

网络环境下编目与 MARC 的未来*

刘 莎 (武汉大学信息管理学院)

【摘要】 论文通过透视编目的发展演变和网络时代的编目趋势,分析 MARC 应用于网络资源编目的优势与不足,比较其与众多元数据方案的异同,提出改进方案和优化措施,预测 MARC 在网络时代的命运。

【关键词】 MARC;元数据;元数据对象描述模式;可扩展标记语言;联机公共目录查询

【Abstracts】 Through perspecting the development of cataloging and its trends in the age of Internet, this paper analyzes the advantages and disadvantages of MARC applying in network resource cataloguing, compares it with many other metadata schemes, proposes some improvement programs and optimization measures, and predicts the fate of MARC in the Internet age.

【Keywords】 MARC; Metadata; MODS; XML; OPAC

MARC最初是为了简化和加快编目过程而设计的,然而,随着计算机和通信技术的发展,MARC 的应用范围日益扩大。目前,MARC 已广泛应用于国际间的数据交流,已成为一种通信格式。它具有以下特点:适用于所有类型的信息载体;有足够的灵活性可适应除书目生产外的各种应用;可用于多种自动化系统中^[1]。MARC 的命运与编目的发展息息相关,网络时代的 MARC 已经走到了命运的转折点。

1 编目的发展演变

文献信息资源编目是图书馆一项十分重要的基础业务,编目活动是文献信息服务得以顺利开展的根本保证。图书馆实行自动化管理后,建立自动化的图书馆目录系统更是图书馆各项业务工作实现计算机管理的核心工作。

图书馆编目技术经过了传统的手工编目、计算机编目,发展到现在的联机合作编目^[2]。面对互联网的迅猛发展,图书馆的编目工作逐渐将网络资源纳入著录范围,这个阶段的编目与传统的编目有着不同的特点,我们可以将网络时代的编目趋势简单概括为以下几个方面。

(1) 编目格式简单化

出于编目成本效益的考虑,编目界一直在探索如何在满足用户需求的前提下适当地简化编目。核心级记录(Core Level)就是在这种背景下出现的,MARC 记录级别的简化是互联网时代的必然选择。

(2) 编目外包化

外包部分甚至全部编目任务给图书馆以外的机构成为很多图书馆的选择,以往选择外包较多的是缺乏专业编目人员的中小图书馆,随着联合编目的发展,套录在编目工作中所占比重日益增加,需要较高专业知识和经验的原始编目减少,外包逐渐扩大到大型图书馆。

(3) 数据来源多样化

在网络环境下,数据来源的多样化是为了解决外包质量不高的问题。Lorcan Dempsey 总结图书馆中使用的描述性元数据有 4 种:专业的、贡献的、程序获取的以及收集用户使用的数据。

(4) 联合目录本地化

联合目录本地化不但要求图书馆是联合目录的成员单位,而且还要求图书馆的所有馆藏信息都提交给联合目录,这对联合目录及成员馆都是一个很大的挑战^[3]。

MARC 是计算机编目的产品,是详细的、严谨的、专业性要求高的一种元数据。MARC 家族分支众多,派系林立。从 20 世纪 60 年代初的 LCMARC (俗称 MARC1),到 1968 年推出 MARC2,1973 年国际标准化组织 (ISO) 将 MARC 格式结构作为国际标准 (ISO2709-1973) 正式颁布,到了 80 年代中期,MARC 格式针对不同类型文献资源分别推出了 7 个系列格式。为了解决不同格式、不同载体的著录问题,LC 开始研制一体化 MARC——USMARC。之后随着时代的变化,MARC 一直保持着与时俱进的特色,不断更新版本,1999 年已经更新至 MARC21。这又是编目发展史上的一个

* 本文系国家社科基金项目“国际性编目规则及其发展研究”(项目编号:11BTQ005)的研究成果之一。

里程碑——一套适用于网络时代的、通用的、开放性的书目信息格式。

在四个编目未来发展趋势中,其中编目格式简单化是顺应时代的产物,然而后面的三种变化都会对 MARC 在编目中地位的弱化产生或大或小的影响。网络环境下用户选择的多样化加剧了这种不利局面的扩展,在这种环境下,“MARC 过时论”的说法引起了广泛而激烈的争论。

2 网络环境下 MARC 的优劣及其与元数据的对比

MARC 自上世纪 60 年代产生于美国国会图书馆以来,已经历了半个世纪的实践检验。这期间,由于书目数据的制作和应用环境发生了很大的变化,如数据库技术、计算机技术和网络技术的高速发展,文献数量的迅速增加以及网上资源的大量出现,图书馆的文献组织方式与服务手段都发生了根本性的变化,使 MARC 在实际运用过程中,遭遇部分使用者的批判,如“MARC 已经过时”、“继续采用 MARC 标准,将阻碍图书馆事业的发展”等观点。网络环境的急剧变化,让 MARC 遇到了自诞生以来最大的挑战。

2.1 MARC 进行网络资源编目的优势

无序海量的网络信息资源只有经过有序加工才能发挥它的应有价值,MARC 作为一种用于传统文献资源编目的元数据,对于网络信息资源整序同样可以发挥它的作用。当前网络信息资源管理模式可以分为三种。

(1) 书签式的管理模式(Bookmark)

用户可以通过创建书签来将自己所需的超文本集中链接管理,但这种方式只能为用户提供浅层链接,还需点击才能进入网站浏览网页,它不仅没有提供检索查询功能,也没有进行深层挖掘,在链接大量信息资源时有明显的局限性。

(2) 搜索引擎式的管理模式

利用搜索引擎,用户可以方便地浏览网页从而查找自己所需要的信息。目前大部分搜索引擎是通过自然语言结合布尔逻辑进行检索,结果按词频、更新时间及关键词在文中出现的位置排序,很少有提供控制词检索及字段检索。同时限于语义或句法,关键词检索和全文检索很容易检出许多与检索主题无关的文献。

(3) 目录式管理模式

顾名思义这种模式就是将网页信息进行分析归入不同学科或主题中。表面看来,这种管理模式似乎是比较理想的方法,然而实际运用发现并不易操作。用户为查询所需信息,需要逐级访问各级类目,直到找到所要找的目标。这种检索模式会让用户无功而返,因为用户只能试探性点击自己认为对的目录,一旦与系统归类不一致,最终只能得到南辕北辙的检索结果。

以上三种网络信息资源管理模式尽管能在一定程度缓解网络资源的混乱状态,但是相比于 MARC 对于文献信息的揭示功能还是存在欠缺。书签式只能链接表层资源,搜索引擎式的自然语言检索模式局限颇多,目录式又缺少与用户的交互反馈,它们都存在着不同程度的信息标引障碍和信息检索困难,所以网络信息资源完全可以借鉴成功而且成熟的传统文献编目方法。

MARC 是用于描述、存储、交换、控制和检索的一套机读书目数据标准,主要针对于印刷型书本的描述。它的结构严谨,能确保其数据元素组成具有统一性,有利于数据交换。而且 MARC 还提供了灵活的著录级别,同时在 MARC 格式的书目数据里可供检索的部分都有清楚的标识,多种检索点可以有效地支持多途径检索。随着网页 OPAC 的发展,利用 MARC 对网络上的资源进行编目,用户可从 OPAC 中同时查到馆藏及网络上的资源。并且在标准的规范下,数据一方面可以进行交换,另一方面能促进联合编目的发展;透过同一界面,读者可以从一个检索系统获得各种不同形式的资源,这是实现网络信息资源与传统信息资源一体化检索的最简单方法^[4]。另外,利用 MARC 编目网络信息资源比利用 DC 元数据描述网络信息资源更加具体、准确,更能取得用户信赖。

从世界范围来看,绝大部分的书目记录都是依据 MARC 和 AACR 方式编制的,只有 2% 左右的数据采用了其它著录方式^[5]。不论从数据描述的丰富性,还是多数据检索的查准率看,MARC 和 AACR 都是极具优越性的。尤其是对 MARC 的不断调整完善解决了虚拟信息资源的编目难点,使 MARC 格式更加适应了网络信息资源管理模式的发展。因此,MARC 的发展前景很好,也将继续保持其在资源著录上的优越性。

2.2 MARC 进行网络资源编目的缺陷

机读目录可以方便地实现一次输入、多项检索,且检索速度快、质量高,如若再配以各种先进的现代化通讯设备,还可以实现联机检索。同时,MARC 格式经过多年来的发展,是迄今为止系统最完善、类目最复杂、标准最严格的元数据格式。但也正是 MARC 所具有的这些优点,使它在实际应用时受到局限。

(1) 受制于卡片目录的思维

MARC 编制从一开始就深深地打上了卡片目录的烙印,从实际效果来看还存在着重复与冗繁,对于网络资源的编目若沿用 MARC 的传统格式难免产生诸多不适。

(2) 机读格式复杂的结构与标识

复杂的结构与标识使得书目数据的提供者局限于专业层面,并且格式中各项规定过于硬性、详细、没有文本标识灵

活^[9]。网络环境下,搜索引擎还不能实现对这些目录的检索,限制了 MARC 对网络资源进行编目。

(3) 规定繁杂,应用难度大

在理解运用 MARC 编目时,由于其繁杂的规定很容易导致理解上的差异。繁杂的规定还要求图书馆系统的编程人员必须是 MARC 专家。程序员对 MARC 的理解也各有差异,这一差异更加削弱了 MARC 详尽编目的优势。可以预见,网络资源本身的海量无序加之 MARC 编目的不一致,必将导致资源标识十分混乱。

(4) MARC 是封闭的而不是开放的

受习惯性思维影响,MARC 是无法与外部数据库系统相互操作的封闭性系统。MARC 资源无法直接面向万维网开放,无法通过搜索引擎向因特网用户提供链接。此外,MARC 在所规定的字段和字节外,不可能随意添加字段与内容。

(5) 编目对象的单一化

MARC 诞生之初只是针对纸质文献编制的,但是网络时代编目的对象已经发生了翻天覆地的变化^[10]。

实际上,MARC 局限于书目数据的控制,而不利于对原始状态的评论、索引和音像文件进行处理。对于那些现在和将来目录中需要进一步强化著录的附加信息,目前并无一个公认的便于交流的标准形式,所以数据缺失现象严重。

2.3 元数据与网络资源编目

由于 MARC 对网络资源编目的缺陷,众多专门用来描述网络资源的数据格式,即元数据应运而生。当前应用十分广泛的有 DC 元数据、MODS 等。这些元数据有自己的特色,在描述网络资源方面与 MARC 相比有很大的优势。

(1) DC (DublinCore)

DC 是在人们为解决网络资源无序化而提出的,是为网上资源的辨识、检索而制定的一种著录格式。DC 自 1995 年产生后即深受瞩目,其以建立一套简洁而灵活,可使非专业图书馆员也易于掌握和使用的资源著录格式为目标,所以其特色为语意明确、简单易用性、语义的互用性、国际一致性、灵活性,一度被部分学者视为即将取代 MARC 的后设资料。

表 1 DC 与 MARC 的比较^[8]

元数据 特征	Dublin Core	MARC
易用性	简单易用,非专业人士亦可用	复杂繁琐,需专业人士方可用
灵活性	灵活简便,可适应于不同数据承载模型	硬性规定众多,需专门的输入系统
著录对象	主要是面对网络信息资源的组织与整理	侧重对传统印刷型文献的标引编目
著录目的	可供网络搜索引擎读取、链接,并建立相应的索引信息	提供给图书馆读者检索馆藏图书之用
揭示关系	不能解释资源之间相互关系	可部分揭示资源间各种关系

然而经过近 10 年的考验,DC 虽简洁明了,但是其字段的定义趋于模糊,容易形成理解及操作上的不一致,加上缺乏得以提供参照结构的方法,没有规范控制,导致检索率低。为便于实际利用,DC 相关的描述项目及描述语也在一次次的会议中逐步增加,于是它原先为数位物件描述所设计的特点,反而成为难以克服的障碍。

(2) MODS (Metadata Object Description Schema)

元数据对象描述模式 (MODS) 是美国国会图书馆于 2002 年 6 月开发的一种新的文献编目元数据,它是继 MARC 之后开发的第二种以 MARC 为基础的文献编目元数据。MODS 是 MARC 元素的子集,但比 MARC 格式简单,它的字段标识符不是数字而是语言,是使用语言字段标识符描述实体的模式。MODS 的检索点一方面是重视图书馆应用而开发的元数据;另一方面,MODS 是在 MARC 格式著录元素的基础上利用 XML 语言,希望能在很多领域应用的元数据,也就是说,任何形式的资料都可以用 MODS 著录。已有的用 MARC 格式著录的图书馆馆藏信息利用 MARC-CXML 转换软件可以互相转换,并能够进行校正。

MODS 与 DC 比较,DC 的 Relation 元素几乎与 MODS 的 relateditem 元素表示同样的内容,但 DC 中的 Relation 元素不具有下位结构,关联资源的元素只能用任意文本表示,所以,不能区别和分析数据,只能依靠关键词检索。从 MARC 转换成 MODS 的过程中,不会丢失数据,如果转换成 DC 就会丢失数据。

3 MARC 的改进

质疑 MARC 繁琐与落后的理由主要是针对它的结构,因此需要加强的一是封装技术问题,二是字段构建问题。针对这两个问题,我们可以从以下几个方面着手对 MARC 进行改进。

对于可读性差,MARC 可以在编制软件时做出一些补充,通过查表让编目人员知道数字的含义是什么即可。其次,MARC 字段设置繁琐易造成理解的不一致,这方面可以采用统一的规范控制处理。再次,MARC 三段式的设计结构和字节式的存储模式都是早期遗留下的缺点,在计算机软件和硬件技术高度发达的今天,这些都不足以成为问题。最后关于 MARC 不能被网上的通用搜索引擎检索到的问题,有相关学者提出用 XML 解决,并给出了相应的算法,可以预见,这一问题的解决不会太遥远。当然,修改 MARC 格式要遵循一条基本准则:向后兼容。最理想的情况是使现有的系统无需做任何修改就能识别新的 MARC 数据——尽管在大多数情况下这一条过于苛刻^[9]。

尽管 MARC 格式存在着许多问题,但 MARC 格式为图书馆提供了非常优秀的揭示文献书目信息的工具,到目前为

止还没出现能取代它的格式。种种迹象表明,1998年IFLA颁布的《书目记录的功能需求》(Functional Requirements for Bibliographical Records)和《可扩展标记语言》(Extensible Markup Language)将是后MARC时代的走向。促进MARC改进的因素很多,现试举其中影响最大的三项分析如下。

(1) FRBR对MARC的冲击

FRBR认为编目对象不能停留在传统的平面层次上,应根据用户的需求将编目对象分成若干层次。MARC数据区虽然设置并定义46-多层次字段来反映条目等级关系的连接,但这种连接是线性的、平面的,虽然MARC本身的结构复杂,但对复杂的层级关系的适应性却较弱,MARC只是描述性元数据,难以适应FRBR的多层次等级立体新框架。建立FRBR模型后,我们可以规范元数据的互操作。FRBR的提出将使MARC面临巨大挑战,其冲击将是深远的。

(2) 可扩展标记语言 (Extensible Markup Language, XML)

XML来源于SGML,是一种简洁、灵活的文本格式。用XML编写的著名内容格式是《元数据目标描述方案》(MODS),它是MARC21格式的简化形式,使用语言化描述而非字段码,MODS更接近MARC21的解决方案。然而,它对于目前MARC21存在的大部分问题都没有给出解决方法。目前,最新的发展动态是用SGML或XML来对MARC进行重新包装,这将使书目数据格式从底层就具备面向WWW的开放性特征,保证了从内到外,从处理到交换、检索的一致性,有利于实现广域网环境下的资源共享^[10]。

(3) OPAC

MARC最终是通过OPAC来显示结果的。伴随着信息技术的不断发展,图书馆OPAC呈现着新的发展趋势:增加任意关键词搜索入口,为用户提供更多的检索途径;提供相关度分级及确切的题目匹配,提高检索结果的准确性;提供多种浏览方式,用户可以自行定制搜索结果的显示方式;提供个性化服务功能;可以解决多语种输出;可以通过掌上电脑、手机等来使用等。OPAC的发展必然推动着MARC的改进。

4 MARC的未来

MARC伴随着图书馆目录的发展变迁一路走到了21世纪,它的“生死问题”从未像今天这样引起如此激烈广泛的争论。1997年美国国会图书馆对400万条MARC数据进行了字段使用统计,即TOP 33:The 33 most frequently used fields in LC USMARC data“用得最多的33个USMARC字段”^[11]。2007年12月31日,耗时3年的对MARC指定内容利用率的调查项目全部完成。MCDU(MARC Content Designation Utilization)项目用于完成编目员关于

MARC指定书目内容的利用情况的实证调查。这些实证数据将会有助于解决图书馆机构当前信息检索系统中对MARC未来命运和它在快速变化的网络信息中的地位的疑问。调查发现,编目员对MARC记录格式的使用影响着供应商系统和终端用户系统的感知行为以及最终用户成功检索、选择和获取信息的效率^[12]。

MCDU的调查数据不仅可以说明在编目过程中编目员通常使用的字段和子字段,帮助用来构建更加便于数据交换的简洁版MARC;也可以结合FRBR思想,将作品与载体分开,验证在FRBR中定义的五条用户任务——发现、识别、选择、获取、导航;同时,它还可以用来鉴定在书目记录中广泛使用的基于特殊格式的记录集与现存的由国会图书馆推荐使用的国家级、核心级和迷你核心级数据。

表2 至少使用一次的字段的数量与百分比
(数据基于国会图书馆书目数据库)^[13]

Field Tag	Number of Records Where Field is Used at Least Once	Percentage of Records Where Field is Used at Least Once
001	7,595,887	100.000%
003	7,595,887	100.000%
005	7,595,887	100.000%
006	974	0.013%
007	210,789	2.775%
008	7,595,887	100.000%
010	7,595,726	99.998%
013	1	< 0.001%
015	1,178,815	15.519%
016	224,984	2.962%
017	821	0.011%
018	2	< 0.001%
019	943,393	12.420%
020	3,845,934	50.632%
022	147	0.002%

这些数据说明MARC仍具有强大的生命力——MARC是为了适应从手工编目向计算机编目的转变而产生,现在为了适应联机合作编目和网络资源的编目,它同样会做出积极的改变和尝试,但是究竟这种改变能否保证它不被时代抛弃,抑或是确保它在未来元数据编目中的一席之地,答案需要实践检验才能给出。

5 结语

MARC经过多年的发展,其自身的数据量姑且不论,单看其对自身规则的调整:从91条到字典式目录规则,到柳别茨基的编目规则与原则,再到巴黎原则到书目记录功能需求,无不是顺应时代与技术的变化而变化的,从简单的目录编制到联机联合编目,无不是为图书馆及用户的需求而不断地调整着自身的基因,这就说明了一条,也是最重要的一条,MARC的存在不是固步自封而是在不断地创新。未来的MARC将是一个完整的文献信息组织的最(下转第18页)

题 名		著 者	
财务会计习题集		王君彩	
学科相关性	10	入藏时间	2009
流通次数	6	预借次数	3
内容	7	排架号	F234.4/121
读者水平	教师(0*10)+学生(6*5)=30		
Value	$10 \times 6 \times 30 \times 3 \times 7 \div (2011-2009)=18900$		

表4 采访经费预算(100万元)

中图法人类	Value	所占比例
F	$(18900+671666.6667) \div 2$	33.7
TP	$(226059.9409+158567.25) \div 2$	18.77
I	$(503471.0472+35867.667) \div 2$	26.32
...	...	...

4 结论

本文针对现在图书馆采购的一些所见,并结合一些自己

的工作经验,加以总结得出上述结论,但这些结论中也存在着不足,需要更加深入地去理解分析。比如说没有加入外部信息的数据模型,所采用的全是馆内分析的方法,分析结果不完整。馆内分析的相关因素不全面,存在一些客观因素无法加入。举例来说,本来很有价值的书,可是却因为这本书知道的人少,所以流通量很小;或者本来很有价值的书,可是专业性太强,无法产生大的流通量。这些因素都无法加入到该模型中。这些还有待进一步研究。

参考文献

- 1 王煜.浅论高校图书馆的采访工作[J].科技信息,2009(5)
 - 2 程晓红.网络环境下图书馆采访工作[J].科技信息,200(5)
 - 3 向桂林,吴金艳,辛希孟.采访质量控制数学模型研究[J].中国图书馆学报,2003(1)
- (作者信息:广东科技学院图书馆馆员,邮编:523083 收稿日期:2011-04-18。)

编校:彭 飞

(上接第22页)好方式。

参考文献

- 1 冷红中.MARC之我见[J].图书馆建设,2008(5):29~32
- 2 饶思军.从 MARC 编目到资源整合[R].武汉:华中农业大学图书馆,2007.
- 3 编目精灵.编目的未来[M].北京:国家图书馆,2010.3~7
- 4 樊晓红.浅析利用 MARC 格式进行网络资源编目[J].图书馆学刊,2005(2):67~68
- 5 张海英,张艳玲.网络环境下信息资源 MARC 格式编目[J].图书馆学研究,2002(4):57~59
- 6 唐志诚.图书馆 MARC 现状及其发展[J].前沿,2004(1):191~194
- 7 冷红中.MARC 的未来[J].图书馆建设,2008(5):29~32
- 8 罗红燕,李章平,陈绍兰.MARC、DC、MODS、FRBR 等文献编目元数据比较[J].图书馆学刊 2009(12):25~27
- 9 傅立云,刘新.MARC 格式中存在的不足及其改进[J].图书情报工作,2005,49(11):85~87
- 10 罗军.MARC 未来及质量控制[J].现代情报,2009,29(3):216~218
- 11 杨莉.MARC 字段使用频率统计对编目发展和文献检索的启示[J].山东图书馆季刊,2007(2):82~98
- 12 MCDU.MARC Content Designation Utilization: Inquiry and Analysis[EB/OL].<http://www.mcd�.unt.edu/>,2011-05-07.

- 13 William E M, Shawne D M.Format Content Designation Analysis: Data Report Frequency Counts for Books, Pamphlets, and Printed Sheets Records Created by Library of Congress (Set 01_B_LC)[R].Texas: University of North Texas, Texas Center for Digital Knowledge, 2006.

(作者信息:武汉大学信息管理系 2010 级图书馆学硕士研究生,邮编:430072。收稿日期:2011-06-23。)

编校:彭 飞

