# DC元数据年会综述（2013）

□ 柴苗岭 / 中国科学院国家科学图书馆成都分馆 成都 610041

摘要：文章介绍了都柏林元数据2013年年会的基本情况，即会议议程和会议内容，并在此基础上，就会议的部分主题，如e-Science、CAMP-4-Data主题研讨以及DCMI年会首次举行的最佳论文和最佳年度报告作重点介绍。在文章最后对DCMI当前的工作主题及其进展也做了一个小结。

关键词：都柏林元数据，DCMI，年会，开放关联数据，e-Science，科学数据
DOI：10.3772/j.issn.1673—2286.2013.12.003

2013年9月2日至6日，DCMI（Dublin Core Metadata Initiative）的第21届国际会议与iPRES（International Conference on Preservation of Digital Objects）在葡萄牙首都里斯本联合召开。

## 1 DC概况

DC元数据即Dublin Core Metadata（都柏林核心元数据），是一套用于网络资源描述和资源发现的描述性元数据。它产生于20世纪90年代中期。随互联网的蓬勃发展，互联网上产生了大量的虚拟文档，当时OCLC（OCLC，Online Computer Library Center，联机图书馆中心）和NCSA（National Center for Supercomputer Applications，美国超级计算机应用中心）看到这类文档的发展前景。他们认为有必要对这部分文档进行管理，于是，1995年两个机构联合在加拿大Dublin（都柏林）召开了第一届元数据研讨会。在会议上，工作组确定了两个目标，其中之一就是“在描述网络资源核心元数据元素集上达成共识”[1]。会后决定建立一个能对网上各种电子资源信息进行描述的、简单易操作的核心元数据元素集，即Dublin核心元数据。

DC元数据从最早的13个核心元素发展到15个核心元素以及DCTERM、词表，共计55个元素，成为一套由资源模型、描述集模型和词汇模型组成的DC抽象模型。这一发展过程经历了17年之久。

其后会议陆续召开，但都不固定，直到2001年，保持每年召开一次的频率，2008年开始，DCMI独立主持DC元数据学术研讨会。

从DC的发展历程上来看，可以分为三个阶段。

（1）1995年至2000年的基础理论完善期

1995年、1996年4月和1996年9月的三次元数据会议，确定了DC的核心元素，把从最早设计的13个增加到现在的15个元素的固定规模，并确定了其数据模型为Warwick框架的元数据结构等。

在1997至2000年的短短的三年里，DC元数据会议总共召开了五次，先后确定了附加的DC修饰词，即堪培拉修饰词，推出了著名的关联元素和1：1原则等。这一系列的举措，搭建了DC的元数据框架，使DC在理论上更加完善起来。

（2）2000年-2007年的实际应用期

2000之后，DC元数据研究会议基本上保持了一年开一次会议的频率。2000年至2003年期间，共召开了三次会议。其逻辑模型结构也于2005年形成，并于2007年对该模型进行了升级，推出了DC1.1。这一阶段是DC的成熟期。在DC1.1方案推出后，不断被研究和使用，标志着DC元数据真正走向了应用。

（3）2008年以后的全面应用，新方法和思路的吸收

自2008年，DCMI不再依附于OCLC，开始独立主持DC元数据学术研讨会。随着语义网和关联数据的兴起，DC也把注意力转向了语义网、语义揭示和建立其他元数据的映射。DC在其随后的发展过程中吸收了本体、关联数据的理论、方法等，受语义网技术的影响，DC采用了属性、子元素、抽象模型、新加坡框架等，但它同时也摒弃了一些老的方法，如修饰词和Warwick框架都没有再使用，始终保持着旺盛的生命力。

DCMI2013年会与代表长期存取的重要国际会议iPRES联合举行，会议主题为LINKING TO THE FUTURE，主要探讨持续性、维护和保存元数据和描述性词汇等问题。描述的对象种类繁多，包括了文化遗产和科学数据、电子政务、金融和电子商务等。

## 2 2013年DCMI年会概况

本次会议时间共6天，聚集了来自38个国家和地区，370名参会人员，包括了DC和iPRES的参会人员。这次参加会议的有两名来自中国大陆，且都来自于中国科学院国家图书馆系统，分别参加了DC和iPRES的会议。

DC2013按照惯例分为了三个阶段：会前会、大会主会议和会后会。

会前会包括培训和博士研讨会。大会主会议包括了3天主会议。DC和iPRES两个会议有交叉环节，如会议的开幕、3个主题报告、POSTER环节、会议闭幕和总结，都是大会联合召开。两个会议各有其程，按主题分组并行，并鼓励交叉选听。大会期间，陆续召开了5个工作组会议。第三阶段是大会主会议之后的主题研讨会和DCMI咨询委员会和监督委员会的会后会。

笔者在本次会议上作为POSTER一文*The Research of Open Conference Resources Organization Based on RDA Description*作者出席会议。除参加大会主会议以外，选听了6日的CAMP-4-Data（Cyber-infrastructure & Metadata Protocols）的主题研讨。

本次会议内容丰富，引人入胜，下面就会议的全程作简介。

### 2.1 会前培训及博士研讨会

DC-2013在9月2日按照惯例对不熟悉本次会议主题的人员，邀请了4位专家进行180分钟的教学培训。培训课程从开放关联数据、本体和本体术语、元数据溯源（Metadata Provenance）和W3C溯源本体（The W3C Provenance Ontology）四个角度出发。

除培训之外，博士研讨会也是会前会的内容之一。这个研讨会上，博士研究生将有机会展示他们的研究，并得到专家们建设性的批评和指导。

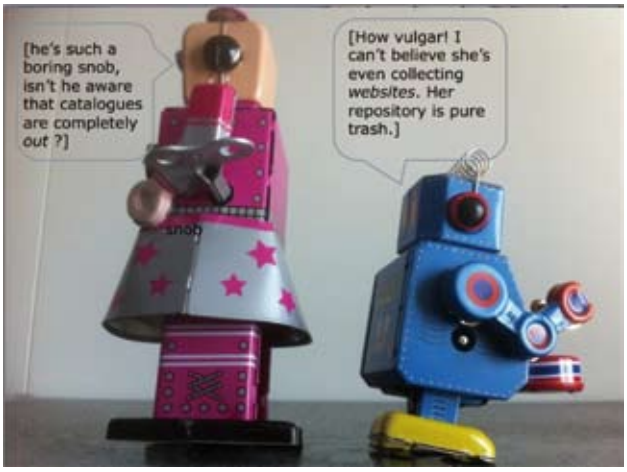
### 2.2 主题演讲

大会的主题演讲都有3个，DC和iPRES两个会议的参会人员共同参与这个环节，这种共享的方式方便了两个领域研究人员的交流。

本次大会的第一位演讲者是Gildas Illien。他演讲的题目是“Darling, we need to talk”。

Gildas Illien是Bibliothèque nationale de France（法国国家图书馆，简称BNF）的文献与数字化信息服务部主任。他在BNF从事了6年网页归档和呈缴本制度，现在负责推动开放关联数据环境下使用data.bnf.fr（2013斯坦福图书馆创新奖得主[2]）发现图书馆数据和目录。他作为European Interest Group on RDA（EURIG）的副主席促进了FRBR规则和RDA模型在欧洲的实施。

在演讲中，Gildas Illien用拟人的方式将开放存取和元数据比作两个互不了解的恋人，忙于和关注各自的工作，但工作都得不到对方的认可，但因为“data”有了共性。这个主旨演讲从管理、政策和宣传，组织、人员、角色和技能，以及从开放数据世界的所有权和身份认同等三个方面进行了阐述。生动的讲解时不时让两个领域的参会者会心一笑。




### 2.3 大会会议

大会开始，DCMI执行主席Stuart A. Sutton先生首先对DC的历史短暂地追忆了一下，然后对本次年会的流程作简介，随后会议分场地开始交流。

此次会议内容从关联数据（Linked Data）、eScience、eScience的应用和数字图书馆（eScience Applications and Digital Libraries）、词汇表设计

（Vocabulary Design）、元数据协同（Metadata Collaboration）、元数据重用（Metadata Reuse）、教育（Education）、图书馆（Libraries）等八个主题交流了会议论文、项目报告和海报。

2013年，会议共收录了12篇论文、11篇项目报告、7篇海报，相比2012年在亚洲举行的年会来说，收录量翻了一番。这也从侧面说明，亚洲在DC元数据、元数据的研究上还有很多上升空间。

从投稿来看，这一次年会中，内容不仅收录了当前的理论研究成果，还吸取了世界范围内的元数据实践者的见解，特别是收到许多e-Science领域的投稿。这些投稿主要是包括关于这一领域的研究数据、元数据以及它们的出版和供应。这也说明了在图书馆、教育、电子政务、运输、经济学统计等领域元数据的发展和管理都面临着挑战。

因为此次主题较多、内容丰富，不能在一篇文章中完全展示交流成果，笔者就以“e-Science”这个主题为例介绍一下会议内容，希望读者能管中窥豹，见微知著。

e-Science主题报告在3日下午交流，共有三篇与会论文和报告，主要是关于科学数据的研究。他们分别从科学数据的元数据设计、早期科学数据的元数据设计和科学数据与图书馆目录的关联的角度发言，受到高度的关注。

第一个发言的是来自雪城大学的秦健教授，她的交流论文的主题是*How Portable Are the Metadata Standards for Scientific Data? A Proposal for a Metadata Infrastructure*，论述了近年在科学元数据的研究方面的成果。秦健教授认为一个涵盖所有目前科学数据的元数据标准的方法是不能跟上日益增长的数据需求的，有其局限性。然后以科学数据领域的元数据标准为例，并认为需要元数据基础设施。该研究收集了4400种以上来自16个标准的独特元素，并将这些元素划分为9类[3]。秦健教授的专题讨论引起了与会代表的浓厚兴趣，不少听众据此选择了她参与的6日的CAMP-4-Data的主题研讨。

第二篇交流的文章是Dominique Ritze和Katarina Boland的*Integration of Research Data and Research Data Links into Library Catalogues*。作者就科学数据和传统数据如何提供一个整合的检索环境，检索出版物和科学数据进行了介绍。

而最后一篇题目为Collaborate, Automate, *Prepare, Prioritize: Creating Metadata for Legacy Research Data*，是Inna Kouper、Stacy R Konkiel等共同撰写的。如何通过改善提供必要的元数据打捞遗产数据是本文提出的主要问题，该报告从通过创建域的特定元数据的定量和定性指标，提出了四管齐下的遗产科学数据的元数据框架，阐释了解决之道。

值得一提的是这个主题下的三篇文章分别入围了DCMI首次推出的最佳论文和最佳项目报告，虽然遗憾都没有摘得桂冠，但这个主题下的研究魅力和受关注度可见一斑。这也说明科学数据的元数据研究是当前世界上的研究热点之一。但从国内研究论文关于“科学数据元数据”的产出情况来看，1989年至2012年研究论文仅51篇，其中2010年达到研究高峰，即8篇，之后2011、2012都每年仅一篇论文[4]，与国际关注度差距较大，可见在这一领域，元数据工作者们还大有可为。

### 2.4 特别会议和主题研讨会

#### 2.4.1 特别会议

本次会议的special session由W3C、都柏林核心元数据组织团体和任务组所赞助，该部分着重于一些当今元数据生态学（metadata ecology）中最迫切的问题进行研讨。范围从元数据词汇的保存和长期管理，到本体和应用配置的作用和关系。即长期保存和RDF词汇管理（Long-term Preservation and Governance of RDF Vocabularies）；LOD中的数据充实和转换（Data enrichment and transformation in the LOD context）；应用程序配置文件作为OWL本体的替代文件（Application Profiles as an alternative to OWL Ontologies）；为什么使用Schema.org（Why Schema.org?）

#### 2.4.2 主题研讨会

主题研讨会从9月6日召开，分别围绕了VocDay 2013：词汇管理（Managing Vocabularies）和CAMP-4-Data：信息基础架构和元数据协议（Cyber-infrastructure & Metadata Protocols）举行。这两个为期一天的主题研讨主要围绕元数据管理和支持科学数据的基础设施的设计和应用、元数据标准的开放和可

持续利用的政策这两个主题进行。

笔者选听CAMP-4-Data的主题研讨，这个组是由Jane Greenberg教授主持的，于大会闭幕后第二天，9月6日举行，历时一天。

会议开始时，Jane Greenberg教授请参会者作自我介绍，气氛从轻松的环境下开始，有的参会者在介绍完自己的来历后，将此次会议比作自己的第二个暑假。从介绍情况来看，以欧美国家的从事元数据工作的研究人员为主。

会议分为发言和讨论。发言从介绍（Introduction）、基础设施模型和框架（Infrastructure Models & Frameworks）、使用和用途跟踪（Usage & Tracking）、永久标识符（Persistent Identifiers）、应用（Applications）5个部分介绍，发言过后是分组讨论。

笔者挑选了几个比较感兴趣的发言作简介。

一是针对开放数据的可持续的数据发现和互操作性的解决方案研究。其中来自Data Observation Network for Earth's（DataONE's）[5]的Rebecca Koskela作了名为*The Metadata Zoo*的发言，这引起了广大参会者的兴趣。DataONE采集了地球上生命科学和环境科学的开放数据，支持该领域数据的持续发现和通过映射的方式实现互操作。

其次，在这个group中雪城大学的秦健教授发言，她就如何构建基于元数据的知识本体进行了介绍。介绍中，她将元数据分为语义、工具和规则块，集成后，按照元数据引用、元数据发现、元数据收割、元数据质量、元数据出处、元数据管理这几个部分生成输出，最后形成一个由DC、VIVO、FOAF、ORCID、Geographic metadata standards和Context metadata elements后台支撑形成的本体接口，并通过元数据构架生成器最后以XML、RDF或者其他方式输出[6]。

第三，来自罗格斯大学图书馆的Grace Agnew和Mary Beth Weber就*Open Source Metadata Application Profile and Research Object Handling for Research Data*为题，介绍了其开发的一个开源的工作流管理系统，该系统包括一个编目实用和一个复合的对象处理系统，它能够创建元数据并智能的对象处理、支持记录和分享研究数据。在会后，我对其项目作了一个跟踪，在其知识组织的方式上，罗格斯大学库采取独特而可扩展的方式支持研究数据生命周期管理罗格斯大学图书馆。这种方法包括利用现有的RUcore机构库，同时扩展到支持的研究数据，以及重新设计和开发团队提供所需要的专业知识，从而支持那些繁忙的研究者。

有趣的研究除以上介绍的之外，还有很多。本会议支持OA，会议录可以在这个地址全文下载：http://dcevents.dublincore.org/public/dc-docs/2013-Master.pdf

### 2.5 闭会研讨

大会会议及主题研讨会结束后，是监事会、应用委员会和咨询委员会会员时长两天的会后研讨，就着重研讨处理元数据管理和架构设计、应用和政策（支持开放、持续获取用来管理科学数据的元数据标准）等进行沟通和讨论。

### 2.6 最佳论文奖及最佳项目报告奖

DCMI2013会议，还首次颁发了最佳论文奖和最佳项目报告，这也是DC会议历史上首次颁发这个奖项，包括最佳论文和最佳项目报告。这两个奖项各自先确定3篇入围论文，然后根据选票确定获奖论文。

最后最佳论文分别由Antoine Isaac等和来自日本的Biligsaikhan Batjargal获得。

#### 2.6.1 最佳论文

最佳论文出自9月4日会议第五场的“Metadata Collaboration”中交流的第一篇论文：*Achieving Interoperability between the CARARE Schema for Monuments and Sites and the Europeana Data Model*。

在文中，作者Antoine Isaac等就不同数据模型中数据聚合上下文（data aggregation context）映射时存在的互操作性问题进行了讨论，即他们在CARARE schema设计的针对考古和建筑古迹和遗址映射到Europeana Data Model（EDM）时遇到的问题和解决方案。

这个映射的目的是把整个欧洲国家保护收藏平和数据库（national monument collections and databases across Europe）中超过两百万以上的元数据记录整合到欧盟数字图书馆（Europeana digital library）。

最后作者认为将数据映射到EDM是比映射到Europeana schema更难的一种方法，但它能让数据越接近原来的元数据，且数据可根据语境、语义与其他数据关联。

其他两篇入围的论文是秦健老师在9月3日交流的论文*How Portable Are the Metadata Standards for Scientific Data? A Proposal for a Metadata Infrastructure*，前面已经作了简介，这里不再赘述。另外一篇是Jihee Beak和Richard P. Smiraglia撰写的*With a Focused Intent: Evolution of DCMI as a Research Community*。

### 2.6.2 最佳项目报告

这个报告是日本的Biligsaikhan Batjargal等撰写的*Linked Data Driven Dynamic Web Services for Providing Multilingual Access to Diverse Japanese Humanities Databases*。

作者建立了一个信息检索原型系统，该系统基于LOD资源、个人姓名规范数据、主题词和其他关联数据资源的链接。此外，还尝试证明这种方法如何在实际检索系统中集成，以及如何连接和访问不同类型的数据库并提高利用可用的LOD资源。

研究还涉及使用不同语言访问数据库，如图1所示。

Web NDL的规范数据能直接链接到美国国会图书馆主题词表（LCSH），在日语和英语之间能进行跨语言的联结。当提问“浮世絵”时，用户能通过使用链接“Ukiyoe”，指向LCSH，而获取“Hashirae”、“Pillar Prints”和“Ukiyo-e”的记录。即使用户并没有“Hashirae”的相关知识，也能通过此法获取相关知识。

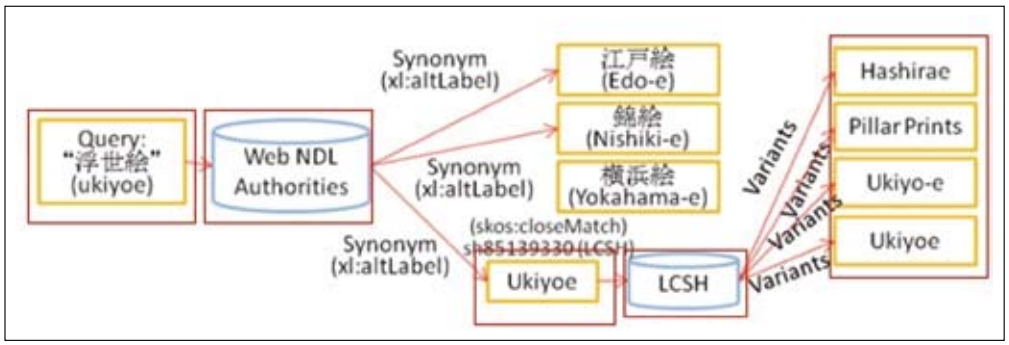



图1 使用主题词进行多语种访问实例图

其他入围论文是前面提到的9月3日e-Science主题下的第二和第三个交流论文。

## 3 DCMI当前工作主题

2007年、2010年、2011年及2012年已经就DCMI的组织构架、机构变化及专题社区和工作组作了十分详细的介绍，在此就不再赘述。下面就当前的工作主题分组罗列一下。

都柏林核心元数据组织团体有一系列“工作主题”，它们主要关注元数据生态学中所有的和正在变化的东西。这些DCMI支持的主题工作如同一个机构一样接受来自DCMI的关注并向其承诺。

### 3.1 平台独立应用程序配置文件 (Platform-independent Application Profiles)

DCAM（DCMI抽象模型），由DCMI2007年开始推广，它为打包语义Web的兼容数据可验证的记录格式提供抽象语法。DCAM旨在为无限的关联数据图形的现代范式和可验证的元数据记录更相似范式，建立一个桥梁，通过使用大量的软件平台和应用技术，进行本地化管理和限制。

5年来，DCAM广泛部署、采集经验，同时核心RDF标准也持续发展。“platform-independent application profiles”这一主题正在重新评估能传达元数据设计模式的通用语言的需求和要求。这种通用语言，既能作为关联数据可兼容数据格式的范本，又能为连贯元数据在社区内的话语和实践创建使用作参考点。

### 3.2 映射多样性词汇

虽然DCMI元数据术语和其他核心词汇通过提供分享的参考点，提高了元数据的连贯性，但多样、重复的词汇不可避免地威胁到元数据仓储的建立。

解决这个问题的关键一步是，创造机器可读的映射。“映射多样性词汇”这个工作主旨在于将DCMI元数据术语映射到其他词汇表的相关术语中。由于这样一个映射暂无先例可循，这个工作期望建立一个能被其他词汇表维护人员所广泛采用的工作流程和出版先例。着手点就是建立一个指向被Schema.org所定义的术语的映射。

### 3.3 可持续词汇

作为应用的基础，任何给定词汇表的价值都基于两个方面，一是确定性，即词汇表（机器可读模式和人工阅读文本）随时间推移有高可靠性。二是它的统一资源标识符（URIs）不会被出售、重新定义或忘记。

为了达成这项工作的共识，DCMI制定并达成了FOAF项目协议。FOAF由个人拥有，应急预案紧急情况下，进行长期或短期转移持续性控制的计划。为了形成最佳实践，并最终确保我们的词汇表能被社会长期保存机构保存，这项工作测试了关于词汇的可持续性和管理方面的问题。

## 4 结语

DCMI这些年来，一直在不停发展。在国内的研究也从理论到图书馆领域的应用再到其他领域的发展，DC在应用上的优势和劣势也愈发明显。

DCMI采用开放的态度与其他元数据交流，并吸取其有用的东西，这也是十余年来DC得以发展、壮大的原因。DC对自己定位始终是在数字资源的描述，其未来发展也是值得期待的。

本文在撰写过程中得到DCMI的CEO Stauart A. Sutton先生的大力帮助，在此表示诚挚的感谢。

**参考文献**

[1] 吴建中.DC元数据[M].上海:上海科学技术文献出版社,2004.
[2] 2013 Stanford Prize for Innovation in Research Libraries.
[3] QIN JIAN, LI KAI. How Portable Are the Metadata Standards for Scientific Data? A Proposal for a Metadata Infrastructure [EB/OL]. [2013-10-28]. http://dcevents.dublincore.org/IntConf/index/pages/view/2013-peerAbstracts#Qin.
[4] 周波.我国科学数据元数据研究综述[J].图书馆学研究,2013(2).
[5] Dataone主页[EB/OL]. [2013-11-08]. http://www.dataone.org/about.
[6] QIN JIAN, LIU XIAOZHONG, CHEN MIAO. Ontology-Enabled Metadata Schema Generator: The Design Approach [EB/OL]. [2013-11-08]. http://dcevents.dublincore.org/public/special/dc2013/08-Qin-Liu-Chen-camp4data.pdf.

**作者简介**

柴苗岭（1978-），女，硕士，中国科学院国家科学图书馆成都分馆馆员。研究方向：元数据、开放资源、开放科学数据。E-mail: chaiml@clas.ac.cn

A Review on Dublin Core 2013

Chai Miaoling / Chengdu Branch of National Science Library of Chinese Academy of Sciences, Chengdu, 610041

Abstract: At the beginning of the paper, the basic information of 2013 DCMI Annual conference is introduced. Contents are composed of conference agenda and conference content. Based on this, parts of main topics like e-Science, CAMP-4-Data discussion are introduced in detail. Also, best paper and annual best report which are awarded for the first time are mentioned. At the end of the paper, the current research topic and development of DCMI are concluded.

Keywords: Dublin Core Metadata, DCMI, Annual conference, Linked open data, e-Science, Research data

（收稿日期：2013-11-26）