

# 关联数据发布技术及其实现<sup>\*</sup>

——以 Drupal 为例

夏翠娟 刘 炜 赵 亮 张春景 徐 昊 朱雯晶

**摘 要** 关联数据是一种轻量级的语义网实现技术,其重要价值在于通过 RDF 数据模型,将网络上的非结构化数据和采用不同标准的结构化数据转换成遵循统一标准的结构化数据,以便机器理解。文章梳理了关联数据的实现方式,目前关联数据的发布模式有:静态发布、批量存储、调用时生成、事后转换(D2R),发布关联数据的工具包括:VoID 词表、前端转换工具、OWL 及 SKOS 相关工具、Web services、Web 应用框架、CMS 及 RDFa。并详细介绍了开源 CMS 平台 Drupal。Drupal 作为一直关注语义网技术的开源 CMS 平台,其高度模块化的架构为实现关联数据的发布打下了良好的基础,RDF、SPARQL、SPARQL\_Endpoint 及其周边模块构成了一套较为成熟的关联数据发布与消费的框架,最后利用该框架实现将“中国历史纪年和公元纪年对照表”发布成关联数据。表 1。图 5。参考文献 21。

**关键词** 关联数据 内容管理系统 语义网 Drupal

**分类号** G254 TP391

## The Current Technologies and Tools for Linked Data: A Case of Drupal

Xia Cuijuan, Liu Wei, Zhao Liang, Zhang Chunjing, Xu Hao & Zhu Wenjing

**ABSTRACT** Linked Data is a lightweight implementation of Semantic Web technology based on RDF data model. With Linked Data, the unstructured data on the web can be transferred to structured data for machine understanding. Common examples of publishing models include static publishing, bulk storage, Real-time generation, and post-conversion (D2R). Most often used tools to publish linked data are VoID vocabulary, front-end conversion tool, OWL and SKOS tools, Web services, Web application framework, CMS and RDFa. Drupal, a popular open-source Content Management System, is also one of the tools with highly modular architecture most critical to Linked Data implementation. The main modules such as RDF, SPARQL, SPARQL\_Endpoint and other surrounding modules compose the architecture of publishing and consuming Linked Data. Finally, building on the architecture of Drupal, “Chronological tables of Chinese history and the corresponding western calendar” was published into Linked Data. 1 tab. 5 figs. 21 refs.

**KEY WORDS** Linked Data. Content Management System. Semantic Web. Drupal.

### 1 引言

互联网上存在着大量非结构化数据和采用不同标准的结构化数据,关联数据是一种简单的语义网实现技术,其重要价值在于通过资源

描述框架(简称 RDF,下同)数据模型,将网络上的非结构化数据和采用不同标准的结构化数据转换成遵循统一标准的结构化数据,以便机器理解。图书馆界如美国国会图书馆,大型媒体公司如 BBC、纽约时报,商业巨头如百思买(Best

<sup>\*</sup> 本文系国家社科基金项目“关联数据的理论和应用研究”(编号:11BTQ041)的研究成果之一。

通讯作者:夏翠娟,Email:cjxia@libnet.sh.cn

Buy)都纷纷试水关联数据,大量传统网页上的数据(包括维基百科)正在被自动半自动地转换成关联数据,甚至英国政府也有意提供语义网(关联数据)信息服务<sup>[1]</sup>。可以认为,关联数据使互联网迈开了向语义网(Semantic Web)进化过程中重要的一步。“文档的网络(the Web of Document)”向“数据的网络(the Web of Data)”转变,已经是大势所趋<sup>[2]</sup>。

仅从技术上看,关联数据的实现并不复杂。但是,要让当前互联网上海量的基于超链接的“文档的网络”转换成基于关联数据的“数据的网络”,是一项浩大的工程,需要简便可行的模式和方法。众多的内容管理系统(简称 CMS,下同)是目前网络资源发布的主力,其中 Drupal 是一款技术领先、用户众多、特色鲜明的开源软件,它基于完全开源的 XAMPP 架构,具有高度模块化、优良的用户友好性等特点,并拥有庞大的网络社区支持,成为 Web2.0 时代的翘楚和 CMS 中的佼佼者。难能可贵的是,它很早就开始了对语义网技术的支持,使得它在新时代得以引领潮流。2009 年,Drupal 已形成了一整套支持关联数据发布(publish)和消费(consume)的完整机制。2011 年 1 月,Drupal7 正式发布,已经将关联数据作为核心模块<sup>[3]</sup>。本文对关联数据的原理、实现技术和工具作了较为全面的梳理,主要以 Drupal 为工具,通过案例,详细阐述关联数据的发布思路、技术和流程。

## 2 关联数据及其实现

### 2.1 概述:如何在 Web 上发布关联数据

Tim Berners-Lee 于 2006 年在 *Linked Data*<sup>[4]</sup> 一文中提出了“关联数据”(Linked Data)的概念和关联数据的四原则<sup>①</sup>,成为“关联数据”理念的基石。

2007 年 7 月, *How to Publish Linked Data on the Web*<sup>[2]</sup> 一文发布,成为关联数据实现技术的经典,后来这篇技术报告又被原作者扩展为一本专著。文中将关联数据定义为“一种在万维网上发布和链接结构化数据的方式”,作者认为,“关联数据是利用万维网来创建来自不同数

据源的语义链接”,“数据价值和效用随着其链接到其他数据的增多而大大增加”<sup>[5]</sup>。也就是说,数据之间的关联越是丰富,数据则越有价值,数据的价值就越是得到体现。

从技术上看,关联数据是语义网的一种轻量级的实现方式,它可以让 Web 代理通过简单通用的 HTTP URI 规范直接访问文档中的“数据”,是实现“数据的网络”的关键技术。发布关联数据从技术上来说是很简单的,只需要遵守两条基本准则:

第一,利用资源描述框架数据模型在万维网上发布结构化数据;

第二,利用 RDF 链接不同数据源的数据。

W3C 的 RDF<sup>[6]</sup> 是指由“主体(subject)—谓词(predicate)—客体(object)”三部分(称为“三元组”<sup>[7]</sup>)组成的资源描述通用模型。其中“主体”是有统一标识符(URI)的资源,也可以是没有命名空间的空白节点如 DOI、ISBN 等;“客体”可以有 URI 的资源、空白节点、也可以是字符串值;而“谓词”表示了主体和客体的关系,即第二条原则中的 RDF 链接(RDF Links)<sup>[2]</sup>。根据关联数据的原则,关联数据的三元组都应该尽可能地以 URI 来表达,以 RDF 来编码。尤其是主体,必须能够以开放的 HTTP 方式来存取,数据的 RDF 描述中应包含其他更多数据的链接。

关联数据完全建立在已有的 Web 技术(HTTP、URL 和 HTML)基础上,以四个基本原则对 Web 技术规范作了进一步的规范和限定。关联数据的发布可简单概括为如下步骤:

①用 RDF 数据模型描述要发布的数据资源,为其生成 HTTP URI(通常是一个 Cool URI),并生成资源的 RDF 描述文档。

②在数据与数据之间建立 RDF 链接。

③在 Web 上发布 RDF 文档,可采用以下两种方法之一:支持 HTTP 的内容协商机制(Content Negotiation),能根据客户端信息请求的类型(text/html 还是 application/rdf+xml)决定返回 HTML 的表示形式还是 RDF 的表示形式;或支持采用带“#”号(hash)的 URI 方式定位到 RDF 中具体的数据资源。

① 四原则的具体内容请参见刘炜等撰写的《RDA 与关联数据》

④提供一个标准开放的访问接口,支持使用 RDF 的标准化检索语言 SPARQL<sup>[8]</sup>对 RDF 数据库进行检索,供远程调用本地数据。

## 2.2 现状:关联数据的实现技术和工具

关联数据是一种新兴的技术,大部分工具尚在发展之中,关联数据的四原则决定了关联数据的实现独立于任何具体的技术和工具,只和相关的 Web 标准有关。因此,在多数情况下,关联数据的生成、发布和消费无需改变原有的内容发布系统,只需在原有系统之上架构一层支持关联数据的应用模块即可。

华东师范大学娄秀明的硕士论文《用关联数据技术实现网络知识组织系统的研究》<sup>[9]</sup>对发布关联数据的技术方案作了综述;沈志宏、张晓林在《关联数据及其应用现状综述》<sup>[10]</sup>一文中也总结了一些关联数据发布的技术方案。

根据数据量的大小、数据的更新频率、数据的存储方式和访问方式的不同,一般可考虑采用以下几种方式来发布关联数据:①发布静态的 RDF 文件,适用于数据量很小的情况。②将 RDF 文件存储在 RDF 数据库中,并采用 Pubby 等服务器作为关联数据服务的前端,适用于数据量大的情况。③在请求数据时根据原始数据在线生成 RDF 数据,适用于更新频率大的情况。④D2R 方式,即从关系数据库到 RDF 数据转换,适用于将关系数据库存储的数据内容发布成关联数据。笔者将这四种方式总结为:静态发布、批量存储、调用时生成、事后转换。

W3C 在其“图书馆关联数据应用孵化小组”的官方网站<sup>[11]</sup>上总结了一些主要的关联数据实现技术、工具和资源,可资参考:

①描述关联数据集的 VoID 词表。虽然通过 Cool URI 获取的 RDF 数据能够很容易转化为可读的 HTML 格式,但无法应对大规模的网络检索。VoID 词表<sup>[12]</sup>是一个 RDF Schema 词表,是描述关联数据集的元数据,作为关联数据集和用户(人或机器)之间的桥梁,它包含数据发现、数据目录和数据存档,用于帮助用户发现如 DBpedia 这样的海量关联数据集中的可用信息。

②传统数据库的前端转换工具。可将关系数据库中的数据加装为关联数据的软件 D2R

Server,关系数据库中的内容可藉此在浏览器上浏览,并支持 SPARQL 语言查询。该软件还可以利用 Apache 的 rewrite 规则,从已有的 SRU 数据库中生成关联数据,如果某些信息可通过一个 SPARQL endpoint 获取,那么像 Pubby 这样的关联数据前端工具可以自动生成资源的 URIs。

③OWL 及 SKOS 相关工具。OWL 本体语言<sup>[13]</sup>和简单知识管理系统 SKOS<sup>[14]</sup>是关联数据发展的两个重要推动力。OWL 是一套 RDFS 词表,扩展了 RDF/RDFS 的描述能力。OWL 和 SKOS 的相关技术发展较为成熟,许多 OWL 工具<sup>[15]</sup>和 SKOS 工具<sup>[16]</sup>都可为关联数据所用。

④用于关联数据的 Web services。理论上说,大部分基于特定领域的 Web 服务 API 都可以经过二次处理支持关联数据,但即使在原有数据库的前端植入关联数据 URI,使后台数据库支持 SPARQL 检索也绝非易事,性能和健壮性方面的考虑也是一种阻碍。而 Web services 技术可以提供一种可选的解决方案,为了方便第三方代理的接入,它有着规范的服务描述文档。已经有少数关联数据应用开始利用 Web services 进行资源的发现和重用了,通常都是以 API 的形式,如 W3C 正在开发之中的 Linked Data API, OCLC 的 VIAF,美国国会图书馆的技术中心搜索服务等。

⑤Web 应用框架。典型的 Web 应用框架通常采用流行的 Model-View-Controller (MVC) 模式、REST Architectural Style 和面向资源架构 (Resource Oriented Architecture, ROA) 等,这些架构能很好地支持应用的创建、维护和重用。他们通常都包含一个通用的组件,即实现 URI 路由机制的模块,允许软件开发者定义一个 HTTP 形式的 URI,并将其映射到控制层,而后使用合适的视图 (views) 和模式 (models) 生成一个 HTTP 响应。这一点促进了 Cool URIs 的应用,并迫使开发者将资源发布在 Web 上。关联数据正是用 HTTP URI 来命名资源,对人和机器分别使用不同的显示方式,如对人输出 HTML 格式,对机器输出 RDF 格式,这一点与 Web 应用框架的特点不谋而合。因而可以利用 Web 应用框架,方便生成 RDF 格式的数据,并不比生成 HTML、XML 和 JSON 复杂。目前基于不同程序语言和

不同操作系统的 Web 应用框架已经得到了十分广泛的应用。

⑥CMS 及 RDFa。CMS 通常建立在一个成熟的 Web 应用框架之上,对于 URI 的支持自然应该不在话下,又由于 CMS 对内容结构的定义有着完善的支持,其后台数据库中存在着大量的结构化信息。而 RDFa 技术可让这些数据表达为 RDF 格式并嵌入 HTML,对人和机器提供服务。

### 3 Drupal 对关联数据的支持

#### 3.1 Drupal 与语义网

早在 Drupal 诞生之初的 2001 年,Drupal 的创始人就在 Drupal 核心代码中引入了支持 RDF 的功能<sup>[17]</sup>,在 Drupal 核心代码包中导入了一个名为 rdf.php 的源文件,提供了面向对象的、基于 RDF 的 XML 解析器和一个 RDF 标题生成器,为 Drupal 支持语义网奠定了基础。在 2007 年的 Drupal 年会上,Drupal 的拥趸者们在 Drupal 开源社区上建立了语义网群组,致力于为 Drupal 增添语义功能。当关联数据概念提出时,Drupal 的开发团队反应迅捷,很快于 2007 年 10 月成立了 RDF 项目组。2008 年 Drupal 年会上正式推出了 Drupal 对关联数据的支持功能<sup>[18]</sup>。随后在不到一年的时间里,Drupal6 上支持关联数据的模块就陆续发布,2009 年的 Drupal 年会上,这些模块已能够全面、规模化地支撑关联数据的生成、发布和利用<sup>[19]</sup>。2011 年 1 月,Drupal7 正式版发布,更是把 RDF 相关模块作为 Drupal 的核心。

表 1 Drupal 与语义网大事

| 日期          | 大事记                                           |
|-------------|-----------------------------------------------|
| 2001 年      | 引入 rdf.php                                    |
| 2007 年年会    | Drupal 开源社区语义网群组成立                            |
| 2007 年 10 月 | Drupal 开源社区 RDF 项目成立                          |
| 2008 年年会    | 宣布对关联数据的重视和支持,在 Drupal6 上支持关联数据的模块陆续公开发布并提供下载 |
| 2009 年年会    | 所有支持关联数据的模块都开发完成并在年会上演示                       |
| 2011 年 1 月  | Drupal7 正式版发布,RDF 相关模块作为核心模块                  |

#### 3.2 Drupal 支持关联数据的原理

##### 3.2.1 Drupal 利用 RDF 数据模型在万维网上发布结构化数据

Drupal 对关联数据的支持是从对自身结构化数据的 RDF 化开始的。在用 Drupal 建立的网站中,存在着大量的结构化数据。Drupal 作为一个内容管理系统,可管理任意不同的内容类型(content type),每一种内容类型都有自己的结构化数据,包含一个字段集(content field),如博客(blog)是 Drupal 的一种常见的内容类型,包含“标题”、“创建者”、“创建时间”、“内容”这些字段。不同的内容类型包含不同的字段集。利用 Drupal 的 CCK 模块,可自定义任意新的内容类型,包含任意组合的字段集。Drupal 站点的所有内容类型及其字段构成了站点的内容模型(site content model)。在 Drupal 中,基于某个内容类型创建的实例叫作节点(node),例如,博客在 Drupal 中作为一种内容类型,博客的标题是这种内容类型的一个字段,某作者写的一篇博文则是一个节点。Drupal 提供了一种方式,使内容类型、字段和节点分别映射到 RDF 三元组模型中的“类”、“属性”和“对象”,即主体、谓词、客体,如图 1 所示。

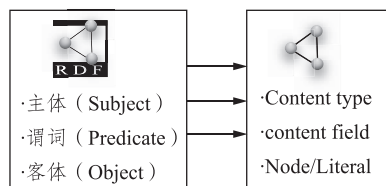


图 1 RDF 数据模型与 Drupal 内容模型的映射关系

这样,Drupal 的站点内容模型就可以映射到基于 RDFS 的本地词汇表(Vocabulary):

site:Blog a rdfs:Class;rdfs:label "Blog";

site:Title a rdf:Property;rdfs:label "Title";

当定义好站点的内容模型后,系统会自动生成基于 RDFS 的本地词汇表(Vocabulary),命名空间(Namespace)的前缀默认为“site”,地址默认为“http://siteurl/ns#”,当一个节点创建之

后,系统会自动赋予一个 http URI,一般默认为: `http://siteurl/node/nodeno`(siteurl 为站点的 URL 地址,nodeno 为系统按内容节点创建时间顺序自动生成的流水号),指向该博文的 html 文档。同时,系统也会自动生成该博文的 RDF 数据,RDF 数据中术语的命名空间为“site:`http://siteurl/ns#`”,RDF 数据也自动获得一个 http URI,一般默认为“`http://siteurl/node/nodeno/rdf`”,Drupal 的 Web 服务器 Apache 支持接收和发送 `application/rdf+xml` 类型的指令,因而在关联数据浏览器上访问 `http://siteurl/node/nodeno/rdf`,可返回 RDF 数据,并可人工浏览 RDF 数据的全貌。有了这样的机制,Drupal 站点中的所有内容都可以自动转换成基于 RDF 数据模型的语义数据,并自动获得 http URI,发布在互联网上。

将站点内容模型映射到 RDF 数据模型是 Drupal 支持关联数据的第一步,仅仅如此是远远不够的。RDF 数据模型是关联数据的基础,但在具体应用中,难免会用到一些领域本体 (Ontology),如 FOAF, DublinCore, SKOS, OWL, SIOC 等等。Drupal 也支持导入外部本体,并使站点管理员可以定义这些外部本体与本地站点内容模型之间的映射关系。

### 3.2.2 Drupal 利用 RDF 链接将来自不同数据源的数据联系起来

仅仅生成本地站点的 RDF 数据对于关联数据的应用来说显然是不够的,还必须能够将本地站点的 RDF 数据以标准化的形式开放给远程站点,同时整合其他数据源的 RDF 数据,并在各类 RDF 数据之间建立关联关系。RDF 数据的特性就是可开放、可利用、可关联。目前,Web 上最常用的开放利用 RDF 数据的方式是 Web 上公开的 SPARQL 查询终端 (public SPARQL Endpoint),如 DBPedia 的 SPARQL 查询终端,在 Web 上可公开访问,支持人机查询。可人工在 SPARQL 查询终端的 Web 界面上输入 SPARQL 查询语句,返回结果可以 html 格式显示在传统浏览器上,也可以 N3、JSON、XML/RDF 等格式的文档导出。SPARQL 查询终端也支持机器查询,SPARQL 语句可嵌入程序代码之中,返回结果也可编程选择、提取、分析。Drupal 可为本地站点

提供 SPARQL 查询终端,公开发布后供人机调用。另一种利用 RDF 数据集的方式是批量导出和导入 RDF 数据,Drupal 同样也提供这种方式,在本地站点管理后台,可导入来自其他数据源的 RDF 数据,N3、JSON、XML/RDF 等常用的格式都支持,也可将本地站点的 RDF 数据导出为上述格式的文件。

如果说超文本链接是传统的“文档的 Web”链接机制,那么,对于“数据的 Web”来说,这种机制即是 RDF 链接。RDF 链接能使用户在不同的数据源之间有目的地穿行。Drupal 同样也提供了这样一种链接机制。在 Drupal 原有内容模型中,就支持字段值为站点内的另一个节点,也就是说,节点 A 和节点 B 之间的关系,可用字段指示。因此,当字段被映射为一个 RDF 属性,而节点被转换成 RDF 数据实例之后,一个由节点转换而成的 RDF 数据实例 A 和另一个由节点转换而成的 RDF 数据实例 B 之间,就由 RDF 属性来定义它们之间的关系,同时建立它们之间的关联。如下所示:

节点 A——字段——节点 B

RDF 数据实例 A——RDF 属性——RDF 数据实例 B

另外,在 Drupal 管理后台,可批量处理存储在 RDF 仓储中的数据,在它们之间建立 RDF 链接。

## 4 如何利用 Drupal 发布关联数据

### 4.1 Drupal 发布与消费关联数据的总体思路

Drupal 的 RDF 项目开发团队以自己建立的 Demo 站点 (PROJECT BLOGS),说明了 Drupal 发布与消费关联数据的总体思路。

首先,新建了“Person”这一内容类型,自定义了“姓”、“名”、“个人主页”等字段,同时导入著名的 FOAF 本体,定义好“Person”的内容模型与 FOAF 本体之间的 RDF 映射,为团队中的每一个人创建一个“Person”类型的节点,系统自动生成了每个人的 html 文档及其 URL 以及 RDF 数据及其 http URI,所有人的 RDF 数据都存储在后台的 RDF 仓储之中。

其次,站点建立了自己的 SPAQL Endpoint 接口,可向远程站点 (REMOTE DRUPAL SITE) 提供关联数据服务。

同时,站点也消费互联网上开放的关联数据集如 DBPedia 中的关联数据。从 DBPedia 中提取所需的关联数据,用 RDF links 链接到本地站点中(见图2)。

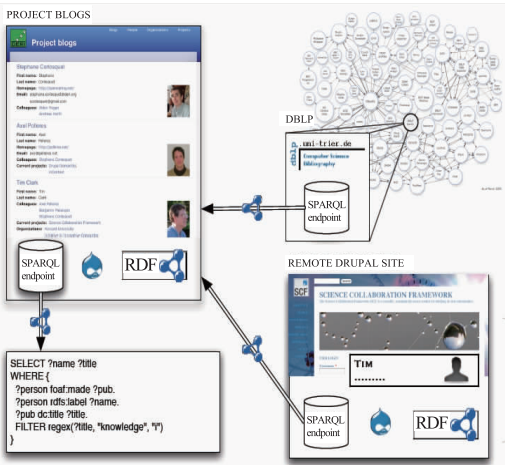


图2 Drupal发布与消费关联数据的总体思路<sup>[20]</sup>

## 4.2 Drupal支持关联数据的模块

Drupal对关联数据的支持,得益于它的模块化设计思想。模块化的设计思想,最早源于管理信息系统(MIS)开发中的“原型法”,在Web2.0时代,则与“永远的Beta版”这种思维方式有异曲同工之妙。但Drupal的模块化设计思想,远不止如此简单。Drupal的每一项特定功能,是以模块的形式开发的,也是以模块的形式管理的。系统管理员可以通过启用或关闭某一模块来决定某一功能是否可用,也可以导入一个新模块来增加一个新功能。作为一个开放源代码软件,Drupal具有标图2 Drupal发布与消费关联数据的总体思路<sup>[20]</sup>化的开发框架,新的功能可以模块的形式封装起来,提供遵循Drupal开发标准的接口,以即插即用的方式融入Drupal的核心模块。有了核心模块的支持,针对特定功能开发的模块一般都有轻量、简洁的特征,易于开发和维护。

Drupal支持RDF的模块,根据其功能的不

同,大致分为以下几组:

第一组:RDF,ARC2。RDF模块和ARC2系统为Drupal提供最根本的支持,是Drupal的内容模型向RDF数据模型转换的关键。在Drupal中安装并启用了RDF模块和ARC2系统后,Drupal的内容模型就可以自动转换成RDF数据模型。当节点被创建后,系统自动生成的RDF数据则存储在ARC2系统中,ARC2是存储RDF数据的仓储。

第二组:RDF CCK, evoc。RDF CCK和evoc这两个模块使Drupal具备了将Drupal的内容模型转换成各种领域本体的功能。RDF CCK模块建立在Drupal的另一个重要模块——CCK模块的基础之上,CCK模块支持管理员自定义任何类型的内容类型,是Drupal之所以具备强大可扩展性的灵魂,也是Drupal应用如此广泛的主要原因之一。RDF CCK模块是RDF模块的补充,它提供一个Web管理界面,利用CCK模块扩展的内容类型和自定义字段,可以在这个管理界面上设置与各种领域本体之间的映射。而evoc模块则负责为本地站点导入外部本体。安装和启用这两个模块后,管理员可以将某个内容类型映射到一个本体的Class,让字段映射到Class的Property。这里用到的Class和Property则需要依赖evoc模块导入已有的外部本体。这两个模块为Drupal的CCK模块插上了语义的翅膀。

第三组:SPARQL, RDF SPARQL Endpoint。SPARQL和SPARQL Endpoint这两个模块为Drupal站点中的RDF数据提供检索支持,SPARQL语言是RDF数据的查询语言,Drupal的SPARQL模块在Drupal站点里内置了RDF数据的搜索引擎,更重要的是,SPARQL Endpoint模块即是上文提到的SPARQL查询终端,其功能与DBPedia的SPARQL查询终端一致,为本地站点的RDF数据向互联网开放提供了人机可用的标准化支持。

第四组:RDF Proxy。RDF Proxy模块从其他提供关联数据的站点获取关联数据。如果说RDF SPARQL Endpoint模块为Drupal站点提供了一个供外界(人或机器)访问的接口,那么RDF Proxy模块则是由外而内吸纳RDF数据的通道。

4.3 将“中国历史纪年和公元纪年对照表”发布成关联数据

笔者基于对 Drupal 发布和消费关联数据的了解以及对历史文献数字化资源利用的考虑,利用 Drupal 和 W3C 的 OWL 时间本体<sup>[21]</sup>将“中国历史纪年和公元纪年对照表”发布成关联数据。在中文历史文献数字化资源的利用中,不可避免地要处理中国历史纪年和公元纪年对照

的问题。“中国历史纪年和公元纪年对照表”是一种典型的结构化数据,规则清晰,非常适合作为关联数据发布,也很有利用价值,W3C 的 OWL 时间本体包含了多种与时间相关的类和属性以及处理时间问题的规则和方法,而 Drupal 提供了一个易操作的平台,支持本体的重用,因而不必要再“重新发明轮子”来自定义本体。

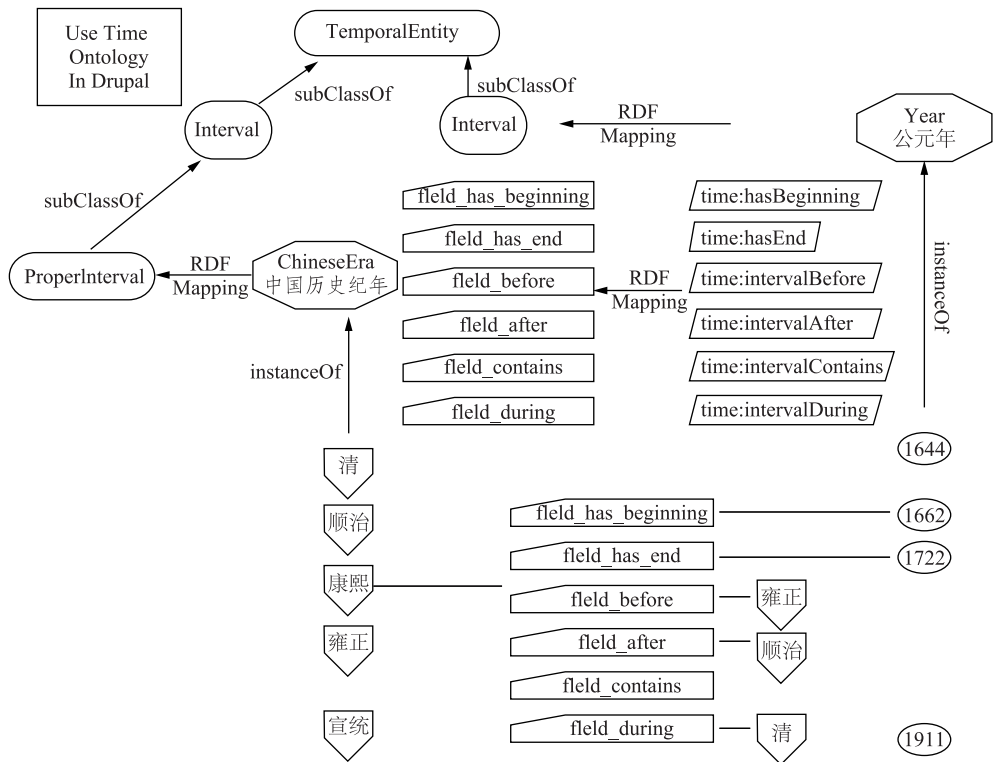


图3 定义内容结构和本体映射方案

首先,根据需求定义内容模型和本体映射方案,利用 CCK 模块自定义“公元年”和“中国历史纪年”两种内容类型,利用 evoc 模块导入 OWL 时间本体,OWL 时间本体包含一个全面的类和属性集,实际使用时可以只选取所需的部分。对于中国历史纪年中的“朝代”如“明”、“清”,年号如“康熙”、“雍正”等,都可以作为一个“时间段”来处理,而对于公元年如“1644”,则可以作为“时间点”来对待,时间段与时间段之间,时间点与时间点之间,存在着包含与被包含、先与后的关系。

在 OWL 时间本体中,“时间段”对应的类有“Interval”及其子类“ProperInterval”,“时间点”对应的类为“Instant”,由于“ProperInterval”表示一个有起点和终点的时间段,并且有“hasBeginning”、“hasEnd”、“intervalBefore”、“intervalEnd”这样的属性来表示关系,因此选择将“中国历史纪年”映射到 OWL 时间本体的“ProperInterval”,而非其父类“Interval”。定义好类的对应关系,同时也定义好字段与属性之间的对应关系,具体见图3。

其次,利用 RDF CCK 模块在 Drupal 站点中作

好两种内容类型与时间本体之间的映射(见图4)。

最后,创建“对照表”中所有中国历史纪年和公元纪年的节点,系统自动为其生成 RDF 数据。至此,“中国历史纪年和公元纪年对照表”

已作为关联数据发布在 Drupal 站点之上,并可利用 SPARQL Endpoint 模块向远程站点提供数据开放的接口,远程站点可编写 SPARQL 语句查询并获取本站 RDF 数据(见图5)。

RDF class:  
time:ProperInterval  
choose the RDF class this content type will be mapped to.

| Label        | Name                | Type           | RDF property          |
|--------------|---------------------|----------------|-----------------------|
| Title        | Node module form,   |                | rdfs:label            |
| hasbeginning | field_has_beginning | Node reference | time:hasBeginning     |
| hasEnd       | field_has_end       | Node reference | time:hasEnd           |
| During       | field_during        | Node reference | time:intervalDuring   |
| Contains     | field_contains      | Node reference | time:intervalContains |
| Before       | field_before        | Node reference | time:intervalBefore   |
| After        | field_after         | Node reference | time:intervalAfter    |
| Description  | Node module form,   |                | rdfs:comment          |

图 4 定义内容模型和 RDF 数据模型之间的映射

hasBeginning:  
1644  
hasEnd:  
1911  
Contains:  
世宗 雍正  
世祖 顺治  
仁宗 嘉庆  
圣祖 康熙  
宣宗 道光  
宣统  
德宗 光绪  
文宗 咸丰  
穆宗 同治  
高宗 乾隆  
After:  
明

http://localhost/ld/node/15/rdf

圣祖 康熙

changed

2010-08-18T10:05:00Z

chineseera After

http://localhost/ld/node/12#this

chineseera Before

http://localhost/ld/node/5#this

chineseera During

http://localhost/ld/node/4#this

chineseera Has beginning

http://localhost/ld/node/13#this

chineseera Has end

http://localhost/ld/node/14#this

chineseera Title

圣祖 康熙

created

2010-08-18T10:05:00Z

type

Chineseera

label

圣祖 康熙

Beginning

http://localhost/ld/node/13#this

End

http://localhost/ld/node/14#this

interval After

http://localhost/ld/node/12#this

interval Before

http://localhost/ld/node/5#this

interval During

http://localhost/ld/node/4#this

图 5 创建内容节点生成 RDF 数据

5 结语

有别于 RDF 数据的静态发布、批量存储、调用时生成、事后转换(D2R)这几种发布方式,Drupal 为关联数据的发布提供了另一种思路 and 模式,具体表现在:①可事先定义好本地内容模型与 RDF 数据模型的映射,支持本体的重用;②在内容创建时即生成 RDF 数据,受数据量大小和更新频率的影响不大;③RDF 数据保存在专

用的 RDF 数据库中,与原有系统的关系数据库之间在物理上彼此独立,在逻辑上相互映射;④以 SPARQL Endpoint 模块提供对外的数据访问接口,不依赖第三方 D2R 工具,更重要的是,它以各种相互独立又密切配合的模块、遵循关联数据的基本原则来实现关联数据的发布和消费,不影响原有内容模型和系统架构。

作为一个 CMS 平台,Drupal 对关联数据的支持为 Web2.7 时代的遗留系统向语义网进化作出了贡献。但如果作为一个专用的关联数据

发布的工具,Drupal 还存在着一定的不足,它必须依赖于自身的内容模型,依赖于在内容模型和本体之间建立的映射关系,而在这种映射的过程中,难免会损失部分语义。

随着关联数据应用范围的扩大,关联数据的应用效果日益显现,其实现技术和工具也会越来越成熟和易于应用。

## 参考文献:

- [1] 刘炜. 关联数据:意义及其实现[OL]. [2011-05-10]. <http://www.kevenlw.name/archives/1435>. (Liu Wei. Linked Data: what for and how to[OL]. [2011-05-10]. <http://www.kevenlw.name/archives/1435>.)
- [2] Chris Bizer. How to publish Linked Data on the Web [OL]. [2011-05-10]. <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>.
- [3] Stéphane Corlosquet. How to build Linked Data sites with Drupal 7 and RDFa. SemTech 2010 [OL]. [2011-05-10]. <http://www.slide-share.net/scorlosquet/how-to-build-linked-data-sites-with-drupal-7-and-rdfa>.
- [4] Tim Berners-Lee. Linked Data[OL]. [2011-05-15]. <http://www.w3.org/DesignIssues/LinkedData.html>.
- [5] 史亚光. 如何在万维网上发布关联数据[OL]. [2011-05-15]. [https://docs.google.com/View?id=ajjq7zprkrz8\\_0dfhmj8cs](https://docs.google.com/View?id=ajjq7zprkrz8_0dfhmj8cs). (Shi Yaguang. How to Publish Linked Data on the Web. [OL]. [2011-05-15]. [https://docs.google.com/View?id=ajjq7zprkrz8\\_0dfhmj8cs](https://docs.google.com/View?id=ajjq7zprkrz8_0dfhmj8cs).)
- [6] W3C RDF Working Group. Resource Description Framework (RDF) [OL]. [2011-05-15]. <http://www.w3.org/RDF/>.
- [7] W3C RDF Working Group. Resource Description Framework (RDF): Concepts and abstract syntax [OL]. [2011-05-15]. <http://www.w3.org/TR/rdf-concepts/#section-triples>.
- [8] Eric Prud'hommeaux, Andy Seaborne. SPARQL query language for RDF[OL]. [2011-05-21]. <http://www.w3.org/TR/rdf-sparql-query/>.
- [9] 娄秀明. 用关联数据技术实现网络知识组织系统的研究[D]. 华东师范大学,2010. (Lou Xiuming. Study on using Linked Data technology to achieve networked knowledge organization system [D]. East China Normal University,2010.)
- [10] 沈志宏,张晓林. 关联数据及其应用现状综述[J]. 现代图书情报技术,2010(11):1-9. (Shen Zhihong,Zhang Xiaolin. Linked Data and its applications: An overview[J]. New Technology of Library and Information Service,2010(11):1-9.)
- [11] W3C. Draft Relevant Technologies[OL]. [2011-09-02]. [http://www.w3.org/2005/Incubator/lld/wiki/Draft\\_Relevant\\_Technologies](http://www.w3.org/2005/Incubator/lld/wiki/Draft_Relevant_Technologies).
- [12] Keith Alexander. Describing Linked Datasets with the VoID vocabulary[OL]. [2011-05-21]. <http://www.w3.org/TR/void/>.
- [13] Michael K Smith,Chris Welty,Deborah L McGuinness. OWL web ontology language guide[OL]. [2011-05-23]. <http://www.w3.org/TR/owl-guide/>.
- [14] Alistair Miles,Sean Bechhofer. SKOS simple knowledge organization system namespace document-HTML variant[OL]. [2011-05-23]. <http://www.w3.org/TR/skos-reference/skos.html>.
- [15] W3C OWL Working Group. Web ontology language (OWL) [OL]. [2011-05-23]. <http://www.w3.org/2001/sw/wiki/OWL>.
- [16] W3C Semantic Web Deployment Working Group. Simple Knowledge Organization System (SKOS) [OL]. [2011-05-26]. <http://www.w3.org/2001/sw/wiki/SKOS>.
- [17] Benjamin Melançon. RDF's long lineage in Drupal [OL]. [2011-06-02]. <http://data.agaric.com/node/2642>.
- [18] Benjamin Melançon. Bringing Linked Data to Drupal;bringing the Semantic Web to the world[OL]. [2011-06-02]. <http://data.agaric.com/node/2700>.
- [19] Stéphane Corlosquet. Publishing and consuming Linked Data with Drupal[OL]. [2011-06-02]. <http://paris2009.drupalcon.org/session/publishing-and-consuming-linked-data-drupal>.
- [20] Stéphane Corlosquet. Produce and consume Linked Data with Drupal[OL][2011-06-02]. <http://openspring.net/sites/openspring.net/files/corl-et-al-2009iswc.pdf>.
- [21] Jerry R. Hobbs,Feng Pan. Time ontology in OWL [OL]. [2011-06-02]. <http://www.w3.org/TR/owl-time/>.

夏翠娟 上海图书馆系统网络中心研究开发部工程师。通讯地址:上海市淮海中路1555号。邮编:200031。

刘 炜 上海图书馆副馆长,研究员。通讯地址同上。

赵 亮 张春景 徐 昊 朱雯晶 上海图书馆。通讯地址同上。

(收稿日期:2011-09-05;修回日期:2011-09-20)