

主编：缪其浩

元数据 與 图书馆

理论学术年刊

上海科学技术文献出版社

元数据

與

主编：缪其浩

图书馆

理论学术年刊

上海科学技术文献出版社

图书在版编目(CIP)数据

元数据与图书馆/缪其浩主编. —上海: 上海科学技术文献出版社, 2005. 5

(图书馆杂志理论学术系列)

ISBN 7-5439-2551-6

I. 元... II. 缪... III. 元数据—应用—图书馆工作 IV. G250.7

中国版本图书馆CIP数据核字(2005)第029473号

责任编辑: 陈宁宁

封面设计: 通 文

元数据与图书馆

缪其浩 主编

*

上海科学技术文献出版社出版发行

(上海市武康路2号 邮政编码200031)

全国新华书店经销

江苏常熟人民印刷厂印刷

*

开本890×1240 1/32 印张11.625 字数322 000

2005年5月第1版 2005年5月第1次印刷

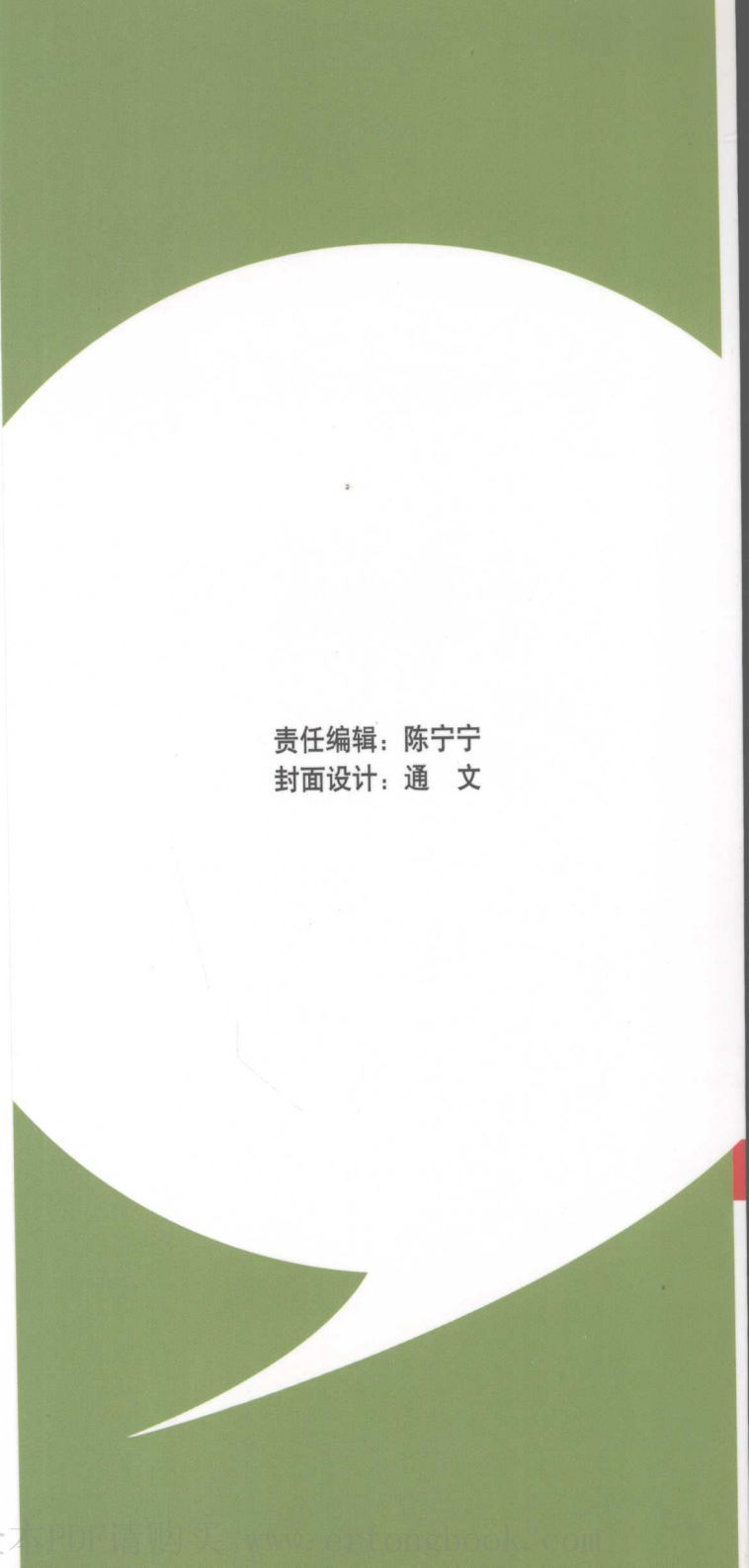
ISBN 7-5439-2551-6 / G · 659

定价: 38.00元

<http://www.sstlp.com>



主 编：缪其浩
副 主 编：王宗义
编 辑：方子丽 曹慧雯 郭敏
特邀编辑：刘 炜 赵 亮 梁颖
英文编辑：许桂菊

A large white speech bubble is centered on the right side of the page, set against a solid green background. The bubble has a rounded top and a sharp, pointed tail at the bottom. The text is printed in black within the bubble.

责任编辑：陈宁宁
封面设计：通 文

《元数据在图书馆的应用》专辑

主持人的话：刘 炜 赵 亮

图书馆界对新名词的“审美疲劳”一般两三年一个周期，能够沉淀下来，大多可算作精华。“元数据”就是这样一个概念。

自90年代中后期，现代元数据观念开始进入中国，从生疏、误读到解构、消化乃至研究应用，已经过去了七八年的历程。从最早试图在“中国试验数字图书馆项目”中运用，到国内一系列数字图书馆项目中的研究与发展；从最早的MARC与DC之争，到对各种元数据方案的耳熟能详，国内同行的认识水平也是在不断提高。特别是《中国数字图书馆标准与规范研究》项目的启动，把国内图情行业的元数据研究与实践者几乎一网打尽，使国内元数据标准规范的制定和研究提高到国际水平。

现在似乎渐入佳境，还没到应该总结的时候。《图书馆杂志》理论学术年刊（2004）推出这一组8篇元数据文章作为专题，是想为这一“佳境”添砖加瓦。一组文章不能像专著那样对一个主题进行系统深入的阐述，但这组文章有一个鲜明的特色，就是强调应用性。我们希望通过理论与应用相结合的文章，为国内元数据的应用实践作一些贡献。

张晓林基于他在DC2004国际元数据年会上的主旨发言提供了一篇论文：《以元数据的发展促进数字图书馆的前进》，从宏观上论述了元数据标准体系的组成要素及其相互关系，以及对于数字图书馆的基础作用。刘炜等的《DC元数据的历史、现状与未来》试图通过“史论”，对DC元数据在应用中的困境和人们理解上常有的误区进行一个了断。赵亮的《元数据的结构与句法》则涉及元数据应用中不得不逾越的一道鸿沟：元数据如何表达。接着三篇论文是对科技部项目研究成果的一些总结，涉及最具应用性的三个主题：专门资源、网络资源和资源集合。其余两篇则完全是元数据的具体应用，分别论述了在“打造服务主导型数字图书馆”和“建设学科

专业门户”时对于元数据方案的完整考虑。

年刊的这组专题文章大多是专为年刊而写的,少数论文曾发表于本刊的“元数据应用研究”专栏,但也特为年刊做了大幅度的修改和充实。我们非常感谢各位作者在百忙中能够欣然赐稿。遗憾的是由于篇幅所限,我们无法收录已经在刊物上发表过的托马斯·贝克先生的《DC词表的维护:实践、策略与模型》和杉本重雄先生的《专业元数据词表研究》两篇专稿,另外两篇曾经刊发的《上海图书馆名人手稿元数据方案》和《元数据与知识本体》,虽经作者大幅修改和扩充,也未能收入年刊,在此向广大读者表示歉意。诚挚地希望感兴趣的读者能够一如既往地关注本刊,并欢迎以各种方式提出您的意见和建议。

目 录

“元数据在图书馆的应用”专辑

以元数据的发展促进数字图书馆的前进·····	张晓林(1)
DC元数据的历史、现状及未来·····	刘 炜等(13)
元数据规范应用框架与编码·····	赵 亮(41)
专门数字对象描述元数据规范设计指南·····	肖 琰等(69)
网络资源描述元数据规范及其设计·····	楼向英等(89)
信息资源集合描述的现状、模型、方法和实践·····	刘 炜等(111)
服务主导型数字图书馆中元数据的制定和应用·····	郑巧英等(145)
“经济学学科知识门户”元数据方案的设计与实现···	马文峰等(167)

国外数字参考咨询服务的发展——美、加、英三国数字

参考咨询服务的主要经验·····	柯 平 王宸君(189)
美国数字参考服务标准研制进展评述·····	韩冬梅 詹德优(205)
文献编目的新结构：FRBR·····	计漫红(219)
新世纪情报检索语言研究综述·····	沙淑欣(233)
我国图书呈缴本制度的变迁及存在的问题·····	廖腾芳(245)
20世纪中国图书馆学研究“引进”问题探析···	李 卓 张晓红(255)

编撰《中国家谱总目》弘扬中华历史文化·····	王鹤鸣(265)
温州家谱文化的历史与现状·····	丁 红(283)
李芳馥：圆上一个世纪之梦·····	陈燮君(301)
唐代官方抄书考略·····	曹 之(313)
亚洲文会图书馆探讨·····	王 毅(321)
古籍版本目录学的研究方法和路径·····	王世伟(341)

后记·····	《图书馆杂志》编辑部
---------	------------

Contents

Zhang Xiaolin: Developing Metadata to Promote the Construction of the Digital Library	(1)
Liu Wei and Lou Xiangying: The History , the Current Situation and the Future of Dublin Core Metadata Initiative	(13)
Zhao Liang: Structure and Encoding of Metadate Schema Application	(41)
Xiao Long and ect.: Design Guide of Descriptive Metadata Standards	(69)
Lou Xiangying and Zhao Liang: Metadata Scheme for Description of Internet Resource and its Design	(89)
Liu Wei and etc.: The Current Situation , the Model , the Way and the Practice of Metadata Schemas for Collection Description ..	(111)
Zheng Qiaoying and etc.: Metadata and its Application in Service-oriented Digital Libraries	(145)
Ma Wenfeng and etc.: Design and Implementation of Metadata for the Project of "Digital Library of Economics"	(167)
Ke Ping and Wang Chenjun: Digital Reference Services in Libraries Abroad	(189)
Han Dongmei and Zhan Deyou: A Review of the Establishment and the Status of Digital Reference Service Standards in U.S.A.	(205)
Ji Manhong: The New Model for the Library Cataloging: FRBR	(219)
Sha Shuxin: Review of the Research on Information Retrieval Language in New Century	(233)
LiaoTengfang: The Changes and its Problems of the System of Legal Deposit Copy in China	(245)

Li Zhuo and Zhang Xiaohong: Analysis on the Introduction of the Theory of Foreign Library Science to China in 20 th Century ...	(255)
Wang Heming: An Introduction to " The General Catalogue of Chinese Genealogy"	(265)
Ding Hong: History and Present Situation of the Genealogy Culture In Wenzhou	(283)
Cheng Xiejun: Mr.Li Fangfu and the New Shanghai Library	(301)
Cao Zhi: A Research on the Handcopied Books in Tang Dynasty	(313)
Wang Yi: On the Library of the North China Branch of The Royal Asiatic Society	(321)
Wang Shiwei: The Research Ways of the Editions Bibliography for Ancient Chinese Books	(341)

张晓林

以元数据的发展促进数字图书馆的前进

文 摘 简单地分析了开放的数字图书馆环境之后,作者详细描述了数字资源的生命周期及我国数字图书馆标准规范(CDLS)项目的分阶段战略目标,论述了CDLS开放元数据机制如何适应多种元数据和动态变化的各个应用领域,探索如何在一个开放的标准框架下建立CDLS扩展元数据机制以适应开放服务环境的挑战。

Abstract:Beginning with a brief analysis of the open digital library environment, the author describes the Digital Resource Life Cycle model and the phased strategies of Chinese Digital Library Standards project, explains the CDLS open metadata approach to accommodate multiple metadata and dynamic fields, explores the CDLS extended metadata mechanism within an open standards framework to meet the challenge of an open service environment.

张晓林 研究员,中国科学院文献情报中心主任,博士生导师。主要研究方向:数字图书馆技术与系统,元数据,知识组织体系,信息管理与信息服务基础理论。有《元数据研究与应用》、《走向知识服务》等代表作。

近些年来,我国图书馆领域在数字图书馆建设方面获得了惊人的发展,从20世纪90年代由国家图书馆联合六家大型公共图书馆参加的“中国试验型数字式图书馆”工程^[1]开始,发展到目前已经在全国范围内实际服务的中国高等教育文献保障系统(CALIS)^[2]、国家科技图书文献中心(NSTL)^[3]、中科院国家科学数字图书馆(CSDL)^[4]等,普及到许多图书馆或信息机构进行自己的数字图书馆项目。此外,商业化的内容生产商也提供了几乎覆盖所有学科领域的数字资源。图书馆越来越普遍地将数字资源作为馆藏的极为重要部分,用户也越来越习惯于直接在线获取他们所需要的信息资源。在那些重点研究机构和高等学府,数字化已经是发展的首选。

面对迅速发展的数字资源与服务,我国图书馆界很快认识到数字图书馆是在分布、异构、动态变化的环境中发展的,而且任何一个真正有效的数字图书馆必须以一种能确保其资源可广泛获取、系统可互操作及服务可持续发展的方式建设及管理。因此,我们需要开放的数字图书馆构架范式,它应该体现以下设计原则^[5]:

(1) **模块化(Modularity)**: 任何的数字图书馆,不论是自行建设的还是第三方提供的,都只是用户数字环境中的一个模块。它仍可作为一个自治的独立系统支持特定的服务,但更应能够被当作整体的一部分整合到别的模块中,从而在多方位的数字服务中支持不同模块的多种组合甚至动态组合。

(2) **开放集成(Open integration)**: 任何模块在集成其他资源时(不论这一资源是内部的还是外部的),都将这些被集成的资源看成逻辑上的第三方模块,独立于具体位置和系统平台。然后,通过基于开放语言的标准描述和捆绑机制来组织这些资源从而使得模块的组织与集成具有相当好的扩展能力。

(3) **开放服务(Open service)**: 每个数字图书馆的服务机制本身也应采用开放语言,设计成公共的第三方服务模块,并能够根据不同的情况进行定制。这样,服务本身也可以被发现、重用、个性化及嵌入到不同的系统中,甚至动态地构建出新的服务或系统。

(4) 可伸缩可扩展(Scalability and Extensibility): 数字资源和服务应基于公共的核心机制和标准的扩展方法, 这样新的资源、服务和实现的技术机制都可以很容易地增加或关联, 从而实现从简单的、小规模的系统顺利地过渡到复杂的、大规模的系统。

在这样一个开放数字图书馆框架的远景目标下, 为确保数字资源及服务的可广泛获取、可互操作及可持续发展, 在开放数字图书馆标准框架的基础上, 国内各大主要图书馆和数字图书馆联盟开始进行我国数字图书馆标准与规范项目(Chinese Digital Library Standards, 简称 CDLS^[6])。

1 数字资源的生命周期模型和 CDLS 数字图书馆标准所采取的策略

CDLS 尝试着发展能够支持如下目标的开放数字图书馆标准框架:

(1) 发展可被互联网上用户(包括作为用户的服务和系统)有效访问和应用的数字资源;

(2) 发展可被互联网上用户(同上)有效访问和应用的网络服务;

(3) 发展互联网上各类用户代理可健壮整合多层次、多方面资源与服务的集成机制;

(4) 发展能有效使用的互操作机制从而使数字图书馆提供的服务可以与其他在线服务(e-services), 如 e-learning、e-science 和 e-government 合作;

(5) 发展常规的、可持续的组织机制以跟踪、组织及协调数字图书馆标准规范的建设。

这样的远景目标至关重要。传统图书馆更多的表现为一个被动的相对封闭的、往往局限于机械的文献检索与获取的信息机构, 但数字图书馆不应该是传统图书馆的数字拷贝。数字图书馆标准应该推动我们去建设一个以用户为中心、开放集成、动态交互的信息环境, 提升用户的信息生产率和知识生产率。立足于这样的目标,

CDLS项目超越了单纯的描述性元数据或检索协议的范围,着眼于以数字资源生命周期模型作为概念框架来构建一个更容易理解、更为有效的数字图书馆标准规范环境。

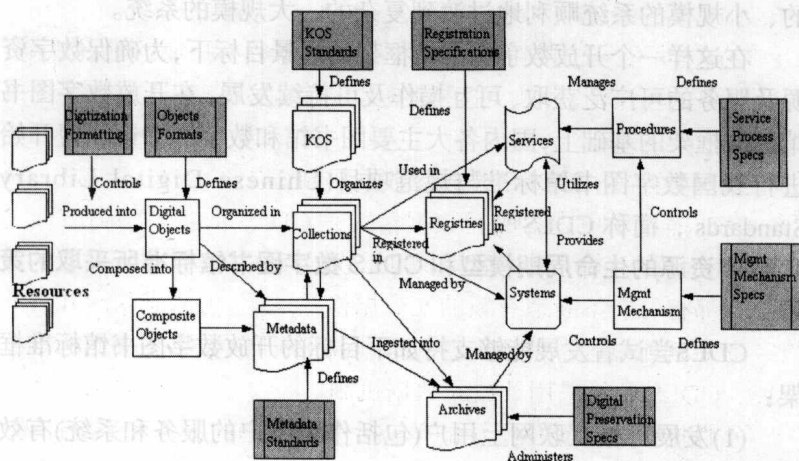


图 1: 数字资源的生命周期模型

如模型中所示,传统的或原生数字形态的信息资源,经数字化或格式转换处理,形成数字对象或复合的数字对象;这些数字对象被组织进一定的资源集合(馆藏),这些资源集合又按照相应的知识组织体系(knowledge organization systems, KOS)在结构与语义等方面进一步组织;这些资源集合由应用系统按照一定的管理机制进行管理,按照一定的服务程序提供相应的服务。此外,数字资源将来还可能在数字档案系统中进行长期保存。所有的这些集合、服务和系统都可以在开放登记系统中注册以便于发现与调用。

这一模型提供了数字图书馆标准规范长远发展的一个基本框架,但我们需要逐步推进的实施策略与目标。为此,CDLS项目制定了如下三个阶段目标^[7],逐步构建既能满足我国目前数字图书馆的建设与服务需要,又能在将来有效支持数字资源与服务的开放集成、合作共享及可持续发展:

(1) 阶段一: 建立数字资源建设和基本检索服务等方面的核心标准规范, 建设一个制定标准规范的强有力的开放平台;

(2) 阶段二: 建立针对复杂资源与服务时的标准规范, 逐步建立支持互操作的公共支撑机制;

(3) 阶段三: 在适当的时候, 发展开放服务与集成的标准规范, 为跨领域的互操作、公共支持机制和可持续地发展和完善标准规范搭建平台。

2 描述对象内容: CDLS 开放元数据机制

设计描述对象内容的元数据格式是 CDLS 的主要努力方向之一。事实上, 目前存在着相当多的元数据格式, 而且任何时候、任何领域都有可能不断地有新的元数据格式产生。为适应我国开放的数字图书馆环境这一整体远景目标, 同时又兼顾元数据领域的动态发展, CDLS 项目采用开放元数据机制。

元数据格式已经从传统的单一、大一统趋向多样化、个性化。人们对元数据的要求也发生了变化, 比如重视满足互操作的需要甚于重视单独的功能, 强调可扩展性甚于强调完整性; 而且, 我们越来越清晰地认识到发展的趋势是分布式而不是集中式, 元数据格式不再象传统那样是独立而封闭的, 它需要为满足将来与其他系统的交换、交互要求做好准备。基于上述考虑, 我们采用了开放元数据机制^[8]。开放元数据机制遵循“模块化”、“可扩展性”、“分布性”、“递归性”四大原则^[9]。模块化原则是指每个元数据格式都只是一个这样的模块: 它只描述对象特定的某一方面, 仅支持对象特定的某项功能; 它可根据功能需求与其他模块互联从而为动态的扩展提供 plug-n-play 能力。可扩展性原则是指整个元数据体系结构和每个元数据模块都可通过复用、嵌入、链接、修改等方式对已有的元数进行扩展。分布性原则是指元数据的结构应当能支持远程的其他元数据系统嵌入, 从而能很容易地复用和组合出所需要的元数据。递归性原则是指元数据也是一种数据, 它本身也可有对应的元数据格式(或者为区别起见, 称为元元数据)来描述。而且, 为描述不同

层次的对象及对象之间的关系,我们需要应用不同层次的各类元数据,下文中我们还会详细阐述这一论点。开放元数据机制构建的元数据格式系统参见图2所示。从图2中,我们可以看到,有一个基本核心元数据集(Basic Core Metadata, BCM)作为最通用的互操作工具,它应该足够简单、足够通用,能支持信息对象的发现、标识及定位。在BCM之上,作为应用的一部分,有一系列标准的扩展规则与扩展机制(即图中的Standard Extension Methods),为BCM的应用进行详细约束与指导。这样就使得BCM能以一种开放的、标准的方式进行扩展并模块化。在BCM的基础上及在标准的扩展机制指导下,我们可以通过复用BCM中的元素、复用其他专门的元数据格式或定义新的元素,为某一大类下的所有类似的资源类型(比如古文献系列下的古籍、家谱等,论文系列或各类代理),设计出适合本大类下各子类型均通用的元数据格式,即图示中的Basic Special Metadata,简称BSM。(在CDLS子项目“专门数字对象描述元数据规范”中,BSM中不属于BCM的元素称为类型核心集,而BCM元素被称为基本核心集)。BSM也是以应用纲要(Application Profile Methods,如图最右边所示)的形式进行扩展的,它的扩展遵循应用纲要的规则与方法。

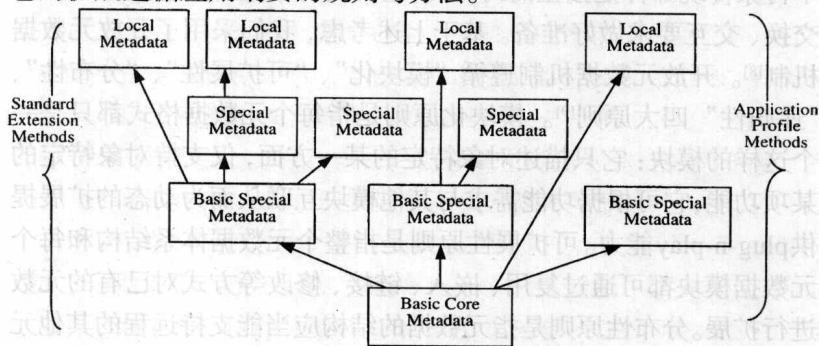


图2: CDLS 的开放元数据机制

在BSM的基础上,针对特殊的专门信息对象元数据格式可以同样的方式产生,比如古文献系列中的古籍、家谱、拓片等的元数

据格式。即：在BSM的基础上和扩展机制的约束下，复用BSM中的元素、复用其他现有的专门元数据格式中的元素或定义新的元素，形成描述专门信息对象的元数据格式，也就是图2中的Special Metadata，简称SM。最后，为满足揭示本地的资源特性或服务特性的需要，应用系统可定义自己的本地元素(可从BCM、BSM、SM及其他元数据格式中复用或自定义)。为确保模块化和可扩展性，在整个过程中，都需要采用应用纲要的方法和标准的扩展机制。

遵循上面的开放元数据机制，CDLS采用简单的DC^[10]作为它的BCM，目前已经为古文献与论文两个系列制定了BSM，为舆图、拓片、古籍、地方志、家谱、学位论文、期刊论文、会议论文、网络资源、电子图书、音频资料等十一种资源对象制定了SM，在项目中称为“专门数字对象描述元数据规范”，这十一种描述元数据规范目前已经在全国范围内进行开放试验。另外，CDLS也制定了发展BSM与SM的设计指南，目前文档状态为草案。

为实现上述元数据格式快速而顺利地得到应用，CDLS正在建设一个开放的应用环境，为每一种元数据格式的应用开发一系列的工具与指南，包括置标规则、编码体系、著录规则、输入模板、与其他元数据格式的映射等，这些工具目前还在开发与试验中，一俟完成，元数据格式的应用将有较好的一致性。此外，为支持上述元数据格式甚至是其他的元数据格式的应用、互操作和长期维护，DCMI元数据注册系统的中国镜像实验系统^[11]也已经建立。

3 将元数据的发展机制应用于开放、整合的数字图书馆服务中

从图1所示的数字图书馆生命周期模型中，我们可以发现在一个开放的数字图书馆环境中，仅仅为内容对象设计开放的描述元数据格式是远远不够的。事实上，资源、服务甚至系统本身的获取、互操作和整合需要各个层次上的元数据。这里我们借用Berners-Lee的语义万维网模型^[12]的思想，以图3来体现CDLS对开放数字图书馆环境下开放的标准规范框架的理解。