

DC元数据年度进展 (2007)

□ 刘炜 / 上海图书馆 上海 200031

摘要: 都柏林元数据2007年年会以“元数据应用纲要的理论与实践”为主题,总结了一年来所取得的进展,讨论了DCMI在组织架构和运作方式上的一些变化,特别宣告了DCMI将与新加坡国家图书馆局合作进行公司化运作。在元数据标准规范方面提出了被称为“新加坡框架”元数据应用规范。这个规范遵照刚刚修订完善、正式发布的“DC元数据抽象模型”,规定了一整套元数据方案所应包含的五类文档,其中必备的“领域模型”和“元素集描述”要求以较为严格的机读形式编码(如UML和XML/RDF),以保证元数据应用系统的各项形式化约束具有严格的一致性和延续性。虽然在描述规范细节方面此次会议并没有推出进一步规范,甚至没有达成一致意见,但“新加坡框架”的提出无疑将成为DC元数据应用推广历程中的一个重要的里程碑。

关键词: DC元数据, 进展, 综述, DC-2007

1 年会概况

2007年度的“都柏林核心(DC)与元数据应用国际研讨会”于8月27日至31日在新加坡洲际大饭店(InterContinental Singapore)顺利举行,一共有来自33个国家和地区的近两百名代表参加了会议。这是继东京(2001)、佛罗伦萨(2002)、西雅图(2003)、上海(2004)、马德里(2005)、墨西哥曼萨利尼(2006)之后的第七届以“元数据国际研讨会”冠名的会议,作为都柏林核心元数据的系列会议,这已经是第十五次了^[1]。

本次会议的主题是“元数据应用纲要的理论与实践(Application Profile: Theory and Practice)”,经过数年的研究、实践和标准化工作的推动,本次会议在元数据应用纲要的形式化方面取得突破性进展,提出了一整套以DC元数据抽象模型为基础的元数据形式化方案,被归纳为“新加坡框架(Singapore Framework)”,为下一阶段开发元数据软件工具,以及编制可用性和互操作性强的元数据方案提供了一个统一的模型和编码基础,有望使元数据应用长期滞后于实践需求的状况得到根本性的改善。

与往届会议相比,本次会议并不算人数最多、规模最大,交流的论文也只有16篇,但取得的成果是令人瞩目的。本次会议一共进行了3场大会交流

发言和2场分会场发言,包含了此次会议文集所收录的全部论文。会议还跟往年一样,进行了4场专题培训、两场主旨报告、15场专题组讨论和3场特别会议。其中一场主旨发言由是我国“文化资源共享工程”管理中心副主任张晓星博士所作,题为“基于DC的文化资源的组织、共享和过滤(DC-Based Organization, Sharing and Filtering of Cultural Resources)”,引起了与会代表的浓厚兴趣。这是我国第一次在国际上全面详细地介绍这个初具规模的信息化工程,许多听众都将这项工程定位于一个国家级的数字图书馆。

会议还决定下一届DC元数据年会将于2008年9月22-26日在德国柏林举行,主题为“语义与社会性应用中的元数据(Metadata for Semantic and Social Applications)”,由德国哥廷根大学、德国国家图书馆等5家德国研究机构主办。会议的主办方代表在会上介绍了筹备计划和当地的情况。

本次会议的另一个重要成果是DCMI决定与新加坡国家图书馆局进行深层次的合作,在2008年6月之前合作成立非营利性公司,以进一步促进元数据的研究和推广应用。可以预见,这是DCMI组织机构和运营方式的一次意义深远的变革,对今后都柏林核心元数据的发展将产生重大影响。

2 DCMI标准化工作

DCMI是一个“虚拟的”标准化组织，近年来以“使信息查找更容易(Making it easier to find information)”作为其组织理念。说它“虚拟”，是因为到目前为止它还不是任何法律意义上的实体，没有在任何国家、经过任何形式的登记注册，只是一群人，为了制定一些标准规范，利用网络作为工具组成的一个学术团体。

DCMI是一个“不断发展的有机体”，一直致力于对其组织架构和运作形式进行变革，使其更加高效并富于成果。这或许与MBA出身、在国际著名咨询公司供职多年的执行经理(Managing Director) Makx Dekkers很有关系。今年DCMI提出与新加坡国家图书馆局合作成立非营利性的公共公司(Public Company)，预示着DCMI将告别“虚拟”时代，转而以某种“实体”形式存在。目前对于成立公司之后DCMI如何运作、是否会影响到目前的组织架构尚不清楚，但肯定能够更加有效地推动DC元数据的研究、应用和普及，更加富于成果，这一点是毫无疑问的。

DCMI的主要工作可以笼统地概括为三个方面：资源描述研究、标准规范制定、元数据推广应用。其中标准规范的制定是三者的核心，是资源描述研究工作的目的和结果，也是元数据推广应用的基础和前提。

早期DCMI致力于一套元素集的标准化，不论“简单DC”(即DCMES)还是“扩展DC”，都只是定义了属性元素集。相对而言，信息系统所要描述的大千世界，其丰富性是无穷无尽的，DCMI逐渐发现，寻求普适的属性描述集合无异于刻舟求剑、缘木求鱼，一套标准永远不可能适用于所有应用，不可能满足所有已知世界、甚至还有未知世界资源描述的需要。属性元素集虽然重要，但无限扩展下去是徒劳无益的，也失去了“核心”元数据的设计初衷。“核心”元素如何更好地起到共性描述的作用，作为“语义互操作”的基础，是一个非常值得探讨的问题。

于是DCMI将工作重点转向了领域应用，一方面保证了核心元素集的相对稳定，另一方面属性元素的扩展可以在一定的领域范围内进行。这就促成了“应用纲要”的提出。“DC元数据应用纲要(Dublin Core Metadata Application Profile)”规定了在领域应用

中对于资源描述属性元素的选取可以不局限于DC词表(即核心集和扩展集的并集)，还可以从其它相关标准规范中选取。应用纲要不仅是给人参考、应用和执行的，更重要的是要给计算机“读”，作为计算机处理数据的基础。每一个具体的专业或主题领域，甚至某一个具体的应用，都可以编制自己的“元数据应用纲要”，作为既符合标准、又满足个性化需求的元数据方案。

应用纲要的提出可以说是DCMI对于元数据标准规范建设和应用的一种创举。然而编制应用纲要应该如何选取元素？如何定义元素之间的关系？如何编制文档，并在“机读”和“人读”之间取得平衡？这些都需要遵循一定的原则和方法，并做出具体的规定，否则在应用中不可能达到基本的一致性，使“标准规范”应用成为了一句空话。

对于上述问题的探讨，包括提出元数据抽象模型、规范元数据应用纲要的构成、探讨元数据方案(Schema)的形式化，是DCMI近年来工作的核心。因此我们可以说，虽然DC元数据以15个核心元素而名扬天下，但实际上已远远超出了15个核心元素的范畴，甚至也不能局限于一套元数据词表，而是包括了一整套原则和方法的、在互联网时代进行元数据应用的标准规范体系。当然，目前这套体系还在不断发展完善中，还有待在应用中推广，在实践中检验。

3 DCMI年度进展述要

除了回顾修订ANSI/NISO 39.85:2001文本，使其以ANSI/NISO39.85:2007发布，回顾修订IETF RFC 2413，并以IETF RFC 5013发布新版标准之外，DCMI在2006-2007年度还修订了DCM元数据术语(DC Terms 2007年7月)、DC-Text草案(2007年4月)、XML编码指南草案(2007年6月)、RDF编码指南草案(2007年6月)、命名域政策推荐意见(2007年7月)、领域与范围(Domains and ranges)草案(2007年7月)、DCMI抽象模型推荐标准(2007年6月)和描述集纲要(Description Set Profile)草案(2007年8月)等八个标准规范(包括草案)文本，可以说取得了显著的进展，也印证了上文所说的，DC元数据应该被看成是一套标准规范

体系，而不仅仅是一套词表的说法。

在上述十项修订或草拟的规范中，就重要性来说，最后两项无疑具有非常突出的地位。“DCMI抽象模型推荐标准”在过去的一年里经过了多次严格的审定，在结构、内容和措辞上作了巨大的改动，最终由应用委员会审核通过，成为DCMI推荐标准。作为元数据应用纲要的模型基础，此标准的意义非常重大，严格来说未来不符合抽象模型的元数据规范，其元素是不能被应用纲要所采纳的。由于篇幅关系，对于这个模型的解读将另文阐述，这里不再赘述。最后一项“描述集纲要”（简称DSP）第一次对元数据应用纲要的形式化内容进行了明确的规定，以前的CWA14855 都柏林核心应用纲要指南只是一份给人读的说明性的文档，未涉及任何形式化定义问题。

DSP被认为是元数据应用纲要的核心内容。那么应用纲要还应该包括哪些文档呢？经过这一年以来的讨论，DCMI也初步有了一致性意见，即一套完整的元数据方案，应该包含以下五个部分：

- 功能需求说明（需要desirable）
- 领域模型（必备mandatory）
- 元素集描述（DSP: Description Set Profile）（必备mandatory）
 - 应用指南（可选）
 - 编码句法指南（可选）

这个元数据应用纲要的结构，被此次会议归纳为“新加坡框架”。具体内容及各部分之间的联系可以用下图¹表示。

4 DCMI现状报告

去年DCMI的基本组织构成已经发生了重大变化，由“工作组(Working Group)”变为“社区(Communities)”和“任务组(Task Groups)”两种形式。当然也有一个例外，Architecture Working Group由于其重要性，转变成了唯一一个论坛：Architecture Forum。一年以来DCMI围绕这两类组织开展运作，每个社区或任务组在原先的邮件列表基础上，纷纷建立起了Wiki平台，既具有文档标识、保存和版本管理的

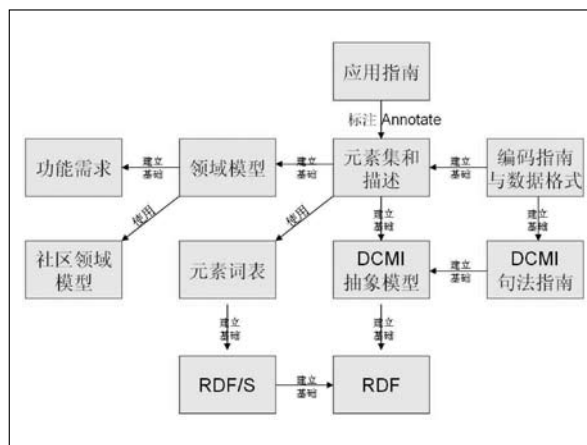


图1 元数据应用纲要“新加坡框架”示意图

功能，又满足社区“共笔”的需求。

组织架构上的变化确实给DCMI带来了新的变化，首先是前述DCMI三个方面任务（资源描述研究、标准规范制定、元数据推广应用）能够得到更加明确的贯彻。社区基本上负责承担领域应用的讨论和宣传推广，或对于任务组的具体成果（例如社区应用模型）进行讨论。任务组则进行领域元数据有关的方案制定或具体问题的研究和攻关，制定专门元数据应用纲要的任务一般交由任务组完成，某些任务组（例如DCMI/RDA）还可以“有偿”地完成专题研究。“架构论坛”已经不是一般的任务组，各任务组相关的元数据标准规范制定的问题（例如编码问题，一致性问题等）都可以提交架构组讨论，它主要负责技术规范的和解释。

上述组织架构结合DCMI原有的三角形管理架构：董事会、应用委员会和咨询委员会，形成一整套从规范动议、研究、提出草案、讨论审议、修订审批到最后发布推广利用的运行机制。另外，年会中与相关机构或其它标准化组织合作进行的“联合会议（Joint Session）”、“特别会议（Special Session）”等，都是发现问题或新的需求、宣传推广已有成果的绝佳机会。可以说在组织运作方面，DCMI显得非常成熟、自信，值得学习和借鉴的地方很多。

过去的一年中DCMI一共有11个任务组，14个社区进行了卓有成效的工作，两个任务组因为与社区或新建立的任务组有重叠而计划取消。年中

¹ 根据Tom Baker等图示翻译，原图参见：<http://dublincore.org/architecturewiki/SingaporeFramework>

陆续提出6个新建社区/任务组的提案，已有学术交流（Scholarly Communications）社区和标识符（Identifier）社区获准建立，根据DCMI的惯例，其它四个一般也都能获准通过。

表1列出了DCMI所有的“二级组织”的简要情况。详细情况请参考DCMI官方网站。

表1 DCMI的“二级组织”

论坛(Forum):	
架构论坛 (Architecture Forum)	负责讨论、制定和解释符合Web应用规范的元数据模型 (DC抽象模型)、应用指南、以及各种编码方式等规范。
任务组(Task Groups):	
RDF规范小组	探讨DC元数据的RDF编码问题。已停止运作，任务移交至架构论坛。
可获得性	由于没有明确的任务，信息交流和讨论交由同名社区执行。因此本任务组计划停止运作。
机构 (Agent) 描述	评估已有描述方案 (如FOAF) 能否满足功能需求。考察ePrints等项目对于机构描述的方案。
资源集合描述应用纲要	修订目前的资源集合描述元数据应用纲要提交应用委员会评估。完成NISO草案。评估与目前DCMI的抽象模型和DSP的一致性。
日期	考察各种日期描述需求，推荐并确认规范的日期描述和编码格式，考虑成为ISO/W3C日期格式规范的可能性。
教育资源描述应用纲要	建立教育资源描述的元数据应用纲要DC-Ed AP。考察领域应用模型、功能需求、资源的类型词表以及术语、已有的元数据方案、最佳实践等。与JISC、IEEE LTSC等相关项目密切合作。

参考文献

[1] 刘炜, 等. DC元数据的历史、现状及未来[M/OL]//《图书馆杂志理论学术年刊2005: 元数据与图书馆》. 上海: 上海科技文献出版社 2005:13-40 [2007-09-16]. <http://eprints.rclis.org/archive/00003408/>.

[2] 刘炜. DC元数据年度进展[J/OL]. 数字图书馆论坛2006(11):24-28. <http://eprints.rclis.org/archive/00010456/>.

[3] NILSSON.M. A presentation of Description Set Profiles and the Singapore Framework[EB/OL]. <http://dublincore.org/architecturewiki/SingaporeFramework>.

[4] DEKKERS M. Closing Presentation on DC-2007 [EB/OL]. http://dublincore.org/workshops/dc2007/Dekkers_DC2007_Final_20070830.pdf.

作者简介

刘炜(1966-), 男, 上海图书馆数字图书馆研究所所长, 研究员。通讯地址: 上海市淮海中路 1555号 200031

DCMI/IEEE LTSC 联合小组	更新LOM-to-DCAM映射表、LOM词表定义、示例等制定形式化的LOM元数据描述集纲要(DSP), 并建立从LOM-XML转换LOM-DCAM-in-RDF的GRDDL。
政府信息资源描述应用纲要	寻求可能的资助或者支持。
核心(Kernel)描述	继续完善目前草案, 翌年提交应用委员会审议。
图书馆应用纲要	征求意见, 继续完善, 提交应用委员会目前正在考虑或需要与其它小组协调的事项有: 复用MODS元素的问题, 馆藏地描述问题, 日期描述问题, 是否采用FRBR作为领域模型问题, 建立DSP问题, 编写应用指南和最佳实践问题, 考虑与RDA协调问题等等。
DCMI/RDA 联合小组	定义建立应用纲要所需的RDA RDF词表。定义RDA模型实体。以RDF或SKOS定义出版物有关的属性值词表。基于FRBR和FRAD制定RDA的DC元数据应用纲要。
推荐新成立的组织:	
可获得性社区	图书馆应用社区
资源集合描述社区	本地化与国际化社区
教育资源社区	资源保存社区
环境资源社区	登记注册应用社区
全球企业应用社区	社会性标签社区
政府信息社区	标准化讨论社区
核心(Kernel)描述社区	软件工具描述社区
推荐新成立的组织:	
标识符社区	主题图(Topic Maps)社区或任务组
注册登记任务组	软件工具描述应用纲要任务组
学术信息社区	书写系统描述任务组

(收稿日期: 2007-10-15)

Report on Dublin Core Metadata Annual Conference 2007

Liu Wei / Shanghai Library, Shanghai, 200031

Abstract: This is an annual report on Dublin Core Metadata Initiative and Review of DC-2007 Conference. The theme of this year's DC conference is "Application Profiles: Theory and Practice", focuses on the theory and practice of developing application profiles. It concluded with a specification named Singapore Framework, which consists of a set of documents providing the means of encoding to let the machine and human both readable. It also introduced with discussion of the major progresses on standardization activities by all the DCMI communities and tasking groups.

Keywords: Dublin Core metadata, DCMI, Annual report, DC-2007