 类增加 123 个，属性减少 92 个，增加 31 个；
- 与 BIBFRAME 1.0 完全相同的类 39 个、属性84 个
- 部分改名，其余被取消

BIBFRAME 2.0的变化

2、部分属性改为类

- BIBFRAME1.0词表中，标识符有ISBN、ISSN、LCCN等不同标识符类型，题名有作品题名、实例题名、缩略题名等不同题名类型，是用属性来表达的
- 2.0词表中，这些类型改用类来表达

BIBFRAME 2.0的变化

- 有50个1.0的属性改为2.0的类，包括：
- 分类的4个分类法属性改为分类的子类
- 标识符的原37个属性改为标识符的子类
- 原提供者（provider）的4个属性发行、制作、出版、生产改为提供活动的子类
- 题名（Title）的原4个不同题名类型属性作品题名、实例题名、识别题名、缩略题名改为题名的子类

BIBFRAME 2.0的变化

- BIBFRAME 2.0 词表共有 58 个与属性对应的类，1.0 原有11个，2.0 新增 47个
- 通过首字母大小写来区分分类和属性，类首字母大写，属性首字母小写
- 2.0 词表中有少数类改为属性的，如原资源间关系类（Related）改为关系属性（related To），具体关系属性为其子属性

BIBFRAME 2.0的变化

3、复用外部成熟词表

- 直接使用 `rdf:Resource` 而取消顶级类 `Resource`
- 取消 `label` 和 `identifierValue` 属性而采用 `rdfs:label` 和 `rdf:value`
- Web 注释数据模型能满足处理书目数据的需要而取消注释类等

对比分析——修改

1、调整核心类

- ① 1.0 的核心类包括作品 (`Work`)、实例 (`Instance`)、规范 (`Authority`)、注释 (`Annotation`) 四大类
- ② 2.0 的核心类调整为作品 (`Work`)、实例 (`Instance`)、单件 (`Item`) 三大类，取消规范和注释，增加单件

对比分析——修改

1、调整核心类

- 其他子类：封面、评论、摘要、目次保留，作为顶级类的子类
- 原作品的11个子类
- 原实例的10子类
- 子类合集
- 单件

对比分析——修改

2、修改题名类及属性

- 1.0 题名类无子类，题名信息属性包括：题名 (`title`)、部分编号 (`partNumber`)、部分题名 (`partTitle`)、副题名 (`subtitle`)、缩略题名 (`abbreviatedTitle`)、作品题名 (`workTitle`)、实例题名 (`instanceTitle`)、识别题名 (`keyTitle`)、立法日期 (`legalDate`)、音乐的调 (`musicKey`)、原创日期 (`originDate`)、原创地 (`originPlace`) 等各种题名和题名附加信息共 26 个

对比分析——修改

2、修改题名类及属性

• BIBFRAME 2.0 :

原作品题名和实例题名由属性变为题名的子类

新增题名子类变异题名 (VariantTitle) , 原识别题名和缩略题名由属性改为变异题名的子类

新增变异题名子类并列题名 (ParallelTitle) 和总题名 (CollectiveTitle)

对比分析——修改

2、修改题名类及属性

- 其他类 : (针对视听资料) 容器题名、书脊题名、封面题名、附加题名页题名

对比分析——修改

2、修改题名类及属性

• 属性情况

部分被取消, 仅保留题名、副题名、 部分编号、部分题名 (由 part Title 改名为 part Name)

新增题名类的属性主要题名 (mainTitle)

新增变异题名类的属性变异类型 (variantType)

对比分析——修改

2、修改题名类及属性

• 总体情况 :

题名类包括作品题名、实例题名、变异题名 3 个子类

题名、主要题名、副题名、部分号、部分题名 5个属性

变异题名有缩略题名、识别题名、总题名、并列题名 4 个子类及变异类型1个属性

对比分析——修改

3、重新定义事件类

- 1.0 事件类 (Event) 定义为：事件的时间或地点，属性包括事件代理 (event Agent)、事件日期 (event Date)、事件地点 (event Place) 等与事件相关的人物、时间、地点等内容
事件属性 (event) 定义为事件的地理区域或时间范围，作为作品的主题

对比分析——修改

3、重新定义事件类

- 2.0事件类 (Event) 定义为：
被一种资源记录的发生在特定时间或地点的事情
事件与作品有关
事件类及一组对应关系属性事件内容支持事件描述，
取消原属性eventAgent、eventDate、eventPlace、event

对比分析——修改

- 4、改变职能描述相关的类和属性
- 5、标识符属性变子类

对比分析——删除

1、删除的类

- 1.0 的53个类中，除完全相同的39个外，改名的3个：
- Provider改名为Provision Activity
- Related改为为属性related To
- Relator改名为Contribution，其余11个被删除

对比分析——删除

1、删除的类

- 取消顶级类Resource，直接使用rdf:Resource；
- 取消规范和注释核心类
- 取消注释类的子类拥有资料（HeldMaterial）及其子类拥有单件（HeldItem），两者合并为单件类

对比分析——删除

1、删除的类

- 删除原实例中与发行方式有关的子类：
 - 专著（Monograph）
 - 多部分专著（Multipart Monograph）
 - 连续出版物（Serial）
 - 集成资源（Integrating）
 - 增加发行方式类（Issuance）

对比分析——删除

1、删除的类

- 取消类别（Category）
- 原属性载体类别（carrier Category）、内容类别（content Category）、媒介类别（media Category）、体裁（genre）分别改为载体（carrier）、内容（content）、媒介（media）、体裁形式（genre Form），并增加4个对应的类
- 取消描述管理信息（Description AdminInfo）

对比分析——删除

2、删除的属性

- 原BIBFRAME 1.0的289个属性中，除完全相同的84个外，有50个属性改为类
- 部分改名

对比分析——新增

- 1、增加管理元数据
- 2、增加通用类及属性
- 3、新增作品描述信息相关类及属性
- 4、新增载体描述的类和属性
- 5、新增音乐相关的类及属性

PART ONE

02

BIBFRAME研究热点

国外的BIBFRAME 2.0 研究

- | | |
|-----------------------------------|------------------------------|
| ① BIBFRAME2.0核心类模型的阐释与分析。 | ④ BIBFRAME2.0的相关比较研究。 |
| ② MARC21书目数据和BIBFRAME2.0模型与词表的映射。 | ⑤ BIBFRAME与关联数据的关系研究。 |
| ③ 不同资源类型的MARC21格式与BIBFRAME的转换与映射。 | ⑥ 与BIBFRAME有关的实验项目和研究项目进展汇报。 |

国内的BIBFRAME 2.0 研究

- | | |
|--|---|
| ① 2.0与1.0的比较研究，多侧重在核心类的演变与比较研究及词表的变化与分析研究。 | ③ CNMARC格式与2.0的映射比较与转换测试。 |
| ② 2.0与FRBR、RDA的关系研究。聚焦于BIBFRAME资源类型与RDA内容类型的对应及BIBFRAME2.0核心类与FRBR第一组实体的映射 | ④ 基于BIBFRAME2.0的相关探索。主要涉及语义联合目录模式建构和测试、数字图书馆语义搜索框架的核心功能模块的建立。 |

国内外研究存在的差异

- ①BIBFRAME对关联数据技术的运用研究有待深入。
- ②BIBFRAME在资源编目体系中的相关性研究不足。
- ③BIBFRAME 2.0的微观研究有待深入。
- ④ BIBFRAME在国内本地化研究有待加强。

BIBFRAME实践现状

PART ONE
03

EDM项目

BNB项目



LD4L项目 (linked data for libraries)



LD4Prod项目 (linked data for production)



上海图书馆的家谱本体设计

CNMARC到BIBFRAME的映射研究

- 建立CNMARC字段、子字段以及指示符与BIBFRAME词汇之间的映射关系
- BIBFRAME官网
- CNMARC与MARC 21 的差别

CNMARC到BIBFRAME的映射研究

(1) 映射策略

- BIBFRAME 2.0：三个层次：作品-实例-单件 (work-instance-item)
- CNMARC记录：标志块、编码信息块等10个功能块构成的平面层次结构
- 实现二者映射：以核心及子类为基础、对字段及子字段解构重组，并映射到类及属性

CNMARC到BIBFRAME的映射研究

(1) 映射策略

- 元数据转换说明：规定修定内容、满足元数据内容结构
- 取值内容要保留原始的子字段顺序
- 说明目标格式内容

CNMARC到BIBFRAME的映射研究

(2) 映射关系

- 一对一关系
- 表示一个源术语的语义信息与唯一一个目标术语的语义有极高的相似性或完全等同。
- 例如 101 文献语种字段，映射结果为 “Work-language-Language”

CNMARC到BIBFRAME的映射研究

(2) 映射关系

- 一对多关系
- 有多个目标术语的语义在其语义范围内
- 例如 200 题名字段，存在 500 统一题名字段的情况下映射到实例类，结果为 “Instance-title-Title”；不存在 500 统一题名字段的情况下应同时映射到作品类和实例类，结果为 “Work-title-Title” 及 “Instance-title-Title”；

CNMARC到BIBFRAME的映射研究

(2) 映射关系

- 一对多关系
- 同一字段的数据在进行三元组的转换时：
 - ① 106字段
 - ② 105字段
 - ③ 711/712字段

CNMARC到BIBFRAME的映射研究

(2) 映射关系

- 特殊的一对多关系
 - ① 字段在指示符不同的时候，含义不同，同一个字段就会映射到不同的三元组
 - ② 定长字段同一栏位根据取值需要映射到不同或多个三元组

CNMARC到BIBFRAME的映射研究

(2) 映射关系

- 多对一关系
- 源术语的语义外延较宽窄，这些术语的语义可以被语义外延较宽的一个目标术语所包含
- 映射优先级判断或对关系类型进行说明

CNMARC到BIBFRAME的映射研究

存在的问题

- (1) BIBFRAME 词表的很多属性词只用“Used with”说明其建议的定义域，而没有严格定义。

CNMARC到BIBFRAME的映射研究

存在的问题

(2) 关联书目数据的发布先要解决作品的识别问题

CNMARC到BIBFRAME的映射研究

存在的问题

(3) 对于无总题名的汇编类数据，在CNMARC格式中启用新字段著录作品的责任者与题名信息；或者考虑众包方式，将这类作品的识别由用户人工完成

PART ONE

04

中文图书/文献编目应用 BIBFRAME的可行性分析

1、BIBFRAME的特点和优势

(1) 特点：

- 适应性广
- 面向语义网
- 以关联数据为基础

1、BIBFRAME的特点和优势

- (1) 兼容MARC格式
- (2) 对书目数据的抽象内容和其载体表现进行严格界定
- (3) 开放性和普适性
- (4) 数据关联、语义化

2、BIBFRAME对图书馆的影响

- (1) 以关联数据为基础
- (2) 编目与检索
- (3) 书目数据融入万维网
- (4) eg.馆藏作品的维护
- (5) eg.更新作品的关联属性
- (6) 对检索结果

3、对图书馆界的挑战

- 编目的主流是MARC，很多成熟的图书馆系统都是基于MARC开发，很多国家在使用不同的MARC格式编目
- BIBFRAME主要数据通过MARC转换
- 经济投入
- 考虑价值
- 图书馆业务
- 如何实现规范控制

3、对图书馆界的挑战

- CNMARC编目，累积了大量的书目数据和规范数据
- 适应中国的编目环境、如何借用成熟的本体模型或RDA词汇进行扩展
- CNMARC向BIBFRAME的转换中的具体问题

4、如何应对BIBFRAME

(1) 图书馆

- 整合有限的资源
- 格式转换
- 做法：
基础设施投入
加强培训

(2) 编目员

- 转变观念
- 学习新知识、发展新技能
- 多关注、多参与、多研究

PART ONE

05

总结与展望

BIBFRAME 优势

- 1900 万件作品，2400 万个实例，2260 万件单件以及 43 亿三元组
- (1) 描述范围广，适用性强
- (2) 采用 RDF 方式，语义性强
- (3) 开放关联，便于共享
- (4) 与 MARC 格式兼容
- (5) 方便灵活，扩展性强

THANKS