

●王松林

## DC-Lib——我国数字图书馆元数据的首选

**摘要** 建立图书馆新的元数据标准既要考虑传统的元数据标准,又要考虑到非图书馆人员的实际应用情况,要把建立新的元数据标准提高到建立一种新的交换格式的高度来认识。《DC图书馆应用纲要》既符合中文文献特点又符合我国数字图书馆需求的元数据格式,可作为我国数字图书馆首选元数据。表4。参考文献10。

**关键词** 数字图书馆 DC图书馆应用纲要 元数据 网络信息资源  
**分类号** G254

**ABSTRACT** The author thinks that when we establish a new library metadata standard, we should consider both traditional metadata standards and practical conditions of non-librarian professionals, and should regard the new metadata standard as a new communication format. DC-Lib (DC-Library Application Protocol) can meet both the characteristics of Chinese document processing and the needs of digital libraries in China, and can be the first choice of metadata for digital libraries in China. 4 tabs. 10 refs.

**KEY WORDS** Digital library. DC-Library Application Profile. Metadata. Network information resource.

**CLASS NUMBER** G254

在众多的元数据项目中,DC在图书馆界和情报界可以说是应用最广、影响最大的一个国际性项目<sup>[1]</sup>。但是即使两家单位同属图书情报服务界(服务领域相同),均都使用DC(采用的元数据相同),由于对DC元素的语义理解不同,并使用了不同的扩展机制,他们之间仍不能较好地数据进行数据交换。为此,DCMI后以DC 15个元素为基础,并根据各自领域制定各自元数据的应用纲要(Application profiles)<sup>[2]</sup>。

目前,先后开发的DC应用纲要有:《DC-Education元数据》<sup>[3]</sup>、《DC政府应用纲要》<sup>[4]</sup>和《DC图书馆应用纲要》等。其中,《DC图书馆应用纲要》(DC-Library application profile,以下简称“DC-Lib”)草案最初于2001年8月发布于DCMI的网站上,之后于10月12日作了修订,其修订稿在2001年10月22~24日东京举行的第9次DC会议的DC图书馆工作小组会议上成为讨论的主要内容<sup>[5]</sup>。2002年,DC-Lib进行了两次修订,日期分别为2002年4月16日和2002年9月24日<sup>[6-7]</sup>。经过几次修订,DC-Lib现已趋于稳定。本文先简单介绍DC-Lib最新版的体例结构及内容,然后提出可将它作为我国数字图书馆首选元数据的理由。

### 1 DC-Lib的体例结构

在DC-Lib中,每一元素、子元素(SubProperty)和修饰

词(含元素细化和编码系统)都以统一的体例结构给出,即每一元素、子元素和修饰词都呈如表1的形式。

表1 DC-Lib统一的体例结构

名称	分配给元素修饰词的惟一标记。
标签	分配给元素修饰词的人读标签。
命名域选取	http://purl.org/dc/elements/1.1/, http://purl.org/dc/terms/, http://purl.org/dc/dcmitype/, http://dublincore.org/usage/terms/dc/ current-schemes/,或 http://www.loc.gov/mods
DC元素细化	用于DC-Lib的DC元素细化:这些修饰词使一个元素的意义更加狭窄或专指。一个被细化的元素虽然带有未经修饰的元素含义,但其范围更为限定。
…的子元素	来自被细化的DCMES中的元素,相当于“老的”DC元素细化。
DC-Lib元素细化	本应用纲要早先版本中的DC-Lib元素细化现已成为DC元素细化,或成了另一命名域的元素。
DC编码系统	这些修饰词标识帮助说明元素值的各种系统。这些系统包括各种受控词表和规范标记或解析规则。因此,一个使用某一编码系统表达的值可能会是选自某一受控词表的标志(例如取自一部分类法或一

续表

DC 编码系统	套主题词表的标志)或一串根据规范标记格式化的字符(例如作为日期标准表达的“2000-01-01”)。即使一个编码系统不为客户机或代理所理解,其值对读者来说还是有用的。需要指出的是,有些编码系统现在虽然还未注册,但对 DC-Lib 来说又是必需的。这些编码系统今后将作为 DC 编码系统被 DCMI 使用委员会进行注册或通过。
元素应用级别	在本应用纲要中,元素应用级别可以是:必备(M)、有则必备(MA)、强烈推荐(R)及供选择(O)。必备(M)确保某些元素总被使用,有则必备(MA)表示若能获得有关信息则该元素一定被使用。一个应用级别为必备(M)的元素,肯定有一个值。强烈推荐(R)和可供选择(O)的元素指如果信息适合该资源则应用值填入,如果信息不合适该资源,它们则可省略。
DC 定义	元数据的 DC 定义。
DC 注释	对特定元数据的 DC 注释。
DC-Lib 定义	元数据的 DC-Lib 定义。
DC-Lib 注释	对特定元数据的 DC-Lib 注释。
最佳操作	用于 DC-Lib 的特定元素的最佳使用介绍。
未解决的问题	有关特定问题、附注。

从表 1 命名域选择的单元可以看出,DC-Lib 的命名域由 5 个命名域组成。其中,http://purl.org/dc/elements/1.1/为《DC 元数据元素集,1.1 版》;http://purl.org/dc/terms/为《DC 修饰词》;http://purl.org/dc/dcmitype/为《DC 类型词汇表》;http://dublincore.org/usage/terms/dc/current-schemes/为《DC 编码系统》;http://www.loc.gov/mods为《用于〈DC-Lib 应用纲要〉中的 MODS 元素》。

所谓命名域(Namespace),是用来对 XML 中由不同域(Domain,例如是由不同团体所定义的不同标记体系)所定义的标记元素作出明确的注释机制。即若引用了某一个集合的元素标记,就应指明其命名域,以说明引用的元素标记的语义定义在哪个地方找到,以致于可以在同一个文档内用命名域机制引用不同元数据集的元素<sup>[8]</sup>。

需要指出的是,并非所有 DC-Lib 元素、子元素或修饰词都有表 1 所列单元。

## 2 DC-Lib 的基本内容

作为一种元数据,DC-Lib 共有 18 个元素组成(见表 2)。

表 2 DC-Lib 的 18 个元素

元素名称	标签	元素名称	标签	元素名称	标签
题名	Title	日期	Date	关联	Relation
创建者	Creator	类型	Resource Type	覆盖范围	Coverage
其他责任者	Contributor	格式	Format	权限	Rights
出版者	Publisher	标识符	Identifier	读者对象	Audience
主题	Subject	来源	Source	版本	Edition
说明	Description	语种	Language	馆藏位置	Location

这 18 个元素,除“读者对象”的命名域选择 http://purl.org/dc/terms/(即《DC 修饰词》)、“版本”和“馆藏位置”的命名域选择 http://www.loc.gov/mods(即《用于〈DC-Lib 应用纲要〉中的 MODS 元素》外,前 15 个元素的命名域均选择 http://purl.org/dc/elements/1.1/(即《DC 元数据元素集,1.1 版》)。这说明 DC-Lib 主要基于 DC 和 MODS,而 MODS(Metadata Object Description Schema,元数据对象描述模式)

则是 LC 下属的网络发展与 MARC 标准机构正在研制的一种针对书目记录元素集的 XML 模式(eXtensible Markup Language schema)<sup>[9]</sup>。

在 DC-Lib 以上 18 个元素中,题名、说明、格式、日期、关联和覆盖范围等 6 个元素还有各自的子元素(SubProperty)。所谓子元素,如上所述,是指“来自被细化的 DCMES 中的元素,相当于‘老的’DC 元素细化”。以上元素所含有的子元素见表 3。

表3 题名等6个元素的子元素

元素	子元素	元素	子元素
题名	交替题名(alternative)	说明	文摘(abstract)
关联	是…另一版本(isVersionOf)		目次表(tableOfContents)
	是…的另一格式(isFormatOf)	日期	创建日期(created)
	有格式(hasFormat)		有效日期(valid)
	由…替代(isReplacedBy)		获取日期(available)
	替代(replaces)		发行日期(issued)
	是…的组成部分(isPartOf)		修改日期(modified)
	有组成部分(hasPart)	格式	篇幅(extent)
	需要(requires)		媒体(medium)
	由…参照(isReferencedBy)	覆盖范围	空间(spatial)
	参照(references)		时间(temporal)

从表3 DC-Lib的子元素设置情况看,它们的命名域均选择 <http://purl.org/dc/terms/>(即《DC修饰词》)。但与DC一般元素修饰词所不同的是,DC-Lib的子元素可以直接使用,而并非一定要与其所属的元素名称捆绑在一起。试比较DC-Lib子元素“交替题名”(alternative)和原DC-Lib的题名元素修饰词“题名|交替题名”(Title|Alternative)。这样无疑将更方便网络信息资源提供者或编目人员使用DC-Lib。需要强调的是,DC-Lib中的子元素并非新创出来,而

全部都是从原DC-Lib中的修饰词转换而来。这与DC所走的道路不同,即DC先用的是“堪培拉修饰词”和“子元素”,后来用的是“DC修饰词”。

除了上述元素及子元素外,DC-Lib的主题、说明、日期、类型、格式、标识符、来源、语种、关联、权限和馆藏位置等11个元素另外还有各自的修饰词(含元素细化和编码系统)。由于它们的命名域来源复杂,表4同时列出这些元素的修饰词及其命名域。

表4 主题等11个元素的修饰词

元素名称	修饰词
主题	主题  DC编码系统(subject   DC encoding scheme(s),命名域选择 <a href="http://dublincore.org/usage/terms/dc/current-schemes/">http://dublincore.org/usage/terms/dc/current-schemes/</a> )
说明	说明  URI(description   URI,命名域选择 <a href="http://dublincore.org/usage/terms/dc/current-schemes/">http://dublincore.org/usage/terms/dc/current-schemes/</a> (将注册))
日期	版权日期(dateCopyrighted,命名域选择 <a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a> ); 提交日期(dateSubmitted,命名域选择 <a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a> ); 接受日期(dateAccepted,命名域选择 <a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a> ); 截获日期(dateCaptured,命名域选择 <a href="http://www.loc.gov/mods/">http://www.loc.gov/mods/</a> ); 日期  ISO 8601(date   ISO 8601,命名域选择 <a href="http://dublincore.org/usage/terms/dc/current-schemes/">http://dublincore.org/usage/terms/dc/current-schemes/</a> (将注册))
类型	类型  DCMI类型(type   DCMIType,命名域选择 <a href="http://purl.org/dc/dcmitype/">http://purl.org/dc/dcmitype/</a> ); 类型  DC编码系统(type   DC encoding scheme(s),命名域选择 <a href="http://dublincore.org/usage/terms/dc/current-schemes/">http://dublincore.org/usage/terms/dc/current-schemes/</a> (将注册))

续表

元素名称	修 饰 词
格 式	格式   IMT(format   IMT,命名域选择 http://purl.org/dc/terms/)
标识符	标识符   URI (identifier   URI,命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/); 标识符   DC 编码体系(identifier   DC encoding scheme(s),命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/(将注册)); 标识符书目引证(identifierCitation,命名域选择 http://purl.org/dc/terms/(已提出))
来 源	来源   URI(source   URI,命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/); 来源   DC 编码系统(source   DC encoding scheme(s),命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/(将注册))
语 种	语种   ISO 639-2(language   ISO 639-2,命名域选择 http://purl.org/dc/terms/); 语种   RFC 1766 (language   RFC 1766,命名域选择 http://purl.org/dc/terms/); 语种   RFC 3066(language   RFC 3066,命名域选择 http://purl.org/dc/terms/)
关 联	关联   URI(relation   URI,命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/); 关联   DC 编码系统(relation   DC encoding scheme(s),命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/(将注册))
权 限	权限   URI(rights   URI,命名域选择 http://dublincore.org/usage/terms/dc/current-schemes/(将注册))
馆藏位置	馆藏位置   DC-Lib 编码系统(location   DC-Lib encoding scheme(s),命名域选择 http://lcweb.loc.gov/marc/sourcecode/organizations/)

### 3 DC-Lib 作为我国数字图书馆首选元数据的理由

DC-Lib 的引言中说,DCMI 的图书馆工作小组对图书馆界及相关应用领域的 DC 元数据元素集的多种运用作了探索,并对它们的可能运用作了如下展望:①在运用不同元数据标准格式的不同体系之间充任一种交换格式;②用于挖掘(Harvesting)图书馆内外数据源的元数据;③支持对不同系统内的资源简单建立图书馆目录记录;④通过对 DC 的转换使 MARC 数据为其他团体所识别;⑤考虑到从使用 DC 的非图书馆作者那里获取资源发现元数据。这些告诉我们,建立图书馆新的元数据标准既要考虑传统的元数据标准,又要考虑到非图书馆人员的实际使用情况,更为重要的是要把建立新的元数据标准提高到建立一种新的交换格式的高度来认识。

DC-Lib 的引言还说,一种图书馆应用纲要成为一种规范,需作如下定义:①必需的元素;②许可的 DC 元素;③许可的 DC 修饰词;④许可的系统与值(例如某一特定受控词表或编码系统的使用);⑤来

自别的不同命名域的图书馆界元素;⑥从其他应用纲要移植来的附加元素、修饰词(如从 DC-Ed 移植来的读者对象);⑦标准定义的元素细化。这些定义(参 DC-Lib 元素、子元素或修饰词的应用级别)的提出,也为我国图书馆建立中文元数据规范指明了方向。同时也看到,DC-Lib 只是诸多 DC 元数据的一种。虽然目前它还是一种可能的应用纲要,但它足以说明 DC 元数据元素集完全适用于图书馆及其相关应用领域或项目。

应用纲要这一概念源自 DCMI 内部,作为一种方法,它阐明什么命名域的什么元素可用于某一特定应用领域或项目。即应用纲要被定义为一种由从一种或多种命名域中抽取出来的数据元素所构成的模式,这些元素被实施者组织在一起,可有效地用于某一个特定应用领域。这就告诉我们,我国图书馆在发展中文元数据的问题上,要跟踪国际元数据及其应用的发展动向,制订出既符合国际发展潮流、又结合我国图书馆特点的中文元数据规范。这一点,DC-Lib 的元素“读者对象”从 DC-Ed 借鉴过来,以及元素“版本”和元素“馆藏位置”从 MODS 借鉴过来,足以引起我们的思考。现在既然有了一种既符合中

文文献特点又符合我国数字图书馆需求的元数据格式,我们完全可以采取“拿来主义”的方针<sup>[10]</sup>。

除了以上将 DC-Lib 作为我国数字图书馆首选元数据的理由外,还需强调的是:①现在人们大多将 DC 作为数字图书馆元数据首选,殊不知制定 DC 的初衷是想将其嵌入到所描述的资源对象中去(一般采用 HTML 置标语言),而主要不是将其作为描述既有资源对象的元数据格式,尽管它能这样做。但从以上论述看,DC-Lib 主要是作为描述既有资源对象的元数据格式,而且在运用不同元数据标准格式的不同体系之间充任一种交换格式,尽管也能像 DC 那样将其嵌入到所描述的资源对象中去。②DC 主要用于描述源网络信息资源,所以在它的 15 个元素中不包括诸如读者对象、版本和馆藏位置等元素及修饰词。而 DC-Lib 既适用于描述源网络信息资源,也适用于描述非数字化的实体信息资源,这从它在 DC 15 个元素的基础上又增加了版本尤其是馆藏位置元素可以看出。

为了加深对 DC-Lib 的特点认识,再对 DC-Lib 中的读者对象、版本和馆藏位置这 3 个元素作些分析:

读者对象(Audience)在 DC-Lib 中被 DC 定义为:资源面向或对之有用的实体级别;被 DC 注释为:读者对象实体级别可由创建者或出版者或第三方确定。而 DC-Lib 对其注释为:2001 年 10 月 DCMI 使用委员会通过。未解决的问题是:需要评估哪些元素细化或编码系统适合 DC-Lib 使用;MARC 中的读者对象代码可作考虑。从增设读者对象情况看,DC-Lib 除受 DC-Ed 影响,也深受 MARC21 除连续出版物、测绘资料和混合性资料以外的 008 字段的字符位 22“读者对象”,以及 UNIMARC 100 字段字符位 17-19“读者对象”的影响,并且拟从 MARC 中抽取读者对象代码。

网络信息资源极易被修改或更改,而且一些网络信息资源并不注明修改或更改的次数和时间。受时效性及各种不确定因素的干扰和影响,有些网络信息资源还一直处于波动和无规律状态,其创制、更迭和消亡情况一般难以预料。而 DC-Lib 有了版本(Edition)元素就能详细说明特定作品的版本(Version or edition),这对成功发现资源和判明资源异同,尤其对频繁发生变化的资源往往十分关键和特别重要,但版本元素不用于不同物理格式意义上的版本(例

如文本资源的 PDF 版);目前的元素(edition)在 MODS 中作为 <publicationInfo> 子元素使用;2002 年 5 月,DCMI 使用委员会已通过决议。

网络信息资源编目的重点由拥有(Ownership)转向存取(Access)。而现在 DC-Lib 有了馆藏位置(Location)元素,就可用以标识藏有该资源或从其存取该资源的机构。对物理文献使用馆藏位置元素,可使用用户在 URI 不适用时(例如物理形式的文献不能以电子形式获取)找回物理形式的文献,同时也便于在 URI 不能找回任何东西或仅能找回有限书目信息时进行检索;但在一个馆藏位置中若有更深的识别内容(例如索取号、登录号),可以考虑使用标识符。在 MODS 中,使用目前的元素 <location>;2002 年 5 月,DCMI 使用委员会也通过决议。

#### 参考文献

- 1 吴建中.战略思考:图书馆十大热门话题.上海:上海科学技术文献出版社,2002
- 2 真濂.矛盾重重的元数据世界(二).图书情报工作,2002(10)
- 3 Education Working Group:Draft Proposal. <http://dublincore.org/documents/2000/10/05/education-namespace/>
- 4 DC-Government Application Profile:DC-Gov. <http://dublincore.org/groups/government/profile-200111.shtml>
- 5 Report of the DCMI Libraries Working Group sessions at the DC 2001 conference in Tokyo, Japan. <http://www.jisemil.ac.uk/cgi-bin/wa/exe?A2=ind0111&L=dc-libraries&D=0&p=58&F=P>
- 6 Guenther, Rebecca. DC-Library application profile :DC-Lib. <http://dublincore.org/documents/2002/04/16/library-application-profile/>
- 7 Guenther, Rebecca. DC-Library application profile :DC-Lib. <http://dublincore.org/documents/2002/09/24/library-application-profile/>
- 8 吴建中.DC 元数据.上海:上海科学技术文献出版社,2000
- 9 The Library of Congress. Network Development and MARC Standards Office. Metadata Object Description Schema. <http://www.loc.gov/standards/mods/>
- 10 王松林.元数据及有关思考.情报学报,2002(4)

王松林 南京政治学院上海分院信息管理系教授,硕士研究生导师。通讯地址:上海市。邮编 200433。

(来稿时间:2003-03-25)