

# RDF数据生成与转换工具介绍

数字人文项目开发工程师 李一斌

2017.5.10



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。

RDB2RDF

More...

ETL

More...

# RDB2RDF

wySwyG  
R2R

从关系数据库到关联数据  
国家自然科学基金青年项目13CTQ008成果之一

- 数据存储于RDB中，关联信息难以利用
- D2R工具：配置复杂，学习成本高
- RDB2RDF
- What you see is what you get
- 通过图形化配置的方式进行关系型数据到关联数据的转换

# RDB2RDF

wySwyG  
R2R

从关系数据库到关联数据  
国家自然科学基金青年项目13CTQ008成果之一

- 图形化配置
- RDF结果预览，下载
- 图数据可视化
- Sparql查询

# RDB2RDF

wySwyG  
R2R

从关系数据库到关联数据  
国家自然科学基金青年项目13CTQ008成果之一

- 图形化配置
  - 数据库连接配置
    - 数据库地址
    - 数据库名称
    - 用户名密码
  - D2R映射配置
  - RDF格式配置
    - RDF/XML
    - Ntriples
    - N3
    - Turtle

# RDB2RDF

wySwyG  
R2R

从关系数据库到关联数据  
国家自然科学基金青年项目13CTQ008成果之一

- D2R映射配置
  - Direct Mapping
    - 资源uri前缀



A Direct Mapping of Relational Data to RDF

W3C Working Draft 29 May 2012

- 通过数据主键来建立资源uri
- 通过字段名来建立谓词，字段值来建立数值型宾语
- 通过外键来建立资源间的关系

# RDB2RDF

People

PK		→ Address(ID)
ID	fname	addr
7	Bob	<a href="#">18</a>
8	Sue	NULL

Addresses

PK		
ID	city	state
18	Cambridge	MA

```
@base <http://foo.example/DB/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<People/ID=7> rdf:type <People> .
<People/ID=7> <People#ID> 7 .
<People/ID=7> <People#fname> "Bob" .
<People/ID=7> <People#addr> 18 .
<People/ID=7> <People#ref-addr> <Addresses/ID=18> .
<People/ID=8> rdf:type <People> .
<People/ID=8> <People#ID> 8 .
<People/ID=8> <People#fname> "Sue" .

<Addresses/ID=18> rdf:type <Addresses> .
<Addresses/ID=18> <Addresses#ID> 18 .
<Addresses/ID=18> <Addresses#city> "Cambridge" .
<Addresses/ID=18> <Addresses#state> "MA" .
```

# RDB2RDF

wySwyG  
R2R

从关系数据库到关联数据  
国家自然科学基金青年项目13CTQ008成果之一

- D2R映射配置
  - R2RML
  - 通用的映射配置



R2RML: RDB to RDF Mapping Language

W3C Working Draft 29 May 2012

- 资源uri配置
- 谓语句uri配置
- 资源间关系配置



# RDB2RDF

wySwyG  
R2R

从关系数据库到关联数据  
国家社会科学基金青年项目13CTQ008成果之一

EMP

EMPNO	ENAME	JOB	DEPTNO
INTEGER PRIMARY KEY	VARCHAR(100)	VARCHAR(20)	INTEGER REFERENCES DEPT (DEPTNO)
7369	SMITH	CLERK	10

DEPT

DEPTNO	DNAME	LOC
INTEGER PRIMARY KEY	VARCHAR(30)	VARCHAR(100)
10	APPSERVER	NEW YORK

```
@prefix rr: <http://www.w3.org/ns/r2rml#>.
```

```
<#TriplesMap1>  
  rr:logicalTable [ rr:tableName "EMP" ];  
  rr:subjectMap [  
    rr:template "http://data.example.com/employee/{EMPNO}";  
    rr:class ex:Employee;  
  ];  
  rr:predicateObjectMap [  
    rr:predicate ex:name;  
    rr:objectMap [ rr:column "ENAME" ];  
  ].
```

```
<http://data.example.com/employee/7369> rdf:type ex:Employee.  
<http://data.example.com/employee/7369> ex:name "SMITH".  
<http://data.example.com/employee/7369> ex:department <http://data.example.com/department/10>.
```

```
<http://data.example.com/department/10> rdf:type ex:Department.  
<http://data.example.com/department/10> ex:name "APPSERVER".  
<http://data.example.com/department/10> ex:location "NEW YORK".  
<http://data.example.com/department/10> ex:staff 1.
```

表名

ancestor

主语配置

主语uri规则

主语类别

http://data.library.sh.cn sh:Person

谓词短语配置

谓词

+1 foaf:name

宾语生成方式

指定列

宾语值

name

宾语修饰方式

无

# RDB2RDF

wySwyG 从关系数据库到关联数据  
R2R 国家自然科学基金青年项目13CTQ008成果之一

- 操作演示

# ETL

Any2RDF  
ETL

上海图书馆开放数据平台  
数据采集转换工具

- 互联网数据质量差，难以利用
- 关联数据构建复杂
- 集成数据采集、加工、转换的一站式Web解决方案

数据采集

数据加工

数据转换

# ETL

- 数据采集
  - 百度百科，互动百科
  - CBDB（中国历代人物传记）

李清照是一个多义词，请在下列义项上选择浏览（共12个义项）

- 宋代女词人
- 2014年王超导演话剧《李清照》
- 汪苏泷演唱歌曲

[lǐ qīng zhào] ◆

## 李清照（宋代女词人） [编辑](#)

李清照（1084年3月13日—1155年？），号易安居士，汉族，齐州济南（今山东省济南市章丘区）人。宋代女词人，婉约词派代表，有“千古第一才女”之称。

李清照出生于书香门第，早期生活优裕，其父李格非藏书甚富，她小时候就在良好的家庭环境中打下文学基础。出嫁后与夫赵明诚共同致力于书画金石的搜集整理。金兵入据中原时，流寓南方，境遇孤苦。所作词，前期多写其悠闲生活，后期多悲叹身世，情调感伤。形式上善用白描手法，自辟途径，语言清丽。论词强调协律，崇尚典雅，提出词“别是一家”之说，反对以作诗文之法作词。能诗，留存不多，部分篇章感时咏史，情辞慷慨，与其词风不同。

有《易安居士文集》《易安词》，已散佚。后人有《漱玉词》辑本。今有《李清照集校注》。<sup>[1-3]</sup>

### 人物关系

编辑



丈夫 赵明诚



父亲 李格非

本名	李清照	去世时间	约1155年
字号	号易安居士	主要作品	《声声慢 寻寻觅觅》《一剪梅 红藕香残玉簟秋》《夏日绝句》等
所处时代	宋朝	主要成就	诗词创作、诗词理论最佳有“千古第一才女”之称
民族族群	汉族	职业	诗人、词人
出生地	山东省济南市章丘区明水街道	配偶	赵明诚
出生时间	1084年3月13日		

# ETL

- 数据加工
  - 数据抽取
    - 字段名
    - 字段值来源（infobox，摘要，标签，正则，jsonpath）
  - 数据过滤
    - 基于规则的数据筛选
  - 多源数据整合
    - 自定义优先级
  - 数据加工实时结果预览

# ETL

Any2RDF  
ETL

上海图书馆开放数据平台  
数据采集转换工具

- 数据转换，导出
  - Excel，CSV数据导出，上传
  - RDF数据转换
    - 类R2RML的RDF配置
    - 多格式下载

# ETL

**Any2RDF** 上海图书馆开放数据平台  
ETL 数据采集转换工具

- 操作演示