

关联书目数据基础

上海图书馆·许磊
福州 2019-10-14

目录 CONTENT

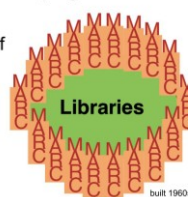
- 01 背景与需求
- 02 关联数据基础
- 03 BIBFRAME书目模型
- 04 CNMARC中作品的识别与建模
- 05 关联书目数据发布
- 06 现在与未来
- 07 BIBFLOW转换路线图
- 08 Folio的元数据管理与编目

01 背景与需求

背景与需求

Escaping Fortress MARC

Rest of
world



Most of
our users,
most of
the time

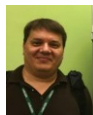
背景与需求



MARC must die
—Roy Tennant



If it isn't on Google, it
doesn't exist
—Jimmy Wales



MARC isn't Dead,
but it is a Dead
End
—Peter Murray



library needs linked data
platforms. It's one of our
last chances to embark on
innovations
—Andrew K. Pace

背景与需求

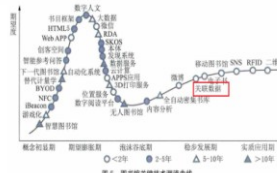


图9 图书馆关键技术演进曲线

新环境下的元数据管理需求

- **完整生命周期管理。**支持元数据从采集、编目、加工、服务、增值、交换、统计分析、评估的完整生命周期是元数据管理最迫切的需求
- **资源类型兼顾。**图书馆资源类型的多样化要求元数据能灵活地描述纸质、数字化、电子、网络资源，不受文本、图像、音视频等。
- **标准规范兼容。**不同资源类型和不同需求环境制定了不同元数据标准规范，需要支持灵活可扩展的、适合各种资源类型的元数据标准规范。
- **互联网环境下的书目与规范控制。**书目控制和规范控制是图书馆的核心职能。基于字符串的规范控制只在图书馆领域内发挥作用。在开放环境下，需要转换到基于URI的身份管理。
- **基于内容的知识组织与知识融合。**应用语义技术实现内容相关而资源类型无关的知识组织与跨领域的知识融通

02 关联数据基础

章节目录

CONTENT

2.1 什么是关联数据

2.2 URI设计

2.3 RDF

2.4 本体设计

2.5 关联数据发布规范与建议

2.6 关联数据的最佳实践

2.1 什么是关联数据

采用RDF数据模型，利用URI（统一资源标识符）命名数据实体，发布消费数据的最佳实践。并且通过HTTP内容协商揭示并获取这些数据，同时强调数据的相互关联、相互联系以及有益于人机理解的语境信息。

URI

HTTP URI

RDF

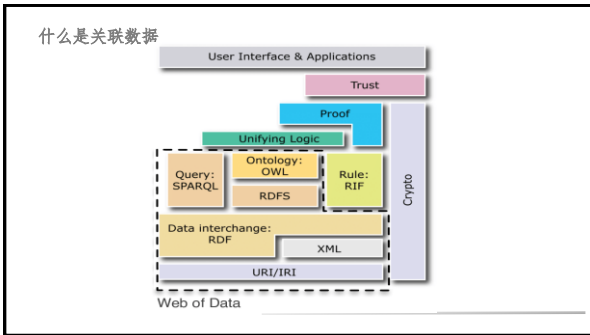
关系

原则1

原则2

原则3

原则4



什么是URI

- 什么是URI（统一资源标识符，Uniform resource identifier）

事物没有名称就可以被认为是不存在的——古希腊哲学家巴门尼德。

网络上的资源，其唯一且必需的属性是URI。

URI是一组按照特定语法规范构成的字符串，为网络资源的标识提供了一种简单和可扩展的方法。

统一：保证了不同类型的资源标识符可以采用统一的使用方式，并为新的资源类型提供统一的标识框架；也保证了这些不同类型的资源标识符在相同的语法规范下有统一的语义解释，并允许在不同的环境下重用。

资源：是指能够被标识的任何对象，不只限于网络文档和数字内容，也包括现实世界的物理实体和抽象概念。

标识符：一个对象可以引用的名称。

URI作用

● 标识和定位

URI的本质是一种标识符。所谓标识符，是指对象的名称标识，是特定背景下资源的明确指向。URI可以进一步地分为统一资源定位符（URL）、统一名称（URN）或两者的结合。ISBN号就只是一个URN。在语义网语境下，URI可被视为定位符，不仅可以标识一个资源，也提供该资源的定位信息。

● 访问与获取

在语义网语境下，URI同时作为统一名称和定位符而存在，它不仅命名了该资源，还提供了定位该资源的主要访问机制。如<http://www.ietf.org/rfc/rfc3986.txt>，不仅标识定位RFC3986这份文档资源，也可以通过HTTP协议进行访问并获取相关内容。

URI作用

● 参引

参引是指为了获取引用资源的相关信息，在万维网上查找URI的过程。可参引的URI提供了使用任何的网络传输协议获取被标识资源的表示形式的信息检索机制。如，传统网页检索中，对URL的参引可以直接返回HTML页面。而在关联数据环境中，对URI的参引可以返回HTML、XML或JSON等形式。因此，参引一个URI，将指向一个资源的描述，这个描述中可能会有更多的URI链接，同时这些URI也是可参引的，最终个体资源互相关联形成一个数据网。

URI作用

● 规范控制

传统的规范控制工作是运用信息资源标目法，确定人名、团体、地名、题名、主题等检索点及其规范形式，同时建立和维护规范文档的过程。

URI标识的实体概念将独立于其表示形式，各种语言或符号的表达都可以作为标签属性。任何形式的符号在检索和显示中都将具有相同的优先级。另外，URI可以标识任何对象，也就是说书目数据中的任何属性都是可控制的，传统意义上的规范数据将得到极大的丰富。在元数据生产过程中，规范控制也将无处不在。

URI设计原则

● 无变量

URI字符串中应避免使用文件扩展名、所有权信息、版本信息等变量内容，避免包含主题词、自动增量，避免与具体时间的技术捆绑。如

<http://example.com/people/v01/123456>、<http://example.com/people/123456.rdf>就不是“酷”的URI。

● 稳定性

研究发现学术出版物的URL在发布7—10年后，有一半会失效。图书馆作为人类文化知识的保存和传播中心，在发布数据时，应该为标识对象分配稳定长久的URI，以确保资源长久地可获取和重用。

URI设计原则

● 使用HTTP URI

URI语法规定其以表示资源访问协议的[`schema:`]开头，常见的有`http`、`ftp`、`mailto`、`file`等。依据关联数据四原则，选用HTTP作为URI的访问协议。HTTP作为主流的网络传输协议，保证了最广泛的工具和系统可以访问。任何资源均可由URI标识。因此，当客户端参引HTTP URI时，需要根据所标识对象的类型和表示形式支持303转向和内容协商机制。



URI设计原则

● 解读性

URI作为标识对象实体的一组字符串，在保障唯一性和稳定性前提下，需要简短易记、利于入读，同时尽量使用小写字母（本体中的类名除外），以便于管理和使用。

● 国际化

URI的使用是全网域的，因此为了便于URI的复用和解析，应使用UTF8编码，不使用特殊字符，如汉字、下划线等。

Hsah URI or Slash URI



URI设计模式

● 类型

根据本地的数据类型，URI分为6种类型：

数据集URI：用于特定资源的URI，如家谱数据集URI；

本体URI：家谱本体、书目本体、古籍本体、档案本体等；

规范词表URI：主要标识对象为主题、分类等；

取值词表URI：主要标识对象为责任方式、版本类型、装订方式等取值词汇；

非信息资源URI：即标识人、地、时、团体等实体对象的URI；

信息资源URI：则标识作品、实例、单件等对象。

URI设计模式

● 基本模式

http://{domain}/{type}/{concept}[type]/{IRI}

URI模式由域名、类型词（URI类型）、概念（本体中的类名）或类型词、内部唯一标识符四个部分组成。

(1) domain=Base URI+[dataset prefix] 当所发布的数据为诸如机构、地名、时间等通用的规范数据时，dataset prefix省略

表6 上海图书馆 URI 模式 domain 示例

类型	domain	备注
本体	http://www.library.sh.cn/ontology/ 或 http://ontology.library.sh.cn/	同一机构所有本体使用相同 domain
关联数据	http://data.library.sh.cn/	
数据集	http://data.library.sh.cn/jp/	家谱数据集
面向读者服务的网站	http://jp.library.sh.cn/ 或 http://jppsc.library.sh.cn/	

URI设计模式

● 基本模式

http://{domain}/{type}/{concept}[type]/{IRI}

表7 上海图书馆 URI 设计模式中的标识对象所属类型词(type)定义列表

标识对象类型词	说 明
ontology	本体
authority	规范词表, 如名称规范档、主题词、分类法等
entity	非信息资源: 人、地、机构等实体
resource	信息资源: 如图书、连续出版物等文献, 以及文档、图像、视频等数字化的资源对象。
vocab	取值词表, 如纂修方式、先祖类别、版本类型

URI设计模式

● 基本模式

http://{domain}/{type}/{concept}[type]/{IRI}

(3) concept或type. concept取本体中的类名, type为自定义的类型词。如, edition表示版本取值词表, person表示非信息资源的人, work表示信息资源的作品。

(4) IRI, 系统唯一内部标识符。可以服用原先数据的标识符。也可以重新分配。根据数据的不同, IRI分为两种类型, 一种是标识本体、规范数据和取值词表, 另一种是信息资源和实体。前者使用术语的英文或拼音形式, 后者使用16位随机符号, 由26个小写字母和1-9数字的混合。

URI设计模式

● 数据集URI模式

数据集URI主要标识对象是特定数据集, 如家谱数据集、名人档案数据集、书目数据集等。

URI模式为http://data.library.sh.cn/{dataset}/。dataset是数据集类型词汇, 如家谱数据集URI为http://data.library.sh.cn/jp/

● 本体词表URI模式

http://www.library.sh.cn/ontology/[Class][property]。

其中, Class是首字母大写的类名, 如
http://www.library.sh.cn/ontology/Person;
property是首字母小写的属性名, 如
http://www.library.sh.cn/ontology/name

URI设计模式

● 规范数据URI模式

http://{domain}/authority/{type}/{term}

如: http://data.library.sh.cn/authority/FDC/jing

● 取值词表URI模式

根据取值词表的适用范围,上海图书馆将取值词表分为通用取值词表和专用取值词表。前者如版本类型的取值词表包括“石印本”“木活字本”“铅印本”等,装订方式的取值词表包括“经折装”“蝴蝶装”“散叶”等;http://data.library.sh.cn/vocab/binding/jing-zhe-zhuang 后者如先祖类型为家谱专用取值词表,其取值词表包括“始祖”“始迁祖”等。http://data.library.sh.cn/jp/vocab/ancestor/shi-qian-zu

URI设计模式

● 信息资源URI模式

http://{domain}/resource/{concept}/{IRI}。

Concept表示资源类型,如work表示作品,item表示单件;

IRI为16位字母和数字随机组合的内部唯一标识符,如

http://data.library.sh.cn/jp/resource/work/ullep7uwac4n8cbt

● 非信息资源URI模式

URI模式与信息资源URI相同,如

http://data.library.sh.cn/entity/person/mwelui6bww62a417

RDF (Resource Description Framework, 资源描述框架)

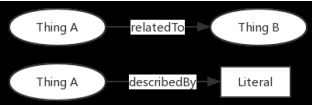
● 什么是RDF

关联数据是实现数据融合,底层的数据模型要保持一致。这个就是RDF。

RDF是一种用于识别或描述网络上的事物(也称为实体)关系或特征的数据模型。

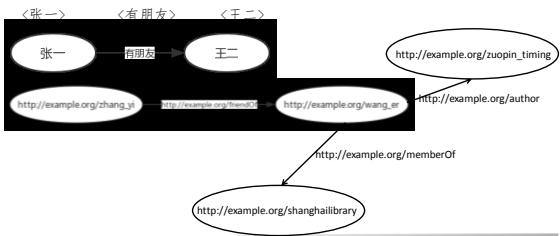
RDF语句,也称为RDF三元组,由主语、谓语和宾语组成。主语和宾语指代两个独立的事物,谓语描述主语和宾语之间的关系。它的格式是这样的:

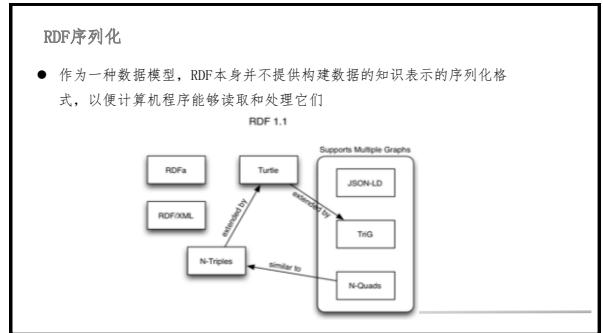
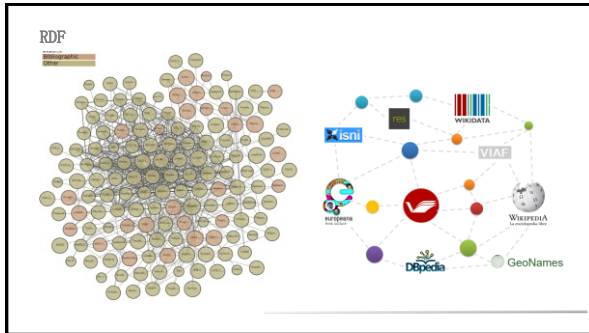
<subject> <predicate> <object>



RDF (Resource Description Framework, 资源描述框架)

张一有个朋友叫王二





本体（轻量级领域本体）

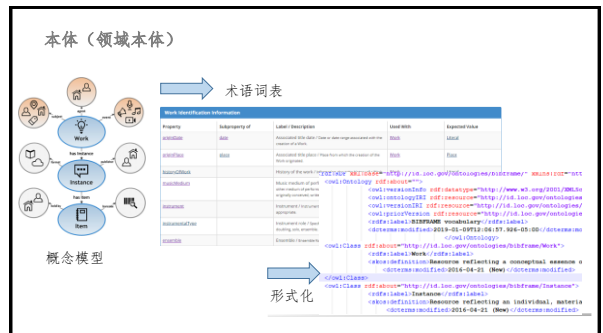
对领域知识共享的、形式化的概念模型, 提供了描述领域知识的结构化术语词表。

概念模型：是指允许描述世界上相关事物的抽象模型，通常用概念实体、属性和关系表示。

形式化:是指本体是机器可读,也就是说用RDFS、SKOS、OWL语言进行编码。

术语词表: 定义了领域内存在着的所有对象类型或概念及其属性和相互关系的专业术语。

共享的：领域内被一致认同的知识，不是私有的个人化的知识。



本体使用-术语词表

● 类

注意点：定义：类间关系。

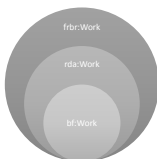
上位类包含下位类，下位类可继承用于上位类的属性。

● 属性

注意点：定义域（domain，主语）、值域（range，宾语）、属性间关系。

定义域决定了属性可以用于描述哪些实体。值域决定了属性值时另一个类还是字符串。

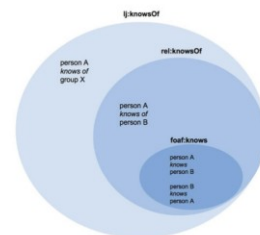
上位属性可描述的实体类（定义域）或属性值（值域）比下位属性更广泛。即继承关系。



本体使用-术语词表

Property:	provisionActivity
Label:	Provision activity
Definition:	Place, name, and/or date information
Used with:	Instance
Expected Value:	ProvisionActivity

Property:	code
Label:	Code
Definition:	String of characters that serves as a code reference
Used with:	Unspecified
Expected Value:	Literal



本体设计-基本原则

● 基本原则：复用+自定义

不要重复发明轮子。复用已有的成熟的、被大量使用的词表。只要合适，不同来源的词表都可以混合复用。

好处：简化建模流程、增加数据互操作性；扩充语义描述的精度与粒度。

复用标准：稳定的URI；可在网络上获得；由一个可识别和可信任的机构发布得；使用发布的最佳实践；遵循标准格式及遵守版权控制政策数据和文档；

复用考虑因素：原数据集类型、元数据格式、项目目标等。

然后，不足之处或更细粒度的描述再考虑自定义术语。

自定义术语的最佳实践：复用特定领域的非本体论资源中的术语，如规范词表、词典等规范性、标准化的行业文献；一般作为已有术语的父或子术语。

本体设计-词表选择

● 词表查询

- (1) OCLC关联数据调查
- (2) Linked Open Vocabularies (LOV)
- (3) 上海图书馆本体服务中心

● 常用的书目模型

BF、Schema.org、DC系列、ISBD、FRBR、RDA等

● 其他会用到的词表

FOAF、Rel、GEO、ORG、SKOS、MADS、RDFS、RDF、PROV等

RDF Vocabularies/Ontologies Used	2018	2015
Schema.org	44 40%	30 30%
SKOS	42 44%	59 90%
Dublin Core Terms	36 41%	51 52%
FOAF	36 38%	55 56%
DCMI Metadata Terms	35 37%	49 49%
RDF Schema	35 37%	45 45%
BibFrame	26 27%	15 15%
Local vocabulary	21 22%	19 19%

[illegible][illegible]

Individual and Group Entities

mo:MusicGroup

mo:Label

Organization

mo:MusicSoloArtist

Solo Music Artist

foaf:Image

bf:Text

schema:AudioObject

Audio

Image

Text

Video

Release

mo:Instrument

schema:VideoObject

mo:Release

Spatial Entities

mo:CorporateBody

Record Label

Place

Venue

schema:EventVenue

schema:MusicVenue

Event

Session

event:Event

schema:Event

mo:RecordingSession

Temporal Entities

Work [Song]

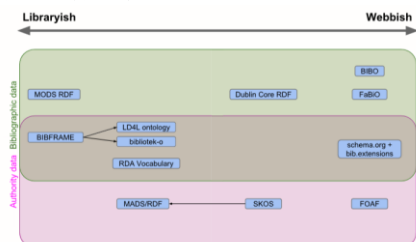
bf:Work

schema:CreativeWork

[illegible]

本体设计-书目词表复用

主要应用
环境，图
书馆还是
网络？



本体设计-词表自定义

● 类 or 属性???

如果是新的描述对象，对领域重要的信息，需要对其进行更细粒度的特征或关系描述，就定义为类。如果是对已有对象新的特征或关系描述，则定义为属性。

● 类

自定义类与已有类的上下位关系，或独立的类，或等同关系。

是子类，则复用已有上位类的所有属性；父类，则不可以复用，而且自定义的属性可以被已有的子类复用。

● 属性

- (1) 数据类型属性，对象类型属性；
- (2) 定义域与值域；
- (3) 与已有属性的上下位关系，或独立的属性，或等同关系。

本体设计-词表自定义

● 术语命名

- 类以大写字母开头，并且总是单数
- 属性以小写字母开头
- 如果一个术语有多个单词，可以使用camelCase
- 数据类型属性是名词/名词短语
- 对象属性是动词/动词短语
- 在命名时考虑逆反属性

本体设计-词表自定义

● 案例一：古籍文献中的藏书印怎么描述？类还是属性???

藏书印有形状、刻法、字体、引文等信息，要不要记录？

● 方案一

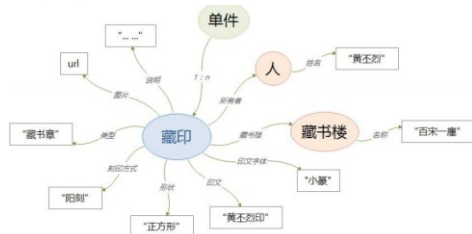
项目需求只需要记录某本书有没有藏书印即可。以后也不准备对藏书印进行更多说明。就简单化，新增属性即可。

ex:Instance--ex:hasSeal--Literal

● 方案二

不仅需要记录有没有印章，还需要描述印章。就定义为类，并定义相关的属性。

本体设计-词表自定义



本体设计-词表自定义

● 案例二：

复用表示人际关系的属性 `rel:knowsOf` 和 `rel:influencedBy`
两个属性的值域与定义域都是 `foaf:Person`。

但实际情况是，一个人不是只能认识人或被人影响，Ta也可以被某个团体影响，或知道某个团体。

● 方案：

自定义属性的连接的不仅仅是 `foaf:Person`，更合理的是其上位类 `foaf:Agent`。因此自定义的属性是 `rel:knowsOf` 和 `rel:influencedBy` 的上位属性。

本体设计-词表自定义

```
ex:knowsOf
  rdfs:domain foaf:Person
  rdfs:range foaf:Agent
rel:knowsOf rdfs:subPropertyOf ex:knowsOf
```

```
ex:influencedBy
  rdfs:domain foaf:Person
  rdfs:range foaf:Agent
rel:influencedBy rdfs:subPropertyOf ex:influencedBy
```

本体设计-注意事项

● 弗兰肯斯坦本体 (Frankenstein ontologies)

科学怪人弗兰肯斯坦，是用不同尸体的各个部分拼凑成一个巨大人体。

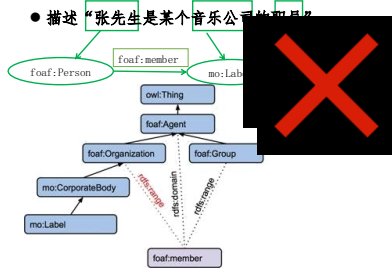
复用与自定义的术语之间一定要兼容。

复用时检查术语语义是否被正确的保留，相近或相同的术语之间语义是否冲突。

特别是复用的是强语义的重量级本体，它对术语的定义都有严格约束与规定。比如属性的定义域、值域；类的相交或不相交。

本体设计-注意事项

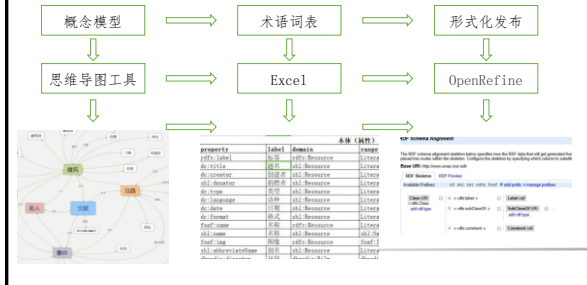
- 描述“张先生是某个音乐公司的成员”



本体设计流程

1. 实体关系分析。分析原始的数据结构，根据项目需求提取其中的实体及其特征与关系。
2. 模型词表调研。领域内有哪些合适的词表。
3. 复用词表。注意属性的domain与range，复用时不要改变原始的外延与内涵。
4. 词表设计。类、属性的定义。关键是属性的domain与range，一个类可以同时继承多个类。
5. 形式化发布。以机器可读的编码，并在网上发布。

本体设计工具



本体发布

属性

```
ex:xxx a owl:ObjectProperty/owl:DatatypeProperty ;
    rdfs:label "XXXX" ;
    rdfs:comment "XXXX" ;
    rdfs:domain ex:xxx ;
    rdfs:range ex:xxx/Literal ;
    skos:definition "XXXX" ;
    owl:inverseOf ex:xxx ;
    rdfs:subPropertyOf ex:xxx ;
    dcterms:modified "XXXX" .
```

本体发布

● 类

```
ex:XXX a owl:Class ;
    rdfs:label " XXXX" ;
    rdfs:comment " XXXX" ;
    skos:definition " XXXX" ;
    rdfs:subClassOf ex:XXY ;
    dcterms:modified " XXXX" .
```

本体设计-总结

复用
RDFS描述的轻量级本体
遵循LD四原则 (六星标准)

本体与元数据

	元数据	本体
相同点	对信息的结构化描述	
不同点	定义及构成 “关于数据的数据。” 元数据方案（元素集）、置 标方案、元数据记录	“领域的可共享的概念及概念间 关系的形式化定义。” 模型、词表（类、属性）、实体/ 实例对象
描述对象	文献	内容
描述重点	文献的外在特征	内容对象的特征及其相关关系
词表定义	元素	类、属性
标准规范	MARC DC	FRBR BIBFRAME FOAF schema.org
编码	RDF XML	RDFs OWL
存储与查询	RDB SQL	RDF Store Graph DB SPARQL
联系	元数据元素可作为本体的属性，本体可看做“元”元数据。	

本体与元数据映射

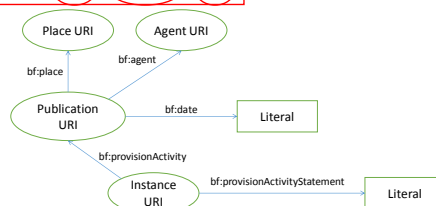
- 元数据是对某个潜在信息性对象做出的陈述。一个元数据陈述就是含有主谓宾的三元组陈述。在元数据模式中，谓词常被称为元素，它通过元素-值配对构成关于某个资源的唯一陈述
- MARC元数据是对整个文献资源中各种信息对象的陈述集合。MARC格式中，字段子字段就是谓词元素。分解到每一个字段或子字段及其取值，就是对多个或单个信息对象做出的唯一陈述。
- 本体与MARC元数据之间的映射，就是把“记录-字段-子字段-值”的结构解析为“类-属性-值”结构。

本体与元数据映射

- 是否映射
 - 很少使用的MARC字段可以考虑不映射
 - 在MARC记录外没有意义的字段不映射
- 映射到哪一层？Work, Instance, Item?
- 映射关系
 - 一对一
 - 一对多
 - 多对一。取值优先级，或者都保留？
 - 无关系。新增术语、舍弃？

本体与元数据映射

=210 \\\$北京\$科学出版社\$2012



关联开放数据七星标准

- ★ 以任何个和发布到网上，任何人可以查看、查找、存储、改变
- ★★ 数据以结构化的方式访问。如excel，而非扫描图片
- ★★★ 开放的、非私有的格式，如CSV
- ★★★★ 使用URI标识对象
- ★★★★★ 链接到更多的数据
- ★★★★★★ 使用的词表与数据集一起发布
- ★★★★★★★ 发布词表在数据集中的应用情况



关联数据发布的最佳实践

- 步骤1，让利益相关者做好准备。通过解释创建和维护关联开放数据的过程让利益相关者做好准备。
- 步骤2，选择高价值的数据集。特藏资源。理想情况下，当这些信息与其他开放数据结合在一起时，可以提供更大的价值。
- 步骤3，建立数据模型。建立关联数据模型描述数据集中的对象及其关系。
- 步骤4，指定适当的开放许可协议。如果对已发布数据的来源、所有权和使用相关条款有明确的说明，则更有可能发生数据重用。图情领域一般使用CC协议（知识共享，Creative Commons）。如，cc2.0协议（署名-非商业性使用-相同方式共享，BY-NC-SA）。
- 步骤5，Cool URI。

关联数据发布的最佳实践

- 步骤6, 使用标准词汇表。尽可能使用先前定义的词汇表。在必要时进行扩展, 创建时尽可能遵循最佳实践。
- 步骤7, 数据转换。将数据转换为关联数据表示形式。RdFa、JSON-LD、Turtle、RdF/XML。
- 步骤8, 提供对数据的机器访问。为搜索引擎和其他应用程序提供使用标准的 Web机制访问数据的各种方法。
- 步骤9, 向公众发布。在一个权威的数据注册网站发布数据集。
- 步骤10, 社会责任。数据一旦发布, 你在维护数据方面的责任。确保数据集能够保持在组织所说的位置, 并且得到持续的维护。

关联数据发布流程



03 BIBFRAME书目模型

BF诞生前的重要论点

通用性、可扩展性、开放性、透明性、低门槛、高上限、协作管理、模块化、层次结构、粒度和优雅失败搜索处理

基于xml的传输模式

Roy Tennant 2004

数据链接池, 图书馆应该准备目录和/或 ILS(综合图书馆系统)内外的链接

强调互操作性和数据共享

Karen Calhoun 2006

书目控制比传统的图书编目领域更为广泛: 书目宇宙需要更多的参与者。

制定实现FRBR的计划: 强调了规范控制的价值; 重视跨社区管理作品、名称、概念和对象之间的关系

On the Record 2008

创建了她自己的RdF编目规则模型, 将FRBR的实体、属性以及与RdF的主语、宾语和谓词的关系等概念关联起来

Martha Yee 2009

用一个新的书目框架来取代 MARC, 这个框架将是 “一个环境, 而不是一个格式”, 对编目规则不可知”

Bibliographic Framework Initiative 2010

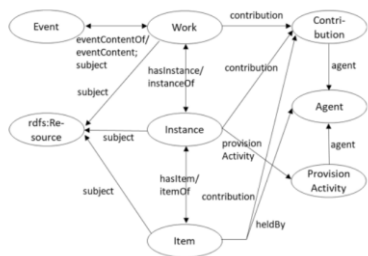
BIBFRAME为什么不采用FRBR的模型

- 严格遵守FRBR的作品与内容表达的层次结构可能会有问题
- 作品概念在不同的社区有不同的含义
- FRBR对于视听资源来说并不是最优模型，因为作品与内容表达是不容易分离的
- FRBR对必须在期刊和文章级进行描述的连续出版物也不是最优模型

BIBFRAME1.0-2.0发展

- 属性合并，减少冗余
- 取消Authority类，调整取消相关属性
 - LD环境下，基于字符串的规范控制是不合时宜的。
 - URI标识任何实体，全局唯一的；
 - 真实世界的事物与描述它的文档是两个不同的对象。
- 取消Annotation类
 - 第一种是传统的注释视角，但确不需要作为核心类存在。
 - 第二种，Item对图书馆馆藏管理是极其重要的实体。不应该作为子类。并且其他本体词表都定义为核心类。

BIBFRAME2.0核心模型

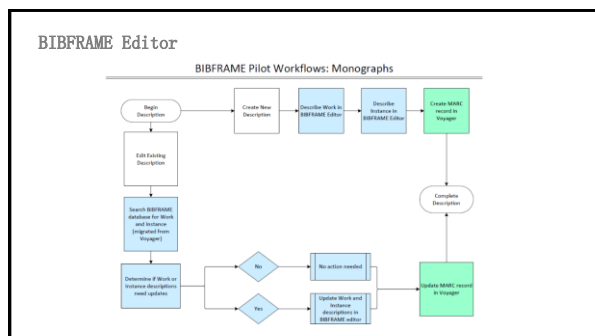


BIBFRAME2.0属性

- 定义域限定是Work、Instance、Item中的某一个；
- 定义域是Work、Instance、Item中的至少两个；
- 不声明定义域，可用于任何类；
- “Used with” “Expected value” 的描述性语句表示属性的定义域和值域，并不是要强制性的。

Property:	IllustrativeContent
URI:	http://id.loc.gov/vocabulary/bibliframe/illustrativeContent
Label:	Illustrative content information
Definition:	Information about content intended to illustrate a resource.
Used with:	Used with Work or Instance
Expected Value:	Illustration
Change Notes:	2016-04-21 (New)

Property:	HasDerivative
URI:	http://id.loc.gov/vocabulary/bibliframe/hasDerivative
Label:	Has derivative
Definition:	Resource that is a modification of the described work.
Used with:	Used with Work or Instance
Expected value:	Expected value Work or Instance
SubProperty Of:	relatedTo
SubProperties:	translation derivation adaptation
Inverse Of:	derivativeOf
Change Notes:	2016-04-21 (New) 2016-04-29 (skold inverse)

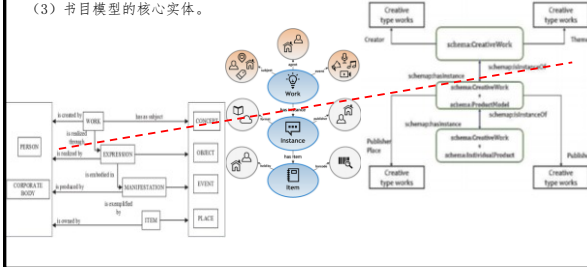


项目的前置步骤

- **项目目标**
为什么这样做？想要达成的目标？
- **数据集选择**
高价值数据。小规模的特色资源。
- **数据分析**

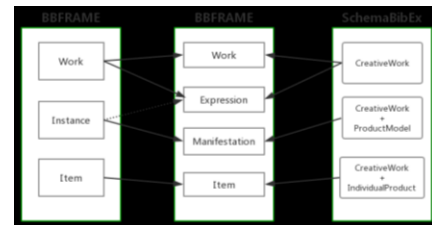
作品识别的重要性

(3) 书目模型的核心实体。



作品识别的重要性

(3) 书目模型的核心实体。



作品识别的核心元素

- 文本的同一性是作品识别的基础原则。

书目世界定义的作品抽象模糊,实践层面需要落实到一个具体的对象上。作品只有落实到具体的文本上,才能为读者所感知。读者到图书馆是寻找含有最适合自己内容的那本书,也就是威尔逊所说的“最好的文本”。

Shoichi Taniguchi, Pisanski J 与 Žumer M, Tanselle, Alves Mey, Leazer 与 Furner 等人同样认为的作品识别最好的方法就是文本内容, 文本的任何改变都会导致一部新作品的诞生。

作品文本是一组具有特定语义和概念的字符串产生的有序集合。

作品识别的核心元素

- 文本是作者思想构造的具现化，作品的识别就不得不说到作者。

借鉴了著作权法的作者认定以及出版界的作者署名做法。对文本的选择和组织也是一种创造性的活动,授予该人作者身份是完全恰当的。因此,作品识别中的作者不仅限于主要责任人,还包括了编者、译者、注释者、插图者等任何对作品内容做出贡献的人。

- 语种也需要参与作品识别。

对单一语言环境下的读者来说，对阅读有一种语言偏好。作品只有被翻译之后才真正地存在。翻译也是在“重新创造文本”。

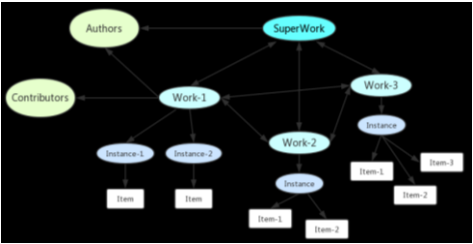
作品识别的核心元素

综上所述，书目数据中作品识别的基本元素包括了指称作品的正题名、分卷号和分卷题名，对作品文本内容负责的所有责任者，以及作品的语种。

作品-超级作品模型

- 作品
即BF作品，是广义上的创作者独创性的思想、情感和观点的表现形式，它通过字母-数字、音乐或舞蹈标记、声音、图像、实物、运动等内容表达方式具现化。近似于FRBR作品与内容表达。
- 超级作品
作品模型中最顶层的抽象实体，作为 bf:Work 的子类，剥离了作品的文本 内容特征，实现了对不同内容表达的作品聚合。近似于 FRBR 作品。
- 关系
bf:relatedTo或其子属性进行连接

作品-超级作品模型



CNMAC数据中作品识别-字段分析

● 本地数据的字段分析

抽取上海联编 2016 年 142793 条 BKAA《中文普通图书》数据为样本，进行分析。
表 1 BKAA 数据字段使用百分比

百分比 (%)	字段
99-100	LDR, 001, 005, 010, 039, 100, 101, 102, 105, 106, 200, 210, 215, 690, 801
60-99	330, 606, 701, 905
10-60	035, 205, 225, 300, 461, 510, 517, 702
1-10	049, 304, 305, 306, 307, 312, 314, 320, 327, 462, 540, 600, 601, 605, 607, 610, 711, 712
0-1*	011, 016, 091, 301, 308, 423, 512, 516, 692, 696, 830, 910, 961
0*	009, 020, 092, 093, 094, 110, 120, 135, 206, 208, 230, 303, 310, 311, 333, 334, 337, 393, 410, 500, 501, 515, 518, 541, 609, 639, 660, 686, 693, 728, 730, 856, 890, 902, 978, 998, 999

注：0-1 表示几百条数据使用此类字段，0 表示少于一百条数据使用此类字段。

CNMARC数据中作品识别-作品键值

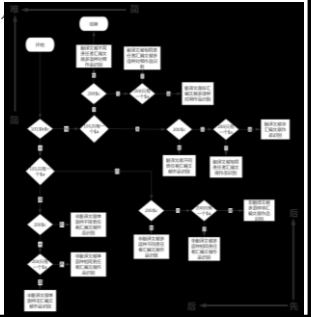
● 作品键值

bf:Work= language+200\$ahi+7XX
ex: SuperWork=200\$a+7X1

- 另外，在作品匹配时需要删除题名字符串中的符号，如空格、引号、顿号、间隔号等。
- 而 7XX 责任者字段，则可以先 进行规范控制处理，作品识别时直接使用其规范控制号即可

CNMARC数据中作品识别-数据分

在实践中，可以根据101\$abc判断是否是翻译作品，是否是多语种对照，根据200\$ac判断是否是汇编文献。依此对CNMARC 数据进行分包处理，由简入难，依次对每类数据制定相应的作品识别方案。



CNMARC数据中作品识别-单语种单一作品识别示例

(1) 当101没有\$bc，并且101\$a只有一个，200只有一个\$a子字段。

原始 CNMARC 数据		作品识别	
题名	责任者	bf:Work	ex:SuperWork[a]
三国演义 白话插图本	罗贯中原著,夏清影译	W1	SW1
三国演义	罗贯中著	W2	SW1
三国演义 100% 大长篇小说	罗贯中原著,陈建东编	W3	SW1
三国演义 100% 大长篇小说	罗贯中原著,陈建东编	W4	SW1



CNMARC数据中作品识别-单语种同责任者多个作品识别示例

(2) 当101没有\$bc，并且101\$a只有一个，200有多个\$a子字段，并且没有\$c。

原始 CNMARC 数据		作品识别		
题名	责任者	bf:Work	ex:SuperWork[a]	ex:SuperWork[b]
陶庵梦忆	张岱	W1	SW1	
陶庵梦忆	张岱著;夏清影译注	W2	SW1	
陶庵梦忆·西湖梦寻	张岱著;胡志泉注	W3	SW2	
陶庵梦忆 西湖梦寻	(明)张岱著	W1、W4	SW1、SW3	SW2
陶庵梦忆 西湖梦寻	(明)张岱著;韩柯柯校注	W5、W6	SW1、SW3	SW2

SW1=“陶庵梦忆+张岱”，SW2=“陶庵梦忆西湖梦寻+张岱”或“西湖梦寻陶庵梦忆+张岱”，SW3=“西湖梦寻+张岱”

CNMARC数据中作品识别-作品识别示例



CNMARC数据中作品识别-翻译作品识别示例

(3) 101有\$c或\$b

在关联数据环境下,作品实体通过URI标识,不再有首选题名的概念,不同语种题名或变异题名都是作品实体可选的题名标签。因此,即使在CNMARC数据中识别的原语种作品的题名是中文形式,在后续的数据关联或增强中,也可以通过DBpedia、VIAF、WikiData等多语种数据集实现跨语种的数据融合。如《鲁滨逊漂流》,识别了英文作品Work1,其题名为中文“鲁滨逊漂流”。然后通过中文题名形式匹配到WikiData的Robinson Crusoe (Q483034),从而获得多语种题名形式。当然,也可以通过510字段\$z与101\$c或\$b判断,如果两个相同,则原语种作品的题名可以从510字段获取。

CNMARC数据中作品识别-翻译作品识别示例

翻译类文献的作品识别在题名与责任者的基础上,再加上语种。即,每一个101\$a或\$b或\$c就需要识别出一个BF作品。具体识别规则如下:

- (1) bf:Work[$\$a$]=101 $\$a$ [1]+200 $\$ahi$ +7XX; bf:Work[$\$a$][2]=101 $\$a$ [1]+200 $\$ahi$ +7X1
(2) 如果510 $\$z$ =101 $\$b$, 则bf:Work[$\b]=101 $\$b$ +510 $\$ahi$ +7X1; 否则
bf:Work[$\$b$]=101 $\$b$ +200 $\$ahi$ +7X1;
(3) 如果510 $\$z$ =101 $\$c$, 则bf:Work[$\c]=101 $\$c$ +510 $\$ahi$ +7X1; 否则
bf:Work[$\$c$]=101 $\$c$ +200 $\$ahi$ +7X1;
(4) 如果101有多个 $\$a$, 除第一个 $\$a$ 之外的所有 $\$a$ 分别与 $\$b$ 、 $\$c$ 进行对比, 如果相同,
则不识别; 如果没有则识别为新的 $\$b$ 。bf:Work[$\$a-n$]=101 $\$a$ [$a-n$]+200 $\$ahi$ +7XX;
(5) ex:SuperWork[a]=200 $\$a$ +510 $\$a$ +7X1.

CNMARC数据中作品识别-翻译作品识别示例

原始 COMARC 数据	作品识别		
题名 / 责任者	bfWork	bfWork	ex:SuperWork(p)
①001 \$aSchongk ②005a 卡斯特勒市长(城)瓦托斯·哈代第3g 基 雷瓦, 沈正明, 刘德刚译	W1	W2, W3	SW1
①001 \$aSchongk ②005a 卡斯特勒市长(城)瓦托斯·哈代第3g 纪 飞等编译	W4, W3	W2	SW1
①105a \$aMoor of Castreidra \$ong			

W1=“chi+卡斯特桥市长+托马斯·哈代+郭国良+沈正明+刘澹娟”；
W2=“chi+卡斯特桥市长+托马斯·哈代”；W3=“eng+卡斯特桥市长+托马斯·哈代”；W4=“chi+卡斯特桥市长+托马斯·哈代+纪飞”；
SW1=“卡斯特桥市长+托马斯·哈代”，或“Mayor of Casterbridge+托马斯·哈代”。

CNMARC数据中作品识别-翻译作品识别示例



CNMARC数据中作品识别-翻译作品识别示例

源作品CNMARC数据	作品识别				W21=“chi+动物园+什克洛夫斯基+赵晓彬+郑艳红”;
题名/责任者	010001	010002	010003	010004	W31=“chi+第三工厂+什克洛夫斯基+赵晓彬+郑艳红”;
0101 1/zhongguo	W1	W2, W3	SW1		W22=“rus+动物园+什克洛夫斯基”;
0101 2/zhongguo	W4, W5	W2	SW1		W32=“rus + 第三工厂+什克洛夫斯基”;
0101 3/zhongguo	W6, W7	W2	SW1		SW2=“动物园+什克洛夫斯基”;
0101 4/zhongguo	W8, W9	W2	SW1		SW3=“第三工厂+什克洛夫斯基”;
0101 5/zhongguo	W10, W11	W2	SW1		SW4=“动物园第三工厂+什克洛夫斯基”

CNMARC数据中作品识别-翻译作品识别示例



05 关联书目数据发布

章节目录
CONTENT

4.1 关联书目数据发布流程

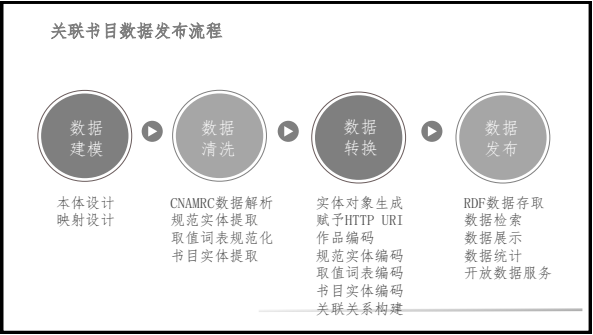
4.2 数据分析

4.3 数据建模

4.4 数据清洗

4.5 数据转换

4.6 数据发布



数据建模

● 数据建模

为RDF数据定义主语所对应的类、谓语所对应的属性、宾语的取值约束，并用形式化的术语词表规范表达，形成书目本体。包括了本体设计与映射设计。

● 本体设计

模型设计是通过分析书目数据中的实体关系，来提炼抽象概念，定义概念间的关系。

词表设计是用规范的术语词表来表达模型中概念、概念特征和概念间的关系。概念用类来表示，概念特征与关系用属性表示。

● 映射设计

数据建模-本体设计

● 基本原则：复用+自定义

不要重复发明轮子。

简化建模流程、增加数据互操作性。

根据数据集类型、原元数据格式、项目目标等复用已有的本体词表作为核心模型的基础。

只要合适，不同来源的词表都可以混合复用。

然后，不足之处再自定义类和属性满足本地特殊需求。

数据清洗-规范实体

- 人名（上海图书馆人名规范库）
 - 将国图规范档中上海 联编中心书目数据中的人名融合到上图人名规范库，使上海联编中心书目数据的责任者可直接参引人名规范库中的人物实体
- 地点（上海图书馆开放数据平台）
 - 地名可直接参引上图地理名词表
- 团体（团体责任者、出版发行者、会议）
 - 机构团体名称、会议名称为上海联编中心书目数据所独有，单独发布

数据清洗-取值词表

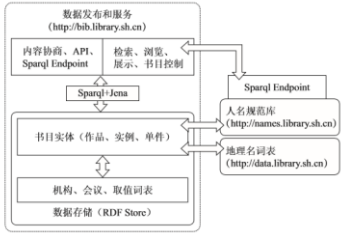
- 发行方式
- 读者对象
- 文献类型
- 图表类型
- 语种
- 责任方式

http://lib.library.sh.cn/vocab/mediatypelabel	http://www.w3.org/1999/02/22-rdf-syntax-ns#rdfs	http://www.w3.org/2004/02/07/owl/core#rdfs
http://lib.library.sh.cn/vocab/mediatypelabel	http://www.w3.org/1999/02/22-rdf-syntax-ns#rdfs	http://id.loc.org/ontologies/bibframe/ExtendedDistance
http://lib.library.sh.cn/vocab/mediatypelabel	http://www.w3.org/2000/01/rdf-schema#label	xbl:*, general
http://lib.library.sh.cn/vocab/mediatypelabel	http://www.w3.org/2000/01/rdf-schema#label	普通书人
http://lib.library.sh.cn/vocab/mediatypelabel	http://www.w3.org/2000/01/rdf-schema#rdfs:label	http://id.loc.org/vocabulary/mediatypelabel

数据清洗-书目实体

- 作品
 - 分成：普通图书、汇编文献、年鉴和集刊、丛书
 - 普通图书：是否是翻译作品，分开处理
 - 汇编作品：同一责任者不同责任者分开处理
 - 丛书作品，只有著录丛编责任者的才发布
 - 年鉴和集刊，在前期编目工作基础上，作为超级作品发布，聚合相关作品
- 实例
 - 一条MARC数据一个实例
- 单件
 - 上海联编数据没有馆藏信息，主要属性就一个索书号。

数据发布与服务



战略规划与行动计划

- 在内容标准、数据格式、编码形式；在数据的生产、存储、显示、检索、消费等图书馆数据生态环境酝酿变革的时期，在图书馆传统编目日渐式微的当下，我们需要结合图书馆的战略规划（“十三五”规划），制定元数据部门的战略规划与行动计划，促进编目范式的转变。为新环境下编目工作指明方向、制定标准、开发工具、提供培训等。形成最佳实践，重塑图书馆编目部门的职责与价值。



PCC战略规划

随着制定新的标准和机制，使图书馆数据与关联数据兼容，PCC认识到自己在促进整个PCC机构对语义查询和数据结构的共同理解方面可以发挥独特的作用。

由于许多技术环境仍在不断变化，PCC将采用迭代方法进行学习和决策。

- 2015-2017 【偏理论】
 - SD1：关联数据培训课程；SD2：与社区外数据实践保持一致；SD3：在规范控制从主要基于创建文本串的方法，转换到专注于管理身份与实体的方法中，提供领导
- 2018-2021 【转向实践】
 - SD3：应用对关联数据的理解；SD4：加速在网站层面向无处不在的标识符创建和身份管理的转变；SD5：扩展使用关联数据取值词表，以增强并在可行的情况下替换现有元数据实践

PCC任务组

- Work Entity
- URI in MARC
- BIBFRAME
- FRBR-LRM
- Linked Data Advisory
- Identity Management in NACO
- Linked Data Training

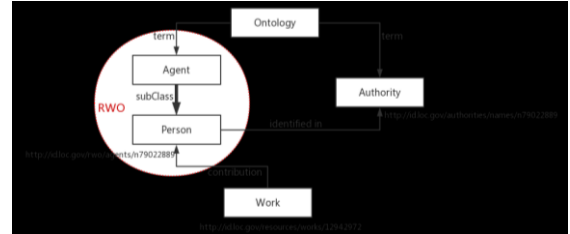
PCC-URI in MARC

- 好处：
 - URI有助于去重、去歧，并且通常使规范的维护更加容易，这是文本字符串所不能做到的。
 - URI对于从MARC到RDF中描述的实体的转换也是必要的。
- 注意点：
 - Real World Object (RWO) vs Authorities
 - <http://id.loc.gov/authorities/names/n79022889>
 - <http://id.loc.gov/rwo/agents/n79022889>

PCC-URI in MARC

- 好处:
 - URI有助于去重、去歧, 并且通常使规范的维护更加容易, 这是文本字符串所不能做到的。
 - URI对于从MARC到RDF中描述的实体的转换也是必要的。
- 注意点:
 - Real World Object (RWO) vs Authorities
 - <http://id.loc.gov/authorities/names/n79022889>
 - <http://id.loc.gov/rwo/agents/n79022889> ← 描述

PCC-URI in MARC



PCC-URI in MARC21

- \$0 Authority record control number
- \$1 RWO URI
- \$4 关系词表
- 字段
 - 代理 (人、团体、会议): 1XX, 7XX等
 - 主题: 6XX
 - 取值词表: 责任方式、语种、类型等
 - 作品: 758

PCC-URI in MARC21样例

- 人:
 - 100 1# Charpentier, Marc-Antoine, \$d 1643-1704, \$e composer.
 - 100 1# Charpentier, Marc-Antoine, \$d 1643-1704, \$0
 - <http://id.loc.gov/authorities/names/n79135306> \$e composer. \$4
 - <http://id.loc.gov/vocabulary/relators/cmp>
- 作品
 - 700 12 \$a Wilson, August. \$t Fences.
 - 700 12 \$i is container of (work): \$a Wilson, August. \$t Fences.
 - 700 12 \$4 <http://rdaregistry.info/Elements/w/P10147> \$i Is container of (work): \$a Wilson, August. \$t Fences. \$1
 - <http://viaf.org/viaf/198566585>

PCC-URI in MARC21样例

- 作品：
 - 758 ## \$1 <http://worldcat.org/entity/work/id/806896805>
- 主题
 - 655 #7 \$a Academic theses \$0
<http://id.loc.gov/authorities/genreForms/gf2014026039> \$2 lcgft
- 规范数据
 - 024
 - 7#\$a500010879\$0<http://vocab.getty.edu/ulan/500010879>\$1[http://vocab.getty.edu/ulan/500010879-agent\\$2](http://vocab.getty.edu/ulan/500010879-agent$2)gettyulan

URI in CNMARC书目数据现阶段建议

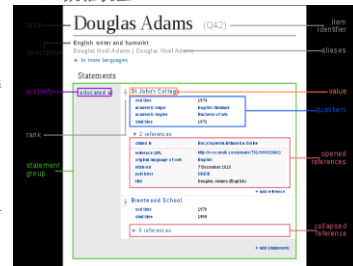
- 用取值词表：责任方式等。与RDA、LC取值词表对照。
- 责任者字段用IRI标识
- 使用统一题名，500
- 作品间关系，4XX

OCLC Project Passage

- 基于MediaWiki和Wikibase的关联数据编目工具。试验在传统编目流程中开发面向普通编目员的编目工具。
- Wikibase是用于存储和管理结构化数据的MediaWiki扩展。MediaWiki一个免费的开源wiki软件包。
- Wikibase是不同于传统图书馆编目工具：成熟、开放、众包、透明。
 - 支持人类和机器的输入，页面编辑历史；
 - 编辑讨论；
 - 无需专业技术即可在用户友好的图形界面中编目；
 - 软件功能、数据模型可定制扩展。
- 其他优势：支持完成RDF数据发布流程、多语言、用于读写数据的API，管理用户权限、非结构化文本文档索引、整合Wikidata等

OCLC Project Passage-Wikibase数据模型

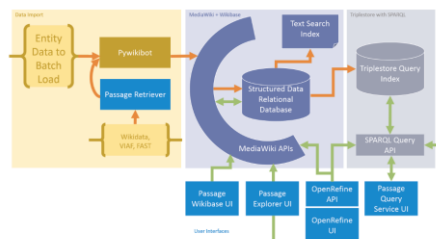
- Item：所说的实体（Thing）。人、概念、地点等等任何事物；
- fingerprint: label（首选名称形式）与description构成唯一字符串识别；aliases，其他语种的名称形式。
- qualifier：对属性的限定说明。
- Rank：同一个声明，检索返回结果的优先级，共3个级别，默认显示最高的。



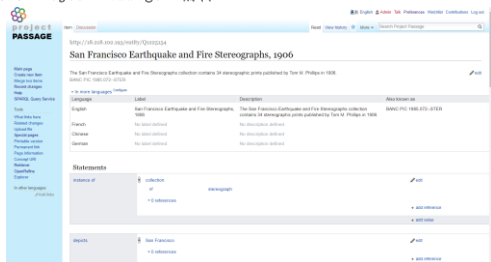
OCLC Project Passage—创建数据

- 一般流程：
 - 声明一个新实体；
 - 添加指纹数据；
 - 添加一个或多个结构化语句，必要时使用属性限定符进行修改；
 - 引用一个声明来提高它的排名。
- 新增属性：如果需要一个以前没有定义的属性，授权的编辑器可以修改 Wikidata 的本体。当这个修改得到更多人的支持，就会由管理者正式采纳发布。

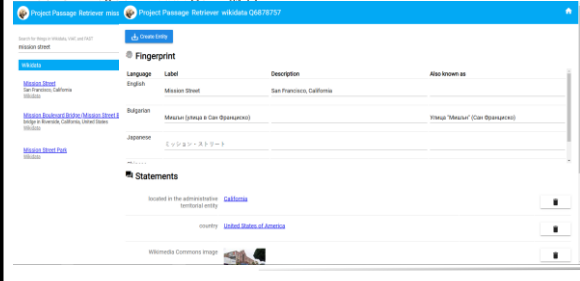
OCLC Project Passage—系统框架



OCLC Project Passage—编辑



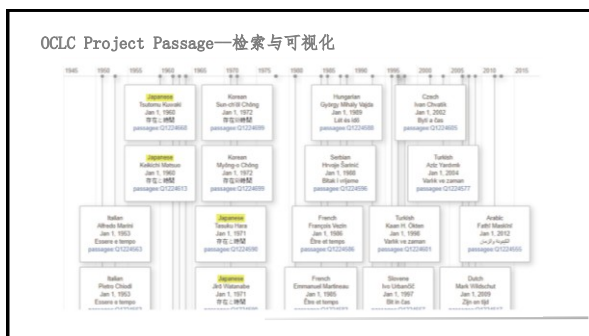
OCLC Project Passage—编辑



OCLC Project Passage—检索与可视化



OCLC Project Passage—检索与可视化



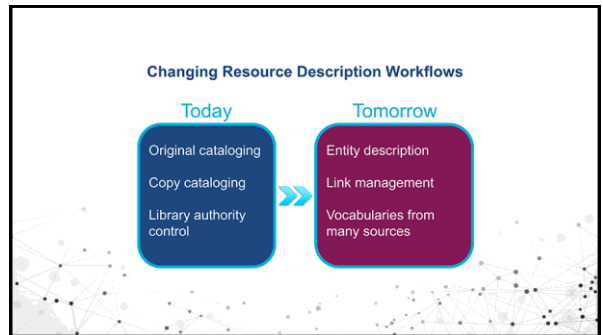
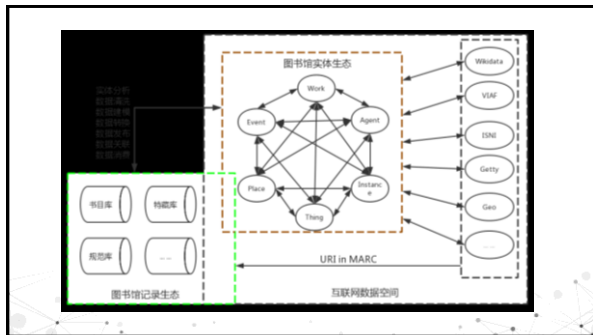
OCLC Project Passage—经验

- 从人类可读记录到知识图谱的转变是一种范式转换：
 - 识别“重要的实体”将推动未来书目领域的建设。而非当前描述“手中的文献”；
 - 基于字符串管理的规范数据和书目数据之间的传统区别消失。
 - 并不是放弃规范控制，仍然坚持图书馆数据的规范性和权威性。
- 众包：最大化众包的潜力，同时最小化风险。数据质量控制。

面向对象的编目

识别实体及其关系
基于URI的身份管理
开放、协作、众包

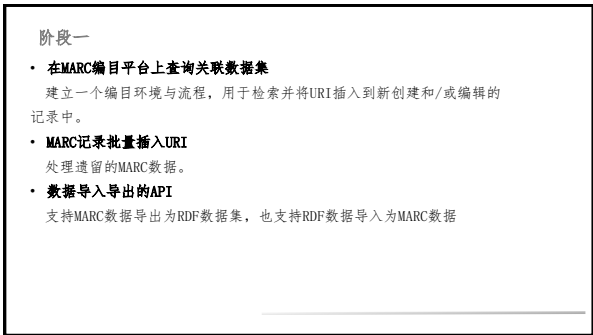
书目数据知识库



转换基础

- 转换到LD最重要的基础是在记录中添加共享的、公认的唯一标识符 (URI)，用于计算机消歧。







FOLIO (The Future of Libraries is Open)

• 什么是FOLIO

Folio致力于开发一个创新的开源图书馆服务平台，更是一个由图书馆、开发商、供应商及其他利益机构共同组成的协作社区。

• 国内FOLIO社区

序号	工作组名称	牵头单位	备注
1	元数据管理	上海图书馆	制订元数据标准、定义管理功能等
2	资源管理	上海交通大学图书馆	纸质资源的采、编、发现等
3	资源获取	北京大学图书馆、CALIS	流通、ILL等
4	用户管理	中国人民大学图书馆	
5	系统配置管理	CALIS	
6	宣传与培训	CALIS、北京大学图书馆	

FOLIO (The Future of Libraries is Open)

• 上图FOLIO社区

上海图书馆、复旦大学、交通大学、南京仰格、科图、麦达、阿法迪、超星、嘉图等

社区文档, <http://sig.library.sh.cn:8090>

FOLIO演示网址, <http://demo.folio.library.sh.cn:4000/>

FOLIO (The Future of Libraries is Open)

• 最重要的特色

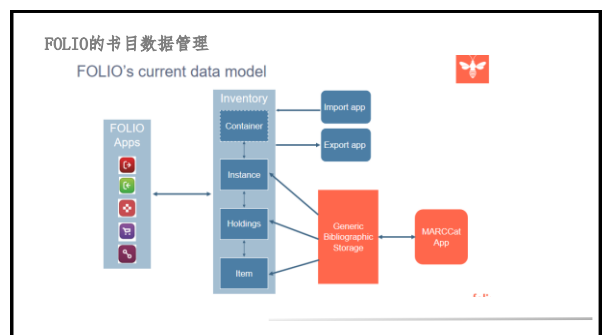
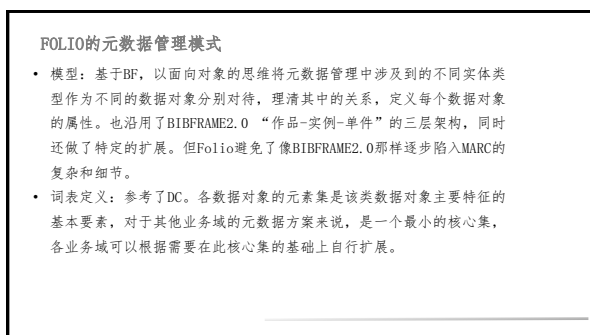
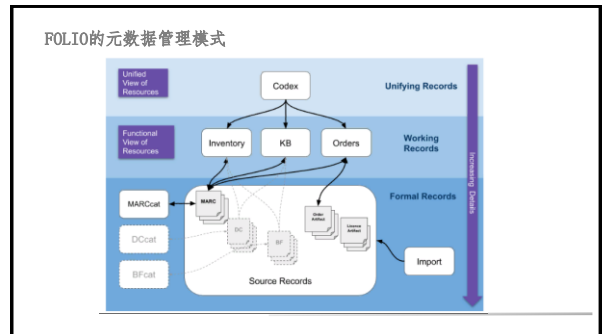
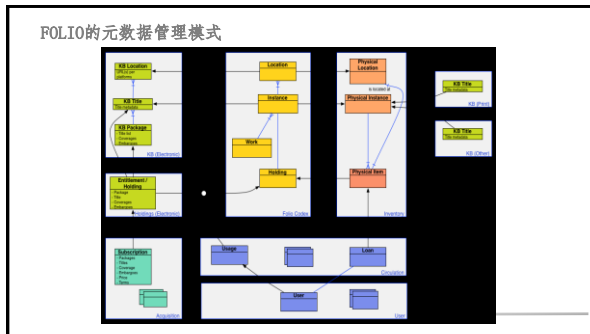
开放的平台

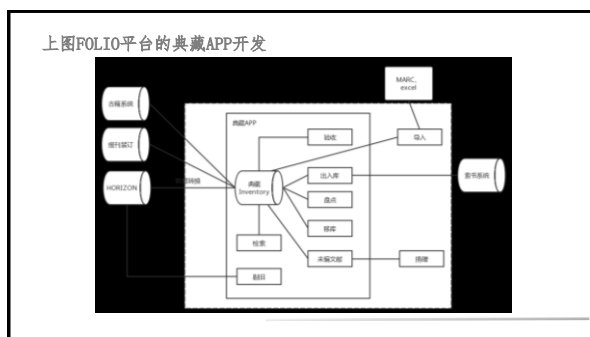
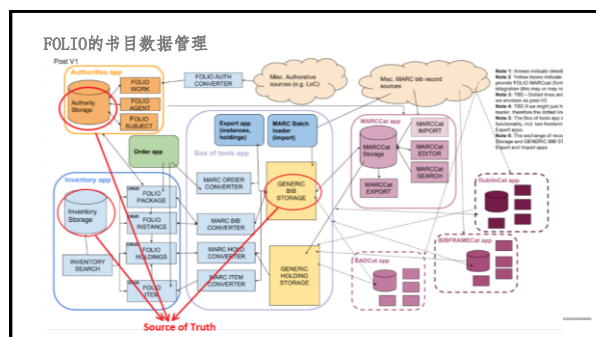
社区驱动

微服务架构

高度的自主权与定制化开发







项目

- (1) W3C图书馆关联数据孵化小组
- (2) LD4L (2014-2016), LD4L Labs (2016-2018), LD4P (2016-2018), LD4PE (2016), LD4P2 (2018-2020)
- (3) OCLC Research's work with linked data
Passage, SchemaBibEx, Work Entity, Person Entity, VIAF, FAST, Worldcat-LD
- (4) LC BIBFRAME Pilot, Phase 1 (2015.9-2016.3), Phase 2 (2017.6-)
- (5) PCC
URIs in MARC, Identity Management in NACO, Work Entity, Creation and Function of Name Authorities in a Non-MARC Environment等
- (6) BIBFLOW

项目与工具

- (1) BFEEditor Sinopia (LD4P)
- (2) BFEEditor (LC)
- (3) Omeka (数字文化遗产的元数据编辑发布平台)
- (4) SHARE-VDE (整个LBD的环境)
- (5) ALIADA (图情档关联数据发布)
- (6) RML
- (7) OpenRefine
- (8) Karma RMLEditor (可视化的RDB2RDF工具)
- (9) Skosmos, OpenSKOS (规范、取值词表RDF数据发布工具)
- (10) 上海图书馆转换工具RDB2RDF, 上海图书馆关联数据服务平台LDSP

主要会议

- (1) LC BIBFRAME Update Forum
- (2) LD4 Conference on Linked Data in Libraries
- (3) SWIB
- (4) European Bibframe Workshop
- (5) The International Linked Open Data in Libraries, Archives and Museums (LODLAM) Summit
- (6) MTSR: Research Conference on Metadata and Semantics Research

主要参考资料

- (1) BIBFRAME Training at the Library of Congress <https://www.loc.gov/catworkshop/bibframe/>
- (2) From MARC to BIBFRAME: Linked Data on the Ground - ALCTS馆藏与技术服务协会 <http://t.cn/AiQ4R1zE>
- (3) Getting Started with Linked Data - OCLC <http://t.cn/AiQ43tt7>
- (4) Linked Data for Librarians by Seth van Hooland and Ruben Verborgh
<https://course.freemetadata.org/>
- (5) Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description
<http://t.cn/AiQ4RZIN>
- (6) Linked Data: Evolving the Web into a Global Data Space <http://linkeddatabook.com/editions/1.0/#toc3>
- (7) How to Publish Linked Data on the Web <http://t.cn/AiQ4nEGT>
- (8) Best Practices for Publishing Linked Data <https://www.w3.org/TR/ld-bp/#MODEL>
- (9) Report on Available Linked Data Training Resources-PCC
<https://www.loc.gov/aba/pcc/sect/documents/PCSCSTFinalReportonAvailableLinkedDataTrainingResources.docx>

主要参考资料

- (10) BIRFRAME Training Plan for Cohort Institutions--LD4P
https://docs.google.com/document/d/1NtFYSLLa0KEmMcngCz7WSGd_mZUUt0B81AL3e2f1Q/edit
- (11) 关联数据：万维网上的结构化数据 9787115472649
- (12) 语义网与数字图书馆 9787305195846
- (13) 基于关联数据技术的数字人文项目建设-夏翠娟（上图数字人文博客） <http://dhhlog.library.sh.cn/zh-hans/node/7>
- (14) SPARQL培训-陈涛（公众号：关联数据） <https://dwz.cn/9bRfNXb7>
- (15) 2018属兔啦-刘炜、夏翠娟、陈涛的关联数据培训 <https://share.weiyun.com/5dL2SfX>
- (16) A New Methodology for Building Lightweight, Domain Ontologies <http://www.mkbergman.com/908/a-new-methodology-for-building-lightweight-domain-ontologies/>
- (17) Ontology Best Practices http://wiki.opensemanticframework.org/index.php/Ontology_Best_Practices
- (18) Ontology engineering in the era of linked data
<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/bult.2015.1720410407>
- (19) 众源数据<http://usources.cn/>
- (20) OpenKG, CN

