

全国图书馆联合编目中心2019年度工作会议 贵阳

## 文献编目：从数字化到数据化

胡小菁（华东师范大学图书馆）

### 大纲

- 数字化 vs 数据化
- MARC的与时俱进
- 国际编目界的选择：关联数据
- 书目模型、编目规则、数据格式的数据化转变
  - IFLA的FRBR系列，RDA，MARC，BIBFRAME
  - 规范控制：从名称选择到实体管理
- 数据化实践
- 国际合作编目机构的战略

编目相关标准  
与实践的进展

### 数字化 vs 数据化

### 数字化 vs 数据化：定义

- 数字化：将信息转为**计算机可读**的数字格式的过程
- 数据化：把信息变为数据的过程
  - 指一种把现象转变为**可制表分析的量化形式**的过程
    - 迈克尔·舍恩伯格、维克多·迈耶·齐泽克《大数据时代》，2013
- 数据
  - 数据代表着对某件事物的**描述**
  - 数据可以**记录、分析和重组**
    - 迈克尔·舍恩伯格、维克多·迈耶·齐泽克《大数据时代》，2013

结构化

## MARC : 数据化先行者

- 作为**结构化数据**，既是数字化成果，也天然具有数据化属性
- 不仅**描述/记录**书目、规范等信息，也可用以进行一定的**量化分析**。

```
200 1 红楼梦 [zhong lou meng] (清) 曹雪芹, 高鹗著
210 北京 [c:中华书局] d2018
215 5册 (13, 1037页) | d21cm
225 2 中华大辞典 [zhong hua da zi jin dian]
410 0 112001 | d中华大辞典
606 0 章回小说 [zhang hui xiao shuo] y中国 | z清代
606 0 古典小说 [gu dian xiao shuo] y中国 | z清代
606 0 长篇小说 [chang pian xiao shuo] y中国 | z清代
690 1242-47 | y2
701 0 曹雪芹, [cao xue qin] (约1715-1783) | 著
701 0 高鹗, [gao e] (约1738-约1815) | 著
```

## MARC的与时俱进



## Roy Tennant : MARC必须死 ( 2002 )

- 用**数字编码**
- 粒度：元素精细**层次不一**
- MARC是**扁平的**，但目录是**分层的**
- 无法容纳图书封面**图片、评论等**
- 多语言文字难以**正确处理**
- 技术边缘化、过时的标准


推荐：XML





## MARC : 2709格式

```
00882cam 2200253 a 45040010008000000000000410000001000130004
902000150006203500013000770400018000900490009001080500022001
17082001900139100001800158245012100176260004900297300002500
34644000500037150400510042165000640047265000470053670000190
05837000026006021320053192060801992 nyu bf 001 0
eng l $a920210871 $a15557012051 $a261328111 $aDL$cdL
C$dRGLI $aHNK0100$a2678. 95b. C58 1992100$a025/ 00285$22011
$aCohn, John M.110$aPlanning for automation :$ba how-to-do
it manual for librarians /$cJohn M. Cohn, Ann L. Kelsey, Ke
ith Michael Fiels. l $aNew York :$bNeal Schuman Publishers,
$c1992. l $avi. 116 p. ;$c28 cm. l 0$aHow-to-do it manuals
for libraries :$uno. 25. l $aIncludes bibliographical refer
ences and index. l 0$aLibraries$automation$management$han
dbooks, manuals, etc. l 0$aLibrary planning$Handbooks, manu
als, etc. l10$aKelsey, Ann L.110$aFiels, Keith Michael. l x
```

LDR	——nam0 22—— 450
001	010111547
005	20190610062504.0
010	ja 978-7-5013-6638-5 b 精装 ja CNY580.00
100	ja 2019031602018 emky0chy50 ea
1010	ja chi
102	ja CN b 110000
105	ja 8 spz 001yy
106	ja r
2001	ja 国家图书馆年鉴 jh 2018 jh 专著 jh guo ja tu shu guan nian jian
210	ja 北京 jk 国家图书馆出版社 ja 2018
215	ja 418页 k 彩图 ja 29cm
264	ja 英文共同题名, National Library of China yearbook
314	ja 主编, 汪东波, 原刊书部, 国家图书馆编
330	ja 本书总结2017年国家图书馆事业的发展状况, 由综述、特别记数、基础业务工作、服务工作、国家数字图书馆建设与服务、重点文化工程及项目、文献整理与编辑出版、学术与科研、对外交流与合作、机构与人事管理、设施建设与安全保障、党建与群团工作、统计资料等部分构成。
80102	ja 中国国家图书馆 jz 2018 j 年鉴
990	ja G259.251-54 jv 5
701 0	ja 汪东波 j (1962-12-) jk 主编 jk wang dong bo
71202	ja 国家图书馆 14 编 jh guo ja tu shu guan
090	ja G259.2 b web
096	ja G259.2 b pdf

	
<pre> &lt;?xml version="1.0" encoding="UTF-8" ?&gt; &lt;collection xmlns="http://www.loc.gov/MARC21/slim" xmlns:xs="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/MARC21/slim.xsd"&gt;   &lt;record id="A1Kw" xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/MARC21/slim.xsd"&gt;     &lt;leader&gt;00000nam a2200000 a 4500&lt;/leader&gt;     &lt;controlfield tag="001" s1s-label="MARC"&gt;99229402&lt;/controlfield&gt;     &lt;controlfield tag="008" s1s-label="MARC"&gt;9906041997 cc 0 b a00&lt;/controlfield&gt;     &lt;datafield tag="020" ind1="1" ind2="1" s1s-label="ISBN"&gt;       &lt;subfield code="a"&gt;9628077171&lt;/subfield&gt;     &lt;/datafield&gt;     &lt;datafield tag="040" ind1="1" ind2="1" s1s-label="MARC"&gt;       &lt;datafield tag="041" ind1="1" ind2="1" s1s-label="MARC"&gt;         &lt;datafield tag="049" ind1="1" ind2="1" s1s-label="MARC"&gt;           &lt;datafield tag="090" ind1="1" ind2="1" s1s-label="CALL NUMBER"&gt;             &lt;subfield code="a"&gt;G259.2&lt;/subfield&gt;           &lt;/datafield&gt;         &lt;/datafield&gt;       &lt;/datafield&gt;     &lt;datafield tag="245" ind1="1" ind2="0" s1s-label="TITLE"&gt;       &lt;subfield code="a"&gt;国家图书馆年鉴 2018&lt;/subfield&gt;       &lt;subfield code="b"&gt;国家图书馆 2018 年&lt;/subfield&gt;       &lt;subfield code="c"&gt;国家图书馆 2018 年&lt;/subfield&gt;     &lt;/datafield&gt;     &lt;datafield tag="246" ind1="1" ind2="1" s1s-label="OTHER TITLE"&gt;       &lt;datafield tag="246" ind1="1" ind2="1" s1s-label="OTHER TITLE"&gt;         &lt;subfield code="a"&gt;[著录]&lt;/subfield&gt;         &lt;subfield code="b"&gt;国家图书馆 2018 年&lt;/subfield&gt;         &lt;subfield code="c"&gt;国家图书馆 2018 年&lt;/subfield&gt;       &lt;/datafield&gt;     &lt;datafield tag="300" ind1="1" ind2="1" s1s-label="DESCRIPTION"&gt;       &lt;datafield tag="410" ind1="1" ind2="0" s1s-label="SERIES"&gt; </pre>	

ID 号	010111547
通用数据	2019031602018 emky0chy50 ea
题名与责任	●国家图书馆年鉴 2018 [专著]
出版项	●北京: 国家图书馆出版社, 2018
载体形态项	418页: 彩图; 29cm
语言	chi
题名责任注	英文共同题名, National Library of China yearbook
内容提要	本书总结2017年国家图书馆事业的发展状况, 由综述、特别记数、基础业务工作、服务工作、国家数字图书馆建设与服、重点文化工程及项目、文献整理与编辑出版、学术与科研、对外交流与合作、机构与人事管理、设施建设与安全保障、党建与群团工作、统计资料等部分构成。
主题	●中国国家图书馆 -- 2018 -- 年鉴
中图分类号	●G259.251-54
著者	●汪东波 (1962-12-) 主编
附加款目	●国家图书馆 编

	
Metadata Object Description Schema <a href="#">Official Web Site</a>	
<pre> &lt;titleInfo&gt;   &lt;title&gt;Sound and fury&lt;/title&gt;   &lt;subTitle&gt;the making of the punditocracy&lt;/subTitle&gt; &lt;/titleInfo&gt; &lt;name type="personal" authorityURI="http://id.loc.gov/authorities/names" valueURI="http://id.loc.gov/authorities/names/n92101908"&gt;   &lt;namePart&gt;Alterman, Eric&lt;/namePart&gt;   &lt;role&gt;&lt;roleTerm type="text"&gt;creator&lt;/roleTerm&gt;&lt;/role&gt; &lt;/name&gt; &lt;typeOfResource&gt;text&lt;/typeOfResource&gt; &lt;genre authority="marcgt"&gt;bibliography&lt;/genre&gt; </pre>	
	

### MARC家族的局限

- 受格式限制，书目/规范等信息以“记录”形式作为一个整体存在。
- 机读目录中的单个“数据”不具有重组能力，因而离真正的数据化还有不小的距离。

### 国际编目界的选择：关联数据

关联数据：2006 Tim Berners-Lee

以2008年瑞典国家图书馆以关联数据发布联合目录Libris为重要标志

### 选择关联数据取代MARC

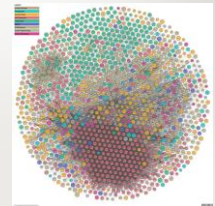
- 为摆脱MARC格式的束缚，国际图书馆学界在新兴的互联网技术中选择了关联数据，力图使图书馆的书目数据由机器可读（machine-readable）走向机器可操作或可执行（machine-actionable），从而融入互联网全球数据库。

机器可操作：是指以一致的方式构造信息，以便机器或计算机可以针对结构进行编程。

### 图书馆目录与规范的关联数据发布

2008-

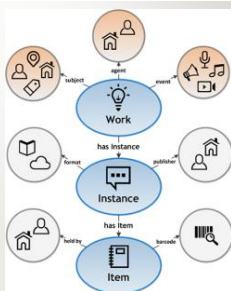
- 瑞典国家图书馆
- 虚拟国际规范档（VIAF）
- 匈牙利国家图书馆
- 德国国家图书馆
- 不列颠图书馆
- .....



<https://www.hol-cloud.net/version/2019-03-20/hol-cloud.png>

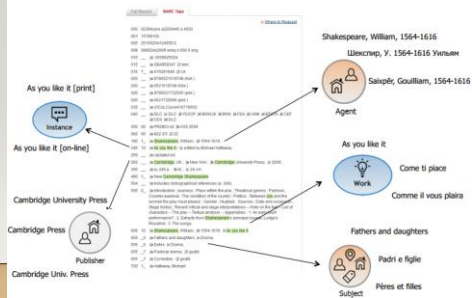
## 书目框架：BIBFRAME

- 2011-
- 美国国会图书馆
- 计划取代MARC



BIBFRAME 2.0模型 (2016-4-21)

## The new revolution: from record to entity



Tiziana  
Possemato. From  
MARC to  
BIBFRAME in the  
SHARE-VDE  
project.  
BIBFRAME  
Update Forum,  
ALA Annual  
Conference 2018

## 不是发布为关联数据，就可以自动关联了

- 名称（文本/字符串）**关联的RDF** → 实体URI/IRI
- 莎士比亚 <http://id.loc.gov/rwo/agents/n78095332>
- 皆大欢喜 <http://viaf.org/viaf/178316948>
- As you like it

数字化 → 数据化



## 书目模型、编目规则和数据格式的数据化转变

Roy Tennant：MARC语法、数据元素、编目规则交织在一起  
伴随着近年图书馆相关领域关联数据应用的探索，在21世纪的第2个十年，  
已经为数据化升级做好了准备。

## IFLA的FRBR系列：数据化建模

1. 从记录到数据
2. 名称：从属性到实体

## 记录与数据：定义（以书目为例）

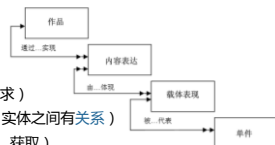
- 书目记录：与图书馆目录和国家书目中描述的各种**实体**相关的**数据集合**。
- 书目数据：数据集合中包含的**数据单元**。

——《书目记录功能需求》（FRBR）

## 书目模型的开始

- 1997：FRBR（书目**记录**的功能需求）
  - 实体-关系模型（**实体有属性**、实体之间**有关系**）
  - 用户任务（查找、识别、选择、获取）
  - 国家书目记录基本需求

图 3.1：第一组实体与基本关系



## 从关注**记录**到关注**数据**

- 编制FR系列模型
- 1999：FRANAR（规范**记录**的功能需求与编号）
- 2005：FRSAR（主题规范**记录**的功能需求）
- 发布
- 2009：FRAD（规范**数据**的功能需求）
- 2010：FRSAD（主题规范**数据**的功能需求）

模型“着重于数据本身，不注重如何打包数据”（为记录）——FRAD

## FRBR系列统一版：LRM IFLA图书馆参考模型：书目信息的概念模型

- 2013编制：FRBR LRM
- 2017发布：IFLA LRM
- FRBR（书目记录的功能需求）
- 实体-关系模型
- 用户任务
- 国家书目记录基本需求
- 实体-关系模型

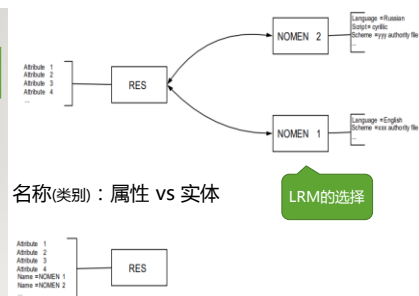
注重“数据/信息”不注重“记录”  
影响RDA：取消“核心元素”

## 名称与实体

- 实体：指任何事物（Thing）。
- 人们通常以名称指代实体，但名称与实体不能混为一谈。
- 数据化要求区分名称与实体
- 1. 我不知道周树人(实体)是谁
- 2. 鲁迅(实体)原名周树人(名称)

## FRBR系列模型：名称(类别)从属性到实体

- FRBR：书目记录3类实体：书目、责任、主题
- 名称作为实体的属性（如作品的题名、个人的名称）
- FRAD：规范数据的对象及其名称
- 实体：名称(name)、标识符、受控检索点
- FRSAD：主题的名称
- 实体：命名(nomen)



《IFLA图书馆参考模型》图4.1：命名(Nomen)的交替实体-关系模型

## LRM确定名称为实体的原因

- 实体-关系模型无法声明属性间关系，如果名称作为属性，则无法表达多个名称间关系
  - LRM定义的nomen间关系：等同、部分、派生
- 以命名(nomen)为实体，可以为其定义语言、文字、来源等属性
  - Lu, Xun@en
  - 鲁迅@zh

对编目规则的影响：nomen与其他实体并列，检索点作为nomen，可以应用于所有实体（比如出版者）。

## MARC21的数据化改造

与 URI/IRI 有关的新增与重定义

## 控制子字段\$0和\$1：捕获 URI

\$0 描述URI  
\$1 实体URI

- \$0：规范记录控制号或标准号（省略前缀(uri)） **2016修改**  
100 1#\$aTrollope, Anthony.\$d1815-1882.\$0(isni)0000000121358464
- \$1：真实世界对象(RWO, Real-World Object) URI **2017新增**  
710 2#\$aCalifornia Poets in the Schools (Project).\$eissuing body,  
\$epublisher.\$0http://id.loc.gov/authorities/names/n85319780  
\$1http://id.loc.gov/rwo/agents/n85319780

MARC >> Bibliographic >> Appendix A - Control Subfields  
<http://www.loc.gov/marc/bibliographic/ecbcdcntf.html>

**2017新增**

书目资源=作品、内容  
表达、载体表现、单件

\$a标签  
\$i关系信息  
\$4关系：代码或URI

## 758字段：资源标识符

100 1#\$aBeard, Henry.  
245 10\$aBored of the rings :\$ba parody of J.R.R. Tolkien's *The lord of the rings* /\$cby Henry N. Beard and Douglas C. Kenney of the Harvard lampoon.  
758 ##\$4http://rdaregistry.info/Elements/m/P30135  
\$iHas work manifested:\$aBored of the Rings  
\$1http://www.wikidata.org/entity/Q1613936 **作品 URL**  
758 ##\$4http://rdaregistry.info/Elements/w/P10197  
\$iParody of (work):\$aTolkien, J. R. R. (John Ronald Reuel), 1892-1973.  
Lord of the rings\$0http://id.loc.gov/authorities/names/no97079452

**戏仿原创作品 描述URL**

758 - Resource Identifier (R)  
<http://www.loc.gov/marc/bibliographic/bd758.html>



## 758的作用

- 扩大MARC支持**关联数据**标识符.....是图书馆将关联数据构建到元数据生产过程的**迁移路径**的一部分
- 将书目描述与适当的作品或其他资源实体**关联**，是将MARC书目数据转换为关联数据的主要挑战。
- 记录MARC记录中相关资源的URI，使得当MARC记录**转换为关联数据表示**时，URI可以作为对象包含在RDF语句中。
- 在无URI可用的情况下，非URI标识符对于正确的数据调和（实体解析）仍然是有价值的。

MARC PROPOSAL NO. 2017-09  
<http://www.loc.gov/marc/marc/2017/2017-09.html>

## 用 URI/IRI 强化MARC记录

```
.....
100 0#$aBridget,$cof Sweden, Saint,$dapproximately 1303-1373.
$0http://id.loc.gov/authorities/names/n50044402
240 10$aRevelationes.$lEnglish$s(Searby)
$0http://id.loc.gov/authorities/names/n84007202
.....
650 #0$aVisions.$0http://id.loc.gov/authorities/subjects/sh85143882
700 1#$aSearby, Denis Michael,$etranslator.
$0http://id.loc.gov/authorities/names/nr98021028
.....
```

<http://id.loc.gov/tools/bibframe/compare-id/full-ttl?find=13910411>

## MARC更新的意义

- 暂时或**相当一段时间内**MARC会继续使用
- 在BIBFRAME实施前，需要软件在两个格式间**来回转换**
  - 具有URI对转换处理很重要

## 附：UNIMARC相关更新（2012-2018）

- | 书目（作品、内容表达相关）         | 规范（个人标识符相关） |
|-----------------------|-------------|
| • 506 首选检索点-识别作品      | • 010 ISNI  |
| • 507 首选检索点-识别内容表达    | • 017 其他标识符 |
| • 576 名称-首选检索点-识别作品   | • 如ORCID    |
| • 577 名称-首选检索点-识别内容表达 |             |

类似758（MARC21）

## 新RDA（3R计划）：一切皆可选

《资源描述与检索》，2017- 基于《IFLA图书馆参考模型》(LRM)

1. 取消核心元素
2. 4种记录方法

## 新RDA：所有元素均为可选

- 2010：RDA（资源描述与检索）
  - **核心元素**（基于FRBR“国家书目记录的基本需求”）
- 2011：ISBD统一版
  - 各著录单元的必备状态
- 2017：IFLA-LRM
  - 不含FRBR中原有的“国家书目记录基本需求”
- 2018：测试版RDA/新RDA
  - 放弃核心元素：所有元素都是**可选**的

## 新RDA：元素可选

eg. 载体表现：联机资源

- 载体类型
- 载体表现数量
- 数字文件特征

引用编号53.04.89.90

**Online resources**

**OPTION**  
Record online resource as a Manifestation; **carrier type** for a manifestation that is accessed using a communications network.

**CONDITION**  
An online resource is complete.

**OPTION**  
Record a Manifestation; **extent of manifestation**.

**CONDITION**  
An online resource consists of more than one file.  
A description of the characteristics of each file is considered important for identification or selection.

**OPTION**  
Record a separate Manifestation; **digital file characteristics** of each file.

## 新RDA：条款可选

可选=原RDA：

- 正文条款
- 交替
- 可选增加
- 可选省略
- 例外

• 新RDA例

- 团体首选名称：首冠词

引用编号42.98.97.06

## Omissions and changes to a name

### Initial articles

#### CONDITION

A value of a name includes an initial article.

#### OPTION

Record a value that includes an initial article.

Example

#### OPTION

Record a value that omits an initial article.

Example

## 对照原RDA：团体首选名称：首冠词

- 原RDA
  - 正文
  - 替代条款
    - 选用时说明
- 新RDA
  - 选项
  - 必须说明采用哪个选项

11.2.2.8 Initial Articles [\[RDA\]](#)

When recording a preferred name for corporate body, include an initial article, if present.

**EXAMPLE** [\[RDA\]](#)

The Library Association  
Der Wahnschaffungs

**Alternative** [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#) [\[RDA\]](#)

Omit an initial article (see appendix C [\[RDA\]](#)) unless the name is to be accessed under the article (e.g., a corporate name that begins with an article that is the first part of the name of a person or place: [\[RDA\]](#)).

**EXAMPLE** [\[RDA\]](#)

Amis de la terre du Moyen  
Les Amis de la terre du Moyen

## Recording methods

The data value of an RDA element can be recorded using one or more of the following methods:

### 新RDA：4种记录方法

- 记录非结构化描述  
照录/转录
  - 记录结构化描述  
规定格式，如检索点、结构化附注
  - 记录标识符  
如标准号码、本地编号
  - 记录IRI（国际资源标识符）
- 每个元素适用一种或多种记录方法

## 新RDA：记录方法可选

eg. 载体表现：记录方法

如何选择记录方法？  
如何集成元素（数据）形成一条“记录”？  
由数据制作机构通过“政策声明”或“应用配置文件”决定。

**Recording an unstructured description**

Record an unstructured description as *Manifestation title of manifestation* [\[RDA\]](#).

For general guidance on unstructured descriptions, see Recording methods, *Recording an unstructured description* [\[RDA\]](#).

**Recording a structured description**

Record a structured description as *Manifestation access point for manifestation* [\[RDA\]](#).

For general guidance on structured descriptions, see Recording methods, *Recording a structured description* [\[RDA\]](#).

**Recording an identifier**

Record an identifier as *Manifestation identifier for manifestation* [\[RDA\]](#).

For general guidance on identifiers, see Recording methods, *Recording an identifier* [\[RDA\]](#).

**Recording an IRI**

Record an IRI for the entity as a real-world object.

For general guidance on IRIs, see Recording methods, *Recording an IRI* [\[RDA\]](#).

引用编号 35.04.54.16

## 新RDA：无核心元素，有最低描述

eg. 记录载体表现

- 载体表现的称谓（至少其一）
  - 题名：正题名
  - 检索点
  - 标识符
- 表现的**内容表达**和/或**作品**（至少其一）

**Recording**

**Minimum description of a manifestation**

A minimum description of a manifestation must include at least one of the following elements:

- *Manifestation: appellation of manifestation* [\[RDA\]](#)
  - *Manifestation title of manifestation* [\[RDA\]](#)
    - *Manifestation title proper* [\[RDA\]](#)
  - *Manifestation access point for manifestation* [\[RDA\]](#)
    - *Manifestation authorized access point for manifestation* [\[RDA\]](#)
    - *Manifestation identifier for manifestation* [\[RDA\]](#)
- *Manifestation: expression manifestation* [\[RDA\]](#)
- *Manifestation work manifestation* [\[RDA\]](#)

A minimum description of a manifestation must also include at least one of the following elements:

引用编号 22.85.24.59, 53.57.76.31

## 新RDA：数据多样性。

“一致” 本来就不存在  
eg. 不同的编目语言

同一个RDA，不同的选项

- 可扩展性
  - 比如用本地控制词表代替工具包中提供的词表
- 选项和配置
  - 没有自上而下、一刀切的“全球书目控制”系统的全球协议
  - 必须与应用纲要一同使用，选择本地应用的元素和记录方法

——RSC时任主席Gordon Dunsire

Gordon Dunsire, Accommodating local cataloguing traditions in a global text.  
Diversity of Data: RDA in the international context, National Library of Malaysia, 2018-8-23

## 应用纲要/应用配置文件

- RDA本身基本上成为一本数据字典，指定元素、其定义以及在描述中提供它们的指导原则。使用RDA的机构决定使用哪些元素、遵循哪些可选准则以及用于系统创建完整描述的结构化路径，并记录在应用纲要中。



——西班牙国家图书馆Kevin M. Randall  
摘自RDA-L邮件组

## RDA Registry

### 元素集

- RDA Registry (Home)
- Elements (RDA element sets)
  - Classes
  - Agent properties
  - Expression properties
  - Item properties
  - Manifestation properties
  - Notation properties
  - Place properties
  - Time-span properties
  - Work properties
  - RDA Entity properties
  - Media-element properties
  - Unconstrained properties

### 取值词表

- RDA/ONIX Framework elements
- Values (value vocabularies)
  - RDA values
  - RDA/ONIX Framework values
- Data (Linked data using RDA vocabularies)
  - Curie prefixes (Abbreviations for compact URIs, XML namespaces, etc.)
  - Examples (Single resource)
  - R-Bails (Multiple resources)
  - Datasets (Multiple resources)
- Tools
  - Maps (RDF maps between RDA vocabularies and other namespaces)
  - Alignments (Alignment tables for RDA vocabularies and other namespaces)
  - Profiles (Application profiles using RDA vocabularies)
  - RIMAF (RDA data editor)

## 注册URI/IRI

RDA对应用  
关联数据是  
认真的！

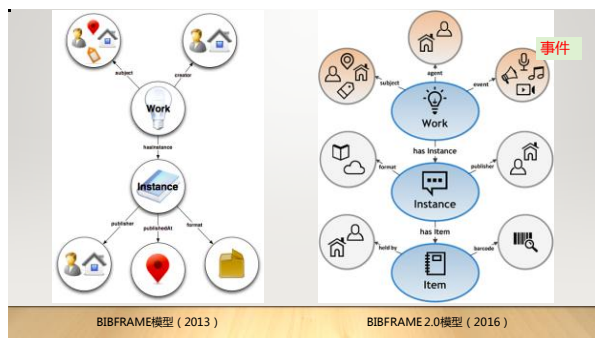
## RDA Reference value vocabularies

- RDA Aspect Ratio Designation
- RDA Bibliographic Format
- RDA Broadcast Standard
- RDA Carrier Element Unit
- RDA Carrier Type
- RDA Cartographic Data Type
- RDA Colour Content
- RDA Configuration of Playback Channels
- RDA Content Type
- RDA Conventional Collective Title
- RDA File Type
- RDA Font Size
- RDA Form of Musical Notation
- RDA Form of Notated Movement
- RDA Form of Tactile Notation
- RDA Format of Notated Music
- RDA Frequency
- RDA Generation
- RDA Groove Pitch of an Analog Cylinder
- RDA Groove Width of an Analog Disc
- RDA Illustrative Content
- RDA Layout
- RDA Material
- RDA Media Type
- RDA Mode of Issuance
- RDA Polarity
- RDA Presentation Format
- RDA Production Method

## BIBFRAME：从1.0到2.0

1.0：2013模型+2014词表

2.0：2016模型&词表



### 非主题事件：演出（柴可夫斯基第5交响曲）

以物件中心描述（盒式录音带）      以事件为中心描述（演出）

- 主要责任者
  - 柴可夫斯基 作曲
  - 鲁道夫 肯普 指挥
  - 柏林交响乐团 演奏
- 题名：柴可夫斯基第5交响曲
- 出版：英国：百代音乐公司出版，1960
- 表演者
  - 演出时间
  - 演出地点
  - 职员（灯光、录音等）

### BIBFRAME核心类

#### BIBFRAME1.0

- 作品
- 实例
- 规范
- 注释
  - 封面.....
  - 馆藏

#### BIBFRAME2.0

- 作品
- 实例
- 单件

#### FRBR第1组实体

- 作品、内容表达
- 载体表现
- 单件

### BIBFRAME核心类演变分析

- 取消规范核心类：规范控制不再依赖规范检索点
- 取消注释核心类：复用成熟关联数据词表
- 新增单件核心类：跨越书目和馆藏格式

## 规范控制不再依赖 规范检索点

- 统一资源标识符URI
- 虚拟国际规范档VIAF  
<http://viaf.org/viaf/29537230>
- 国际标准名称标识符  
ISNI 0000 0003 6868 2258



## 复用成熟关联数据词表

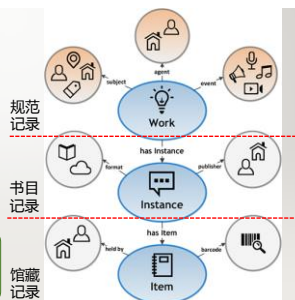
- RDF/RDFS
  - bf:Resource -> rdf:Resource
- 考虑用Web注释数据模型 (Web Annotation Data Model)
- 细粒度描述可能使用专门词表
  - 保存
  - 音像
  - .....



## 跨越书目和馆藏格式

- 在关联数据环境中，“记录”只是RDF三元组的一种组合，明确区分书目与馆藏的必要性也就不存在了

**资源-属性-值**  
**主体-谓词-客体**  
某电子刊-访问政策-校园网  
某电子刊-访问政策-开放获取



## BIBFRAME词表：类和属性数量

- BIBFRAME 1.0 : 类54, 属性270, 合计324
- BIBFRAME 2.0 : 类186, 属性195, 合计381
- 理由：对于机器处理而言，用类相比属性有3个优点：可重用、易查询、可柔性降级。

参见：BIBFRAME 2.0类的变化  
<http://catwizord.net/posts/2016050220754.html>

### BIBFRAME词表：属性的使用（灵活定义）

- 使用范围（类）
  - 定义域(domain)：102（用于单个类）
  - 用于(used with)：92（适用于多个类）
- 取值范围
  - 值域(range)：152（用于单个类）
  - 期望值(expected value)：39（适用于多个类或不指定Unspecified）
- 多个类：Work、Instance、Item中的2个或更多

### 规范控制：从名称选择到实体管理

名称、实体及其标识，真实世界对象及其描述

### 名称与实体（及其标识）

- 关联数据建立在资源描述框架(RDF)之上，在RDF中：
  - 名称以文字(literal)或字符串(string)来表达（用引号括起）
    - “鲁迅” “周树人”
    - 名称匹配 ≠ 相同对象
  - 实体作为“资源”，以统一资源标识符(URI)标识（直接引用）
    - <https://viaf.org/viaf/29537230>
    - <http://data.library.sh.cn/entity/person/x5xly94zjd7u351t>

鲁迅（维基数据）

<https://www.wikidata.org/wiki/Q23114>

### 实体的不同侧面

- 实体
  - 指任何事物(Thing)
  - RDF中的“资源”
- 真实世界对象（RWO）
  - 一个实体（例如人、地点等），可以是实际的或概念性的，常被称为事物(Thing)
- 实体=RWO
  - 名称（名称的不同形式）
  - 描述（如规范记录）
  - 标识符（URI）

URI FAQs / PCC URIs in MARC Task Group

<https://www.loc.gov/aba/pcc/bibframe/taskGroups/URI%20FAQs.pdf>

## 区分实体及其描述

- Thing/资源/实体/真实世界对象
- 描述。如规范记录，包含：
  - 描述元数据（**实体的描述**）：个人的姓名、生卒年、使用语言等，作品的题名、创作者、创作年代等
  - 管理元数据（**规范记录的描述**）：如编制机构、记录编制与更新日期等。

描述的描述

## LC名称规范档

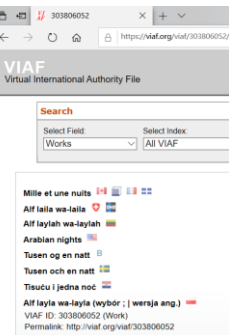
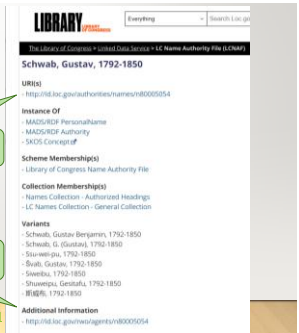
- 规范记录URI

<http://id.loc.gov/authorities/names/n80005054>

- 真实世界对象URI

<http://id.loc.gov/rwo/agents/n80005054>

<http://id.loc.gov/authorities/names/n80005054.html>



## 不同形式的翻译题名1

一千零一夜

الف ليلة وليلة

Thousand nights and one night

Mille et une nuits

Alf Laila wa-Laila

Arabian nights

天方夜谭

<https://viaf.org/viaf/303806052/>

## 不同形式的翻译题名2

红楼梦

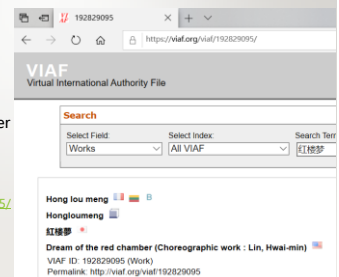
石头记

Dream of the red chamber

Hong lou meng

Honglouloumeng

<https://viaf.org/viaf/192829095/>





## 从相同名称形式到唯一标识符

- IFLA全球书目控制目标要求每个人对书目全球使用**相同形式**是不现实的

——IFLA最小级规范记录工作组最终报告（1998）

### 唯一标识符

- 个人标识符：ISNI、ORCID、Researcher ID
- 作品标识符：ISWC、ISAN（音像）、ISTC（文本）
- 载体表现标识符：ISBN、ISSN、ISMN（音乐）、ISRC（录音）
- 统一资源标识符URI/国际资源标识符IRI

## 鲁迅的各种标识符



### Wikidata

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines. Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikisource, Wikiversity, Wikinews, and others. Wikidata also provides support for many other sites, and services beyond just Wikimedia projects! The content of Wikidata is available under a free license, exported using standard formats, and can be interlinked to other open data sets on the linked data web.

owl:sameAs

Identifiers

(QID ID) 118118447

+ 1 reference

Library of Congress authority ID

00047088

+ 1 reference

VIAF ID

29537240

+ 2 references

CANL ID

4028410413

+ 0 references

NEJ Auth ID

80112747

中文名称规范联合数据库检索系统  
Chinese Name Authority Joint Database Search System

分列 → 合并

当前检索条件：未知检索词 "Schwab, gustav" 共检索到 7 条记录，当前显示前 7 条记录 保存检索式 重新检索

全部 输出 加入收藏夹

序号	规范名称	标识符	记录来源
1	规范名称 Schwab, Gustav 1792-1850	个人名称 CALS	
2	变异名称 Schwab, Gustav 1792-1850	个人名称 CALS	
3	变异名称 Schwab, Gustav Benjamin 1792-1850	个人名称 CALS	
4	变异名称 曲斯德夫·施瓦布 (Schwab, Gustav 1792-1850)	个人名称 NLC	
5	规范名称 施瓦布 (Schwab, Gustav 1792-1850)	个人名称 NLC	
6	变异名称 施瓦布 (Schwab, Gustav Benjamin) 1792-1850	个人名称 CALS	
7	变异名称 施瓦布 (Schwab, Gustav 1792-1850)	个人名称 NLC	

全部 输出 加入收藏夹

友情链接： 国际通用规范文档 (VIAF) 美国国会图书馆规范 中文名称规范一站式检索 相关网站 关于我们

中国图书馆文献标准委员会(CALIS)管理中心开发研制 | 技术支持

<http://cnass.cccna.org>

制订机构	标准/规则	实体/类	规范名称 规范检索点	实体标识符	发布年份	说明
IFLA	FRAD	名称 Name	受控检索点(实体)	标识符(实体)	2008	引入：名称(实体)
	LRM	命名 Nomen	类别属性LRM-E9-A1 受控检索点	类别属性LRM-E9-A1 标识符	2017	命名实体类别：名称、受控检索点、标识符
LC	BIBFRAME 1.0	规范(类) bf:Authority	规范检索点(属性) bf:authorizedAccessPoint	文本串	2013-2014	规范(类)：核心类
	BIBFRAME 2.0	多种实体	通用标签 rdfs:label	URI	2016-2017	取消：规范(类) 取消：规范检索点(属性)
JSC/RSC	RDA	多种实体	规范检索点	标识符(元素)	2010	引入：标识符(元素)
LC	新RDA/3R项目	多种实体	首选名称 规范检索点	结构化描述(含首选名称、规范检索点) 标识符, URI	2018	4种记录方法：增加URI
	MARC 21	子字段\$0	规范记录控制号	子字段\$0 \$1 规范记录控制号	1998	区分：实体及其描述

规范控制：从名称选择到实体管理 《数字图书馆论坛》2018年第1期  
表1 规范名称与实体标识在编目标准中的演变(有修改)

## 数据化实践

图书馆编目领域的关联数据应用经历了开发、试验，正面临实践推广的最后阶段，书目信息的数据化成为可见的发展方向。

## 关联数据应用的2个阶段

1. 转换现有书目/规范数据、以关联数据形式发布
  2. 原生关联数据：以关联数据制作书目/规范数据
- 会有很长一段时间处于混合阶段

## 从发布到原生：瑞典国家图书馆一路领先

- 2008以关联数据发布其联合目录 Libris
- 2018上线关联数据原生系统 LibrixXL
  - Libris XL的格式基于BIBFRAME 2.0，并辅以及其他模型的元素



Linked data or die! (SWIB13主题演讲)

<http://commonplace.net/2013/12/20/linked-data-or-die/>

## 美国国会图书馆：BIBFRAME进展



2017.5：全部MARC目录转换到 BIBFRAME 2.0

- 1700+万书目记录转换为BF作品、实例和单件
- 120万统一题名规范记录转换为BF作品
- BIBFRAME数据库每日更新：由来自MARC书目和规范记录的新作品/实例/单件描述

2017.6- 第2阶段试验

- 60名编目员，BIBFRAME原编

2018.6：BIBFRAME数据集

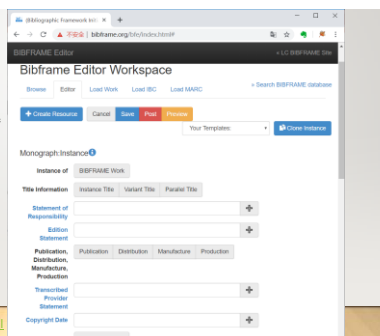
- 作品1900万+实例2400万+单件2260万=43亿三元组
- 作品、实例提供批量下载

2018.11：BIBFRAME编辑器代码、编辑器演示网站新发布

<http://www.loc.gov/bibframe/>

## BFE

- BIBFRAME编辑器



## bf: BIBFLOW

## BIBFLOW: 转换编目工作流程

- 加州大学戴维斯校区“编目再发明：未来图书馆运作模式”项目
  - 转变编目工作流程试验
  - 涉及套录和原编
  - 特别针对连续出版物和规范记录
- 2017年发布最终报告《BIBFLOW路标》
  - 总结二阶段路径：从MARC生态系统中的关联数据（为MARC批插入URI）到原生关联数据生态系统（关联数据原生编目）

<https://bibflow.library.ucdavis.edu/>

## 大型关联数据项目：LD4L系列(梅隆基金资助)

- 2014-2016 (初步探索)
  - LD4L: 创建一个既可以在各个机构内部工作，又可以通过协调、可扩展的关联开放数据网络运行的模型
- 2016-2018 (开发工具、探索工作流程、针对不同特藏资源对BIBFRAME词表进行扩展)
  - LD4L Labs: 帮助图书馆使用关联数据改善对学术资源信息的交流和理解
  - LD4P: 转变技术服务生产工作流程
- 2018-2020 (实现以关联数据来描述图书馆资源，云环境生成编目数据)
  - LD4P2: 为编目社区建立一条途径，开始转向关联数据描述图书馆资源

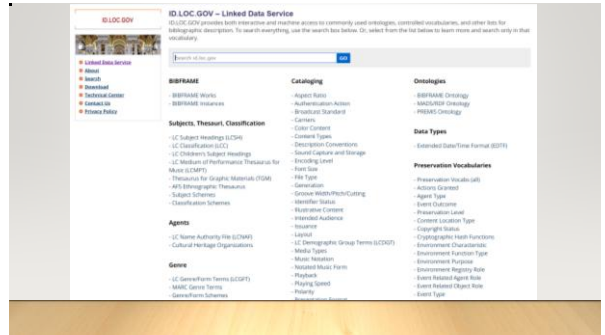
Linked Data for Libraries  
<http://www.ld4l.org/>

## SHARE-VDE (2016- )

- 意大利厂商 Casalini 和 @Cult 是欧美很多BIBFRAME项目背后的支持厂商
- 与十多家北美研究图书馆合作开发的 SHARE-VDE环境
  - 以外部来源URI强化MARC记录
  - 将1亿条书目和规范记录转换为BIBFRAME2.0后发布
  - 支持数据自动更新
- LD4P2项目采用SHARE-VDE环境，作为合作生成关联数据的支撑平台

关联数据消费

<http://www.share-vde.org>



```
graph LR; A[RDA Registry (home)] --> B[RDA element sets]; A --> C[RDA element properties]; A --> D[RDA Framework elements]; A --> E[RDA values]; B --> F[RDA注册]; C --> F; D --> F; E --> F;
```

The diagram illustrates the RDA registration process. At the top left, a green oval labeled "RDA Registry (home)" contains four items: "RDA element sets", "RDA element properties", "RDA Framework elements", and "RDA values". Arrows from each of these items point to a large green oval on the right labeled "RDA注册" (RDA Registration). Below the "RDA Registry (home)" oval is a green box labeled "元素集" (Element Sets). Below the "RDA Framework elements" oval is a green box labeled "取值词表" (Value Tables).

[illegible]

## 上海图书馆 开放数据平台



<http://data.library.sh.cn>

## 更多何处获取IRI/URI

- PCC《创制和获取URI：常用词表和参考源指南》
  - 总结了40个图档领域常用的词表和参考源的信息

参见：创制和获取URI的常用词表和参考源指南  
<https://catwizard.net/posts/20180302205719.html>

## 国际合作编目机构的战略

## PCC战略方向

- PCC (Program for Cooperative Cataloging)
  - 以美国为主的国际性合作编目组织
- 2015-2017期间，PCC接受了向**关联数据**的过渡。
- 2015年，PCC制定了从**基于记录的环境**转变为**基于元数据声明/语句和身份管理的动态集合的环境**的进程。

### PCC《2018-2021年战略方向》

SD1、扩大我们的范围，使其更加多样化，包容性，以及与研究和文化遗产社区中的元数据创建者和其他机构和项目合作伙伴协作。

SD2、最大化PCC的运营效率

SD3、应用对**关联数据**的理解

SD4、加速在网络层面向无处不在的**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

SD6、为PCC提供资源，支持PCC的战略方向和计划的整体可持续性

PCC (Program for Cooperative Cataloging) Strategic Directions, January 2018-December 2021  
<https://www.loc.gov/aba/pcc/about/PCC-Strategic-Directions-2018-2021.pdf>

### 参见

- 文献编目：从数字化到数据化  
《中国图书馆学报》2019年第5期
- 规范控制：从名称选择到实体管理  
《数字图书馆论坛》2018年第1期
- BIBFRAME核心类演变分析  
《中国图书馆学报》2016年第3期

