

# 轻量级书目本体关联数据 建设实践

郭振英

北京工业大学图书馆

2015-11-6

# 内容

问题的提出

技术路线

实践思考

# 问题的提出

- ▶ 关联数据功能：
- ▶ 帮助图书馆自身融入数据网络，扩大资源范围，强化与外界信息的共享与交换，提升服务能力。

# 问题的提出

▶ 关联数据离我们有多远？

# 问题的提出

## ▶ 当前高校图书馆探索举例：

利用OPAC 系统分别同时向DBpedia、上海图书馆和本地数据的SPARQL端点发送数据查询请求，然后获取返回的数据，并通过脚本程序实时显示在OPAC 的页面中。

— 东莞职业技术学院图书馆 钟远薪

# 问题的提出

## ▶ 当前高校图书馆探索举例：

提取有效的MARC字段,从Wiki等网站获取相关信息,参照FRBR模型,对书目数据进行实体属性的层次解析,利用SQL Server、D2R等工具实现部分书目数据语义化

-西安航空学院图书馆 樊秋妮

# 问题的提出

高校图书馆：

作为关联数据发布者和消费者，有必要积极探索、开展关联数据的应用实践，培育图书馆的关联数据环境，立足用户需求，从关联数据的角度定义和发现资源。

# 技术路线

构建轻量级书目本体

基于轻量级书目本体重新组织书目数据

通过D2Rserver发布关联数据

利用关联数据进行内部信息聚合和外部数据发现



# 技术路线

- ▶ **选择合适的书目本体**
- ▶ **馆藏书目数据关联数据化前提：**
- ▶ 书目数据采用基于MARC建立的一维线性组织形式，语义化程度较低，和语义网的发展方向相悖的。把馆藏书目数据发布为关联数据是对书目数据重新组织的过程，是馆藏书目融入语义网的重要环节。

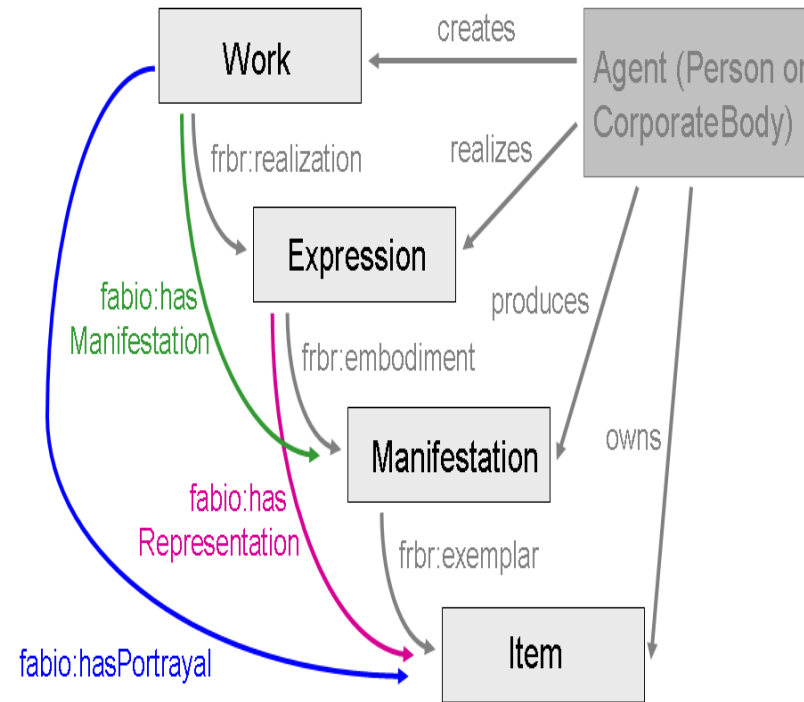
## 技术路线

### ► 书目本体

### ► BIBO、FaBIO

- FaBIO在2013年3月发布的1.7.3版本中，共定义了256
- 象属性和63个数据属性

New Work-Expresson-Manifestation-Item relationships in FaBiO



内容庞杂，不利于实际应用

Classes

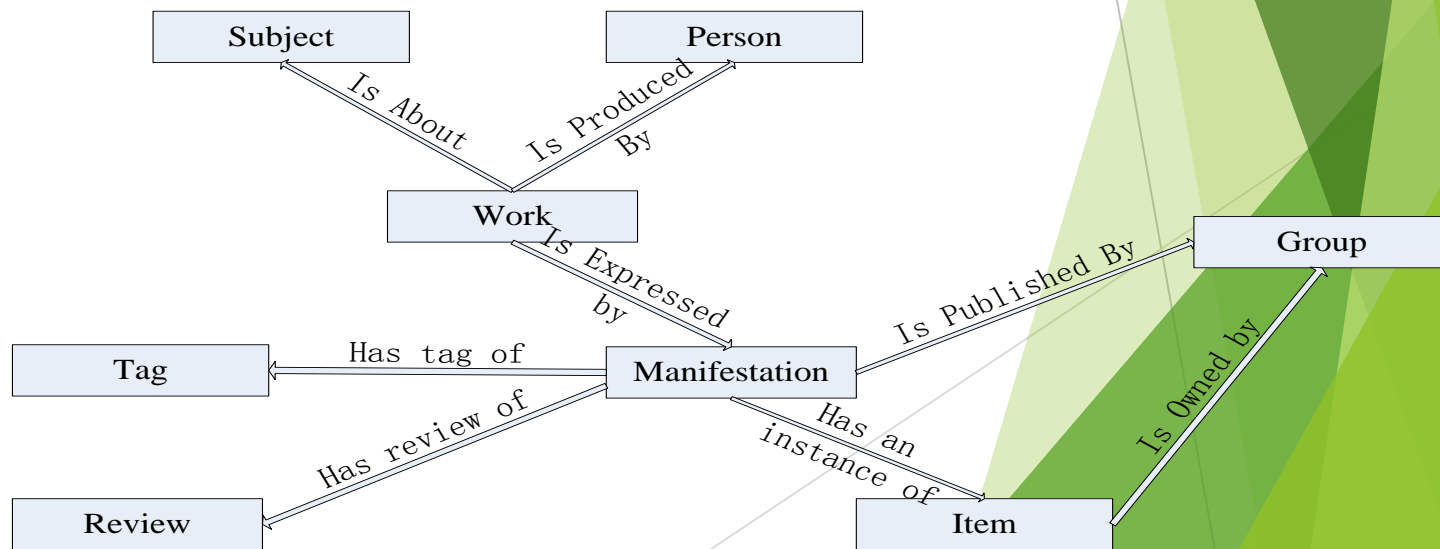
[abstract](#) [academic proceedings](#) [addendum](#) [algorithm](#) [analog item](#) [analog manifestation](#) [analog storage medium](#) [announcement](#) [anthology](#) [API](#) [application profile](#) [article](#) [artistic work](#) [audio document](#) [authority file](#) [bachelor's thesis](#) [bibliographic database](#) [bibliographic metadata](#) [biography](#) [blog](#) [blog post](#) [book](#) [book chapter](#) [book review](#) [book series](#) [brief report](#) [call for applications](#) [case for support](#) [case for support document](#) [case report](#) [catalog](#) [chapter](#) [citation metadata](#) [clinical case report](#) [clinical guideline](#) [clinical trial design](#) [clinical trial report](#) [collected works](#) [comment](#) [complete works](#) [computer application](#) [computer file](#) [computer program](#) [concept](#) [concept scheme](#) [conference paper](#) [conference poster](#) [conference proceedings](#) [controlled vocabulary](#) [correction](#) [corrigendum](#) [cover](#) [critical edition](#) [data file](#) [data management plan](#) [data management policy](#) [data management policy document](#)

# 技术路线

## ► 轻量级书目本体构建

- 考虑用户认知和简化层次，结合FRBR模型，采纳用户熟悉的元素，并考虑社会化元数据，构建基于八个基本类的轻量级书目本体：
- “作品”、“载体表现”、“单件”，“团体”，“主题”，“个人”，“书评”，“标签”。

项目	包含元素
简 明 书 目 记录元素	管理记录号、出版年、书目索书号、作者、题名、出版社、ISBN
OPAC 详细 页面元素	题名、著者、ISBN、作品语种、出版发行、载体形态、主题、分类号、馆藏地
分面分类	语种、年份、主题词、馆藏、分类、作者、关键词、格式
外部API揭示	封面、目次、书评、摘要、随书光盘



# 技术路线

## ▶ 书目数据重新组织

- ▶ 基于轻量级书日本体重组书目数据
- ▶ 构建数据库Schema “U\_LINKDATA”
- ▶ 选择使用率较高的TP3类别中西文MARC共58304条，聚合作品类的规范记录共55444条，并与单件记录208921条建立关联关系。

名称	数据类型	大小	小数位	可否为空?	默认值
WORK_ID	CHAR		9		
WORK_URL	VARCHAR2	255		✓	
WORK_TITLE	VARCHAR2	255		✓	
WORK_AUTHOR	VARCHAR2	255		✓	
WORK_SUBJECT	VARCHAR2	255		✓	
WORK_CREATEDATE	CHAR	14		✓	
WORK_LASTUPDATE	CHAR	14		✓	
WORK_NOTE	VARCHAR2	255		✓	

# 技术路线

## ▶ 基于 D2RServer 关联数据发布

- ▶ D2RServer 提供了基于映射文件将数据库发布成 RDF 三元组的机制，构建 D2R 环境，下载与目前使用的数据库版本相匹配的驱动（如数据库服务器 oracle11g 11.1.0.7 上 jdbc 的驱动文件 ojdbc6.jar），将其 jar 文件放在 D2R Server 的 lib 目录下面，通过命令生成映射文件启动 D2R Server

# 技术路线

Snorql: Exploring http://localhost:2020/sparql

**SPARQL:**

```
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX rdvocab: <http://rdvocab.info/Elements/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xc: <http://www.extensiblecatolog.info/Elements/>
PREFIX ifla-frbr: <http://iflastandards.info/ns/fr/frbr/frbrer/>
PREFIX rdarole: <http://RDVocab.info/roles/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rel: <http://purl.org/vocab/relationship/>
PREFIX vocab: <http://localhost:2020/vocab/resource/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX map: <file:/c:/d2r-server-0.7/linkdata-mapping.ttl#>
PREFIX db: <http://localhost:2020/resource/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
```

```
SELECT DISTINCT ?manif_id ?manif_title ?manif_author ?manif_lang ?manif_version ?manif_publisher
| ?manif_publisheryear ?manif_isbn WHERE {
?manif vocab:M_INFO_WORK_ID ?work_id. FILTER (?work_id = "000026448")
?manif dc:title ?manif_title.
?manif dc:creator ?manif_author.
?manif dcterms:language ?manif_lang.
?manif vocab:M_INFO_MANIF_VERION ?manif_version.
?manif vocab:M_INFO_MANIF_PUBLISHERINFO ?manif_publisher.
?manif vocab:M_INFO_MANIF_PUBLISHYEAR ?manif_publisheryear.
```

Results:

**SPARQL results:**

manif_id	manif_title	manif_author	manif_lang	manif_version	manif_publisher	manif_publisheryear	manif_isbn
"000024724"	"Java编程思想 = Thinking in Java"	"埃克尔,"	"chi"	"第2版"	"机械工业出版社"	"2002"	"7-111-10441-2"
"000509017"	"Java编程思想 : 评注版 = Thinking in Java"	"埃克尔"	"chi"	"第4版"	"电子工业出版社"	"2011"	"978-7-121-13521-7"
"000410570"	"Thinking in Java = =Java编程思想"	"Eckel, Bruce."	"eng"	"4th ed."	"China Machine Press,"	"2007"	"9787111212508"
"000210205"	"Thinking in Java = =Java 编程思想"	"Eckel, Bruce."	"eng"	"3rd ed."	"China Machine Press,"	"2003"	"7-111-13165-7 :"
"000293260"	"Java编程思想 = Thinking in Java"	"埃克尔,"	"chi"	"4th ed."	"机械工业出版社"	"2007"	"978-7-111-21382-6"
"000383394"	"Thinking in Java = =Java编程思想"	"Eckel, Bruce."	"eng"	"2nd ed."	"China Machine Press,"	"2002"	"7111092864 :"

# 技术路线

## ► 关联数据内部信息聚合和外部知识发现

通过建立基于本体的关联数据重新组织书目数据，实现内部聚合，可以揭示同一主题、同一著者、同一版本、同一语种等本馆馆藏，尤其是不同MARC格式的版本聚合，例如可以检索第4版本的“thinking in java”共2种，合并了OPAC中文文献和西文文献不同页面的揭示。

通过“owl: sameAs”等关系可建立和DBpedia的连接，实现对著者的进一步揭示。在DBpedia 关联数据集中对人名URI命名规则是“http: // dbpedia.org / page / 人名”，中国人人名是以“\_”分隔的姓名拼音。对于“thinking in java”的作者Eckel, Bruce的URI为[http://dbpedia.org/page/Bruce\\_Eckel](http://dbpedia.org/page/Bruce_Eckel)，建立URIs相关属性相似性连接，根据作者名称构建链接指向DBpedia的著者信息，可进行外部著者信息的发现。

# 实践思考

- ▶ 馆藏书目本体的标准化是关联数据实现的前提。
- ▶ 规范数据是实现信息聚合的基础，非一馆可以实现。
- ▶ 关联数据建设如何和当前图书馆自动化集成系统融合。



敬请各位老师批评指正！

谢谢！