

数字图书馆标准 与规范建设

中国时代科技出版社

数字图书馆标准与规范建设

上 卷

《数字图书馆标准与规范建设》项目组 编写

中国时代科技出版社

图书在版编目(CIP)数据

数字图书馆标准与规范建设

《数字图书馆标准与规范建设》项目组 编写

—北京:中国时代科技出版社.2008.3

ISBN 978 - 7 - 5487 - 4638 - 5

I. 数… II. 数… III. 图书 IV. TQ582.15

中国版本图书馆(CIP)数据核字(2008)第 10089 号

数字图书馆标准与规范建设

ISBN 978 - 7 - 5487 - 4638 - 5

中国时代科技出版社出版发行

北京市朝阳印刷厂 新华书店发行经销

2008 年 3 月第 1 版 2008 年 3 月北京第 1 次印刷

开本:787 × 1092 1/16 印张:70.5

字数:1380 千字 定价:498.00 元(精装二册)

编委会成员名单

《数字图书馆标准与规范建设》项目组 编写

主 编 许 松
副主编 付小琼 曹 霞
编 委 (以姓氏笔画排列)

丁先敏	丁 锐	王明宁	王宏北	王 涛
马 光	叶 川	叶光毅	付小琼	田风德
刘建琼	刘云涛	刘建波	江德风	汪冬才
许 松	冯希毅	沙 辰	汪明辉	曹 霞
袁 宁	袁 成	郭宏德	郭光英	梁敏建
廖东雄	熊 敏	穆 霞	穆中春	

目 录

第一篇 数字图书馆标准规范	1
数字图书馆标准规范的发展趋势	3
数字图书馆资源与服务集成标准规范研究现状	61
第二篇 服务标准和协议应用	85
LDAP WHOIS++在数字图书馆中的应用案例及其分析	87
WEB 服务标准和协议应用案例分析及应用指南	109
第三篇 数字资源标准规范	229
《数字资源加工标准》项目报告	231
数字资源唯一标识符的现状与发展	271
第四篇 元数据标准规范	321
专门数字对象元数据项目元数据著录实验系统	323
CDLS 元数据名称空间方案	351
OAI 元数据获取协议综述报告	365
第五篇 基本元数据标准规范	387
基本元数据标准应用规范	389
基本元数据著录规则	407
第六篇 专门元数据标准规范	453
国内专门数字资源对象元数据实例列表分析报告	455
专门数字资源对象的格式标准推荐分析报告	529
第七篇 古籍元数据规范及其著录规则	567
古籍元数据规范	569
古籍元数据著录规则	605
第八篇 论文元数据规范及其著录规则	655
学位论文元数据规范	657
学位论文描述元数据著录规则	683
第九篇 电子图书	733
电子图书元数据著录规则	735
电子图书元数据规范	771
第十篇 网络资源	811
网络资源元数据规范	813
网络资源元数据著录规则	859
第十一篇 音频资料	909
音频资料描述元数据规范	911
音频资料元数据著录规则	951

第一篇

数字图书馆标准规范

数字图书馆标准规范 的发展趋势

科技部科技基础性工作专项资金重大项目 研究成果

项目名称：我国数字图书馆标准规范建设

子项目名称：我国数字图书馆标准规范发展战略与基本框架

项目编号：2002DEA20018

研究成果类型：研究报告

成果名称：数字图书馆标准规范的发展趋势

成果编号：CDLS-S01-001

成果版本：总项目组推荐稿

成果提交日期：2003 年 3 月

撰写人：张晓林、鞠文红、李珍（中国科学院文献情报中心）

项目版权声明

本报告研究工作属于科技部科技基础性工作专项资金重大项目《我国数字图书馆标准规范建设》的一部分，得到科技部科技基础性工作专项资金资助，项目编号为 2002DEA20018。按照有关规定，国家和《我国数字图书馆标准规范建设》课题组拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

本报告可以复制、转载、或在电子信息系统上做镜像，但在复制、转载或镜像时须注明真实作者和完整出处，并在明显地方标明“科技部科技基础性工作专项资金重大项目《我国数字图书馆标准规范建设》资助”的字样。

报告版权人不承担用户在使用本作品内容时可能造成的任何实际或预计的损失。

作者声明

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性和可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。科技部有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

《我国数字图书馆标准规范建设》课题组网站 (<http://cdls.nstl.gov.cn>) 作为本报告的第一发表单位，并可向其他媒体推荐此作品。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

数字图书馆标准规范的发展趋势

目 录

1. 总论

- 1.1 数字图书馆标准规范的意义
- 1.2 数字图书馆标准规范的建设
- 1.3 数字图书馆标准规范的范围与结构

2. 关于数字内容创建的标准规范

- 2.1 内容编码
- 2.2 数据格式
- 2.3 内容标识
- 2.4 复合数字对象的建立与封装

3. 关于数字对象描述(元数据)的标准规范

- 3.1 元数据发展现状
- 3.2 开放元数据机制
- 3.3 元数据选择与应用原则
- 3.4 关于内容主题描述语言的选择

4. 关于资源组织描述的标准规范

- 4.1 资源组织描述的发展与要求
- 4.2 对资源集合的描述要求
- 4.3 资源集合知识组织体系、管理机制和资源组织过程规范

5. 关于数字资源系统服务的标准规范

- 5.1 关于接入条件、传输条件和数据应用条件的标准规范
- 5.2 关于检索条件的标准规范
- 5.3 关于信息服务业务过程的标准规范
- 5.4 分布数字服务机制的标准规范

6. 关于数字资源长期保护的标准规范

- 6.1 数字资源长期保存的基本框架
- 6.2 开放存档系统参考模型
- 6.3 数字资源长期保存的元数据

7. 建立标准规范描述体系的目标与任务

- 7.1 描述体系的目标
- 7.2 描述体系的任务
- 7.3 描述体系的开放性质

参考文献

附录 1: 英文缩略语

附录 2: 电子政务互操作框架 (E-GIF)

1. 总论

1.1 数字图书馆标准规范的意义

数字图书馆作为基于网络环境提供数字信息资源和服务的系统机制,需要建立和遵循关于数字化加工、资源描述、资源组织、资源互操作和资源服务等方面的标准和规范,需要采用和遵循内容编码、数据通信、计算机系统、安全、管理、知识产权、服务运营等方面的标准和规范,才能保证所建立的资源和服务的可使用性、互操作性和可持续性。可使用性,指所建立的资源或服务能够在广泛的网络环境和复杂的技术条件下被用户方便地使用。互操作性,指所建立的资源或服务能够在更大系统范围上、能与其它资源或服务方便、有效的交换、转换、整合,从而为用户提供逻辑上集成的服务。可持续性,指所建立的资源或服务能够在变化的技术与运行机制下长期保存和使用,能够被集成入未来的资源与服务环境。

随着数字信息资源的迅速发展和网络化信息服务的不断丰富,数字图书馆系统正走向开放、集成、并与其他数字信息系统灵活融合的发展阶段,任何一个数字图书馆都将作为整个数字信息环境中的一个模块,按照整体环境的标准规范来组织资源、提供服务,才能保证自己系统的资源和服务的可使用性、互操作性和可持续性,也才能有效利用整体环境中的其它资源与服务来提高对自己用户的服务能力和效率。可以说,在数字化网络化环境下,任何孤立、封闭的数字图书馆系统都将失去生存和发展的能力,而实现开放、互操作和集成的基础则是标准规范。

1.2 数字图书馆标准规范的建设

数字图书馆作为网络形态下的信息资源与服务系统,一直对标准规范建设比较重视。在美国,由于数字图书馆建设由多种机构在多个领域开展,因此不同机构都积极开展了数据加工格式和描述元数据等方面的研究与应用,许多机构或项目都规定了数据加工格式的标准,图书馆、博物馆、教育、档案、政府机构等都制定了自己的元数据格式,图书馆界和数字图书馆建设领域还提出了若干检索服务协议。而且,随着数字图书馆建设的规模化和逐步成熟,合作研究建立标准规范逐步成为主流,而且元数据互操作标准的研究明显加大了力度。

吸取美国数字图书馆建设早期的自适应性和分散性方式所带来的问题,英、加等国以及美国在数字图书馆建设中更为注意系统地 and 集中地建设相关标准与规范,逐步以标准规范体系(Frameworks)、指南(Guidelines)、最佳实践建议(Best Practices)和应用协议(Application Profiles)等方式提出了系统化的数字资源建设标准规范要求,对数字信息资源建设和服务所涉及的数字化加工、资源描述、资源组织、资源互操作和资源服务等方面的标准、规范及其应用要求进行系统描述,

许多国家性、地区性或行业性数字信息系统都提出了标准规范应用指南,全面详细地规定数字信息资源创建、描述、组织、服务、长期保存等方面应采取的标准与规范,其中许多规定具有法律或行政效力。进一步地,有些大规模建设项目还对描述数字信息内容和集合的元数据格式及其扩展方式进行了规定。有些数字图书馆项目还对检索和互操作标准进行了描述。还有些数字图书馆建设体系面向更大环境,对整个数字信息服务涉及的通信、系统、资源、安全、管理、知识产权、服务、运营等多方面的标准与规范进行系统描述,这些标准规范体系为数字图书馆高效率建设、共享、可持续发展奠定了坚实基础。

从具体的标准规范看,越来越多的数字图书馆标准规范不断被提出、完善和应用,许多已经被广泛地确认和接受,有些已经或正在成为国家或国际标准,相关应用技术也在逐步完善。例如,多数数字图书馆体系确认和接受主流(尤其是开放)图像数据格式为图像数据数字化的标准格式,接受开放的标记语言格式及纯 ASCII 文本为文字内容的格式标准;在资源描述方面,若干元数据格式等已被广泛接受,而各个专门领域也逐步确认了自己的通用或事实描述标准;在资源检索和调度方面,基于 URI 的唯一标识符机制已被普遍接受,开放检索或搜索协议也逐步得到普遍应用,数字资源长期保存 OAIS 模型成为资源保护元数据设计和长期保存系统建设的事实标准。

在此基础上,发达国家数字图书馆界充分利用WEB网络标准规范和电子商务技术的发展成果,在承认数字图书馆建设与应用将是基于分布、异构和自主的系统环境的同时,在支持分布环境下多主体数字图书馆自主建设、支持和容纳已有的多元化标准规范成果的同时,通过开放语言、标准链接与扩展机制、公共登记系统等方法,积极探索基于XML/RDF^[31]的对标准规范的开放描述、互操作和共享机制,从而支持标准规范的发现、解析、转换、复用,便于有效利用由众多过去(以及未来)的分布、异构和自主系统构成的复杂数字图书馆环境。这种努力为人们灵活、经济地建设和利用标准规范提出了有效机制,为数字图书馆在技术上可持续和应用上可扩展、为用户综合利用包括数字图书馆在内的多样化信息系统打下基础。

从数字图书馆标准规范建设方式来看,最初由单个数字图书馆分散进行标准规范研究与建设的状态,已经基本转变由多个数字图书馆(甚至包括其它领域的机构与专家)联合起来,以开放、合作、共享的方式共同开展研究,使得研究成果更具备开放性、可靠性和广泛的应用性,也促进了数字图书馆标准规范建设过程的工程化、规范化和有效的合作交流。多数与数字图书馆相关的标准规范制定机制,例如W3C(World Wide Web Consortia)^[1]、IETF(Internet Engineering Task force)^[2]、DCMI(Dublin Core Metadata Initiative)^[3]、OAI(Open Archive Initiative)^[4]、Z39.50 Implementation Group^[5]等,都由多方专家组成工作组,采取全程开放过程

来随时吸收各方意见,按照规范程序进行标准规范的起草、制定、试验、推荐、审批等工作,通过专门维护网站支持广泛的发布、合作与交流,保证标准规范制定过程的开放性和标准规范成果的开放性,推动标准规范的尽早、广泛和公共应用。

1.3 数字图书馆标准规范的范围与结构

数字图书馆领域本身处于不断发展的过程中,而且数字图书馆与其他数字信息系统(例如数字科研、远程教育、电子政务、电子商务、数字博物馆等)不断相互融合,因此数字图书馆的相关标准规范也在不断发展之中。不过,从数字图书馆建设和服务应用的角度,人们可以从至少三个不同角度来界定数字图书馆标准规范的范围与结构。

(1) 数字资源加工与描述标准体系

数字图书馆发展初期,数字资源本身的建立和描述成为数字图书馆的核心任务,建立数字资源加工、标识和描述的标准规范成为当务之急,因此许多数字图书馆系统都将数字资源加工与描述标准体系作为自己的标准规范建设任务,以保障自己所建立或引进的数字资源能被有效地标识、描述、组织和处理。当然这类标准规范体系所涉及的数字资源已经超过传统图书情报领域所熟悉的文献资料。

这类体系的典型代表包括美国国会图书馆数字资源格式规范体系^[6]、美国国会图书馆数字资源检索与互操作规范体系^[7]、OhioLink多媒体资源标准体系^[8]、加州数字图书馆数据格式^[9]、加州数字图书馆元数据与编码标准^[10]、美国NSDL元数据标准体系^[11]、新西兰国家图书馆元数据标准框架^[12]、UN/FAO农业信息资源检索元数据框架^[13]、CEN/ISSS元数据体系^[14]、INDECS元数据框架模型^[15]、英国电子政务元数据框架(E-Government Metadata Framework, e-GMF)^[16]、加拿大政府信息元数据框架(Government Online Metadata Standard, GOMS)^[17]等。

这类标准体系的内容主要涉及:数字文件格式标准、数字文件标识标准、数字对象描述性元数据模式框架和具体描述格式、元数据内容描述体系(例如在描述过程中所使用的主题词表或分类法)。随着数字图书馆的进一步发展,更为广泛和系统化的标准规范体系被确立,这些基于资源加工和元数据描述的标准体系正逐步被融入新的上层体系中。

(2) 数字资源生命周期标准框架

数字图书馆依赖数字信息资源来提供服务,但仅仅局限于数字资源加工和描述并不能完整地支持数字资源的服务;尤其在分布、异构和自主的数字资源不断涌现、数字图书馆越来越依赖开放环境来集成各种分布的数字资源的时候,需要有更广泛更系统的视角来认识和组织数字图书馆标准规范。因此,人们提出了数字资源生命周期思想,围绕数字资源创建、描述、组织、检索、服务和长期保存的

整个生命周期来规划、设计、组织标准规范,一方面针对图书情报领域的数字图书馆资源建设提供系统化基础性的指导和约束,另一方面保证各类数字资源的广泛可使用性,保障分布资源和服务的互操作,促进数字资源的长期保存和可持续利用。

这类标准规范框架的典型代表包括:英国公共图书馆领域的NOF/People's Network项目标准与指南^[18];英国分布国家电子资源项目(DNER)标准体系^[19];加拿大文化在线项目(CCOP)标准与指南^[20];美国IMLS数字资源建设指南框架^[21];美国RLG/CMI数字化指南^[22]。当然,有些系统在建立自己的数字图书馆标准规范体系时还进一步涉及了数字资源的选择、版权、使用政策、以及系统管理等方面的内容,例如英国DNER系统互联指南^[23]、美国亚利桑那州数字化项目指南^[24]和美国科罗拉多州数字化项目(CDP)指南^[25]。美国国家信息标准局(NISO)最近发表的图书馆系统标准指南也属于这类框架^[26]。

这类框架一般都将数字资源生命周期所涉及的标准规范分为多个层次,形成标准规范的层次结构。例如,NOF/People's Network按照生命周期分为数字对象生产(Creation)、管理(Management)、资源建设(Collection Development)、使用(Access)和复用(Re-use)5个层次;IMLS从数字资源建设角度分为资源集合(Collections)、资源对象(Objects)、元数据(Metadata)和资源建设项目(Projects)4个层次;CCOP分为内容生产(Content Creation)、编目与元数据(Cataloguing and Metadata)、词汇与词表(Terminology and Controlled Vocabularies)、数据库结构(Database Structure)、项目网站(Project Web Site)、长期保存与记录管理(Preservation and Records Management)等6个方面;CDP分为数字资源加工(Scanning/Digital Audio)、元数据(Metadata)、法律问题(Legal Issues)、资源政策(Collection Policies)、项目建设(Projects)等方面。

按照数字资源生命周期来规划和建设数字图书馆标准规范,便于更为系统地认识和组织从数字资源创建到长期保存的整个进程中的各种标准规范,促进这些标准规范的相互支撑和互操作,从而保障数字资源以及建立在数字资源上的服务(而不仅仅是元数据或某种特定的裸数据)在网络环境和整个生命周期中的可使用性。同时,这种认识和组织数字图书馆标准规范的做法并不是将诸如网络通信、计算机系统与数据库、安全、知识产权等更基础的技术或政策标准排斥在数字图书馆标准规范体系中,而是承认这些技术或政策标准主要由非图书情报领域的有关机构制定,在数字图书馆建设中对这些标准的选择与应用主要是由系统平台提供方、内容资源提供方和数字图书馆建设方协调进行。

(3) 数字信息系统互操作标准体系

实际上,网络环境下数字信息系统的可使用性、可互操作性和可持续性不仅涉及信息资源的组织、描述、利用和保存,还必然涉及到网络通信、计算机系统与