DOI: CNKI:11 - 2746/G2. 20121105. 2336. 001

关联数据发布流程与关键问题研究
——以科技文献、科学数据发布为例*

沈志宏　刘筱敏　郭学兵　张晓林　黎建辉

摘　要　本文结合中国科学引文数据库和中国生态系统研究网络通量数据的关联数据发布，以关联数据的发布技术框架为研究对象，采取实例阐释的方法，提出了关联数据发布过程中可参考的标准化流程，并详细分析了其中的关键问题。研究表明，关联数据发布流程可以分解成数据建模、实体命名、实体 RDF 化、实体关联化、实体发布、开放查询六个关键步骤，发布过程中需要考虑到多语种问题、值词表的发布、RDF 词表的发布等关键问题。关于利用 D2R Server 发布数据，本文建议：不要采用空白节点；尽量做好关系型数据库的前期设计；指定非文本属性的数据类型；适当进行实体表的拆分与合并。图 5。表 3。参考文献 31。

关键词　关联数据　关联开放数据　数据发布　科技文献　科学数据　D2R Server

分类号　G350　TP393

A Research on Publishing Workflow and Key Issues of Linked Data: Experience with Publishing Scientific Literature and Scientific Data as Linked Data

Shen Zhihong, Liu Xiaomin, Guo Xuebing, Zhang Xiaolin & Li Jianhui

ABSTRACT　Combined with the publication of linked data in Chinese Science Citation Database (CSCD) and the flux data in Chinese Ecosystem Research Network (CERN), this paper studies the standardized publishing workflow of linked data and key issues in such process by studying on the technical framework of publishing linked data with examples. The standardized publishing workflow of linked data consists of such six critical steps as data modeling, entity naming, turning entity into RDF, make connections between entities, entity publication and open-type inquiry. Some key issues need to be considered in the publishing process, such as dealing with multilingual values, publishing value vocabularies and publishing RDF vocabularies. In addition, this paper gives some suggestions in case D2R Server is used to publish linked data: not using blank note; trying the best to make pre-design of relational database better; specifying the data type of non-text attributes; splitting and combining table entities properly. 5 figs. 3 tabs. 31 refs.

KEY WORDS　Linked data.　Linking open data.　Data publishing.　Scientific literature.　Scientific data.　D2R server.

1　引言

随着关联数据理念的推广和关联开放数据运动的不断深入，越来越多的信息库采取了关联数据的形式对外发布并提供访问服务。发布关联数据的途径往往因资源内容的特征分成三种[1]：如果数据量很小（几百条 RDF 三元组或者更少），可以直

*　本文系中国科学院信息化专项“科技数据资源整合与共享工程”（XXH12504）和国家科技基础条件平台建设项目“基础科学数据共享网—理化天文空间生物”课题“标准规范及共享服务平台建设”（BSDN2009 - 17）的研究成果之一。

通讯作者：沈志宏，Email: shenzhihong@ mail. las. ac. cn

接采用静态的 RDF 文件(静态发布);如果数据量很大,则需要将它们放进 RDF 库中,并选择 Pubby[8] 服务器作为关联数据服务的前端;如果数据的更新频率很大,就需要引入更新机制,或者在请求数据的时候再根据原始数据在线生成(on-the-fly translation) RDF。其中的第三种方式,即在线映射,往往会借助于一些映射工具,如:D2RQ 平台[9]、Virtuoso RDF Views[4]、Sparqlify[5] 等。此外,W3C 还有一个 RDB2RDF 工作组[6],从事 R2RML 映射语言的研究[7]。

很多传统的信息都借助于关系型数据库进行存储,另外由于 D2RQ 的使用简单方便,因此 D2RQ 在很多场合都得到了应用。作为关联数据发布的标准教程,文献"*How to Publish Linked Data on the Web*"就重点推荐了 D2RQ,并介绍了它的软件架构和使用方法。D2RQ 平台包括 D2R Engine、D2R Server 以及 D2RQ 映射语言,基于 D2R Server,目前已经有很多数据源对外开放了关联数据的接口(即变成了关联数据的数据集),如 DBLP 书目库[8]、CIA Factbook[9]、欧洲国家地区统计信息库[10] 等。

采用 D2R Server 将关系型数据发布成关联数据,软件操作流程如下:

① 准备 Java 环境,下载某个版本的 D2R Server,如:d2rq-0.8;

② 执行 generate-mapping 工具,连接至数据库,生成映射文件,如:mapping.n3;

③ 根据发布需求,在以上生成的 mapping.n3 文件基础上进行修改与完善;

④ 以 mapping.n3 为参数,启动 d2r-server。

经过以上简单四步,D2R Server 即可提供关联数据的访问服务,这些服务包括:用户可以浏览某类实体的列表(directory),并通过每个 URI 访问到某一条实体的网页描述和 RDF 描述。同时,D2R Server 还提供了 SPARQL 查询接口,并提供了一个 Web 化的执行界面(SPARQL Explorer),用户可以在该界面输入 SPARQL 查询语句,并查看到执行结果。

在多年的科研活动中,科研人员和科研设备产生了大量宝贵的科技文献资源和科学数据资源。近年来出版界对科学数据出版(data publication)、语义出版(semantic publishing)的关注,也突出了科研人员对科学数据的开放访问以及与科技文献互操作的强烈需求。可以看到,在关联开放数据的环境下,大量的图书馆资源和科学数据及其描述信息开始以关联数据的方式提供开放访问服务,图书馆资源如美国国会图书馆发布的 LCSH 的 SKOS 版本[11]、瑞典联合目录 LIBRIS[12]、德国国家图书馆发布的规范文档[13] 等,科学数据资源如 Linked Life Data[14]、Diseasome Map[15]、Linked Sensor Data[16] 等。作为知识的不同载体,科技文献与科学数据无论在内容上还是在语义描述模型上都有各自的特点,另外科学数据还带有强烈的学科领域特征。在这种背景下,如何基于关联数据机制,研究以标准化流程完成这些科研资源的发布,并推广至其他类型的科研资源,具有切实的指导意义。本文以科技文献和科学数据的发布为例,研究基于 D2R Server 发布关联数据的流程和其中的关键问题。其中,科技文献选取了中国科学引文数据库(Chinese Science Citation Database,CSCD)[17] 的部分论文元数据(以下简称 CSCD-SW),科学数据选取了中国生态系统研究网络(Chinese Ecosystem Research Network,CERN)[18] 的长期观测通量数据的描述信息(以下简称 FLUX-SW)。

2 关联数据发布

2.1 发布原则与流程

Tim Berners-Lee 制订了关联数据关于内容描述的四项基本原则[19]:

① 使用 URI 来标识事物;

② 使用 HTTP URI 使人们可以访问到这些标识;

③ 当有人访问到标识时,提供有用的信息;

④ 尽可能提供关联的 URI,以使人们可以发现更多的事物。

对应四项要求,本文认为,包括科技文献和科学数据在内,各类信息的关联数据化发布,可以分解成六个关键步骤,即数据建模、实体命名、实体

RDF 化、实体关联化、实体发布、开放查询，每个步骤的含义解释如下：

① 数据建模：选取待发布的实体，选择或设计 RDF 词表（RDF vocabulary），定义待发布实体之间的语义关系；

② 实体命名：为每个实体赋予一个永久的 URI（Cool URI）[20]；

③ 实体 RDF 化：采用 RDF 来描述每一个实体；

④ 实体关联化：采用 RDF link 来描述实体之间的关联；

⑤ 实体发布：配置发布服务器，负责解析每个实体的 URI，并根据内容协商原则（Content Negotiation）返回正确的网页描述和 RDF 描述；

⑥ 开放查询：配置 SPARQL 服务端（SPARQL endpoint），对外开放 SPARQL 语义查询接口[21]；

各步骤的实现目标及该步骤的阶段性产出如表 1 所示。

表 1 关联数据发布的关键步骤

步 骤	阶段目标	预期产出
数据建模	使实体的关联结构形式化、规范化	数据模型、RDF 词表
实体命名	使每个实体具有一个“Web 上可访问”的名字	实体命名规范
实体 RDF 化	使实体的描述达到“程序可理解”	实体的语义描述
实体关联化	使数据集具有跨实体发现的能力	
实体发布	使实体的描述达到“Web 上可访问”	实体的 HTTP 访问接口
开放查询	使数据集具有语义查询的能力	数据集的查询接口

2.2 词表、值词表、RDF 词表

在各内容对象 RDF 化的过程中，常常需要用到一些标准化的词汇，这些词汇可以来源于多个词表（mixed vocabularies）。在发布 RDF 数据的时候，应尽量采用一些已经为人熟知的 RDF 词表所定义的词汇，如：dc: title。然后在这些已有词表的基础上，根据实际需求进行扩展。

图书馆关联数据孵化小组（Library Linked Data Incubator Group）[22] 明确将词表区分成值词表（value vocabulary）和 RDF 词表（RDF vocabulary），值词表即传统的“词表”，如：LCSH 主题词表、AGROVOC 叙词表、GeoNames 给出的地理位置名称等。发布 RDF 属性值的时候会用到值词表，图书馆管理数据符号小组的报告列举出已发布成关联数据的不同用途的值词表[22]（见表 2）。

表 2 已关联数据化的值词表

分 类	词 表
分类体系	杜威十进制分类法 DDC、通用十进制分类法 UDC
主题词表	美国国会图书馆主题词表 LCSH、法国国立图书馆主题词表 RAMEAU、德国国立图书馆受控词表 SWD、日本国立国会图书馆主题词表 NDLSH
名称规范数据	虚拟国际规范档 VIAF、艺术家联合名录 ULAN、美国国会图书馆名称规范档 LC/NAF、地名数据库 GeoNames
叙词表	STW 经济学叙词表、AGROVOC、Eurovoc、图形材料叙词表 TGM
其他受控词表	DCMI 类型词表、MARC 编者关系代码表、PRONOM、CC 许可集等
其他资源	Wordnet、Freebase、DBPedia 等

与图书馆领域中传统的“词表”不同，RDF 词表用以指定在采用 RDF 描述实体信息时所使用的词汇。W3C 给出 RDF 词表的定义[23]：在语义网中，RDF 词表用来描述和代表关注领域的概念和关系（即 RDF 类名和属性名）。并且认为：词表和本体之间并没有严格的区别，通常采用本体来指代

复杂的、比较正式的词汇集，而词表则意味着较为宽松的要求。

RDF 词表大部分来自于各元数据元素集，通常采用 RDF Schema (RDFS) 和 OWL 本体语言 (OWL Web Ontology Language) 建模语言提供的结构进行描述。比较常用的 RDF 词表如 Dublin-Core、SKOS、FOAF、FRBR 等，图 1 表示常见的 RDF 词表标签云图[14]。

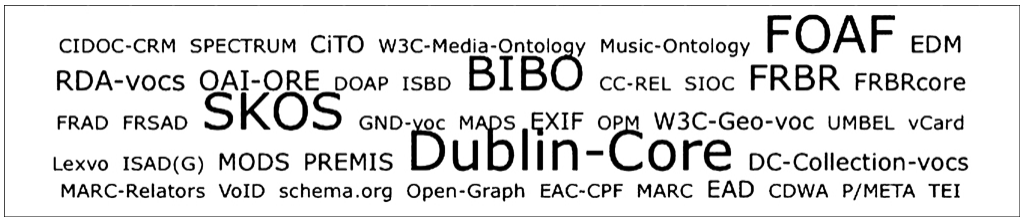

图 1 RDF 词表标签云图

2.3 CSCD-SW 信息的发布

参照上文总结的关联数据发布流程，CSCD-SW 信息的发布流程如下:

① 数据建模

选取的实体内容主要包括: 论文、人员、机构、基金项目、地址、期刊。

关于文献信息的元数据标准比较丰富，CSCD-SW 采用了 Dublin-Core、DC-TERMS、PRISM 多个词表描述论文信息，另外采用了 VCARD 和 FOAF 词表描述人员和地址信息，采用了 ARPFO[15] 词表描述基金项目的信息。此外，CSCD-SW 还自定义了一个 ARA 词表，用以描述论文与基金项目之间的产出关系。

CSCD-SW 各实体之间的关联关系如图 2 所示。

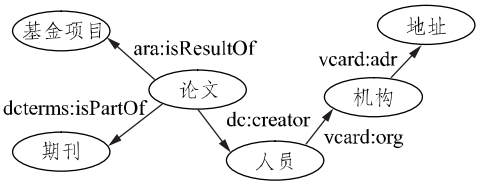

图 2 CSCD-SW 实体关联模型

②实体命名

分配每个实体的 URI 格式如下:

<baseUri> / <entityTypeName> / <entityId>

确定基地址(baseUri) 为 http://semweb.csdb.cn/cscd，因此，ID 为 3490804 的论文 (entityTypeName 为 article) 具有如下 URI:

http://semweb.csdb.cn/cscd/resource/article/3490804

③实体 RDF 化

以论文的元数据为例，其包含的基本信息包括标题、摘要、关键字等，图 3 表示(见下页) 某篇论文的 RDF 图。

④实体关联化

根据图 2 添加实体之间的 RDF link。基金项目与论文、论文与人员之间的关系主要通过关系型数据库 E-R 模型中的主外键关联关系映射得到。

此外，CSCD-SW 同时将每篇论文的分类代码描述成 RDF link，如: 原分类代码 Q948 被改造成一个指向 http://semweb.csdb.cn/cs/resource/clc/Q948 的链接。图 4(见下页) 表示添加了多个 RDF link 的论文元数据 RDF 图。

⑤实体发布

选取 D2R Server，开放每一个实体的访问接口。CSCD-SW 严格遵循了内容协商的原则: 当发现 HTTP 请求头中包含 Accept: text/html(通常由网页浏览器发出) 时，返回普通的 HTML 页面; 而当发现 HTTP 请求头中包含 Accept: application/rdf+xml(通常由关联数据的消费程序发出) 时，返回 RDF/XML 为描述语言的 RDF 描述。

⑥开放查询

CSCD-SW 开放了 SPARQL 查询接口，地址如下:

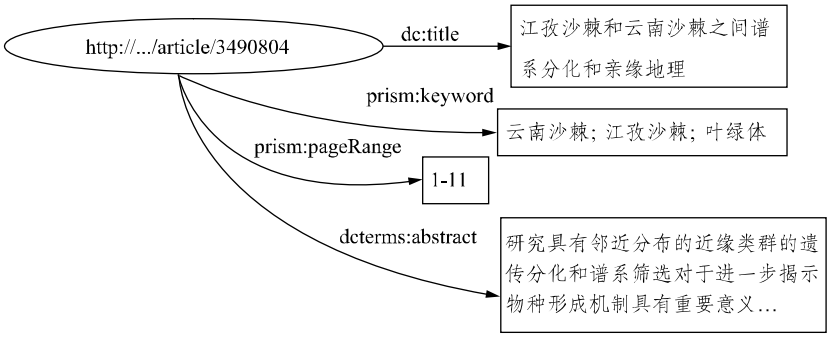

图3　CSCD 论文描述的 RDF 图

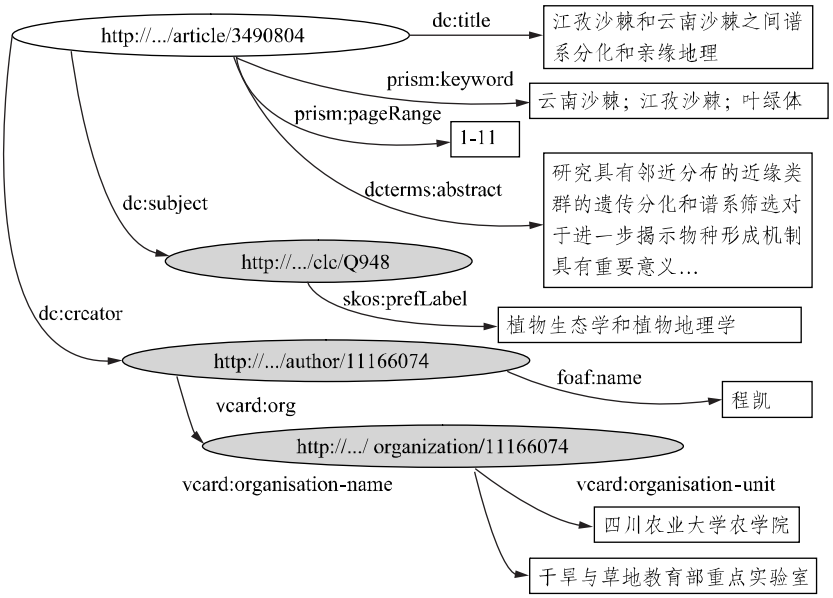

图4　包含 RDF 链接的论文基本信息

http://semweb.csdb.cn/cscd/sparql

作为例子，如下查询可以得到在“西北师范大学”以“NCET-05-0886”为基金项目发表的论文及作者列表：

```
SELECT ?article ?author WHERE {
?article dc:creator ?author.
?funding ara:result ?article.
?funding dc:title ?fundName.
?author vcard:org ?org.
?org vcard:organisation-name ?orgName.
filter(REGEX(?orgName, '西北师范大学') &&
REGEX(?fundName, 'NCET-05-0886'))
}
```

2.4　FLUX-SW 信息的发布

FLUX-SW 信息的发布流程如下：

①数据建模

选取的实体内容包括：野外观测台站、通量塔、监测指标、数据实体、数据属性。其中，数据实体与数据属性分别对应于某项科学数据实体（一张关系型数据表，或者一个数据文件）的基本元数据及其数据结构描述，数据实体元数据往往还包含一项

dc: source 属性,指向原始的数据表或者数据文件的 HTTP URL。

在 RDF 词表上,FLUX-SW 主要采用了国家生态系统观测研究网络、中国生态系统研究网络行业标准"长期生态学数据资源元数据标准"[8],同时还采用了 Dublin-Core、DC-TERMS、PRISM 等词表。此外,由于包含了空间信息,FLUX-SW 采用了 GEO 词表[9],用以描述台站和通量塔的空间位置信息。

FLUX-SW 各实体之间的关联关系如图 5 所示。

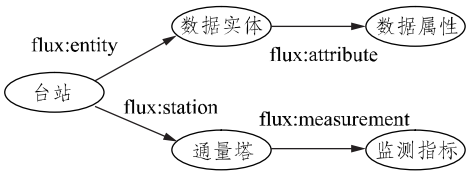

图 5 FLUX-SW 实体关联模型

②实体命名

分配每个实体的 URI 格式如下:

< baseUri > / < entityTypeName > / < entityId >

确定基地址(baseUri) 为 http: //semweb. csdb. cn/flux,因此 ID 为"CBS"(长白山) 的台站(entityTypeName 为 station) 具有如下 URI:

http: //semweb. csdb. cn/flux/resource/station/CBS

③实体 RDF 化

添加每一条实体的属性内容,以长白山台站的记录为例,发布的属性包括台站名称、空间信息、气候特征、土壤种类、植被等。

④实体关联化

根据图 5 添加实体之间的 RDF link。台站与数据实体、台站与通量塔、数据实体与数据属性、通量塔与监测指标之间的关系主要通过关系型数据库 E-R 模型中的主外键关联关系映射得到。

⑤实体发布

选取 D2R Server,开放每一个实体的访问接口。FLUX-SW 同样严格遵循了内容协商的原则。

⑥开放查询

FLUX-SW 开放了 SPARQL 查询接口,地址如下:

http: //semweb. csdb. cn/flux/sparql

作为例子,如下查询可以得到可监测到"CO2 通量"的台站列表:

```
SELECT distinct(?station) ?stationName
WHERE {
?measurement flux: varname ?varname.
?measurement flux: tower ?tower.
?tower flux: station ?station.
?station dc: title ?stationName.
filter (regex(?varname, 'CO2 通量'))
}
```

3 关键问题

3.1 多语种属性

科技文献和科学数据的描述中常常会包含多语种信息,如: CSCD-SW 论文的标题、关键词、摘要具有中英文两种语言描述。同样,FLUX-SW 中数据实体的标题、测量指标的变量名称,也具有多语种的属性值。表 3 表示 FLUX-SW 测量指标变量名称的中英文两种语言描述。

表 3 多语种属性文本示例

测量指标编号	变量名称(中文)	变量名称(英文)
#1	土壤含水量	Soil moisture
#2	风向	Wind direction
#3	CO2 通量	CO2 flux

本文认为: 不管变量名称的文本值采用什么语种,"变量名称"该谓词都应该采用一致的"flux: varname"(因为它们的语义是一致的) 。这也就意味着,在 FLUX-SW 中不会分别定义 flux: varname_en 和 flux: varname_zh 这样的谓词。相反通过采用 XML 的 xml: lang 属性来指定属性文本的语种。如下 RDF/XML:

< flux: varname xml: lang = "en" > CO2 flux < / flux: varname >

< flux: varname xml: lang = "zh" > CO2 通量

</flux: varname >

构造 SPARQL 查询时，可以采用类似于@ LANG 的表达方式来匹配指定语种的属性值，如下查询将匹配 varname 属性文本为英文版本的“CO2 flux”：

```
SELECT ?measurement ?varname
WHERE {
?measurement flux: varname ?varname.
filter( ?varname = 'CO2 flux'@ en)
}
```

由于正则匹配函数 REGEX( ) 可以匹配不同语种的文本，因此如下查询可以同时匹配到“CO2 flux”和“CO2 通量”文本：

```
SELECT ?measurement ?varname WHERE {
?measurement flux: varname ?varname.
filter( REGEX( ?varname ,'CO2') )
}
```

当然，也可以利用 LANG( ) 函数设定指定的语种：

```
SELECT ?measurement ?varname WHERE {
?measurement flux: varname ?varname.
filter( REGEX ( ? varname , 'CO2 ')  && LANG ( ?varname) = 'en')
}
```

3.2 值词表的发布

如 2.2 小节所述，在 Web 上已经有越来越多的分类体系、主题词表等资源以关联数据的方式发布。但对于中文资源来说，目前可用的关联数据资源非常有限。因此在从事数据内容的发布之前往往还需要做一些额外准备，值词表的发布就是其中的一个步骤。

本研究基于互联网上开放的《学科分类与代码》(国标 GB/T 13745) 与《中国图书馆分类法》(第四版) 资源，通过数据整理入库，基于 D2RQ 提供了分类代码的关联数据版本(简称 CS-SW) ，仅限研究与学术交流之用[8] 。

值词表的发布同样需要遵循 2.1 小节所提出的流程，其中比较重要的是 RDF 词表的选择。CS-SW 采用 SKOS[9] 作为 RDF 词表，SKOS 可以理解成一种编码方式[10] ，它定义了 skos: Concept 来表达一个词汇概念，并定义该概念可以具有的属性，如: skos: prefLabel、skos: altLabel、skos: hiddenLabel 等，SKOS 同时定义了两个概念之间的语义关系，如: skos: broader、skos: narrower、skos: related 等。

如下是针对“A1 马克思、恩格斯著作”的 RDF 描述(语言采用 N3) ：

```
<http: //semweb. csdb. cn/cs/resource/clc/A1 >
a      cs: clc , skos: Concept;
rdfs: label "A1 马克思、恩格斯著作";
dc: identifier "A1";
skos: broader  < http: //semweb. csdb. cn/cs/re-source/clc/A > ;
skos: narrower  < http: //semweb. csdb. cn/cs/resource/clc/A12 > , < http: //semweb. csdb. cn/cs/resource/clc/A11 > , < http: //semweb. csdb. cn/cs/resource/clc/A16 > , < http: //semweb. csdb. cn/cs/resource/clc/A18 > , < http: //semweb. csdb. cn/cs/resource/clc/A15 > , < http: //semweb. csdb. cn/cs/resource/clc/A13 > , < http: //semweb. csdb. cn/cs/resource/clc/A14 > ;
skos: prefLabel "马克思、恩格斯著作".
```

可以看出，其中不仅描述了该分类的标签文本、所属的 RDF 类，还给出了该类的上位分类 A、下位分类 A11 ~ A18 的链接。

3.3 RDF 词表的发布

在发布关联数据的同时，除了采用通用的 RDF 词表之外，往往还需要创建新的词汇(包括类名和属性名) ，根据关联数据的四项基本原则，这些 RDF 词表也需要发布成关联数据。

D2RQ 默认会发布用户新建的 RDF 词表，一个 RDF 词汇会具有如下形式的 URI:

http: //semweb. csdb. cn/flux/vocab/resource/ < RDF 词汇 >

在制定 RDF 词表时，本文建议 RDF 类名采取首字母大写的约定，如: Station、JournalIssue 等; RDF 属性名采取首字母小写的约定，如: author、isPartOf

等。如下表示在 FLUX-SW 中台站 Station 类的 RDF 描述:

```
flux: Station
a rdfs: Class;
rdfs: comment "represents a station" ;
rdfs: label "station"@en, "观测台站"@zh.
```

其对应的映射文件代码为:

```
map: station a d2rq: ClassMap;
d2rq: dataStorage map: database;
d2rq: uriPattern "station/@@station. S_CODE|urlify@@";
d2rq: class flux: Station;
d2rq: classDefinitionLabel "station"@en;
d2rq: classDefinitionComment "represents a station";
d2rq: classDefinitionLabel "观测台站"@zh;
```

3.4 其他 D2R Server 发布建议

在采用 D2R Server 发布关系型记录的时候,为了简化工作量并提高发布数据的质量,可遵循如下建议:

① 不要采用空白节点(blank node)。尽管 D2R 配置文件提供了 d2rq: bNodeIdColumns 用以指定哪些列可以用来映射成空白节点,但建议避免使用空白节点。由于空白节点的局域性,会造成跨数据集的空白节点无法关联,同时在多数据集的数据合并(如:溯源信息的回溯)时,空白节点也会带来其他问题。

② 尽量做好关系型数据库的前期设计。在执行 generate-mapping 工具之前,建议构建好数据库中的主外键关联关系,这样 generate-mapping 会在生成映射文件的过程中自动将这种关联映射成 RDF 链接的生成规则。另外,D2R Server 对数据库的设计还具有一些额外的要求,如: D2R Server 不允许用于生成 URI 的主键值中包含下划线等特殊字符,这时候往往需要通过新增一个自动增长的列作为主键来解决。

③ 指定非文本属性的数据类型。generate-mapping 通常会忽略数据库的列值类型,而统一将各种列值视为文本类型处理,这样发布出来的数据在 SPARQL 查询时则无法支持数值的比较和运算。因此,建议采用 d2rq: datatype 来指定数值、日期属性的类型。

④ 适当进行实体表的拆分与合并。由于数据组织的灵活性,原始的数据模型与 RDF 数据模型往往会存在着不匹配,如在 CSCD 中,论文的期刊、卷信息与论文元数据的原始记录存储在一张表中。另外一个相反的例子,分类类目(skos: Concept 对象)的信息会分别存储在不同的表(学科分类表、中图分类代码表)里,这时候又需要合并来自多张表的数据。针对这种情况,建议充分进行数据建模,不要拘泥于原有物理表的存储结构,而应该根据 RDF 实体之间的关系重新组织,并通过映射规则完成这种转换。

4 结语

本文结合中国科学引文数据库和中国生态系统研究网络通量数据的发布,提出了关联数据发布流程中的六个关键步骤,并结合多语种问题、值词表的发布、RDF 词表的发布等关键问题进行了详细的分析,最后给出利用 D2R Server 发布数据的建议。由于关联数据还没有引起国内数据库领域足够的关注,关联数据在国内尚没有形成有影响力的或者成熟的应用,基本处于起步探索阶段,因此本研究具有较强的实践意义。

当然,在关联数据的发布过程中,也暴露出 D2RQ 的一些不足,如:在跨越多张表进行 SPARQL 关联查询的时候,D2R Server 的性能会比较慢; HTML 发布界面过于简单,不够美观,缺乏分页控制,用户体验不够友好;无法在映射之前进行有效的数据转换处理等。再以科学数据为例,由于其内容除了关系型记录外,大部分体现为数据文件,因此除了 D2RQ 之外还需要寻找一种高效的文件系统 RDF 映射框架,同时还应考虑到两者之间的无缝集成。此外,考虑到 D2RQ 仅仅用以显式的映射,因此还需要采取类似于 Silk [9] 等关联发现框架,来发现不同实体之间的隐性关联。

尽管如此，作为一种致力于关系型数据库的RDF映射框架，D2RQ由于其对环境（操作系统、数据库版本等）的适应性、操作简便性，以及灵活的可配置性，仍不失为对现有数据内容完成关联数据化发布的最佳选择。

参考文献：

[1] Chris Bizer, Richard Cyganiak, Tom Heath. How to publish linked data on the Web [EB/OL]. [2012-10-10]. http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/.

[2] Cyganiak R, Bizer C. Pubby—A linked data frontend for SPARQL endpoints [EB/OL]. [2012-10-10]. http://www4.wiwiss.fu-berlin.de/pubby/.

[3] Christian Bizer, Andy Seaborne. D2rq: Treating non-rdf databases as virtual rdf graphs [R/OL]. [2012-10-10]. http://www.wiwiss.fu-berlin.de/suhl/bizer/pub/Bizer-D2RQ-ISWC2004.pdf.

[4] OpenLinksoftware [EB/OL]. [2010-09-10]. http://virtuoso.openlinksw.com.

[5] AKSW/Sparqlify: Overview [EB/OL]. [2012-10-10]. https://github.com/AKSW/Sparqlify.

[6] W3C RDB2RDF working group. [EB/OL]. [2012-10-10]. http://www.w3.org/2001/sw/rdb2rdf/.

[7] Sahoo SS, Halb W, Hellmann S, et al. A survey of current approaches for mapping of relational databases to rdf [R]. W3C RDB2RDF XG Incubator Report 2009: W3C.

[8] Bizer C, Cyganiak R. D2r server-publishing relational databases on the semantic web [C] // ISWC 2006, 2006: 26.

[9] D2Rserver for the CIA factbook [EB/OL]. [2012-10-10]. http://www4.wiwiss.fu-berlin.de/factbook/.

[10] D2Rserver for eurostat [EB/OL]. [2012-10-10]. http://www4.wiwiss.fu-berlin.de/eurostat/.

[11] 刘炜. SKOS版的LCSH [EB/OL]. [2012-10-10]. http://www.kevenlw.name/archives/655. (Liu Wei. LCSH in SKOS format [EB/OL]. [2012-10-10]. http://www.kevenlw.name/archives/655.)

[12] LIBRIS [EB/OL]. [2012-10-10]. http://libris.kb.se.

[13] Hannemann J, Kett J. Linked data for libraries [C] // Information Technology, Cataloguing, Classification and Indexing with Knowledge Management. Gothenburg, Sweden: 2010-8.

[14] Momtchev V, Peychev D, Primov T, et al. Expanding the pathway and interaction knowledge in linked life data [C] // Proceedings of International Semantic Web Challenge, 2009.

[15] Diseasome. Map: Explore the human disease network, dataset, interactive map and printable poster of gene-disease relationships [EB/OL]. [2012-10-10]. http://diseasome.eu/map.html.

[16] Barnaghi P, Presser M, Moessner K. Publishing linked sensor data [C] //ISWC 2010, 2010-11-07, Shanghai, China.

[17] 中国科学文献服务系统 [EB/OL]. [2012-02-10]. http://sdb.csdl.ac.cn/. (China sciences document service system [EB/OL]. [2012-02-10]. http://sdb.csdl.ac.cn/.)

[18] 中国生态系统研究网络 [EB/OL]. [2012-02-10]. http://www.cern.ac.cn. (Chinese ecosystem research network [EB/OL]. [2012-02-10]. http://www.cern.ac.cn.)

[19] Tim Berners-Lee. Linked data——Design issues [EB/OL]. [2012-10-10]. http://www.w3.org/DesignIssues/LinkedData.html.

[20] Cool URIs don't change [EB/OL]. [2012-02-10]. http://www.w3.org/Provider/Style/URI.

[21] SPARQL protocol for RDF [EB/OL]. [2012-10-10]. http://www.w3.org/TR/rdf-sparql-protocol/.

[22] Library linked data incubator group: Datasets, value vocabularies, and metadata element sets [EB/OL]. [2012-02-10]. http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025.

[23] Ontologies-W3C [EB/OL]. [2012-02-12]. http://www.w3.org/standards/semanticweb/ontology.

[24] Metadata element set tag cloud [EB/OL]. [2012-02-12]. http://www.w3.org/2005/Incubator/lld/wiki/File:LLD-MetadataElementSetTagCloud.png.

[25] Academic research project funding ontology (ARPFO) [EB/OL]. [2012-02-10]. http://vocab.ox.ac.uk/project-funding.

[26] 长期生态学数据资源元数据标准 [EB/OL]. [2012-02-10]. http://cerndis1.cern.ac.cn/cms/upload/2008-03-24/1206323915797.doc. (Long-term ecological data resource metadata standard [EB/OL]. [2012-02-10]. http://cerndis1.cern.ac.cn/cms/upload/2008-03-24/1206323915797.doc.)

[27] Basicgeo (WGS84 lat/long) vocabulary [EB/OL]. [2012-02-10]. http://www.w3.org/2003/01/geo/.

[28] Classfication and code—Linked data server [EB/OL]. [2012-10-27]. http://http://semweb.csdb.cn/cs/.

[29] SKOS simple knowledge organization system reference [EB/OL]. [2012-02-10]. http://www.w3.org/TR/2009/REC-skos-reference-20090818/.

[30] 刘炜. SKOS 不是 KOS,Linked Data 不是 Data [EB/OL]. [2012-10-18]. http://www.kevenlw.name/archives/2124. (Liu Wei. SKOS, NOT a KOS,linked data, NOT a kind of data [EB/OL]. [2012-10-18]. http://www.kevenlw.name/archives/2124.)

[31] Volz J, Bizer C, Gaedke M,et al. Silk—A link discovery framework for the web of data [C] // LDOW 2009, Madrid, Spain.

沈志宏 中国科学院计算机网络信息中心高级工程师。
通讯地址:北京海淀区中关村南四街4号。邮编:100190。
刘筱敏 中国科学院国家科学图书馆研究馆员。
通讯地址:北京中关村北四环西路33号。邮编:100190。
郭学兵 中国科学院地理科学与资源研究所高级工程师。
通讯地址:北京朝阳区大屯路甲11号。邮编:100101。
张晓林 中国科学院国家科学图书馆馆长,研究员,博士生导师。
通讯地址:北京中关村北四环本路33号。邮编:100190。
黎建辉 中国科学院计算机网络信息中心正高级工程师,博士生导师。
通讯地址:北京海淀区中关村南四街4号。邮编:100190。

(收稿日期:2012-07-06)