

基于Atom的数字图书馆资源描述框架

□ 王莉 / 中国科学技术信息研究所 北京 100038

摘要: 文章从元数据互操作的角度出发, 提出一个适用于数字图书馆领域的资源概念模型, 采用Atom连锁格式描述, 希望能够利用Atom的内建特性简化数字图书馆系统之间的数据集成工作, 既保证原始数据的完整性, 又可以实现完全无损失的元数据互操作。文章从对元数据互操作技术的概述开始, 介绍了Atom的基本特性, 阐明选择Atom描述资源的理由。在分析Atom在数字图书馆领域的应用基础上, 提出基于Atom的数字图书馆资源描述框架, 并给出期刊文献的描述样例, 同时给出一个基于Atom机制构建的数字图书馆资源网络, 指出重构资源网络框架将面临的风险及规避方法。文章结尾指出下一步的主要研究内容。

关键词: DL-RDF/Atom, Atom, AtomPub, 数字图书馆, 资源描述框架

DOI: 10.3772/j.issn.1673-2286.2010.06.005

元数据是数字图书馆建设中用来描述和组织信息资源必不可少的工具。随着数字图书馆建设的蓬勃发展, 出现了许多元数据格式, 例如DC、MARC、EAD、CDWA以及VRA等, 大部分都是针对特定资源类型、具有特定用途的专用格式。应用经验表明: 由于数字图书馆资源类型的多样化, 很难有一个统一的元数据格式可以满足所有类型的信息资源描述需要, 并全面支持所有的数字图书馆应用。元数据之间的互操作是数字图书馆建设中不可回避的关键问题。

1 元数据互操作概述

业界对于元数据互操作技术的研究主要分为语义互操作、语法与结构互操作和协议互操作三类^[1]。语义互操作实际上是在具体的元数据格式之间建立映射关系, 技术上实现很容易, 难度主要集中在元素

对照上。由于各种元数据方案在设计角度和描述粒度上存在差异, 简单的映射必然存在歧义, 难免会丢失信息, 甚而可能出现对照错误的情况。这种实现方式的优点在于一旦建立起良好的映射关系之后, 系统之间可以真正做到无缝连接; 而最大的缺点则是映射关系难以复用, 出现一种新的元数据方案, 就需要研究新的映射关系。语法与结构互操作的基本思路是通过一个标准的描述框架兼容所有的元数据格式。因此, 系统只需要解析这个标准描述框架, 就能够解读相应的元数据, 显然比一一建立对照关系更为通用。目前的研究主要集中在RDF/XML方面, 也有提出基于本体技术来实现元数据互操作^[2,3]。协议互操作通常采用OAI-PMH、Z39.50、OpenURL等标准协议从服务层面实现元数据的互操作。

语法结构和协议这两类互操作技术在目前应用较为广泛, 但是

存在不同程度的技术难点。从理论上说, 建立标准框架可以很好地解决异构数据的互操作问题, 但是资源的共性、服务的共性是非常难以提取和把握的。普遍认同的框架模型必然是模糊的、抽象的、一般化的, 复杂的问题抛给了实际应用。同样, 单纯从服务层面解决元数据互操作问题, 简单的协议(如OpenURL)无法完全满足实际的应用需求, 而一个周全的协议(如Z39.50)实现起来也非常复杂。

笔者提出一种基于Atom的数字图书馆资源描述框架(简称DL-RDF/Atom), 采用Atom连锁格式描述资源, 与Atom发布协议结合, 快速解决异构数字图书馆系统间的元数据互操作问题, 是后两类技术的一种融合。文章第二节介绍Atom的基本特性, 指出其在数据表示和管理方面具有独特的优越性; 同时, 分析Atom在数字图书馆领域的应用, 阐明选择Atom描述资源的

理由。第三节提出基于Atom的数字图书馆资源描述框架,并给出期刊文献的描述样例。第四节描述基于Atom机制构建的数字图书馆资源网络,指出采用Atom重构资源描述框架可能遇到的风险及相应的规避方法。第五节给出下一步的主要研究内容。

2 Atom^[4-7]

Atom是继RSS之后推出的内容聚合规范,由Atom连锁格式(Atom Syndication Format)和Atom发布协议(Atom Publishing Protocol, AtomPub)两部分组成。Atom连锁格式是一种基于XML的文件格式,通过定义明确的内容类型标签以及动态可扩展的链接机制构造了一个健壮、灵活、没有歧义的内容模型,具有良好的可扩展性。Atom发布协议构建在Atom连锁格式之上,通过简单的HTTP操作发布和管理内容,其核心是可编辑资源“集合”(collection)。每一个集合都有一个唯一的URI,所有的操作都是通过向URI发送请求完成的。

2.1 Atom在数据表示与管理方面的应用

Atom连锁格式和发布协议最初只是为了聚合和发布站点内容而设计的,内建特性使其在论坛讨论、日程安排、照片共享、维基网站、文件管理等多个领域得到广泛应用,甚至有很多Web应用开发都采用其表示数据、管理数据,用于快速构造轻量级的Web服务。GData^[8](Google Data API)就是

这样一个典型的应用。GData实际上是Google提供的一套API,使用Atom和RSS两种聚合协议,通过对Atom发布协议进行扩展,实现外部程序对支持GData接口的服务(如Google博客、日历和网络相册)发送查询或更新请求,让客户端软件能在浏览器外的环境中获取信息。另外一个典型应用是AtomServer。AtomServer^[9]是一个可扩展的开源的Atom存储框架,采用Atom条目存储通用数据,将Atom发布协议(借鉴GData作了适当扩展)作为一种通用的前端API实现各种数据存储服务。

2.2 Atom在数字图书馆中的应用

目前,国内图书馆界对Atom的研究主要集中在Web2.0/Lib2.0的应用领域,采用该技术实现“新闻报道、服务推送、书目或其他数据的定题查询定制(类似传统的SDI服务等)”^[10]。国外则领先一步,开始关注如何将Atom发布协议应用于图书馆的资源和服务,Jangle(just another generic library environment)框架就是在这样的研究背景下产生的。

Jangle^[11]是一个开放源码软件项目,其1.0版规范草案于2008年11月5日推出,提出了一套基于Atom发布协议的图书馆网络服务设计规范,旨在通过图书馆服务层建设“另一种通用的图书馆环境”,与图书馆现有系统连接,方便快捷地构造新的REST风格的图书馆服务。

Jangle框架的核心在于利用Atom发布协议机制读写数据,通

过该协议从服务层面解决元数据互操作问题。在数据建模方面,该框架只是从图书馆信息模型的角度提出了参与者(Actors)、资源(Resources)、项目(Items)和集合(Collections)四个实体概念,并没有深入到“资源”和“项目”的内部作更多的定义,关注的焦点是REST风格的面向图书馆服务的体系结构。正因为如此,Jangle项目组在应用实现中甚至提出一种直接移植现有数据到新系统中的建设方案,将解决数据异构的问题抛到整个项目建设之外,只专注于构建REST风格的Web服务。

3 基于Atom的数字图书馆资源描述框架

本研究课题的提出是为了解决“元数据互操作”问题,关注的对象是“文献”^①,核心目标是建立一个面向数字图书馆资源的领域框架,要求这个框架是适用于数字图书馆领域的模型,而不是一个一般化的、完全抽象的资源描述框架。

“实体-关系”分析法是建立概念模型普遍采用的方法,而一个概念模型是否具有生命力最关键的因素是其是否能够清晰地映射基础存储数据。实体对象的确立是起点,也是最重要的环节。

3.1 实体

追溯导致互操作性差的原因,文献对象存储和管理方式上的差异是其源头。因此,笔者从数字图书馆对文献资源的组织存储展开研究,以文献品种作为实体分析的起点,定义了三种具有层次关系的实

^① 文献是记录知识的一切载体。具体地说,在数字图书馆中,文献资源包括期刊、会议集、报告、学位论文、单册图书、丛书、工具书等类型,从载体形式上看,有印刷本、音像资料、缩微胶片、实物、分散的网络资源、数据库产品等。

体。

● 种 (varieties)

“种”即文献资源的品种，其含义与资源采购中常用的“品种”概念相似，可以是一种期刊、一卷丛书，或者一个图书品种。一般来说，一个品种即一个编目对象，关于品种的描述形成一条书目记录。

● 册 (items)

“册”是一个单本概念，指某个实际存在的资源，可以是一本期刊、丛书中的一个分册，也可以是一个电子资源的URL链接。一般来说，关于册的描述形成一条馆藏记录。

● 片段 (fragments)

“片段”是根据单个文献资源的内容结构特征对其进行的细分，可以是一本图书中的索引、章节或段落，也可以是某本期刊中的一篇文章。甚至一篇文章中的图表，如果有单独识别的必要，也可以根据应用需求将其也定义为“片段”，加以描述。

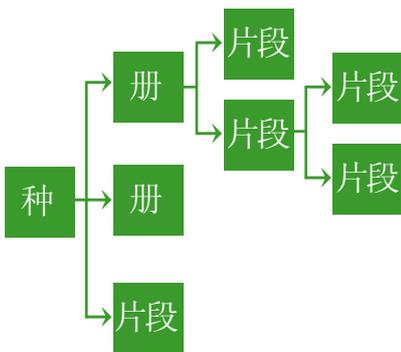


图1 三种实体的层次模型

图1展示了三种实体的层次结构，这实际上是一种“整体/部分”关系。对于期刊这种资源类型，这种层次关系是完整的，形如：期刊一本（年卷期）一篇。对于单册图书来说，“图书”是品种实体，

直接由一个或多个章节（片段）组成，没有“册”这一层次；如果一个图书品种存在复本，也可以将不同的复本定义为不同的“册”，这种描述方式能够非常方便地支持馆藏外借服务。

3.2 实体的全局唯一标识

本框架采用URI形式为每一个实体建立全局唯一标识，基本语法格式为：`{entity}/{id}`。其中，`{entity}`可取的值包括varieties、items和fragments，分别表示三种不同的实体类型。`{id}`是实体在本地系统内部使用的唯一标识，可以是单个值，也可以用列表形式（以逗号分隔）或者起止范围（用连字符分隔）表示多个实体。例如：`items/00018,00032,00076`；`items/00081-00095`；`items/00018,00032,00076,00081-00095`。

这种URI形式简单直观，与实体类型间的层次结构结合起来，就不仅仅是斜杠分隔的字符串，而是具有层次关系的目录结构。例如，可以用`varieties/{varietieid}/items`得到某种期刊（由`{varietieid}`标识）的卷期列表。如果在`/items/`后面加上`{id}`，能够很容易地得到具体卷期的信息，继而获取单篇文献信息。

“易于理解且可预测”是该URI结构最大的特点，对于资源的寻址非常方便，接口开发人员只需花很少时间了解其构造规则即可理解其含义，直接获得相关的资源。

需要注意：实体的全局唯一标识应该是静态的，在资源发生更改或应用系统发生变化时，URI链接保持不变。显然，URI的层级结

构反映的是概念模型中实体间的关系，与资源本身的存储位置无关。

3.3 属性

目录结构的URI标识反映的仅仅是“整体/部分”这一相对简单、易识别的层次关系。现实中实体间的关系纷繁复杂，可以类似于URI这样在单一记录中直接标识出来，更多的则需要通过实体特征之间的比较才能发掘。

实体特征是通过属性或属性组反映出来的，不仅是区分不同实体类型的关键，也用来识别同类实体的不同实例。定义属性的过程实际上是制定元数据方案的过程，可以参考的资料很多，包括各种复杂的编目规范，以及针对不同文献类型制定的特殊的元数据描述规范。过多考虑属性的全面与详尽，很容易陷入细节，最终成为一个编目标准，接下来“概念模型—存储模型”的映射也会变得复杂而困难，更是丢失了Atom简单易用的特性。因此，本框架为三类实体定义了一组通用属性，详细说明见表1^②。这些属性按照Atom连锁格式定义，在AtomPub协议机制下使用，可以根据应用需要进行扩展，扩展的部分要求建立单独的命名空间进行规范管理。

表1是关于实体通用属性的说明，并给出了相应的Atom描述。实体与实体的差异与联系通过属性来识别，其中Link元素非常关键。

Link是建立实体间链接关系的重要元素，Atom格式中为该元素规定了href、rel、type、hreflang、title、length等6个属性，其中rel属性描述当前实体与href指定实体

② 注：星号标记的属性表示对标准Atom连锁格式有所扩展。

表1 实体属性及其描述格式

实体通用属性	Atom描述	是否必备	可否重复	说明
ID	feed/{entry}/id	Y	N	实体的全局唯一标识, 采用URI形式表示
Title	feed/{entry}/title	Y	N	题名
Subtitle	feed/{entry}/subtitle	N	N	其他题名
Author*	feed/{entry}/author @name @uri @email @responsibility @organization @address	Y	Y	主要责任者 - 姓名 (必备) - URI - E-mail - 责任方式 - 所属机构 - 通信地址及邮编
Contributor*	feed/{entry}/contributor @name @uri @email @responsibility @organization @address	N	Y	其他责任者 - 姓名 - URI - E-mail - 责任方式 - 所属机构 - 通信地址及邮编
Language*	feed/{entry}/lang	Y	N	第一语种
Resources Type*	feed/{entry}/type @label @scheme	Y	N	资源类型 (内容) - 描述 - 编码体系
Resource Code*	feed/{entry}/code @codeno @scheme	Y	Y	资源标识 - 标识码 - 编码体系
Description*	feed/entry/content @abstract @tableofcontents	N	N	说明 - 文摘 - 目录
Category Info*	feed/{entry}/category @term @scheme @label	N	Y	主题类别说明 - 主题词/分类号 - 编码体系 - 描述
Link*	feed/{entry}/link @href @rel @type @hreflang @title @length	N	Y	链接关系 - 目的链接 (必备) - 链接关系的类型说明 - 媒体类型 - 链接到实体的语种 - 链接到实体的题名 - 链接到实体的内容长度
Format Info*	feed/{entry}/formatinfo @type @extent @medium @version @href	Y	N	载体形态 - 载体类型 (必备) - 大小 - 媒体 - 版本 - 获取链接

之间的关系, 可以表示alternate、related、self、enclosure以及via等5种最常见、最一般的链接类型, 示例1—3给出了如何灵活运用这些链接类型描述实体间关系的范例。

示例1.

```
<link rel="alternate" href="http://example.com/varieties/varietyid?format=cnmarc" />
```

说明: href链接指向当前实体 (由varietyid标识) 的CNMARC记录, 建立不同元数据格式之间的关系。

示例2.

```
<link rel="related" href="http://example.com/varieties/varietyid/items/" />
```

说明: href链接指向当前期刊 (由varietyid标识) 的卷期列表, 支持期刊目次浏览服务。

示例3.

```
<link rel="enclosure" type="audio/mpeg" title="MP3" href="http://example.com/varieties/varietyid.mp3" length="5214" />
```

说明: 采用Link元素描述多媒体形式的非书资料, 其中type属性指定了href链接中的媒体类型为MP3, 其后的链接提供音频下载。

除了以上内建链接之外, 还可以对rel取值进行扩展, 用Link元素

续表1 实体属性及其描述格式

实体通用属性	Atom描述	是否必备	可否重复	说明
Page Info*	feed/entry/pageinfo @firstpage @lastpage @totalpages @sequencenumber	N	N	页码信息 - 起始页码 - 结束页码 - 总页数 - 顺序号
Publisher Info*	feed/{entry}/publisherinfo @name @address @frequency	N	N	出版发行 - 名称 - 地址 - 出版周期
Publish Date*	feed/{entry}/published @year @month @day @date	N	N	出版时间 - 年 - 月 - 日 - 日期 (RFC 3339)
Copyright*	feed/{entry}/rights @copyright @accessrights	N	N	版权信息 - 版权声明 - 访问权限
Generator*	feed/{entry}/generator @label @uri @version	N	N	记录加工方 - 描述 - 加工系统URI - 加工系统版本
Last Updated Date	feed/{entry}/updated	Y	N	记录更新时间 (RFC 3339格式)
Icon	feed/icon	N	N	可用来存储资源封面

描述更多的联系类型。最常见的扩展是利用Link元素实现检索结果集的跳转，例如：可以规定rel的取值为first/last/prev/next，分别表示href链接指向当前结果集的首页/末页/前页/下一页。还可以规定rel的取值为cite/cited，分别表示当前实体与目标实体之间的引用/被引用关系。

一般来说，实体属性是实体的静态特征，其取值是相对固定的，在数字化过程中完成，存储在数据库中。但Link元素是一个特例，它反映的某些链接关系是在系统运行过程中产生的。例如，用Link元素表示文献与参考文献之间的引用关

系是静态的，用其实现检索结果集的跳转则是实时的。

示例4以期刊为例，给出了“刊一期一篇”三个层次描述片段。

示例4.

```
Varieties:
<id>varieties/10433</id>
<title>European Journal of
Ageing</title>
<subtitle>Social, Behavioural
and Health Perspectives</subtitle>
```

```
<author>
  <name>Springer-Verlag</
name>
  <address>Berlin/
Heidelberg</address>
</author>
<lang>En</lang>
<category>
  <term>Social Sciences,
general</term>
  <term>Psychology,
general</term>
  <term>Public Health/
Gesundheitswesen</term>
  <term>Demography</term>
  <term>Aging</term>
  <term>Social Sciences,
general</term>
</category>
<type>
  <label>J</label>
  <scheme>GB3469-83</
scheme>
</type>
<code>
  <codeno>1613-9372</
codeno>
  <scheme>ISSN(ISO3297:20
07)</scheme>
</code>
<code>
  <codeno>1613-9380</
codeno>
  <scheme>EISSN(ISO3297:2
007)</scheme>
</code>
<!-- 链接该刊下的所有卷
期-->
  <link ref="related"
type="application/atom+xml"
```

```
href="http://example.com/
openbiblio/varities/10433/items"/>
  <publisherinfo>
    <Name>Springer-Verlag</
Name>
    <Address>Berlin/
Heidelberg</Address>
    <frequency>monthly</
frequency>
  </publisherinfo>
  <generator>
    <uri>http://example.com</
uri>
  </generator>
  <updated>2009-4-
02T14:39:37Z</updated>
  .....

Items:
  <id>items/10433-2008.5:1</
id>
  <title>European Journal of
Ageing</title>
  <subtitle>Vol.5, No.1</
subtitle>
  <!--链接到刊的详细记录-->
  <link ref="related"
type="application/atom+xml"
href="http://example.com/
openbiblio/varities/10433"/>
  <!--链接该期的所有文章-->
  <link ref="related"
type="application/atom+xml"
href="http://example.com/
openbiblio/items/10433:2008.5:1/
fragments"/>
  <published>
    <year>2008</year>
    <month>2</month>
    <day>25</day>
```

```
</published>
.....

Fragments:
  <id>fragments/10433-2008.5:1
-76</id>
  <title>Intergenerational contact
beyond the dyad: the role of the
sibling network<</title>
  <author>
    <name>Galen, R.I. van</
name>
    <email>rubenvangaalen@
yahoo.com</email>
    <organization>Statistics
Netherlands (CBS), SAV/SET</
organization>
    <address>P. O .
Box 4000, Voorburg, The
Netherlands(2270JM)</address>
  </author>
  <lang>En</lang>
  <category>
    <term>Frequency of
contact</term>
    <term>Parent-child
relationships</term>
    <term>Sibling network</
term>
  </category>
  <content>
    <abstract>this paper we aim to
reach beyond the dyadic perspective
on intergenerational contact and
examine the influence of the sibling
network on parent-child contact.
.....
  </abstract>
  </content>
  <!--文章的DOI号-->
```

```
<code>
  <codeno>10.1007/
s10433-008-0076-6</codeno>
  <scheme>DOI(ANSI/NISO
Z39.84:2000)</scheme>
</code>
  <pageinfo>
    <firstpage>19</firstpage>
    <lastpage>29</lastpage>
    <sequencenumber>3</
sequencenumber>
  </pageinfo>
  <!--链接到刊的详细记录-->
  <link ref="related"
type="application/atom+xml"
href="http://example.com/
openbiblio/varities/10433"/>
  <!--链接到期的详细记录-->
  <link ref="related"
type="application/atom+xml"
href="http://example.com/
openbiblio/items/10433:2008.5:1"/>
  <!--链接到参考文献-->
  <link ref="cite"
type="application/atom+xml"
href="http://example.
com/openbiblio/fragments/
search/?DOI=10.1111/
j.1741-3737.2000.01288.x"/>
  <link ref="cite"
type="application/atom+xml"
href="http://example.com/
openbiblio/fragments/search/
?DOI=10.1093/esr/jcl032"/>
  .....
  <rights>
    <copyright>Springer-Verlag,
2008</copyright>
  </rights>
  .....
```

4 基于Atom机制重构资源描述框架

采用Atom连锁格式对数字图书馆资源进行描述，目的在于利用

Atom发布协议快速构建数字图书馆网络服务，解决元数据互操作问题。图2描述了一个基于Atom机制建立的数字图书馆资源网。

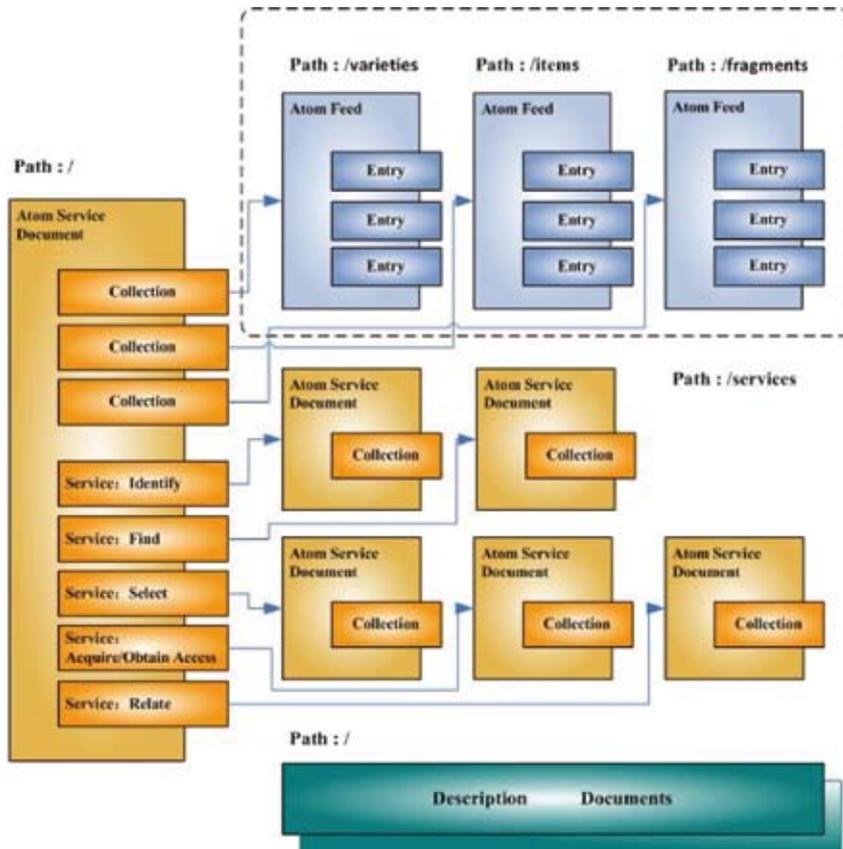


图2 基于Atom机制的数字图书馆资源网

基于Atom的数字图书馆描述框架是资源网络的基础，主要作用于图2中虚线框出的范围。该范围标识的是整个文献资源集合，是Atom机制中应用服务访问的数据源。

关于资源集合的说明存放在根路径下的Atom服务文档中。该文档是Atom机制的核心，不仅明确定义了可用的资源集合，还包括对应用服务的说明。图2中列出了识别（Identify）、查找（Find）、选择（Select）、获得/存取（Acquire/Obtain Access）和关联（Relate）

五个方面的应用。关于应用服务的说明实际上为客户机提供了一个访问端点。服务器返回的结果同样以Atom服务文档的格式描述，不同的是它包含的是符合要求的资源集合列表。这些资源集合是根据客户机发出的服务请求动态生成的，可以看作数据源的动态视图。虽然它们只是整个数字图书馆文献资源集合的子集，但是可能包含更丰富的链接关系，而且这些链接关系是在实际应用中产生的，因此具有进一步分析研究的价值。

描述文档也是Atom机制中不可缺少的，大都是对数据格式的说明，用以支持各类应用服务。例如，实现查找功能就需要检索说明文档的支持。图2中描述文档的路径标识为“/”，实际应用中并不要求一定部署在根路径下，但是需要在相关应用中给出文件链接。

基于Atom机制重构数字图书馆资源描述框架主要面临两大风险：

风险一：实体对象的分类与标识是整个框架的基础，通用属性的定义目前只是一个草案，可以利用Atom提供的扩展机制按需定义。然而，一旦出现各自规定的可能，在现实应用中就必然存在不同的元素与表示方式，尤其在面临复杂资源的时候，这种风险更大。

逐步完善框架应用规范是减少风险的可行方法，主要包括以下方面：

- 必备的元素；
- 元素修饰词；
- 所采用的编码体系修饰词及值；
- 有可能使用的、取自其他命名域的元素；
- 有可能使用的、取自其他应用纲要的元素或元素修饰词；
- 各种限定约束的规范、标准、无歧义。

风险二：现实环境中存在着多种元数据描述格式，并且已经积累了大量宝贵资源。采用Atom连锁格式重构资源描述框架，绝不可能丢失这些信息财富。如何继承现有数据，是信息重构必须首先解决的问题。

MARC在图书馆领域有多年的应用历史，笔者选取CNMARC格式，与本框架之间进行对照，见表2^[12-14]。

表2 CNMARC与DL-RDF/Atom格式对照

CNMARC	DL-RDF/Atom
001 本地系统控制号	feed/{entry}/id
200\$a正题名	feed/{entry}/title
200\$d并列题名\$e副题名	feed/{entry}/subtitle
责任者名称: 701\$a / 702\$a / 711\$a / 712\$a	feed/{entry}/author@name
著作责任: 701\$a / 702\$a / 711\$a / 712\$a	feed/{entry}/author@responsibility
任职机构/地址: 701\$p / 702\$p	feed/{entry}/author@organization
任职机构/地址: 701\$p / 702\$p	feed/{entry}/author@address
101\$a作品语种	feed/{entry}/lang
105 / 110 数据编码 (内容类型)	feed/{entry}/type
010\$aISBN	feed/{entry}/code
011\$aISSN	
016\$aSRN	
017\$a其他标准号	
022\$a政府出版物号	
040\$aCODEN	feed/entry/content
330\$a提要、文摘或全文	
610\$a非控主题词	
606\$a受控主题词	feed/{entry}/category@term
分类号: 675\$a / 676\$a / 680\$a / 686\$a / 690\$a / 692\$a	feed/{entry}/category@scheme
分类法版本: 675\$v / 676\$v / 680\$v / 686\$v / 690\$v / 692\$v	
215\$e附件	feed/{entry}/link
856\$u电子资源URL	rel="enclosure"
106\$a物理介质	feed/{entry}/formatinfo@type
215载体形态	feed/{entry}/formatinfo@extent @medium
其他文献特殊细节:	
135 / 230电子资源	
194中国古籍	
206测绘制图资料	
208录音资料与印刷乐谱	feed/{entry}/formatinfo@version
205\$a版本说明	
215\$a页数	feed/entry/pageinfo@totalpages
210\$c出版发行者名称	feed/{entry}/publisher@name
210\$a出版发行地\$b出版发行者地址	feed/{entry}/publisherinfo@address
326\$a出版频率 (连续出版物)	feed/{entry}/publisherinfo@frequency
210\$d出版发行日期	feed/{entry}/published@date
801\$a国家\$b机构	feed/{entry}/generator@label
801\$c处理日期	feed/{entry}/updated

表2是从CNMARC格式到本框架的对照关系。由于MARC格式本身只是对文献著录项目、格式、顺序的规定,因此派生出若干针对特定文献类型的著录规则,如GB/T3792系列国家标准、国家图书馆编著的《中国文献编目规则》。更多实际应用又在标准框架下按需扩展。显然,信息重构不是一个简单的对照转换过程。期刊文献的著录就是一个典型的例子。在MARC标准中只规定了如何描述期刊,并没有关于一本刊里文献的著录说明,因而衍生出不同的解决方案。例如,国家科技图书文献中心的文献加工细则规定采用463字段连接篇与刊的MARC记录,在篇级著录中463\$v存放物理本的年卷期,形如\$v2000,vol.30,no.4。类似这样的著录细则,在做信息重构时就需要编写特定的转换规则。

由于MARC源于传统的图书馆编目工作,虽然为适应数字化、网络化发展陆续进行了扩展,但是仍然难以满足数字图书馆多资源类型、多载体格式的描述需求,尤其在网络信息资源的描述方面存在先天缺陷。在此背景下DC元数据标准应运而生,并逐渐在业界得到广泛应用。笔者选取DC图书馆应用纲要(DC-LIB)2004年9月版本^[15],给出了核心元数据与本框架之间的对照关系,见表3。

5 结语

从资源描述的角度来看,如果要完全解决元数据互操作问题,需要尽可能地兼容各种元数据格式,或者穷尽细节做到全面覆盖,或者采用抽象的方法,通过宽容和松散

表3 DC LIB与DL-RDF/Atom格式对照

DL LIB		DL-RDF/Atom
dc:identifier		feed/{entry}/id
title	dc:title	feed/{entry}/title
	dcterms:alternative	feed/{entry}/subtitle
creator	dc:creator	feed/{entry}/autor@name
	dcterms:role	feed/{entry}/author@responsibility
contributor	dc:contributor	feed/{entry}/author@name
	dcterms:role	feed/{entry}/author@responsibility
dc:publisher		feed/{entry}/publisherinfo@name
subject	dc:subject	feed/{entry}/category@term
	dcterms:DDC/LCC/LCSH/MESH/UDC	feed/{entry}/category@scheme
description	dc:description	feed/entry/content
	dcterms:abstract	feed/entry/content@abstract
	dcterms:tableofcontents	feed/entry/content@tableofcontents
date	dcterms:issued	feed/{entry}/published@date
	dcterms:modified	feed/{entry}/updated
dc:type		feed/{entry}/type
format	dc:format	feed/{entry}/formatinfo
	dcterms:extent	feed/{entry}/formatinfo@extent
	dcterms:medium	feed/{entry}/formatinfo@medium
dc:source		feed/{entry}/link rel="related"
dc:relation		feed/{entry}/link rel="related"
dc:language		feed/{entry}/lang
dc:coverage		feed/{entry}/category@term
rights	dc:rights	feed/{entry}/rights@copyright
	dcterms:accessrights	feed/{entry}/rights@accessrights
dc:edition		feed/{entry}/formatinfo@version
dc:location		feed/{entry}/code@codeno

来支持应用。这两种技术路线是正确的，沿途也布满了荆棘，沿着任何一条路线坚持走下去都可能构造出理想的、自圆其说的体系，然而却直接面临着应用上的困难，或者实现起来过于复杂，或者需要自定义的部分太多，最终互操作问题仍然得不到解决。在“特殊”和“抽象”之间寻找平衡，是框架走向实用的关键。

本框架构想从文献资源的“整体/部分”关系进行梳理，目的在于利用Atom机制快速构造数字图书馆网络服务，解决元数据互操作问题。实体对象的分类与标识是整个框架的基础，通用属性的定义目前只是一个草案，可以利用Atom提供的扩展机制按需定义。接下来的主要工作是对实体属性持续不断地扩展与改进，使其更结构化，便于理解，从而有利于机器的识别与处理。

资源描述框架研究工作只是一个开始。Atom机制的实现需要Atom连锁格式与Atom发布协议相结合，本框架的研究属于对Atom连锁格式的领域应用与扩展，除了对其完善改进之外，还需要对Atom发布协议进行扩展，以支持数字图书馆领域内的典型应用。同样，寻找合适案例，开发原型系统，也列入下一步研究计划。

参考文献

[1] 毕强,朱旺玲. 元数据标准及其互操作研究[J]. 情报理论与实践,2007,30(5):666-670.
 [2] 孙雨生. 基于ontology的数字图书馆互操作机制研究[J]. 情报资料工作, 2007,3:79-81.
 [3] 刘炜. 基于本体的数字图书馆语义互操作[D]. 上海:复旦大学,2006.
 [4] The Atom Syndication Format [EB/OL]. [2009-4-22]. <http://www.ietf.org/rfc/rfc4287.txt>.
 [5] SNELL J. An overview of the Atom 1.0 Syndication Format [EB/OL]. (2005-08-02) [2009-4-27]. <http://www.ibm.com/developerworks/xml/library/x-atom10>.

html?S_TACT=105AGX52&S_CMP=cn-a-x.

- [6] The Atom Publishing Protocol [EB/OL]. [2009-4-22]. <http://www.ietf.org/rfc/rfc5023.txt>.
- [7] SNELL J. Getting to know the Atom Publishing Protocol, Part 1: Create and edit Web resources with the Atom Publishing Protocol [EB/OL]. [2006-10-07]. http://www.ibm.com/developerworks/xml/library/x-atompp1/index.html?S_TACT=105AGX52&S_CMP=cn-a-x.
- [8] Google Data API概述[EB/OL]. [2009-4-27]. <http://code.google.com/intl/zh-CN/apis/gdata/overview.html>.
- [9] AtomServer 2.1.9 [EB/OL]. [2009-4-27]. <http://atomserver.codehaus.org/index.html>.
- [10] 刘炜,葛秋妍. Web2.0 技术图书馆应用分析[EB/OL]. [2009-4-27]. <http://www.libnet.sh.cn/sztsg/fulltext/reports/2006/libraryTech20.pdf>.
- [11] Jangle Specification Version 1.0 [EB/OL]. [2009-02-24]. <http://jangle.org/drupal/spec/1.0>.
- [12] 新版中国机读目录格式使用手册[M]. 北京:书目文献出版社,2005.
- [13] CALIS联机合作编目中心标准与文件[EB/OL]. [2009-6-17]. <http://www.calis.edu.cn/CALIS/lhml/lhml.asp?fid=FA0307&class=2>.
- [14] 国家科技图书文献中心文献数据加工细则[M]. 北京:腾图电子出版社,2001.
- [15] DC图书馆应用纲要[EB/OL]. (2004-9-10) [2009-6-17]. http://dc.library.sh.cn/dc_lib.htm#Abstract.

作者简介

王莉(1974-),女,毕业于北京师范大学信息管理系,硕士,现就职于中国科学技术信息研究所信息技术支持中心,高级工程师。通讯地址:北京复兴路15号中国科学技术信息研究所信息技术支持中心 100038。E-mail: wangli@istic.ac.cn

A Lightweight Resource Description Framework of DL Based on Atom Syndication Format

Wang Li / Institute of Scientific & Technical Information of China, Beijing, 100038

Abstract: In the view of metadata interoperability, this paper provides a resource description framework in the field of digital library. The DL-RDF/Atom is based on the Atom Syndication Format in order to use the built-in features to simplify the procedure of data integration, both to ensure the integrity of the original data and can achieve metadata interoperability without loss of data. First, this paper outlines the technologies of metadata interoperability, introduces the basic characteristics of Atom and explains the reason of choosing Atom. After the analysis of Atom applications in the field of DL, this paper describes the Digital Library Resource Description Framework based on Atom, gives a journal literature sample, and analyzes a DL resource network based on Atom mechanism. This paper points out the risk of DL-RDF/Atom and methods to minimize the risk. Finally, this paper lists the contents of the future research.

Keywords: DL-RDF/Atom, Atom, AtomPub, DL, RDF

(收稿日期: 2009-04-28)