

# CNMARC与BIBFRAME映射及其实现

## --以上海联编中文普通图书数据为例

2017.11.23

全国文献编目工作研讨会



# 目录

- 关联数据基础
- BIBFRAME
- 映射
- 上海图书馆关联数据项目
- 小结



# 关联数据基础

# 1.1 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

## 1.2 关联数据四原则

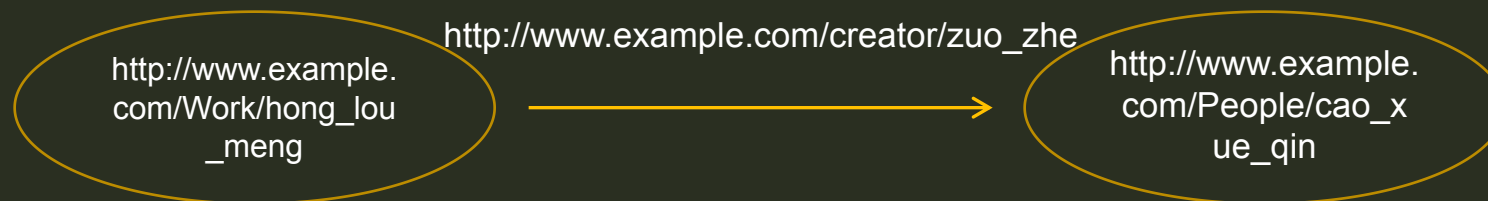
- 使用URI作为任何事物的标识名称
- 使用HTTP URI使任何人都可以访问这些标识名称
- 当有人访问某个标识名称时，提供有用的信息
- 尽可能提供相关的URI，以使人们可以发现更多的事物

# 1.3 RDF三元组

- 通用的资源描述框架
- 资源：任何在网络上被标识的对象，包括文档、人、物理对象、抽象概念等
- 被描述的资源具有一些属性特征，这些属性各有其值。
- RDF使用主-谓-宾结构（subject-predicate-object）的三元组表示资源-属性-值的陈述



# 1.4 RDF三元组



# 1.5 URI

- Uniform Resource Identifier, 统一资源标识符
- 一组按照特定语法规则构成的字符串, 为资源的标识提供了一种简单和可扩展的方法
- 作用: 标识和定位, 访问与获取, 参引, 规范控制





## 1.6 本体类和属性


- 本体：对领域知识的抽象化、概念化。提供了描述领域知识的结构化术语词表
- 词表：类，属性
- 类：抽象层面的泛指，通常是对一类实体对象共性
- 属性：对象类的各种特征，用于表示类与类之间的关系。

## 1.7 本体类和属性

- 属性在RDF三元组中是作为谓语，它连接了主语和宾语。
- 本体词表通过RDF Schema结构的属性公理定义了属性的定义域（domain）和值域（range）特征，规定了其连接的主语和宾语的类型。
- 根据定义域和值域的不同，属性可以分为对象属性（Object properties）和数据类型属性（Datatype properties）。对象属性连接实体到实体。数据属性连接的是实体到数据值。

# 1.7 本体类和属性

Property:	provisionActivity
Label:	Provision activity
Definition:	Place, name, and/or date information
Used with:	<a href="#">Instance</a>
Expected Value:	<a href="#">ProvisionActivity</a>

Property:	code
Label:	Code
Definition:	String of characters that serves as a code representation
Used with:	Unspecified
Expected Value:	<a href="#">Literal</a> 

bf:Instance--bf:provisionActivity--bf:ProvisionActivity

rdfs:Resource--bf:code--Literal

2

BIBFRAME

## 2.1 BIBFRAME简介

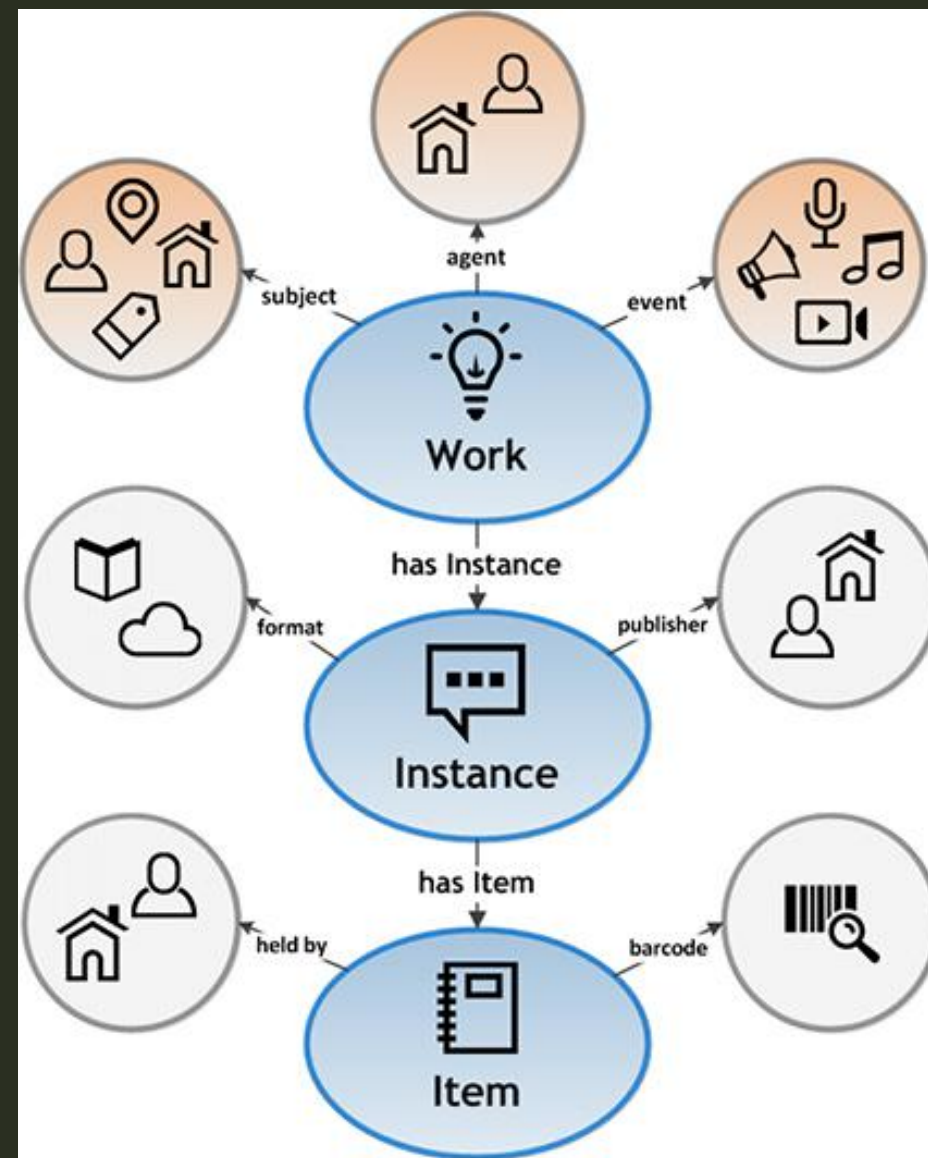
- 书目框架（BIBFRAME）是美国国会图书馆牵头，以取代 [MARC](#) 为基本目标而开发的一种面向“未来网络世界中书目描述”的书目本体
- 2011.05，开始实施
- 2012.11，BF1.0
- 2016.04，BF2.0

## 2.2 BIBFRAME模型

Work  $\approx$  FRBR Work (+ Expression)

Instance  $\approx$  FRBR Manifestations

Item  $\approx$  FRBR Item



## 2.3 BIBFRAME vs MARC

- 结构标准 vs 结构、编码标准
- 实体关系属性 vs 记录
- 使用URI进行实体标识和控制 vs 字符串
- 开放、可扩展 vs 专属封闭



3

映射



# 3.1 Work层映射

BIBFRAME词汇三元组						CNMARC
subject	predicate	object	subject	predicate	object	字段
Work	title	Title	rdfs:label	Literal	500,200\$c200\$a	hi,225\$a
			mainTitle	Literal	500\$a,200\$a	\$c
			partNumber	Literal	500\$h,200\$h,225\$h	
			partName	Literal	500\$i,200\$i,225\$i	
	language	Language			101,500\$m	
	natureOfContent	Literal			105\$a	第4-7,500\$l
	genreForm	GenreForm			100\$a	第20105\$a
	intendedAudience	IntendedAudience			8/9/11/12	
					100\$a	第17-19
			note	Note	333	
	summary	Summary			330	
	classification	shlib:ClassificationCl c	classificationPortion	Literal	690\$a	
			edition	Literal	690\$v	
			agent	Agent	7XX	
	contribution	Contribution	role	Role	7XX\$4	
			rdfs:label	Literal	101ind2=2,311,313,314,322,323	
			noteType	Literal		

## 3.2 Instance层映射

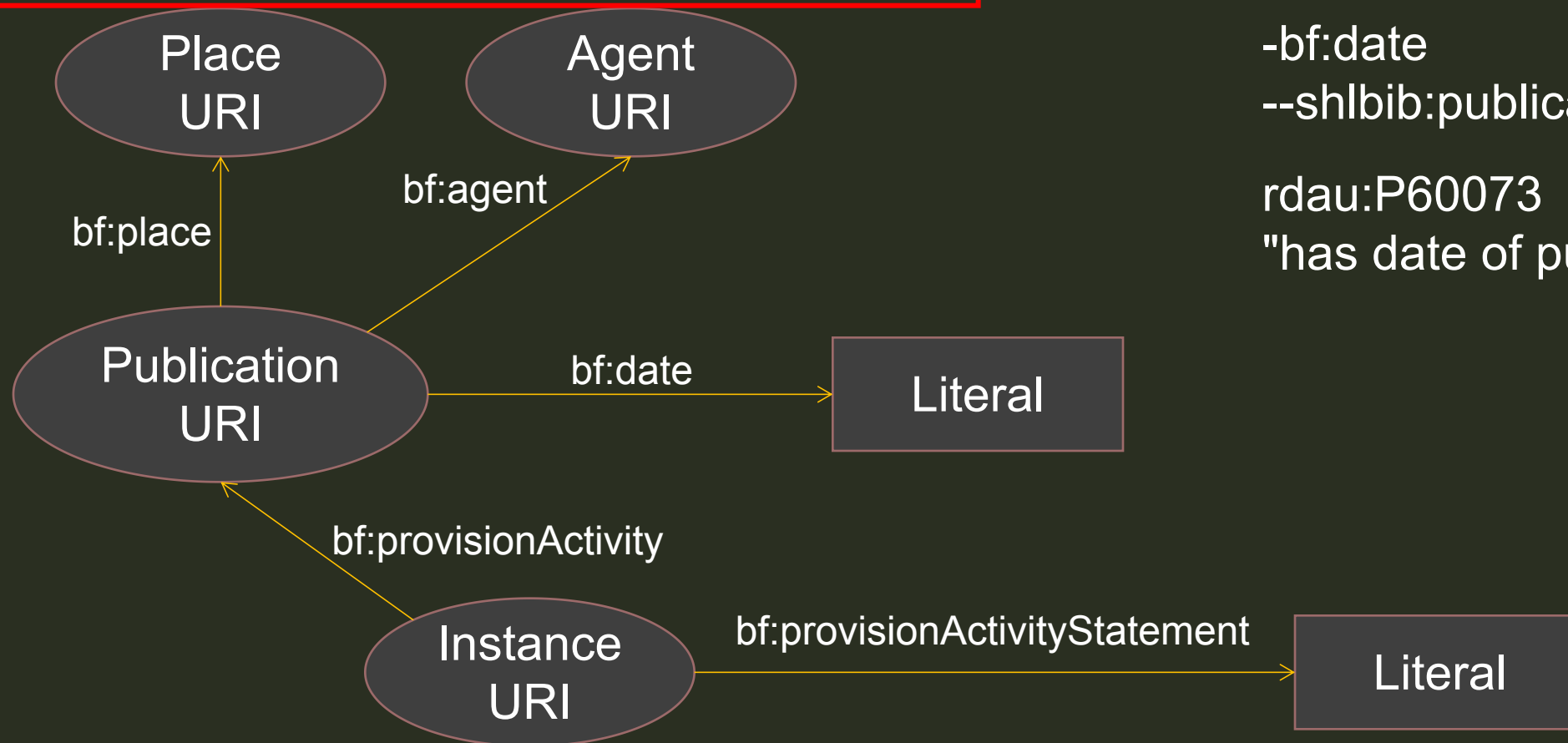
BIBFRAME词汇三元组						CNMARC
subject	predicate	object	subject	predicate	object	字段
Instance	title	Title		rdfs:label	Literal	200\$acdehinv9,225\$adehi
				mainTitle	Literal	200\$a,225\$a
				partNumber	Literal	200\$h,225\$h
				partName	Literal	200\$i,,225\$i
		VariantTitle		rdfs:label	Literal	512-518,540,541,545
				variantType	Literal	
				mainTitle	Literal	\$a
				subtitle	Literal	\$e
		responsibilityStatement	Literal			200\$fg,225\$f
		identifiedBy	Isbn		rdf:value	Literal
				qualifier	Literal	010\$b
	acquisitionTerms	Literal			010\$c	
	editionStatement	Literal			205	
	provisionActivityStatement	Literal			210	
	provisionActivity	ProvisionActivity		place	Place	210\$a
				agent	Agent	210\$c
				date	Literal	100\$a9-16, 210\$d
				rdfs:label	Literal	215\$a
	extend	Extend			215\$d	
	dimensions	Literal			105\$a0-3	
	illustrativeContent	Illustration			327	
	tableOfContents	TableOfContents		rdfs:label	Literal	部分3XX,215\$c,
	note	Note		rdfs:label	Literal	
				noteType	Literal	

# 3.3 Item层映射

BIBFRAME词汇三元组					CNMARC	
subject	predicate	object	subject	predicate	object	字段
Item	shelfMark	ShelfMark		rdfs:label	Literal	905\$s
	electronicLocator					856\$u
	note	Note		rdfs:label	Literal	316-318, 345
				noteType	Literal	

## 3.4 CNMARC-BF映射实例

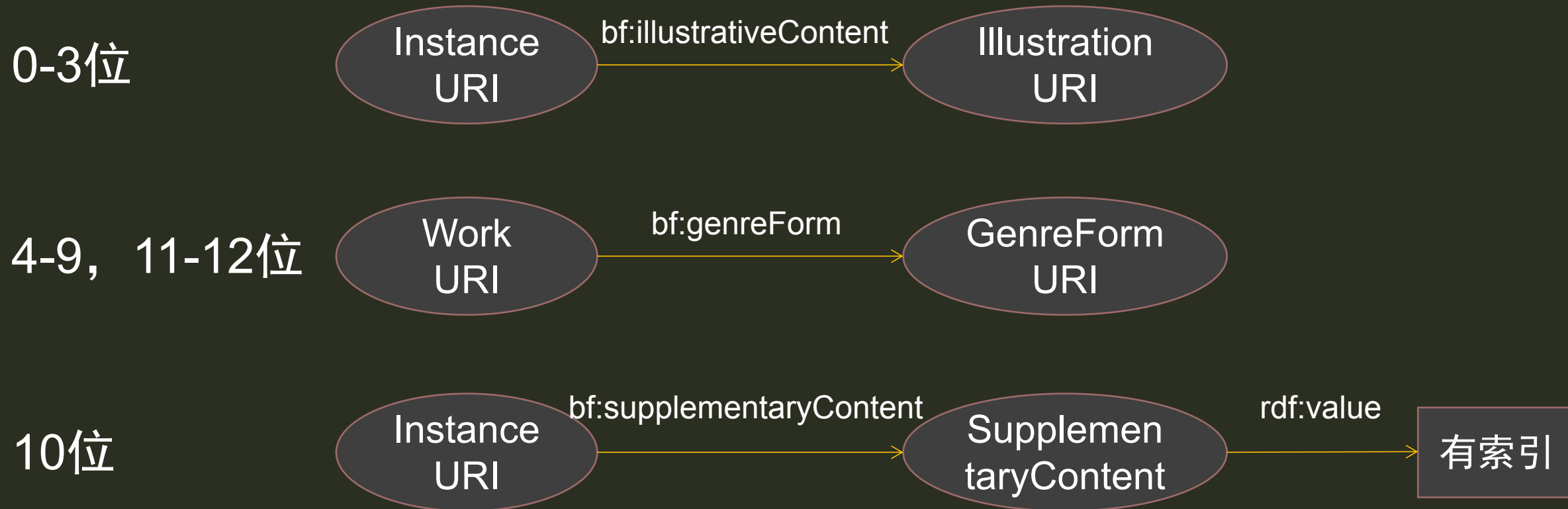
=210 \ \$a北京\$ c科学出版社\$d2012



-bf:date  
--shlbib:publicationDate  
rdau:P60073  
"has date of publication"

## 3.5 CNMARC-BF映射实例

=105 \ \$aa z 000ay



# 4

## 上海图书馆关联数据项目

## 4.1 数字人文项目



上海图书馆  
名人手稿档案库



上海图书馆  
SHANGHAI LIBRARY

盛宣怀档案知識庫

A  
B  
C  
D  
E  
F  
G  
H  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
W  
X  
Y  
Z

G 1811 gan  
先祖名人  
家谱文献 11

1932 gongzhi  
先祖名人  
家谱文献 10

1824 quaejia  
先祖名人  
家谱文献 19

### 中文古籍聯合目錄及循證平台

Chinese Ancient Books Union Catalogue and Evidence-based Platform

在此輸入標題/責任者/分類/館藏地...

☒ 上圖館藏 ☐ 柏克萊館藏

中文古籍聯合目錄及循證平臺是上海圖書館數字人文平臺的一個試驗型項目，目前收錄有上海圖書館和加州柏克萊大學圖書館的善本館藏，以及一些在歷史上有一定影響的官修目錄、史志目錄、私家目錄和版本目錄等，輔之以人名、地名、印章、刻工等額外規範數據，並將結合即將開發完成的內容分析統計、時空及社會關係分析和可視化工具。該項目旨在藉由分佈式雲平臺技術和關聯語義技術，實現各館現存古籍珍藏的聯合查詢、規範控制，並提供學者循證版本、考鏡流藏之功用，未來希望更多的圖書館攜手加入，以嘉惠學林、澤被後人。

Chinese Ancient Books Union Catalogue and Evidence-based Platform is a pilot project for the Shanghai Library digital humanity infrastructure. Currently it includes the collections from the Shanghai Library and the ... [more↓](#)



記  
，是  
錄

2000

## 4.2 规范数据项目

上海圖書館  
人名規範庫

人物

地点

输入姓名/异名/民族/性别/职衔搜索

高级搜索

上海图书馆开放数据平台

本网站是上海图书馆的开放数据平台，将陆续以关联开放数据（Linked Open Data）的方式向互联网公开发布各种术语词表、规范档、书目数据以及各种数据清洗和转换工具等。提供各种数据消费接口供开发人员调用，以促进数据的开发获取、共享和重用。

人名规范库

More...

华人姓氏列表

More...

中国历史纪年表

More...

地理名词表

More...

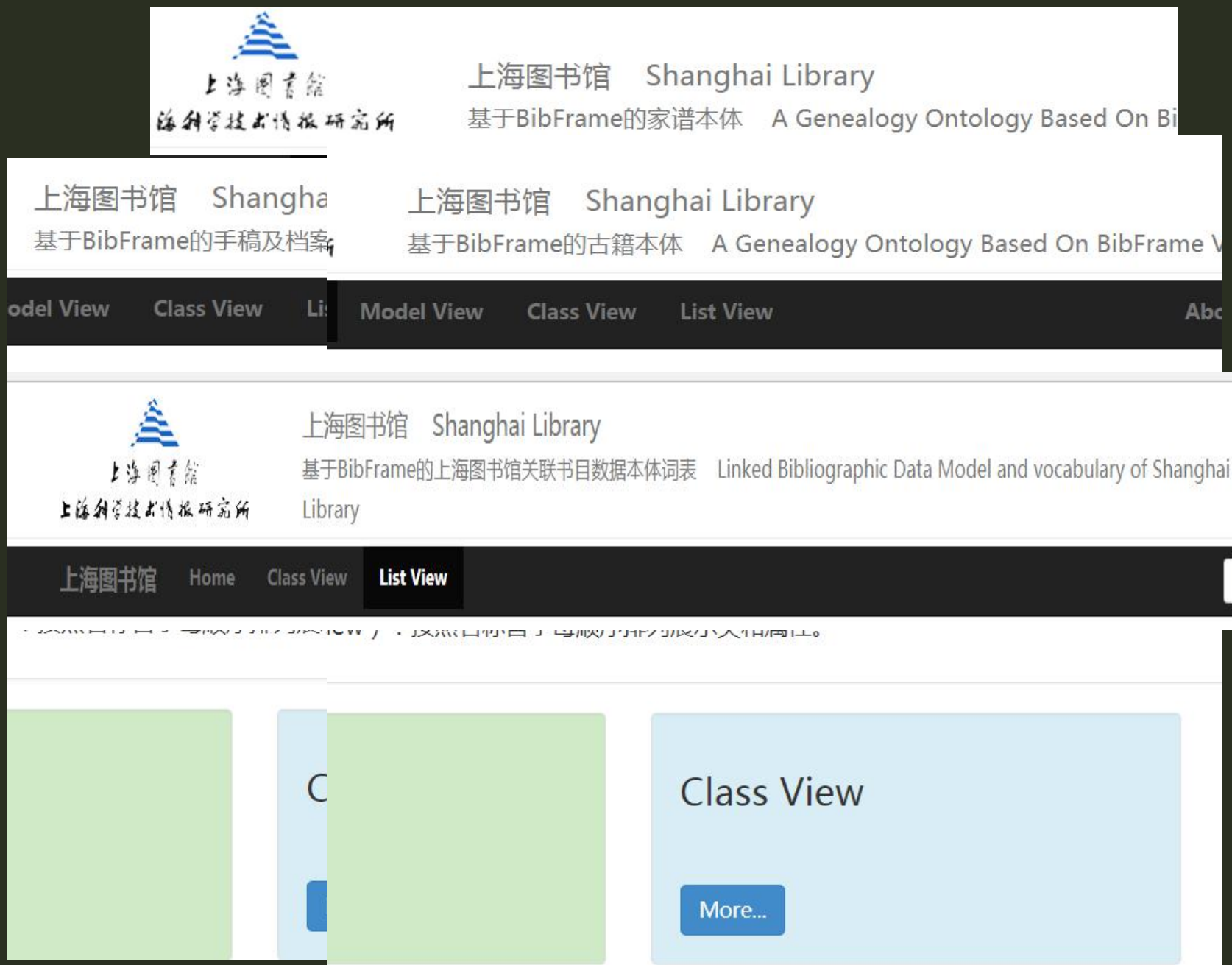
机构名录

More...



## 4.3 本体词表

- 家谱本体
- 手稿及档案本体
- 古籍本体
- 书目本体



## 4.4 关联书目数据项目

- 2017年2月开始讨论将联编书目数据发布为关联数据，探索关联书目数据的应用可能性，实验BIBFRAME词表在中文编目环境中的适用性，为编目技术、流程的变革积累经验与技术。
- 采编中心、系统网络中心协作开展

## 4.5 项目关键

- 联编CNMARC数据分析
- CNMARC与BIBFRAME的映射
- CNMARC数据解析
- 实体识别：作品，人，团体，会议，地点
  - 作品的识别与去重☆（作品的内涵）

## 4.6 项目成果

- CNMARC与Bibframe映射表，发布基于Bibframe的书目本体
- 发布文献类型、责任方式、语种等取值词表
- 发布人、团体、会议、地点等规范实体
- 发布150W关联书目数据
- BF数据编辑器--BibframeEditor（基于LC的BIBFRAME Editor本地化修改）
- 关联书目数据的检索、展示平台

## 4.6 发布平台



## 4.6 词表, 规范数据

上海图书馆 首页 书目 作品 人名 地名 机构团体 取值词表 本体

发行方式 读者对象 文献类型 图表类型 语种

发行方式[3]

集成性资源 TRIPLES  
integrating resource

专著 TRIPLES  
monographic

# 5.3 作品

题名：

☐ 丛书

作者：

☐ 概念检索

查询

读者对象

普通成人(1078362)

青年(1025881)

科研人员(61445)

少年(34491)

小学生(29469)

全部显示

文献类型

小说(54817)

传记(43824)

书目(27240)

散文(25592)

试题集(17791)

全部显示

共匹配到 1111813 条记录

红楼梦

著者：曹雪芹，高鹗著

分类：I242.47

语言：汉语

介绍：作品以贾、史、王、薛四大家族为背景，以贾宝玉、林黛玉的爱情悲剧为主要线索，着重描写了贾家荣、宁二府由盛到衰的过程，广泛反映了当时的社会矛盾和阶级斗争，对封建礼教进行了批判。

相关书目，匹配到[29]条

[1] 孙温绘《全本红楼梦》 - (清)曹雪芹，(清)高鹗原著；(清)孙温绘

北京：现代出版社，2016

RDF

[2] 红楼梦注释 - (清)曹雪芹著；贾太宏译注

北京：西苑出版社，2016

RDF

[3] 红楼梦 - 曹雪芹，高鹗著

长春：长春出版社，2016

RDF

[4] 红楼梦 - (清)曹雪芹著

北京：煤炭工业出版社，2016

RDF

[5] 红楼梦 - (清)曹雪芹，(清)高鹗著

哈尔滨：哈尔滨出版社，2016

RDF

[6] 红楼梦：一百二十回全本 - 曹雪芹，高鹗著

长春：时代文艺出版社，2016

RDF

[7] 红楼梦：乾隆间程甲本 - (清)曹雪芹著

北京：中国书店，2015

RDF

## 5.4 概念检索

☐ 丛书      作者：茅盾      ☒ 概念检索

共匹配到 60 条记录



茅盾文集

☐ 丛书      作者：沈雁冰      ☐ 概念检索

共匹配到 0 条记录

☐ 丛书      作者：沈雁冰      ☒ 概念检索

匹配到 60 条记录



茅盾文集



# 5.4 书目

上海图书馆

首页

书目

作品

人名

地名

机构团体

取值词表

本体

题名：

作者：

☐ 概念检索

出版年： ~

识别码：

出版年

共匹配到 1465455 条记录

9999(184)

9941(1)

3201(1)

3014(1)

3011(1)

▼全部显示

出版地

北京(734876)

上海(93770)

武汉(39534)

南京(31020)

广州(29087)

▼全部显示

传感器与信号处理

传感器

著者：

出版：

尺寸：

页码：

定价：

素食与健康

素食与

著者：

出版：

尺寸：

页码：

定价：

脂砚斋重评石头记

庚辰本【一】

石头记

题名：脂砚斋重评石头记：庚辰本

著者：曹雪芹(著)

版本：[影印本]

出版者：人民文学出版社

出版地：北京

出版年：2010

尺寸：22cm

页码：4册(1973页)

定价：CNY290.00

丛编：红楼梦古抄本丛刊

识别码：

1. (A100000NLC)004483271[OSCN]

2. (A310000SHL)012010036101[OSCN]

RDF

6

小结

文献 → 数据

载体 → 内容

# 谢谢

许磊 [leixu@libnet.sh.cn](mailto:leixu@libnet.sh.cn)