

DC 元数据年度进展

□ 刘 炜 / 上海图书馆 上海 200031

摘要: 本文详细介绍了都柏林核心 (DC) 元数据 2006 年年会的基本情况, 以及 DC 元数据制订和维护机构 DCMI 近一年来的活动和所取得的成果, 对 DCMI 正在发生的变化和近期的活动进行了展望。

关键词: 元数据 都柏林核心 年会 DCMI

1 DC-2006 会议概述

“2006 年国际都柏林核心及元数据应用” (DC2006 年年会) 于 10 月 3 日至 6 日在墨西哥西海岸小镇曼萨尼约 (Manzanillo) 召开, 主办单位是墨西哥著名的科利马大学。会议以 “Metadata for Knowledge and Learning” 为主题, 吸引了大约 250 名元数据研究者和应用专家 (注册代表 187 名), 分别来自 24 个国家和地区, 在规模上超过了以往各届 DC 年会。

曼萨尼约是一个人口只有十余万的小城, 地处中美洲太平洋海岸, 火山众多、地貌丰富、风景秀丽、交通便捷, 是一个天然良港, 也是一个度假胜地。会议主办方将会址选在该城设施最为齐全的 Karmina Palace 酒店, 完善的服务和近乎封闭的半岛环境让代表们心无他骛、专事 “核心”。

会议沿用 2001 年日本 DC 年会以来的一贯程式, 四天的正式会议包括培训 (Tutorial Session)、主旨报告 (Keynote Speech)、大会论文宣读 (Plenary Session)、分组论文交流 (Paralleled Session)、工作组专题会议 (Workshop)、特别会议 (Special Session)、联合会议 (Joint Session) 等, 正式会议的前后还有应用委员会 (Usage Board)、咨询委员会 (Advisory Board) 和理事会 (Board of Trustees) 的会议各一天, 所以整个一周都是在紧张的会议中度过。年会是 DCMI (Dublin Core Metadata Initiative, 都柏林核心元数据首创计划) 最重要的活动形式, 除此之外只有几个委员会和少数工作组需要在两次年会之间再开一次会议, 其余的活动几乎完全依赖网络进行。

大会安排了两场主旨报告, 紧扣本届年会的主题: 为了知识和教育的元数据。第一场主旨报告题为 “资源网络和信息流的共享管理” (Shared Management of Source Networks and Information Flows), 由拉丁美洲地区著名的情报专家、毕业于美国雪城大学的 Abel L. Packer 主讲, 他目前担任国际卫生组织拉丁美洲信息中心主任 (Director of Latin American and Caribbean Center on Health Sciences Information)。其人不愧是个情报专家, 从网络知识交流的角度, 探讨了学术情报传播和利用的一些规律性的现象, 发人深省。

另一场主旨报告安排在会议最后一天, 即 10 月 6 日上午, 由华盛顿大学信息学院的 Michael Crandal 主讲。Mike 是 DCMI 的老朋友和积极参与者, 曾作为盖茨美琳达基金会华盛顿大学项目的负责人, 华盛顿大学是 DC 研究和应用的重镇, 特别在教育元数据方面颇有建树, Mike 的主旨报告题目就是 “DC 教育: 产品、过程和人 (DC-Education: Product, Process and People)”, 高屋建瓴地回顾了 DC 对于领域应用的历史, 总结了取得的成就, 并对目前的问题和令人振奋前景进行了展望。

会议共交流了 27 篇论文, 其中 12 篇大会交流, 分三个主题: 元数据模型 (Metadata Models)、本体与控制词表 (Ontology and Controlled Vocabularies) 和教育元数据 (Metadata for Education); 小组讨论的论文有 15 篇, 分成四场: 元数据方案实施 (Implementation)、调整与共享 (Adapting and Sharing)、应用部署 (Deployment Issues) 以及领域中的资源

(Resources in Context)。大会交流论文大多是由DCMI的成员参与，因此水平较高，代表了DC元数据研究和应用的最新进展。

DC年会中的培训是DC普及推广和本地化的重要举措，虽然每年的培训内容大同小异，基本上都包括DC元数据语义(DCMES、DCQ标准的介绍)、DC元数据编码(XHTML、XML及RDF，今年增加了模型的介绍)、规范词表(如SKOS等)、应用纲要四个方面，但每年都有新的内容，而且由这些元数据大师来介绍他们的孩子，别有一番深意，特别体现在提问与回答阶段，深入浅出，挥洒自如。

会议结束后提出一个“DC2006曼萨尼约宣言(Manzanillo Resolution 2006)”，主要基于许多的拉丁美洲(西班牙语和葡萄牙语区)参会代表的提议。网络文化一直是英美强势，DC元数据也不例外，亚洲、拉丁美洲都一直处于被动跟随地位，非洲更是一片荒漠，几乎连参会的成员都没有，因此这些地区的文化特殊性就不可能反映在元数据标准的制订中，对于其它语种无疑是不公平的，对于文化的多样性来说也是一种损失。近两届DC年会都在西班牙语区召开，因此本届会议顺理成章地提出这个宣言，对于促进拥有4亿人口的拉丁美洲地区来说是一个重要的进步。宣言提出在拉丁美洲地区促进元数据的普及和推广，倡议设立专门的实体或组织，召开专题会议、组织培训，建立西语DC网站，利用邮件列表、Wiki等技术手段展开全面的交流。所以这个宣言同时也是一个行动纲领。

DC年会前后的DCMI的工作会议(包括前述的三个委员会会议)构成了DC年度进展的重要内容，其讨论的内容涉及内部工作，具体成果将在下一年度工作进展中体现，在这里就不多做介绍了。

2 DCMI 现状报告

DCMI每年发布两次“现状报告”，每届年会之前的现状报告是DC元数据年度进展的重要参考，发布这个报告之前Makx Dekkers都会催促各委员会和工作组负责人对上年的工作进行及时的清理和总结，以便年会中汇总意见，并进行讨论决策。

2006年4月底应用委员会召开过一次中期会

议，以下内容根据2006年9月DCMI最新发布的现状报告整理。

◆ 对DCSV与DC抽象模型不一致的地方进行了修订。2005年DCMI抽象模型正式成为DC的推荐标准(Recommendation)，之前成为推荐标准的DCMI DCSV, DCMI Period, DCMI Point和DCMI Box就需要修订，主要是去除了以前暂时规定和列举的语法和描述，并在术语方面与DCMI抽象模型保持一致。详见<http://dublincore.org/usage/decisions/2006/2006-01.DCSV-revisions.shtml>。

◆ 对于DCMI类型词表(DCMI Type Vocabulary)在语义描述和使用指南方面进行了行文格式和术语的规范化(Editorial Changes)，详见<http://dublincore.org/usage/decisions/2006/2006-02.dcmitype-changes.html>。

◆ 因为要准备对NISO Z39.85-2001(都柏林核心元数据标准DCMES)进行五年一次的修订，基于与DCMI抽象模型保持一致的考虑，在行文上也作了少量修改。并于2006年8月28日至9月25日公开征求意见。

◆ 5月30日至6月27日对《都柏林核心元数据XML编码(Expressing Dublin Core metadata using XML)》和《都柏林核心元数据RDF编码(Expressing Dublin Core metadata using the Resource Description Framework (RDF))》两个文档进行了公开征求意见，这两个文档由DCMI的技术核心Andy Powell、Pete Johnston、Michael Nilsson等主笔，目前虽然还都是工作草案(Working Draft)，尚未提交应用委员会讨论，但与主持Web标准的W3C保持一致是非常重要的，因此这两个文档的意义和影响非同小可，可以澄清很多应用中的不确定性和随意性的问题，各类元数据应用在编码实现时都应该参考。这两个文件经过短期征求意见所获得的反馈意见可参见<http://dublincore.org/architecturewiki/DCRDFTaskforce/PublicCommentJune2006>。

◆ 决定对“DC资源集合描述应用纲要”于10月在曼萨尼约年会上进行发布和提交前的讨论，准备提交ISO、CEN和NISO等标准化组织，以满足应用的需要。

◆ 要求DCMI架构工作组对于与DCMI抽象模型相关的几个编码规范(DC-XML, DC-RDF and DC-TEXT)进行审定，主要是希望对词表编码体系修饰

词 (Vocabulary Encoding Schemes) 进行进一步的明确, 对于其应用的领域和范围 (domain 和 range) 进行规范。

◆ 鉴于 DCMI 的推荐标准在语义和形式上更为严格, 更加考虑机器处理的形式化需求, 增加和修订一批应用指南和著录规则也已成为当务之急。

◆ DCMI 的“附属会员 (Affiliates)”计划, 在 2006 年 4 月又增加了韩国国家图书馆, 此前已有芬兰赫尔辛基大学图书馆、芬兰国家图书馆、英国博物馆图书馆档案馆委员会 (MLA)、联合信息系统委员会 (JISC)、新加坡国家图书馆局、新西兰国家图书馆、新西兰政府档案馆和新西兰国家服务委员会等, 它们分别代表芬兰、英国、新加坡和新西兰参加 DCMI。

DCMI 目前有 17 个工作组, 作为一个纯志愿性的学术组织, DCMI 每年取得的大量进展都应该归功于这些工作组成员无私的辛勤劳动。过去一年这些工作组主要进行了以下工作:

① DCMI “可获得性”工作组 (Accessibility WG) 积极参与了国际标准 ISO/IEC JTC1 SC36 以及 W3C 相关工作组的活动, 主要成果体现在提出了一个普遍适用的可获得性描述框架: AccessForAll。国外对于信息系统 (包括网页) 的可获得性非常重视, 但是由于对可获得性的描述缺乏统一的规范, 在应用中颇为不便。DCMI 的可获得性工作组与 IMS 教育元数据组织 (IMS Global Accessibility)、欧洲标准化组织 (CEN ISSS Learning Technologies Workshop) 和国际标准化组织 (ISO JTC1 SC36) 一起制定了这个标准, 目前进展顺利, ISO 国际标准将于 2007 年推出。

② DCMI “代理”工作组 (Agents WG) 的工作仍主要在明确功能需求方面, 1998 年就成立的这个工作组至今还没有一个像样的描述“代理”的功能需求文档出台。相对于 FOAF、hCard 等标准格式的不断出台并得到普及应用, 虽然它们只满足了某一方面的需求, 但也走在了 DCMI 前面。

③ DCMI “架构”工作组 (Architecture WG) 的工作显然是非常重要和卓有成效的。除了前述的有关 DC 元数据编码的工作草案之外, 架构工作组还展开了命名域管理政策方面的讨论和研究 (结果参见: <http://dublincore.org/architecturewiki/namespacepolicy>), 以及探索应用纯文本进行元数据编码 (DC-Text) 的一套体系规范。命名域问题是国内元数据标准规范制定和应用中一个非常头疼的问题, 这个文档颇值得

我们参考。纯文本在描述元数据方面有着非常独特的优势, 具有很强的编码独立性, 正是由于目前各类编码语言的能力局限, 以及应用中的障碍, 使得 DC-Text 有望成为一种互操作性很强的中间编码格式, 此次 DC 年会 Pete Johnston 专门有一篇文章介绍 DC-Text。由于 DCMI 架构工作组日益体现出其重要作用, 特别是在今后与 Web 标准制定组织 W3C 将开展更紧密地合作, “架构”方面的标准、规范、规则、推荐意见及最佳实践将成为今后 DCMI 的主要工作, 所以该工作组即将升格为“架构论坛”, 以期能够发挥更大的作用。

④ DCMI “引文”工作组 (Citation WG) 在参考了各种正式或事实引文格式标准之后, 希望编写一份“书目引文 XML 编码指南” (Guidelines for the Dublin Core in XML Encoding of Bibliographic Citations), 目前该工作组的推荐规范是 2005 年 6 月 13 日版“DC 元数据进行书目引文编码指南 (Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata)”, 见 <http://dublincore.org/documents/dc-citation-guidelines/>。

⑤ DCMI “集合描述”工作组 (DCMI Collection Description WG) 可以说是一年中工作最勤奋的工作组, 所推出的“资源集合描述应用纲要”已经是草案第十二版 (参见 8 月 24 日版本)。许多内容比初期的草案已发生了很大的变化, 大多变化是由于需要同时满足揭示对象属性和集合属性两层描述而来, 例如扩展了“集合类型”元素 `cid:CDType` 以及其下的五个元素, 区分了“资源集合的描述”和“资源集合描述的描述”两个层次等, 这些变化经过了充分的讨论, 在本届年会上提交应用委员会审议, 目前已成为 DCMI 正式认可的第一个应用纲要正式版 (参见第 12 版的 status 状态: This is a Dublin Core Application Profile), 即将提交 ISO 和 CEN 审议。该应用纲要是目前 DCMI 在研的系列标准中最成熟, 能够很快成为国际标准的一个规范草案。但由于资源集合描述涉及的环境很复杂, 问题很多, 达成一致的困难也不小, 目前该工作组还在努力之中。

⑥ DCMI “教育”工作组 (Education WG) 过去一年的主要工作有两项, 一是继续完善 DC 教育资源描述应用纲要, 以期在尽可能短的时间内拿出推荐草案; 另一项是与 IEEE LOM LTSC 元数据标准化组

织合作, 提出一个复用 IEEE LOM 中的元素的推荐方案。DCMI 与 IEEE LOM 的合作一直很密切, 双方都可基于 RDF 的元数据抽象模型, 因此这两套元数据规范在互操作方面还是比较容易达成一致的。然而对于具体应用来说 DC 太过简单, 而 IEEE LOM 又太过复杂, 应用中一致性存在较大问题, 所以教育领域对一个灵活的、中等复杂程度的元数据方案的需求还是十分迫切的, 因此从 DCMI 方面来说尽快拿出一套适中的应用纲要规范仍然十分必需。

⑦ DCMI “政府”工作组 (Government WG) 继续完善应用纲要, 目前的草案是 2005 年 8 月的版本。除此之外, 该工作组重点讨论了“服务”的描述和规范词表的应用问题。该工作组建立了一个 Wiki 用于工作交流, 参见: <http://dublincore.org/dogapwiki/>。

⑧ DCMI “图书馆”工作组 (Library WG) 一直是一个人数众多、富于成果的小组, DC 图书馆应用纲要也是最早提交应用委员会讨论的规范草案, 目前的版本是 2004 年 9 月 4 日公布的, 但主要是碍于“应用纲要”这种形式的规范性和支持其互操作性的 DCMI 抽象模型一直没有成熟, 使得图书馆应用纲要还处在工作草案阶段, 从一定程度上影响了 DC 在图书馆应用的一致性和互操作性。过去一年中主要进行了两项工作: 其一是对 RDA 进行评估。由于美国国会图书馆推出了 AACR2 的数字升级版: 资源描述与存取 RDA, 与 DC-Lib 的应用领域有所重叠, 于是有人动议对 RDA 进行研究, 于是 DCMI 成立了一个子工作组, 与美国图书馆协会的 CC:DA (the Committee on Cataloging: Description & Access) 进行了接触。目前发现 RDA 还不足以对 DC-Lib 构成威胁。第二项工作是探讨在 DC-Lib 应用纲要中使用 MODS 元素的规范化问题, 由于这两者背后的模型相差太远, 给元素的语义互操作带来不少问题, 目前还没有很好的折中方案, 问题尚在讨论之中。

⑨ DCMI “本地化和国际化”工作组 (Localization and Internationalization WG) 一直致力于 DC 标准文本的翻译, 在世界各地的应用推广等工作, 它与登记注册体系工作组关系密切, 并且配合“DCMI 附属会员计划”的实施做了许多工作。

⑩ DCMI “保存”工作组 (Preservation WG) 一直试图明确保存性元数据描述的功能需求, 并提出了一套基于保存性需求的资源描述应用纲要, 然而

该小组取得的成果很少, 本届年会对是否继续保留这个工作组 (转变为社区) 也进行了讨论, 最后大家一致认为这个工作组还是很有必要的, 特别是可以结合 OAIS 模型提出一套应用纲要。

⑪ DCMI “注册”工作组 (Registry WG) 自从建立了 DCMI 元数据登记注册系统之后似乎就没什么事情了, 目前各类不同功能和作用的登记注册系统都陆续开发出来并投入使用, DCMI 的注册系统只实现了简单的术语维护和参考查询功能, 要真正实现基于 Web 的多种元数据服务, 道路还很漫长。

⑫ DCMI “工具”工作组 (Tools WG) 应该说是 DC 元数据普及应用的一个很重要的论坛, 然而长期以来该工作组定位不是十分明确, 也可能由于涉及面太广而难以集中在某一方面做些卓有成效的工作。过去的一年主要在六月份召开的美国“联合数字图书馆国际会议” JCDL2006 上组织了一次有关元数据应用工具的专题会议, 并在美国数字图书馆杂志 2006 年 7 月号发表文章对会上交流的内容进行了回顾总结。

⑬ DCMI 其它几个工作组基本上已处于停滞状态。“标准”(Standard) 工作组的主要职能已于 DC-2005 年会之后交由“DCMI 领导”小组 (DCMI Directorate) 承担, “ORDL/DCMI Profile 联合”工作组、“日期”(Date) 工作组、“环境”(Environment) 工作组以及“核心 Kernel”工作组都没有新的进展。

邮件列表是各工作组的主要交流和沟通形式, 近两年增加了 Wiki 方式。从邮件列表统计目前 DCMI 所有工作组加起来大约有 2239 名成员, 其中最大的 DC-General 邮件列表可以看成关于 DC 元数据的综合组, 拥有最多成员, 达 934 人。专门的工作组中人数较多的有图书馆工作组 (335 人)、教育工作组 (289 人)、政府工作组 (177 人)、架构工作组 (158 人) 和集合描述工作组 (144 人)。目前 DCMI 网站的月访问量大约在 15 万次左右, 唯一访客 (unique visitors) 每月在 8 万人左右。

3 DCMI 的近期动态

DCMI 在基层组织的组成形式方面正在发生重大变化。长期以来 DCMI 开展活动的基层组织“工作组 (Working Groups)”从 2006 年 11 月开始被“社区

(Communities)”和“任务小组 (Tasking Groups)”所取代,除了 Andy Powell 领导的“架构工作组 (Architecture Working Group)”升格为“架构论坛 (Architecture Forum)”之外,其它大多工作组都转为社区,并暂时拥有一个任务小组,而 AGENTS、DATE 和 KERNEL 三个工作组将直接转为任务小组(没有对应社区)。任务小组并不附属于社区,而具有更为明确的工作目标和时限,目前看来制定和维护相关领域应用的元数据应用纲要 Application Profile 和编写应用指南、著录规则等相关文档 (deliverables) 是任务小组 (TG) 的主要工作。而社区的构成比较简单,类似于一个俱乐部,职能主要是参与讨论、反馈意见等等。不一定所有的社区都拥有任务小组,某些社区可以拥有多个任务小组,某个任务小组也可以(例如日期 DATE 小组)向多个社区负责(主要是征询意见、达成共识)。

社区依旧以邮件列表(讨论组)作为交流方式,信息发布以传统网站为主要方式(目前各工作组的网站直接转为各社区的主页),而任务小组的主要工作载体则为 Wiki,也可拥有邮件列表,这并不是绝对的。所有的社区和任务组都要在现有工作组任务书的基础上重新编写任务宗旨说明文件(Charter),以阐明其设立的意义、任务、目的等等。预计 DCMI 的转型将于 2007 年 3 月结束。

进行这项改革的主要原因是为加强 DCMI 基层组织的活动,使 DCMI 更加活跃和富于成果。DCMI 将继续以前的开放性,并使其成果更符合 Web 技术(特别是语义 Web 技术)的发展潮流,使 DC 元数据规范真正成为互联网资源描述不可或缺的一环。

由“理事会 (Board of Trustees)”、“咨询委员会 (Advisory Board)”和“应用委员会 (Usage Board)”组

成的 DCMI 的“领导”架构将继续维持不变(当然领导“核心”一直是由 Makx Dekkers 和 Tom Baker 组成的指导委员会 Directorate)。其近年来积极推行的“附属会员机构 (Affiliates)”将依旧通过原来的机制(即参与理事会和咨询委员会的形式)参与活动,各社区和任务小组的负责人也通过参与咨询委员会的活动而影响决策。

下届 DC 年会将在新加坡召开,新加坡国家图书馆局已成为“DCMI 附属会员计划”的成员,召开年会也是会员所需承担的义务之一。在此次曼萨尼约年会上,新加坡国家图书馆管理局的信息技术主管 Raju 先生已经向大家展示了精心制作的多媒体宣传片和初步建设的网页,给大家留下了深刻的印象。下届年会估计正式会议的会期将缩短到三天,培训将集中于一天进行。对于扩大征文的范围,提高征文的质量也将采取一定的举措,例如设立一定的奖项鼓励优秀论文投稿等等。

由于 DC 元数据年会坚持开放原则,论文集必须在网上开放存取,所以不可能获得 Springer 之类出版商的支持,对国内需要发表 SCI 被引论文的大学教授、硕博博士研究生来说就失去了吸引力。但是 DCMI 偏重应用的风格应该成为我国图书情报领域科研人员走向世界的一个良好窗口,国内 DC 元数据的研究和应用也曾轰轰烈烈,取得了不少成果,但是随意进行语义扩展、命名域与编码的本地化规定稍嫌过多,与 DCMI 不一致的地方不在少数。如果能够通过参与 DCMI 的活动,一方面深入学习领会精髓,另一方面也可以把我们的需求带到 DCMI,通过参与讨论甚至影响标准规范的制订,岂不更好? 中图学会是否可以联合国内一些有实力图书情报机构,设立专门的奖金,资助一些中青年科研人员走出国门、参加 DC 年会呢?

参考文献

- [1] DC-2006 会议网站. [2006-11-1] <http://dc2006.ucoi.mx/>
- [2] Makx Dekkers. DCMI 现状报告 (2006 年 9 月). [2006-11-1] <http://dublincore.org/news/communications/statusreport-200609.shtml>
- [3] 刘炜等. DC 元数据的历史、现状及未来. 见: 图书馆杂志理论学术年(2005). 上海: 上海科技文献出版社, 2005
- [4] Makx Dekkers. DC-2006 Proceedings of the International Conference on Dublin Core and Metadata Applications, University of Colima (Mexico) and DCMI

作者简介

刘炜, 上海图书馆数字图书馆研究所所长, 研究员, 从事数字图书馆、知识组织、元数据研究。通讯地址: 上海图书馆数字图书馆研究所 200031

Annual Progress Report 2006 of Dublin Core Metadata Initiative

Liu Wei / Shanghai Library, Shanghai, 200031

Abstract: DCMI (Dublin Core Metadata Initiative) had just held a successful annual conference (DC-2006) at Manzanillo, Mexico on Oct.3-6, 2006. This paper reported the ongoing activities during the conference and through out the whole past year, outlined the annual progresses of DCMI and all its working groups, and made a brief introduction on the recent restructuring of DCMI working mechanism and the announcement of DC-2007 conference.

Keywords: Metadata, DC, Annual conference, DCMI