

DC元数据年会综述 (2012)

□ 赵亮 张春景 / 上海图书馆 上海 200031

摘要: 文章详细介绍了都柏林核心元数据组织 (DCMI) 2012年年会情况与特点, 也总结了历届年会的历史沿革与变化, 并通过重点介绍年会交流中如主旨发言等内容, 阐述了在元数据生态环境变化背景下, DCMI 应如何拓展思路, 应对未来发展路径的难题。文章还通过笔者对DCMI CEO和CIO等关键人物的访谈, 介绍了DCMI这一年的工作成绩与今后的发展方向思考, 也介绍了DCMI以元数据设计模式、词表映射和词表长期保存为主的工作方向。最后笔者提出了中国的元数据社区的专业人士应如何应对发展路径的思考以及如何对国际元数据界作出贡献的思路。

关键词: 都柏林元数据, DCMI, 年会, 定位, 元数据设计模式, 词表映射, 词表长期保存

DOI: 10.3772/j.issn.1673—2286.2012.11.007

2012年9月3日至9月7日, 都柏林核心元数据组织 (Dublin Core Metadata Initiative, 以下简称DCMI) 的年度大会 (以下简称DC-2012) 在马来西亚的古晋举行, 这正好是DCMI第20次年会。由于近年国际经济形势环境的恶化, 远离欧美的本届年会可能是DCMI会议从一个区域性的会议升格为国际会议后规模最小的一次。同时这届年会, 也标志着DCMI今后的工作方向将有比较大的转变, 这就如本届年会中由Dan Brickly所作的主题演讲题目中所设问的: 我们还有什么可做的 (What is left to do)?

1 年会历史回顾

1.1 DC 20的历程

翻开DCMI的历史, 在其18年的历史进程中, 本届年会正好是其第20次年会 (在1996年及1997年都开了两次会议)^[1]。自1995年起, DCMI的年会是元数据专家们的聚会。在2001年之前, DCMI年度会议的定名方式也是以序号来命名, 分别是DC-1一直排列到DC-8。DC-1到DC-7的会议都定名为“DC元数据研讨会” (Dublin Core Metadata Workshop), 然而在加

拿大渥太华举行的DC-8会议已然更名为“第八届DC元数据国际研讨会” (The 8th International Dublin Core Metadata Workshop)。从2001年在日本东京举办的第九届DCMI年会起, 年会不再以序号命名, 而直接以公元年号来命名。每年年会的名称也与其主题相关, 例如DC-2001的会议名称是“DC与元数据应用国际大会” (International Conference on Dublin Core and Metadata Application), 在意大利佛罗伦萨举办的DC-2002年会的名称就是“为数字社区服务的元数据: 支持多样性与相似性” (Metadata for e-Communities: Supporting Diversity and Convergence)。从DC-2005起, DCMI年会改为在“DC与元数据应用国际大会”的名称下, 增加反映当年年会主题的副标题。另外, 从DC-2001开始, DCMI年会基本保持了亚欧美三个洲轮办的规律。继2001年在东京, 2004年在上海, 2007年在新加坡, 2009年在首尔举办之后, 今年DC-2012又一次回到了亚洲, 在马来西亚沙捞越州的首府古晋 (Kuching) 举办。

1.2 本刊历年的会议介绍综述

本刊自2006年起, 每一年都对DCMI的年会与年度

进展进行了总结介绍。2006年DCMI的年会在墨西哥的曼萨尼约举行,由刘炜撰写的《DC元数据年度进展》^[2]除了较详细的会议介绍外,还介绍了DCMI的年度进展与相关热点动态。

2007年DCMI的年会在新加坡举行,由刘炜撰写的《DC元数据年度进展(2007)》^[3]还对DCMI的标准化工作的发展方向与原则以及年度工作进展作了介绍,其中重点是介绍了影响深远的元数据应用纲要“新加坡框架”的内容。最后,在介绍DCMI的现状中,此文较详细地介绍了DCMI的组织架构体系。2007年在DCMI的发展历史上,是一个非常重要的年份,不仅有元数据应用纲要“新加坡框架”的发布,也与RDA开展了深度合作,这一合作意味着DCMI被主流图书馆元数据界所认可。同时DCMI也定下了走向实体化的方向,DCMI也在2008年底正式在新加坡注册成为一个非盈利的实体组织。

2008年的年会在德国柏林举行,由刘炜、夏海撰写的《DC元数据年度进展(2008)》^[4]除了年会介绍,还讨论了DCMI年度的主要进展,以及结合会议内容讨论了相关热点。2009年在韩国首尔举行,由黄田青与刘炜撰写的《DC元数据年度进展(2009)》^[5]这篇论文对于2009年的年会着墨不多,主要是介绍DC的年度进展与社区及组织架构等等。

2010年的年会在美国宾夕法尼亚州的匹兹堡市举行。由沈芸芸、冯英、刘秀文撰写的《DC元数据年度进展(2010)》^[6]仍然沿袭了前面几年的大致结构,以年会概况、年度进展以及现状报告的形式开展。其现状报告对于DC的管理与运营架构介绍之详细程度是前几年年会的介绍论文中所不可比的,不仅对DC的组织结构作了清晰的描述,而且对DCMI的各个工作组与社区作了非常详尽的介绍。

2011年的年会在荷兰海牙举行,由赵亮、杨佳撰写的年会介绍论文《DC元数据年会综述(2011)》^[7]首次将本刊每年对DC的综述报告从“年度进展”更名为“年会综述”,因为这些综述性的文章都是由每年的年会参会者在参会后撰写的,其所描述的DC会议与年度发展内容基本上都来自于参会的心得、网站的资料,以及更重要的是与DCMI的成员进行交流的结果。这篇论文除了较详细地介绍会议的交流情况以及简单地描述年度进展外,最主要的变化是根据会议主办方对几位DCMI的主要成员所作的访谈内容,以及本文作者自己

与DCMI的CIO Tom Baker所作的访谈内容,做了一个对DCMI的历史与发展讨论的综述。

1.3 本届会议的特点

本届年会的主题是“面向全球性挑战的元数据”(Metadata for Meeting Global Challenges)。在会议网站上^[8],对这个主题有着如下的解释:“DC-2012将探索在应对诸如食品安全、数字鸿沟以及可持续发展方面全球性挑战时,元数据应用在全球、国家与区域中的作用。”应该说这个主题还是立意高瞻的,因为在应对所有这些全球性挑战时,信息系统几乎影响着一切。而在信息系统中,元数据的作用至关重要。

在会议形式上,继DC-2010之后,DC-2012是第2次与其他的机构组织合办会议。DC-2010是DCMI与美国信息科学技术协会(ASIS&T)一同举办的^[9]。但DC-2010和ASIS&T年会的合办属于一前一后,而DC-2012则是完全嵌入由MIMOS^①(MIMOS是隶属于马来西亚科技部的国有的信息与通讯技术研发中心,以下简称MIMOS)^[10]举办的2012知识技术周(Knowledge Technology Week 2012,以下简称KTW)^[11]活动中。KTW是马来西亚人工智能以及语义技术研究与实践者最重要的交流平台。

2012年的KTW于2012年9月3日至9月7日举办,一共有三个正式的大会。其中包括有“第12届环太平洋地区人工智能国际会议”(PRICAI 2012, The 12th Pacific Rim International Conference on Artificial Intelligence),“第15届多智能体系统的原理与实践国际会议”(PRIMA 2012, The 15th International Conference on Principles and Practice of Multi-Agent Systems)以及“都柏林核心元数据和元数据应用国际会议”(DC-2012, International Conference on Dublin Core and Metadata Applications)等。KTW还有一个人工智能的专题研讨会组合(AIW2012, Artificial Intelligence Workshops),其中包括多个国际论坛与主题研讨会。另外KTW还有相关一系列相关的培训、主题论坛、专题会议与演示活动等,其中包括“人工智能培训”、“博士生论坛”等等一系列活动。KTW是一个规模庞大、内容纷繁复杂的会议。在笔者拿到的会议秩序册的封面上,就列有包括DC-2012在内的九大项会议与活动内容。

相较于KTW的规模,嵌入在其中的DC-2012却很

^① MIMOS的全称是MIMOS Berhad,原名缩写含义为马来西亚微电子系统研究所。现在该机构仅仅是沿用了原来MIMOS的名称。

可能是自DC-2001以来规模最小的一次。由于近年来经济形势恶化的影响,欧美的许多研究与实践者难于承受远距离旅行所需的时间与经济成本,再加上马来西亚当地的元数据研究与实践者参会人数不算很多,造成本届DCMI年会参会者数量比往年要少很多,据介绍仅有70多位。相比在德国柏林举办的DC-2008创纪录的300多名以及在墨西哥小镇曼萨尼约举办的DC-2006约250名左右的参会人数,本届年会在规模上缩减了许多。人数的缩减也使得本届会议在讨论形式与内容上难于出新,类似DC-2011大量开放主题的非会议(un-conference)讨论就难于出现了。很多时候会议的讨论与交流也显得冷清许多。

2 本届年会概况

DC-2012的完整日程与KTW 2012完全重合。从9月3日至9月7日的进程包括2天的培训以及与之并行的各种内部或公开的管理层面或工作组层面的会议,和3天的正式大会安排。在正式大会的安排中,包括论文的交流

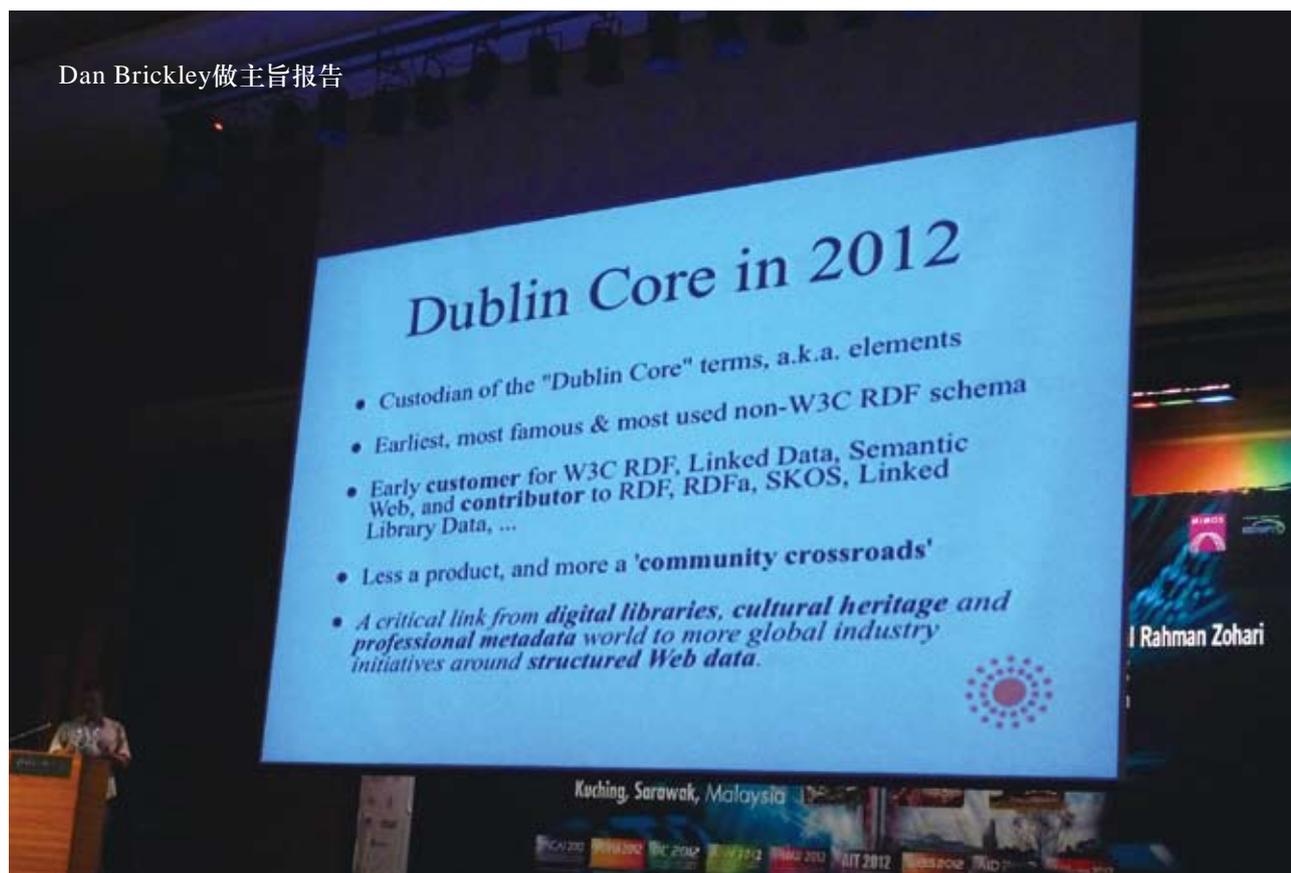
报告、项目报告、海报展示以及主旨报告、特别会议等。

2.1 培训与专门会议

DC-2012在9月3日、9月4日分别以“面向语义万维网的元数据”以及“元数据应用纲要与本体设计”为主题安排了两个整天的培训。9月3日的培训课程教师为图书馆界的独立咨询师Karen Coyle,她也是本届年会的主旨演讲人之一。9月4日的讲师为Sam Oh,他是韩国首尔成均馆大学(Sungkyunkwan University)图书馆与信息科学系的教授。他作为DCMI会员韩国国家图书馆的代表担任DCMI的理事会理事,同时他也是DCMI监事会(Oversight Committee)的成员。

由于我们多年未在海外举行的DCMI年会中参加其培训项目,因此本届年会笔者特意注册了这两天的培训课程,以体验了解精彩的培训内容。Karen Coyle的培训课程从元数据的基础出发,深入浅出地讲述DC元数据和语义万维网的基本概念及相关知识。她对于一些理念的清晰阐述以及在培训中多样化的讨论与问答

Dan Brickley做主旨报告



甚至是实践环节,使笔者这样在国内尚算有一些元数据理论与实践经验的学员受益非浅。例如她在教学中提到,元数据是一种“具有实际应用目的的数据”,我们在设计与实践中必须要真正从实际应用的需求出发。如果你在元数据应用纲要中每增加一个属性(元素),你可能就需要增加一个美元的成本。而当你的这条元数据属性被使用时,你可能就获得了一美元的回报。

Sam Oh教授的培训课程更多地是面向系统设计与开发人员。他介绍了XML Schema、命名域、主题图(Topic Maps)、主题图的本体建模、RDF、本体语言OWL等几乎全套的语义万维网技术。一天的培训内容相当丰富,技术含量充足。作为大学老师的Sam Oh教授具有丰富的教学经验,条理清晰,也同样寓教于实践,在课程中安排了许多问答与练习的环节。由于Sam Oh教授的培训内容更偏技术,与KTW的参会人员的背景更贴近,因此他的培训课程也吸引了更多的学员参与。

培训课程这两天属于会前会的阶段。和往年DCMI的年会一样,与培训课程同时举行的还有各种开门与闭门的各类社区与工作会议。9月3日上午是闭门的监事会会议,下午是开门的亚洲推进工作组的会议。9月4日上午是开门的foaf(友之友, friend of a friend)词表工作组,下午是闭门的咨询委员会会议。这里介绍的两个工作组会议,其工作组都是新成立的。

2.2 论文与会议交流

如前文所述,本届年会的规模较小,因此论文的数量也不多。本次年会共收录了10篇论文、3个项目计划报告以及1个展示海报。本次会议的交流发言共分5场,其主题分别是开放链接数据(Linked Open Data)、元数据信任(Trust)、科学数据(Scientific Data)、元数据转换(Transformation)、文化遗产(Cultural Heritage)。由于论文以及项目计划报告的提交者都在会议上作了交流发言,因此每个主题专场安排2到3个论文或报告。这五场交流在9月6日、9月7日的两天会议进程中都已经安排完了。展示海报环节则是在9月6日下午4:30和整个KTW的所有展示海报一起出展。

和过去的会议相比,本届会议的论文、项目计划报告等比较平淡,亮点不多。大多数论文与报告都是实践的案例与总结,较少如架构、理念、模型等方面思考的理论性文章。本次会议能够称得上理论性与探索意义的论文可能就属秦健老师的“元数据的功能与架构需

求:支持科学数据的发现与管理”^[12]。这可能也从一个侧面说明DCMI的发展已经不是或者不可能在架构与模型等理论、标准层面作很大的拓展,而是要重在实践与应用推广。本届论文与关联数据相关的内容却是有一定的数量,也说明关联数据的应用与元数据界是息息相关的。本届会议论文的另一个特点是全球性。总计论文、计划报告及展示14篇文献中,其作者竟然有来自斯洛文尼亚、德国、葡萄牙、英国、爱尔兰、意大利、巴西、美国、加拿大、日本、泰国、马来西亚以及澳大利亚等13个国家之多。因此可以说本届DC年会在文化多样性上显得非常有特色。

2.3 两个专题会议

本届年会另外还举行了两个小型的专题专题会议,一个是在计划外的,一个是在计划中的。

第一个小型专题专题会议并没有被列在会议日程表中,是新成立的亚洲推进工作组的主持人Shigeo Sugimoto教授临时召集的。Shigeo Sugimoto教授来自日本筑波大学图书馆、信息与媒体研究学院,他也是DCMI监事会的成员。DCMI成立亚洲推进工作组的目的是想以亚洲为试验田,进一步增强地区性的合作,更紧密地倾听各地区用户真正的需求——无论是本地化还是应用方面的需求。同时,多搞一些区域性的活动,有助于降低DCMI推广的成本,提高效率。因此,Shigeo Sugimoto教授在9月5日会议正式议程结束后,临时召集来自亚洲的代表举行了这个专题会议。在将近一小时的会议中,来自日本、中国、马来西亚、泰国、印度、新加坡、韩国等亚洲各个国家的代表分别介绍了自己所在国元数据研究与实践的一些经验与实例。

第二个小型专题专题会议排在会议日程的最后一天日程表中,其会议主题是“关联数据的词表与Drupal”。Drupal是一个著名的开源内容管理平台,许多网站采用Drupal平台来开发管理。Drupal平台的另一个优势是它管理与发布关联数据都有相应的模块支撑。DCMI的网站已经有多年的历史,技术相对比较陈旧,也只有数位成员可以更新内容。DCMI正在资助一个开发团队,想通过基于Drupal平台的开发建设一个新的DCMI网站,目标是使DCMI的网站能成为一个大家交流的社区,可以在上面共享资源、讨论和学习。此外,在KTW 2012的大会中举办这

样一个Drupal平台开发的特别会议也有其独特的优势,因为一些作为开发人员的参会者因为参加其他的会议来到了这里,这也使得这样一个技术性很强的专业会议在DC-2012能够保证一定的参与度。

2.4 精彩的主题演讲

在KTW 2012中,所有会议的主旨演讲都是共享的,这使得会议所邀请的主旨演讲的成本通过共享的方式得以降低,也提供了跨领域的知识交流机会。KTW 2012的主旨演讲一共有六场,其中有两场是由DCMI邀请的^②。

2.4.1 Dan Brickley: 我们还有什么可做的 (What is left to do) ?^[13]

DCMI邀请的第一位主旨演讲者是Dan Brickley,他的演讲题目是“我们还有什么可做的?”。

Dan Brickley是DCMI社区早期的参与者之一。他也是W3C社区中很活跃的一份子,也是使用广泛、影响深远的foaf词表的制定者之一。Dan Brickley目前在谷歌公司工作,参与了Schema.org的推广与相关工作。

Dan Brickley首先从回顾DCMI的历史出发,总结了DCMI社区的定位与特点。在1995年,DC标准制定者的设想是由每个网页的制作者自己生产元数据,而不是像图书馆那样,由专业人士来创建元数据。虽然如此,DC所能描述的对象还只是万维网的一部分,其中包括学术性的文章、报纸、联机的电子书、学习材料、以及其他一些特别的集合,如媒体的存档、视频的注释、课程目录、书目记录集合、预印材料、科学数据集、博物馆、档案馆以及政府文件等等。我们并没有一个很好的名称去称呼这样一个子集,或者用专业、资源集合、公共领域或其他什么名称。我们没有一个简单的方法在万维网中划定DCMI所关注的信息领域。DCMI就像是定位在传统图书馆、档案馆以及其他传统公共机构的元数据需求与万维网无数各种各样的信息管理需求中间的一个桥梁。DCMI过去的关注点并不是很明确,但它确实应该有自己的更明确的关注点,而不是着眼整个万维网。我们可以说是其中特殊的一部分集合、信息仓储或者是档案类的内容,那些内容在这个信息社会中会起到一些特

别的或重要的作用。也可以说是由一批专业人员维护的对社会有用的各种各样的信息集合。

Dan Brickley提到几乎在DCMI一开始,就有人提出DCMI的使命已经终结的预言。Dan Brickley参加DCMI的年会正好是1997年芬兰赫尔辛基举行的DC-5会议。当时就有人开玩笑说,在赫尔辛基开的这次会议,可以被称为“芬兰终结”(Finnish Finish)会议,因为DC元数据集已经基本定型,其术语定义也已经达成共识,这方面的工作已经结束。实际上,后来事实证明,芬兰会议远远不是一个结束,而应该算是一个重要的开始。因为1997年DC-5开会时,正好是RDF草案发布的时间。当时DCMI社区的一些关键人物对这一语义万维网的基石性的技术标准制定起到了关键性的作用。

无论从理论模型还是架构标准,在现今的万维网以及关联数据发展的大环境下,DCMI还有什么可做的?拿Schema.org这样的计划来举例,它并没有直接采用DC元数据标准。那么DCMI社区该如何来认识Schema.org这样的决定呢?如果我们仅仅将DC看成是一个产品,只是一个元数据术语的静态集合的话,那Schema.org不使用DC元数据集应该很令人失望。如果我们将DC看成是DCMI社区的代称,是一个由个人、团体、项目组成的带来许多合作机会的社区,那么Schema.org就是一个巨大的机会。作为Schema.org的实际建设参与者,Dan Brickley也详细介绍了Schema.org计划的方方面面与进展情况。

最后,对于回答“我们还有什么可做的”这个问题,Dan Brickley给出了自己的回答。他认为DCMI社区已经建立了与数字图书馆界、公共的数据集、各种文化遗产、博物馆、电子政务以及受控词表等独特的关系,我们应该在这样的基础上找寻我们的方向。在词表的映射、关联与长期保存,加强DCMI的教育职能,通过更好地应用最新的语义万维网相关的技术以及与传统书目元数据界的合作,都可以是我们去努力的方向。

2.4.2 Karen Coyle: 非同凡“想”(Thinking Different)^[14]

DCMI邀请的第二位主旨演讲者是Karen Coyle,她的演讲题目是“非同凡‘想’”。

Karen Coyle曾经是加州大学图书馆的馆员,她参

^② KTW 2012共有三个大会:PRICAI-2012、PRIMA-2012以及DC-2012,6个主旨演讲是由每个大会各分摊两个。

与了著名的加州数字图书馆的建设。目前她是一个涉猎许多数字图书馆领域的独立咨询专家。她也服务于许多标准委员会,其中包括MARC标准组织(MARBI)、OpenURL以及作为美国图书馆协会(ALA)在电子书标准制定工作中的代表等等。

首先Karen Coyle解释了她的演讲题目,这是来自Apple的广告语。这句话在语法上是有问题的,因为正确的说法是:“think differently”。当然乔布斯肯定知道语法的问题,他之所以这么说,是因为他要将“Different”作为一个名词。简言之,如果是“think differently”,不管你如何去想,你只是想法的不同。而“Think Different”要表达的,是去想完全不同的东西。

她以一些实例来阐述如何“非同凡‘想’”。以The Internet Archive的Open Library这个社会化的书目信息网站为例,这个网站从亚马逊取数据,也从图书馆拿数据,也让用户来编辑数据。他们在书目信息管理中碰到一个问题就是如何做字顺排列。对于传统的图书馆编目来说,字顺排列是一个非常重要的原则。然而Open Library最后用了一个完全不同于图书馆界做法的思路,那就是不采用字顺排列方法。

在卡片目录时代,无论是题名还是作者标目,字顺几乎是唯一的发现手段。然而在网络时代,我们是否还需要如此来做呢?通过全文以及结构化的搜索手段,我们可以从各个角度发现我们所需要的文献,以各种不同于传统管理的手段来聚类展示我们的文献,我们为什么需要采用复杂的规则来做好字顺排序呢?Google也好,Amazon也好,都不以字顺排列显示结果。甚至连OCLC的WorldCat也不是缺省按字顺排列结果的。接着Karen Coyle又以一些形象的实例解释了名称、字顺其实与知识组织无关,而按语义聚类才是更自然的知识组织方法。

Karen Coyle提到的另一个有趣的实例是关于书的内容定位。在纸书的时代,书有固定的页码结构。你可以清晰地告诉朋友你看到了某本书的第几页,或者说你引用的某句话来自某本书的第几页等等。在电子书时代,如何定位这些信息呢?在流式的电子文本中,页码是没有任何意义的。电子书在不同屏幕大小、不同字号间距的情况下,没有办法以固定的页码来确定内容的定位。你难道可以告诉朋友说:“你那本书看到多少了?我看了17%。”其实这个问题可能在印刷术发明之前就解决了,过去的手抄本《圣经》标识了每一个章与节。

另一个例子是哲学家路德维希·维特根斯坦(Ludwig Wittgenstein)喜欢在他的著作上给每一个段落编上号。在电子书时代也完全可以这样处理,并且可以设定阅读时隐藏,而在需要的时候调出来显示当前内容的位置。正是因为我们还是以纸书的理念与架构去处理电子书,而没有将电子书看成是不一样的东西,才产生了电子书页码缺失这样一个伪难题。试想完全结构化编码的电子书内容给每个内容块加上标识是件难事吗?

对于图书馆而言,也有如何真正做到“非同凡‘想’”的问题。例如传统的图书阅读,在作者、书、读者之间很少有交互。我们知道各种书可以通过书目信息、推荐书单或甚至是他们所在的相同架位等共同点找到他们之间的关系。我们还可以通过书的作者与读者,通过查询他们目前的状态找到人与人之间交流的机会。这样书被阅读时就处于鲜活的状态,并且通过书,我们可以关联到和书一样重要的读者。以这样的观念延伸去看什么是数字阅读,其实不仅仅是载体的不同,内容模式上的不同,更多的是交互方式的不同,阅读从单向变成了双向甚至是多向的。阅读本身不再是线性的逻辑,而是动态变化甚至是网状的。

但是很可惜,图书馆还是将书当成需要管理与控制的物理对象,而不是可以互相关联的知识。确实,在19世纪电报发明之前,是书主导了信息世界。那时我们习惯于信息存在于一个实实在在的、预先打包的实体中,以一种线性的方式来管理。现在这样的观念过时了。电报、电话、广播、电视甚至现在的因特网,完全取代了书在信息世界的主导地位。几年前我们能够想象电话与因特网结合而成为一个多媒体信息与交流的工具吗?

那图书馆应该如何去做呢?我们必须将图书馆从“物”的组织管理转变到知识发现上来。这意味着图书馆只能成为这个信息世界中的一个部分,因为大量的信息是存在于图书馆之外的。这也意味着要更强调事物之间的关联,而不是它们线性排列的位置。同时也意味着要将图书馆的用户视为整个知识海洋的贡献者而非仅仅是消费者。它意味着很多很多,但最重要的是要放弃对信息的控制,这也将是图书馆员们最难做到的事。

最后Karen Coyle以从2维到5维(2-D to 5-D)的形象比喻指明了图书馆可以努力的方向:

2维就是目前的图书馆,我们以线性的方式排列书籍与目录。

3维就是加上链接,给图书馆的“物”、“物”之间以及与外界广阔世界的信息加上关联。

4维就是时间线,我们要能在图书馆看到随着时间的变化所产生的不同思想及创意的变化,包括在不同的时间点上哪些文献与内容影响了哪些作者与发明者,又催生了哪些新的发现。

5维就是人,图书馆中的一切都是由人来创造,并且人们会以你意想不到的方法来使用图书馆的服务与资源。通过使用图书馆,人们会理解或创造新的知识。这些知识可能混合了来自他个人的思想以及以前接触的信息,而人与人之间的交互为这些新知识的创造提供了更好的条件。

3 年度进展与发展方向

和去年一样,笔者除了查询相关文献与DCMI网站,更主要是采访了CEO Stuart Sutton以及CIO Tom Baker两位先生,由两位先生亲口阐述对于DCMI本年度的工作进展评述以及今后发展方向与工作重点的思考。

3.1 年度进展与现状

目前DCMI的CEO Stuart Sutton先生上届年会前不久刚刚上任这个职务,因此这一年可以说是他的履新之年。他在本届年会的开幕致辞上以理念、组织与社区为题清晰地介绍了DCMI的使命以及历史与各个发展阶段。

DCMI一开始试图寻求一个对于数字信息的描述机制,寻找一个解决方案,并希望这个解决方案简单有效,能够应用在所有领域,能够描述各种类型的资源。从这个原点出发,1995年DCMI推出了DC元数据,希望可以帮助用户在万维网中更容易找到信息。在1998年,DC元数据成为一个成熟的DC元数据元素集(DCMES, The Dublin Core Metadata Element Set)。DCMI起步的视角仅仅是万维网HTML的世界,然而很快就走向了被广泛应用的更结构化的XML。自2000年以来,DC也和以RDF为基础的语义万维网及关联数据有了更多的关联,DC也从一个面向万维网的元数据集转变为面向语义万维网和关联数据的核心词表。

从组织机构来说,自1995年的发端一直到2008年,DCMI一直作为OCLC的一个项目而存在。到了2008年12月,DCMI在新加坡注册成为一个非盈利的公共的有担保的有限公司。然而DCMI一直是作为一个社区而存

在的,它是一个对元数据专业知识进行众包的国际化社区。一开始,这个众包所向的“大众”其实是非常小众且多样化的慢慢地转变成一个开放的由共识驱动的社区。在DCMI社区中,有着许多面向主题的社区与工作组,他们在各个领域以及跨领域的范围中寻找元数据设计的基本原则,以及关于这些原则的描述与最佳实践。

鉴于前几届的年会介绍论文(尤其是2007年与2010年)已经对DCMI的组织架构作了十分详细的介绍,2011年的年会介绍也对目前的机构变化作了阐述,因此这里不再赘述DCMI机构组织方面的详细内容。我们在这里罗列一下DCMI目前存在的专题社区(Communities)和工作组(Task Groups)供大家参考。社区与工作组的区别是社区负责某些DCMI关注的问题讨论,工作组则专注于问题研究和具体解决方案的设计。

这一年,DCMI延续了过去一两年来所制定的工作方向,虽然没有在标准制定及理论研究方面有什么大的进展,然而持续关注维护一些重要词表域名的长期可获得性,以及开放的、由共识驱动的面向主题的元数据创新等等。因此去年定的两个工作主题——词表的长期维护包括域名的保护,以及与其他项目进行元数据匹配的工作都得到很好的推进。

除此以外,DCMI这一年还有两项重要的进展。一是在伦敦开了一个2007年伦敦会议的5周年纪念会。2007年的4月底5月初,在大英图书馆举行了一次名为“数据模型讨论”的会议,这是由来自DCMI、语义万维网以及RDA社区的专家们一起参加的一个小型会议^[16]。这个会议促进了DCMI、语义万维网社区以及RDA工作组的深度合作,为DCMI更好地为书目元数据界服务打下了基础,也促使RDA更多地向语义万维网的技术与规范靠拢。在那个会议上,RDA的维护管理机构同意将以RDF来描述RDA,并且寻找可以和DC进行互操作的方案。

DCMI这一年另一项重要的工作是将DC术语集以关联数据的形式发布。DCMI在其术语网站上以RDFa的形式发布所有术语的符合关联数据发布规范的RDF描述。这样,对普通用户来说,你看到的是术语内容介绍的HTML网页,而对关联数据的消费者(语义浏览器、语义搜索引擎、各种解析工具等)而言,能获得的就是内嵌的可以直接关联的结构化数据。以关联数据的方式发布DCMI的元数据方案,将扩大DC在语义网领域的

应用范围和应用效率。这也意味着DCMI在关联数据的应用领域坚实地跨出了一大步。

3.2 发展方向与工作重点

今年年会主旨演讲人Dan Brickley的设问实际上也反映了DCMI社区相当一批参与者的困惑：在今天的万维网，语义万维网以及关联数据的大环境下，这个社区还有什么可做的？在上世纪90年代，早期的DCMI社区做了很多工作与努力，但是当时的DC元数据并不为外界所熟知。现在DC元数据已经变得非常有名，应用广泛，可是外界包括DCMI社区自身的许多参与者却不知道DCMI在做什么或应该干什么。

在回答这个问题时，Stuart Sutton先生首先回顾了DCMI的历史发展。他说DCMI有三个发展阶段：

第一个阶段是自1995年起，这时DCMI社区的目标重点是放在一个核心元素集的建设上，是面对万维网环境寻找一个简洁的元数据描述方案，最后生产的也是一条条元数据记录。

第二个阶段是自2000年起，DCMI提出了应用纲要的解决方案，应用纲要的思想是混搭来自不同命名域的不同词表中的元素，制定针对某一个特定需求的信息描述元数据的解决方案。这时DC从一个简单并稍嫌封闭的核心元素集转变为一个词表。

第三个阶段则就是目前的状况，DC将作为一个在异构的词表生态环境中作为沟通桥接功能的词表，主要用

表1 DCMI专题社区^[15]

译名	频率
可访问性社区	Accessibility Community
教育社区	Education Community
政务社区	Government Community
核心社区	Kernel Community
知识管理社区	Knowledge Management Community
图书馆社区	Libraries Community
本地化与国际化社区	Localization and Internationalization Community
长期保存社区	Preservation Community
注册社区	Registry Community
学术交流社区	Scholarly Communications Community
科学与元数据社区	Science and Metadata Community
社会化标签社区	Social Tagging Community
工具社区	Tools Community
词表管理社区	Vocabulary Management Community

表2 DCMI工作组^③

译名	频率
抽象模型评估工作组	Abstract Model Review Task Group
书目元数据工作组	Bibliographic Metadata Task Group ^④
文化遗产元数据工作组	Cultural Heritage Metadata Task Group
教育元数据应用纲要工作组	Education Application Profile Task Group
核心任务工作组	Kernel Task Group
图书馆应用纲要工作组	Libraries Application Profile Task Group
元数据溯源工作组	Metadata Provenance Task Group
DCMI与网络知识组织系统工作组	DCMI/NKOS Task Group
注册工作组	Registry Task Group
Schema.org 元数据匹配工作组	Schema.org Alignment Task Group
翻译工作组	Translation Task Group
用户文档工作组	User Documentation Task Group ^⑤
词表管理架构联合工作组	Vocabulary Management Infrastructure Joint Task Group

于在多种词表术语中进行语义的映射，以便支持关联数据和不断生长的语义万维网。

^③ 此表内容取自DCMI网站，有些新成立的工作组尚未体现在DCMI的网站上，例如我们前面介绍过的亚洲推进工作组（Asia Forward Task Group）以及foaf词表工作组。

^④ 由以前的DCMI/RDA工作组更名而来。

^⑤ 由以前的用户指南工作组（User Guide Task Group）与词汇工作组（Glossary Task Group）合并而成。

对DCMI发展阶段的把握以及定位的考虑,决定了DCMI今后的工作重点与发展方向。Stuart Sutton介绍说,DCMI今后的工作方向有这么三个:一是继续做好元数据设计模式的研究,使DCMI的理论基础即DC抽象模型要不断与时俱进,不断进化;二是要做好词表的映射(Vocabulary Mapping)工作,以发挥DC元数据作为一个应用最广泛的核心元数据集的桥梁作用。这个工作的突破口就是做好DC术语与schema.org之间的映射,或者称为元数据匹配(Metadata Alignment)的工作;三是继续做好词表的长期保存,除了foaf词表的长期保存与其域名的维护机制这个示范性项目外,DCMI也在寻求其他可以合作的对象。

今年DCMI的另一个创举是将这些拟定的重点工作方向以“工作主题”(Work Themes)集的形式发布,分别为“平台独立的应用纲要”、“多种词表映射”和“词表的可持续性”对应上面三个工作方向。这些工作主题是DCMI为了适应元数据生态环境变化而提出的聚集点所在,也是跨越单个DCMI单个社区与工作组的具有广泛性议题的重要任务,这些任务的完成会得到DCMI组织层面的资源支撑。有些新成立的工作组也是为了帮助更好地完成这些任务,如foaf工作组。另外,在今天的会议上也提到了一个拟议中的主要工作方向,那就是与书目元数据界的合作,帮助图书馆界将书目数据可以与关联数据兼容,可以被翻译成关联数据的语言,这样书目数据也可以被图书馆以外的其他领域数据相关联。这一工作方向也缘于DCMI与RDA的紧密合作基础。

在具体实施层面,DCMI会关注两个职能的拓展。一是进一步推进区域合作,多举办地区性的活动,包括更多的地区性会议,甚至是以各种非英语语种为主的会议,推动区域的合作与推广。目前由于网络数字技术的成熟,即使地区性的会议也一样能够吸引更多的远程参与者。这不仅能够应对由于经济状况变化而带来的成本压力,同时也可以真正解决地区性、多样化所带来的各种需求。前面提到的DCMI新成立的“亚洲推进工作组”正是为了这个任务而诞生的。二是进一步加强DCMI的教育功能。DCMI在教育方面的目标是提升业界对于元数据的认识与关注度,同时帮助元数据设计者与实践者更好地完成他们的任务。教育的内容也可以不限于元数据,其他相关的新技术应用培训也可以纳入其中。并且教育的对象也不仅仅限于用户,而可以是对教师的培训,通过有层次的培训体系更高效的

推动DCMI教育功能的拓展。另外,地区性的合作组织将是DCMI今后教育培训很好的平台。如前面提到的新成立“亚洲推进工作组”就能为东部与西部的元数据应用需求和专业经验之间打开一扇双向交流的窗口。

4 结语

对于如何回答“我们还能做什么”这样的问题,DCMI以及其他元数据专业社区的参与者或多或少都有一些疑虑。但通过参与本届年会,通过学习与交流,笔者认为这个问题并不难回答,甚至根本还谈不上是个问题,因为在很多方面我们都处于发展路径的中间阶段,还远远谈不上走到了终点。DCMI社区对这个问题的回答是要顺应元数据生态环境的变化,找准自己的定位,发挥自己的特长,通过抓住重要的“工作主题”使自己得以可持续发展。

其实国内的元数据界近年来也面临着发展路径的疑虑,随着几个大型的元数据课题项目走向尾声的脚步声,大家同样也面临剩下来还能干什么的疑惑。然而笔者认为,国内的元数据生态环境还远远没有达到DCMI社区所认识的高度,我们在理论研究、实践应用以及符合国际规范等诸多方面尚有许多课要补。其次我们可以学习DCMI的发展策略与思路。DCMI与RDA以及Schema.org的合作展示了原来定位在传统元数据界与网络知识海洋中间的DCMI元数据社区向两翼进行拓展的战略思路。DCMI将成为两者之间真正的桥梁与粘合剂。而我们如何通过这个工具及其平台更好地跨出我们的“围城”,这也是值得我们认真去思考的话题。

最后,我们不仅仅是要更多地参与DCMI社区的各种交流与任务,加强与国际的合作,而且可以通过参与其推动区域性发展及强化教育职能的任务中发挥更大的作用。DCMI也鼓励地区性的合作伙伴举办一些教育性质的活动,共享本地化的学习资源。DCMI也期待有一些更多的地区性的活动能在中国举行,比如小型研讨会、培训班、针对一些确定主题的小型会议等。

推动区域合作与发展,DCMI并没有非常具体的目标与做法。这只是打开了一扇窗,然后通过我们共同的努力,去看看我们可以领略到怎样的风景。

注:本文的著述获得了DC2012会议的两位主旨演讲人Dan Brickly先生与Karen Coyle女士以及DCMI的CEO Stuart Sutton和CIO Thomas Baker两位先生的大力帮助,在此一并表示诚挚的感谢。

参考文献

- [1] Conferences and Workshops [EB/OL]. [2012-10-22].<http://dublincore.org/workshops/>.
- [2] 刘炜.DC元数据年度进展[J].数字图书馆论坛,2006(11):24-28.
- [3] 刘炜.DC元数据年度进展(2007)[J].数字图书馆论坛,2007(11):19-22.
- [4] 刘炜,夏海.DC元数据年度进展(2008)[J].数字图书馆论坛,2008(11):46-49,63.
- [5] 黄田青,刘炜.DC元数据年度进展(2009)[J].数字图书馆论坛,2009(12):70-74.
- [6] 沈芸芸,冯英,刘秀文.DC元数据年度进展(2010)[J].数字图书馆论坛,2010(12):40-45.
- [7] 赵亮,杨佳.DC元数据年会综述(2012)[J].数字图书馆论坛,2011(12):47-53.
- [8] DC-2012, Kuching, Sarawak, Malaysia [EB/OL]. [2012-10-22]. <http://dcevents.dublincore.org/index.php/IntConf/dc-2012>.
- [9] DC-2010 Conference Home [EB/OL]. [2012-10-22]. <http://dublincore.org/workshops/dc-2010/>.
- [10] Home - MIMOS [EB/OL]. [2012-10-22]. <http://www.mimos.my/>.
- [11] Knowledge Technology Week 2012 [EB/OL]. [2012-10-22]. <http://dcevents.dublincore.org/index.php/IntConf/index/pages/view/ktw2012>.
- [12] QIN J, BALL A, GREENBERG J. Functional and Architectural Requirements for Metadata: Supporting Discovery and Management of Scientific Data [EB/OL]. [2012-10-22]. <http://dcevents.dublincore.org/index.php/IntConf/dc-2012/paper/view/107/61>.
- [13] BRICKLEY D. What is left to do? [EB/OL]. [2012-10-28]. <http://dublincore.org/workshops/dc2012/DCMI2012danbri-keynote.pdf>.
- [14] COYLE K. Thinking Different [EB/OL]. [2012-10-28]. <http://dublincore.org/workshops/dc2012/DCMI2012coyle-keynote.pdf>.
- [15] DCMI Work Structure [EB/OL]. [2012-10-22]. <http://dublincore.org/groups/>.
- [16] LOWERY J. Cataloguing Standards Used in the British Library [EB/OL]. [2012-10-22]. <http://www.bl.uk/bibliographic/meeting.html>.

作者简介

赵亮 (1964-), 副研究馆员, 研究方向: 数字图书馆, 元数据, 数字出版。E-mail: lzhaol@libnet.sh.cn
张春景, 助理研究员。E-mail: cjzhang@libnet.sh.cn

A Review on Dublin Core 2012

Zhao Liang / Shanghai Library, Shanghai, 200031

Abstract: This is a review of DC-2012 annual conference of Dublin Core Metadata Initiative and the paper describes in detail DC-2012 and its features. And also the paper gives a brief overview of the history and evolution of all the annual conferences before 2012. By highlights such as what was being told by keynote speaker in this annual conference, and by introducing how the metadata ecology evolves, the paper tells us how DCMI to develop ideas to deal with the problems of the future development path. The writer of the paper interviewed the CEO and CIO of DCMI who told us what main achievements are made by DCMI in this year and what is DCMI's idea of future consideration. The paper outlines the main work theme of DCMI, such as Metadata Design Patterns, Vocabulary Mapping and Vocabulary Preservation. Finally, the paper proposes what should be considered of future development direction by professionals in Chinese metadata community and how we make contribution to the international metadata community.

Keywords: Dublin Core metadata, DCMI, Annual conference, Goal, Metadata design patterns, Vocabulary mapping, Vocabulary preservation

(收稿日期: 2012-10-31)