

都柏林核心简介

Diane I. Hillmann

国家科学数字图书馆

2004年10月11日，DC2004培训

什么是元数据？

- 我们可以从三个比较熟悉的例子得出结论：
 - 数据库管理系统（“关系数据库模式”）
 - 传统的图书编目(MARC & AACR2)
 - 万维网（自二十世纪九十年代中期以来）
 - 为DC提供了发展环境

元数据的类型

- 管理性
- 描述性
- 访问/使用
- 保存性
- 技术/结构性
- 其它?

图书馆与网络的比较

- 传统图书馆:
 - 从卡片目录到MARC经历了100多年的发展
 - MARC 记录包括:
 - 描述（所有对象）
 - 主题和分类（关于主题）
 - 收藏信息（位置）
 - 管理记录所必须的管理性信息
 - 基于共同遵守标准的分布式元数据创建

万维网：一个分布式图书馆？

- 二十世纪九十年代中期的状况：
 - 成千上万的信息提供者，使用各种各样的元数据方案
 - 搜索引擎提供大量但精确度非常低的信息
 - 资源的不稳定性，如地址的变更，甚至是消失等
 - 资源的数量和类型呈指数速度增长

15个核心元素 (1996)

创建者	题名	主题
其他责任者	日期	描述
出版者	类型	格式
覆盖范围	权限	关联
来源	语种	标识符

元数据的作用

资源发现

管理文档

控制 IP 权限

标识版本

保证权限

说明状态

标记内容结构

标识空间位置

描述过程

DC的特点

- 所有的元素都是可选的
- 所有的元素都是可重复的
- 可以以任何顺序来显示元素
- 可扩展的
- 国际性的

向数据模型转变

- 机器可处理性的实现需要一致性的数据模型
- 1996年: 在第二届DC工作组提出了“沃维克框架”: 将DC定义为元数据框架中一个专门的模块 (“资源发现”)
- 1997年: 提出“修饰词”进行语义限定
 - 一些早期的使用者将修饰词扩展到很复杂的地步
“DC.Creator.telephone-number”
- 1998年: DCMI涉足新兴的”资源描述框架”(RDF) 的研究并且为DC明确了一种清晰简单的数据模型
- 2000年: 正式批准发布了第一套修饰词集

DC数据模型的目前进展

- 提供了对资源明确的定义
- 将DC的原则和实践与DCMI以外的发展结合起来
- 明确清晰DC“信息包”与其他元数据包的关系
- 为DCMI将来的发展奠定基础

DC原则

- 向上兼容
- 一对一
- 最为恰当的值

向上兼容

- 十五个核心元素可以单独使用，也可以和修饰词一起使用
- 修饰词使元素更为精确：
 - 元素修饰词缩小了元素的含义而不是延展
 - 编码体系给出元素取值的环境
- 如果你的系统遇到不熟悉的修饰词，可以向上兼容—或者干脆忽略它

一对一原则

- 用一条记录来描述一个资源的表现形式
 - 例如：蒙娜丽莎的一幅数字化图像如果和原画是一模一样的，则不描述数字化图像
- 将资源和资源的责任者分开描述
 - 例如：创建者的电子邮件地址和机构都是创建者的属性而不是资源的属性

最为恰当的值

“对特定的元素或者限定词的最佳实践都可以随着环境的变化而变更，不过一般来说，使用者不能假定处理译解元数据的一定是机器，这就可能会导致对元素的构造强加一些约束，但是要谨记检索的可用性需求”

-- 来自 “DC使用”

简单或修饰版的DC?

- 简单DC仅限于最初的15个元素
- 修饰版DC则额外包括：
 - 新元素
 - 元素限定
 - 编码体系值

元素限定

- 使元素的含义更狭义，更明确：
 - *创建日期*和*修改日期*
 - *被替代*和*替代关联*
- 依据所选的句法，修饰词可以单独使用而不是必须和元素一起使用
 - `<dct:created>2002-10-04</dct:created>`, 替代:
 - `<dc:date><dct:created>2002-10-04 </dct:created></dc:date>`
 - 需要一个模式将“*创建日期*”向上兼容到“*日期*”
 - DC足够简单可以用于支持这两种用法

有关值的编码体系

- 表示所取的值一般取自：
 - 来自受控词表的术语（例如美国国会图书馆标题表）
 - 以特定标准格式编码的字符串（例如“05/02”表示5月2日，而不是2月5日）
- 即使软件不能识别的编码体系代码，值也应该是合适的，可以用于资源发现。

DC的成长与变化

- DCMI委员会比较重视参与的开放性
 - 会议，工作组，讨论列表
- 随着使用者不断的杜撰新的术语和使用方式的出现，DCMI术语集也在不断的发展
- DCMI应用委员会审核所提议的新的元数据术语

DC应用委员会

- 应用委员会接受新的元素，限定，编码体系和类型术语的提议
 - 依照语法规则，有用性，定义是否清晰，是否与现存的术语含义重叠等原则来评估提议
- 批准状况的层级模式：须遵循，推荐使用，已不使用，已登记

DCMI命名域和相关政策

- 在以下三个命名域中，所有的DCMI元数据术语都给出了唯一的标识：
 - <http://purl.org/dc/elements/1.1/> - 传统的15个DC元素
 - <http://purl.org/dc/terms/> - 所有其他的元素/修饰词
 - <http://purl.org/dc/dcmitype/> - 类型词表
 - 例子: <http://purl.org/dc/elements/1.1/title>
- 从政策上推动命名域名（URIs）的长期稳定性
 - 不涉及到语义上实质性变更的情况（例如修订）将不会导致改变命名域的统一资源标识符（URIs）

应用纲要与互操作

- 使用者会想知道他们的同行如何来设计元数据—避免“重新发明轮子”的重复劳动
- 信息提供者为了在自己的领域内提高访问能力，需要协调元数据的用法，例如：
 - 不同的国家之间（北欧元数据项目）
 - 预印本知识库（OAI协议）
 - 主题门户(Renardus)
 - 数学和物理(MathNet, PhysNet)

元数据记录的编码

- 20世纪90年代中期：可以在网页中嵌入HTML标记
 - 简单，易于配置但是不够灵活、很难维护
 - 一些类似*DC.Creator.eyecolor*这样不好的标记使人误以为能够提供其实并不存在的嵌套与实体区分的支持
- 2000年以来：更好的XML/RDF选择
 - RDF元数据支持复杂的结构，但并不破坏简单DC的语法规则
 - OAI协议促成了大量简单DC、非修饰词版DC的XML schema出现，并且提供了专门的协议支持它们

元数据收割的世界

- 服务提供者从多个内容提供者处收割和整合元数据
 - 预设标准元素集合，记录格式和收割方法
- 1999年以来：OAI开始作为学术联盟的预印本提供者
 - 今天：使用OAI-PMH创建的集合创造了元数据服务的新模式

为什么收割模式可行

- 挖掘出的元数据可以促进：
 - 现有元数据的复用
 - 附加增值服务的创建
- 低廉的初始成本是实施数字图书馆基础设施的关键所在
 - OAI新的静态资源库版本更是大大降低了初始成本标杆
- 个别领域已经开始定制通用架构
 - 简单DC的分支扩展

例子：国家科学数字图书馆 (USA)

- 原则：“创建新的元数据非常的昂贵，尽可能使用现存的元数据”
- 早期策略：
 - 支持标准格式，建构于修饰词版DC基础之上的元数据库
 - 收集DC和本地元数据（尽可能）
 - 将所用的元数据装配到中心资源库中
 - 将所有这些记录对元数据收割者开放
 - 将有限的人力用在资源集合元数据方面
 - 尽可能自动生成元数据

NSDL Search Results for

(Displaying results: 1 - 20 of 60) [Next Results](#) >>>

Title/Description	Resource Format	Found in Collection
<p>Plant nematodes the grower should know [electronic resource] / G. Steiner. [Reprinted from the Fourth proceedings of the Soil Science Society of Florida]. more info [Archived Version]</p>		
<p>The evolutionary position of nematodes Abstract Background The complete genomes of three animals have been sequenced by global research eff ... more info [Archived Version]</p>		
<p>Biodiversity of nematode assemblages from the region of the Clarion-Clipperton Fracture Zone, an area of commercial mining interest Abstract Background The possibility for commercial mining of deep-sea manganese nodules is currently ... more info [Archived Version]</p>		
<p>Nematode Movies This web page presents four videos of nematodes that show their twisting movement. These videos are ... more info [Archived Version]</p>	[text]	
<p>Analysis and functional classification of transcripts from the nematode <i>Meloidogyne incognita</i> Abstract Background Plant parasitic nematodes are major pathogens of most crops. Molecular character ... more info [Archived Version]</p>		
<p>The Effect of Nematode Isolate and Soil Environment on the Tobacco cyst Nematode (<i>Globodera tabacum solanacearum</i>), a Pathogen of Flue-cured Tobacco and Other Solanaceous Crops Tobacco cyst nematode (TCN), <i>Globodera</i></p>	[text] [pdf]	

More Information

Title	Nematode Movies
Creator	Michael W Davidson
Subject	Life Science
Subject	Animals
Subject	Microorganisms
Subject	Microscopy
Subject	Worms
Description	<p>This web page presents four videos of nematodes that show their twisting movement. These videos are from a collection of digital movies of microscopic organisms taken at a typical North Florida pond. They were taken under darkfield illumination at magnifications of 100X and 200X. Users may access these videos, ranging from 21.1 seconds to 45.4 seconds, by using a playback format at the following connection speeds: 28.8k (modem), 56.6k (modem), or T1/Cable/DSL; or downloading the video clip in .AVI format. (Author/LSR) Copyright 2003 Eisenhower National Clearinghouse (ENC).</p>
Publisher	Florida State University
Contributor	Michael W Davidson
Date	2003
Format	text/html
Identifier	http://micro.magnet.fsu.edu/moviegallery/pondscum/nematode/index.html
Language	en
Rights	Copyright 1995-2002, by Michael W. Davidson and The Florida State University. All Rights Reserved.

NSDL的稳步发展

- 建立评估元数据质量的方法论
 - “元数据必须能很好的协同”
- 发展对大量不同的数据进行标准化的技术
- 从处理元数据记录发展到处理“资源的描述性文本”

走向语义万维网

- 简单链接数据模型
 - 网络链接指的是“与...有关”
 - 自1997年以来，以资源描述框架(RDF)的方式逐渐形式化
- XML: 通用数据交换格式
- 统一资源标识符：万事万物都有唯一的地址
 - 资源和元数据术语都可以用来描述统一资源标识符
 - XML命名域：元数据词表术语的唯一标识符
 - 统一资源标识符可以作为数据整合的锚点

更多DC信息

- DCMI 网址
 - <http://dublincore.org>
- “DC使用”
 - <http://dublincore.org//documents/usageguide/>
- 参与工作组
 - <http://dublincore.org/groups>
- 提问！
 - <http://askdcmi.askvrd.org/>

提问

谢谢大家的参与！

Diane I. Hillmann

NSDL图书馆服务与运作部主任

DCMI应用委员会，顾问委员会成员

“DC使用”杂志编辑

AskDCMI网络管理员

DIH1@cornell.edu (第四个字符是“一”)