

DC元数据年度进展 (2010)

□ 沈芸芸 / 北京大学图书馆 北京 100871

冯英 / CALIS管理中心 北京 100871

刘秀文 / 北京大学图书馆 北京 100871

摘要: 文章介绍了都柏林元数据 (DC) 2010年年会基本情况, 以及DC管理机构DCMI近一年的研究成果, 其中主要是对关联数据理论与实践的研究, 以及在关联数据和互操作环境下对DCMI推荐文档进行的相关审议和修订。文章也总结了DCMI过去一年工作与组织进展, 探索了DCMI的发展方向。

关键词: 都柏林元数据, DCMI, 关联数据, 语义, 互操作

DOI: 10.3772/j.issn.1673-2286.2010.12.007

1 DC年会概况

2010年度的“都柏林核心 (DC) 与元数据应用国际研讨会”于10月20日至22日在美国宾夕法尼亚州的匹兹堡市举行。本次会议继续由都柏林核心元数据组织 (DCMI) 主办, 并与美国信息科学技术协会 (ASIS&T) 年会一同举办。共有来自美国、欧洲、亚太等23个国家和地区的约150名代表参加了此次会议。代表们分别来自大专院校、图书馆、政府机关、数据库商、国际组织、研究与专业协会、软件开发商等不同领域。出席人数最多的是东道主美国, 约占参会代表总数的2/3, 其次是欧洲代表, 亚洲代表主要来自中日韩、新加坡和台湾, 约1/3的参会代表第一次参加DC年会。

本次会议主题为“使元数据更好地工作: 庆祝DC15年”, 为探讨DC如何更好地发展提供了更为开放的平台。会议日程安排与形式继续沿用前几次研讨会模式, 多种形式并举, 设计周全而开放, 使DCMI各组织的会议日程与研讨交流议程高效有序。

1.1 培训

会议安排了两场基础培训和两场高级培训。基础培训^[1,2]总结了DC的历史、关键概念和进展情况, 高级培训^[3,4]阐述了从DC到语义网需要经过的6个步骤——使用DC核心集及映射、扩展核心集及跨库检索功能、使用可管理的取值列表进行分面、使用标准词表、使用RDF和互联网标识符以及整合本地信息到关联数据, 完成DC到语义网的进程, 最后为过渡到关

联数据奠定基础。

1.2 大会主旨发言

本次大会共有2场主旨发言, 分别是: Stuart Weibel做的关于DC发展评估的报告^[5], 他对DC的管理、标准化、合作、新加坡框架 (互操作层次)、DC数据模型以及最新的关联数据等都做了详细而中肯的评价, 其中对DC的国际化、标准化、兼容性以及指导性文件等评价较高; Mike Bergman做的关于实现数据互操作途径的报告^[6], 主要论述了关联数据的作用以及从数据、结构化数据到语义网的发展, 总结了实现互操作的条件与方式, 最后对社会的不同层面如W3C、DCMI等在其中所扮演的角色进行定位, 认为在互操作方面DCMI是最好的游戏者 (但可能不是仅有的)。

1.3 各类主题报告

大会安排了内容丰富的各类主题报告, 涉及DC的原则、应用、实践、研究进展等各方面的内容:

- 大会交流报告: 包括4个主题共11场报告, 4个主题分别是: 元数据—原则与创新、元数据—图书馆应用与FRBR模型、社区与元数据, 以及都柏林核心实践。

- 专门研讨会: 包括8场会议, 主题包括: 关联数据—领域建模、词表选择与开发、关联数据应用纲要—建模与需求、都柏林在英国、ISO教育元数据MLR等。由此可以看出, 本次会议最受关注的主题是关联数据。

- 项目报告：会议安排了所有提交的6个项目的报告，交流了各自项目对DC的应用实践和心得体会，其中包括笔者提交的《为中国数字图书馆建立元数据应用规范：国家图书馆元数据总则项目》报告。

- 海报展示与介绍：9项。

1.4 工作会议和社区会议

在一些主题报告的同时，大会还安排了全部DCMI内部组织的会议，共有15个工作组会议和社区会议同时举行。

本次会议是DCMI独立运作开展学术交流和推广的第三个年头，整体运作方式完全参照企业的业务流程运行模式，从而使DCMI各组织分工明确，通过小范围的高质量的学术交流和讨论，明确了各自需要完成的具体任务与对应的时间表，会后，DCMI的相关工作组发布了最新的会议报告。

整体而言，本次研讨会在总结DC历史、展望未来的同时，肯定了DCMI对提升数据互操作所做的贡献和所扮演的重要角色，并指明了DC的发展方向：强化DCMI在社区中的指导和培训作用、互操作层次的提升、关联的理论和实践等等，这就预示着DC的应用、与其他元数据方案和编码体系的关系以及DC在语义网和数据关联中的作用研究都将日益深入，实践范围将日益广泛。会议最后对DCMI的内部组织进行了人员调整。

本次会议共出版学术论文10篇，项目报告6篇，海报8篇（见<http://dcpapers.dublincore.org/ojs/pubs>）。

2 DCMI年度关注及重要进展

应该说，DCMI在2010年度的关注重点无疑是在关联数据的理论与试验以及语义网方面，围绕着这个主题，在DCMI术语和词表管理、DC抽象模型、其他元数据方案、模型及编码体系（如LOM、FRBR、RDA、SKOS）等各个方面探讨如何更好地描述网络资源以更有效地管理和发现它们。为此目的，DCMI内部机构设置也做了相应调整。

2.1 关联数据

本次会议的热点话题之一是关联数据与语义网。从

大会主题报告、培训、会议论文及特别会议都有涉及关联数据，其中有4个关于关联数据的特别会议，分别讨论关联数据的领域建模（Domain Modeling）、词表选择及开发、关联数据应用框架的模型及需求以及下一步工作计划。可见DC社区对关联数据的关注程度。

如何利用关联数据实现从文档万维网到数据万维网的转换是本次会议探讨的热点。Jennifer B. Bowen在其论文中给出了Rochester大学图书馆在将图书馆目录数据转为关联数据方面所做的实践，通过开源软件XC（eXtensible Catalog）可将MARC数据转为关联数据。特别会议参考DBPedia和BBC项目的实践，从关联数据的应用目标及功能方面探讨需要建模的领域以及建模时需考虑的方面。在关联数据的词表选择方面，与会人员考虑了RDA、FOAF、VIVO、DC等，指出不同的应用及用户社区可采用不同的词表。

虽然大家都在关注关联数据，但其应用发展并不如想象中那么快，应用也不够广泛。Michael K. Bergman在大会主题发言中谈到虽然关联数据是一项有价值的技术，但目前实际以关联数据形式发布的数据集并不多，且缺乏互操作性，很难出版（publish）。他认为应通过对照映射提高关联数据的互操作性，通过构建参考概念（reference concepts）等方式更好地表现数据万维网中的“资源”。

2.2 DCMI词表管理工具作为开源项目发布

2010年5月，DCMI词表管理工具（DCMI's Vocabulary Management Tool）作为开源项目发布在基于web的托管服务Github上。该工具用来生成DCMI元数据术语的文档文件，即一个web文档和一个RDF概要。DCMI希望借此修改系统生成面向用户的web文档，该文档将会带有内嵌的、使用W3C标准RDFa的RDF表示，并邀请开源社区对此感兴趣的成员改编此工具以适应所用词表，并帮助完善Github上的编码。

2.3 DCMI术语修改与维护

DCMI应用委员会于2010年10月发布了DCMI元数据术语最新维护。最新版反映了自2008年1月以来的编辑方面的变化，包括：

- 删除术语dcterms:creator、dcterms:contributor和dcterms:publisher中易被误解的应用注释；

- 指定dcterms:title和dcterms:alternative的形式范围 rdfs:Literal;

- 勘误：术语dcterms:accrualMethod、dcterms:accrualPeriodicity和dcterms:accrualPolicy的所属域为dcmitype:Collection，而非dcterms:Collection;

- 增加新的数据类型dcterms:RFC5646;

- 在dcterms:DCMIType的描述中将“See:”参照增加到<http://purl.org/dc/dcmitype/>;

- 指定dcam:memberOf的形式范围 dcam:Vocabulary EncodingScheme;

- 声明dcterms:creator在属性上等同于foaf:maker (<http://xmlns.com/foaf/0.1/maker>)。

2.4 DCMI内部组织的变动

为了配合年度的工作，DCMI自2009年末开始调整其内部组织，包括工作组的成立和撤消、人员的变动等。

- 成立DCMI用户指南工作组，修订常用却显过时的文档“DC使用指南”（Using Dublin Core）。

- 成立DCMI抽象模型审议工作组，准备对DCMI抽象模型进行审议。

- 成立DCMI元数据起源工作组，其目标是定义关于描述语句或描述集声明的应用规范，创建描述元数据语句集合所需的数据元素的共享模型，以便于集中导入、获取、利用和出版与质量、权利、及时性、数据源类型、委托机构等有关的事实。

- 撤消DCMI代理工作组和政府应用纲要工作组，这是DCMI咨询委员会会议的决定。

- 成立DCMI/NKOS工作组，以期为KOS（知识组织系统）资源建立DC应用纲要。

3 DCMI现状报告

DCMI的使命、原则、管理模式基本没有变动，内部机构设置除2.4中根据工作任务和主题发展而做的变动外没有大的改动。下面重申DCMI的使命、运作原则、管理模式^[7]，并通过DCMI主要机构的现状报告^[8,9]对DCMI在过去一年的主要学术活动做管中窥豹。

3.1 DCMI使命

DCMI的使命是提供简单的标准规范以方便信息的

发现、共享和管理。这一使命的达成要通过：开发和维护描述资源的国际标准，支持全球范围内的用户和开发者社区，以及提升DC方案的广泛使用。DCMI作为一个机构的主要特性表现在：独立性、国际化和影响力。

3.2 DCMI运作原则

- 开放性：DCMI社区向所有对元数据感兴趣的团体和个人开放。

- 国际视野和参与：DCMI从非正式的工作组已经发展到全球社区的参与，表现在一系列DC文档的翻译、DC研讨会的召开地点以及DCMI成员和领导机构代表的地区多样性。

- 非营利：DCMI元数据标准可以根据实际需要公开使用，并鼓励公私机构采用其标准。

- 技术独立：DCMI希望维护其标准规范的语义的独立，并方便这些已达成共识的语义的编码表达。

- 超领域：建立伊始，DCMI的创立原则就是穿越互联网和内部网的信息仓储的资源发现。

3.3 管理

DCMI的管理有以下几个层次，其结构见图1。

- 执行者（领导者）：首席信息官、执行主席
- 监事会
- 咨询委员会和应用委员会
- 社区和工作组
- 成员
- 合作者（Partner）

3.4 DCMI主要机构年度研究总结

3.4.1 监事会

通过DCMI合作项目（Partnership Program）继续争取合作与经费支持，增加了一个新的合作层次——支持者（Supporter）。在继续激发人们对DCMI关注的同时，根据最新的研究进展讨论决定管理、财政和战略发展问题。

3.4.2 应用委员会

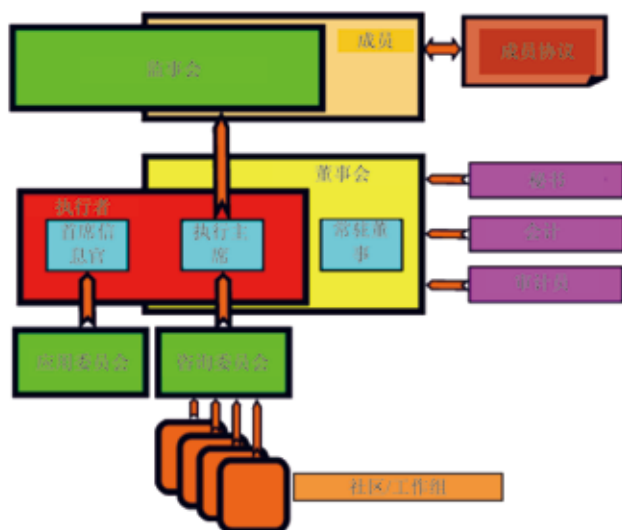


图1 DCMI组织机构图^[7]

于2010年10月发布了DCMI元数据术语最新维护，并修订了DCMI的FAQ和DC使用指南。委员会重审了近期的大量活动，从DCMI元数据术语的常规维护到应用纲要的技术审议，再到新的元数据术语的审批，以及大量的指南材料的审议，这些指南解决了面向用户的、关于元数据使用和形成最佳实践的基本问题。

应用委员会将保持对应用纲要的评估，期间将先对DCMI抽象模型和相关的规范文档进行审议，因为到目前为止它们已经形成了应用委员会审议标准的基础。

应用委员会还讨论了该委员会的未来发展观，这可能使一些工作得以结合在一起，如维护DCMI元数据术语、审议基本用户指南和再次以明确的标准审议应用纲要。

3.4.3 咨询委员会

在人事方面有一些调整，增选包括曾蕾在内的5位元数据专家进入咨询委员会，Sarah Rice和Sarah Shreeves辞去委员资格。至此，该委员会包括来自17个国家的50余名委员。

3.4.4 构架论坛

开展了“关联数据应用模型：建模和需求”的讨论，其关键点在于审议DCMI抽象模型和相关规范。考虑DCMI抽象模型的修订，此次修订将合并DC-RDF指南。按预期，修订后的抽象模型将替代一些DCMI的专有术语以利于参照RDF，这基于关联数据被更广泛理

解的基础上。它也将包含一些更易于非专家理解的图表以及关于DCMI抽象模型的基本原理和角色的解释信息。

论坛还将开会讨论关联数据应用规范的最新途径——通过记录描述元数据的使用模式来提高元数据应用设计的兼容性，最大化关联数据环境下元数据的一致性。

3.4.5 集合描述应用纲要工作小组

修改DC资源集合元数据应用纲要使其符合DCMI抽象模型。

3.4.6 教育社区和教育应用纲要工作小组

讨论教育应用纲要的模型，讨论点转向属性，Wiki上已经发布了带有应用案例注释和交叉参照的推荐属性列表。

3.4.7 术语工作组

修改术语表，删除了一些非DC元数据专有的、无关的款目。术语表中的概念对于理解DC社区元数据论述的发展——从基础的15个描述性元素到现在的关联数据环境下的互操作方法——是很关键的。

3.4.8 用户指南工作组

开展了广受欢迎的指南的修订——“DC使用指南”，使它的说明文字和实例与DCMI的属性和分类相适应。

3.4.9 核心(Kernel)社区和核心工作小组

该工作组研究方向较为单一，主要是完成Kernel/ERC(电子资源引文)1.2版草案文档，并为用户提供相应的开源应用系统。DCMI核心工作组开发了可以实现RDF和Kernel/ERC元数据互操作的软件，这个软件能够把包含在ANVL格式(类似于e-mail头文件)里的核心元素(Kernel elements)转换成RDF/Turtle，也可以把JSON转换成XML。这种独立的开源ANVL/ERC和核心元数据(Kernel metadata)程序包发布于2010年5月，已经放在DCMI工具网页上。它的兼容性非常好

(Windows、Mac、Linux上面都可以安装)，在125个不同的平台上通过测试。

XTF (eXtensible Text Framework) 索引和检索系统最新发布的2.2版本也可以实现对Kernel/ERC元数据的支持。加利福尼亚数字图书馆新推出的两项服务：EZID (Easy Identifiers) 和Merritt仓储服务也可以实现对Kernel/ERC元数据的支持。这些发展都提供了很好的从ERC/ANVL到RDF的映射，所有这些都必须遵守语义网协定 (Semantic Web conventions) 和DCMI抽象模型。

3.4.10 图书馆社区和图书馆应用纲要工作组

二者是DCMI比较活跃的组织，讨论和研究的问题非常广泛，对DC元数据在图书馆中的应用提出了更具操作性的方案。研究问题如下：

- 修订DCMI图书馆应用纲要：审议功能需求、概念模型、描述集合纲要和使用指南
- 跟踪DCMI/RDA的最新进展
- 跟踪W3C图书馆关联数据研发小组的最新进展及案例
- 关联数据的授权

3.4.11 本地化和国际化社区、翻译工作组

主要讨论和规范DC文档的翻译出版，使作者知道在哪里、怎样在本地或者国际上出版这些文档，以保证非英语的DC术语翻译的质量，并促进文档的发现。工作组计划为希望翻译DCMI文档的人建立使用指南，该指南的内容将包括：DCMI和DC发展的背景、DCMI文档的翻译状况、翻译的益处、推荐的优先翻译者、发展本地翻译并出版翻译文档、维护本地翻译文档以及案例。

3.4.12 元数据起源工作组

该工作组为新成立的工作组，根据其目标设定工作计划如下：跨越几个词表来回顾总结与元数据起源相关的词表元素和术语；生成和获取关于元数据起源信息的功能需求；在特定环境下（如在RDF或关联数据中）应用纲要的使用指南。

3.4.13 DCMI/NKOS工作组

该工作组的于2009年12月成立，同时在两个主要领域开展工作：一是开发简单的实体关系模型，使应用纲要可以基于这个模型建立；二是开发KOS类型表，它将包含各种类型的KOS，并基于特性（如结构和复杂性、概念间的关系以及历史作用）来定义。该工作组讨论了DC应用纲要的前三个部分，即功能需求、领域模型及描述集纲要和使用指南；同时还讨论了语义关系的注册。工作组还强调应关注其他协会、机构和工作组的相关活动，以避免重复工作，保证与其他KOS资源描述和相关注册工作的一致性和互操作性。该工作组提出了KOS资源应用纲要，并计划下一步的测试。

3.4.14 DCMI/RDA工作组

工作组讨论了RDA元素集和词表的注册，以及更广泛使用的RDA元素的生成。“RDA词表：过程、结果和使用”一文已发表，描述了在注册工作中遇到的挑战和所采取的解决方案。

该工作组与词表映射框架项目 (Vocabulary Mapping Framework, VMF) 有密切的交流。VMF发布了矩阵，为书目元数据元素间的计算机自动最佳映射提供工具。该矩阵包含RDF的三位关系，在TTL格式中表达了约2500个角色和11500个关联概念，还有映射自第三方词表（包括RDA和DC）的800余条术语。

该工作组继续与FRBR审查组密切联系，FRBR审查组维护FRBR和FRAD模型。二者在一起讨论FRBR实体关系模型的注册草案和FRAD模型的注册，以及FRBR和FRAD元素的映射，这些元素由RDA使用并被注册为RDA元素集的一部分。

3.4.15 注册社区和注册工作组

该工作组年度研究内容集中于如何开发对术语注册系统进行登记注册的注册系统。在调查现有的注册系统和注册用户关于注册的具体需求的基础上，1) 建立一套与现有注册系统共同的功能需求，2) 设计一套注册交互的使用案例，并考虑注册系统和其他应用系统（目标相同）间的互操作精度（报告、通知、版本等），3) 设计一套与注册系统转换语言相关的功能需求，包括互操作精度，它可以在注册系统间以及与其他应用（如基于同样目的的词表管理应用）一起共享信息。

3.4.16 学术交流社区

去年DCMI学术交流社区相对平静，邮件列表主要用作事件通知和偶尔被版主更新。最近有成员建议开始重新审议知识库文档，如DRIVER使用指南 2.0（或者即将出台的OpenAIRE使用指南），以期发现怎样使数字知识库达到更高的互操作层次，如DC元数据互操作层次中所定义的那些层次。

DCMI的社会标签、标准化社区、集合资源描述社区、政府社区、知识管理社区、长期保存社区也开展了相关的研究和推广工作。

3.5 网站和微博

根据统计，DCMI网站的每月访问量是35265，年度网站访问量比前年增加了4.5%，页面访问量增加了2.7%。年度访问者人数增加了1.4%，达到277941人。

DCMI网站访问最多的文档是：元数据基础概述

页面、DCMI元数据术语、DC核心元数据元素集（1.1版）、元数据规范详述页面以及DC使用指南。

2010年3月，DC Twitter账户数达到1466个，比去年的报告增加了52%。

4 结语

从过去一年的发展可以看出，DCMI社区的工作更规范，更具影响力，对于互操作的研究和应用已经使互操作的第1个层次取得巨大的成功，即我们已经知道数据在哪里以及目前需要什么样的服务，为将来走向更广泛的互操作打下基础；目前的工作正在转向更多的关联（如人、主题、地点），希望对词表的信任和持久性问题以及更好的关联实践找到解决方法。

下一步，DCMI计划发挥社区和工作组的能量，更好地帮助社区，提供支持互操作性的指南和培训，DCMI的角色应当是参考词表（reference vocabularies）的权威机构^[10]。

参考文献

- [1] GREENBERG J. Dublin Core: History, Key Concepts, and Evolving Context (part one) [EB/OL]. [2010-11-23]. http://dublincore.org/resources/training/dc-2010/Tutorial1_basic_Greenberg.pdf.
- [2] PHIPPS J. Dublin Core: DCAM, Syntax, and Semantics (part two) [EB/OL]. [2010-11-23]. http://dublincore.org/resources/training/dc-2010/Tutorial2_basic_Phipps.pdf.
- [3] DANIEL R, Jr. Six Step SAFARI from the Dublin Core to the Semantic Web [EB/OL]. [2010-11-23]. http://dublincore.org/resources/training/dc-2010/Tutorial4_SAFARIDaniel.pdf.
- [4] COYLE K. Semantic Web & Linked Data [EB/OL]. [2010-11-23]. http://dublincore.org/resources/training/dc-2010/Tutorial3_transitional_Coyle.pdf.
- [5] WEIBEL S. An irreverent look at the history of the Dublin Core [EB/OL]. [2010-11-23]. http://dublincore.org/workshops/dc2010/DC-2010_20101021_Weibel_keynote.pdf.
- [6] BERGMAN M K. Bridging the Gaps: Adaptive Approaches to Data Interoperability [EB/OL]. [2010-11-23]. http://dublincore.org/workshops/dc2010/DC-2010_20101022_Bergman_keynote.pdf.
- [7] DCMI By-laws [EB/OL]. [2010-11-28]. <http://dublincore.org/about/bylaws/>.
- [8] DEKKERS M. Dublin Core Metadata Initiative - Status report March 2010 [EB/OL]. [2010-11-28]. <http://dublincore.org/news/communications/statusreports/2010/03/>.
- [9] DEKKERS M. Dublin Core Metadata Initiative - Status report September 2010 [EB/OL]. [2010-11-25]. <http://dublincore.org/news/communications/statusreports/2010/09/>.
- [10] DEKKERS M. DC-2010: Conclusion and closing [EB/OL]. [2010-11-23]. http://dublincore.org/workshops/dc2010/DC-2010_20101022_Dekkers_Closing.pdf.

作者简介

沈芸芸 (1972-)，研究方向为元数据及数字加工规范。E-mail: yyshen@lib.pku.edu.cn
 冯英 (1970-)，研究方向为数字图书馆技术与服务。E-mail: fengy@calis.edu.cn
 刘秀文 (1983-)，研究方向为元数据及数字加工技术与规范。E-mail: liuxw@lib.pku.edu.cn

An Annual Report on Dublin Core 2010

Shen Yunyun / Peking University Library, Beijing, 100871
 Feng Ying / Administrative Center of CALIS, Beijing, 100871
 Liu Xiuwen / Peking University Library, Beijing, 100871

Abstract: This annual report is on DC2010 Conference. Besides, it introduces research activities done by DCMI which is responsible for DC management, mainly including the research and practice on Linked Data and the review and revision of DCMI Recommended resources on the environment of Linked Data and interoperability. Meanwhile, the report reviews all activities by DCMI communities and task groups, and discusses the next steps in DCMI, too.

Keywords: Dublin Core metadata, DCMI, Linked data, Semantic Web, Interoperability

(收稿日期: 2010-12-07)